Kinetics of Irreversible Protein Aggregation: Analysis of Extended Lumry−Eyring Models and Implications for Predicting Protein Shelf Life

Christopher J. Roberts*

Department of Chemical Engineering, University of Delaware, Newark, Delaware 19716

Received: August 25, 2002; In Final Form: November 15, 2002

A general phenomenological kinetic model of irreversible protein aggregation in aqueous solution proceeding via non-native conformations is examined with a view toward the development of techniques to predict in vitro protein shelf life quantitatively. Approximate analytical solutions are derived for interpreting experimental data and agree with explicit numerical simulations over a broad range of physically relevant conditions. Results are presented in terms of the observed kinetics of monomer loss to provide a basis for direct comparison with typical experimental measures of aggregation as well as to provide a basis for quantitative prediction of protein shelf life. Similar to classic Lumry−Eyring models, experimentally observed aggregation kinetics are described as a combination of reversible conformational transitions and the intrinsic kinetics of aggregation via non-native states. However, in contrast to classic models, the extended Lumry−Eyring model developed here includes a detailed description of the intrinsic aggregation kinetics. The results indicate that apparent or observed aggregation kinetics can be grouped into one of four qualitatively distinct classes on the basis of their apparent reaction orders with respect to monomer concentration, each of which may be realized in practice depending on the experimental conditions and the relative rates of the intrinsic aggregation steps for the protein in question. The results also provide a basis for predicting shelf life by extrapolating high-temperature ("accelerated") aggregation kinetics to lower-temperature, long-term storage conditions. Illustrations are given within the context of a model pharmaceutical protein for which sufficient data are available. Generalizations to systems with more complex conformational transitions are discussed briefly.

I. Introduction

Irreversible protein inactivation occurs through a number of routes, including aggregation, adsorption, oxidation, deamidation, and hydrolysis.[1−3] For many proteins, aggregation is a dominant degradation route[2−12] and is important in a variety of different fields, including biotechnology,[12−14] food processing,[10,11] pharmaceuticals,[3−7] and biology and biochemistry.[8,9,15−19] There is also increasing evidence that a number of diseases due to protein deposition in vivo are inherently linked to changes in the kinetic stability of proteins relative to that of aggregated states.[9,16−18] As irreversible aggregates are ultimately lower in free energy than unaggregated protein in solution, quantitative models of the underlying kinetics are critical for the control and manipulation of aggregation-prone systems and may provide new insights into the nature of aggregation processes in general. Prediction and control of protein degradation kinetics, and ultimately of shelf life or storage stability, is also critical to rational design and selection of viable commercial protein products.[1−7]

Protein aggregation can be reversible or irreversible, depending on the conditions under which aggregates are formed. The present report is concerned with the kinetics of irreversible aggregation, referred to simply as aggregation, rather than the reversible situation, referred to here as self-association. Broadly speaking, the factors that control aggregation are qualitatively understood, but the prediction of aggregation rates remains problematic.[3,9,20−22] Presumably, many of the same factors[23−26] that direct protein folding, protein−ligand recognition, and protein association also determine aggregation propensities or rates. Hydrophobic or solvent-mediated interactions are specifically implicated in many cases, based in large part on structural and kinetic evidence[6−9,12,16,17,25,27−31] indicating that aggregation proceeds via the molten globule state or some other non-native conformational state[32] in which a degree of secondary structure is retained. Interestingly, the nature of the aggregation-prone state(s) appears to be quite diverse in terms of its proximity to the native state on the "folding landscape".[33,34] Examples exist to indicate the aggregative state may be slightly perturbed or near-native,[6,7,29,35] molten-globule-like,[28,36] a partly folded folding intermediate,[8,12,19,37,38] or even the "unfolded" state.[8,39] Therefore, it is not clear whether it is most appropriate to treat aggregation as proceeding through essentially one dominant "aggregative" conformational state for a given protein system or whether multiple states and subsequently multiple pathways should be considered.[25] However, in the absence of direct evidence to the contrary, the simpler case of a dominant aggregation-prone conformational state has been an effective approximation for many systems[6−8,19,28,29,35−38] and is used throughout this work.[43]

Aggregation kinetics for systems such as those alluded to above have typically been rationalized qualitatively on the

* Corresponding author. E-mail: roberts@che.udel.edu. Tel: 302-831-0838. Fax: 302-831-1048.

10.1021/jp026827s CCC: $25.00
Published on Web 01/14/2003

basis of so-called Lumry−Eyring models[44−46] that consider how irreversible processes (e.g., aggregation) that proceed through some non-native state affect the observed or apparent thermodynamics of reversible protein conformational transitions or self-association.[46−48] Historically, the "reactive" state with respect to the (unspecified) irreversible process was assumed to be the unfolded state *U*, although knowledge of the precise identity of the reactive state was not needed to answer the types of questions typically addressed. More recently, this qualitative picture was extended to treat quantitatively the effects of irreversible aggregation on the observed or apparent thermodynamics of folding−unfolding transitions in thermal unfolding experiments for systems in which protein unfolding becomes the rate-determining step at high temperature.[46−48] The qualitative effects of temperature, pressure, and added cosolvents/solutes on protein aggregation kinetics have also been rationalized using similar approaches.[7,10,35,49,50]

A fundamental limitation of classic Lumry−Eyring models from the perspective of shelf-life prediction[51,52] is their neglect of the detailed kinetics of the irreversible step(s) (e.g., aggregation). In this report, an extended Lumry−Eyring description of protein aggregation proceeding through an aggregation-prone or "reactive" state R_A is developed. Specifically, the model includes both the detailed kinetics of conformationally mediated irreversible aggregation and the effects of reversible conformational processes. The extended Lumry−Eyring model is analyzed in terms of the different qualitative types of kinetic behavior that may arise and is used to derive approximate analytical expressions to describe experimental aggregation kinetics in terms of the monomer concentration (M) as a function of time. The observed kinetics take one of four distinct mathematical forms, depending on the details of the intrinsic aggregation kinetics as well as shifts in conformational equilibrium between aggregation-prone and "inert" conformational states. From a practical standpoint, it is important to distinguish the different types of kinetic behavior when attempting to make shelf-life predictions (e.g., by extrapolating from high to low T). Such predictions have historically been problematic due at least in part to highly non-Arrhenius kinetics over narrow T ranges. Recently, the model developed here was employed to enable quantitative shelf-life predictions for a model pharmaceutical protein that exhibits such non-Arrhenius behavior.[51]

In general, the results here and in a separate report[51] suggest that protein aggregation kinetics under physiological and commercial storage conditions may be predictable from knowledge of the kinetics at more experimentally tractable conditions, provided both the thermodynamic (conformational equilibrium) and intrinsic kinetic components can be accurately measured or calculated. Furthermore, the results indicate that the most important physicochemical quantities for predicting shelf life in such systems are: k_{11}, the intrinsic rate coefficient for the aggregation of two R_A monomers; K_{NR}, the equilibrium constant describing the $N-R_A$ transition; and in some cases $\langle k_{1j} \rangle$, the average intrinsic rate coefficient for aggregation involving R_A molecules and intermediate or higher molecular weight soluble aggregates.

II. Modeling

The reaction scheme of a general model of irreversible aggregation is shown below in eq 1.

$$
\begin{aligned}
& N \overset{K_{NR}}{\longleftrightarrow} R_A \\
& K_{NR} = \left(\frac{R_A}{N}\right)_{\text{equil}} = \frac{k_{\text{u}}}{k_{\text{f}}} \\
& R_A + R_A \xrightarrow{k_{1,1}} A^{(2)} \\
& A^{(2)} + R_A \xrightarrow{k_{1,2}} A^{(3)} \\
& A^{(3)} + R_A \xrightarrow{k_{1,3}} A^{(4)} \\
& \quad \vdots \\
& A^{(n^*-1)} + R_A \xrightarrow{k_{1,n^*-1}} A^{(n^*)} \\
& A^{(2)} + A^{(2)} \xrightarrow{k_{2,2}} A^{(4)} \\
& A^{(2)} + A^{(3)} \xrightarrow{k_{2,3}} A^{(4)} \\
& \quad \vdots \\
& A^{(i)} + A^{(j)} \xrightarrow{k_{i,j}} A^{(i+j)} \qquad i, j < n^*
\end{aligned} \tag{1}
$$

N and R_A are the monomeric protein in the native and aggregation-prone (non-native) states, respectively. K_{NR} is the equilibrium constant for the $N-R_A$ transition at a given set of experimental conditions. $A^{(j)}$'s are soluble aggregates consisting of j monomers (i.e., "j-mers"). On the time scale of interest, R_A is the only monomeric state prone to aggregate appreciably. The exact structural identity of R_A is deliberately not specified, although as mentioned in section I there is significant indirect evidence that secondary structure in non-native states predisposes them toward aggregation. Such considerations notwithstanding, eq 1 remains general so long as one defines R_A appropriately as the (sub)ensemble of molecular arrangements (protein and solvent) that constitutes the aggregation-prone conformational state. R_A molecules can aggregate with other R_A molecules and with already formed j-mers. k_{ij} is the intrinsic rate constant for the aggregation of an i-mer with a j-mer ($i = 1$ is equivalent to R_A, and $k_{ij} = k_{ji}$ by definition). Protein association is not included above but can be incorporated. In addition, eq 1 explicitly assumes irreversible aggregation events. This restriction can be relaxed, but the resulting derivations become more complicated without illustrating significantly different physics under conditions of interest here (see also section IV). n^* defines the size cutoff for protein aggregates having appreciable solubility or reactivity with respect to aggregation and is expected to depend on the solution conditions. For example, solubility and therefore n^* typically decrease as pH approaches the pI or as solution ionic strength increases.[53] Changing n^* can have interesting consequences that will be discussed further in sections III and IV. Alternatively, n^* may be determined by a kinetic "bottleneck" in that j-mers with $j \geq n^*$ are essentially unreactive with respect to aggregation but remain in solution.[21,54]

From eq 1, the following set of differential equations with respect to time (t) may be written.

$$\frac{dN}{dt} = -k_u(N - K_{NR}^{-1}R_A) \quad (2.1)$$

$$\frac{dR_A}{dt} = k_u(N - K_{NR}^{-1}R_A) - 2k_{1,1}R_A^2 - (\sum_{j=2}^{n^*-1} k_{1,j}A^{(j)})R_A \quad (2.2)$$

$$\frac{dM}{dt} = \frac{dN}{dt} + \frac{dR_A}{dt} \quad (2.3)$$

$$\frac{dA^{(2)}}{dt} = k_{1,1}R_A^2 - k_{1,2}A^{(2)}R_A - 2k_{2,2}(A^{(2)})^2 - (\sum_{j=3}^{n^*-1} k_{2,j}A^{(j)})A^{(2)} \quad (2.4)$$

$$\vdots$$

$$\frac{dA^{(n^*-1)}}{dt} = k_{1,n^*-2}A^{(n^*-2)}R_A + \sum_{j=2}^{(n^*-1)/2} k_{j,(n^*-1-j)}A^{(j)}A^{(n^*-1-j)} - k_{1,n^*-1}A^{(n^*-1)}R_A - (\sum_{j=2}^{n^*-2} k_{j,n^*-1}A^{(j)})A^{(n^*-1)} - 2k_{n^*-1,n^*-1}(A^{(n^*-1)})^2 \quad (2.n^*+2)$$

The two terms in eq 2.1 represent the forward and reverse reactions for the conformational equilibrium between N and R_A states. The equilibrium conditions are governed by K_{NR}, which also determines the relative values of the folding and unfolding (or reverse and forward) first-order rate coefficients, k_f and k_u, as a function of T. No a priori assumption is needed as to whether this equilibrium is ever established. However, for conditions of practical interest (e.g., those favoring N), equilibrium is established much faster than the time scale for aggregation.[51,54] Specifically, conformational equilibrium is expected to hold for T below a crossover temperature T_x, above which unfolding becomes the rate-determining step.[51] Empirically, T_x is close to or greater than T_M (the midpoint unfolding temperature[55,56]) for many proteins under practical conditions.[37,46−48] In the remainder of this paper, equilibrium between N and R_A is assumed to be maintained throughout the aggregation process. This assumption is expected to hold for systems of practical interest from the standpoint of long-term storage[59] but may be violated for proteins subjected to extreme conditions (e.g., $T > T_x$).[6,46−48] The third term in eq 2.2 corresponds to the aggregation of two R_A monomers. The fourth term is due to the aggregation of R_A with aggregates that formed earlier or were added to the system. Similar terms are present in eqs 2.4 to 2.(n^*+2). For a system initially devoid of aggregates, all $A^{(j>1)}$ must of course accumulate initially as a result of the aggregation of two R_A molecules. Equation 2.3 states that we typically cannot distinguish between N and R_A monomers experimentally (e.g., in a size-exclusion HPLC assay); therefore, we observe the total (soluble) monomer, M.

In addition, any or all of the aggregate species may have finite solubility limits below the levels reached during aggregation. This constraint can be captured in eq 2 by imposing a limit on each aggregate species at its corresponding saturation level, $A^{(j),SAT}$, under the solution conditions of interest.[53] This is to be distinguished from the influence of solubility on determining n^* discussed above, for which

$$A^{(j\geq n^*),SAT} \rightarrow 0 \quad (3)$$

Considering the highly specific conditions under which native proteins crystallize,[58] we expect precipitated aggregates to appear more as amorphous, flocculated colloids than as ordered structures.[54] High molecular weight, fibrillar, and amyloid precipitates are obvious exceptions to this.[9,60] For use later in this section, we introduce one additional model parameter, j', which is defined as the lowest value of j for which $A^{(j)}$ levels would exceed their saturation point during the course of aggregation if precipitation did not intervene. Note that j' is to be distinguished from n^* in that $A^{(j'),sat}$ is not zero, and $j' \rightarrow n^*$ if $A^{(j'),sat} \rightarrow 0$.

Equations 2.1 and 2.2 are combined to give

$$\frac{dM}{dt} = -2k_{11}R_A^2 - (\sum_{j=2}^{n^*-1} k_{1j}A^{(j)})R_A \quad (4)$$

Writing eq 4 in terms of M (the experimentally measurable quantity) rather than R_A gives

$$\frac{dM}{dt} = -2k_{11}f_R^2M^2 - f_R(\sum_{j=2}^{n^*-1} k_{1j}A^{(j)})M \quad (5)$$

where f_R is the fraction of M existing as R_A and is given by

$$f_R \equiv \frac{K_{NR}}{1 + K_{NR}} \quad (6)$$

It is useful at this point to cast the model equations in dimensionless form. Using M_0 ($\equiv M_{t=0}$) as the characteristic concentration and $(k_{11}M_0 f_R^2)^{-1}$ as the characteristic time scale for the system gives

$$\frac{dm}{d\tau_{agg}} = -2m^2 - f_R^{-1}(\sum_{j=2}^{n^*-1} k'_{1j}a_j)m \quad (7)$$

$$\frac{da_j}{d\tau_{agg}}|_{2\leq j<n^*} = f_R^{-1} \times [(\sum_{\substack{\nu<w \\ \nu+w=j}} k'_{\nu w}a_\nu a_w)f_R^{-1} - (\sum_{i=2}^{n^*-1} k'_{ij}a_i)a_j f_R^{-1} - k'_{jj}a_j^2 f_R^{-1} - k'_{1j}a_j m] \quad (8)$$

where $m \equiv M/M_0$, $\tau_{agg} \equiv k_{11}M_0 f_R^2 t$, $k'_{ij} \equiv k_{ij}/k_{11}$, and $a_j \equiv A^{(j)}/M_0$. Equation 7 and 8 are exact, but the second term on the right-hand side of eq 7 is problematic with respect to analyzing experimental data since the set $\{a_j\}$ is not typically known within a reasonable degree of accuracy. There are, however, a number of approximate analytical forms to eq 7 that are more amenable for use with typical experimental data, and these will be derived in the remainder of this section.

II.1. Approximate Solutions

Type I Behavior. For systems in which k_{ij} is only weakly dependent on i and j ($ij \neq 1, 1$), we replace k'_{ij} in eq 7 and 8 with its average value κ_{ij} ($\equiv \langle k'_{ij} \rangle$). Furthermore, for n^* and/or κ_{ij} significantly greater than 1, a_j levels will reach pseudo-steady state levels rapidly compared to the overall time scale for aggregation (i.e., at $\tau_{agg} \ll 1$), and so we may approximate $da_j/d\tau_{agg} \approx 0$. Summing eq 8 over $2 \leq j < n^*$ and collecting terms gives (Appendix A)

$$0 \cong m^2 - \frac{\alpha_I}{2}\kappa_{ij} f_R^{-2}\left(\sum_{j=2}^{n^*-1} a_j\right)^2 - \frac{\kappa_{1j}}{n^*-2} f_R^{-1}\left(\sum_{j=2}^{n^*-1} a_j\right)m \quad (9)$$

Equation 9 is quadratic in the summation over a_j and thus can be solved to give

$$\sum_{j=2}^{n^*-1} a_j \cong f_R\left(\frac{2}{\alpha_I\kappa_{ij}}\right)^{1/2} m \qquad \text{if } \kappa_{ij}/\kappa_{1j} > \text{ca. 1 and/or } n^* \to \infty \quad (10a)$$

$$\sum_{j=2}^{n^*-1} a_j \cong f_R\frac{(n^*-2)}{\kappa_{1j}}m \qquad \text{if } \kappa_{ij}/\kappa_{1j} \ll 1 \quad (10b)$$

where α_I is a constant of order 1 in eq 10a and κ_{1j} is the equivalent of κ_{ij} with the restriction that one of the reactants is an R_A molecule ($i = 1$). Inserting eq 10a in eq 7 gives

$$\frac{dm}{d\tau_{agg}} = -2\left(1 + \sqrt{\frac{\kappa_{ij}}{2\alpha_I}}\right)m^2 \quad (11a)$$

Inserting eq 10b into eq 7 gives

$$\frac{dm}{d\tau_{agg}} = -n^*m^2 \quad (11b)$$

Equation 11a describes kinetic behavior that we will denote as Type IA whereas that characterized by eq 11b will be denoted Type IB. Physically, Type IA behavior occurs when all a_j remain soluble (for $j < n^*$), when the intrinsic kinetic barriers to aggregation are similar for any step involving a previously formed aggregate (i.e., k_{ij} is essentially independent of ij, k_{11} excepted), and when $\kappa_{ij} \approx 1$ or higher. In this case, pseudo-steady-state a_j levels are effectively determined by m (actually R_A, which is proportional to m), cf. eq 10a, and thus the steady-state kinetic profiles appear as pseudo-second order with respect to m. Note that this result is quite similar to that obtained for radical polymerization with a second-order initiation step.[57] The main difference is that combining an i-mer with a j-mer does not constitute a termination step in the above model, so we have an added coefficient of $\alpha_I^{-1/2}$. Type IB behavior is a subset of Type IA behavior for the case in which soluble aggregates are reactive with R_A molecules but not with other aggregates ($k'_{ij} \approx 0$ unless i or j is 1). In this case, the pseudo-steady-state levels are simply

$$a_{j,ss} \approx \frac{f_R}{\kappa_{1j}}m \quad (12)$$

and each j-mer grows by the sequential addition of R_A molecules.

Note that if $\kappa_{ij} \ll 1$ one expects that not all a_j levels will reach steady-state values until $\tau_{agg} \gg 1$. Concomitantly, at least some of the lower molecular weight j-mers will achieve levels comparable to m on the experimental time scale and therefore will also be measurable by the same technique(s) used to quantitate m. In this case, one should exclude those species from the summation(s) in, for example, eqs 7 and 9, and should treat those species explicitly as was done above for m and R_A.

Type II Behavior. Equation 7 can be rearranged in the following form:

$$\frac{dm}{d\tau_{agg}} = -\left[2m + f_R^{-1}\sum_{j=2}^{n^*-1} k'_{1j}a_j\right]m \quad (13)$$

If the second term in the brackets in eq 13 is dominant, then eq 13 will be well approximated by

$$\frac{dm}{d\tau_{agg}} = -f_R^{-1}\kappa_{1j}\left(\sum_{j=2}^{n^*-1} a_j\right)_{ss}m \quad (14)$$

where again each k'_{1j} has been replaced with its average value. Of course, the term in parentheses is a constant only after an initial lag time before reaching steady state (denoted by the subscript ss).

Physically, Type II behavior does not occur because aggregation steps that consume monomer by addition to preexisting aggregates are dominant, as this in fact occurs for Types IA and IB as well. Rather, it occurs because the overall concentration of those aggregates is so large compared to R_A itself that addition of monomer causes a relatively minor change in their overall level. Thus, the loss of monomer is primarily via addition to a pseudoconstant "background" concentration (on a number, not size basis) of aggregates.

Type III Behavior. Type I and II behaviors assume that aggregates remain at levels below their respective solubility limits for *all* lower molecular weight species (i.e., all $j < n^*$) over the time scale of interest. When n^* is determined by solubility limits, we have the further restriction that all j-mers larger than n^* have essentially zero solubility (relative to overall monomer levels). We may relax these assumptions by realizing that in many systems j-mer solubility limits ($a_{j,sat}$) will be finite and will lie significantly below the maximum levels ($a_{j,max}$) that would be achieved solely as a result of the competition between consecutive reaction steps that produce and consume each j-mer species. As mentioned earlier, j' is by definition the lowest value of j for which $a_{j,sat}$ is significantly below $a_{j,max}$ for a given set of κ_{ij}, f_R, and n^*. Equation 7 is then rearranged to read

$$\frac{dm}{d\tau_{agg}} = -2m^2 - f_R^{-1}\kappa_{1j}\left(\sum_{j<j'} a_j\right)m - f_R^{-1}\kappa_{1j}\left(\sum_{j\geq j'}^{n^*-1} a_j\right)m \quad (15)$$

Equation 15 simplifies greatly under steady-state conditions for the following reasons. First, the summation over $j < j'$ can be approximated in a manner similar to that done in the derivation of eq 11, with α_I being replaced by α_{III} since the number of terms in the summation over a_j will in general be different if j' is significantly less than n^*. Second, the summation over $j \geq j'$ will be essentially constant once $a_{j'}$ reaches its saturation value. The resulting expressions are

$$\frac{dm}{d\tau_{agg}} = -2\left[1 + \sqrt{2\kappa_{ij}/\alpha_{III}}\right]m^2 - \kappa_h^{(app)}m \quad (16a)$$

or

$$\frac{dm}{d\tau_{agg}} = -2\left[\frac{j'}{2}\right]m^2 - \kappa_h^{(app)}m \quad (16b)$$

with

$$\kappa_h^{(app)} \equiv f_R^{-1}\kappa_{1j}(n^* - j')a_{j',sat} \quad (17)$$

In the above expressions we have taken the simplest case in which only $a_{j'}$ is constrained by a finite solubility limit. It is straightforward to extend this description to the case where multiple species reach saturation limits merely by separating the summation over $j \geq j'$ into separate summations with upper and lower bounds determined by the identity of those j-mers

that reach their saturation limits. Such added complexity does not change the qualitative results and is expected to have only minor quantitative effects for realistic parameter values.

Finally, converting eqs 11, 14, and 16 from a dimensionless form to their equivalent expressions for $M(t)$ gives

$$\frac{dM}{dt} = -2k_{11}^{(app)}\left(1 + \sqrt{\frac{2\kappa_{ij}}{\alpha_I}}\right)M^2 \qquad \text{(Type IA)} \quad (18a)$$

$$\frac{dM}{dt} = -2k_{11}^{(app)}\left(\frac{n^*}{2}\right)M^2 \qquad \text{(Type IB)} \quad (18b)$$

$$\frac{dM}{dt} = -k_1^{(app)}M \qquad \text{(Type II)} \quad (19)$$

$$k_1^{(app)} = f_R\langle k_{1j}\rangle\left(\sum_{j=2}^{n^*-1} A^{(j)}\right)_{ss} \quad (20)$$

$$\frac{dM}{dt} = -2(1+\eta)k_{11}^{(app)}M^2 - k_h^{(app)}M \qquad \text{(Type III)} \quad (21)$$

with $\eta = (2\kappa_{ij}/\alpha_{III})^{1/2}$ or $(j'/2 - 1)$ depending on the relative κ_{1j} and κ_{ij} values. In the above expressions, $\langle k_{1j}\rangle$ is the average k_{1j} value, $k_{11}^{(app)}$ is defined as

$$k_{11}^{(app)} \equiv f_R^2 k_{11} \quad (22)$$

and $k_h^{(app)}$ comes from $\kappa_h^{(app)}$ (eq 17):

$$k_h^{(app)} \equiv f_R(n^* - j')\langle k_{1j}\rangle A^{(j')}_{sat} \quad (23)$$

III. Results

In this section, the results of numerical simulations of the complete model (eq 2) will be presented as a function of the model parameters. These results illustrate both the rich set of kinetic behaviors the model produces as well as the level of agreement between the approximate expressions derived in section II and the exact results. The results are presented primarily in terms of $m(\tau_{agg})$ profiles to facilitate direct comparison with typical experimental measures of protein aggregation in commercial formulations. In many cases, the associated analyses and data transformations are readily implemented with even limited experimental data. The model parameters that are adjustable within each kinetic type are n^*, κ_{ij} or κ_{1j}, and K_{NR} (or f_R, cf. eq 6) for all types, with j' and $a_{j',sat}$ as additional parameters for Type III scenarios. For the purposes of simulating the profiles numerically, κ_{ij} indicates the value that is assigned to all k'_{ij} (except k_{11}, which equals 1 by definition), and κ_{1j} indicates the value assigned to all k'_{1j} with the constraint that all k'_{ij} $(i, j \neq 1)$ are set to zero. All simulations were performed using the initial conditions $m = 1$, $a_{j>1} = 0$. Using different initial a_j levels does not dramatically impact the results shown below, although the parameter values separating Type IB and II behaviors (see below) will change slightly and lag times in Type I and II behaviors can be shortened.

In general, a numerical solution of eq 2 gives qualitative results illustrated by the concentration profiles in Figure 1 for $n^* = 5$, with the other parameter values given in the Figure caption. Characteristic of these profiles is the buildup of successive soluble aggregate species ($j < n^*$), with each eventually reaching pseudo-steady-state levels $A^{(j)}_{ss}$ that are on the order of or lower than R_A levels.

Type I Behavior. Figure 2 shows the effect of n^* on Type IA kinetic profiles, with the associated second-order plot as an inset. The values of κ_{ij} and K_{NR} are given in the caption. The key points to note are that after an initial lag time the profiles show pseudo-second-order kinetics, as anticipated by eq 11a, and the lag time increases with increasing n^*. The apparent second-order rate coefficient is obtained directly from the limiting slopes in the inset and is an increasing function of n^*. Equation 11a illustrates the dependence of the limiting slope on κ_{ij} and α_I. Note that κ_{ij} is the same for all curves in Figure 2, thus α_I is a function of n^*. As mentioned in section II, α_I accounts for differences in the number of terms retained after the expressions for $da_j/d\tau_{agg}$ are summed over j. As a result, α_I changes with n^* but ultimately reaches a limiting value at high n^*. Figure 3 shows α_I values as a function of n^*, calculated from the slopes in Figure 2 (inset) using eq 11a and the known value of κ_{ij}. The limiting value of α_I is approximately $^1/_2$. Thus, to a good approximation, the parenthetical term in eq 18a can be replaced by $(1 + 2\kappa_{ij}^{1/2})$. If one could reasonably estimate the ratio of $\langle k_{ij}\rangle$ to k_{11} (e.g., by calculating relative "reaction" rates for sticky spheres of different radii and with different sized "sticky patches" [61]), it might be possible to correct for this term when determining $k_{11}^{(app)}$ values from experimental $M(t)$ data.

The effects of κ_{ij} on Type IA kinetics at fixed n^* and K_{NR} are shown in Figure 4. The limiting behavior in eq 11a is maintained across a broad range of κ_{ij}, and increasing κ_{ij} increases the pseudo-second-order rate coefficient. Perhaps less apparent in Figure 4 is the effect of κ_{ij} on the lag time until the summation in eq 11a reaches steady state. To better illustrate this point, the inset shows the values of the slopes for each

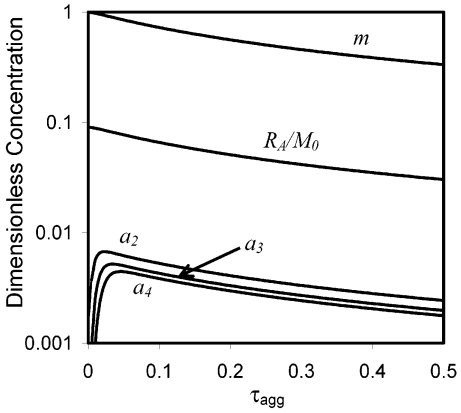

Figure 1. Canonical kinetic profiles for eq 2 (dimensionless form) with $n^* = 5$, $\kappa_{ij} = 10$, and $K_{NR} = 0.1$.

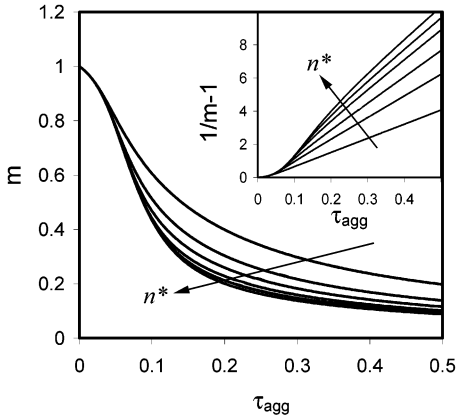

Figure 2. Effect on n^* on Type I kinetics at fixed $\kappa_{ij} = 10$ and $K_{NR} = 0.2$ plotted as $m(\tau_{agg})$ (main panel) with the curves (uppermost to lowermost) for $n^* = 5$, 10, 20, 50, 100, and 200. The inset shows a second-order plot with the same values of n^* increasing from the lowermost to uppermost curves.

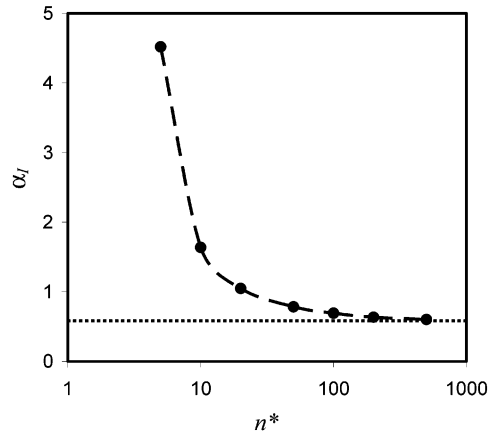

Figure 3. Asymptotic behavior of α_I as n^* increases. Values for the effective second-order rate coefficient in eq 11a are taken from the limiting slopes in Figure 2(inset), and α_I values are then calculated using the known value of κ_{ij} for each data set. The dot−dashed line indicates a value of $\alpha_I = {}^1/_2$.

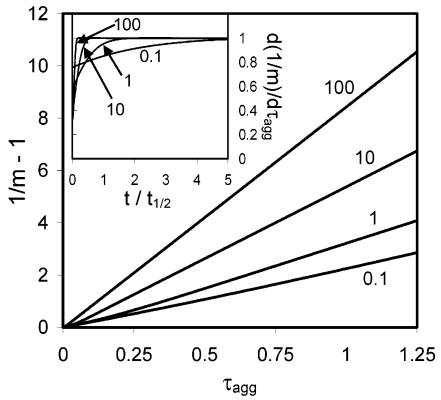

Figure 4. Effect of κ_{ij} on Type I kinetics with $n^* = 10$ and $K_{NR} = 0.2$. Each curve is labeled with the value of κ_{ij} used to generate it. Although not easily observable on the scale of the plot, at early τ_{agg} there is a detectable lag time before the limiting second-order rate of monomer loss is achieved, which decreases rapidly with increasing κ_{ij}. The inset shows the effect of κ_{ij} on the lag time to achieving the limiting behavior of eq 11a shown in the main panel. The value of the first derivative of $(1/m - 1)$ scaled by its long-time value is plotted on the ordinate. The curves are labeled by their respective κ_{ij} values. For clarity, t for the inset is scaled by the respective $t_{1/2}$ for each curve.

curve in Figure 4 as a function of τ_{agg}, normalized against their associated steady-state values. Interestingly, the time scale for steady-state behavior is much less than $\tau_{agg} = 1$ only for κ_{ij} on the order of 10 or higher. Despite this, the observable lag time from the perspective of $m(\tau_{agg})$ is much less than $\tau_{agg} = 1$ even for the lower κ_{ij} values. Thus, it appears that the approximate solutions derived in section II may hold within experimental precision even under conditions that would not be anticipated from only theoretical considerations.

The effects of K_{NR} on Type IA kinetics are relatively minor, with effectively ideal pseudo-second-order behavior at K_{NR} less than ca. 0.01 and slight deviations from pseudo-second-order behavior as K_{NR} approaches 1. The inset to Figure 5 illustrates this with a second-order plot (cf. Figures 2 and 4) at fixed n^* and κ_{ij} with t scaled by $t_{1/2}$, the time for m to reach a value of 0.5. This scaling provides a convenient means to discriminate between the different types of kinetic behavior exhibited by the model and is also easily applied to experimental data.

Similar results occur for Type IB behavior (not shown), except that now the limiting values of the second-order rate coefficients are found to follow eq 11b. The effects of n^*, κ_{1j}

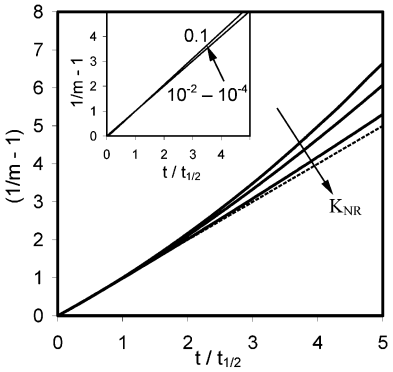

Figure 5. Effect of K_{NR} on Type IA (inset) and Type III (main panel) kinetics. The Type IA profiles are with $\kappa_{ij} = 10$ and $n^* = 10$, and the curves are labeled with their respective K_{NR} values. For K_{NR} from 0.01 to 0.0001, the curves are indistinguishable on the scale of the plot and show ideal pseudo-second-order behavior. For the Type III profiles, the solid curves indicate the simulated profiles ($K_{NR} = 0.001$, 0.01, and 0.1 from uppermost to lowermost), and the dashed curve indicates the profile that would be expected if only the second-order term in eq 16 or 21 were significant. The curves are plotted relative to $t_{1/2}$ for ease of comparison and to aid in the discussion in section IV.

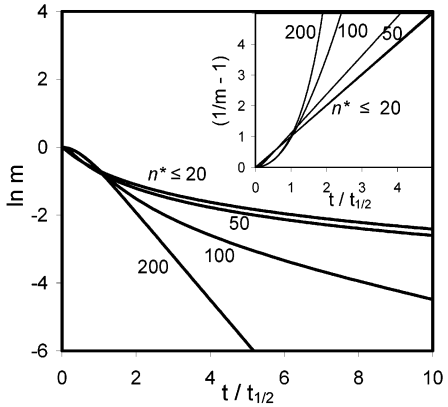

Figure 6. Effect of n^* on Type II kinetics; $K_{NR} = 0.01$ and $\kappa_{ij} = 100$. The curves are labeled with their respective n^* values. Note the transition from Type IB to Type II behavior as n^* increases. The first-order dependence characteristic of Type II behavior is more easily seen in the main panel where ln m rather than $1/m - 1$ (inset) is plotted versus $t/t_{1/2}$.

(rather than κ_{ij}), and K_{NR} on Type IB behavior are similar to those just described for Type IA, with one dramatic exception. Namely, a transition occurs from Type IB to Type II kinetics as n^* and K_{NR} are increased and as κ_{1j} is decreased. This transition and how Type IB and II behaviors are related will be discussed in the next subsection. It is worth noting that a similar transition is detectable for IA to II but occurs only at extremely low κ_{ij} (i.e., $\kappa_{ij} < 0.01$) and therefore is irrelevant for our purposes. This difference is due to much lower steady-state a_j levels in Type IA systems relative to Type IB given the same value of κ_{ij} and κ_{1j}. Physically, this is due to the many more parallel aggregation steps available for the consumption of a given j-mer in IA kinetics compared to those available in IB kinetics since aggregation steps that involve only aggregates are not present in the latter case.

Type II Behavior and the IB/II Transition. Figure 6 illustrates a typical pattern of how Type II behavior evolves from Type IB behavior as n^* increases at constant κ_{1j} and K_{NR}. The inset shows this on a second-order plot, and the main panel provides a first-order plot. Together, they clearly show the transition from pseudo-second-order behavior (Type IB) to pseudo-first-order behavior (Type II, eq 14), which occurs over

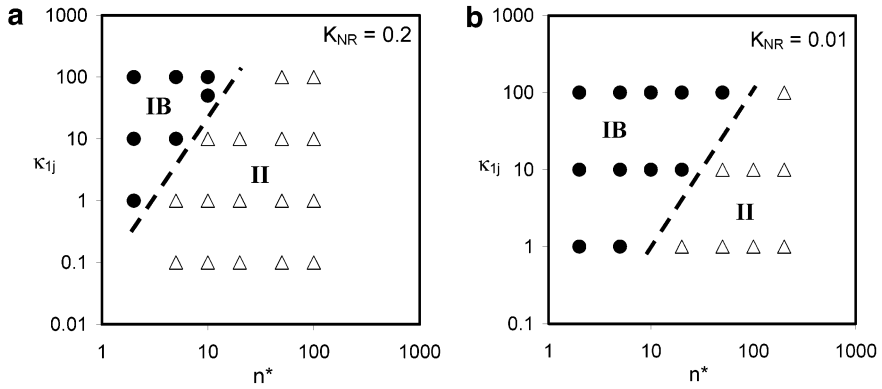

Figure 7. State diagrams for Type II kinetics showing boundaries the between IB to II behavior as a function of κ_{ij} and n^* at (a) $K_{NR} = 0.2$ and (b) $K_{NR} = 0.01$.

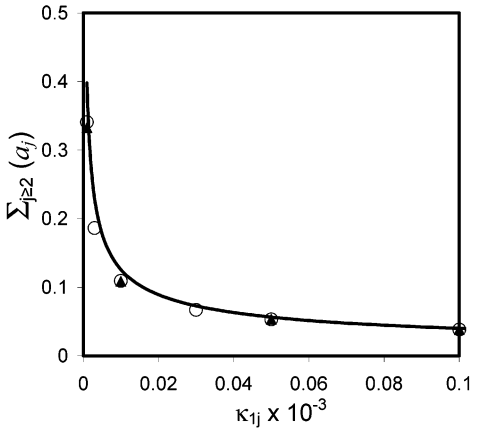

Figure 8. Dependence of $\Sigma_j(a_j)$ on κ_{1j} for Type II kinetics ($K_{NR} = 0.2$) with $n^* = 50$ (○) and $n^* = 100$ (▲). The dotted line is a fit to $\Sigma_j(a_j) = c_1 + c_2\kappa_{1j}^{-1/2}$ (see also discussion in text).

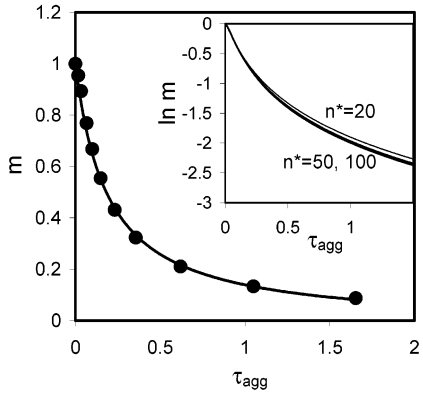

Figure 9. Effect of n^* on Type III behavior at fixed $\kappa_{ij} = 10$, $K_{NR} = 0.1$, j'=10, and $a_{j',sat}/a_{j',max} = 0.1$. The inset shows ln $m(\tau_{agg})$ versus τ_{agg} with curves labeled by n^*. The curves also illustrate the characteristic profile for Type III behavior: mixed second- and first-order dependence on m. The main panel shows a fit of eq 16 (—) to $n^* = 50$ points (●); intermediate points were deleted from the $n^* = 50$ data so as to simulate the limited number of points that would typically be available experimentally. See Figure 10 for values of the fitted parameters.

a relatively narrow but appreciable range of n^*. A similar effect is seen as κ_{1j} is decreased at constant n^* and K_{NR}. Figure 7 presents parameter "state diagrams", showing the boundary between Type IB and Type II behaviors in terms of a 2D parameter-space plot at constant $K_{NR} = 0.2$ (Figure 7a) and 0.01 (Figure 7b). Clearly, the region of Type IB behavior increases in size with decreasing K_{NR}. These results are consistent with the derivation presented in section II, which argued that eq 14 would be the correct limiting form when the second term in eq 13 dominates, as will be the case if the a_j levels in the summation are sufficiently high and as the number of terms in the summation increases. Decreasing κ_{1j} induces high steady-state a_j levels, and increasing n^* increases the number of terms in the summation. This line of reasoning is confirmed by inspection of the intermediate a_j levels for Type II kinetics, which are found to be on the order of R_A levels in the simulations for the τ_{agg} range of interest (data not shown).

Finally, we note that $\Sigma_{j>1}(a_j)$ and κ_{1j} are not independent quantities since increasing κ_{1j} necessarily lowers the average a_j values. Figure 8 illustrates this effect by showing $\Sigma_{j>1}(a_j)$ versus κ_{1j} for $K_{NR} = 0.2$ and two values of n^* (50 and 100). Empirically, $\Sigma_{j>1}(a_j)$ scales reasonably well as $\kappa_{1j}^{-1/2}$ (see fitted curve in Figure 8 and caption) and is only weakly dependent on n^* within this range of n^*. Thus, a further simplification to eq 20 is

$$k_1^{(app)} \cong \chi f_R \langle k_{1j} \rangle^{1/2} \tag{24}$$

where χ is a constant of order 1. This same reasoning also explains why decreasing K_{NR} promotes Type IB behavior over Type II. That is, decreasing K_{NR} decreases R_A levels (at fixed M_0) and thus lowers the average a_j levels in the system at $t > 0$. The exception to this is if the initial a_j levels are artificially high (i.e., if the system is "seeded" with preformed aggregates). In this case, one anticipates that Type IB behavior may be suppressed relative to Type II behavior when κ_{ij} and n^* are low and high enough, respectively.

Type III Behavior. Figure 9 shows typical profiles for Type III behavior (parameter values in caption), which is characterized by a mixed first- and second-order dependence with respect to m (inset). The relatively weak dependence of the profiles on n^* is also illustrated in the inset. As anticipated from the derivation of eq 16, increasing n^* causes the pseudo-first-order character of the profiles to increase, albeit only slightly, due to the greater number of terms in the summation in $\kappa_h^{(app)}$. The relative contributions from second- and first-order terms can be quantified in terms of the apparent rate coefficients from a numerical fit of the simulated data to an expression such as eq 16 or 21. The main panel of Figure 9 illustrates this for the simulated $n^* = 50$ data from the inset, with data points at intermediate times removed so as to imitate the discrete nature of experimental data. The relative values of the fitted coefficients are shown as part of Figure 10 (see discussion below).

Increasing j' has an effect on the observed kinetic profiles analogous to but more pronounced than n^*, as shown in Figure 10. The inset shows the associated values of η and $\kappa_h^{(app)}$

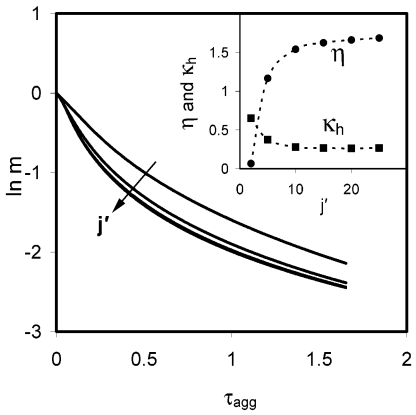

Figure 10. Effect of j' on Type III kinetics at fixed $\kappa_{ij} = 10$, $K_{NR} = 0.1$, $n^* = 50$, and $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}} = 0.1$. The curves correspond to $j' = 2$, 5, 10, 15, 20, and 25 (from uppermost to lowermost). The curves for j' greater than ca. 15 are indistinguishable on the scale of the plot. Inset: the effect of j' on the coefficients in eq 21 on the basis of least-squares fits of eq 21 to the data in the main panel.

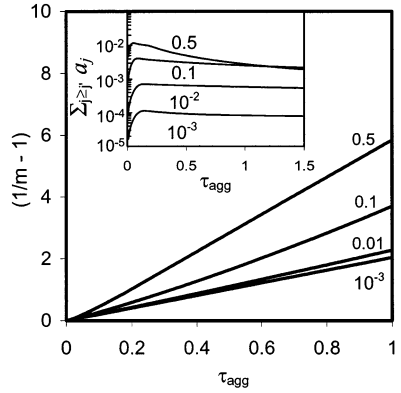

Figure 11. Effect of $a_{j',\mathrm{sat}}$ on Type III kinetics; $\kappa_{ij} = 10$, $K_{NR} = 0.05$, $n^* = 25$, $j' = 2$, and $a_{2,\mathrm{max}} = 0.0035$. The curves are labeled by the ratio $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$ used to calculate them. The main panel shows $(1/m - 1)$ versus τ_{agg}, with the inset showing corresponding values of the summation in the second term of eq 16 as a function of τ_{agg} and $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$. Note that a pseudo-steady-state level is not maintained if $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$ is too high, in which case $\kappa_h^{(\mathrm{app})}$ in eq 16 is not reasonably independent of τ_{agg}, and behavior more akin to that of Type I is produced at long τ_{agg}.

$(= k_h^{(\mathrm{app})}/k_{11})$ determined from least-squares fits of the curves in Figure 10 to eq 21 as a function of j'. As expected, η approaches 0 at low j' and increases asymptotically with increasing j'. η and $\kappa_h^{(\mathrm{app})}$ have asymptotic values of orders 1 and 0.1, respectively. Note that j' cannot be increased to arbitrarily large values, as its value must by definition remain below n^* and should be much less than n^* to produce kinetics significantly different from Types I and II. However, this constraint does not appear have a significant effect since the asymptotic behaviors are reached at a relatively small ratio of j' to n^*.

Increasing κ_{ij} does not qualitatively affect Type III kinetics over the range $1 < \kappa_{ij} < 1000$, as $m(t/t_{1/2})$ profiles were found to superimpose effectively up to $t/t_{1/2} \approx 5$ (not shown). In contrast, changing $a_{j',\mathrm{sat}}$ has significant effects not only on the quantitative value of $\kappa_h^{(\mathrm{app})}$ (cf. eq 17) but also on the qualitative kinetic behavior. These effects are illustrated in Figure 11. The profiles are shown for different ratios of $a_{j',\mathrm{sat}}$ to $a_{j',\mathrm{max}}$. As expected, the inset to Figure 11 explicitly shows that changing $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$ shifts the steady-state value of the summation term in $\kappa_h^{(\mathrm{app})}$. However, an additional change occurs as $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$ approaches 1. Unlike at lower values of $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$, the summation in $\kappa_h^{(\mathrm{app})}$ achieves a pseudo-steady-state value for only a limited range of τ_{agg}, after which it declines according to the same kinetics that would occur in the absence of the saturation limit. In particular, it switches from Type III behavior at early times to Type I (A or B) behavior at longer times. The duration of the Type III behavior decreases as $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}}$ approaches 1, and Type I behavior is recovered at all t when $a_{j',\mathrm{sat}}/a_{j',\mathrm{max}} = 1$.

Finally, the effects of K_{NR} on Type III kinetics are illustrated in the main panel of Figure 5 over a broad range of K_{NR}. The most striking result is the increased deviation from ideal pseudo-second-order behavior with decreasing K_{NR}. This is in stark contrast to the behavior observed for Type I kinetics, for which the opposite trend is observed.

IV. Discussion

Distinguishing between Types IA, IB, II, and III. For a system obeying pseudo-second-order kinetics such as eqs 11a and 11b, it is easily shown that a plot such as Figure 5 should yield a straight line with a slope of 1. This is precisely the behavior we observe for Type IA in the inset to Figure 5 for all K_{NR} below ca. 0.01, with deviations from this behavior becoming more pronounced as K_{NR} approaches 1, including a detectable lag time at early t. As a point of reference, if we are considering aggregation kinetics as a function of T, then $K_{NR} = 1$ corresponds to the midpoint temperature T_M for the $N-R_A$ equilibrium. Similarly, $K_{NR} = 0.01$ indicates a T of ca. (T_M − 10 °C) if the thermodynamics are similar to those for folding/unfolding for small globular proteins.[62] In contrast, Type III kinetics display the opposite behavior since decreasing K_{NR} enhances the pseudo-first-order contribution to the observed kinetics (Figure 5, main panel). Obviously, Type II kinetics should not appear as a straight line at any $K_{NR}(T)$ since the profiles will be pseudo-first order after the initial lag time has elapsed. Thus, it appears that a plot of experimental data analogous to that in Figure 5 may provide a simple and unambiguous indication of whether a given protein system is exhibiting Type I, II, or III aggregation behavior. Note that T is not the only experimental variable that may independently influence K_{NR}. Consideration should be given to how normalized kinetic profiles appear as a function of additional variables such as pressure and osmolyte concentration.

Of course, a system may also switch behaviors as T (and therefore K_{NR}) changes—typically, one expects K_{NR} to decrease with decreasing T. The results in Figure 7 show that the net effect will primarily be to shift from Type II to Type I (or possibly Type III) with decreasing K_{NR}. Thus, using only an analysis such as that illustrated with Figure 5, one may potentially confuse a II/I switch as T is lowered with a purely Type I system, but a Type III system would remain unambiguous. However, it should be remembered that only a Type II system shows a true long-time first-order dependence on m. Thus, it should be possible to distinguish further among different scenarios provided one collects data over a number of reaction half-lives. Furthermore, since a Type II system will evolve toward a Type IB system with decreasing K_{NR}, it may be possible to distinguish between Type IB and IA by observing whether a transition between first- and second-order kinetics can be induced by adjusting experimental parameters such as temperature, pressure, and osmolyte concentration. That is, the evolution toward ideal pseudo-second-order behavior with decreasing K_{NR} is rather subtle over relatively large changes in K_{NR} for Type IA (Figure 5 inset) whereas a transition from Type

II to Type IB would appear as a dramatic change from pseudo-first to pseudo-second order on a similar plot. The difference between IA and IB behavior is not necessarily a concern from the standpoint of extrapolating *down* in T provided α_I or n^* is not a strong function of T (see next subsection). However, from the standpoint of accurately determining k_{11}, the value of k_{11} would be overestimated by an approximately constant factor of $[1 + (\kappa_{ij}/\alpha_I)^{1/2}]$ or $[1 + n^*/2]$ for Types IA and IB, respectively, unless one knows whether Type IA or IB kinetics are operative. If Type IA is operative, one might reason that $[1 + (\kappa_{ij}/\alpha_I)^{1/2}]$ is at most on the order of 10 since $\kappa_{ij} > 10^2$ might not be justified on the basis of geometric or diffusive arguments.[63] Conversely, $[1 + n^*/2]$ not only could be on the order of 10^2 or higher but should be estimable, e.g., from dynamic light scattering of soluble high molecular weight aggregates such as those observed in HPLC for some aggregating protein systems[5,21,54] at pH values significantly below their pI values.

In addition, one may consider the initial rate kinetics that will arise for the different scenarios. On the basis of the derivations in section II, one can show that in the limit of $t = 0$ the kinetics for Types I−III will be purely second order with an observed rate coefficient $k_{11}^{(app)}$. This conclusion relies on the assumption that $(a_{j,t=0})$ is zero for $j > 1$. In practice, this condition may not be realizable. If instead the initial a_j levels are close to (or even slightly higher than) their steady-state values, the following expressions for the initial rate (r_0) as a function of M_0 will hold.

$$r_0 \equiv -\frac{1}{M_0}\left(\frac{dM}{dt}\right)_{t\to 0} = 2k_{11}^{(app)}\left(1 + \sqrt{\frac{\kappa_{ij}}{2\alpha_I}}\right)M_0 \qquad \text{(Type IA)} \tag{25}$$

$$r_0 = n^* k_{11}^{(app)} M_0 \qquad \text{(Type IB)} \tag{26}$$

$$r_0 = k_1^{(app)} \qquad \text{(Type II)} \tag{27}$$

$$r_0 = 2(1 + \eta)k_{11}^{(app)} M_0 + \kappa_h^{(app)} \qquad \text{(Type III)} \tag{28}$$

Thus, initial rates (with and without "seeding" of preformed aggregates) may provide an additional means to discriminate among different kinetic scenarios. Similarly, there are additional orthogonal data that will aid in this determination. For example, conditions that favor Type II kinetics will cause significant lag times as well as giving rise to $a_{j,max}$ levels that may be comparable to m and thus presumably detectable by the same techniques used to monitor $m(t)$. Furthermore, since $a_{j,sat}$ levels implicitly depend on M_0, it should in principle be possible to switch to (from) Type III kinetics by increasing (decreasing) M_0.

Finally, it should be noted that in some cases steady state a_j levels may not be reached until $t/t_{1/2} >$ ca. 0.2, which corresponds to a regime beyond that typically of interest from a commercial shelf-life perspective.[1,2] Unless the solution contains soluble aggregates at near-steady-state levels (e.g., as a result of the steps needed to prepare it), initial rates in particular will yield values of $k_{11}^{(app)}$ or $k_1^{(app)}$ that may be appreciably lower than those derived from fits to the kinetics over multiple half-lives. If the initial sample conditions are representative of the (commercial) system of interest, then this poses no significant problem since such early-time rate coefficients are most appropriate. Alternatively, if one instead regresses apparent rate coefficient values from data over multiple half-lives, then there is the potential to *over*estimate the values most appropriate for early times. As a result, extrapolated rate coefficients (see below) would be artificially high, and thus the predicted shelf lives would be artificially low, albeit by an approximately constant amount provided the parenthetical coefficients in eqs 18, 20, and 21 are insensitive to T. This situation, which is undesirable from the perspective of calculating accurate k_{11} values, can perhaps be tolerated for shelf-life prediction since those predictions will offer a lower bound to the expected shelf life (see also below).

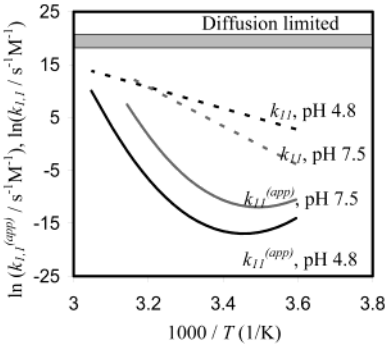

Figure 12. Arrhenius plot of k_{11} and $k_{11}^{(app)}$ for bG-CSF at pH 4.8 and 7.5. Adapted from ref 51.

It should be emphasized that Type I−III kinetics capture the observable behaviors of eq 2 under many conditions commonly encountered by commercial and biological systems. However, the expressions derived in section II are *not* appropriate under conditions where aggregation is faster than the time scale to establish and/or maintain conformational equilibria.[44−48,51] In this case, aggregation is "unfolding-limited"[46] and is well described by a classic Lumry−Eyring description.[44−46] This form of aggregation kinetics appears to be promoted at extreme conditions such as high T, concentration, and pressure.[6,46,50,51]

Prediction of Low-*T* Shelf Life from High-*T* Data. In this subsection, the results for Type III kinetics described earlier are combined with examples adapted from ref 51 to illustrate practical approaches to estimate low-T shelf life quantitatively from high-T data. Techniques such as these have been recently implemented for bG-CSF, and detailed descriptions of the experimental procedures are given elsewhere.[51] The bG-CSF system[49,51,54] is a convenient reference point for much of the analysis presented next, although the discussion below remains general.

To begin, we consider the temperature dependence of $k_{1,1}^{(app)}$ and $k_{1,1}$. Figure 12 illustrates this for a model protein[51] by plotting the logarithm of $k_{1,1}^{(app)}$ and $k_{1,1}$ versus inverse T (i.e., an Arrhenius plot) for pH 7.5 and 4.8. From Figure 12, it can be seen that $k_{1,1}^{(app)}$ displays highly non-Arrhenius behavior whereas $k_{1,1}$ empirically follows an Arrhenius T dependence over the T range from $T \sim T_M$ down to $T \approx 0$ °C. Recall that $k_{1,1}$ is related to $k_{1,1}^{(app)}$ through eq 22. Thus, Figure 12 shows that the non-Arrhenius component of the T dependence of $k_{1,1}^{(app)}$ is quantitatively accounted for by shifts in $N-R_A$ equilibrium with T. Also shown is an estimate of the limiting value of $k_{1,1}$ for bG-CSF if the R_A-R_A reaction were diffusion-limited,[51] indicating that $k_{1,1}$ is well below the diffusion limit for this system.

These results suggest a straightforward procedure for estimating shelf life or equivalently $k_{1,1}^{(app)}$ at low T solely from high-T data by explicitly accounting for the non-Arrhenius T dependence. Namely, one first determines $K_{NR}(T)$, presumably from thermodynamic unfolding measurements.[51,56,62] Because of a combination of practical constraints[5,48,51,64] and theoretical considerations illustrated below, one selects a range of T below T_M (or more accurately, T_x)[51] but above some T_H to measure (apparent) aggregation kinetics directly (e.g., using initial rates

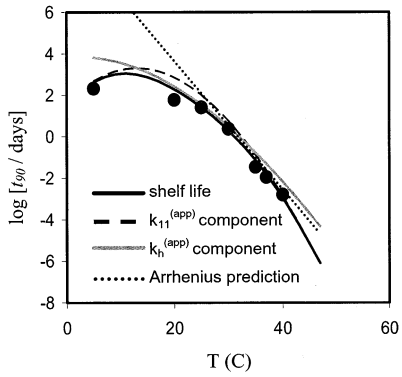

Figure 13. Shelf life of bG-CSF as a function of temperature. Adapted from ref 51.

and eq 28). The apparent rate coefficients from the high-T aggregation data ($k_{1,1}^{(app)}$ and $k_h^{(app)}$) are then converted into their corresponding intrinsic values $k_{1,1}$ and k_h using eqs 22 and 23. $k_{1,1}$ and k_h are extrapolated to low T using the standard Arrhenius relation. For $k_{1,1}$, for example, this is

$$\frac{\mathrm{d} \ln k_{1,1}}{\mathrm{d}(1/T)} = -\frac{E_{a,11}}{R} \tag{29}$$

The extrapolated low-T values of $k_{1,1}$ and k_h values are used in eqs 21−23 along with the corresponding values of K_{NR} either directly from experiment or extrapolated from higher T using standard thermodynamic relations.[51,56,62] Equation 21−23 are used to calculate the time to 10% loss of monomer (i.e., t_{90}, our working definition of shelf life) at the (extrapolated) lower T values. An example of the results of such a procedure for this model protein is given in Figure 13, where T_H was taken to be 30 °C. The solid black curve representing the predicted t_{90} values at $T < T_H$ is in excellent agreement with the experimental data.[51] Figure 13 also illustrates the dramatic non-Arrhenius behavior over the narrow T range between 5 and 30 °C, which is also apparent in Figure 12. The two additional curves in Figure 13 are the contributions to t_{90} from the second- and first-order terms in eq 21, labeled $k_{1,1}^{(app)}$ and $k_h^{(app)}$, respectively. These curves also illustrate the stronger non-Arrhenius T dependence of $k_{1,1}^{(app)}$ compared to that of $k_h^{(app)}$ (cf. eq 22 and 23), although $k_h^{(app)}(T)$ is also significantly non-Arrhenius and should be included for quantitative shelf-life predictions.[51]

The above analysis assumed Arrhenius behavior for $k_{11}(T)$. Our current understanding of protein aggregation indicates that aggregation is driven to a large extent by hydrophobic (hϕ) forces similar to those that drive proteins to fold in aqueous solution.[8,23] Although the exact nature of the rate-determining step in, for example, the hϕ-mediated aggregation of two monomers remains an open question, one can speculate that an analogy between folding and aggregation holds such that k_{11} would have a nonzero heat capacity of activation (Δc_p*) akin to that observed for the folding rate constant (k_f) of simple globular proteins.[65,66] If this is the case, then k_{11} values extrapolated to $T < T_H$ using eq 29 would be related to the true value ($k_{11}^{\Delta C_p \neq 0}$) by the following relation (Appendix B):

$$\ln \frac{k_{11}}{k_{11}^{\Delta C_p \neq 0}} = -\frac{\Delta c_p^*}{R}\left[\left(\frac{T_H}{T} - 1\right) - \ln\frac{T_H}{T}\right] \tag{30}$$

Continuing with the folding−aggregation analogy, Δc_p* will be negative,[65,66] and $|\Delta c_p^*| < |\Delta c_{p,f}^*|$, with $\Delta c_{p,f}$* the activation heat capacity of folding. Thus, from eq 30, we see that the assumption of Arrhenius behavior for k_{11} will overestimate $k_{11}(T)$ (underestimate t_{90}) at $T < T_H$ and thus provide a conservative lower bound on t_{90} extrapolated from higher T. Typical values of T_H are 30−45 °C, and storage temperatures of interest are 5−25 °C. If we assume that on average about half of the hϕ surface area of R_A is buried in attaining the transition state and use a value of −1 kcal/mol K for a typical $\Delta c_{p,f}$*,[65,66] then we have Δc_p* on the order of −0.5 kcal/mol K. The maximum $k_{11}/k_{11}^{\Delta C_p \neq 0}$ values (at $T_H - T = 40$ °C) anticipated from eq 30 are then on the order of 10, indicating the potential for a significant underestimation of shelf life if activation heat capacities for protein aggregation are similar in magnitude to those for folding.

It is not always possible to obtain accurate measures of $K_{NR}(T)$ over the full T range of interest, in part because of experimental difficulties in measuring the thermodynamics of the $N-R_A$ transition as well as ambiguity in positively identifying R_A with a conformational state that can be quantitated accurately. For the case used above, R_A was fortuitously the same state as that achieved by thermally induced unfolding, so $K_{NR}(T)$ was determinable from standard thermal unfolding experiments.[51] That notwithstanding, it is possible to obtain useful quantitative estimates of low-T shelf life even when the nature of R_A is ambiguous. This is achieved by placing an upper (lower) bound on $k_{1,1}^{(app)}$ (shelf life) at low T. A description of the analogous procedure for $k_h^{(app)}$ is foregone, as it is a straightforward extension of the one for $k_{1,1}^{(app)}$. We begin by writing the T dependence of $k_{1,1}^{(app)}$ explicitly as (Appendix C)

$$R \ln \frac{k_{1,1}^{(app)}(T)}{k_{1,1}^{(app)}(T_H)} = -E_{a,1,1}^{(app)}|_{T_H}\left(\frac{1}{T} - \frac{1}{T_H}\right) - 2\Delta c_p\left[1 - \frac{T_H}{T} + \ln\frac{T_H}{T}\right] \tag{31}$$

In eq 31, R is the gas constant, $k_{1,1}^{(app)}(T_H)$ is the experimental value of $k_{1,1}^{(app)}$ determined at T_H, $E_{a,1,1}^{(app)}|_{T_H}$ is the *apparent* activation energy derived from a plot of ln $k_{1,1}^{(app)}$ versus $1/T$ over the T range just above T_H, and Δc_p is the heat capacity change for the $N-R_A$ transition (assumed to be independent of T). For realistic systems, $\Delta c_p \geq 0$ since R_A has a significantly greater solvent exposure of hydrophobic residues than N.[3,6,8,9,25,37] In addition, $\Delta c_p \leq \Delta c_{p,NU}$, with $\Delta c_{p,NU}$ being the heat capacity change of unfolding, such as that measured in thermal unfolding experiments where U is the conformational state achieved at high T in the absence of degradation. From eq 31, we see that using $\Delta c_p = \Delta c_{p,NU}$ provides an upper bound on $k_{1,1}^{(app)}(T < T_H)$ or equivalently a lower bound on t_{90}. To illustrate, Figure 14 shows $k_{1,1}^{(app)}(T)$ as a function of the value of Δc_p used in eq 31, with $T_H = 30$ °C (as in Figure 13) and $k_{1,1}^{(app)}(T \geq T_H)$ from Figure 13 and ref 51. The values of Δc_p corresponding to each curve are labeled on the Figure, with $\Delta c_p = 3.6$ kcal/mol being the value experimentally determined for bG-CSF from thermal unfolding experiments.[51] This procedure can circumvent the need to determine the conformational identity of R_A or even the need to quantitate R_A specifically in experiments. Furthermore, Figure 14 illustrates how large the errors in predicted rate coefficients and shelf life can be if $\Delta c_p/R$ for the system of interest is significantly greater than zero, and one instead uses the practice of assuming an Arrhenius extrapolation of the *apparent* rate coefficient(s).[21,50,67]

Analogous analyses to those presented above are straightforward for Type I(A or B) kinetics with the understanding in this case that $k_{11}^{(app)}$ is derived from eq 18 rather than eq 21, and of course a $k_h^{(app)}$ extrapolation should not be included. Note

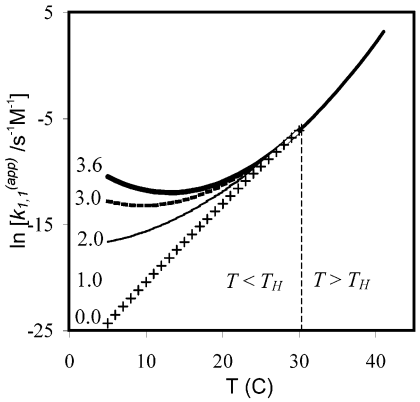

Figure 14. Effect of Δc_p used in extrapolating $k_{11}^{(app)}$ from high T ($> T_H$) to lower T using eq 31. The numbers indicate the value of Δc_p (kcal/mol) used in the extrapolation from $T_H = 30$ °C. The high-T values of $k_{11}^{(app)}$ are the same as those taken from ref 51 in Figure 13.

that initial rates or fits to full kinetic profiles in both Type I and Type III scenarios yield values of $k_{11}^{(app)}$ multiplied by the appropriate coefficient in terms of either κ_{ij} or n^* (cf. eq 18 and 21 and discussion of initial rates in section III). For extrapolating shelf-life data, the exact value of this coefficient is unimportant provided it does not change appreciably over the T (or other extrapolating variable) of interest. The shelf-life extrapolation for Type II kinetics is in principle analogous to that described above for Type III kinetics, except that $k_1^{(app)}$ is related to $\langle k_{1j} \rangle$ via a factor of f_R rather than f_R^2 since it is a first-order rather than a second-order apparent rate coefficient (cf. eq 24). Interestingly, this also suggest that $\langle k_{1j} \rangle$ (and its T dependence, if any) might be estimable for a given protein system if Type II kinetics can be induced. However, caution should be taken when performing shelf-life extrapolations from high T (high K_{NR}) to low T (low K_{NR}) if Type II kinetics are operable because of the potential for a change from Type II to Type IB kinetics (section IV) within the extrapolation regime.

Impact of Additional Non-Native Conformational States. Section II did not explicitly include additional non-native conformational states other than N and R_A. The effects of additional non-native configurational states are considered briefly here to show how the results presented above may be expected to change as well as to point out much that remains unchanged for practical purposes.

The incorporation of additional non-native states is depicted schematically in eq 32 for the scenario akin to that used in eq 1.

$$N \leftrightarrows I \leftrightarrows R_A \leftrightarrows U$$
$$R_A \rightarrow \cdots \rightarrow \alpha \qquad (32)$$

In eq 32, a double-lined arrow indicates a reversible transition whereas a single lined arrow indicates an irreversible one, and α indicates the (irreversibly) aggregated state. The irreversible steps are assumed to involve aggregation and to consist of multiple steps (indicated by the ellipses and multiple single-lined arrows), (e.g., as described by eq 1). The additional states that are considered are the "unfolded" state U and an arbitrary folding intermediate I. Note that the exact arrangement of eq 32 is not intended to specify the relative order of states on the folding pathway, although the order as depicted is reasonable for some systems, where I is a relatively stable folding intermediate, R_A is a molten-globule state, and U lies above both on the folding landscape. For our purposes, the quantity of interest is $f_R(T)$, the fraction of M in the R_A state as a function of T. In general, we may write the fractional population of any of the states f_j, ($j = N, U, I, R_A$) as

$$f_N = \frac{1}{1 + K_{NI} + K_{NR} + K_{NU}} \qquad (33)$$

$$f_j = K_{Nj} f_N \qquad j = U, I, R_A \qquad (34)$$

where K_{ij} is the equilibrium constant for the transition between i and j at a given set of solution conditions. Taking the derivative of eq 34 with respect to inverse T for $j = R_A$ gives

$$\frac{d \ln f_R}{d\, 1/T} = f_N \frac{d \ln K_{NR}}{d\, 1/T} + K_{NR} \frac{d \ln f_N}{d\, 1/T} \qquad (35)$$

Equation 35 indicates that the non-Arrhenius T dependence of $k_{1,1}^{(app)}$ will be influenced significantly by the presence of additional non-native states only under conditions where N is significantly depleted as the equilibrium shifts toward I and/or U. Physically, this situation corresponds to conditions where folding intermediates are detectable experimentally and so can in principle be included in the analysis. The only situations in which this poses a serious difficulty are either (1) when (non-reactive) non-native conformational states are significantly populated but cannot be detected or (2) when multiple reactive states exist *and* their populations relative to one another change greatly with T. Note, as the analysis associated with eq 31 illustrates, knowledge of the precise nature and identity of the aggregation-prone state(s) is necessary to determine k_{11} with high accuracy but is not necessarily required to make reasonable extrapolations of shelf life or storage stability from accelerated conditions.

Additional Considerations: Reversible Aggregation Steps and Fibril Formation. There is increasing evidence that many if not all proteins can in principle be induced to form ordered, typically fibrillar, irreversible aggregates analogous to those implicated in a number of degenerative diseases,[9,15−18,38,60] and characterized by highly specific arrays of misfolded monomers having significant β-sheet content. This occurs in some cases even when the native conformational state contains predominantly α-helix rather than β-sheet structure,[68] suggesting that a conformational transition to a rare, off-pathway folding intermediate may be a critical precursor to the formation of ordered irreversible aggregates.[20,38,60,68] Further, available kinetic evidence for the early stages of fibril formation (i.e., protofilament and filament formation) is consistent with a mechanistic picture in which initial aggregation steps are appreciably reversible and grow reversibly up to some critical size.[16,17,20,68] Once that critical size is achieved, the formation of an irreversible, fibril precursor composed of some minimum number of β-sheet-containing monomers is hypothesized to occur.[16,17,20] This fibril precursor is often termed a "nucleus" by analogy with phase transformation kinetics since once such (experimentally unobservable) "nuclei" are formed they appear to grow irreversibly by the addition of monomeric protein at specific sites. Whereas such an analogy to true phase-transformation kinetics has yet to be shown to have a rigorous basis, phenomenologically it provides a very insightful way to describe the observed lag times, near-linear growth regimes, and effects of "seeding" with preformed fibrils on the kinetics of fibril formation in a number of systems.[16,17,20,60]

It is important to note that the model presented and analyzed in this report is not capable of describing such kinetics in its present form. It appears more appropriate to restrict this to protein systems that form disordered aggregates. To capture the

physics described above, the present model must be modified to include effects such as reversible (early-stage) aggregation steps, significant global conformational changes of monomers within a reversible aggregate, "nucleation" of ordered irreversible aggregates or nuclei, addition of monomers (native and non-native) to growing nuclei, and the resulting kinetic competition with those steps leading to amorphous aggregates such as described in the present report. Such considerations were beyond the scope of this article but have been initiated in our labs and will be included as part of a future report.

V. Summary and Conclusions

The predictions and behavior of an extended Lumry−Eyring model of protein degradation via irreversible aggregation were investigated over a broad range of parameter values representing conditions relevant to biological, biochemical, and commercial protein systems. Four distinct types of kinetic behavior were observed via direct numerical simulation and are well described by approximate analytical expressions derived here. Type IA behavior arises when all "reactive" aggregate species remain soluble during the progress of the aggregation process, and all aggregates are similarly reactive with respect to further aggregation steps. Type IB behavior is a subset of Type IA for systems in which aggregation steps involving monomers and (preformed) aggregates are kinetically preferred over those steps involving only aggregates. Type II behavior occurs when intermediate aggregate levels are comparable to or greater than those of the reactive monomer species, and all reactive aggregates remain below their solubility limits. This class of kinetic behavior is anticipated primarily for systems that form large aggregates ($n^* \approx 100$ or larger) and/or that have low initial monomer levels and relatively small rate coefficients for the addition of monomers to preexisting aggregates. Type III behavior occurs whenever one or more reactive, intermediate molecular weight aggregates species is forced to remain constant during a significant portion of the overall aggregation process (e.g., because of saturation at a solubility limit). In terms of the kinetics of monomer loss (a typical experimental quantity for monitoring shelf life), the apparent reaction order(s) with respect to soluble monomer are purely second-order for Types IA and IB, purely first order for Type II, and mixed first- and second-order for Type III. For Type III behavior, the second-order term dominates at higher temperature (larger fractional population of non-native monomers) and becomes comparable to the first-order contribution at very low temperature.

The analysis presented here indicates the primary determinants of the time scale for monomer loss (i.e., shelf life), and more importantly, the temperature dependence of that shelf life includes (1) the intrinsic rate coefficients for aggregation steps involving non-native monomeric protein—either that for irreversible dimerization or for the sequential addition of monomer to preexisting aggregates and (2) the thermodynamics of conformational equilibrium between native and non-native conformations. The remaining kinetic parameters primarily determine only the qualitative kinetic behavior observed. The importance of conformational equilibria on kinetic stability is not a new concept.[5,7,15,19,35,44,45,50] In contrast, the work presented here clearly indicates that the *intrinsic* kinetics of aggregation are an equally important consideration and can play a dominant role in determining both the qualitative and quantitative details of observed protein aggregation profiles with time.

The results in this report also provide a basis for the rational extrapolation of kinetic data on protein aggregation from so-called accelerated conditions to ones representative of realistic long-term storage. The primary conclusions are that shelf-life predictions extrapolated from accelerated conditions should be viable provided that the intrinsic aggregation kinetics for a given protein are accurately categorized, sufficient kinetic and thermodynamic data are available to extrapolate separately the contributions from shifts in conformational equilibrium from the intrinsic aggregation kinetics, and only one intrinsic kinetic rate law holds over the range of the extrapolation variable. It is concluded that conservative bounds on extrapolated shelf life may be possible from typical thermodynamic and accelerated kinetic data, even without knowledge of the precise nature or structural details of the "aggregation-prone" conformational state.

VI. Appendix A: Supplemental Derivation for Equation 9

For illustration purposes, eq 9 is derived here without first converting from R_A to M. Applying a steady-state approximation[57] to all $A^{(j)}$ for $j > 1$ (i.e., $dA^{(j)}/dt = 0$) and summing over eq 2.4 to $2.n^*+2$ gives

$$k_{11}R_A{}^2 = \left[\frac{1}{2}\sum_{i=2}^{n^*-1}\sum_{j=2}^{n^*-(i+1)} k_{ij}A^{(i)}A^{(j)} + \frac{3}{2}\sum_{i=2}^{(n^*-1)/2} k_{ii}(A^{(i)})^2\right] + \left[\sum_{i=2}^{n^*-1}\sum_{j=n^*-i}^{n^*-1} k_{ij}A^{(i)}A^{(j)} + 2\sum_{i=n^*/2}^{n^*-1} k_{ii}(A^{(i)})^2\right] \quad \text{(A.1)}$$

where the first and second brackets each account for aggregation steps that produce aggregates with $j < n^*$ and $j \geq n^*$, respectively. Thus, if n^* is determined by eq 3, the second bracket in eq A.1 gives the rate of production of precipitated protein. Alternatively, if n^* is determined by a kinetic bottleneck (i.e., $k_{ij} \rightarrow 0$ for $i, j \geq n^*$), then this term corresponds to the rate of growth of a (polydisperse) high molecular weight soluble aggregate species such as that sometimes observed in size-exclusion HPLC.[21,54]

Taking the limit as $n^* \rightarrow \infty$ and assuming that k_{ij} is not a strong function of i and j ($ij \neq 1, 1$), we approximate eq A.1 by

$$0 \approx k_{11}R_A{}^2 - \frac{1}{2}\langle k_{ij}\rangle\left(\sum_{j=2}^{n^*-1} A^{(j)}\right)_{ss}{}^2 - \frac{1}{2}\langle k_{ij}\rangle\left(\sum_{i=2}^{n^*-1} A^{(i)}\sum_{j=n^*-i}^{n^*-1} A^{(j)}\right)_{ss} - \langle k_{ij}\rangle(A^{(n^*-1)})_{ss}{}^2 - \langle k_{1j}\rangle(A^{(n^*-1)})_{ss} \quad \text{(A.2)}$$

where $\langle\cdots\rangle$ indicates an average over $j > 1$ and the ss subscript denotes steady state. Arguing that in the limit of large n^* the average $A^{(j)}_{ss}$ levels will not be highly sensitive to j, the second parenthetical term on the right-hand side of eq A.2 can be approximated by

$$\left(\sum_{i=2}^{n^*-1} A^{(i)}\sum_{j=n^*-i}^{n^*-1} A^{(j)}\right)_{ss} \approx \beta\left(\sum_{j=2}^{n^*-1} A^{(j)}\right)_{ss}{}^2 \quad \text{(A.3)}$$

where β is a constant of order 1. The fourth term is negligible, and the fifth term is approximated by

$$\langle k_{1j}\rangle(A^{(n^*-1)})_{ss} \approx \langle k_{1j}\rangle\frac{\left(\sum_{j=2}^{n^*-1} A^{(j)}\right)_{ss}}{n^*-2} \quad \text{(A.4)}$$

The third term is essentially the same as the larger-aggregate contributions captured in the second term (i.e., it arises from

those aggregation events that produce *j*-mers with $j \geq n^*$). Combining eq A.2 with eq A.4 gives eq 9.

VII. Appendix B: Derivation of Equation 30

We express k_{11} in the spirit of classic transition-state theory[57]

$$k_{11} = Ck_BT\exp(-\Delta G^*/k_BT) \quad \text{(B.1)}$$

where k_B is Boltzmann's constant, ΔG^* is the activation free energy for the dimerization of two R_A molecules, and C is an unspecified proportionality constant that depends on the nature of the barrier-crossing event(s) and their efficiency.[57,69] Expanding ΔG^* about T_H and assuming a constant Δc_p^* gives

$$-RT_H\ln\left(\frac{k_{11}}{k_{11}(T_H)}\right) = T_H(\Delta H^*|_{T_H} - \Delta c_p^*)\left(\frac{T_H}{T} - 1\right) + \Delta c_p^* T_H \ln\frac{T_H}{T} \quad \text{(B.2)}$$

where $\Delta H^*|_{T_H}$ is the activation enthalpy for k_{11} at T_H. Taking eq B.2 with $\Delta c_p^* = 0$, subtracting from the analogous expression with a nonzero Δc_p^*, and rearranging terms gives eq 30.

VIII. Appendix C: Derivation of Equation 31

From eq 22 and the definition of f_R, we have

$$\frac{\mathrm{d}\ln k_{11}^{(\mathrm{app})}}{\mathrm{d}(1/T)} = \frac{\mathrm{d}\ln k_{11}}{\mathrm{d}(1/T)} + 2\left(1 - \frac{K_{NR}}{1+K_{NR}}\right)\frac{\mathrm{d}\ln K_{NR}}{\mathrm{d}(1/T)} \quad \text{(C.1)}$$

By treating k_{11} as Arrhenius, considering T significantly below T_M for the $N-R_A$ transition ($K_{NR} \ll 1$), and using the van't Hoff relation for K_{NR}, eq C.1 becomes

$$R\frac{\mathrm{d}\ln k_{1,1}^{(\mathrm{app})}}{\mathrm{d}(1/T)} = -E_{a,1,1} - 2\Delta H_{NR}(T_{ref}) - 2\Delta c_p(T - T_{ref}) \quad \text{(C.2)}$$

where we have assumed that the heat capacity change for the $N-R_A$ transition is independent of T and T_{ref} is an arbitrary reference T. Choosing $T_{ref} = T_H$ and integrating from T_H down to some arbitrary T, we have

$$\ln\frac{k_{11}^{(\mathrm{app})}(T)}{k_{11}^{(\mathrm{app})}(T_H)} = -(E_{a,11} + 2\Delta H_{NR}(T_H))\left(\frac{1}{T} - \frac{1}{T_H}\right) + 2\Delta c_p\int_{T_H}^{T}\mathrm{d}T'\left(\frac{1}{T'} - \frac{T_H}{(T')^2}\right) \quad \text{(C.3)}$$

Equation 31 follows from evaluating the integral in eq C.3 and defining

$$E_{a,11}^{(\mathrm{app})} = E_{a,11} + 2\Delta H_{NR}(T_H) \quad \text{(C.4)}$$

Acknowledgment. P. Tessier is thanked for a critical review of and helpful suggestions regarding the manuscript.

References and Notes

(1) Bummer, P. M.; Koppenol, S. *Drugs Pharm. Sci.* **2000**, *99*, 5−69.
(2) Cleland, J. L.; Powell, M. F.; Shire, S. J. *Crit. Rev. Ther. Drug Carrier Syst.* **1993**, *10*, 307−377.
(3) Middaugh, C. R.; Edwards, K.-L. T. *Exp. Opin. Invest. Drugs* **1998**, *7*, 1493−1500.
(4) Pikal, M. J. In *Freeze-Drying/Lyophilization of Pharmaceutical and Biological Products;* Rey, L., May, J. C., Eds.; Marcel Dekker: New York, 1999; pp 161−198.
(5) Remmele, R. L., Jr.; Bhat, S. D.; Phan, D. H.; Gombotz, W. R. *Biochemistry* **1999**, *38*, 5241−5247.
(6) Grillo, A. O.; Edwards, K.-L. T.; Kashi, R. S.; Shipley, K. M.; Hu, L.; Besman, M. J.; Middaugh, C. R. *Biochemistry* **2001**, *40*, 586.
(7) Kendrick, B. S.; Carpenter, J. F.; Cleland, J. L.; Randolph, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14142.
(8) Finke, J. M.; Roy, M.; Zimm, B. H.; Jennings, P. A. *Biochemistry* **2000**, *39*, 575−583.
(9) Fink, A. L. *Folding Des.* **1998**, *3*, R9−R23.
(10) Baier, S.; McClements, D. J. *J. Agric. Food Chem.* **2001**, *49*, 2600−2608.
(11) Verheul, M.; Roefs, S. P. F. M.; de Kruif, K. G. *J. Agric. Food Chem.* **1998**, *46*, 896−903.
(12) Foguel, D.; Robinson, C. R.; de Sousa, P. C., Jr.; Silva, J. L.; Robinson, A. S. *Biotechnol. Bioeng.* **1999**, *63*, 552−558.
(13) Shi, P.-Y.; Maizels, N.; Weiner, A. M. *BioTechniques* **1997**, *23*, 1036−1038.
(14) Jaenicke, R. *Philos. Trans. R. Soc. London, Ser. B* **1995**, *348*, 97−105.
(15) Kim, Y.-S.; Wall, J. S.; Meyer, J.; Murphy, C.; Randolph, T. W.; Manning, M. C.; Solomon, A.; Carpenter, J. F. *J. Biol. Chem.* **2000**, *275*, 1570−1574.
(16) Wood, S. J.; Wypych, J.; Steavenson, S.; Louis, J.-C.; Citron, M.; Biere, A. L. *J. Biol. Chem.* **1999**, *274*, 19509−19512.
(17) Esler, W. P.; Stimson, E. R.; Ghilardi, J. R.; Vinters, H. V.; Lee, J. P.; Mantyh, P. W.; Maggio, J. E. *Biochemistry* **1996**, *35*, 749−757.
(18) Nielsen, L.; Khurana, R.; Coats, A.; Frokjaer, S.; Brange, J.; Vyas, S.; Uversky, V.; Fink, A. L. *Biochemistry* **2001**, *40*, 6036−6046.
(19) Mulkerrin, M. G.; Wetzel, R. *Biochemistry* **1989**, *28*, 6556−6561.
(20) Lomakin, A.; Teplow, D. B.; Kirschner, D. A.; Benedek, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7942−7947.
(21) Yoshioka, S.; Aso, Y.; Izutsu, K.-i.; Kojima, S. *Pharm. Res.* **1994**, *11*, 1721−1725.
(22) Plaza del Pino, I. M.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 58.
(23) Dill, K. A. *Biochemistry* **1990**, *29*, 7133−7155.
(24) Zaks, A. In *Stability of Protein Pharmaceuticals*; Ahern, T. J., Manning, M. C., Eds.; Plenum Press: New York, 1992; pp 249−272.
(25) Fink, A. L.; Calciano, L. J.; Goto, Y.; Kurotsu, T.; Palleros, D. R. *Biochemistry* **1994**, *33*, 12504−12511.
(26) Petsev, D. N.; Thomas, B. R.; Yau, S.-T.; Vekilov, P. G. *Biophys. J.* **2000**, *78*, 2060−2069.
(27) Kheterpal, I.; Williams, A.; Murphy, C.; Bledsoe, B.; Wetzel, R. *Biochemistry* **2001**, *40*, 11757−11767.
(28) Hammarstrom, P.; Persson, M.; Freskgard, P.-O.; Martensson, L.-G.; Andersson, D.; Jonsson, B.-H.; Carlsson, U. *J. Biol. Chem.* **1999**, *274*, 32897−32903.
(29) Smellar, L.; Rubens, P.; Heremans, K. *Biochemistry* **1999**, *38*, 3816−3820.
(30) Narhi, L.; Wood; S. J.; Steavenson; S.; Jiang, Y.; Wu, G. M.; Anafi, D.; Kaufman, S. A.; Martin, F.; Sitney, K.; Denis, P.; Louis, J.-C.; Wypych, J.; Biere, A. L.; Citron, M. *J. Biol. Chem.* **1999**, *14*, 9843−9846.
(31) Naiki, H.; Hasegawa, K.; Yamaguchi, I.; Nakamura, H.; Gejyo, F.; Nakakuki, K. *Biochemistry* **1998**, *37*, 17882−17889.
(32) We use the term conformational state within its thermodynamic and statistical thermodynamic sense to indicate *all* conformations of a protein and its environment (i.e., solvent and additional solutes) within a given ensemble, or rational subdivision of an ensemble, consistent with a set of macroscopic observables specific to that state.
(33) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.
(34) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *J. Chem. Phys.* **1994**, *101*, 6052−6062.
(35) Krishnan, S.; Chi, E. Y.; Webb, J. N.; Chang, B. S.; Shan, D.; Goldenberg, M.; Manning, M. C.; Randolph, T. W.; Carpenter, J. F. *Biochemistry*, **2002**, *41*, 6422.
(36) Lindner, R. A.; Treweek, T. M.; Carver, J. A. *Biochem. J.* **2001**, *354*, 79−87.
(37) Oliveberg, M. *Acc. Chem. Res.* **1998**, *31*, 765.
(38) Khurana, R.; Gillespie, J. R.; Talapatra, A.; Minert, L. J.; Ionescu-Zanetti, C.; Millett, I.; Fink, A. L. *Biochemistry* **2001**, *40*, 3525−3535.
(39) The term "unfolded" does not imply a state devoid of structure such as the canonical random-coil polymer model.[8,34,38,40−42] On the contrary, the nature of polypeptides in aqueous solvents appears to force significant conformational structuring under all but the most extreme conditions.[38,40,41]
(40) Kauzmann, W. In *Advances in Protein Chemistry;* Anfinsen, C. B., Anson, M. L., Bailey, K., Edsall, J. T., Eds.; Academic Press: New York, 1959; Vol. 14, pp 1−63.
(41) Pappu, R. V.; Rose, G. D. *Protein Sci.* **2002**, *11*, 2437−2455.
(42) Rucker, A. L.; Creamer, T. R. *Protein Sci.* **2002**, *11*, 980−985.

(43) For the purposes of the extended Lumry−Eyring model employed here as well for extrapolating shelf-life predictions, it is not necessary to know the detailed identity (e.g., in terms of structural details) of the aggregation-prone state(s). The major requirement is that it(they) remain the predominant aggregation-prone state(s) over the range of the variable being extrapolated.

(44) Lumry, R.; Eyring, H. *J. Phys. Chem.* **1954**, *58*, 110−120.

(45) Zale S. E.; Klibanov A. M. *Biotech. Bioeng.* **1983**, *25*, 2221−2230.

(46) Sanchez-Ruiz, J. M. *Biophys. J.* **1992**, *61*, 921−935.

(47) Tello-Solis, S.; Hernandez-Arana, A. *Biochem. J.* **1995**, *31*, 969.

(48) La Rosa, C.; Milardi, D.; Grasso, D.; Guzzi, R.; Sportelli, L. *J. Phys. Chem.* **1995**, *99*, 14864−14870.

(49) Kasraian, K.; Kuzniar, A.; Earley, D.; Kamicker, B.; Wilson, G.; Manion, T.; Hong, J.; Reiber, C.; Canning, P. *Pharm. Dev. Technol.* **2001**, *6*, 439−445.

(50) Webb, J. N.; Webb, S. D.; Cleland, J. L.; Carpenter, J. F.; Randolph, T. W. *Proc. Natl. Acad. Sci U.S.A.* **2001**, *98*, 7259.

(51) Roberts, C. J.; Darrington, R. T.; Whitley, M. B. *J. Pharm. Sci.*, in press.

(52) Shelf life is defined for the present report as the time for 10% loss of monomer or 90% remaining (i.e., $M(t_{90}) = 0.9M_0$).

(53) Fields, G. B.; Alonso, D. O. V.; Stigter, D.; Dill, K. A. *J. Phys. Chem.* **1992**, *96*, 3974.

(54) Bartkowski, R. E.; Kitchel, R.; Peckham, N. R.; Margulis, *J. Protein Chem.* **2002**, *21*, 137.

(55) Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859−1877.

(56) Pace, C. N.; Shirley, B. A.; Thomson, J. A. In *Protein Structure*; Creighton, T. E., Ed.; Oxford University Press: Oxford, U.K., 1990; p. 311.

(57) Laidler, K. J. *Chemical Kinetics*, 3rd ed.; HarperCollins: New York, 1987.

(58) Mikol, V.; Giege, R. In *Crystallization of Nucleic Acids and Proteins*; Ducruix, A., Giege, R., Eds.; Oxford University Press: New York, 1992; pp 219−240.

(59) A simple argument serves to illustrate this point. For low extents of reaction ($M/M_0 \approx 0.9-1.0$), a zeroth- or first-order approximation is reasonable. Therefore, $t_{90} \approx 0.1/k_{obs}$, with k_{obs} the observed rate coefficient. For systems of practical interest, t_{90} is on the order of 10 to 100 days and is much longer for commercial products. For unfolding-limited aggregation, $k_{obs} = k_u$, which therefore requires $k_u \approx 10^{-7}$ to 10^{-6} s^{-1} or equivalently the characteristic time scale for unfolding to be on the order of (at least) 10^2-10^3 days.

(60) Teplow, D. B. *Amyloid* **1998**, *5*, 121−142.

(61) Barzykin, A. V.; Shushin, A. I. *Biophys. J.* **2001**, *80*, 2062−2073.

(62) Privalov, P. L. *Adv. Protein Chem.* **1979**, *33*, 167−241.

(63) Solc, K.; Stockmayer, W. H. *Int. J. Chem. Kinet.* **1973**, *5*, 733−752.

(64) Remmele, R. A. Arrhenius Consequences of IL-1R(II) Aggregation in the Vicinity of the Tm. Formulation Strategies for Biopharmaceuticals, IBC USA, Miami, FL, 2002.

(65) Segawa, S.-I.; Sugihara, M. *Biopolymers* **1984**, *23*, 2473−2788.

(66) Oliveberg, M.; Tan, Y.-J.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8926−8929.

(67) Yoshioka, S.; Izutsu, K.; Aso, Y.; Takeda, Y. *Pharm. Res.* **1991**, *8*, 480−484.

(68) Goers, J.; Permyakov, S. E.; Permyakov, E. A.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2002**, *41*, 12546−12551.

(69) Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 11−27.