Nucleation, Growth, and Activation Energies for Seeded and Unseeded Aggregation of α-Chymotrypsinogen A†

Jennifer M. Andrews,‡ William F. Weiss IV,‡ and Christopher J. Roberts*

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

Received September 20, 2007; Revised Manuscript Received November 14, 2007

ABSTRACT: The intrinsic time scales for nonnative aggregate nucleation ($\tau_n^{(0)}$) and chain growth ($\tau_g^{(0)}$) were determined for α-chymotrypsinogen A as a function of temperature under acidic conditions where the resulting aggregates do not appreciably condense. Previous results (Andrews and Roberts (2007) *Biochemistry 46*, 7558) indicated that the product $\tau_n^{(0)}\tau_g^{(0)}$ increases with increasing temperature but could not distinguish $\tau_n^{(0)}$ and $\tau_g^{(0)}$. Separate experimental values of $\tau_n^{(0)}$ and $\tau_g^{(0)}$ are reported here from two approaches based on either (i) combining unseeded monomer loss kinetics with static light scattering of the resulting aggregates or (ii) seeded monomer loss kinetics as a function of number concentration of seed. Values of $\tau_n^{(0)}$ and $\tau_g^{(0)}$ from (i) and (ii) agree quantitatively, and indicate that nucleation has a large, negative effective activation energy (*ca.* −76 kcal/mol) while growth has at most a weak dependence on temperature. The results are consistent with a model in which nucleation requires significant conformational changes within a nonnative oligomer, beyond those for monomer unfolding. The results more generally illustrate the potential utility of approaches (i) and (ii) for quantitatively determining *in vitro* $\tau_n^{(0)}$ and $\tau_g^{(0)}$ values, as well as how the effects of seeding can be predicted purely from unseeded kinetics and static light scattering measurements prior to significant aggregate condensation.

Nonnative protein aggregation is a common problem during the development of protein products such as biopharmaceuticals (*1*), and has established links to neurodegenerative diseases including Alzheimer's and Parkinson's (*2*). Because aggregation is often net irreversible under the conditions that aggregates form *in vitro* or *in vivo*, it is important to be able measure and discriminate among the key stages that determine aggregation kinetics in order to better understand the underlying mechanism. In particular, questions regarding the controlling physics and chemistry for the nucleation stage in nonnnative aggregation remain among the most difficult to address experimentally. This is due at least in part to limitations of biophysical techniques to directly characterize and/or isolate the relatively unstable and poorly populated nuclei implicated in nucleation (*3*). A lucrative experimental approach has been to infer details of the aggregation process from qualitative and quantitative analysis of aggregation kinetics in the context of mechanistic mathematical models (*3*−*11*). This is the approach adopted here, with analyses based on a Lumry−Eyring nucleated polymerization (LENP) model (*4*) that incorporates many of the simpler models available and that reliably describes the experimental aggregation behavior of α-chymotrypsinogen A (aCgn) (*12*, *13*).

aCgn forms soluble, nonnative, noncovalent aggregates at the elevated temperatures, low ionic strength, and acidic conditions in this report (*12*, *13*). The aggregates are uncondensed, linear, semiflexible polymers in which nonnative protein chains are the constituent monomers (*13*). Characteristic dimensions of the aggregates are *ca*. 10 nm (diameter) and hundreds of nm (contour length), with a monomer diameter of *ca*. 4 nm. These dimensions are similar to those observed for a number of systems which form rigid filaments and/or fibrils (*1*, *13*, *14*). The aggregates do not appreciably condense or coalesce to supramolecular structures such as fibrils or precipitates under the conditions of study here (*13*). Soluble, prefibrillar aggregates have been implicated as potentially toxic species in the context of neurodegenerative diseases (*15*−*17*), and are also more typical than their fibrillar counterparts in industrial settings (*18*, *19*). These observations suggest that the insights gained regarding aCgn aggregation, and methods to experimentally quantify its nucleation and growth kinetics, may be of general utility.

Figure 1 depicts the LENP model as it applies for aCgn (*12*, *13*). In stage I, folded (monomer) proteins reversibly unfold to molten globule conformers that make up the aggregation-prone or reactive monomer state (R). The equilibrium constant between native and (partially) unfolded monomer is denoted K_{NR} to remain consistent with previous work (*4*). aCgn unfolding is effectively two-state and equilibrates rapidly compared to aggregation under the conditions considered here (*12*). In stage II, R monomers reversibly self-associate to form oligomers R_j with equilibrium constants K_j (j denotes the number of monomers in a reversible oligomer; $j = 2$ to $x - 1$, where x denotes the number of monomers in the nucleus). Stage III begins with

† Funding Sources: University of Delaware Research Foundation. National Institutes of Health (Grants R01 EB006006, P20 RR015588). Boehringer Ingelheim Pharmaceuticals. Merck and Co.

* To whom correspondence should be addressed: cjr@udel.edu. Tel: 302-831-0838. Fax: 302-831-1048.

‡ Authors contributed equally to this work.

10.1021/bi7019244 CCC: $40.75
Published on Web 01/24/2008

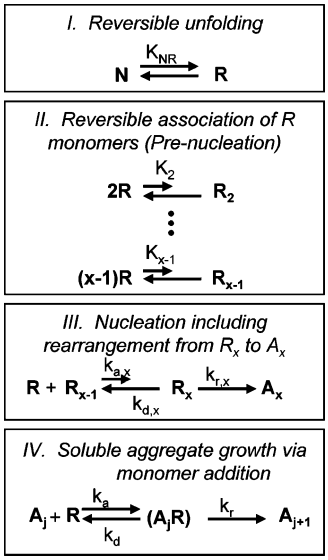

FIGURE 1: LENP aggregation scheme as it applies to aCgn at pH 3.5, 10 mM citrate.

the reversible association of R with R_{x-1} to form the "prenucleus" R_x with forward and reverse rate constants $k_{a,x}$ and $k_{d,x}$, respectively. This is followed by an irreversible structural conversion or conformational rearrangement of the protein chains in R_x, with rate constant $k_{r,x}$, to create a net irreversible aggregate A_x. Aggregate growth by chain polymerization occurs in stage IV. To remain general and consistent with the need for a structural conversion in stage III, soluble aggregates A_j are able to first reversibly associate with monomer R to yield (A_jR). This is followed by an irreversible conformational rearrangement of the R chain to produce the irreversible aggregate A_{j+1}. The corresponding rate coefficients are defined in Figure 1. Aggregate−aggregate association or condensation is also possible in the full LENP model, but is not included here because the aggregates of aCgn do not condense under the solution conditions tested here (*13*).

Figure 1 is a limiting case of the general LENP model, and was denoted as $\delta = 1$, $n^* \rightarrow \infty$ in ref (*4*). Physically, $\delta = 1$ corresponds to aggregate growth via monomer addition, and $n^* \rightarrow \infty$ indicates that aggregates are soluble and uncondensed, and continue to consume monomer until the monomer supply is exhausted. The resulting model equations (eq 1 below) have a relatively simple form that is similar to those for native polymerization models (*3*, *20*) except that those previous models do not account for unfolding or conformational changes required to nucleate and propagate aggregation (*21*).

$$\frac{dm}{dt} = -\frac{x}{\tau_n}m^x - \frac{1}{\tau_g}m\sigma \tag{1a}$$

$$\frac{d\sigma}{dt} = \frac{1}{\tau_n}m^x \tag{1b}$$

In eq 1, m is the concentration of monomer [M] scaled by its initial concentration C_0, t is time, and σ is the sum across the concentrations of all soluble aggregates $\sum_j[A_j]$ scaled by C_0 (*4*). All species concentrations are expressed on a number per volume basis, e.g., molarity. τ_n and τ_g are the overall nucleation and growth time scales, respectively, and are equivalent to the inverse rate coefficients of nucleation and growth. They are defined in eq 2 (*4*).

$$\tau_n \equiv \tau_n^{(0)} f_R^{-x}(C_{ref}/C_0)^{x-1} \tag{2a}$$

$$\tau_g \equiv \tau_g^{(0)} f_R^{-1}(C_{ref}/C_0) \tag{2b}$$

The intrinsic time scale of nucleation is $\tau_n^{(0)} \equiv (k_{nuc}K_{x-1}C_{ref}^{x-1})^{-1}$, with $k_{nuc} = k_{a,x}k_{r,x}/(k_{d,x} + k_{r,x})$. The intrinsic time scale of growth is $\tau_g^{(0)} \equiv (k_g K_{RA}^{\delta-1} C_{ref}^{\delta})^{-1}$, with $k_g = k_a k_r/(k_d + k_r)$. $\tau_n^{(0)}$ and $\tau_g^{(0)}$ are functions of only temperature and solvent composition, and are defined as the values of τ_n and τ_g, respectively, when the unfolded (reactive) monomer fraction (f_R) is 1 and C_0 equals the standard state or reference concentration (C_{ref}) for the free energy of prenucleus formation (*4*). Using the intrinsic time scales eliminates the dependences of τ_n and τ_g on C_0 and on ΔG_{NR}^0 ($= -k_B T \ln K_{NR}$) that can otherwise convolute comparison of aggregation kinetics among different temperatures or protein concentrations (*4*, *11*−*13*, *18*). The need to account for the dependence on ΔG_{NR}^0 or f_R is also a distinguishing feature of nonnative aggregation that precludes it being accurately modeled with native polymerization models (*3*, *20*) when changes in temperature and solvent composition are included (*21*).

In the absence of seeding, and for the canonical case with $\tau_n \gg \tau_g$, the value of σ reaches a limiting, steady-state value (σ_{ss}) before the value of m falls much below 1. As a result, monomer loss profiles over one or more half-lives are well described by (*4*, *12*)

$$\frac{dm}{dt} = -\frac{\sigma_{ss}}{\tau_g}m = -k_{obs}m \tag{3}$$

with k_{obs} defined implicitly in eq 3. Numerical solutions of eq 1 for $\tau_n \gg \tau_g$ show (*4*)

$$\sigma_{ss} \cong 2(\tau_g/\tau_n)^{1/2} \tag{4}$$

$$k_{obs} = 2(\tau_n\tau_g)^{-1/2} \tag{5}$$

Equations 3 and 5 and their associated derivations (*4*) illustrate that it is not generally possible to determine separate values of τ_n and τ_g from any experiment based solely on monomer loss, extent-of-reaction, or mass of protein converted to aggregate (*3*, *4*, *20*). In the absence of seeding, this conclusion holds even as $t \rightarrow 0$ (*3*).

The LENP analysis indicates that σ_{ss}, and thus separate values of τ_n and τ_g, can be obtained from experimental values of m combined with the weight average molecular weight (M_w) of an aggregated sample via (*4*)

$$\frac{M_w}{MW_{mon}} = m + \frac{(1 + c_x^2)(1 - m)^2}{\sigma_{ss}} \tag{6}$$

with MW_{mon} denoting the monomer molecular weight. In eq 6, values of the numerical constant c_x depend only on x and are determined from solution to the full model in which all aggregate species are accounted for explicitly ($c_x = 0.3$, 0.25 for $x = 3$, 4, respectively (*4*)). Physically, c_x is the ratio of the variance to the mean of the distribution of aggregate

Table 1: Summary of Results from Unseeded Monomer Loss Kinetics and Light Scattering and from Seeding Experiments[a]

T (°C)	M_w/MW_{mon}	$\sigma_{ss}C_0$ (nM)[b]	x	f_R	$k_{obs} \times 10^3$ (min^{-1})	$\tau_n \times 10^{-3}$ (min)	$\tau_g \times 10^2$ (min)
55	374 ± 8	120 ± 4	3	0.078	2.5 ± 0.5	92 ± 18[b]	78 ± 16[b]
						33 ± 17[c]	190 ± 40[c]
60	96 ± 2	460 ± 20	3	0.694	46 ± 5	1.3 ± 0.1[b]	16 ± 2[b]
						1.2 ± 0.2[c]	12 ± 1[c]
62.5	104 ± 1	430 ± 10	3	0.920	63 ± 9	1.0 ± 0.1[b]	11 ± 2[b]
65	156 ± 3	280 ± 10	4	0.982	62 ± 4	1.2 ± 0.08[b]	5.7 ± 0.4[b]
						1.2 ± 0.3[c]	4.8 ± 0.4[c]
67.5	140 ± 2	310 ± 10	3	0.996	58 ± 10	1.4 ± 0.2[b]	8.6 ± 1.4[b]

[a] Nucleation and growth time scales in the table are distinct from those plotted in Figure 2; the latter follow from eq 2 after accounting for differences in f_R. Uncertainties correspond to 95% confidence intervals for fitted parameters, or result from both propagation of experimental uncertainties and fitted confidence intervals. [b] Unseeded experiments. [c] Seeded experiments.

sizes (*4*). Equation 6, together with experimental values for M_w and m as $m \rightarrow 0$, provides the value of σ_{ss} for a given sample condition. The τ_n and τ_g values then follow from rearranging eqs 4 and 5 to give $\tau_n = 0.25(k_{obs}\sigma_{ss})^{-1}$ and $\tau_g = \sigma_{ss}/k_{obs}$. In this case, k_{obs} is obtained from regression of unseeded $m(t)$ data versus the integrated form of eq 3. This approach and analysis (*13*) are referred to here as the $k_{obs} - M_w$ method.

Alternatively, the LENP model predicts that separate τ_n and τ_g values can be regressed from monomer loss data for samples that are seeded with controlled levels of preformed soluble aggregates, by accounting for the different initial conditions that this implies for the solution of eq 1 ($m = 1$ and $\sigma = \sigma_0$ at $t = 0$). This also requires knowledge of the number concentration of seeds at time zero (i.e., $\sigma_0 > 0$). Although seeded kinetics are a common experimental approach to infer relative rates of aggregation or qualitative effects of seeding for different proteins or peptides, they are typically obtained with knowledge of only the mass concentration of seeds and little or no knowledge of the initial seed size(s) (*22*, *23*). The lack of information regarding number of seeds precludes such measurements from providing absolute, quantitative rates for growth. Our approach is based instead on determining σ_0 prior to seeding, and globally fitting the seeded monomer loss kinetics as a function of σ_0 in order provide independent values of nucleation and growth rates. This also obviates a need to otherwise employ chemical labels that might influence the measured rates (*24*, *25*).

The $k_{obs} - M_w$ and the seeded monomer loss approaches are both employed in this report to determine the effects of temperature on τ_n and τ_g for aCgn. The effects of temperature on $\tau_n^{(0)}$ and $\tau_g^{(0)}$ are also determined to infer mechanistic details regarding nucleation and growth in nonnative aggregation for aCgn. More generally, the results provide a quantitative test of whether seeded and unseeded experiments can provide self-consistent nucleation and growth kinetics, and an illustration of alternative ways to implement an LENP analysis that may be of use beyond the case of aCgn.

MATERIALS AND METHODS

Protein solutions were prepared as described previously (*12*, *13*). All solution conditions were pH 3.5, 10 mM sodium citrate. Unseeded and seeded isothermal monomer loss kinetics were monitored at selected temperatures (T) from 55 to 67.5 °C, with an initial monomer concentration (C_0) of 1.2 mg/mL. Monomer concentrations, [M], from samples quenched at selected time points were quantified by size exclusion chromatography (SEC) as described previously (*12*, *13*). The injection volume was 100 μL for 1.2 mg/mL samples and 20 μL for 4.8 mg/mL samples to avoid column overloading.

Unseeded kinetics for this C_0 and T range, with data over approximately one or more half-lives, follow eq 3 and yield values for k_{obs} by least-squares regression (*12*). For each sample condition, the resulting fully aggregated samples ($m \leq 0.05$) were analyzed by static light scattering (*13*) to obtain the corresponding values of the aggregate weight average molecular weight (M_w) and the average radius of gyration (R_g) (see also Supporting Information). The previously determined aggregate morphology, as indicated by the scaling of R_g with M_w, along with a lack of detectable condensation was confirmed to hold across the conditions considered here (not shown). Values of σ follow from substitution of the measured m and M_w values in eq 6, with previously determined values of c_x (*4*) and x (*12*, *13*). The second osmotic virial coefficient (B_{22}) at 20 °C for a purely monomeric solution of aCgn was determined from a nonlinear least-squares regression of static light scattering versus protein concentration against the standard Debye equation (*13*, *26*) (see also Supporting Information).

For seeded experiments, a stock of fully aggregated protein with $C_0 = 4.8$ mg/mL (187 μM initial monomer) was prepared by thermostatting an initially monomeric stock solution at 65 °C for a minimum of six times the half-life for monomer loss. The resulting number concentration of aggregates was determined from the M_w and eq 6 to be 2.05 ± 0.03 μM. Aliquots of this stock were combined gravimetrically with appropriate ratios of buffer solution and monomer stock solution (3.0 mg/mL, 117 μM monomer) to produce samples with initial monomer concentrations of 1.2 mg/mL (47 μM) and a series of values for the initial concentration of aggregates (σ_0). Values of σ_0/σ_{ss}^0 used in the seeding experiments were 0, 0.22, 0.48, 0.83, 1.71, 2.59 for 65 °C; 0, 0.13, 0.23, 0.48, 0.95, 1.39 for 60 °C; and 0, 1.67, 3.43, 6.80, 10.25 for 55 °C. The 95% confidence intervals for all σ_0/σ_{ss}^0 values are ±4% of the reported value. σ_{ss}^0 is the value of σ_{ss} for an unseeded sample with $C_0 = 1.2$ mg/mL at the temperature of interest (see Table 1). Monomer loss versus time was monitored for both seeded and unseeded samples at each of 55, 60, and 65 °C. The $m(t;\sigma_0)$ data for each temperature were simultaneously regressed against eq 1 with initial conditions of $m = 1$ along with the corresponding σ_0 values to provide a single value for each of τ_n and τ_g at each temperature. The x values were set at those determined previously (*4*), and do not signifi-

cantly affect the results (see below). The regression provided fitted values and 95% confidence intervals for τ_n and τ_g.

RESULTS

The monomer loss profiles from the 65 °C seeding experiments are shown in Figures 2A and 2B, and are representative of those at 55 and 60 °C (not shown). For clarity, Figure 2A contains the monomer loss profiles for $\sigma_0/\sigma_{ss}^0 = 0$, 0.22, and 0.48, and Figure 2B contains those for $\sigma_0/\sigma_{ss}^0 = 0.83$, 1.71, and 2.59. It is important to note that all data in Figures 2A and 2B were fit simultaneously using only τ_n and τ_g as the two fitting parameters for a given temperature. The fitted curves compare well with the experimental $m(t)$ data sets. The most notable systematic deviations between experiment and model in Figures 2A and 2B occur for the $m(t)$ profiles with low seed concentrations. At present our data do not offer a conclusive explanation for this, but it may be a result of the fact that experimental errors in one data set are being offset in the global fit to the others, and the fact that the fastest aggregation rates are reaching the limit of experimental resolution with our heating/quenching capabilities. This is consistent with the observation that systematic deviations are not apparent if each $m(t)$ data set is instead fit individually to eq 3 (not shown). The fitted values of τ_n and τ_g for each temperature are given in Table 1 with their corresponding 95% confidence intervals from the statistical uncertainty of the global fit. Values from the alternate $k_{obs} - M_w$ method are also provided in Table 1, with uncertainties propogated from the confidence intervals on fitted k_{obs} and M_w values. The results from the two methods are in semiquantitative or quantitative agreement across all temperatures. To the best of our knowledge, this is first time that quantitative agreement for separate rates of nucleation and growth has been shown between two orthogonal methods; one of which is invasive (e.g., seeding), while the other is not.

Figure 2C and its inset show the experimental half-lives (t_{50}) as a function of reduced seed concentration for 55, 60, and 65 °C (symbols). The curves are predicted t_{50} values obtained by solving eq 1 for the corresponding σ_0 values using the τ_n and τ_g values from k_{obs} and M_w for the unseeded experiments. The results are in quantitative agreement, and show that experimental seeded kinetics can be accurately predicted based solely on knowledge of τ_n and τ_g values in the absence of seeding. The relatively weak temperature dependence of the values for τ_n in Table 1 are reflective of the competition between the strong and opposite temperature dependences of f_R and $\tau_n^{(0)}$ (see below).

Figure 3 shows the intrinsic time scales calculated from eq 6, with separate points for values based on unseeded or seeded experiments (open and closed symbols, respectively), using $C_{ref} = 1$ mg/mL (monomer) = 38.9 μM, and the values of f_R and x obtained previously (*12*) and given in Table 1. With the exception of 65 °C, $x = 3$ for all selected temperatures. The calculations for 65 °C used $x = 4$ for consistency with previous work (*12*), although $x = 3$ may also be reasonable (*13*). Because the midpoint unfolding temperature for these conditions is significantly lower than 65 °C (*12*), f_R differs only slightly from 1. In addition, $C_0 \approx C_{ref}$ by our choice of C_{ref}, and therefore the precise value of x has little quantitative effect on the values of $\tau_n^{(0)}$

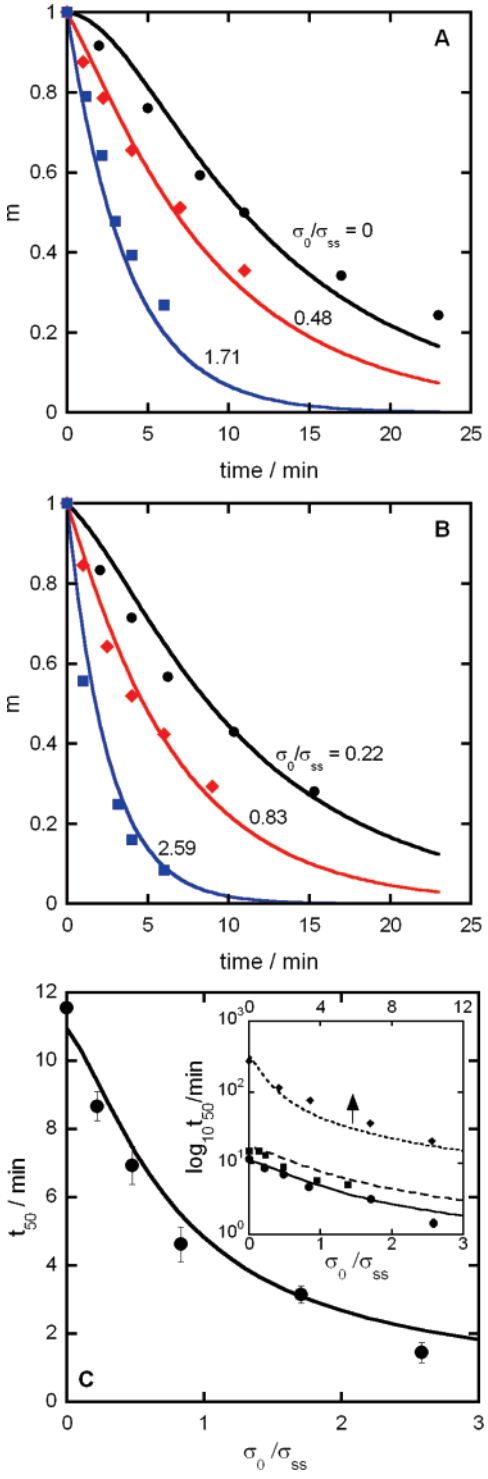

FIGURE 2: (A and B) Monomer concentration versus incubation time for seeding experiments using $C_0 = 1.2$ mg/mL monomer (47 μM), $T = 65$ °C, with different number concentrations of seed. Experimental uncertainties are smaller than the symbols. σ_0/σ_{ss} values are indicated next to the corresponding data sets in the plots. Lines are fits to eq 1 in which each data set has a different initial condition (σ_0) but shares the same values of τ_n and τ_g. (C) t_{50} versus σ_0/σ_{ss} from seeding at 65 °C (circles). The solid line represents the predicted curve from eq 1 with τ_n and τ_g values obtained from k_{obs} and M_w in the absence of seeding. The inset shows $\log(t_{50})$ versus σ_0/σ_{ss} from seeding at 65 (circles), 60 (squares), and 55 (diamonds) °C. The corresponding predicted curves from eq 1 with τ_n and τ_g from k_{obs} and M_w in the absence of seeding are also shown for 65 (solid line), 60 (dashed line), and 55 (dotted line) °C. For improved clarity in the inset, the upper x-axis scale corresponds to the 55 °C data.

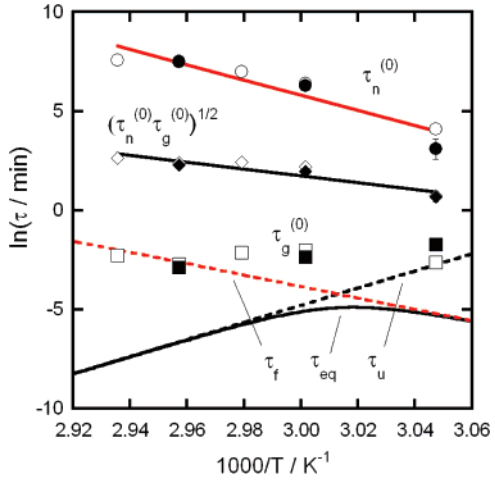

FIGURE 3: Rate diagram comparing intrinsic time scales of folding (τ_f) and unfolding (τ_u) with those for nucleation (circles, $\tau_n^{(0)}$), growth via chain elongation (squares, $\tau_g^{(0)}$), and net monomer loss (diamonds, $[\tau_n^{(0)}\tau_g^{(0)}]^{1/2} \sim k_{obs}^{-1}$). Open symbols denote values from the $k_{obs} - M_w$ method in the absence of seeding, and closed symbols denote values from seeded monomer loss. Confidence intervals are smaller than the size of the symbols if not discernible on the scale of the figure. Solid lines are linear fits to the data as plotted for $\tau_n^{(0)}$ and $[\tau_n^{(0)}\tau_g^{(0)}]^{1/2}$. The dashed red line shows τ_f extrapolated from lower temperature based on stopped-flow refolding kinetics combined with equilibrium unfolding free energies from (*12*).

or $\tau_g^{(0)}$ for 65 °C and higher temperatures. Characteristic time scales for folding and unfolding from ref (*12*) and $[\tau_n^{(0)}\tau_g^{(0)}]^{1/2}$ ($\sim 1/k_{obs}$, eq 3) are shown for comparison. τ_f and τ_u are defined as the inverse of the first-order rate coefficients for folding and unfolding, respectively, and were determined previously (*12*). The values for $\tau_n^{(0)}$, $[\tau_n^{(0)}\tau_g^{(0)}]^{1/2}$, and $\tau_g^{(0)}$ display reasonably Arrhenius behavior over the temperature range of the data, with effective activation energies defined by

$$\frac{d(\ln \tau)}{d(1/T)} = \frac{E_a}{k_B} \qquad (7)$$

where τ is the time scale of interest, E_a is the effective activation energy, and k_B is Boltzmann's constant. The fitted values of the effective activation energy for $\tau_n^{(0)}$ and $[\tau_n^{(0)}\tau_g^{(0)}]^{1/2}$ are -76 ± 27 and -34 ± 14 kcal/mol, respectively (uncertainties are 95% confidence intervals of the fit). $\tau_g^{(0)}$ is weakly dependent on temperature, with a fitted E_a that is not statistically different from zero.

DISCUSSION

We previously showed mathematically based on analysis of the LENP model (*4*) that quantitative determination of τ_n and τ_g from seeded kinetics requires knowledge of the number concentration of seeds, as opposed to only qualitative or semiquantitative values that can be obtained by the more common approach of seeding with fixed mass concentrations of seeds. The results in Figure 1, and their agreement with results from the $k_{obs} - M_w$ approach, provide experimental validation of the LENP analysis for obtaining quantitative, absolute values for τ_n and τ_g from seeded kinetics. This approach required the number concentration of seeds to be known prior to seeding: i.e., σ_0 was known from eq 6 and SLS measurements of the seed material. The results also highlight the utility of characterization and control of the seed material properties (*13*) for use in seeding.

Important features of Figure 3 are (*1*) the large negative E_a value for nucleation (≈ -70 kcal/mol) and its similarity to that for folding ($E_a \approx -60$ kcal/mol) over this temperature range; (*2*) the negligible temperature dependence for growth; (*3*) the large difference in magnitudes of $\tau_n^{(0)}$ ($\sim 10^2 - 10^3$ min) and $\tau_g^{(0)}$ ($\sim 10^{-1}$ min); these differences are even larger when τ_n and τ_g are compared for $f_R \ll 1$ (*cf.* Table 1). The lattermost feature is characteristic of a canonical nucleation and chain-polymerization mechanism in which growth is very rapid compared to nucleation.

The second feature suggests that growth may be diffusion limited, in that the E_a value in this case would correspond to the temperature coefficient of the viscosity of the solvent (*11*, *27*). With water as the solvent, $E_a \approx 3$ kcal/mol (*28*) and therefore is indiscernible from zero within the precision of the data. The predicted diffusion-limited time scale for monomer−aggregate association is $\sim 10^{-7}$ min using a Smoluchowski model in the absence of electrostatic repulsions, and using previously measured hydrodynamic radius for monomer ($R_h \approx 2$ nm) and aggregate ($R_h \approx 30$ nm) (*13*). This assumes only a small fraction (<0.1%) of the aggregate surface is able to add monomer (*29*), approximating the geometric effect of monomers adding only at the termini of freely rotating aggregate chains. To reconcile this value of a diffusion-limited time scale for growth with those for growth in Figure 2 requires a stability ratio with a free energy barrier of *ca.* 14 k_BT (*27*, *30*). Relaxing the above assumptions and thereby assuming 100% of the aggregate surface is "reactive", and accounting for differences between hydrodynamic radii and contact radii for aggregates, yields an upper limit for the energy barrier of *ca.* 18 k_BT (see also Supporting Information). While this range of values constitute a large barrier, it is physically reasonable considering that the ionic strength is low and that the pH of the solution is approximately one unit below the pK_a values of all acidic side chains in aCgn. Therefore the net charge on each protein monomer can be as high as +20 (assuming all acidic and basic groups are protonated), and large electrostatic repulsions between monomers and aggregates are expected. It is not known how the charge is distributed in the aggregate, but the local charge density is presumably similar to that of the free monomer.

The LENP model (*4*) provides a relationship for $\tau_n^{(0)}$ in terms of the equilibrium constant (K_i) for forming an oligomer of $i = x$ or $x - 1$ monomers, the forward ($k_{a,x}$) and reverse ($k_{d,x}$) rate coefficients for association of R with R_{x-1} to form R_x, and the forward rate coefficient ($k_{r,x}$) for conformational rearrangement of R_x to nucleate the (irreversible) aggregate A_x (see also Figure 1). The expressions below (eq 8) for $\tau_n^{(0)}$ follow as limiting cases in which association is either much slower (eq 8a) or faster (eq 8b) than rearrangement of R_x to A_x.

$$\tau_n^{(0)} = [K_{x-1}C_{ref}^{x-1}k_{a,x}]^{-1} \quad (k_{r,x} \gg k_{d,x}) \qquad (8a)$$

$$\tau_n^{(0)} = [K_xC_{ref}^{x-1}k_{r,x}]^{-1} \quad (k_{r,x} \ll k_{d,x}) \qquad (8b)$$

Using eq 8 for $\tau_n^{(0)}$ in eq 7 gives the following relationships for the effective activation energies, and shows E_a has contributions from the thermodynamics of reversible prenucleus formation (K_x or K_{x-1}) as well as the kinetics of either rearrangement ($k_{r,x}$) or monomer−prenucleus association ($k_{a,x}$).

$$E_a = \Delta H_{x-1} + E_{a,assoc} \quad (k_{r,x} \gg k_{d,x}) \tag{9a}$$

$$E_a = \Delta H_x + E_{a,r} \quad (k_{r,x} \ll k_{d,x}) \tag{9b}$$

In eq 9, the enthalpy of reversibly forming R_x or R_{x-1} from monomers is denoted by ΔH_x or ΔH_{x-1}, respectively. The activation energy for $k_{a,x}$ is denoted as $E_{a,assoc}$, while that for $k_{r,x}$ is denoted $E_{a,r}$. The osmotic second virial coefficient (B_{22}) for purely monomer samples at this pH and 20 °C was found to be $14 \pm 6 \times 10^{-4}$ mol mL g^{-2} via light scattering (see Materials and Methods section and Supporting Information). Combined with the long Debye screening length for these solution conditions ($1/\kappa \approx 5$ nm), this result indicates that the dominant reversible interactions between aCgn molecules are repulsive electrostatics, consistent with the acidic pH of the solution conditions. It is therefore reasonable to expect that ΔH_x and ΔH_{x-1} are positive, and so the negative E_a value for nucleation is due either to simple association of R with R_{x-1} or to rearrangement of R_x to A_x. The latter situation is not unexpected if the rearrangement step is analogous to folding of globular proteins at high temperatures, where the free energy barrier can be primarily entropic (*31*). The former situation is difficult to reconcile if association does not occur simultaneously with a conformational change, as otherwise the activation barrier would presumably include a large enthalpic cost due to electrostatic repulsion.

In either case, the data in Figure 2 indicate a significant entropic barrier to the kinetics of nucleation that is not explained simply by the entropy loss of reversible oligomer formation. Together, the data and analysis provide clear experimental evidence to support the view that protein folding and aggregate nucleation may share fundamental similarities (*32*, *33*). If rates of folding and aggregate nucleation are governed by similar mechanisms of conformational searches, it is then anticipated that E_a for aggregate nucleation can change sign and become dominated by enthalpic barriers at lower temperatures than those examined here (*31*).

ACKNOWLEDGMENT

The authors thank A. M. Lenhoff and A. S. Robinson for helpful discussions and suggestions on the manuscript, and E. W. Kaler for use of the light scattering facilities.

SUPPORTING INFORMATION AVAILABLE

Additional details of static light scattering measurements and calculation of diffusion-limited rate coefficients. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Wang, W. (2005) Protein aggregation and its inhibition in biopharmaceutics, *Int. J. Pharm. 289*, 1−30.
2. Dobson, C. M. (2006) Protein aggregation and its consequences for human disease, *Protein Pept. Lett. 13*, 219−227.
3. Ferrone, F. (1999) Analysis of Protein Aggregation Kinetics, *Methods Enzymol. 309*, 256−274.
4. Andrews, J. M., and Roberts, C. J. (2007) A Lumry-Eyring Nucleated-Polymerization Model of Protein Aggregation Kinetics I. Aggregation with Pre-Equilibrated Unfolding, *J. Phys. Chem. B 111*, 7897−7913.
5. Chen, S. M., Ferrone, F. A., and Wetzel, R. (2002) Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation, *Proc. Natl. Acad. Sci. U.S.A. 99*, 11884−11889.
6. Hurshman, A. R., White, J. T., Powers, E. T., and Kelly, J. W. (2004) Transthyretin aggregation under partially denaturing conditions is a downhill polymerization, *Biochemistry 43*, 7365−81.
7. Ignatova, Z., and Gierasch, L. M. (2005) Aggregation of a slow-folding mutant of a beta-clam protein proceeds through a monomeric nucleus, *Biochemistry 44*, 7266−7274.
8. Pallitto, M. M., and Murphy, R. M. (2001) A mathematical model of the kinetics of beta-amyloid fibril growth from the denatured state, *Biophys. J. 81*, 1805−1822.
9. Roberts, C. J. (2003) Kinetics of irreversible protein aggregation: Analysis of extended Lumry-Eyring models and implications for predicting protein shelf life, *J. Phys. Chem. B 107*, 1194−1207.
10. Roberts, C. J. (2006) Non-Native Protein Aggregation: Pathways, Kinetics, and Shelf-Life Prediction, in *Misbehaving Proteins: Protein Misfolding, Aggregation, and Stability* (Murphy, R. M., and Tsai, A. M., Eds.) pp 17−46, Springer, New York.
11. Roberts, C. J., Darrington, R. T., and Whitley, M. B. (2003) Irreversible aggregation of recombinant bovine Granulocyte-Colony Stimulating Factor (bG-CSF) and implications for predicting protein shelf life, *J. Pharm. Sci. 92*, 1095−1111.
12. Andrews, J. M., and Roberts, C. J. (2007) Non-native aggregation of alpha-chymotrypsinogen occurs through nucleation and growth with competing nucleus sizes and negative activation energies, *Biochemistry 46*, 7558−7571.
13. Weiss, W. F., IV, Hodgdon, T. K., Kaler, E. W., Lenhoff, A. M., and Roberts, C. J. (2007) Nonnative protein polymers: structure, morphology, and relation to nucleation and growth, *Biophys. J. 93*, 4392−4403.
14. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation, *Semin. Cell Dev. Biol. 15*, 3−16.
15. Slow, E. J., Graham, R. K., and Hayden, M. R. (2006) To be or not to be toxic: aggregations in Huntington and Alzheimer disease, *Trends Genet. 22*, 408−411.
16. Bucciantini, M., Calloni, G., Chiti, F., Formigli, L., Nosi, D., Dobson, C. M., and Stefani, M. (2004) Prefibrillar amyloid protein aggregates share common features of cytotoxicity, *J. Biol. Chem. 279*, 31374−31382.
17. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J. S., Taddei, N., Ramponi, G., Dobson, C. M., and Stefani, M. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, *Nature 416*, 507−511.
18. Chi, E. Y., Krishnan, S., Randolph, T. W., and Carpenter, J. F. (2003) Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation, *Pharm. Res. 20*, 1325−1336.
19. Wang, K. Y., and Kurganov, B. I. (2003) Kinetics of heat- and acidification-induced aggregation of firefly luciferase, *Biophys. Chem. 106*, 97−109.
20. Oosawa, F., and Asakura, S. (1975) *Thermodynamics of the Polymerization of Proteins*, Academic Press, London.
21. Roberts, C. J. (2007) Non-native protein aggregation kinetics, *Biotechnol. Bioeng. 98*, 927−938.
22. Evans, K. C., Berger, E. P., Cho, C. G., Weisgraber, K. H., and Lansbury, P. T. (1995) Apolipoprotein-E Is A Kinetic But Not A Thermodynamic Inhibitor Of Amyloid Formation - Implications For The Pathogenesis And Treatment Of Alzheimer-Disease, *Proc. Natl. Acad. Sci. U.S.A. 92*, 763−767.
23. Devlin, G. L., Knowles, T. P. J., Squires, A., McCammon, M. G., Gras, S. L., Nilsson, M. R., Robinson, C. V., Dobson, C. M., and MacPhee, C. E. (2006) The component polypeptide chains of bovine insulin nucleate or inhibit aggregation of the parent protein in a conformation-dependent manner, *J. Mol. Biol. 360*, 497−509.
24. O'Nuallain, B., Williams, A. D., Westermark, P., and Wetzel, R. (2004) Seeding specificity in amyloid growth induced by heterologous fibrils, *J. Biol. Chem. 279*, 17490−17499.
25. Bhattacharyya, A. M., Thakur, A. K., and Wetzel, R. (2005) Polyglutamine aggregation nucleation: Thermodynamics of a highly unfavorable protein folding reaction, *Proc. Natl. Acad. Sci. U.S.A. 102*, 15400−15405.

26. Hiemenz, P. C., and Rajagopalan, R. (1997) *Principles of Colloid and Surface Chemistry*, 3rd ed., Marcel Dekker, Inc., New York.
27. Russel, W. B., Saville, D. A., and Scholwalter, W. R. (1989) *Colloidal Dispersions*, Cambridge University Press, New York.
28. Lemmon, E. W., McLinden, M. O., and Friend, D. G. (2005) Thermophysical Properties of Fluid Systems, in *NIST Chemistry Webbook, NIST Standard Reference Database Number 69* (Linstrom, P. J., and Mallard, W. G., Eds.), National Institute of Standards and Technology, Gaithersburg, MD.
29. Barzykin, A. V., and Shushin, A. I. (2001) Effect of anisotropic reactivity on the rate of diffusion-controlled reactions: Comparative analysis of the models of patches and hemispheres, *Biophys. J. 80*, 2062−2073.
30. Prieve, D. C., and Ruckenstein, E. (1980) Role Of Surface-Chemistry In Primary And Secondary Coagulation And Hetero-Coagulation, *J. Colloid Interface Sci. 73*, 539−555.
31. Oliveberg, M., Tan, Y. J., and Fersht, A. R. (1995) Negative Activation Enthalpies In The Kinetics Of Protein-Folding, *Proc. Natl. Acad. Sci. U.S.A. 92*, 8926−8929.
32. Colon, W., and Kelly, J. W. (1992) Partial Denaturation Of Transthyretin Is Sufficient For Amyloid Fibril Formation In vitro, *Biochemistry 31*, 8654−8660.
33. Ohnishi, S., and Takano, K. (2004) Amyloid fibrils from the viewpoint of protein folding, *Cell. Mol. Life Sci. 61*, 511−524.

BI7019244