

A Lumry–Eyring Nucleated Polymerization Model of Protein Aggregation Kinetics:

1. Aggregation with Pre-Equilibrated Unfolding

Jennifer M. Andrews and Christopher J. Roberts*

Department of Chemical Engineering, 150 Academy Street, Colburn Laboratory, University of Delaware, Newark, Delaware 19716

Received: January 10, 2007; In Final Form: April 2, 2007

A mathematical model is presented of the kinetics of non-native protein aggregation that combines Lumry–Eyring and nucleated polymerization (LENP) descriptions. The LENP model is solved for cases in which aggregation rates are slow compared to folding–unfolding equilibration and is shown to be a generalization of a number of previously proposed nucleation-and-growth models for non-native and native protein aggregation. The model solutions exhibit a number of qualitative kinetic regimes. Each regime has a characteristic set of experimental signatures that are related to the relative rates of growth and nucleation as well as to the threshold size at which aggregates condense to form higher-order structures or other phases. Approximate model solutions provide practical rate equations that can be regressed against typical experimental kinetic data to obtain mechanistic parameters characterizing the aggregation pathway. In all kinetic regimes, it is found that observed rate coefficients (k_{obs}) or half-lives (t_{50}) obtained from extent-of-reaction measurements are convolutions of more than one stage in the pathway unless purely seeded growth occurs. Despite this convolution, the combination of apparent reaction order (time domain) and the scaling of k_{obs} or t_{50} with initial protein concentration provides a means to determine a value for the dominant nucleus size in each case. Additional information, such as equilibrium unfolding thermodynamics and the limiting aggregate size distribution, are required to further deconvolute k_{obs} into intrinsic contributions from nucleation, growth, and conformational changes. The model and analysis are expected to be generally applicable to a wide range of proteins and polypeptides that form non-native aggregates.

1. Introduction

Non-native aggregation is one of the primary routes by which unwanted and potentially harmful degradants and impurities form during the manufacture, transport, and storage of biopharmaceutical products.^{1–8} The intermediates or the end products in non-native aggregation are also increasingly implicated as harmful species in the progression of numerous chronic diseases.^{9–14} There also exist situations in which non-native aggregation is desirable.^{15,16}

In general, non-native protein aggregation refers to the process by which originally monomeric or unaggregated proteins or polypeptides spontaneously self-assemble to form intermediate- to high-molecular-weight aggregates that have elevated levels of non-native secondary structure. These non-native structures are often, but not always, rich in intra- and inter-protein β -sheets.^{11,17–19} Although significant β -sheet content may be present in the native or non-native monomeric states for some aggregation-prone proteins, there are also examples where β -sheet formation is not apparent until small to intermediate size aggregates form. Therefore, it remains an open question as to what type and degree of structural perturbation must occur to the native state prior to early stage association to facilitate the later stages of the process.^{13,16–18,20–28} These observations, together with kinetic signatures such as higher-order concentration dependences^{29–33} and in some cases apparent lag phases that are seeding-dependent,^{34–38} suggest that non-

native aggregation commonly occurs through (incomplete) monomer unfolding combined with assembly via nucleated growth,^{17,18,28,39–41} in which the rate-limiting step may or may not be nucleation.^{33,42,43}

In many practical situations, the overall process of aggregation is under kinetic control.^{31,42–46} The observed kinetics may, however, be controlled in part by the thermodynamics of those steps or stages that precede the rate-determining step(s).^{33,45–49} In addition, because a number of intermediates in the process are often transient and/or thermodynamically disfavored, they are fundamentally difficult to directly characterize via experiment. Together, these considerations help to motivate the need for accurate yet tractable mathematical models of non-native aggregation kinetics. When combined with experimental kinetic and thermodynamic data, such models can provide a relatively unique and noninvasive means to gain qualitative and quantitative details of the aggregation mechanism^{30–33,44–46} and can also aid in the design of experiments and additives to more accurately predict or control aggregation rates.^{46,50,51}

A number of mathematical models for protein aggregation kinetics have been previously proposed and/or analyzed. Many of these fall into one or more of the following categories depending on the level of detail or approximation in the model solution: (1) Lumry–Eyring models, in which folding–unfolding transitions are treated in detail, with simplified or qualitative treatment of assembly steps;^{21,43,47–49} (2) native polymerization models, in which nucleation and polymer growth are treated in detail, but conformational changes prior to or during assembly are not included;^{30,31,44,52–54} (3) extended

* Author to whom correspondence should be addressed. Phone: 302-831-0838. E-mail: cjr@udel.edu.

TABLE 1: Key Symbols

name	definition	section	name	definition	section
A_j	aggregate composed of j monomers ^a	2.6	N	native state (monomer) ^a	1
A_x	aggregate nucleus ^a	2.3	n^*	size at which condensation steps begin	2.6
A_jR	reversibly associated A_j and R ^a	2	R	reactive monomer ^a	1
a_j	$[A_j]/C_0$	2.7	R_i	reversible oligomer of i monomers ^a	2.2
C_0	total protein concentration at $t = 0$ on a monomeric basis ^a	2.7	R_x	reversible aggregate prenucleus ^a	2.3
C_{ref}	reference protein concentration ^a	2.7	SD_{agg}	aggregate size distribution variance	3.3
C_M	total uncommitted monomer concentration ^a	2.7	T	absolute temperature ^b	2.1
c_m	C_M/C_0	2.7	t	time ⁱ	2.7
f_R	reactive fraction of monomers	2.1	t_{50}	half-life for monomer loss ⁱ	3
I	intermediate state (monomer) ^a	2.1	U	unfolded state ^a	1
K_{IU}	equilibrium constant for $I \leftrightarrow U$	2.1	x	nucleus size	2.3
K_i	equilibrium constant ^b for $iR \leftrightarrow R_i$	2.2	β_{gn}	τ_n/τ_g	2.7
K_{NI}	equilibrium constant for $N \leftrightarrow I$	2.1	$\Delta G_{\text{IU}}^\circ$	standard Gibbs free energy change j for $I \leftrightarrow U$	2.1
K_{RA}	equilibrium constant ^c for $A_j + R \leftrightarrow A_jR$	2.7	ΔG_i°	standard Gibbs free energy change j for $iR \leftrightarrow R_i$	2.2
k_a	association rate coefficient ^d	2	$\Delta G_{\text{NI}}^\circ$	standard Gibbs free energy change j for $N \leftrightarrow I$	2.1
$k_{a,x}$	k_a for nucleation step ^d	2.3	δ	number of monomers in each growth event	2.4
k_B	Boltzmann's constant ^e	2.1	Φ	m/c_m	2.7
k_d	dissociation rate coefficient ^f	2	$\lambda^{(1)}$	first moment of soluble aggregate distribution	3.1
$k_{d,x}$	k_d for nucleation step ^f	2.3	θ	dimensionless time	2.7
k_g	growth rate coefficient ^d	2.7	σ	$\sum [A_j]/C_0$	2.7
k_{nuc}	nucleation rate coefficient ^f	2.7	σ_0	σ value at $t = 0$	3.5
k_{obs}	observed aggregation rate coefficient ^f	3	σ_{ss}	σ value at steady state	3.1
k_r	rearrangement rate coefficient ^f	2	σ_{int}	σ_{ss} value at $\log \delta\beta_{\text{gn}} = 0$	3.2
$k_{r,x}$	k_r for nucleation step ^f	2.3	τ_g	characteristic growth time scale	2.7
MW_{mon}	monomer MW ^g	3.1	τ_n	characteristic nucleation time scale	2.7
$\langle MW \rangle_w$	weight-averaged molecular weight	3.1	$\tau_g^{(0)}$	τ_g at $[R] = C_{\text{ref}}$	2.7
m	$([N] + [I] + [U])/C_0$	2.7	$\tau_n^{(0)}$	τ_n at $[R] = C_{\text{ref}}$	2.7

^a [mol/volume]. ^b [(mol/volume)⁻¹]. ^c [(mol/volume)⁻¹]. ^d [(mol/volume)⁻¹ time⁻¹]. ^e [energy/kelvin]. ^f [time⁻¹]. ^g [mass mol⁻¹]. ^h [kelvin]. ⁱ [time]. ^j [energy/mol].

Lumry–Eyring models, in which both folding–unfolding and assembly are treated in detail to varying degrees depending on the nature of the experimental data and conditions considered;^{32,33,45,55–57} (4) aggregate condensation and polymerization models that primarily or exclusively treat higher-order assembly of the aggregates that form during the earlier stages of nucleation and/or polymerization;^{58–60} (5) statistical/structural models based on multivariate analysis of relative rates of fibril or filament growth across large databases of polypeptide variants with differing primary sequences but little or no secondary or tertiary structure.^{61–63}

Many of the models just cited^{30,31,33,44,45,54,58} are limiting cases of the Lumry–Eyring nucleated polymerization (LENP) model developed and analyzed in this report or are mathematically equivalent to the LENP model with the appropriate parameter or variable transformations. Some of these instances are highlighted as part of the model derivations or analysis in sections 2 and 3, although not all cases are described in detail for reasons of space constraints. The derivations and model solutions described in the remainder of this report therefore provide a relatively comprehensive and general description of a large number of aggregation-prone protein and polypeptide systems within a common framework. Our analysis of the model behavior provides a more global view of the range of experimentally observable kinetic behaviors captured by the LENP model than has been previously available. It also provides the only systematic investigation of the effects of aggregate insolubility beyond the case considered previously by one of us.⁴⁵ Finally, it generalizes previous work^{33,45} providing quantitative relationships between k_{obs} or t_{50} and experimental quantities such as initial protein concentration (C_0) and unfolding free energy, and extends the analysis to include aggregate size distributions and their quantitative relationship to nucleation and growth rates. As such, it presents a more comprehensive, quantitative description of non-native aggregation kinetics than has previously been available.

2. Model Description and Derivations

Table 1 provides a brief definition of key symbols used throughout the report and the section number in which each symbol is first defined. Figure 1 provides a schematic depiction of non-native aggregation in terms of a minimal set of five stages needed to qualitatively capture experimental observables such as outlined in section 1 as well as to remain sufficiently general to self-consistently capture many of the models referred to above. Double arrows indicate reversible steps, while single arrows indicate effectively irreversible steps. The stages are described in detail in subsections 2.1–2.5 and are translated into a reaction network with corresponding parameters (Figure 2).

Figure 2 and the corresponding model equations are for the case of a three-state monomeric protein in which the native state (N) undergoes a cooperative, all-or-none conformational transition to form an intermediate (I) that undergoes a further conformational transition to result in the unfolded state (U). The monomer conformation that is most prone or reactive with respect to aggregation is denoted as R and may be any of N, I, or U. By appropriately defining one of the model parameters (f_R ; subsection 2.1 and ref 33), the derivations below hold for $R = U, I$, or N , as well as for multistate proteins that unfold through one or more intermediates but do not dissociate upon unfolding.

2.1 Conformational Transitions of Unaggregated Protein (Stage 1). The first stage in Figure 1 includes all transitions between conformational states that are thermodynamically distinct and are therefore separated by sufficiently high free energy barriers that mass-action kinetic rate coefficients can be reasonably defined for each transition step.^{64,65} Figures 1 and 2 depict a scenario in which a true intermediate state exists between the unfolded and native states. The simpler, two-state case is recovered by setting $[U]$ and K_{IU} to zero.

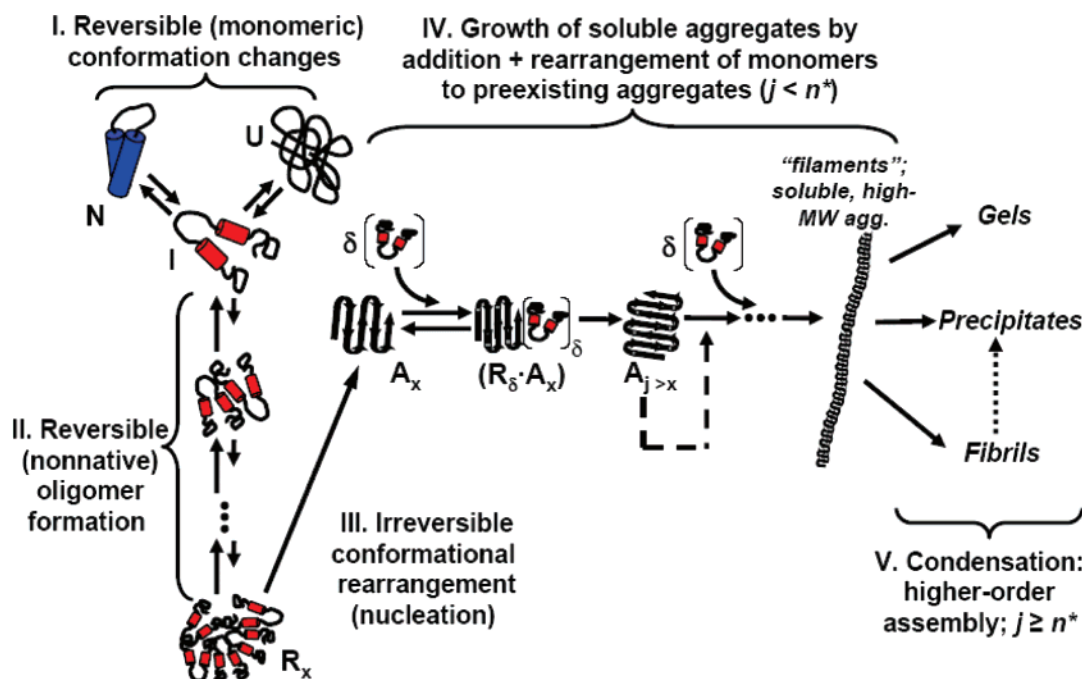


Figure 1. Diagrammatic depiction of the protein aggregation process. Non-native, incompletely unfolded monomers are shown as reactive intermediates for concreteness, although the model permits any monomeric conformational state to be the most prone to aggregation. The five stages noted in the diagram are described in detail in the text.

The reversible unfolding transitions in stage 1 of Figure 2 have equilibrium constants

$$K_{\text{NI}} = \exp\left(-\frac{\Delta G_{\text{NI}}^{\circ}}{k_{\text{B}}T}\right) = \frac{[\text{I}]}{[\text{N}]}_{\text{eq}} \quad (1\text{a})$$

$$K_{\text{IU}} = \exp\left(-\frac{\Delta G_{\text{IU}}^{\circ}}{k_{\text{B}}T}\right) = \frac{[\text{U}]}{[\text{I}]}_{\text{eq}} \quad (1\text{b})$$

In eq 1, the subscript eq indicates the quantity in parentheses is evaluated at conformational equilibrium; all other symbols are defined in Table 1. The model assumes that only one of the conformational states or subensembles is predominantly “reactive” with respect to non-native aggregate formation. The fraction of monomer that is reactive is given by

$$f_{\text{R}} = \frac{[\text{R}]}{[\text{N}] + [\text{I}] + [\text{U}]} \quad (2)$$

Equation 2 is easily generalized to cases with multiple intermediates by including their concentrations in the sum in the denominator. Under conditions where the dynamics of monomer conformational transitions are rapid compared to the rate-limiting step(s) in aggregation, f_{R} is determined solely by the relative free energies of N, I, U, etc. This is the assumption used in the remainder of this report; i.e., all monomer conformational transitions are considered pre-equilibrated. For the case of two-state unfolding with $K = \exp(-\Delta G_{\text{NU}}^{\circ}/k_{\text{B}}T)$, f_{R} is $K/(1 + K)$ if $\text{R} = \text{U}$, or $1/(1 + K)$ if $\text{R} = \text{N}$.^{33,45}

2.2. Prenucleation (Stage 2). Stage 2 in Figures 1 and 2 depicts the reversible association of R chains up to some threshold or prenucleus size, x (where x is the number of chains in the largest, reversible prenucleus). Typical values assumed or experimentally inferred for x range from 2 to 10 or more.^{29–33,39,40} We treat each oligomer of R as a thermodynamically distinct species with a definable equilibrium constant (K_i)

for the conversion of i monomers of R to an oligomer of size i (R_i , $i = 2, 3, \dots, x, \dots$)

$$K_i = \exp\left(-\frac{\Delta G_i^{\circ}}{k_{\text{B}}T}\right) = \frac{[\text{R}_i]}{[\text{R}]^i}_{\text{eq}} \quad (3)$$

Table 1 provides symbol definitions. We assume that the formation of R_i ($i < x$) from R is reversible and rapid compared to later steps and therefore is pre-equilibrated.

2.3. Nucleation of the Smallest Irreversible Aggregates (Stage 3). Stage 3 in Figure 2 is the initiation or nucleation step that essentially commits the prenucleus R_x to a net irreversible aggregate A_x . This step in the aggregation process is arguably the least well understood and the most difficult to characterize experimentally. For the sake of generality we permit the possibility that a rearrangement step ($R_x \rightarrow A_x$) is needed to allow the monomers in R_x to change conformation or realign to create the secondary structure and/or interchain contacts needed to stabilize the nucleus. Treating the rearrangement process as a single irreversible step with a well-defined rate coefficient ($k_{\text{r},x}$) implicitly assumes either that the process results in a significant stabilization of the x -mer or that subsequent growth is so rapid as to preclude A_x significantly reversing to R_x .

In the LENP model, the nucleus (R_x) is composed of x monomers. This is distinct from most previous models^{31,44,45,53,54,58} that assume the forward association step $R_{x-1} + R \rightarrow R_x$ is rate-limiting and therefore use $x - 1$ as the nucleus size. The association-limited case is obtained in the LENP model when $k_{\text{r},x} \gg k_{\text{d},x}$ (subsection 2.7). The difference between these two definitions of nucleus size becomes important only as x approaches 2.^{39,40}

2.4. Growth of Soluble Aggregates via Polymerization (Stage 4). Stage 4 in Figures 1 and 2 is the growth of aggregates by recruitment of one or more R monomers. The number of monomers per growth step is denoted δ . Stage 4 includes reversible association of δ monomers of R with preformed,

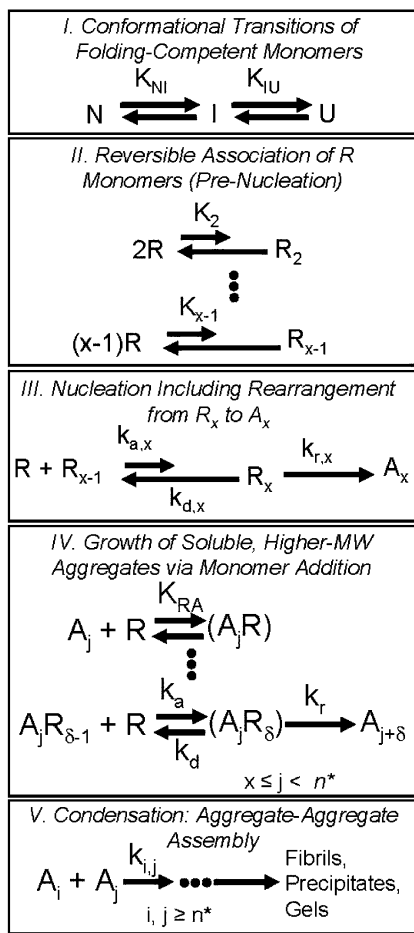


Figure 2. Reaction scheme with associated model parameters for the five key stages in the LENP model solved here. The steps shown in each panel are treated as elementary, reversible (double arrow) or irreversible (single arrow) steps in the reaction pathway when translating them to mass-action kinetic equations (additional details in section 2).

soluble aggregates (A_j , $x \leq j < n^*$), with a subsequent rearrangement or secondary structure change to commit to the growing aggregate. Similar to stage 3, the rearrangement step is treated as irreversible and does not assume either association or rearrangement to be rate-limiting for growth. Rate coefficients and association constants describing the growth stage are independent of the size of A_j for $j < n^*$. This approximation is expected to hold well for aggregates that have a spatial dimensionality near one (e.g., linear polymers, filaments, and protofibrils).^{31,44} Treating the rearrangement steps in stages 3 and 4 as irreversible is anticipated to be valid when describing aggregation kinetics at fixed pH, temperature, and solvent composition, as well as under conditions where the resulting aggregate size distribution remains narrow.^{31,66}

2.5. Condensation: Aggregate–Aggregate Assembly (Stage 5). Stage 5 includes any condensation or aggregate–aggregate assembly processes, such as bundling of protofibrils or filaments (i.e., each A_j species in stages 3 and 4 may represent a filament or protofibril) and aggregate agglomeration leading to noncrystalline precipitation, gelation, or other phase separation. The current LENP model treats stage 5 by assuming that condensation steps do not occur with appreciable rates for soluble aggregates below a threshold size n^* and that aggregates with $j \geq n^*$ are no longer available to consume monomer. Thus, stage 5 is treated only indirectly with respect to monomer loss and is primarily neglected with respect to aggregate size

distributions. Appendix A provides a version of the LENP model that explicitly treats condensation kinetics.

2.6. General Material Balances. For the reaction scheme in Figures 1 and 2 we define the total uncommitted or assayable monomer concentration in solution (C_M) on a number concentration basis as

$$C_M = [N] + [I] + [U] + \sum_{i \geq 2} i[R_i] + \sum_{i=1}^{\delta} i \sum_{j=x}^{n^*-1} [A_j R_i] \quad (4a)$$

With eq 3 and the assumption of pre-equilibrated R_i formation, eq 4a can also be written as

$$C_M = [N] + [I] + [U] + x[R_x] + \sum_{\substack{i \geq 2 \\ i \neq x}} i[R_i] K_i + \sum_{i=1}^{\delta} i \sum_{j=x}^{n^*-1} [A_j R_i] \quad (4b)$$

Equation 4 is valid for any of the following conditions: (1) assay by gel permeation chromatography and polyacrylamide gel electrophoresis under native-favoring conditions such that R_i and R_δ in $A_j R_\delta$ species are primarily reversed to N in the assay itself; (2) assay by spectroscopies, such as far-UV circular dichroism, ThT fluorescence, and infrared, that primarily detect secondary structure changes upon conversion from R_x to A_x , and $A_j R_\delta$ to $A_{j+\delta}$; or (3) R_i and $A_j R_\delta$ species are relatively unstable and therefore give an experimentally undetectable contribution to C_M . Any species with a number concentration that can be accurately quantified in the selected assay does not need to be included in eq 4.

We define dimensionless concentrations by scaling by the initial monomer concentration (C_0) such that $c_m \equiv C_M/C_0$, $a_j \equiv [A_j]/C_0$, $m \equiv ([N] + [I] + [U])/C_0$, and $\sigma \equiv \sum_{j=x}^{n^*-1} a_j$. Square brackets denote number concentrations. Effective rate coefficients for nucleation and growth, respectively, are $k_{\text{nuc}} \equiv k_{a,x} k_{r,x} / (k_{d,x} + k_{r,x})$ and $k_g \equiv k_a k_r / (k_d + k_r)$. We define Φ as the fraction of C_M that exists as free monomer, i.e., $\Phi = ([N] + [I] + [U])/C_M$. In general, the value of Φ depends on the model parameters describing reversible association in stages 2–4 as well as the values of f_R , C_0 , c_m , and each a_j . It is given by the real, positive root of the x th-order polynomial

$$x \left[f_R^x (c_m)^{x-1} \frac{K_{x-1} k_{\text{nuc}} C_0^{x-1}}{k_{r,x}} \right] \Phi^x + \sum_{i=2}^{x-1} i [f_R^i (c_m)^{i-1} K_i C_0^{i-1}] \Phi^i + \delta \left(\frac{k_g K_{\text{RA}}^{-1} C_0^\delta}{k_r} f_R^\delta (c_m)^{\delta-1} \sigma \right) \Phi^\delta + \sum_{i=1}^{\delta-1} [K_{\text{RA}}^{i-1} C_0^i f_R^i \sigma (c_m)^{i-1}] \Phi^i + \Phi - 1 = 0 \quad (5)$$

where Φ implicitly depends on time because the coefficients in eq 5 are functions of c_m .

Equations 6–9 give the dimensionless, dynamic material balances that follow from balancing all forward and reverse terms for each of the species indicated in eq 4b and assuming local steady-state values for $[R_x]$ and each $[A_j R]$. Additional details are provided in Appendix B.

$$\frac{dc_m}{d\theta_n} = -x \Phi^x c_m^x - \delta \beta_{\text{gn}} \Phi^\delta c_m^\delta \sigma = -x m^x - \delta \beta_{\text{gn}} m^\delta \sigma \quad (6)$$

$$\left. \frac{da_j}{d\theta_n} \right|_{x < j < n^*} = \beta_{gn}(a_{j-\delta} - a_j)\Phi^\delta c_m^\delta = \beta_{gn}(a_{j-\delta} - a_j)m^\delta \quad (7)$$

$$\frac{da_x}{d\theta_n} = \Phi^x c_m^x - \beta_{gn} a_x \Phi^\delta c_m^\delta = m^x - \beta_{gn} a_x m^\delta \quad (8)$$

$$\frac{d\sigma}{d\theta_n} = \sum_{j=x}^{n^*-1} \frac{da_j}{d\theta_n} = \Phi^x c_m^x - \beta_{gn} a_{n^*-1} \Phi^\delta c_m^\delta = m^x - \beta_{gn} a_{n^*-1} m^\delta \quad (9)$$

The second equalities in eqs 6–9 follow from the identity $m = \Phi c_m$. β_{gn} is defined as

$$\beta_{gn} = \frac{\tau_n}{\tau_g} = \frac{\tau_n^{(0)}}{\tau_g^{(0)}} f_R^{(\delta-x)} \left(\frac{C_0}{C_{ref}} \right)^{(1+\delta-x)} \quad (10)$$

with the characteristic time scales $\tau_n \equiv \tau_n^{(0)} f_R^{-x} (C_{ref}/C_0)^{x-1}$, $\tau_g \equiv \tau_g^{(0)} f_R^{-\delta} (C_{ref}/C_0)^\delta$, $\tau_n^{(0)} \equiv (k_{nuc} K_{x-1} C_{ref}^{x-1})^{-1}$, and $\tau_g^{(0)} \equiv (k_g K_{RA}^{\delta-1} C_{ref}^\delta)^{-1}$. Dimensionless time in eqs 6–9 is defined as $\theta_n \equiv t/\tau_n$.

τ_n and τ_g are the characteristic timescales of nucleation and growth, respectively, for a given temperature, pressure, solvent composition, and C_0 . $\tau_n^{(0)}$ and $\tau_g^{(0)}$ are the intrinsic timescales of nucleation and growth, respectively, for the reference or standard state conditions in which $[R] = C_{ref}$ is imposed. Defining intrinsic time scales in this way factors out the effects of differences in concentration of reactive monomer due to changes in total protein concentration (C_0) or unfolding free energy (i.e., f_R) at a particular solvent composition, temperature, or pressure.

2.7. Limiting Cases. In this subsection we briefly explain how a number of previously proposed mathematical models of protein aggregation kinetics^{30,31,44,45} occur as limiting cases of eqs 6–9, including that explored in detail in the remainder of this report. We first consider the case where

$$x \left[f_R^x (c_m)^{x-1} \frac{K_{x-1} k_{nuc} C_0^{x-1}}{k_{r,x}} \right] \Phi^{x-1} + \sum_{i=2}^{x-1} i [f_R^i (c_m)^{i-1} k_i C_0^{i-1}] \Phi^{i-1} + \delta \left(\frac{k_g K_{RA}^{\delta-1} C_0^\delta}{k_r} f_R^\delta (c_m)^{\delta-1} \sigma \right) \Phi^{\delta-1} + \sum_{i=1}^{\delta-1} [K_{RA}^{i-1} C_0^{i-1} f_R^i \sigma (c_m)^{i-1}] \Phi^{i-1} \ll 1 \quad (11)$$

This is mathematically equivalent to stipulating $\Phi = 1$ in eq 5. Experimentally, eq 11 is expected to hold under conditions that favor native or unreactive monomer conformations (i.e., $f_R \ll 1$). Independent of f_R , it also holds when initial protein concentrations are sufficiently low that the process of forming reversible oligomers R_i is thermodynamically unfavorable (i.e., $K_i C_0^{i-1} \ll 1$ for all i) and the total number concentration of irreversible aggregates is relatively low ($\sigma \ll 1$).

From eqs 5 and 11 it follows that $\Phi = 1$, $c_m = m$, and eqs 6–9 become

$$\frac{dm}{d\theta_n} = -x m^x - \delta \beta_{gn} m^\delta \sigma \quad (12)$$

$$\left. \frac{da_j}{d\theta_n} \right|_{x < j < n^*} = \beta_{gn}(a_{j-\delta} - a_j)m^\delta \quad (13)$$

$$\frac{da_x}{d\theta_n} = m^x - \beta_{gn} a_x m^\delta \quad (14)$$

$$\frac{d\sigma}{d\theta_n} = m^x - \beta_{gn} a_{n^*-1} m^\delta \quad (15)$$

The solution to eqs 12–15 is a function of only four parameters (x , n^* , δ , and β_{gn}) and the initial conditions ($m = 1$, $a_j = a_{j,0}$, $\sigma = \sigma_0 = \Sigma a_{j,0}$). In eqs 12–15, the m^x term is that for nucleation of A_x from x molecules of R . The terms containing $\beta_{gn} m^\delta$ correspond to polymer growth via addition of δ monomers per net irreversible growth event. Equation 15 indicates that new aggregates form only via nucleation (the m^x term) and that no aggregates are lost unless they grow to n^* or larger.

Equations 12–15 are equivalent to the rearrangement-limited aggregation scenario described briefly in ref 33. If one also imposes the constraints that $f_R = 1$ and $n^* \rightarrow \infty$, then they become mathematically equivalent to the model of native protein polymerization analyzed by Ferrone⁴⁴ and used to interpret experimental kinetics of non-native aggregation.^{39,40} They become formally equivalent to that earlier model if one also assumes nucleation and growth are association-limited ($k_{a,x} = k_a$, $k_r \gg k_d$, $k_{r,x} \gg k_{d,x}$) and therefore uses $x - 1$ as the nucleus size. With a further assumption that nucleation consumes a negligible amount of monomer before growth becomes appreciable (i.e., $\beta_{gn} \gg 1$), eqs 12–15 can be reduced to the nucleated polymerization model solved analytically by Oosawa and Asakura.³¹ Finally, eqs 12–15 are equivalent to the purely irreversible, association-limited, or downhill polymerization model considered previously by one of us⁴⁵ if $x = 2$, $k_r \gg k_d$, $k_{r,x} \gg k_{d,x}$, and one replaces $k_{a,x}$ and k_a with $k_{1,1}$ and $\langle k_{ij} \rangle$ in ref 45.

Equations 6–9 also reduce to a model that is very similar to that presented by Lomakin et al.³⁰ (Appendix C) if one assumes that all free monomers exist in the reactive conformation ($f_R = 1$), all $A_j R$ concentrations and all R_i concentrations other than R_x are arbitrarily small, $k_{r,x} \ll k_{d,x}$, $n^* \rightarrow \infty$, and $x \gg 1$. The original model of ref 30 assumes that nuclei are irreversibly formed as subunits of much larger micelles, R_y ($y \gg x$). As long as y and/or x are large ($\gg 10$), there will be a threshold or critical micelle concentration (c^*) for monomers. Below c^* , aggregation does not occur without seeding.³⁰ If the assumption of large x is relaxed, then one can also recover a version of the model equations similar to a case considered by Goldstein and Stryer⁵³ and by Powers and Powers.⁵⁴

3. Results and Discussion

The model considered throughout most of this section is given by eqs 12–15. It is relevant when $\Phi \approx 1$ holds, e.g., native-favoring conditions or conditions where the total protein concentration can be considered thermodynamically dilute.

Simulations of eqs 12–15 were performed with the following ranges of model parameters: $x = 2, 4, 6$; $\delta = 1$ or 2 ; $\beta_{gn} = 10^{-1}$ to 10^4 ; $n^* = 10$ to 10^3 (effectively $n^* \rightarrow \infty$). Unless otherwise stated, all results are for initial conditions of $m = 1$, $\sigma = 0$, and therefore $a_j = 0$ ($x \leq j < n^*$). Calculations were also performed for additional x and δ combinations at selected values

TABLE 2: Summary of Key Experimental Signatures and Scaling Behaviors for Each Kinetic Type Produced by the LENP Model^a

kinetic type	γ^b	ν^c	α^d	intrinsic time scale ^e	$\langle MW \rangle_w / MW_{\text{mon}}$	
					polydispersity	$m \rightarrow 0$ limit ^f
Ia^g						
low β_{gn}	$x - 1$	x	x	$\tau_n^{(0)}$	high	
high β_{gn}	$x/2$	$x/2 + 1$	$x/2 + 1$	$\tau_n^{(0)}(\tau_n^{(0)}/\tau_c^{(0)})^{1/2}$	high	
Ib	$x - 1$	x	x	$\tau_n^{(0)}$	high	
Ic	$x - 1$	x	x	$\tau_n^{(0)}$	monodisperse	x
Id	$x - 1$	x	x	$\tau_n^{(0)}$	low	$1/\sigma_{\text{ss}}$
II	$(x + \delta - 1)/2$	δ	$(x + \delta)/2$	$(\tau_n^{(0)}\tau_g^{(0)})^{1/2}$	low	$\sigma_{\text{ss}}^{-1} \approx C_0^{(x-\delta-1)/2}$
III^h	$x - 1$	$x, 1$	$x, 1$	$\tau_n^{(0)}, \tau_g^{(0)}$	ppt	ppt
IV	$(x + \delta - 1)/2$	δ	$(x + \delta)/2$	$(\tau_n^{(0)}\tau_g^{(0)})^{1/2}$	low	$\geq n^*$
type	physical scenario					
Ia	low β_{gn} : condensation-dominated growth; nucleation-controlled kinetics					
Ib	high β_{gn} : competing condensation, polymerization, and nucleation					
Ic	polymerization to low-MW, soluble aggregates; condensation dominates thereafter					
Id	nucleation only; negligible polymerization or condensation					
II	nucleation-dominated; slow polymerization; negligible condensation					
III	slow nucleation, fast polymerization; negligible condensation					
IV	type Ib with aggregates saturated at (finite) solubility limits					
	slow nucleation; fast polymerization; condensation appreciable after $t > \text{ca. } t_{50}$					

^a Examples of full profiles for each type except⁴⁵ **Ia** and **III** are given in Figures 4 and 6. ^b γ is defined by $1/t_{50} \sim k_{\text{obs}} \sim (C_0)^\gamma$. ^c ν is defined by $dm/dt = -k_{\text{obs}}m^\nu$ over ca. one or more half-lives. ^d α is defined by $k_{\text{obs}} \sim f_R^\alpha$. ^e $\tau_n^{(0)} = 1/[k_{\text{nuc}}K_{x-1}C_{\text{ref}}^{(x-1)}]$; $\tau_g^{(0)} = 1/[k_gK_{\text{RA}}^{(\delta-1)}C_{\text{ref}}^\delta]$; $\tau_c^{(0)} = 1/[k_cC_{\text{ref}}]$, $k_c = \langle k_{ij} \rangle$ in ref 45. ^f σ_{ss} from solution to eq 27; $\log \sigma_{\text{ss}} \sim [(x-1-\delta)/2] \log C_0$ in type **II**; cases where the model does not provide information have no entry; ppt indicates that precipitates are necessarily present. ^g References 33 and 45 and Appendix A. ^h Reference 45.

of n^* and β_{gn} to verify that the general behaviors reported extend beyond those shown explicitly here. All time profiles are shown with time scaled by t_{50} , the time for 50% loss of initial monomer for that particular set of x , δ , n^* , and β_{gn} . This allows results with widely differing parameter sets to be more equitably compared in the figures provided.

The primary outputs of the model that are directly comparable with experimentally accessible quantities are the monomer loss kinetics (or extent of reaction on a mass basis) and moments of the soluble aggregate populations. The former are typified by $m(t/t_{50})$ profiles and the corresponding values of k_{obs} and effective reaction order (ν) with respect to reduced monomer concentration ($dm/dt \approx -k_{\text{obs}}m^\nu$ when considered over $t/t_{50} > \text{ca. } 1$). The moments considered here are a dimensionless zeroth moment (equivalent to σ defined in section 2), a first moment ($\lambda^{(1)}$, defined below), and a dimensionless second moment or weight-averaged molecular weight ($\langle MW \rangle_w$, defined below). The foremost of these is typically what is measured by any extent-of-reaction-based assay (see also, discussion regarding eq 4). If accessible, the lattermost quantity is typically measured via static light scattering.

Additional physically relevant quantities are also calculable in the LENP model, e.g., total aggregate mass precipitated ($j \geq n^*$ if n^* is determined by aggregate solubility limits). However, certain types of experimental quantities, such as increases in the bound fraction of dye molecules are not unambiguously predicted by this or any mass-action model unless one makes additional assumptions regarding the size-dependent molecular structure and/or specificity of ligand–aggregate interactions.

Experimentally, the value of β_{gn} for a given protein can be manipulated by changing C_0 , temperature, pressure, or solvent conditions. At fixed temperature and solvent composition, increasing C_0 corresponds to decreasing β_{gn} if $x > \delta + 1$, as all other parameters in β_{gn} depend on only temperature and solvent composition (cf. eq 10). Changing temperature or solvent composition will likely change more than one contribution to β_{gn} ; however, increasing temperature or adding chemical

denaturants that increase f_R will tend to decrease β_{gn} if $R \neq N$. Equivalently, decreasing protein concentration and/or changing sample conditions to shift the folding–unfolding equilibrium away from R will increase β_{gn} . Table 2 summarizes many of the results from subsections 3.1–3.6 and may be useful to refer to when considering each section below.

3.1. Model Behavior with Finite n^* . This section presents the global behavior of the LENP model as a function of its model parameters. The results are illustrated primarily with $x = 6$ and $\delta = 1$. The same qualitative behaviors and underlying causes are found for all other x and δ pairs investigated here. Figure 3 shows state diagrams illustrating the range of parameters (x , δ , n^* , β_{gn}) over which different kinetic regimes or types were observed. Different symbols indicate qualitatively different kinetic behaviors explained in more detail below. Final working equations for each case are presented in subsection 3.6.

Figures 4a and 4d show time profiles for the monomer fraction remaining (m) and $\langle MW \rangle_w$ as a function of n^* at $\beta_{\text{gn}} = 10^3$. The $\langle MW \rangle_w$ scaled by the monomer molecular weight (MW_{mon}), assuming $\Phi = 1$, is

$$\frac{\langle MW \rangle_w}{MW_{\text{mon}}} = \frac{m(t) + \sum_{j=x}^{n^*-1} j^2 a_j(t)}{m(t) + \sum_{j=x}^{n^*-1} j a_j(t)} \quad (19)$$

The summations in eq 19 begin at $j = x$ rather than $j = 2$ because the concentrations of each $R_{i=2,\dots,x}$ are not experimentally appreciable for conditions where $\Phi = 1$. Figures 4b and 4c, respectively, show the corresponding time profiles for the dimensionless number concentration of soluble aggregates (σ) and the first moment of the soluble aggregate distribution with $j < n^*$. This moment, $\lambda^{(1)}$, is defined as

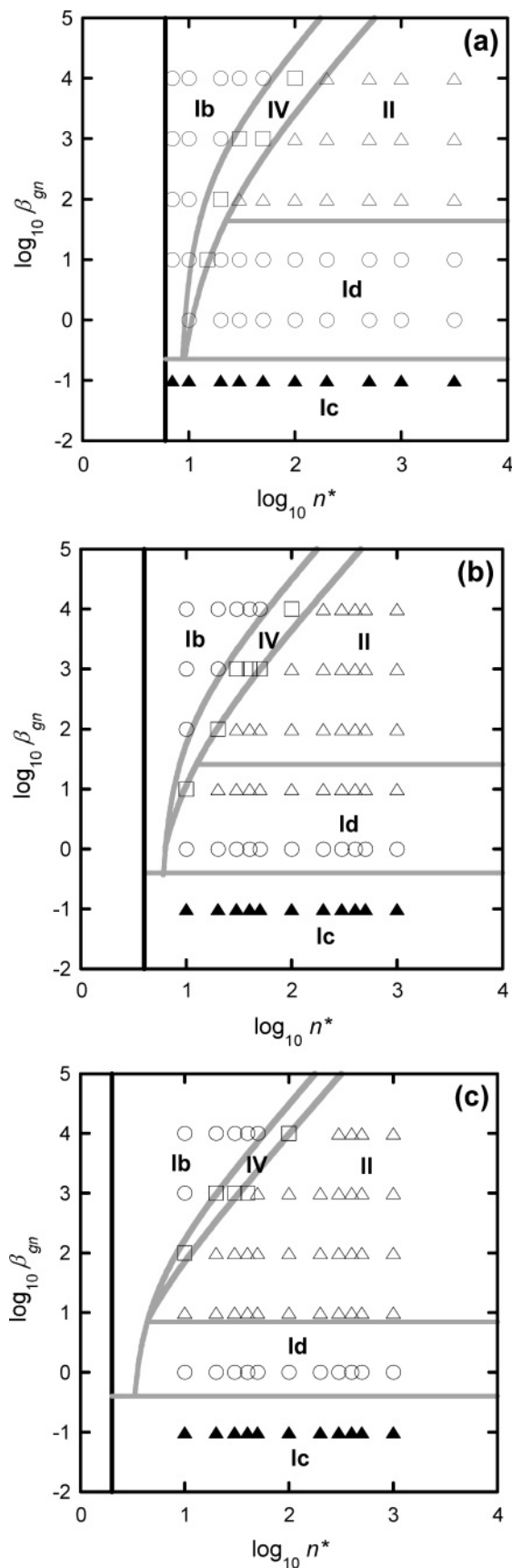


Figure 3. Kinetic state diagrams from the numerical solution of eqs 12–14 for (a) $x = 6$, (b) $x = 4$, and (c) $x = 2$. The qualitatively different kinetic profiles (types **Ib**, **Ic**, **Id**, **II**, and **IV**) are explained in section 3. Symbols represent the results from direct numerical simulation. The boundaries between behavior types are derived from the simple theoretical considerations outlined in the Supporting Information.

$$\lambda^{(1)}(t) = \frac{\sum_{j=x}^{n^*-1} j a_j(t)}{\sum_{j=x}^{n^*-1} a_j(t)} = \frac{\sum_{j=x}^{n^*-1} j a_j(t)}{\sigma(t)} \quad (20)$$

The second equality in eq 20 also provides the relationship between σ and $\lambda^{(1)}$ at a given time t . This first moment may not be experimentally accessible but is considered here because of its utility in later sections of the report.

From Figure 4, three qualitatively different behaviors are evident as a function of n^* at this relatively high value of β_{gn} . The first occurs at low n^* , with the following features (profiles labeled **Ib**): $m(t)$ follows kinetics that are x th-order in m ; σ increases rapidly to a sharp maximum at low extents of reaction ($t/t_{50} \ll 1$), then declines with a scaling of $\sigma(t) \sim [m(t)]^{x-1}$ at longer times; $\lambda^{(1)}$ increases rapidly to a plateau value near $(n^* + x)/2$. Physically, these profiles show that any nuclei that form are quickly grown to $j \geq n^*$ where they are no longer able to consume monomers. As a result, the concentration of each aggregate species (a_j) rapidly reaches an individual steady state (see also Figure 5a). Applying a steady-state Bodenstein approximation to eqs 13 and 14 and substituting the values for $a_{x,ss}$ and $a_{j,ss}$ in eq 12 gives

$$\frac{dm}{d\theta_n} \cong -n^* m^x \quad (\text{type Ib}) \quad (21a)$$

$$\sigma = \frac{(n^* - x)}{\delta \beta_{gn}} m^{x-\delta} \quad (\text{type Ib}) \quad (21b)$$

The δ in the denominator of eq 21b occurs because only $(n^* - x)/\delta$ terms in the a_j distribution will be nonzero at any time if all $a_j = 0$ at $t = 0$.

Type **Ib** behavior is essentially steady-state polymerization for short polymers. It is worth noting that although the Bodenstein approximation is mathematically convenient the results above and later in this section clearly indicate that this is a reasonable approximation only at conditions of rather small n^* .

At the other extreme of n^* , Figure 4 shows the following features (profiles labeled **II**): $m(t)$ follows kinetics that are δ -order ($\delta = 1$ in Figure 4, see also inset to Figure 4a); σ initially increases rapidly and achieves a nearly steady-state or plateau value (σ_{ss}) at relatively low extents of reaction ($t/t_{50} \ll 1$); $\lambda^{(1)}$ initially increases linearly with time ($t/t_{50} < \text{ca. } 1$) but approaches an asymptote at long t/t_{50} . In this case, n^* is significantly larger than the largest appreciably populated soluble aggregates even once all monomer has been consumed. Thus, the profiles are independent of n^* (limit of $n^* \rightarrow \infty$) as n^* becomes sufficiently large at a given β_{gn} value.

Type **II** kinetics correspond to cases in which nuclei form at early t/t_{50} and grow until they exhaust the available monomer pool. However, the resulting aggregates do not reach sufficiently large sizes to appreciably condense. As a result, the a_j distribution never achieves a steady state (cf. Figure 5c), but the sum across that distribution (σ) does.

The relationship between σ_{ss} and the remaining model parameters is the subject of subsection 3.2. For now it suffices to note that eq 12 reduces to eq 22 for type **II** kinetics

$$\frac{dm}{d\theta_n} \cong -\delta \beta_{gn} m^\delta \sigma_{ss} \quad (\text{type II}) \quad (22)$$

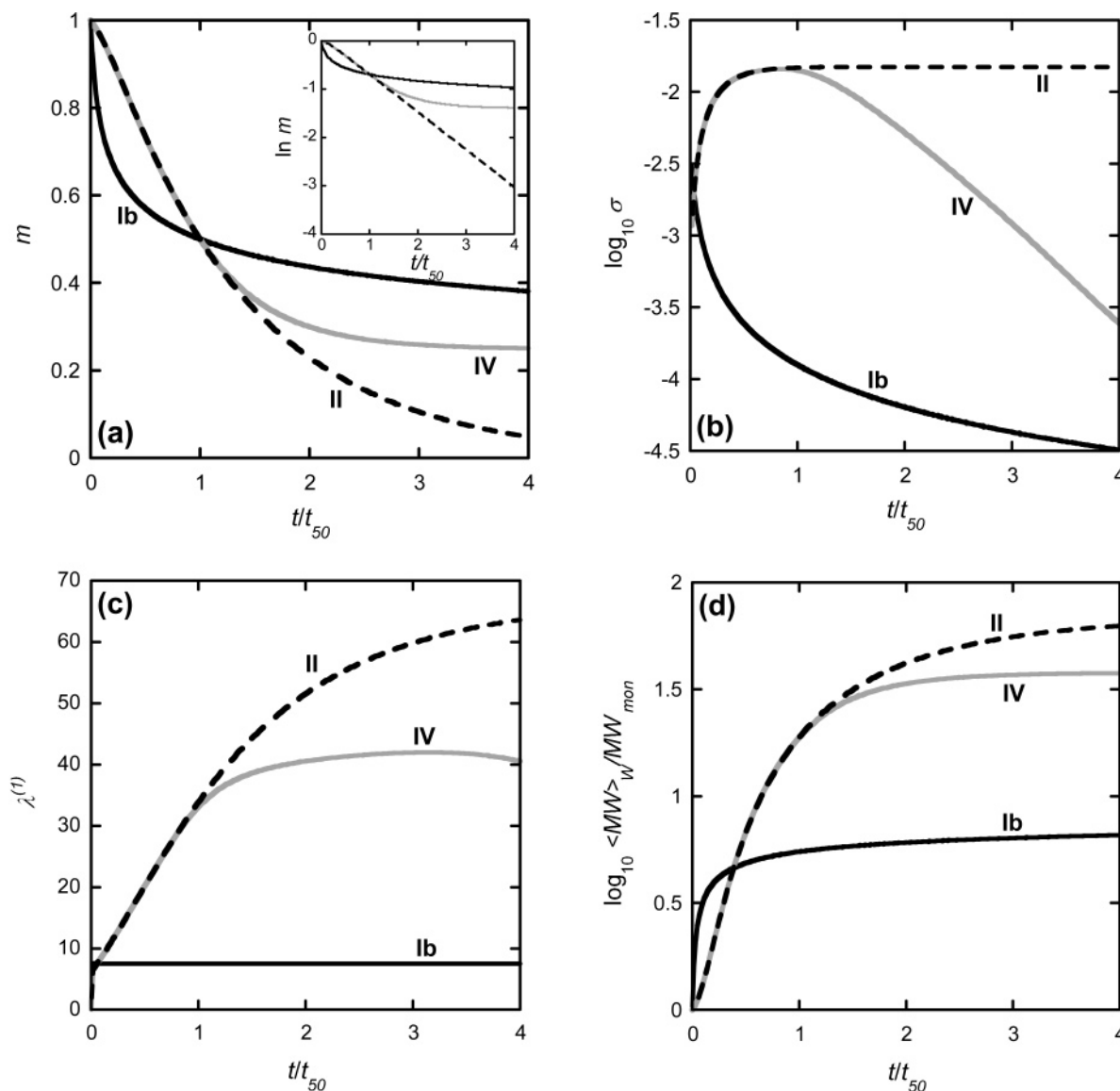


Figure 4. Illustration of the effects of changing n^* on the characteristic aggregation profiles versus time at fixed $\beta_{\text{gn}} = 1000$, $x = 6$, and $\delta = 1$. The panels show the time evolution of: (a) monomer loss; (b) total soluble aggregate (σ) with $j < n^*$; (c) first moment of the aggregate size distribution (eq 20); (d) weight-averaged molecular weight scaled by monomer molecular weight (eq 21). Types **Ib** (solid black), **IV** (solid gray), and **II** (dashed black) correspond to $n^* = 10, 50$, and 100 , respectively.

The remaining case in Figures 4 and 5 is denoted as type **IV**. The profiles in Figure 4 display the following features: $m(t)$, $\sigma(t)$, and $\lambda^{(1)}(t)$ profiles all overlay with those for type **II** at early t/t_{50} ; however, at $t/t_{50} > \text{ca. } 1$, the monomer loss rate abruptly slows, $\sigma(t)$ falls from the plateau level it had previously achieved (Figure 4b), and $\lambda^{(1)}(t)$ decreases at large t/t_{50} .

At low extents of reaction, type **IV** kinetics are exactly those of type **II**. However, unlike type **II** behavior, for type **IV** scenarios the soluble aggregate size distribution overlaps n^* before the large majority of monomer is consumed (remaining $m \approx 0.1$ or higher, cf. Figure 5b). An approximate representation of the $m(t)$ profile for type **IV** is

$$\frac{dm}{d\theta_n} \cong -\delta\beta_{\text{gn}}m^\delta\sigma_{\text{ss}} \quad \text{for } t < t_{\text{IV}} \quad (23a)$$

$$\frac{dm}{d\theta_n} \cong 0 \quad \text{for } t > t_{\text{IV}} \quad (23b)$$

The value of t_{IV} is approximately that at which $\lambda^{(1)}$ reaches n^* .

There are two additional kinetic regimes exhibited by eqs 12–15. These are denoted as types **Ic** and **Id** and are illustrated in Figure 6 by changing β_{gn} at fixed $n^* = 50$. The panels are analogous to those in Figure 4. Curves labeled **Ib**, **II**, and **IV** in Figure 6 all show the features described above, except that $\lambda^{(1)}$ and $\langle \text{MW} \rangle_w / \text{MW}_{\text{mon}}$ reach higher plateau values for **Ib** than for **II** because β_{gn} is not held fixed in Figure 6.

Type **Ic** behavior is independent of n^* and manifests at low ratios of the rates of aggregate growth versus nucleation, i.e., β_{gn} significantly less than 1. In this case the rate of aggregate growth is negligible on the time scale of aggregate nucleation. Therefore, σ reaches a plateau value near $(1 - m_\infty)/x$ because all aggregates are x -mers. The long-time monomer fraction ($m_\infty \equiv m$ at $t/t_{50} \gg 1$) is significantly larger than zero when x is significantly greater than 2. The lack of an n^* dependence follows because no A_j species with $j > x$ are significantly populated, and $n^* \geq x$ by definition. Equation 12 reduces to

$$\frac{dm}{d\theta_n} \cong -xm^x \quad (\text{type Ic}) \quad (24)$$

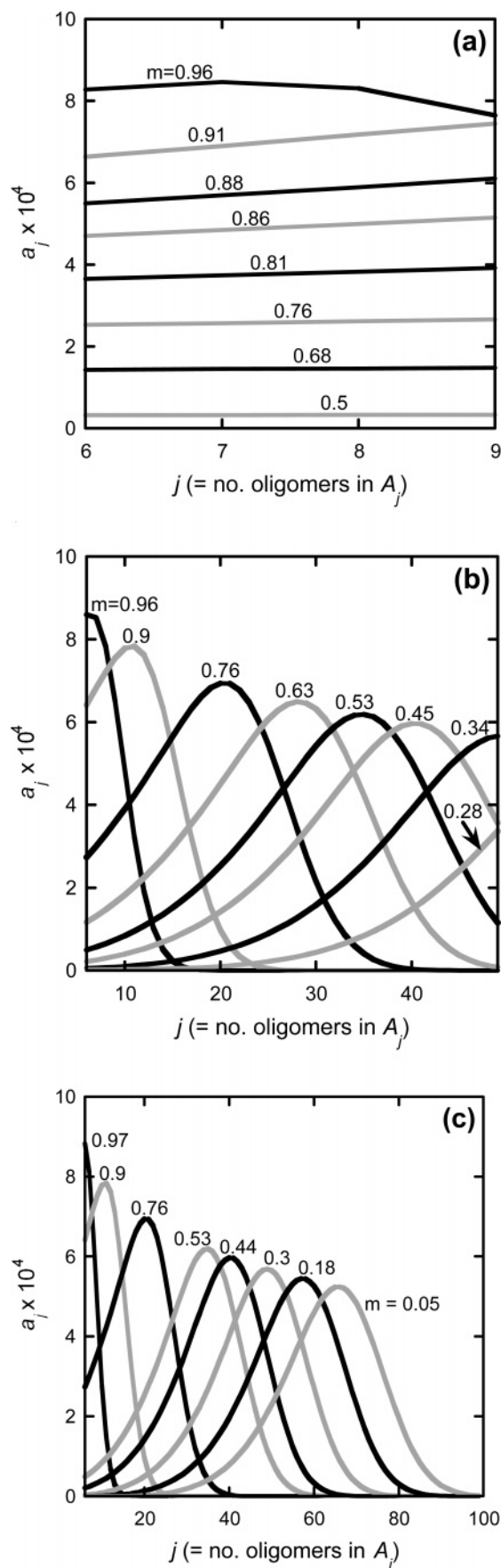


Figure 5. Evolution of the soluble aggregate size distribution, a_j ($x \leq j < n^*$), with extent of reaction ($1 - m$) from each of the simulated profiles in Figure 4: (a) **Ib**; (b) **IV**; (c) **II**. The number beside each a_j distribution indicates its corresponding value of fraction monomer remaining (m).

Types **Ic** and **Ib** are experimentally indistinguishable from their respective monomer loss kinetics alone. It is only through examining their aggregate size distributions at longer times that one may distinguish them experimentally (see also subsection 3.3). Type **Ib** has a broad distribution with a relatively low overall number concentration of aggregates ($\sigma \ll 1$), while type **Ic** is essentially monodisperse with an average size slightly below x .

Type **Id** behavior is intermediate to **Ic** and **II**. $\sigma(t)$ is qualitatively the same as **Ic** and **II**, while $\lambda^{(1)}(t)$ increases slowly and essentially linearly over all t/t_{50} of practical interest. In terms of monomer loss, it displays effectively mixed-order kinetics: $m(t)$ appears x th-order at $t/t_{50} < \text{ca. } 1$ and appears closer to δ -order kinetics at larger t/t_{50} values

$$\frac{dm}{d\theta_n} \cong -xm^x \quad \text{for } t < t_{Id} \quad (25a)$$

$$\frac{dm}{d\theta_n} \cong -\delta\beta_{gn}m^\delta\sigma_{ss} \quad \text{for } t > t_{Id} \quad (25b)$$

In eq 25, t_{Id} is a somewhat arbitrary time at which the second term in eq 12 tends to dominate over the first. To properly capture type **Id** behavior over a number of half-lives, one should use eqs 12 and 15 with the a_{n^*-1} term neglected in the $n^* \rightarrow \infty$ limit rather than eq 25.

For nucleus sizes with $x > \text{ca. } 2$ the qualitative $m(t/t_{50})$ profiles for types **Ib**, **Ic**, **Id**, and **IV** could be mistakenly interpreted as showing biphasic kinetics if one factors in the discrete nature and inherent statistical uncertainty in most experimental aggregation data. Therefore, the qualitative $m(t/t_{50})$ profiles for types **Ib**, **Ic**, **Id**, and **IV** provide alternative interpretations of experimental observations sometimes considered biphasic.^{67,68}

In cases **Ib** and **IV** the residual pool of monomer left after the initial rapid phase of monomer loss is not unreactive or resistive to aggregation. Rather, it is not consumed on experimental time scales simply because: (1) nucleation rates drop dramatically as monomer concentration drops; (2) those aggregates that were initially nucleated are grown rapidly and reach n^* before they consume all remaining monomers. Thus, the monomers that remain after the apparent burst phase could be readily consumed if reactive aggregates were still present. The fact that new aggregates are not appreciably nucleated is simply a consequence of the higher-order dependence of nucleation rates on monomer concentration ($\sim m^x$ in eq 12).

In type **Ic** and **Id** kinetics, the residual pool of monomer is also no less reactive to aggregation than those monomers consumed earlier in the reaction. In these cases the apparent initial burst phase followed by a slower phase of monomer consumption is a consequence of the combination of slow growth kinetics and the higher than second-order dependence of nucleation rates on monomer concentration.

3.2. Steady-State σ Values for Infinite n^* Conditions (Types **Ic, **Id**, and **II**).** As $n^* \rightarrow \infty$, eq 12 remains unchanged, eq 15 becomes eq 26, and eqs 13 and 14 no longer need to be integrated

$$\frac{d\sigma}{d\theta_n} = m^x \quad (26)$$

From the discussion in subsection 3.1 it follows that this case produces only types **Ic**, **Id**, and **II** behaviors. The value of σ_{ss} is a function of x , δ , and β_{gn} . It determines the observed $m(t)$

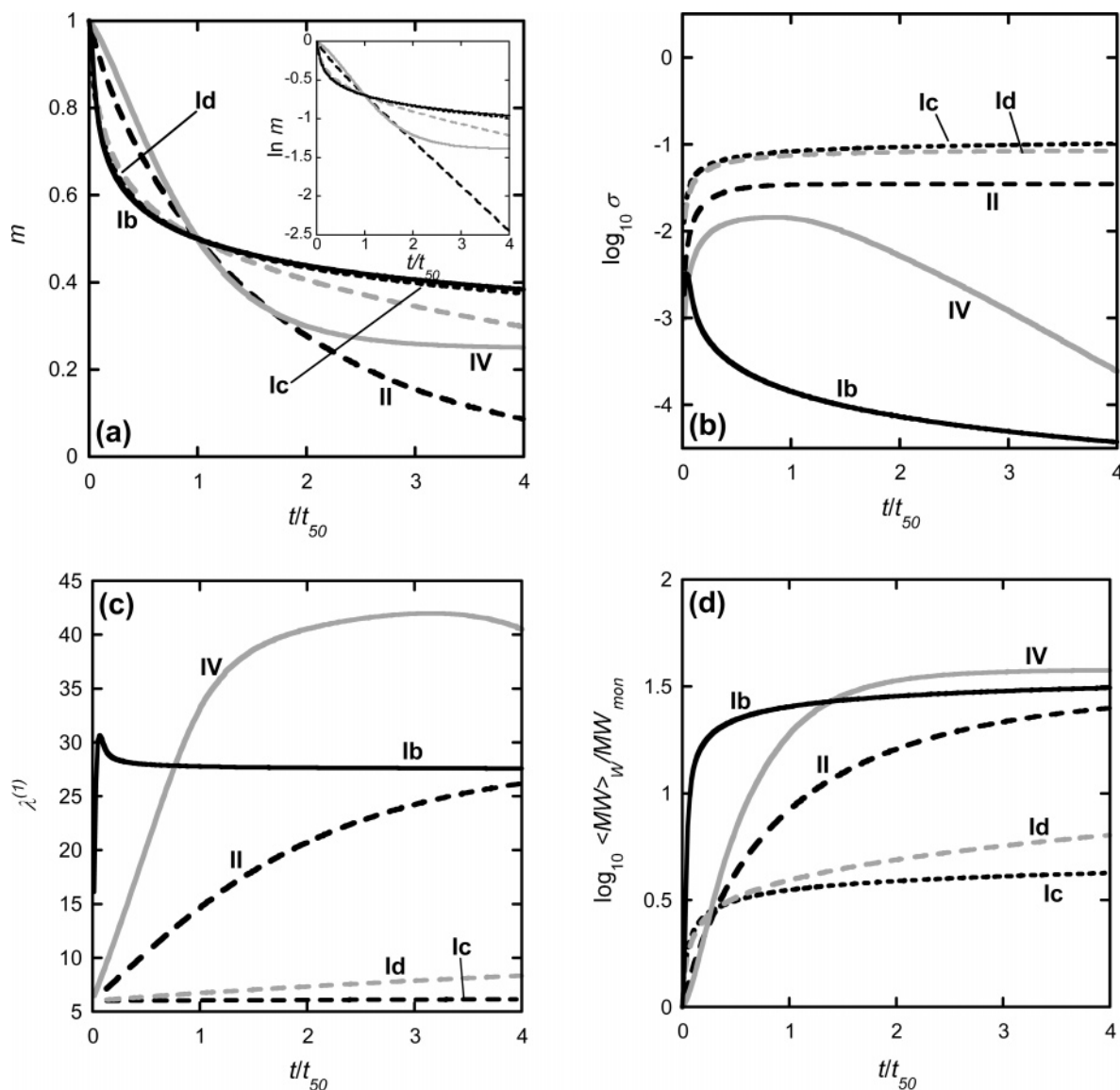


Figure 6. Illustration of the effects of changing β_{gn} on the characteristic aggregation profiles at fixed $n^* = 50$, $x = 6$, and $\delta = 1$. The panels are analogous to those in Figure 4. Types **Ib** (solid black), **Ic** (dotted black), **Id** (dashed gray), **II** (dashed black), and **IV** (solid gray) for this set of (n^* , x , δ) correspond to $\beta_{\text{gn}} = 10^4$, 0.1, 2, 100, and 1000, respectively.

and $\sigma(t)$ profiles for type **II** kinetics as well as other experimentally useful quantities (see also subsections 3.3 and 3.5).

Equations 12 and 26 can be combined to give

$$\frac{d\sigma}{dm} = -\frac{1}{x + \delta\beta_{\text{gn}}m^{\delta-x}\sigma} \quad (27)$$

with the initial condition of $\sigma = 0$ at $m = 1$. Profiles of $\sigma(m)$ are shown in Figure 8 for a range of β_{gn} values for $x = 6$ and $\delta = 1$. For most physically relevant values of β_{gn} (i.e., $\beta_{\text{gn}} > \text{ca. } 1$), a plateau or steady-state value of σ is achieved (σ_{ss}) before approximately 50% monomer loss in all cases. For a given x and δ , σ_{ss} is a function of only β_{gn} . The inset to Figure 8 shows $\sigma_{\text{ss}}(\beta_{\text{gn}})$ for $x = 6$ and $\delta = 1$ obtained by integrating eq 27 from $m = 1$ to $m = 0.001$ for a wide range of β_{gn} values (10^{-3} to 10^6). The profile is plotted versus $-\log \delta\beta_{\text{gn}}$, as this corresponds to increasing C_0 when $x > \delta + 1$ (see also section 3) and doing so permits collapse of multiple profiles when considering different x and δ values (cf. Figure 9). The vertical lines in the inset correspond to the horizontal lines in Figure 3a delineating bounds between types **Ic**, **Id**, and **II**.

Figure 9 shows $\sigma_{\text{ss}}(\beta_{\text{gn}})$ profiles for a range of x with $\delta = 1$ (main panel) and 2 (inset). The logarithm of $\sigma_{\text{ss}}x$ is plotted versus that for $\delta\beta_{\text{gn}}$ to more closely align the profiles for different x and δ values. It is apparent that the high and low β_{gn} regimes of σ_{ss} can be reasonably described by common curves.

$$\sigma_{\text{ss}} \cong \frac{1}{x} \quad \text{at low } \beta_{\text{gn}} \quad (28)$$

$$\sigma_{\text{ss}} \cong \frac{\sigma_{\text{int}}}{x} (\delta\beta_{\text{gn}})^{-1/2} \quad \text{at high } \beta_{\text{gn}} \quad (29)$$

The parameter σ_{int} is the average value of σ_{ss} if the linear, high- β_{gn} regions are extrapolated (on the log–log scale of Figure 9) to the intercept at $\log \delta\beta_{\text{gn}} = 0$. Empirically, σ_{int} is ca. 2 for $\delta = 1$ and ca. 3 for $\delta = 2$.

Equation 28 provides an upper limit for the value of the number concentration of aggregates (σ) and is that in which all monomers have been converted to the smallest irreversible aggregate, a_x , without subsequent growth. This σ_{ss} value is larger than that in Figures 4, 6, and 7 under type **Ic** conditions because

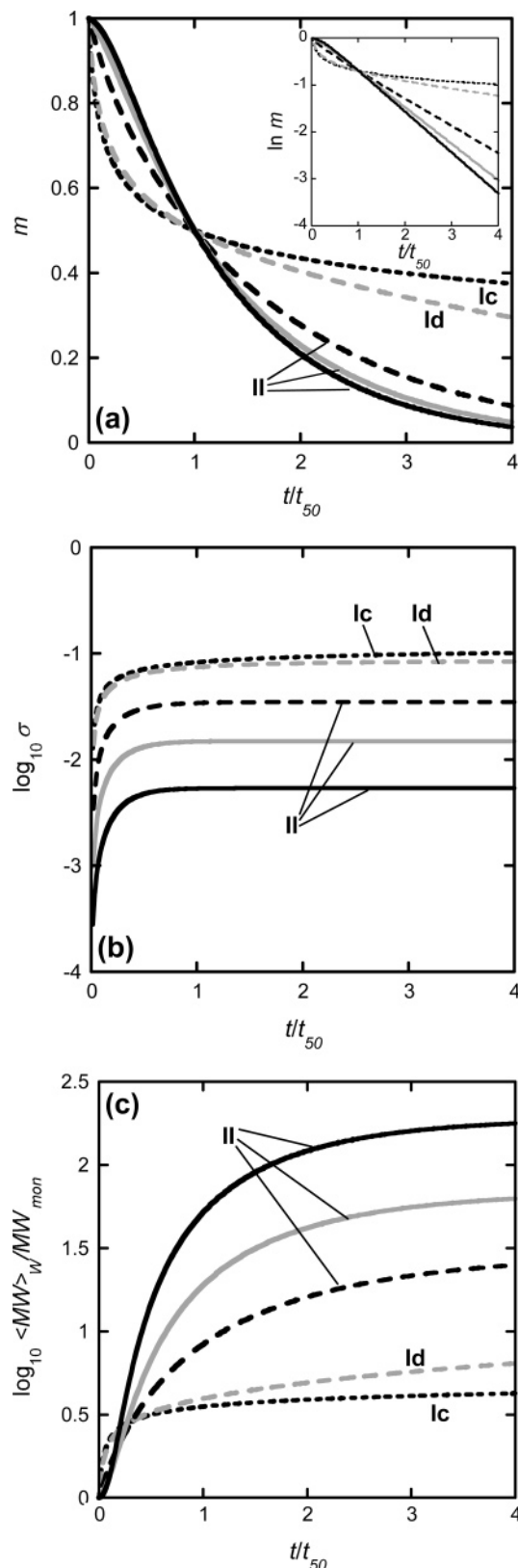


Figure 7. Illustration of the effects of changing β_{gn} on the characteristic aggregation profiles at $n^* = 500$ (effectively infinite n^*) with $x = 6$ and $\delta = 1$. The panels show: (a) time evolution of monomer loss; (b) total soluble aggregate with $j < n^*$; (c) weight-averaged molecular weight scaled by monomer molecular weight. The upper two curves in panels a and b are type **Ic** (dotted black) and **Id** (dashed gray) for $\beta_{gn} = 0.1$ and 2, respectively. The lower three curves in panels a and b are all type **II** with different values for $\beta_{gn} = 100$ (dashed black), 1000 (solid gray), and 10^4 (solid black). The corresponding curves are indicated in panel c.

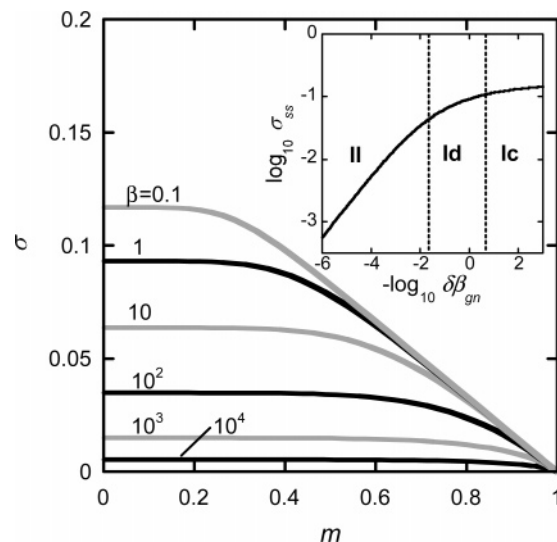


Figure 8. Solution to eq 27 for σ as a function of m ($\sigma_0 = 0$ at $m = 1$) with $x = 6$ and $\delta = 1$, as a function of β_{gn} . Inset: Corresponding behavior of the plateau values (σ_{ss}) over a finer discretization of β_{gn} . The vertical lines in the inset correspond to the β_{gn} values for the horizontal boundaries between **II**, **Id**, and **Ic** in Figure 3a.

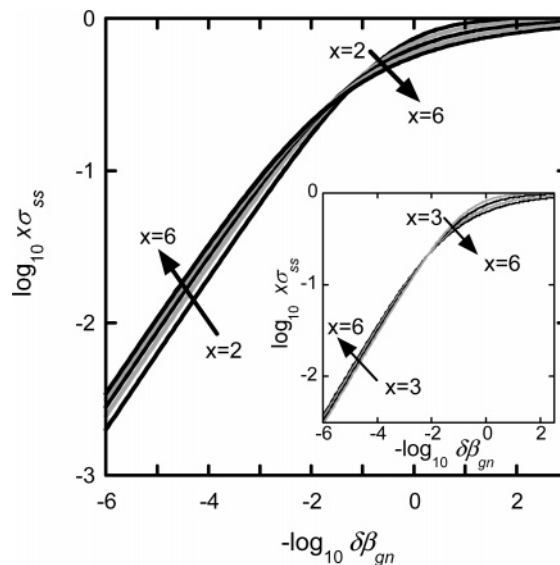


Figure 9. Scaling behavior of σ_{ss} for β_{gn} for a range of x with $\delta = 1$ (main panel) and $\delta = 2$ (inset). Note that $\sigma_{ss}x$ is plotted to scale out the effect of x on the low- β_{gn} (type **Ic**) limiting behavior.

eq 27 was solved for $m \rightarrow 0$. The results in subsection 3.1 did not extend beyond $t/t_{50} = 4$ and so produced $m_{\infty} \approx 0.4$ for type **Ic**.

3.3. $\langle MW \rangle_w$ Profiles and Scaling with β_{gn} . The dependence of $\langle MW \rangle_w$ on time and (n^*, β_{gn}) was shown qualitatively in Figures 4, 6, and 7. In this subsection we examine in more detail the behavior of $\langle MW \rangle_w$ for conditions corresponding to type **II**, **Ic**, and **Id** cases. Before doing so, a discussion of the other two scenarios (**Ib** and **IV**) is first presented. These are cases in which $\lambda^{(1)}$ reaches or approaches n^* before monomers have been effectively consumed (before $m \approx 0$ or $m^* \approx 0$). Therefore, the experimental $\langle MW \rangle_w$ profiles and limiting, long-time behavior will depend on the physical basis for n^* . As a result, care must be taken in interpreting experimental data in the context of the current LENP model in these scenarios.

If aggregates remain soluble and/or suspended at $j \geq n^*$ such that techniques such as light scattering accurately detect them, then the $\langle MW \rangle_w$ profiles shown in Figures 4 and 6 will be

incorrect for type **Ib** at essentially all times and incorrect for type **IV** at long times. In such cases, one must explicitly account for condensation steps (aggregate–aggregate association or agglomeration).^{32,58,60} For the above reasons, we do not anticipate that the present version of the LENP model will quantitatively capture quantities such as $\langle MW \rangle_w$ for **Ib** even at early times. For type **IV** behavior, the long-time behavior will be in error, but the early-time behavior is unaffected by these caveats.

Independent of the preceding considerations, one can qualitatively anticipate rather broad (high polydispersity) aggregate size distributions for **Ib** behavior beginning at early times ($m \approx 1$) (cf. Figure 5a). In contrast, for type **IV** behavior before the plateau in m , the distribution will be relatively narrow (low polydispersity), as illustrated in Figure 5b. Once the plateau in m is reached, the size distribution will broaden significantly as aggregate–aggregate condensation induces greater polydispersity. Alternatively, if aggregates precipitate at $j \geq n^*$ the model predicts that the soluble aggregate fraction will behave as shown in Figures 4 and 6, provided the size distribution in the supernatant can be accurately assessed experimentally.

Given the above discussion, we restrict our quantitative analysis to conditions where aggregate condensation or precipitation does not occur (types **Ic**, **Id**, and **II**); experimentally, this would be evident from both a lack of visible precipitates and a relatively low polydispersity in light scattering, electron micrographs, etc.

For types **Ic**, **Id**, and **II**, $\lambda^{(1)} \ll n^*$ at all practical times. As a result, the denominator in eq 19 is exactly 1, and eq 19 becomes

$$\frac{\langle MW \rangle_w}{MW_{\text{mon}}} = m(t) + \sum_{j=x}^{n^*-1} j^2 a_j(t) \quad (30)$$

For type **Ic** there is no significant growth, and only a_x is populated. Therefore, $a_x(t) = x^{-1}(1 - m)$ and

$$\frac{\langle MW \rangle_w}{MW_{\text{mon}}} = x + (1 - x)m(t) \quad (\text{type Ic}) \quad (31)$$

In this case, $\langle MW \rangle_w / MW_{\text{mon}}$ begins at 1 and increases to a near constant value between 1 and x at long times. Recall from subsection 3.1 that as x increases m does not reach zero until $t/t_{50} \gg 1$. Therefore for practical purposes the long-time value of m in eq 31 is not necessarily close to zero.

For type **II** kinetics growth is rapid compared to nucleation, and nucleation effectively ceases after a brief initial phase as the value of m^x falls significantly below 1 in eq 12. As a result we find that after a short induction time ($t/t_{50} \ll 1$) the a_j distribution has low polydispersity with a variance of $SD_{\text{agg}} = c_x \lambda^{(1)}(t)$, with $c_x \approx 0.1$ – 0.6 depending on x (see also Figure 11). The variance is related via eq 32 to the mean value $\langle j \rangle$ and the expected value of j^2 in the normalized a_j distribution ($E(j^2)$)⁶⁹

$$SD_{\text{agg}}^2 = E(j^2) - \langle j \rangle^2 = \frac{\sum_{j=x}^{n^*-1} j^2 a_j}{\sigma} - (\lambda^{(1)}(t))^2 \quad (32)$$

Using eq 32 and the above proportionality between SD_{agg} and $\lambda^{(1)}$, eq 30 becomes

$$\frac{\langle MW \rangle_w}{MW_{\text{mon}}} \approx m(t) + (1 + c_x^2)(\lambda^{(1)}(t))^2 \sigma(t) \quad (33a)$$

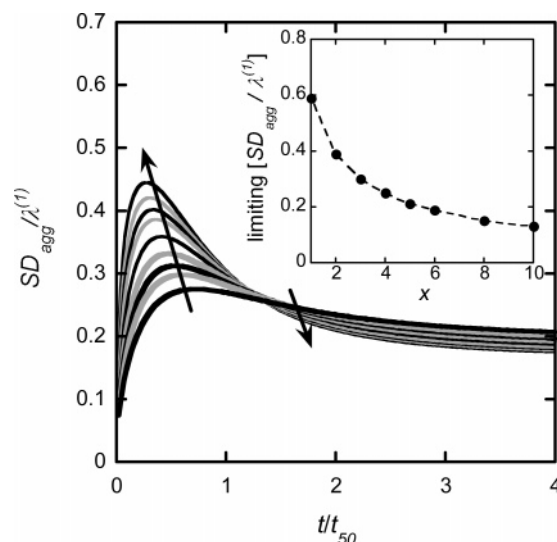


Figure 10. Ratio of SD_{agg} to $\lambda^{(1)}$ as a function of t/t_{50} for $x = 6$ and $\delta = 1$ over a range of β_{gn} values from 100 to 10^4 that fall in the type **II** regime (arrows indicate increasing β_{gn}). Inset: Long-time value of $SD_{\text{agg}}/\lambda^{(1)}$ as a function of x for $\delta = 1$. The variation of $SD_{\text{agg}}/\lambda^{(1)}$ with time at larger t/t_{50} values for each x value is smaller than the size of the symbols in the inset.

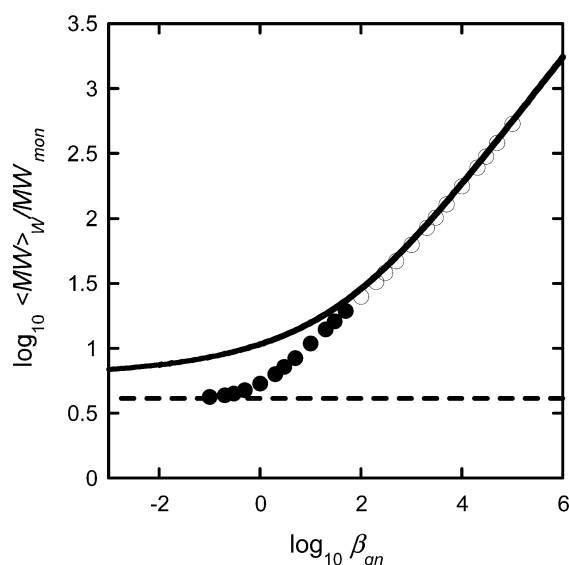


Figure 11. Scaling behavior of $\langle MW \rangle_w$ with β_{gn} in the limit of large n^* . Symbols represent results from numerical simulation of eqs 12–15 with $n^* = 1000$, $x = 6$, and $\delta = 1$ for long times ($t/t_{50} = 4.0$). The β_{gn} values used in the simulations correspond to type **II** (open circles) and types **Ic** and **Id** (closed circles). The curves are analytical results from eq 34 (solid black) and eq 33b (dashed black), respectively. Additional details are provided in subsection 3.3.

Substituting $\lambda^{(1)} = (1 - m)/\sigma$ (eq 20) with $n^* \gg \lambda^{(1)}$, eq 33a can be expressed as

$$\frac{\langle MW \rangle_w}{MW_{\text{mon}}} = m(t) + \frac{(1 + c_x^2)(1 - m(t))^2}{\sigma(t)} \quad (\text{types Id, II, and early time IV}) \quad (33b)$$

The term $(1 + c_x^2)$ becomes important as $x \rightarrow 2$ (see Figure 10 inset for $c_x(x)$). In the absence of knowledge of x , one can use the approximation $(1 + c_x^2) \approx 1$. Finally, taking eq 33b in the limit as $m \rightarrow 0$ ($t/t_{50} \gg 1$) for types **Id** and **II**

$$\frac{\langle \text{MW} \rangle_w}{\text{MW}_{\text{mon}}}|_{m \rightarrow 0} \cong \frac{1 + c_x^2}{\sigma_{ss}(\beta_{gn}; x, \delta)} \quad (\text{types Ib and II}) \quad (34)$$

In eq 34 we have explicitly indicated the dependence of σ_{ss} on β_{gn} , x , and δ . Figure 11 shows a comparison of eq 34 with values for $\langle \text{MW} \rangle_w$ obtained at long time ($t/t_{50} = 4$) from direct simulation of eqs 12–14 with $(n^*, x, \delta) = (500, 6, 1)$. The solid black curve is $-\log \sigma_{ss}$ calculated from eq 27 with $x = 6$ and $\delta = 1$. The small differences between the curve and the values from the direct simulation under type II conditions (open circles) are due to the fact that σ_{ss} was calculated from eq 27 for $m = 0.001$, while the symbols are for $t/t_{50} = 4$. Therefore, the corresponding m values are slightly higher than 0.001. The differences between the solid curve and the direct simulation for $\beta_{gn} < \text{ca. } 100$ occur because at these low β_{gn} values type Ic and Id behaviors are found (cf. Figure 3a at high n^*). Therefore, the value of m is significantly greater than zero even at $t/t_{50} = 4$ (cf. Figures 6a and 7a). The dashed line corresponds to the value for $\langle \text{MW} \rangle_w / \text{MW}_{\text{mon}}$ obtained from eq 31 or from eq 33b using $m(t/t_{50} = 4)$ from the solution to eq 24 and $\sigma(t/t_{50} = 4) = x^{-1}(1 - m(t/t_{50} = 4))$.

We draw the reader's attention to the result that $\langle \text{MW} \rangle_w$ at a given fraction monomer loss scales as $\beta_{gn}^{1/2}$ for type II conditions and thus scales with $C_0^{(x-\delta-1)/2}$ at fixed temperature, pressure, and solvent composition. This result indicates that larger limiting aggregate sizes are predicted as one decreases C_0 if $x > \delta + 1$. If $x = \delta + 1$, then aggregate size distributions at a given m value are predicted to be independent of C_0 with all other conditions held constant. Further, this scaling relationship provides a means other than using initial rates or k_{obs} versus initial protein concentration^{31,39,40,44} to infer the nucleus size x from experimental data. It also highlights that models that assume equal rate coefficients for polymer nucleation and growth ($\beta_{gn} = 1$)^{58,59} severely constrain the aggregate size distributions that can be described.

The monomer loss profiles generated by the LENP model do not display a true lag phase, because aggregation begins at $t = 0$. They do, however, display a region of acceleratory kinetics for high values of β_{gn} , such as those visible in Figure 7a. An acceleratory region is more apparent in $\langle \text{MW} \rangle_w$ versus time profiles (cf. Figures 7a and 7c) because of $(1 - m)^2$ dependence (eq 33b). The absence of a true lag phase is the result of the assumption of pre-equilibration of reversible x -mers to form the prenucleus R_x . If R_x were not immediately available at $t = 0$, then a true lag phase could in principle be observed corresponding to the time required to form R_x , with a time constant determined by those of the forward association steps in stage 2.

3.4. Boundaries in Kinetic State Diagrams. The solid curves in Figure 3 show predicted boundaries based on the considerations described in the Supporting Information. They compare well with results from explicit simulations, indicating that one may reliably use the predicted state diagrams in the absence of direct simulation. Qualitatively, both simulated and predicted boundaries display a number of trends of use when interpreting experimental aggregation kinetics. For large β_{gn} (large τ_n/τ_g), only type Ib, II, or IV behavior occurs, and the range of n^* over which type Ib occurs expands with increasing β_{gn} . The range of n^* over which type IV occurs also expands with increasing β_{gn} and shrinks with decreasing x . For $x = 2$, the type IV regime is sufficiently small to have been overlooked in an earlier report.⁴⁵ Recalling the definitions of β_{gn} and n^* , it follows that types Ib, II, and IV are promoted by experimental conditions that shift the monomer conformational equilibrium

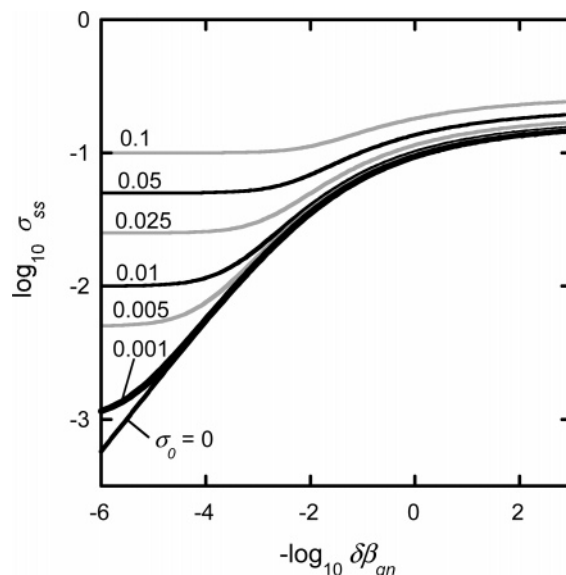


Figure 12. Effect of seeding on σ_{ss} vs $\delta\beta_{gn}$ profiles (large n^* , $x = 6$, and $\delta = 1$) from numerical solution of eq 27 with different σ_0 values. Numbers by each curve indicate the value of σ_0 used as the initial condition with $m = 1$.

away from R if $x > \delta$ and by decreasing the initial protein concentration if $x > \delta + 1$. Interestingly, if $x = \delta + 1$ then β_{gn} is unaffected by increases of initial protein concentration but still increases under conditions that disfavor R. This is the case for association-limited or downhill polymerization ($x = 2$ and $\delta = 1$).⁴⁵ These observations offer additional consistency tests for experimental kinetics when attempting to identify which kinetic regime applies and therefore which working equations should be used to interpret the kinetics.

3.5. Effects of Seeding. Seeding experiments, in which low levels of preformed aggregates are added to the solution at time $t = 0$, can provide an additional means to separate nucleation and growth kinetics. Starting with seeded solutions can increase the observed aggregation rate of systems following type II or IV kinetics by providing seeds for growth and bypassing all or part of the nucleation stage. The analyses in preceding sections indicate that seeding will have little or no perceptible effect on observed aggregation kinetics for type Ib situations. This follows because A_j species introduced as seeds will rapidly be swept to $j \geq n^*$, and therefore not have a large impact on $m(t)$ or the resulting a_j distribution.

The remainder of this section focuses on types Ic, Id, and II. In terms of solutions to the LENP model, the effect of seeding is to change the initial conditions for solving eqs 12–15, i.e., by making some set of aggregate concentrations (a_j) nonzero at $t = 0$. In the case of large n^* , the exact distribution of a_j at $t = 0$ is not needed; only the initial value of σ , defined here as σ_0 , is needed. Figure 12 illustrates the effect of nonzero σ_0 on the steady-state σ values obtained from the integration of eq 27 for a range of $\delta\beta_{gn}$ values for the same conditions as in Figure 5. Also shown is the curve for $\sigma_0 = 0$ indicating where σ_{ss} would lie in the absence of seeding. Not surprisingly, the results show that if σ_0 is significantly smaller than the σ_{ss} value that would be achieved in the absence of seeding, then seeding has no appreciable effect on the resulting σ_{ss} value. However, if σ_0 is at least 10-fold greater than the value of σ_{ss} in the absence of seeding, then $\sigma_{ss} \approx \sigma_0$. In that case, σ_{ss} does not scale with β_{gn} , and k_{obs} is therefore fundamentally different for types II, Id, and Ic. Specifically, eqs 12 and 15 reduce to eq 35, independent of β_{gn} .

$$\frac{dm}{dt} \cong - \frac{\delta f_R \sigma_0 (C_0/C_{\text{ref}})^{\delta}}{\tau_g^{(0)}} m^{\delta} = -k_{\text{obs}} m^{\delta} \quad (\text{seeded growth, high } \sigma_0) \quad (35)$$

The resulting expression for k_{obs} in eq 35 shows that the only time scale probed by experimental kinetics is that for growth $\tau_g^{(0)}$. To ensure purely seeded growth kinetics in experiments, one needs to seed with $\sigma_0 \approx 10\sigma_{\text{ss}}$. Experimentally, if one does not know the value of σ_{ss} for a given population, then the most conservative estimate is to assume $\sigma_{\text{ss}} \approx 0.1$ without seeding (Figure 6) and therefore use a seed-to-monomer number concentration ratio of 1:1. For large β_{gn} , Figures 6 and 12 show $\sigma_{\text{ss}} \ll 1$ in the absence of seeding. In such cases, a lower seed-to-monomer ratio may be sufficient to ensure that eq 35 holds.

3.6. Working Equations and Scaling Behavior. The following working equations for monomer loss and aggregate growth kinetics correspond to eqs 21, 22–25, 31, and 33 after converting from dimensionless time. The effective k_{obs} parameter in each case is defined implicitly by inspection of its monomer loss rate equation. The expressions for $\langle \text{MW} \rangle_w$ each assume that $\sigma \rightarrow \sigma_{\text{ss}}$.

Type **Ib**

$$\frac{dm}{dt} = - \frac{n^* f_R^x (C_0/C_{\text{ref}})^{x-1}}{\tau_{\text{nuc}}^{(0)}} m^x = -k_{\text{obs}} m^x \quad (36)$$

Type **Ic** and Early Time Type **Id**

$$\frac{dm}{dt} \cong - \frac{x f_R^x (C_0/C_{\text{ref}})^{x-1}}{\tau_{\text{nuc}}^{(0)}} m^x = -k_{\text{obs}} m^x \quad (37a)$$

$$\frac{\langle \text{MW} \rangle_w}{\text{MW}_{\text{mon}}} \cong x - \frac{x-1}{(1+(x-1)k_{\text{obs}}t)^{1/(x-1)}} \quad (37b)$$

Type **II** and Early Time Type **IV**

$$\frac{dm}{dt} = - \left(\frac{C_0}{C_{\text{ref}}} \right)^{(x+\delta-1)/2} f_R^{(x+\delta)/2} \sqrt{\frac{\delta \sigma_{\text{int}}^2}{x^2 \tau_g^{(0)} \tau_{\text{nuc}}^{(0)}}} m^{\delta} = -k_{\text{obs}} m^{\delta} \quad (38a)$$

$$\frac{\langle \text{MW} \rangle_w}{\text{MW}_{\text{mon}}} \cong m(t) + \left[\frac{x(1+c_x^2) \sqrt{\delta} \left(\frac{\tau_{\text{n}}^{(0)}}{\tau_g^{(0)}} \right)^{1/2}}{\sigma_{\text{int}}} \times \left(\frac{C_0}{C_{\text{ref}}} \right)^{(1+\delta-x)/2} f_R^{(\delta-x)/2} \right] [1-m(t)]^2 \quad (38b)$$

Types **Ib**, **Ic**, and **Id** all show apparent reaction orders with respect to m of $\nu = x$. For types **II** and **IV** over approximately the first half-life, the apparent reaction order for monomer loss or conversion to aggregate is 1 when growth occurs by monomer addition ($\delta = 1$). Comparing the β_{gn} range in Figure 3 and the scaling of $(\langle \text{MW} \rangle_w / \text{MW}_{\text{mon}})_{m \rightarrow 0}$ with β_{gn} in Figure 10, it can be seen that types **II** and **IV** cover much of the physically relevant parameter space when the resulting aggregates are composed of more than ca. 10 monomers. Together, these results highlight that caution must be taken when assuming that observed first-order kinetics alone permit one to conclude that a monomolecular step is rate-limiting in aggregation.

From eqs 36, 37a, and 38a, it is apparent that k_{obs} depends on several variables and parameters. The only variable that can

be independently controlled in experiments is C_0 . In types **Ib**, **Ic**, and early stage **Id**, k_{obs} scales as C_0^{x-1} . In types **II** and **IV**, k_{obs} scales as $C_0^{(x+\delta-1)/2}$. The contributions from both x and δ in the exponent of C_0 for types **II** and **IV** follow because the majority of monomer consumption is aggregate growth via polymerization, but nucleation rates at early times dictate the steady-state concentration of those aggregates. In other words, k_{obs} in these cases is predominantly a measure of growth once one considers more than the first few percent monomer loss,^{31,44} but the number concentration of aggregates upon which growth occurs is determined by nucleation.

The remaining model parameters within k_{obs} cannot be independently varied in experiment. Any change in temperature, pressure, or solvent conditions will potentially alter the free energy of unfolding, $\tau_g^{(0)}$, $\tau_{\text{n}}^{(0)}$, and possibly x and δ . These considerations notwithstanding, it should be noted that f_R depends exponentially on the free energy of unfolding under conditions where R is not the dominant monomer state, and k_{obs} scales with f_R to a power greater than 1 in all cases with $x > \delta$. Therefore, although folding–unfolding thermodynamics are never the sole determinant of k_{obs} or t_{50} , changing the free energy of unfolding can provide one of the most effective means to quantitatively⁴⁶ or semiquantitatively¹ control relative aggregation rates at fixed C_0 .

Purely on the basis of monomer loss data, or equivalently on mass converted or extent-of-reaction data, one finds from eqs 36–38 that k_{obs} is necessarily a convolution of at least two contributions from among $\tau_g^{(0)}$, $\tau_{\text{n}}^{(0)}$, and f_R for a given x , δ , and C_0/C_{ref} . For type **Ib**, **Ic**, and **Id** kinetics, the intrinsic time scale of nucleation can be extracted from k_{obs} provided that f_R can be determined from folding–unfolding thermodynamics for the conditions of interest. However, no information regarding growth rates or δ is obtainable. bG-CSF at near neutral pH and elevated temperatures⁴⁶ provides an experimental example of predominantly type **Ib** behavior in which $x = 2$, and n^* is low due to limited aggregate solubility.

For type **II** and **IV** kinetics in the absence of seeding, our analysis and those of previous workers with limiting cases of the present model^{31,33,44,45} indicate that $\tau_g^{(0)}$ and $\tau_{\text{n}}^{(0)}$ always occur as a product in k_{obs} . This conclusion holds even for experiments in which apparent lag and growth phases occur, as the magnitude of the rate in the growth phase depends on the number concentration of nuclei formed during the nucleation phase. Thus, the same physics that cause the convolution of $\tau_g^{(0)}$ and $\tau_{\text{n}}^{(0)}$ in the kinetics described in this work will be operative in lag-and-growth experiments, independent of whether the lag phase is a true nucleation phenomenon.^{32,44} On the basis of the results in subsection 3.4 we also conclude that extent-of-reaction measurements alone cannot quantitatively deconvolute nucleation and growth time scales even if one seeds the system, unless one knows the number concentration of the seeds, and assures it is sufficiently high compared to the value of σ_{ss} in the absence of seeding.

All of the above results and discussion hold for $\Phi \approx 1$. If the monomers remain soluble to arbitrarily high concentration, then at sufficiently large C_0 values one will instead need to consider solutions of the more general LENP model (eqs 6–9) in which $\Phi < 1$ is possible. A detailed analysis of that model is beyond the scope of this report. However, many of the conclusions from previous reports^{30,54} on limiting cases for $\Phi < 1$ are expected to hold. For example, above some threshold protein concentration (c^*) essentially all R monomers will exist

as part of larger, reversible oligomers. If the dominant oligomer is R_x or R_{x-1} , and both x and n^* are large, then eqs C2 and C3 in Appendix C hold. For large β_{gn} values, an analytical solution to these equations at low extents of reaction³⁰ shows that the effective $k_{\text{obs}} \sim (\tau_n^{(0)} \tau_g^{(0)})^{-1/2}$ but is independent of C_0 . The solution without $f_R = 1$ shows that $k_{\text{obs}} \sim f_R^{(x+\delta)/2}$ (derivation not shown). Thus, the above discussion regarding the convolution of nucleation and growth time scales in type II kinetics and the dependence of k_{obs} or t_{50} on folding–unfolding thermodynamics is expected to quantitatively hold beyond the particular version of the LENP model analyzed in detail here. However, the precise scaling with C_0 can be significantly different at elevated C_0 .⁵⁴

Finally, examination of eqs 38a and 38b suggests a means to determine separate values for τ_g and τ_n from experimental data. If one independently combines an extent-of-reaction assay (e.g., monomer loss or mass percent converted to aggregates) with corresponding, quantitative $\langle \text{MW} \rangle_w / \text{MW}_{\text{mon}}$ data, it is possible to deconvolute τ_g and τ_n since they occur as a product in k_{obs} and as a quotient in the long-time value of $\langle \text{MW} \rangle_w / \text{MW}_{\text{mon}}$. However, the convolution of multiple parameters from stages 1 and 4 in τ_g and from stages 1, 2, and 3 in τ_n remains (cf. definitions of τ_g , τ_n , $\tau_g^{(0)}$, and $\tau_n^{(0)}$ in subsection 2.7). This analysis further highlights that information on both the kinetics of (mass) conversion and the quantitative details of the aggregate size distribution over a number of half-lives are needed to reliably extract both nucleation and growth time scales when aggregates remain soluble. When aggregates are essentially insoluble at low n^* , eq 36 shows that only τ_n is relevant to the experimental kinetics.

4. Summary

We have presented a mathematical model of non-native protein aggregation that generally and self-consistently incorporates monomer conformational changes, reversible self-association, and (net) irreversible aggregate nucleation and growth via polymerization. The model provides a reasonably comprehensive, computationally simple framework for qualitative and quantitative interpretation of experimental aggregation kinetics under conditions where detailed kinetics of folding or unfolding can be neglected. A number of previously reported models for protein or polypeptide aggregation are shown to be simpler limiting cases of this LENP model.

The global model behavior under the assumption of pre-equilibrated folding–unfolding displays a number of qualitatively distinct kinetic regimes that depend primarily on two model parameters: the ratio of nucleation to growth time scales, β_{gn} , and the threshold size n^* at which aggregate condensation becomes appreciable. The different kinetic types are distinguished experimentally in terms of: (1) apparent reaction order; (2) dependence on initial protein concentration and folding–unfolding thermodynamics; (3) the qualitative and quantitative features of the resulting aggregate size distributions.

The working equations for each kinetic regime provide expressions for estimating the dominant nucleus size from aggregation kinetics over a range of initial protein concentrations. For certain cases, the intrinsic time scale for nucleation can also be determined if monomer unfolding thermodynamics are known. In other cases, intrinsic time scales for both nucleation and growth can be obtained by combining extent-of-reaction kinetics and one or more moments of the resulting aggregate size distribution.

5.1. Appendix A

Here we allow aggregate–aggregate association for $j < n^*$ and treat each $A_i + A_j$ reaction step as effectively irreversible

with a size-independent forward rate coefficient for condensation (k_c)



This parameter is equivalent to the average $k_{i,j}$ value ($i, j > 1$) in ref 45. We show here the case of $\Phi = 1$ and $\delta = 1$. A dimensionless rate coefficient for condensation steps, κ_c , is

$$\kappa_c = \frac{k_c C_{\text{ref}}}{k_{\text{nuc}} K_{x-1} C_{\text{ref}}^{x-1} f_R^x} (C_0 / C_{\text{ref}})^{2-x} = \frac{\tau_n^{(0)}}{\tau_c^{(0)}} f_R^{-x} (C_0 / C_{\text{ref}})^{2-x} \quad (\text{A2})$$

For $k_r \gg k_d$ and $k_{r,x} \gg k_{d,x}$, taking $x = 2$, and equating $k_{a,x} = k_{1,1}$ and $k_a = \langle k_{1,j} \rangle$, this dimensionless parameter reduces to the product $f_R^{-2} \kappa_{ij}$ in ref 45 since $K_1 = 1$ by definition. As in ref 45, with reasonably large n^* ($> \text{ca. } 20$) one observes two different cases under steady-state conditions. If $\kappa_c / \beta_{\text{gn}} < \text{ca. } 1$, then one recovers type II behavior. However, for larger $\kappa_c / \beta_{\text{gn}}$ each a_j rapidly achieves a steady state (i.e., $da_j/d\theta_n = 0$), and one can approximate

$$0 \approx m^x - \frac{\beta_{\text{gn}}}{n^* - x} m \sigma - \frac{\alpha}{2} \kappa_c \sigma^2 \quad (\text{A3})$$

Since $\beta_{\text{gn}} / \kappa_c \ll 1$ and $(n^* - x) \gg 1$ the term linear in σ is neglected, and we have

$$\sigma \approx \kappa_c^{-1/2} m^{x/2} \quad (\text{A4})$$

Therefore eq 12 becomes

$$\frac{dm}{d\theta_n} = -x m^x - \beta_{\text{gn}} \kappa_c^{-1/2} m^{1+x/2} \quad (\text{A5})$$

Converting from dimensionless form gives

$$\frac{dm}{dt} = - \left[\frac{x f_R^x \left(\frac{C_0}{C_{\text{ref}}} \right)^{x-1}}{\tau_n^{(0)} \left(\frac{C_{\text{ref}}}{C_0} \right)} \right] m^x \quad (\text{for } \beta_{\text{gn}} < \text{ca. } 1) \quad (\text{A6})$$

$$\frac{dm}{dt} = - \left[f_R^{(1+x/2)} \left(\frac{C_0}{C_{\text{ref}}} \right)^{x/2} \frac{\sqrt{\tau_c^{(0)} / \tau_n^{(0)}}}{\tau_g^{(0)}} \right] m^{1+x/2} \quad (\text{for } \beta_{\text{gn}} \gg 1) \quad (\text{A7})$$

In eqs A6 and A7, the bracketed terms indicate the k_{obs} coefficients that one would observe for kinetics measured at fixed temperature, pressure, solvent composition, and a given C_0 . Equation A6 is mathematically and physically equivalent to type Ib behavior. For eq A7, if $x = 2$, then it is equivalent to type Ia kinetics in ref 45 except that ref 45 used a simpler limiting case in which $k_{\text{nuc}} = k_{a,x} = k_{1,1}$. For $x = 2$, both eqs A6 and A7 produce k_{obs} with the same dependence on f_R (or $\Delta G_{\text{un}}^\circ$) and C_0 .

5.2. Appendix B

The derivation below employs $\delta = 2$ for concreteness, while illustrating the general procedure for incorporating $\delta > 1$. Taking the derivative of eq 4 with respect to time and realizing that all reversible steps will cancel when summed over all i and j indices gives

$$\frac{dC_M}{dt} = -xk_{r,x}[R_x] - k_r \sum_{j=x}^{n^*-1} 2[A_jR_2] \quad (B1)$$

The individual balance equations for those species not assumed to be pre-equilibrated are

$$\left. \frac{d[A_jR]}{dt} \right|_{x \leq j < n^*} = k_a[R]([A_j] - [A_jR]) - k_d([A_jR] - [A_jR_2]) \quad (B2)$$

$$\left. \frac{d[A_jR_2]}{dt} \right|_{x \leq j < n^*} = k_a[R][A_jR] - k_d[A_jR_2] - k_r[A_jR_2] \quad (B3)$$

$$\left. \frac{d[A_j]}{dt} \right|_{x \leq j < n^*} = -k_a[R][A_j] + k_d[A_jR] + k_r[A_{j-2}R_2] \quad (B4)$$

$$\frac{d[A_x]}{dt} = k_{r,x}[R_x] - k_a[R][A_x] + k_d[A_xR] \quad (B5)$$

$$\frac{d[R_x]}{dt} = k_{a,x}[R_{x-1}][R] - (k_{r,x} + k_{d,x})[R_x] \quad (B6)$$

Taking $[R_x]$ and each $[A_jR]$ and $[A_jR_2]$ to be at steady state gives eqs B7–B9 with the substitution of $K_{RA} = k_a/k_d$

$$[R_x]_{ss} = \frac{k_{a,x}[R][R_{x-1}]}{k_{d,x} + k_{r,x}} \quad (B7)$$

$$[A_jR]_{ss} = \frac{K_{RA}[R][A_j]}{1 + \frac{K_{RA}[R](k_r/k_d)}{1 + (k_r/k_d)}} \cong K_{RA}[R][A_j] \quad (B8)$$

$$[A_jR_2]_{ss} = \frac{K_{RA}^2[R]^2[A_j]}{1 + (1 + K_{RA}[R])(k_r/k_d)} \cong \frac{k_a}{k_d + k_r} K_{RA}[R]^2[A_j] \quad (B9)$$

The rightmost expressions in eqs B8 and B9 hold at conditions similar to those for which the $\Phi = 1$ approximation is valid. Equations B8 and B9 are generalized to larger δ using

$$[A_jR_i]_{ss} \cong K_{RA}^i [R]^i [A_j] \quad (B10)$$

$$[A_jR_\delta]_{ss} \cong \frac{k_a}{k_d + k_r} K_{RA}^{\delta-1} [R]^\delta [A_j] \quad (B11)$$

Equation B11 holds for $\delta = 1$ even without the assumption of $\Phi \rightarrow 1$. Substituting eqs B7–B9 into eqs B1, B4, and B5 and rearranging using the dimensionless variables defined in section 2 gives eqs 6–9 with the substitution $[R] = f_R \Phi c_m C_0$.

5.3. Appendix C

Equations 6–9 reduce to a model that is very similar to that presented by Lomakin et al. for Alzheimer's $A\beta$ aggregation kinetics³⁰ if one assumes that all free monomers exist in the reactive conformation ($f_R = 1$), all A_jR concentrations and all R_i concentrations other than R_x are arbitrarily small, $k_{r,x} \ll k_{d,x}$, and $n^* \rightarrow \infty$. With these simplifications and the identity $m = c_m \Phi$, eq 4 simplifies to

$$m^x \left(\frac{C_0}{c^*} \right)^{x-1} + m = c_m \quad (C1)$$

In eq C1, $c^* \equiv [xK_x]^{-1/(x-1)}$ effectively defines a threshold monomer concentration at which $[R_x] = [R]$. Imposing $x \gg 1$ yields distinctive behaviors of this model.³⁰ For example, at low initial protein concentrations ($C_0 < \text{ca. } c^*$) the first term in eq C1 is essentially zero; $c_m C_0$ therefore equals the free monomer concentration mC_0 , and the concentration of nuclei at any given time is negligible. Significant aggregation therefore occurs only via seeded growth.³⁰

In the alternative regime of high initial protein concentration ($C_0 > c^*$), at early times (i.e., $c_m \approx 1$) the first term dominates the second in eq C1, $m \approx c^*/C_0$, and eqs 6 and 9 become

$$\frac{dc_m}{d\theta_n} = -\frac{1}{K_x C_0^{x-1}} c_m - \beta_{gn} c^* C_0^{-1} \sigma \quad (C2)$$

$$\frac{d\sigma}{d\theta_n} = \frac{1}{xK_x C_0^{x-1}} c_m \quad (C3)$$

Acknowledgment. The authors gratefully acknowledge financial support from the National Institutes of Health (Grant No. P20 RR-015588), Boehringer-Ingelheim Pharmaceuticals, and the University of Delaware Research Foundation.

Supporting Information Available: Locations of boundaries in state diagrams. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F. *Pharm. Res.* **2003**, *20*, 1325.
- Cleland, J. L.; Powell, M. F.; Shire, S. J. *Crit. Rev. Ther. Drug Carrier Syst.* **1993**, *10*, 307.
- Cromwell, M. E. M.; Hilario, E.; Jacobson, F. *AAPS J.* **2006**, *8*, E572.
- Purohit, V. S.; Middaugh, C. R.; Balasubramanian, S. V. *J. Pharm. Sci.* **2006**, *95*, 358.
- Remmele, R. L.; Bhat, S. D.; Phan, D. H.; Gombotz, W. R. *Biochemistry* **1999**, *38*, 5241.
- Rosenberg, A. S. *AAPS J.* **2006**, *8*, E501.
- Shire, S. J.; Shahrokh, Z.; Liu, J. J. *Pharm. Sci.* **2004**, *93*, 1390.
- Wang, W. *Int. J. Pharm.* **2005**, *289*, 1.
- Damas, A. M.; Saraiva, M. J. *J. Struct. Biol.* **2000**, *130*, 290.
- Dobson, C. M. *Protein Pept. Lett.* **2006**, *13*, 219.
- Fink, A. L. *Folding Des.* **1998**, *3*, R9.
- Lansbury, P. T.; Lashuel, H. A. *Nature* **2006**, *443*, 774.
- Murphy, R. M. *Annu. Rev. Biomed. Eng.* **2002**, *4*, 155.
- Ross, C. A.; Poirier, M. A. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 891.
- Cox, D. L.; Lashuel, H.; Lee, K. Y. C.; Singh, R. R. P. *MRS Bull.* **2005**, *30*, 452.
- Haines, L. A.; Rajagopal, K.; Ozbass, B.; Salick, D. A.; Pochan, D. J.; Schneider, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 17025.
- Dobson, C. M. *Semin. Cell Dev. Biol.* **2004**, *15*, 3.
- Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riekel, C. *Acc. Chem. Res.* **2006**, *39*, 568.
- Rousseau, F.; Schymkowitz, J.; Serrano, L. *Curr. Opin. Struct. Biol.* **2006**, *16*, 118.
- Dong, A.; Randolph, T. W.; Carpenter, J. F. *J. Biol. Chem.* **2000**, *275*, 27689.
- Finke, J. M.; Roy, M.; Zimm, B. H.; Jennings, P. A. *Biochemistry* **2000**, *39*, 575.
- Goers, J.; Permyakov, S. E.; Permyakov, E. A.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2002**, *41*, 12546.
- Grillo, A.; Edwards, K.; Kashi, R.; Shipley, K.; Hu, L.; Besman, M.; Middaugh, C. *Biochemistry* **2001**, *40*, 586.
- Khurana, R.; Gillespie, J. R.; Talapatra, A.; Minert, L. J.; Ionescu-Zanetti, C.; Millett, I.; Fink, A. L. *Biochemistry* **2001**, *40*, 3525.
- Stork, M.; Giese, A.; Kretschmar, H. A.; Tavan, P. *Biophys. J.* **2005**, *88*, 2442.

- (26) Von Bergen, M.; Barghorn, S.; Biernat, J.; Mandelkow, E.-M.; Mandelkow, E. *Biochim. Biophys. Acta* **2005**, *1739*, 158.
- (27) Wiseman, R. L.; Powers, E. T.; Kelly, J. W. *Biochemistry* **2005**, *44*, 16612.
- (28) Uversky, V. N.; Fink, A. L. *Biochim. Biophys. Acta* **2004**, *1698*, 131.
- (29) Ferrone, F. A. *J. Mol. Biol.* **1985**, *183*, 591.
- (30) Lomakin, A.; Teplow, D. B.; Kirschner, D. A.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7942.
- (31) Oosawa, F.; Asakura, S. *Thermodynamics of the Polymerization of Protein*; Academic Press: London, 1975.
- (32) Pallitto, M. M.; Murphy, R. M. *Biophys. J.* **2001**, *81*, 1805.
- (33) Roberts, C. J. Non-Native Protein Aggregation. In *Misbehaving Proteins: Protein (Mis)folding, Aggregation, and Stability*; Murphy, R. M., Tsai, A. M., Eds.; Springer: New York, 2006; p 17.
- (34) Harper, J. D.; Lansbury, P. T. *Annu. Rev. Biochem.* **1997**, *66*, 385.
- (35) Hong, D. P.; Ahmad, A.; Fink, A. L. *Biochemistry* **2006**, *45*, 9342.
- (36) Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2001**, *40*, 6036.
- (37) O'Neill, B.; Williams, A. D.; Westmark, P.; Wetzel, R. J. *Biol. Chem.* **2004**, *279*, 17490.
- (38) Wood, S. J.; Wypych, J.; Steavenson, S.; Louis, J.-C.; Citron, M.; Biere, A. L. *J. Biol. Chem.* **1999**, *274*, 19509.
- (39) Chen, S. M.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11884.
- (40) Ignatova, Z.; Gierasch, L. M. *Biochemistry* **2005**, *44*, 7266.
- (41) Brito, R. M. M.; Damas, A. M.; Saraiva, M. J. *Curr. Med. Chem.: Immunol., Endocr. Metab. Agents* **2003**, *3*, 349.
- (42) Hurshman, A. R.; White, J. T.; Powers, E. T.; Kelly, J. W. *Biochemistry* **2004**, *43*, 7365.
- (43) Kendrick, B. S.; Cleland, J. L.; Lam, X.; Nguyen, T.; Randolph, T. W.; Manning, M. C.; Carpenter, J. F. *J. Pharm. Sci.* **1998**, *87*, 1069.
- (44) Ferrone, F. *Methods Enzymol.* **1999**, *309*, 256.
- (45) Roberts, C. J. *J. Phys. Chem. B* **2003**, *107*, 1194.
- (46) Roberts, C. J.; Darrington, R. T.; Whitley, M. B. *J. Pharm. Sci.* **2003**, *92*, 1095.
- (47) Lumry, R.; Eyring, H. *J. Phys. Chem.* **1954**, *58*, 110.
- (48) Sanchez-Ruiz, J. M. *Biophys. J.* **1992**, *61*, 921.
- (49) Zale, S. E.; Klibanov, A. M. *Biotechnol. Bioeng.* **1983**, *25*, 2221.
- (50) Gibson, T. J.; Murphy, R. M. *Biochemistry* **2005**, *44*, 8898.
- (51) Liu, L.; Murphy, R. M. *Biochemistry* **2006**, *45*, 15702.
- (52) Flyvbjerg, H.; Jobs, E.; Leibler, S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5975.
- (53) Goldstein, R. F.; Stryer, L. *Biophys. J.* **1986**, *50*, 583.
- (54) Powers, E. T.; Powers, D. L. *Biophys. J.* **2006**, *91*, 122.
- (55) Kurganov, B. I.; Rafikova, E. R.; Dobrov, E. N. *Biochemistry (Moscow)* **2002**, *67*, 525.
- (56) Wang, K. Y.; Kurganov, B. I. *Biophys. Chem.* **2003**, *106*, 97.
- (57) Buswell, A. M.; Middelberg, A. P. J. *Biotechnol. Bioeng.* **2003**, *83*, 567.
- (58) Speed, M. A.; King, J.; Wang, D. I. C. *Biotechnol. Bioeng.* **1997**, *54*, 333.
- (59) Gast, K.; Modler, A. J.; Damaschun, H.; Krober, R.; Lutsch, G.; Zirwer, D.; Golbik, R.; Damaschun, G. *Eur. Biophys. J.* **2003**, *32*, 710.
- (60) Chi, E. Y.; Kendrick, B. S.; Carpenter, J. F.; Randolph, T. W. *J. Pharm. Sci.* **2005**, *94*, 2735.
- (61) Tartaglia, G. G.; Cavalli, A.; Pellarin, R.; Caflisch, A. *Protein Sci.* **2005**, *14*, 2723.
- (62) DuBay, K. F.; Pawar, A. P.; Chiti, F.; Zurdo, J.; Dobson, C. M.; Vendruscolo, M. *J. Mol. Biol.* **2004**, *341*, 1317.
- (63) Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. *Nat. Biotechnol.* **2004**, *22*, 1302.
- (64) Finkelstein, A. V. In *Slow Relaxations and Nonequilibrium Dynamics in Condensed Matter*, 77th NATO Advanced Study Institute Euro Summer School Ecole Thematique du CNRS, Les Houches, France, July 1–26, 2002; Springer: Berlin, 2003; p 649.
- (65) Privalov, P. J. *J. Mol. Biol.* **1996**, *258*, 707.
- (66) It has been shown for some systems that monomers can be recovered from non-native aggregates. However, this typically occurs appreciably only if the aggregates are subjected to much higher denaturant concentrations,⁷⁰ temperatures,⁷¹ or pressures^{72,73} than those at which they were formed.
- (67) Yoshioka, S.; Aso, Y.; Izutsu, K.; Kojima, S. *Pharm. Res.* **1994**, *11*, 1721.
- (68) Yoshioka, S.; Izutsu, K.; Aso, Y.; Takeda, Y. *Pharm. Res.* **1991**, *8*, 480.
- (69) Roussas, G. G. *A First Course in Mathematical Statistics*; Addison-Wesley: Reading, MA, 1973.
- (70) Callahan, N. A.; Xiong, L. W.; Caughey, B. J. *Biol. Chem.* **2001**, *276*, 28022.
- (71) Sasahara, K.; Hironobu, Naiki, and Yuji, Goto. *Biochemistry* **2006**, *45*, 8760.
- (72) Foguel, D.; Robinson, C. R.; de Sousa, P. C.; Silva, J. L.; Robinson, A. S. *Biotechnol. Bioeng.* **1999**, *63*, 552.
- (73) St John, R. J.; Carpenter, J. F.; Randolph, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13029–13033.