# Aggregation of a Slow-Folding Mutant of a β-Clam Protein Proceeds through a Monomeric Nucleus †

**ARTICLE** *in* BIOCHEMISTRY · JUNE 2005

Impact Factor: 3.02 · DOI: 10.1021/bi047404e · Source: PubMed



**2 AUTHORS**, INCLUDING:

Lila M Gierasch
University of Massachusetts Amherst
**234** PUBLICATIONS **11,684** CITATIONS

# Aggregation of a Slow-Folding Mutant of a β-Clam Protein Proceeds through a Monomeric Nucleus†

Zoya Ignatova‡,§ and Lila M. Gierasch*,‡,||

*Departments of Biochemistry and Molecular Biology and of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003*

*Received December 9, 2004; Revised Manuscript Received March 21, 2005*

ABSTRACT: Mechanistic understanding of protein aggregation, leading either to structured amyloid fibrils or to amorphous inclusion body-like deposits, should facilitate the identification of potential therapeutic intervention strategies for the devastating amyloid-based diseases. Here we focus on the in vitro aggregation of a slow-folding mutant of the β-clam protein, cellular retinoic acid-binding protein I (P39A CRABP I), which forms inclusion bodies when expressed in *Escherichia coli*. Aggregation was monitored by observing the fluorescence of a fluorescein-based biarsenical dye (FlAsH) that ligates to a tetra-Cys motif, here incorporated into a flexible Ω-loop. The fluorescence signal of FlAsH on the tetra-Cys-containing P39A CRABP I is sensitive to whether this protein is native or unfolded, and was used in combination with other techniques to follow aggregate formation. The aggregation time course is compatible with a nucleation-dependent polymerization model, and detailed kinetic analysis showed that the energetically unfavorable nucleus is monomeric. A similar conclusion was reached previously for poly(Gln) species [Chen, S., Ferrone, F. A., and Wetzel, R. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 11884−11889] and points to an unfavorable equilibrium between the misfolded intermediate and the bulk pool of monomers as causative in aggregation. The P39A mutation, which removes a helix-stop signal, may slow closure of the β-barrel in P39A CRABP I relative to the wild type, leaving it vulnerable to aggregation. Wide-angle X-ray scattering showed that the amorphous aggregates formed by the aggregation-prone intermediates of P39A CRABP I contain predominantly β-strands structured in a lamellar fashion with 10.03 Å spacing between adjacent β-sheets.

Two different physiological consequences can be attributed to protein misfolding and resulting aggregation. The first is a loss of protein function, which is often accompanied by mislocalization or degradation of the defective protein. Second, the toxic properties of the aggregates cause certain metabolic alterations and generate cellular abnormalities (*1*). Protein aggregates are commonly viewed as oligomeric complexes of structured, partially folded, or unfolded kinetic intermediates driven to associate by specific intermolecular interactions (*2*−*4*). In the biotechnology and pharmaceutical industries, protein aggregation causes only economic and technical problems, whereas the effects for patients suffering from a variety of diseases linked to protein aggregation can be devastating. Several human diseases have been identified to be associated with protein deposition in the form of structured aggregates (amyloid fibrils) (e.g., Alzheimer's, Parkinson's, Creutzfeldt-Jakob, and Huntington's diseases) (*4*, *5*), and considerable research has been focused on this type of aggregation due to its high pathogenic impact on humans. Recent studies have pointed out a link between amorphous aggregates, similar to inclusion bodies (widely considered as a hallmark of overexpressing bacterial systems), and pathogenesis of some neurodegenerative disorders (e.g., inclusion body myositis, light-chain deposition disease, and cataracts) (*3*, *6*, *7*). Inclusion bodies sediment at low speeds after cell breakage and appear as amorphous dense refractile particles under the electron microscope. Several independent studies provide clear evidence that the amorphous aggregates result from highly specific intermolecular interactions between folding intermediates with a nativelike secondary structure (*3*, *8*−*11*). Aggregation can be triggered by higher-order domain swapping, resulting in the formation of an interdomain interface that is identical to an intradomain interface in the native state (*12*). Inspection of the amorphous aggregates by Fourier transform infrared or Raman spectroscopy shows the presence of significant nativelike secondary structure with an increase in non-native β-sheet content (*3*, *10*, *11*). β-Sheet structure is one of the structural characteristics of stable amyloid fibrils, raising the possibility of a common aggregation mechanism for structured and amorphous protein deposits. The detailed kinetic mechanism of amorphous aggregate formation is, however, not well understood.

Several lines of evidence derived mainly from results obtained in in vitro studies with amyloidogenic proteins suggest that the kinetics of aggregation are generally consistent with nucleation−propagation polymerization with

† This research was supported in part by NIH Grant GM27616 to L.M.G. and in part by DFG Grant Ka 505/9-1 to Z.I.

* To whom correspondence should be addressed. E-mail: gierasch@biochem.umass.edu. Telephone: (413) 545-6094. Fax: (413) 545-1289.

‡ Department of Biochemistry and Molecular Biology.

§ Present address: Department of Biotransformation und -sensorik, TU Hamburg-Harburg, Denickestr. 15, 21071 Hamburg, Germany.

|| Department of Chemistry.

10.1021/bi047404e CCC: $30.25 
Published on Web 04/23/2005

a lag phase, the duration of which depends on the concentration of the protein and can be significantly shortened by seeding with preformed aggregates (*13*−*16*). Nucleated polymerization growth comprises formation of an energetically unfavorable nucleus, followed by an energetically downhill process of elongation of this nucleus via sequential addition of monomers (*13*, *14*, *17*). The initial kinetics of nucleation polymerization aggregation are concentration-dependent, and rates are dependent on time squared (*13*, *17*). Some studies with different amyloid aggregation systems have led to alternative models for explaining deviations from the established simple nucleation polymerization model. These models range from a downhill aggregation with no nucleus (*18*) through formation of micelles as a prerequisite to the nucleation step (*16*, *19*), a secondary nucleation event (*20*), and short-lived metastable assemblies (*21*).

In this paper, we have investigated the aggregation mechanism of a slow-folding, aggregation-prone mutant of cellular retinoic acid-binding protein I (CRABP I).[1] CRABP I (136 amino acids), a member of the $\beta$-clam protein family, is composed of a small helix−turn−helix subdomain and two orthogonal five-stranded $\beta$-sheets, surrounding a central ligand-binding cavity. In vitro experiments suggest that refolding passes through three well-defined kinetic phases: (1) an initial hydrophobic collapse accompanied by formation of considerable secondary structure ($\sim$300 $\mu$s) (J. A. Habink, U. Neu, and L. M. Gierasch, unpublished results), (2) the development of nativelike topology and formation of the ligand (retinoic acid) binding cavity (100−200 ms), and (3) establishment of the hydrogen bonding network (1 s) (*22*−*24*). Mutation of Pro39 to Ala in CRABP I retards both folding and unfolding in vitro (*25*), and the resulting mutant (P39A CRABP I) exhibited an increased propensity to aggregate and a tendency to form inclusion bodies in vivo. In a recent study, we have shown that introduction of a specific probe (a tetracysteine motif) into the internal flexible nonconserved $\Omega$-loop of P39A CRABP I does not perturb its structure; the structure of the mutant P39A tetra-Cys CRABP I is indistinguishable from that of P39A CRABP I (*26*). The recently developed fluorescent dye, FlAsH (*27*), ligates to the incorporated tetracysteine motif and emits a bright fluorescent signal. The quantum yield depends on the conformational state of the protein and reports sensitively on structural rearrangements. In this study, we have exploited these features and used the fluorescence of FlAsH-labeled P39A tetra-Cys CRABP I to explore its in vitro aggregation. Fluorescence measurements confirmed that the P39A tetra-Cys CRABP I follows a nucleation polymerization model reminiscent of that observed in vivo. The energetically unfavorable nucleus can be bypassed by seeding with preformed aggregates. Aggregation progress monitored by the soluble monomer concentration revealed that the number of monomeric units comprising the critical nucleus is equal to one, suggesting that the energetically unfavorable event is unfolding within one molecule. It will be of great interest to relate the nature of this monomeric nucleus to intermediates in the folding reaction of CRABP I and variants that are being characterized. Furthermore, we show that the amorphous aggregates, despite their morphological heterogeneity, have a defined microstructure. As previously reported (*3*, *10*), they are rich in $\beta$-structure locally, and in addition, wide-angle X-ray scattering of the P39A tetra-Cys CRABP I deposits showed a 10.03 Å spacing, consistent with a lamellar $\beta$-sheet organization of the aggregates. Our results have important implications for understanding the mechanism of aggregation as a generic feature of the polypeptide chain and for developing rational strategies for therapeutic treatment of disorders associated with protein aggregation.

## EXPERIMENTAL PROCEDURES

*Protein Expression and Purification.* P39A tetra-Cys CRABP I with the FlAsH-binding motif, constructed as described previously (*26*), was expressed in the *Escherichia coli* BL21(DE3) cells. Expression cultures (1 L of LB medium containing 100 $\mu$g/mL ampicillin) inoculated from overnight cultures at 1:25 dilution were grown at 37 °C to an $A_{600}$ an $\approx$1.0. Protein expression was induced with 0.4 mM IPTG for 4 h, and the cells were harvested by centrifugation at 3700 rpm for 30 min. Cell pellets from 1 L of medium were resuspended in 20 mL of SL buffer [20 mM Tris-HCl and 500 mM NaCl (pH 8.0)] and frozen at −80 °C. After cell lysis and centrifugation, the P39A tetra-Cys CRABP I was purified from the clear supernatant on a Ni−NTA column (Qiagen, Valencia, CA) as described by Clark et al. (*24*). After elution from the column, the sample was dialyzed successively from 10 to 1 mM Tris-HCl containing 2 mM $\beta$-mercaptoethanol. A typical purification from the soluble fraction yields 7−10 mg of protein from 1 L of culture; 150−200 $\mu$M protein solutions were usually stored at 4 °C and were used within 2 weeks. Concentrations were determined spectrophotometrically using an $\epsilon_{280}$ of 21 750 $M^{-1}$ $cm^{-1}$ (a corrected value for wild-type CRABP I with four additional cysteines). Although P39A tetra-Cys CRABP I partitions equally between the soluble and insoluble cell fractions, we used for its isolation only the soluble form to ensure the initial folding homogeneity of the protein population used in the in vitro aggregation experiments.

*Aggregation Kinetics.* Different concentrations of P39A tetra-Cys CRABP I solutions were labeled with FlAsH-$EDT_2$ in an equimolar ratio in 10 mM HEPES (pH 7.8) [containing 1 mM tris(carboxyethyl)phosphine and 1 mM 2-mercaptoethanol] for 60−120 min at 37 °C. The labeled protein was destabilized by dissolving in 1.7 M urea [prepared from stock solutions from 8 to 8.5 M urea and 10 mM HEPES (pH 7.8)], and the aggregation kinetics were assessed by two different methods (FlAsH is marketed now by Invitrogen under the Lumio trade name).

*(1) FlAsH Fluorescence.* Samples (usually a 200−250 $\mu$L total volume) were incubated at 37 °C with no stirring, and periodically, the fluorescence at 530 nm (bandwidth of 2 nm) was measured upon excitation at 500 nm (bandwidth of 2 nm) (Photon Technology International QM-1 fluorometer). The temperature of the cuvette holder was maintained at 37 °C with a water bath. At the time of measurement, the reaction mixtures were gently vortexed prior to withdrawal of the aliquot for measurement (120 $\mu$L) to ensure an equal distribution of the aggregates over the entire volume of the

[1] Abbreviations: CD, circular dichroism; CRABP I, cellular retinoic acid-binding protein I; EDT, ethanedithiol; FlAsH, 4′,5′-bis(1,3,2-dithioarsolan-2-yl)fluorescein; LB, Luria broth; WAXS, wide-angle X-ray scattering.

sample. In parallel, the fluorescence of controls without protein was recorded for all experiments and subtracted from that of the sample.

*(2) Spectrophotometric Assay.* Aliquots of FlAsH-labeled P39A tetra-Cys CRABP I at different concentrations were incubated at 37 °C with no stirring. At the indicated times, the tubes were gently vortexed as described in the previous paragraph, and then 75 $\mu$L aliquots were withdrawn. The aliquots were centrifuged at 14 000 rpm for 20 min at 4 °C, and the clear supernatant was diluted into water and subjected to spectroscopic measurements. The absorbance was converted into concentration using $\epsilon_{280}$ (see above).

*Determination of Nucleus Size.* We utilized the mathematical analysis of nucleation-controlled aggregation kinetics introduced by Ferrone (*17*), which was applied by Chen et al. (*13*) in the description of aggregation of polyglutamine peptides. Kinetic parameters were obtained at the early times of aggregation reaction at different initial protein concentrations. The total concentration of the monomers that have polymerized into aggregates ($\Delta$) is

$$\Delta = {}^1/_2 JJ^* c^* t^2 \quad (1)$$

where $J$ is the elongation rate of aggregates as a function of time $t$, $J^*$ is the rate of elongation of the nucleus, and $c^*$ is the concentration of the nucleus. The elongation rate is proportional to the bulk concentration ($c$) of the monomers:

$$J = k_+ c \quad (2)$$

where $k_+$ is the elongation rate constant. The concentration of the nucleus is given by the following equation:

$$c^* = K_{n^*} c^{n^*} \quad (3)$$

where $n^*$ is the number of monomers in the nucleus and $K_{n^*}$ is the equilibrium constant for the monomer−nucleus equilibrium. Combining eqs 1−3 results in

$$\Delta = {}^1/_2 k_+{}^2 K_{n^*} c^{(n^*+2)} t^2 \quad (4)$$

Equation 4 assumes that the elongation rate of the nucleus ($k_+^*$) is equal to the elongation rate of the aggregates ($k_+$) (*17*). A plot of the changes of the soluble monomer or the rate of aggregate formation (FlAsH fluorescence) versus $t^2$ will yield a slope equal to ${}^1/_2 k_+{}^2 K_{n^*} c^{(n^*+2)}$. The logarithm of the slope is

$$\ln(\text{slope}) = \ln({}^1/_2 k_+{}^2 K_{n^*}) + (n^* + 2)\ln(c) \quad (5)$$

A plot of ln(slope) versus ln($c$) yields a slope of $n^* + 2$ and thus can be used to estimate the number of monomers in the critical nucleus.

*Seeding Experiments.* FlAsH-labeled P39A tetra-Cys CRABP I (25 $\mu$M) was destabilized by 1.7 M urea as described above and incubated overnight at 37 °C without stirring to allow complete aggregation. The aggregates were washed with the same volume of water and 10 mM Tris-HCl (pH 8.0) and sonicated for 2 min to ensure that a homogeneous mixture was being used. Different volumes of seed were added to the aggregation mixture representing 1, 5, and 10% of the initial concentration of labeled P39A tetra-Cys CRABP I. Aggregation was monitored either by FlAsH fluorescence or by absorbance at 280 nm and converted into soluble monomer.

*Fluorescence Microscopy Experiments*. Two microliters of labeled inclusion body suspension isolated either from cells [see (1) below] or grown in the test tube [see (2) below] was immobilized in 1% agarose in LB and imaged using a Nikon Eclipse E600 microscope, with excitation at 485 nm and a 510 nm emission cutoff filter. (1) P39A tetra-Cys CRABP I-expressing cells were labeled with FlAsH-$EDT_2$ as previously described (*26*). Four hours after induction, cells were centrifuged at 3500 rpm for 20 min and used to isolate intact inclusion bodies with a slight modification of the procedure described by Oberg et al. (*10*). Cells from 500 mL of LB were resuspended in 15 mL of 20 mM Tris-HCl (pH 8.0) and frozen. Freeze−thaw cycles were repeated three times. PMSF (1 mM) and lysozyme (0.5 mg) were added to the thawed solution, and the solution was incubated on ice for 15 min and then sonicated at 0 °C for 10 min (50% duty cycle). After centrifugation of the lysed cells at 10 000 rpm for 30 min, the pellet was resuspended in an additional 15 mL of SL buffer, sonicated, and centrifuged as before. The pellet was washed in 25 mL of SL buffer containing 0.1% Triton X-100. After centrifugation (10 000 rpm for 30 min) the freeze−thaw, lysis, centrifugation, and washing steps were repeated until the lysis was complete (the yellow film on the top disappeared). The pellet, representing a fraction of pure inclusion bodies, was resuspended in 1.5 mL of SL buffer containing 0.1% Triton X-100. [2] FlAsH-labeled P39A tetra-Cys CRABP I (30 $\mu$M) was destabilized by 1.7 M urea and incubated overnight at 37 °C without stirring to allow complete aggregation.

*Wide-Angle X-ray Scattering (WAXS).* The same samples used for microscopy experiments were used in the WAXS experiments. Suspensions of inclusion bodies or aggregates were desalted by dialysis overnight at 4 °C in $H_2O$ and were dried on Teflon plates with cylindrical profiles 3 mm in diameter and approximately 1 mm in height. The samples were irradiated with the X-ray beam perpendicular to the sedimentation axis.

## RESULTS

*Aggregation Kinetics.* Our previous experiments with in vivo aggregation of P39A tetra-Cys CRABP I showed that the aggregation is characterized by a lag phase that can be extended by a reduction of the biosynthesis rate (*26*). This is consistent with the widely established nucleation−polymerization model of aggregation. To test the hypothesis that P39A tetra-Cys CRABP I aggregation follows this model in vitro, we performed several experiments in which the kinetics of P39A tetra-Cys CRABP I aggregation were examined by two different biophysical methods: the decrease in the soluble monomer concentration, followed by the absorbance at 280 nm, and the concomitant increase in the fluorescence of the FlAsH-labeled protein.

Introduction of the tetra-Cys motif in the flexible Ω-loop in CRABP I and mutation of Pro39 to Ala causes little perturbation of the equilibrium stability of the native counterpart CRABP I (*26*). Despite the similar equilibrium stability, the replacement of the conserved helix-terminating Pro39 retards both unfolding and folding (*25*); additionally, the Pro39 to Ala mutant protein forms inclusion bodies upon

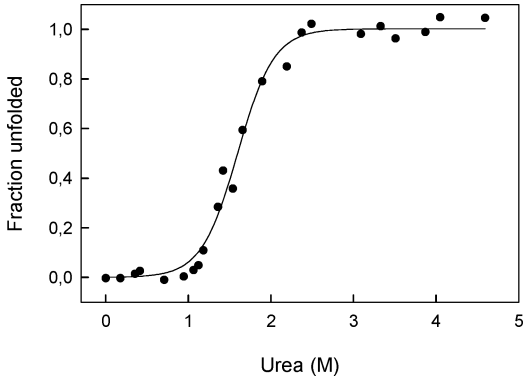


FIGURE 1: Urea-induced unfolding of labeled P39A tetra-Cys CRABP at 37 °C followed by the intrinsic Trp fluorescence (excitation at 280 nm; emission at 355 nm). The $C_m$ value is calculated to be 1.70 ± 0.05 M urea.

expression in *E. coli* under conditions where the parent protein without the Pro39 to Ala mutation does not. We suspect that a stabilized folding intermediate with an extended non-native helix is aggregation-prone in vivo (*26*). The P39A tetra-Cys CRABP I protein once properly folded does not aggregate spontaneously when stored at 4 °C, even at high concentrations, suggesting that partial unfolding is required to initiate aggregation. Titration of P39A tetra-Cys CRABP I with urea at 37 °C monitored by Trp fluorescence resulted in an apparent denaturation midpoint of 1.70 ± 0.05 M urea (Figure 1). At the transition midpoint, P39A tetra-Cys CRABP I aggregates faster than under other conditions, as has been observed for other proteins [e.g., apomyoglobin (*28*)], enabling us to study its aggregation kinetics in the laboratory at reasonable time scales; hence, all of the in vitro experiments reported here were carried out under these conditions (1.7 M urea). At this intermediate concentration of urea, the kinetics of unfolding are more rapid, and the equilibrium between native, partially unfolded, and unfolded states will be shifted away from native. While these conditions will make observation of aggregation and study of concentration more convenient, it is important to recognize that there may be many species present and interconverting under these denaturant conditions. Additional exploration of the dependence of these equilibria on solution conditions is beyond the scope of this study, but will be pursued in an effort to more rigorously characterize the aggregation-prone species.

The unfolded and folded states of the aggregation-prone mutant, P39A tetra-Cys CRABP I, display different quantum yields of FlAsH dye fluorescence, with the unfolded state being hyperfluorescent (*26*). FlAsH fluorescence was used to follow the aggregation in vitro, based on the assumption that hyperfluorescence would be a property of all non-native states including aggregates. Native P39A tetra-Cys CRABP I was prelabeled before the induction of aggregation to ensure a homogeneously labeled starting sample. The curves obtained by fluorescence measurement of the FlAsH-probed P39A tetra-Cys CRABP I are sigmoidal in shape. At early times, there is an apparent lag phase, followed by an exponential increase in fluorescence, and finishing with a plateau (Figure 2A), resembling the curves of the in vivo aggregation behavior of P39A tetra-Cys CRABP I (*26*). The aggregation is clearly concentration dependent. A shorter lag phase and a higher rate in the exponential phase were observed as a consequence of the increase in the initial protein concentration.

The time course of the soluble monomer concentration (based on the absorbance at 280 nm) is roughly the complement of the FlAsH fluorescence curve (Figure 2B), supporting our use of FlAsH to report on concentration of aggregates. The monomer concentration displays a lag phase with the same duration as the lag phase observed in the FlAsH fluorescence experiments. In the next phase, the concentration of the soluble monomer decreases with the same rate as the increase in the amount of aggregate monitored by FlAsH fluorescence (Figure 2A). Overlaying the curves from both experiments demonstrates that the increase in the FlAsH fluorescence occurs at the same time scale as the decrease of the soluble monomer measured by the $A_{280}$ (data not shown). Monitoring of aggregation by fluorescence and of monomer concentration by absorption was carried out with no stirring. At the time of the withdrawal of the sample, the reaction mixtures were gently mixed to ensure an equal distribution of the aggregates throughout the entire volume. This is essential, particularly at the later time points when the size of the aggregates increases sufficiently

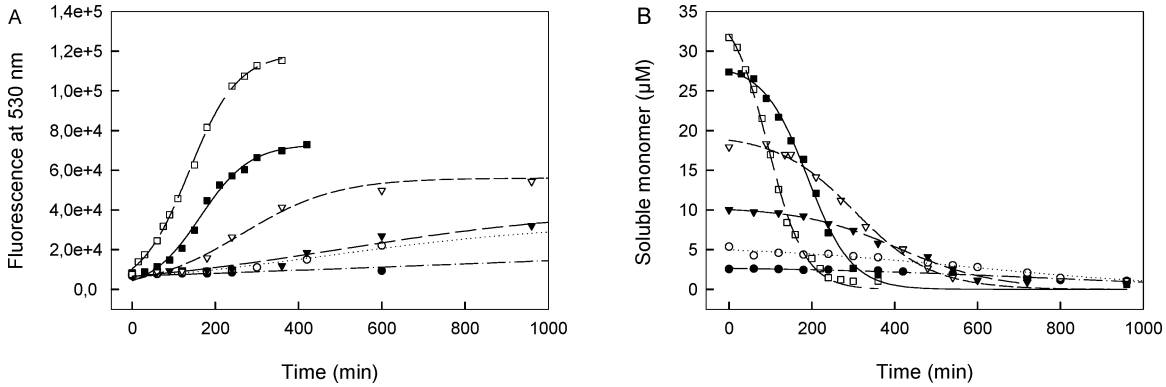


FIGURE 2: In vitro aggregation kinetics of labeled P39A tetra-Cys CRABP at 37 °C followed by the FlAsH fluorescence of the suspension (A) and the decrease of the amount of the soluble monomer (B). The absorption at 280 nm was used to quantify the concentration of the soluble monomer in the clarified supernatant from the aggregation suspension. The initial concentrations of P39A tetra-Cys CRABP I in these experiments were 2.5 (●), 5 (○), 10 (▼), 15 (▽), 25 (■), and 30 $\mu$M (□). The native protein was prelabeled in bulk solution with an equimolar amount of FlAsH for 2 h, diluted into 10 mM HEPES (pH 7.8) containing 1.7 M urea to initiate the aggregation. From each point, a blank (lacking only protein) was subtracted. The $\epsilon_{280}$ value of P39A tetra-Cys CRABP I of 21 750 $M^{-1}$ $cm^{-1}$ was used to convert the absorbance into protein concentration. Lines represent sigmoidal curve fits.

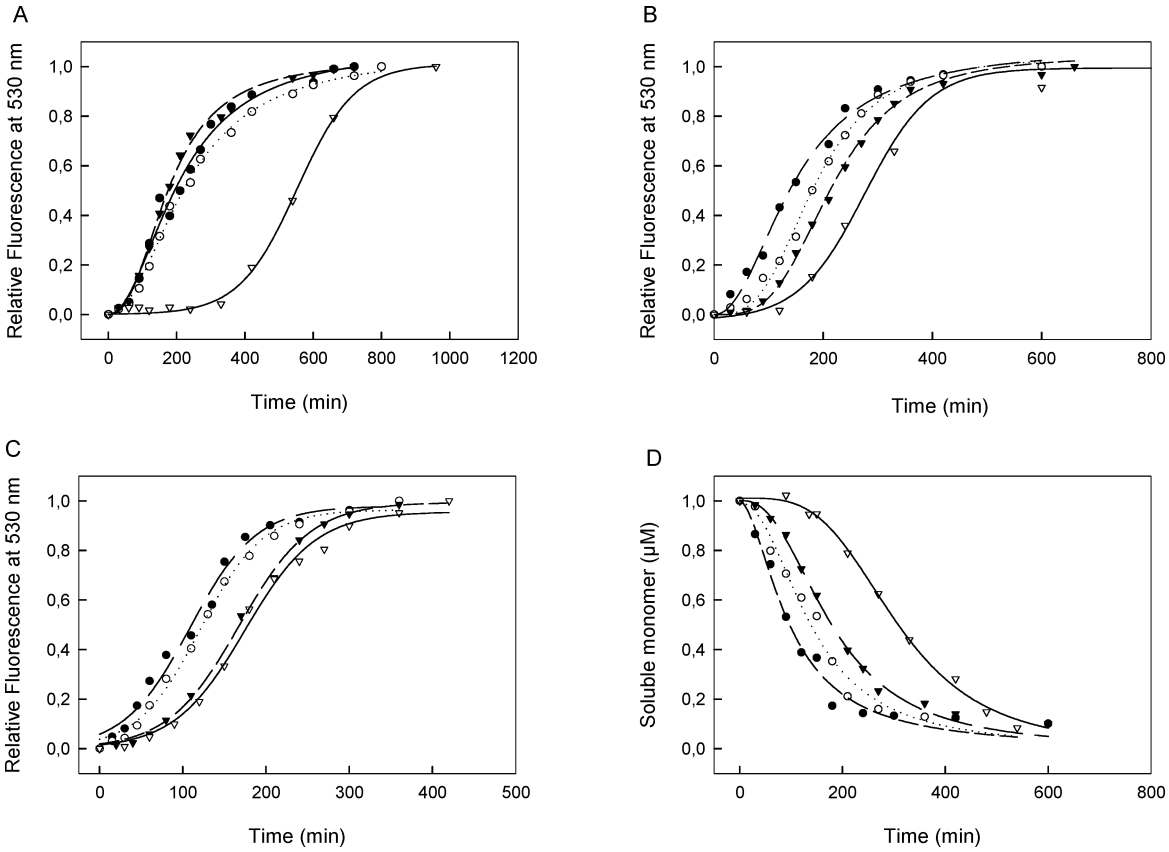


FIGURE 3: Addition of preformed seed shortens the lag phase. Aggregation of 10 (A), 15 (B), and 25 $\mu$M (C) FlAsH labeled P39A tetra-Cys CRABP I with and without preformed aggregates (seed) was monitored by FlAsH fluorescence. Fluorescence values are related to the highest and lowest values in each experiment and converted into relative fluorescence. (D) Aggregation of 15 $\mu$M FlAsH labeled P39A tetra-Cys CRABP I followed by absorbance at 280 nm. Reaction mixtures contained no seed (▽) and 1 (▼), 5 (○), and 10% seed (●) from the initial protein concentration.

such that they precipitate out of solution.

*Seeding Experiments.* Aggregation reaction mixtures were incubated over a prolonged period of time at 37 °C to allow the process to reach completion, and then the product was added to fresh, destabilized FlAsH-labeled P39A tetra-Cys CRABP I to test the effect of seeding on the aggregation reaction. The pre-aggregates were washed and sonicated to increase the number of sites available for further polymerization. Addition of different amounts of preformed aggregates at time 0 shortened significantly the length of the apparent lag phase for all concentrations tested (Figure 3). The rate of growth of the aggregates in the exponential phase appears to be similar to that of the unseeded reaction regardless of the seed concentration. At the limit of high concentration of seed, reaching ~10% of the initial concentration of soluble P39A tetra-Cys CRABP I, the apparent lag phase was entirely eliminated (Figure 3A−C). In the seeding experiments followed by FlAsH fluorescence, the absolute amount of fluorescence at time zero and at the end point increased compared to the unseeded experiment. We attribute this observation to the additive effect of the fluorescence of the aggregates in the seed mixtures, since the aggregates and all unfolded species are hyperfluorescent relative to the native counterpart. We present the data as relative fluorescence to exclude the contribution of the prelabeled seed mixture. Although the seeding curves at 15 $\mu$M when measured by fluorescence or by spectroscopy (Figure 3B,D) were superimposable, a slight shift was seen in the curve without seed, which could be due to the additional centrifugation of the samples subjected to spectroscopic measurements. The effect of the added preformed aggregates in the form of seeds is highest at the lowest concentration that was tested (10 $\mu$M) when the reaction is very slow without a seed (Figure 3A). Addition of a seed shortened the lag phase significantly, whereas at a higher protein concentration (25 $\mu$M), when the lag phase was already shorter, the impact of seed addition was less significant (Figure 3C).

*Nucleus Size.* The detection and characterization of the aggregation nucleus are very challenging. Its transient existence and the speed with which it converts either back to the bulk phase monomer or into elongating aggregates mean that its nature cannot be determined directly and can be inferred from kinetic studies. The kinetics of aggregation monitored either by FlAsH fluorescence or by absorbance at 280 nm and the seeding experiments point out that P39A tetra-Cys CRABP I displays the features of typical nucleated growth polymerization behavior: There is an apparent lag phase in aggregation that can be reduced either by an increased concentration of the monomer or by seeding. The initial phase of aggregation kinetics approximately follows a cosine or square function of time (*17*) and can thus be used to yield a quantitative measure of the nucleus size.

In the time courses of monomer concentration during aggregation (Figure 3D), the slopes account for the rate of elongation of the nuclei and of the aggregates. [In the initial phases, aggregation likely proceeds through small oligomers that are still soluble. Although our experiments did not

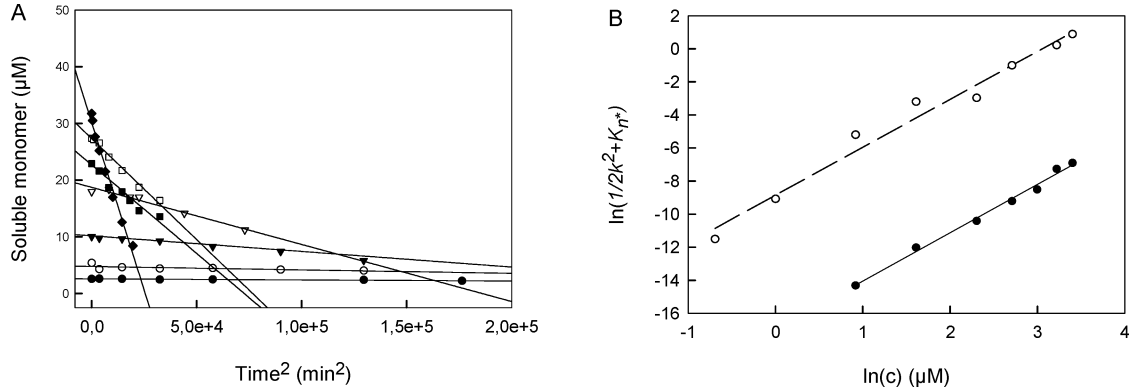


FIGURE 4: Analysis of the nucleated aggregation kinetics of P39A tetra-Cys CRABP I at 37 °C. (A) The changes in the monomer concentrations from the curves shown in Figure 2B are presented as dependent on $t^2$. The initial concentration of FlAsH-labeled P39A tetra-Cys CRABP I was 2.5 (●), 5 (○), 10 (▼), 15 (▽), 20 (■), 25 (□), and 30 $\mu$M (◆). Linear regression of the curves from panel A yielded linear fits with slopes equal to $^1/_2k_+^2K_{n^*}$, the logarithm of which is plotted vs the logarithm of starting P39A tetra-Cys CRABP I concentration in panel B (filled symbols, solid line). The aggregation kinetics followed by the FlAsH fluorescence were fitted to the nucleation kinetics equation (eq 1) in the same way as performed for the soluble monomer data. The logarithmic dependence on the linear fits of the $t^2$ function on the initial monomer concentration is given in panel B (empty symbols, dashed line).

distinguish among these soluble species, size-exclusion chromatography showed that the monomer concentration was changing with the same kinetics as our spectroscopic measurements (data not shown), arguing that the contribution from soluble oligomers was constant and small under the conditions used and validating the mathematical analysis.] For mathematical analysis of the aggregation rates, the changes in the monomer concentration were plotted as a function of the square of time ($t^2$) (Figure 4A) (*13*, *17*). The slopes of the linear fits of these data provide the aggregation rate at the corresponding concentration (see Experimental Procedures for a detailed mathematical description). The correlation coefficients of all linear fits for different concentrations were above 0.973. A plot of the logarithm of the aggregation rate (see Experimental Procedures) versus the logarithm of concentration (Figure 4B) was linear with a regression coefficient of 0.9927 and a slope of 2.9225, which is equal to $n^* + 2$ and is therefore consistent with a monomeric nucleus.

Although the final value of the FlAsH fluorescence was observed to increase with the increase of the initial concentration of P39A tetra-Cys CRABP I, the fluorescent yields determined as arbitrary fluorescent units per micromolar initial protein concentration show the opposite tendency (Table 1). The origin of this effect is not clear, but it is possible that it may arise because of increased intermolecular quenching effects with increased monomer concentration. The FlAsH fluorescence data fit well the nucleated-growth polymerization model, and the $t^2$ function was used to approximate the initial phase of aggregation. Similar to the curve fitting of the soluble monomer data, the kinetic parameters were obtained from the early points of the aggregation monitored by the FlAsH-fluorescence. The slope of the logarithmic plot of these kinetic parameters versus concentration yielded a value of 2.8994 ($n^* + 2$) (Figure 4B). This value rounded to the nearest integer corresponds again to a monomeric nucleus.

The observation of a monomeric nucleus is not inconsistent with the observation that preformed aggregates (seed) accelerate the aggregation process. From nucleated polymerization theory, eq 3, describing the nucleation kinetics (*17*), points out that even in the case of monomeric nucleus the concentration of the nuclei ($c^*$) will depend on the bulk monomer concentration ($c$). In this case, formation of nuclei depends on an energetically unfavorable pre-equilibrium reaction that turn leads to an aggregation-prone, misfolded state within the bulk pool of monomers.

Table 1: Fluorescence Yield of FlAsH-Labeled P39A Tetra-Cys CRABP I Aggregates

| concentration[a] ($\mu$M) | fluorescence yield[b] (arbitrary fluorescence units/ $\mu$M protein) |
|---|---|
| 0.5 | 16354 |
| 1 | 21303 |
| 2.5 | 9680 |
| 5 | 6644 |
| 10 | 4200 |
| 15 | 3905 |
| 25 | 2919 |
| 30 | 3844 |

[a] Concentration of the FlAsH-labeled P39A tetra-Cys CRABP I monomer used in the aggregation experiments (Figure 2A). [b] The fluorescence yield was calculated from the final FlAsH intensity in the aggregation kinetics (Figure 2A) at a corresponding initial concentration of the monomer.

*Aggregate Morphology and Structure.* We used fluorescence microscopy to characterize the morphology of aggregates. The aggregates isolated from cells and those grown in the test tube appear to have the same microscopic morphology (Figure 5). Under the microscope, no obvious structure in the aggregates is distinguishable. The fine structure of the aggregates was inspected by wide-angle X-ray scattering spectroscopy. The aggregates were resuspended in water and sedimented as thin films. Diffraction patterns were taken with an X-ray beam perpendicular to the axis of sedimentation so that any structure organized perpendicular to the plane will contribute. The X-ray diffraction patterns exhibit one outside reflection at 4.75 Å, consistent with the spacing of adjacent $\beta$-strands (Figure 6A). This reflection was expected due to the predominantly $\beta$-sheet organization of native P39A tetra-Cys CRABP I, and the likely nature of partially folded aggregates. The second and higher-order reflections are not visible on the diffraction diagrams (Figure 6). The scattering is not axial as expected for a fibril with long-range order, and the diffuse scattered

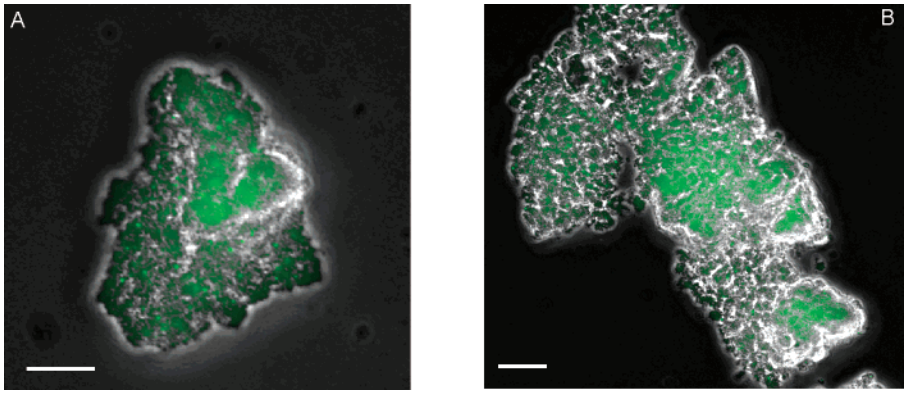


FIGURE 5: Fluorescence images of aggregates grown in vitro from prelabeled P39A tetra-Cys CRABP I at 37 °C (A) and inclusion bodies isolated from labeled cells expressing this protein (B). The intact inclusion bodies from the cells were isolated by using the procedure described by Oberg et al. (*10*).

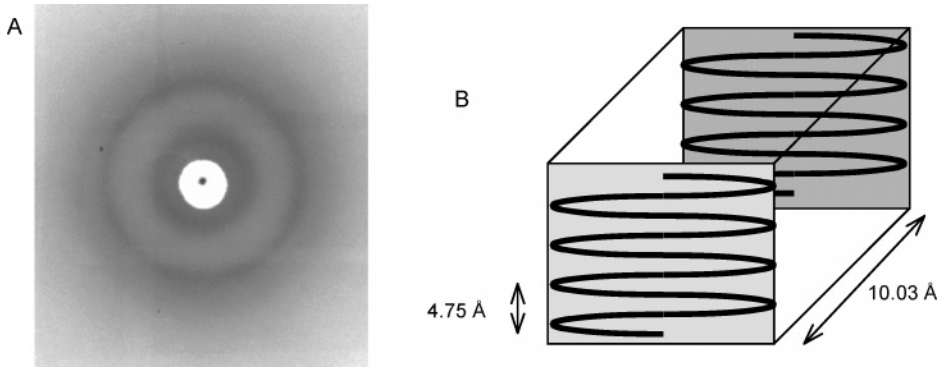


FIGURE 6: Structure of P39A tetra-Cys CRABP I aggregates. (A) Wide-angle X-ray diffraction pattern of aggregates grown in vitro taken perpendicular to the sedimentation axis. The dominant reflections are at 4.75 Å (the outside dense circular reflection) and 10.03 Å (the inside dense reflection). (B) Schematic lamellar model of the aggregate assembly. The 4.75 Å spacing corresponds to the spacing of the adjacent $\beta$-strands, and the perpendicular diffraction at 10.03 Å accounts for the intersheet spacing.

circle at 4.75 Å is indicative of many more locally structured regions (Figure 6A). Because of the geometry of the beam with respect to the sample, we attribute the inside dense reflection at 10.03 Å in the diffraction patterns to a distance between $\beta$-sheets in the *Z*-plane, pointing out a lamellar $\beta$-sheet organization of the P39A tetra-Cys CRABP I amorphous aggregates. This inside reflection is also quite diffuse, indicating that the aggregates contain a mixture of lamellar-like structure regions and largely unstructured regions.

## DISCUSSION

The aim of this study was to characterize using in vitro experiments the mechanism of aggregation of the aggregation-prone mutant P39A tetra-Cys CRABP I, which was observed to form inclusion bodies in vivo. The formation of these inclusion bodies by P39A tetra-Cys CRABP I in vivo showed nucleated polymerization behavior (*26*). The model for this type of aggregation involves an energetically unfavorable nucleus whose limiting rate of buildup and replenishment thus manifests itself in an apparent lag phase, after which sigmoidal aggregation curves ensue. Indeed, these data show that P39A tetra-Cys CRABP I aggregates in vitro at 37 °C via an apparent lag phase, which then is followed by rapid aggregate formation that continues until the monomer is largely depleted in the solution (Figure 2). Several key observations support a nucleation−polymerization model for P39A tetra-Cys CRABP I aggregation in vitro: The rate of the aggregation reaction increased with higher initial concentrations of the monomer, and the apparent lag phase could be bypassed by addition of preformed aggregates or seeds. The effect of the acceleration of the aggregation rate upon addition of preformed aggregates is more pronounced at lower protein concentrations, which is the most conclusive evidence for a nucleated polymerization growth of the aggregates. At low initial concentrations, nucleus formation is more unfavorable than at high concentrations, and addition of seed aggregates helps to overcome the energy barrier of nucleus formation, allowing the reaction to proceed to the more energetically favorable process, growth of the polymer.

The experimental observations for in vivo and in vitro P39A tetra-Cys CRABP I aggregation suggest that this process proceeds through a critical nucleus. The nucleation kinetics analysis is based on a model comprising the formation of the nucleus through establishment of a pre-equilibrium with the bulk phase monomer, elongation of the nucleus and the aggregates, controlled by a second-order rate constant (*13*, *15*, *17*). The results from our kinetic analysis of the aggregation of P39A tetra-Cys CRABP I suggest a monomeric nucleus that is an energetically unfavorable state in an unfavorable equilibrium with the bulk pool of monomers. Although our data firmly support the formation of a nucleus, its size calculation is indirect. A direct monitoring of its formation is, by definition, experimentally very challenging. In terms of the nucleation polymerization theory, the nucleus is the least prevalent species whose size corresponds to the peak of the energy curve through which

phase transition occurs (*17*). This barrier can be considered to be the free energy of nucleus formation. Despite this unexpected size of the nucleus in the context of the nucleation theory, it is no longer an exception. Evidence of a monomeric nucleus has been presented for aggregation of polyglutamine peptides resembling the N-terminal expansion of Huntington (*13*, *29*).

Aggregates of FlAsH-labeled P39A tetra-Cys CRABP I grown in the test tube have a morphology remarkably similar to that of inclusion bodies isolated from overexpressing *E. coli* cells. Bacterial inclusions are often enriched in a single aggregated protein (*30*), consistent with our finding that the formation of in vitro aggregates resembles in vivo inclusion body formation. Although microscopically unstructured, the P39A tetra-Cys CRABP I aggregates show a specific diffraction pattern consistent with well-organized structure (Figure 6). Documented examples in the literature provide clear evidence for an enrichment of β-sheet structure in amorphous aggregates (*10*). Even proteins with predominantly α-helical structure, or a mixture of α and β in their native states, show an increase in β-sheet content in their inclusion bodies (*3*, *11*). It has been suggested that amorphous aggregates involve β-sheet formation in the immediate vicinity of the interaction between the associating molecules (*3*, *8*). Since native P39A tetra-Cys CRABP I itself has a predominantly β-sheet organization which would contribute to the first reflection at 4.75 Å, we cannot distinguish whether any other non-native β-sheet structures are formed. The second reflection at 10.03 Å, which we attribute to a lamellar three-dimensional structuring of the adjacent β-sheets, may arise from intermolecular interactions stabilizing the aggregate structures.

Similar specific reflections in the range of 4.6−4.9 and 8.8−10.6 Å with a more pronounced density in meridional and equatorial direction are a characteristic fingerprint of the diffraction patterns of ordered amyloid fiber structures (*31*). Although the reflections from aggregates of P39A tetra-Cys CRABP I display a more diffuse circular shape and are less defined in the meridional and equatorial directions, the observed diffraction picture indicates that amorphous aggregates and fibrils share a common molecular skeleton.

Our experiments provide insight into the P39A tetra-Cys CRABP I aggregation pathway in vitro, and the overall kinetics mirror the aggregation behavior observed in vivo (*26*). It is remarkable that the kinetics reported here for a protein undergoing an amorphous aggregation match the kinetics described for structured amyloid-like aggregation. Despite differences in the morphology of the microscopic aggregates, the mechanisms of aggregation for both groups share common features. It is therefore fundamentally important to understand how seemingly unrelated proteins can follow a common aggregation mechanism. In the specific case of the slow-folding, aggregation-prone variant of CRABP I, we also have the possibility of relating our detailed understanding of the in vitro folding of the wild-type protein to the nature of the monomeric pre-aggregate. On the basis of our previous work (*22*−*25*), we speculate that the P39A mutation, which removes a helix-stop signal, delays closure of the β-barrel in P39A CRABP I relative to that in the wild type, leaving it vulnerable to aggregation. We are extending these studies currently. Moreover, since the early aggregation species have been thought to be responsible for cytotoxicity and disease pathology in several instances (*1*, *32*, *33*), the description of different scenarios of protein aggregation, including the identification of transition and intermediate states in the aggregation pathway, will help the development of efficient therapeutic strategies for treatment of protein deposition diseases.

## ACKNOWLEDGMENT

We thank Nick Renzette and Steven Sandler for the help with fluorescence microscopy, Ted Atkins for assistance with the WAXS, and Joanna Swain and Steve Eyles for critical reading of the manuscript.

## REFERENCES

1. Stefani, M., and Dobson, C. M. (2003) Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution, *J. Mol. Med. 81*, 678−699.
2. Wetzel, R. (1996) For protein misassembly, it's the "I" decade, *Cell 86*, 699−702.
3. Fink, A. L. (1998) Protein aggregation: Folding aggregates, inclusion bodies and amyloid, *Folding Des. 3*, 9−23.
4. Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem. Sci. 24*, 329−332.
5. Ross, C. A., and Poirier, M. A. (2004) Protein aggregation and neurodegenerative disease, *Nat. Med. 10*, 10−17.
6. Brujin, L. I., Houseweart, M. K., Kato, S., Anderson, K. L., Anderson, S. D., Ohama, E., Reaume, A. G., Scott, R. W., and Cleveland, D. W. (1998) Aggregation and motor neuron toxicity of an ALS-linked SOD1 mutant independent from wild-type SOD1, *Science 281*, 18511−18554.
7. Kosinski-Collins, M. S., and King, J. (2003) *In vitro* unfolding, refolding and polymerization of human γD crystallin, a protein involved in cataract formation, *Protein Sci. 12*, 480−490.
8. Speed, M. A., Wang, D. J., and King, J. (1996) Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition, *Nat. Biotechnol. 14*, 1283−1287.
9. King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A. (1996) Thermolabile folding intermediates: Inclusion bodies precursors and chaperonin substrates, *FASEB J. 10*, 57−66.
10. Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) Native like secondary structure in interleukin-1β inclusion bodies attenuated total reflectance FTIR, *Biochemistry 33*, 2628−2634.
11. Przybycien, T. M., Dunn, J. P., Valx, P., and Georgiou, G. (1994) Secondary structure characterization of β-lactamase inclusion bodies, *Protein Eng. 7*, 131−136.
12. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1994) 3D domain swapping: A mechanism for oligomer assembly, *Protein Sci. 4*, 2455−2468.
13. Chen, S., 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.
14. Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: Mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins, *Annu. Rev. Biochem. 66*, 385−407.
15. Jarrett, J. T., and Lansbury, P. J. (1993) Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell 73*, 1050−1058.
16. Lomakin, A., Chung, D. S., Beneder, G. B., Kirschner, D., and Teplow, D. B. (1996) On the nucleation and growth of amyloid β-protein fibrils: Detection of nuclei and quantification of rate constants, *Proc. Natl. Acad. Sci. U.S.A. 93*, 1125−1129.
17. Ferrone, F. (1999) Analysis of protein aggregation kinetics, *Methods Enzymol. 309*, 256−274.
18. 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*, 7365−7381.
19. Lomakin, A., Teplow, D. B., Kirschner, D. A., and Benedek, G. B. (1997). Kinetic theory of fibrillogenesis of amyloid protein, *Proc. Natl. Acad. Sci. U.S.A. 94*, 7942−7947.

20. Padrick, S. B., and Miranker, A. D. (2002) Islet amyloid: Phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis, *Biochemistry 41*, 4694−4703.
21. Bitan, G., Kirkitandze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B., and Teplow, D. B. (2003) Amyloid $\beta$-protein (A$\beta$) assembly: A$\beta$40 and A$\beta$42 oligomerize through distinct pathways, *Proc. Natl. Acad. Sci. U.S.A. 100*, 330−335.
22. Clark, P. L., and Gierasch, L. M. (1996) Intrinsic tryptophans of CRABP I as probes of structure and folding, *Protein Sci. 5*, 1108−1117.
23. Clark, P. L., Liu, Z. P., Rizo, J., and Gierasch, L. M. (1997) Cavity formation before stable hydrogen bonding in the folding of a $\beta$-clam protein, *Nat. Struct. Biol. 4*, 883−886.
24. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1998) Probing the folding pathway of a $\beta$-clam protein with single-tryptophan constructs, *Folding Des. 3*, 401−412.
25. Eyles, S. J., and Gierasch, L. M. (2000) Multiple roles of prolyl residues in structure and folding, *J. Mol. Biol. 301*, 737−747.
26. Ignatova, Z., and Gierasch, L. M. (2004) Monitoring protein stability and aggregation *in vivo* by real-time fluorescent labeling, *Proc. Natl. Acad. Sci. U.S.A. 101*, 523−528.
27. Griffin, B. A., Adams, S. R., Jones, J., and Tsien, R. Y. (2000) Fluorescent labeling of recombinant proteins in living cells with FlAsH, *Methods Enzymol. 327*, 565−578.
28. De Young, L. R., Dill, K. A., and Fink, A. L. (1993) Aggregation and denaturation of apomyoglobin in aqueous urea solutions, *Biochemistry 32*, 3877−3886.
29. Thakur, A. K., and Wetzel, R. (2002) Mutational analysis of the structural organization of polyglutamine aggregates, *Proc. Natl. Acad. Sci. U.S.A. 99*, 17014−17019.
30. Rajan, R. S., Illing, M. E., Bence, N. F., and Kopito, R. R. (2001) Specificity in intracellular protein aggregation and inclusion body formation, *Proc. Natl. Acad. Sci. U.S.A. 98*, 13060−13065.
31. Serpell, L. C., Fraser, P. E., and Sunde, M. (1999) X-ray fiber diffraction of amyloid fibrils, *Methods Enzymol. 309*, 526−536.
32. Hardy, J., and Selkoe, D. J. (2002) The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics, *Science 297*, 353−356.
33. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J., and Selkoe, D. J. (2002) Naturally secreted oligomers of amyloid $\beta$ protein potently inhibit hippocampal long-term potentiation *in vivo*, *Nature 416*, 535−539.

BI047404E