Formation of Short-Lived Protein Aggregates Directly from the Coil in Two-State Folding†

Maria Silow,‡ Yee-Joo Tan,§ Alan R. Fersht,‖ and Mikael Oliveberg*,‡

Department of Biochemistry, Chemical Centre, Lund University, P.O. Box 124, S-221 00 Lund, Sweden, Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Republic of Singapore, and Cambridge Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, U.K.

Received May 3, 1999; Revised Manuscript Received August 4, 1999

ABSTRACT: Recent results on the 102 residue protein U1A show that protein aggregation is not always slow and irreversible but may take place transiently in refolding studies on a millisecond time scale. In this study we observe a similar aggregation behavior with the classical two-state protein CI2. Since both U1A and CI2 appear to fold directly from the coil at low protein concentrations, it is likely that the aggregates also form directly from the coil. This is in contrast to the behavior of larger multistate proteins where aggregation occurs in connection to "sticky" intermediates.

Protein aggregation is a poorly understood process which has gained an increasing interest in recent years, largely because of its connections to neurodegenerative pathologies, such as Alzheimer's disease and the controversial prion diseases (*1*, *2*). Also, the formation of insoluble inclusion bodies makes aggregation a central problem for large-scale production of recombinant proteins. For a recent review on protein aggregation, see ref *3*. Typically, aggregation is observed under conditions that thermodynamically favor the native protein conformation, e.g., under polypeptide synthesis in vivo (*4*), in the unfolding transition region, or when unfolded protein is suddenly brought to refolding conditions in vitro. Precursors for protein aggregation are often suggested to be folding intermediates (*5*−*7*) or molten globule conformations (*8*−*10*). Largely because such partly structured species accumulate in connection to aggregation (*5*) but also because structurally defined precursors provide a simple explanation for the seemingly specific organization of many protein aggregates (*11*). In disease-related aggregation, however, there are also several examples of specific aggregates arising directly from unfolded peptides and proteins (*12*, *13*). As the visible signs of aggregation manifest themselves over relatively long time scales, it has been possible to follow aggregation kinetics by equilibrium techniques with high structural sensitivity, for example, electrophoresis (*14*−*16*), chromatographic separation (*15*, *17*), and small-angle X-ray scattering (*18*). For faster aggregation events it has often been sufficient to start reactions by manual mixing (*9*, *19*−*21*). The limited time resolution of these experiments, however, has led to focus most efforts on the slow processes in protein aggregation and on conditions where the aggregates are long-lived enough to be detected. Therefore, very little is known about the dynamic aspects of protein aggregation, and there is a severe lack of detail at the microscopic level. Nevertheless, there is clear evidence for a dynamic exchange between aggregates and protein monomers from identification of transient oligomeric species in refolding experiments (*20*, *22*, *23*). This dynamic picture of protein aggregation is supported by rapid-mixing experiments which show that reversible contacts between denatured protein molecules can form also on very short time scales (*7*, *24*). The 102 residue protein U1A (Figure 1) aggregates on a millisecond time scale, sometimes in the dead time of measurements. The aggregates are short-lived and dissociate and fold in a fraction of a second, just a few times more slowly than the normal folding reaction. The propensity to form the transient aggregates is high, and they accumulate extensively also at very low protein concentrations ($\sim$1 μM). In kinetic terms, the rapidly formed aggregates of U1A closely resemble transient folding intermediates (*24*) and produce many characteristics of multistate folding behavior (*25*). At sufficiently low protein concentrations ($<$1 μM), however, U1A appears to fold directly from the coil in a highly concerted two-state process (*24*). An immediate conclusion from these results would be that the precursor for transient aggregates of U1A is the coil, i.e., the starting material for the folding reaction, not "sticky" intermediates as observed in folding studies of multistate proteins (*26*).

In this study we investigate further the role of unfolded protein in transient aggregation by comparing the folding behavior of U1A with a more extensively characterized two-state protein, namely, chymotrypsin inhibitor 2 (CI2) (Figure 1). CI2 was the first protein observed to display two-state kinetics under physiological conditions (*27*), and its two-

† Our work is supported by the Swedish Natural Science Research Council, Sven och Ebba-Christinas Stiftelse, The Royal Swedish Academy of Sciences, and the Sven and Lilly Lawski Foundation for Natural Science.

* To whom correspondence should be addressed: tel +46 46 222 01 07; fax +46 46 222 45 34; e-mail mikael.oliveberg@biokem.lu.se.

‡ Lund University.

§ National University of Singapore.

‖ Cambridge Center for Protein Engineering.

[1] Abbreviations: GdnHCl, guanidine hydrochloride; D, denatured protein; N, native protein.

10.1021/bi9909997 CCC: $18.00
Published on Web 09/17/1999

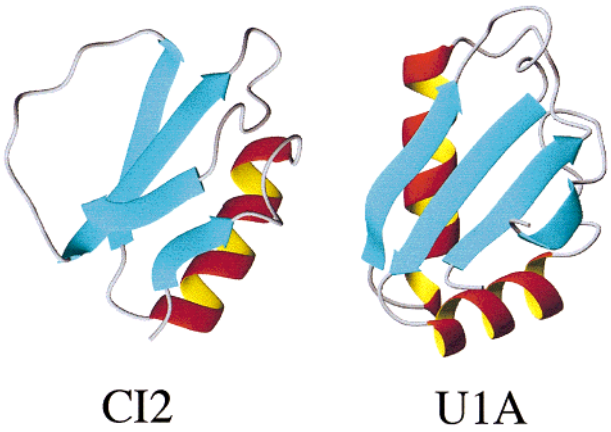

FIGURE 1: Crystal structures of CI2 (PDB code: 1YPB) and U1A (PDB code: 1FHT).

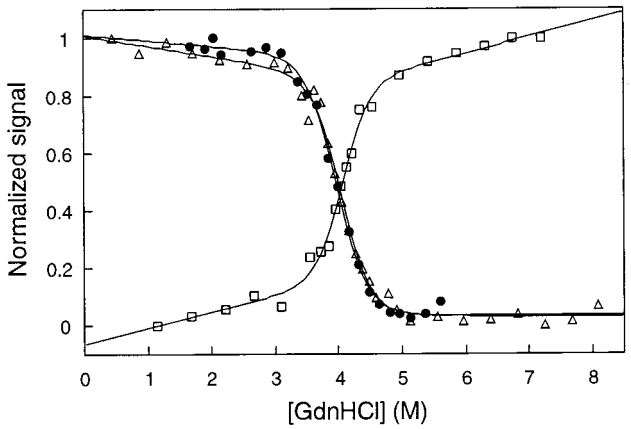

FIGURE 2: Equilibrium denaturation of U1A using fluorescence (●), fluorescence anisotropy (△), and far-UV CD (□). The unfolding midpoints and m_{D-N} are the same within experimental error (Table 1).

Table 1: Stability, Transition Midpoints, and *m*-Values of U1A

measurement	$[GdnHCl]_{1/2}$ (M)	m_{D-N} (kcal mol^{-1} M^{-1})	ΔG_{H_2O} (kcal mol^{-1})
fluorescence 1 μM[a]	3.98 ± 0.03	2.49 ± 0.2	9.9 ± 0.3
fluorescence 10 μM[a]	4.07 ± 0.05	2.32 ± 0.3	9.5 ± 0.6
CD[a]	4.13 ± 0.04	2.26 ± 0.3	9.3 ± 0.6
anisotropy[a]	4.03 ± 0.03	2.89 ± 0.4	11.7 ± 0.8
kinetics	4.08[c]	2.20 ± 0.2[d]	9.1 ± 0.2[b]

[a] The value is determined by standard GdnHCl titration experiments (*31*). [b] $\Delta G_{D-N}^{H_2O} = -2.3RT \log K_{D-N} = -2.3RT (\log k_u - \log k_f^{fast})$ at [GdnHCl] = 0 M (cf. eq 1). $\log k_f^{fast}$ corresponds to refolding at low protein concentration (*24*). [c] Obtained from the minimum of the chevron plot, i.e., where $\log k_u = \log k_f$ (24). [d] From eq 1 it follows that $m_{D-N} = m_{‡-N} - m_{D-‡}$, where $m_{‡-N}$ and $m_{D-‡}$ are derived from the slopes of $\log k_u$ and k_f, respectively (*24*), and ‡ represents the transition state.

state properties have since then been confirmed by a wealth of other studies (*28*−*30*). Our results show that also CI2 can form short-lived refolding aggregates. Since CI2 does not accumulate folding intermediates, this provides additional evidence that aggregates can form directly from unfolded protein. That is, two-state proteins can undergo "two-state" aggregation.

EXPERIMENTAL PROCEDURES

All experiments were done at 25 °C in 50 mM Mes buffer at pH = 6.3 (19 mM Mes acid and 31 mM Mes sodium salt), purchased from Sigma. All solutions were prepared volumetrically.

Materials. The proteins examined in this study are the RNA-binding domain RBD1 (residues 1−102) of the human U1 small nuclear ribonucleoprotein A (U1A) and a truncated version (residues 20−83) of chymotrypsin inhibitor 2 (CI2) from barley (Figure 1) expressed and purified as described in ref *27*.

Equilibrium Unfolding. Unfolding was induced by GdnHCl[1] (ultrapure from Gibco/BRL Life Technologies). Fluorescence measurements were done on a Perkin-Elmer LS50B, CD spectra on a Jasco J-720 using 0.5 mm path-length cuvettes, and fluorescence anisotropy on a Spex fluorolog FL 212. Data points are averages of three scans. Protein stability was derived by standard procedure linear regression, i.e., assuming that protein stability and $\log K_{D-N} = \log([D]/[N])$ depend linearly on the concentration of GdnHCl (*31*).

Kinetics. Kinetic measurements were done with an SX.18MV stopped-flow instrument from Applied Photophysics. Mixing was 1:10 (GdnHCl jump) or 1:1 (pH jump). In refolding by GdnHCl jump, U1A was denatured in 5.1 M GdnHCl (*24*) and CI2 in 5.2 M GdnHCl. In pH jump, CI2 was denatured in 32 mM HCl (pH 1.5) (*32*). Excitation was at 280 nm and detection was with a 315 nm cutoff filter. Curve fitting was done with the SX.18MV software or with the software package Kaleidagraph (Abelbeck Software). Dead-time spectra were recorded with a second monochromator on the emission light. Bandwidths were 2 nm for the excitation and 3 nm for the emission.

RESULTS

Equilibrium Unfolding of U1A Is Two-State. Intermediates in protein folding are often difficult to detect by equilibrium techniques, partly because the denaturant needed to unfold the native protein also unfolds possible intermediates and partly because equilibrium unfolding is insensitive to rapidly formed intermediates. A common way to infer the existence of equilibrium intermediates, however, is from noncoincident unfolding curves measured by two or more techniques with different sensitivities to secondary and tertiary structure (*23*). If the midpoints do not coincide, this indicates that tertiary and secondary structure do not unfold concomitantly and that a partly structured species populates in the unfolding transition. To look for such deviations with U1A, we monitored unfolding by fluorescence, CD, and fluorescence anisotropy (Figure 2). The results show that both the midpoints and the GdnHCl dependence of the protein stability (m_{D-N}) coincide within experimental error (Table 1), suggesting that the unfolding transition is a highly concerted two-state process. The observation is corroborated by Hall and co-workers, using CD and differential scanning calorimetry (*33*). The absence of equilibrium intermediates under partly denaturing conditions, however, does not exclude that metastable intermediates accumulate during folding under physiological conditions.

Folding Kinetics of U1A Changes from Two-State to Apparent Three-State as a Result of Transient Aggregation. Short-lived intermediates that accumulate early in the folding reaction may be indicated kinetically as follows: For a two-state system D ↔ N, the rates of unfolding, k_u, and refolding, k_f, are related to the equilibrium constant K_{D-N} by

$$K_{D-N} = \frac{[D]}{[N]} = \frac{k_u}{k_f}; \qquad \log K_{D-N} = \log k_u - \log k_f \qquad (1)$$

Deviation from eq 1, i.e., that the observed refolding rate is slower than that calculated from K_{D-N} and k_u, is usually taken to indicate that an intermediate has formed in the mixing time of the stopped-flow instrument. Refolding is retarded because the intermediate is a more stable ground state than D and, hence, the activation barrier for folding that must be overcome is higher (*25*).

U1A shows clear deviations from two-state behavior at typical stopped-flow concentrations ([U1A] $\approx$ 3 μM), manifested in a downward curvature in the refolding limb of the chevron plot (*24*) (Figure 3). A distinct feature of this deviation is the positive slope of log k_f versus [GdnHCl] under near native conditions, which signals an expansion upon activation, as seen in the unfolding process (*24*, *34*). At low concentrations of protein, (<1 μM), folding becomes faster and the rate constants coincide precisely with those expected for two-state folding (Figure 3). Another characteristic feature of U1A is the symmetrically curved chevron plot which is seen under two-state conditions. Most other two-state proteins, for example, CI2, show V-shaped chevron plots with linear unfolding and refolding limbs. Still, fitting second-order polynomials to the unfolding and refolding limbs of U1A yields precisely two-state kinetics and a linear dependence of log K_{D-N} vs [GdnHCl] according to eq 1 (*35*, *36*). On this basis, the curvatures have been connected to movements of the transition state ensemble rather than accumulation of folding intermediates. Consistently, similar curvatures have been induced in other two-state proteins, CI2 (*37*, *38*) and S6 (*39*), by mutations that affect the fine structure of the activation barrier and, hence, the behavior of the transition state ensemble. Although it then appears that monomeric U1A is two-state and that the unfolded protein is the aggregating species, the picture is somewhat muddled by the anomalously curved chevron plot. Therefore, we have here applied the same analysis to the better characterized two-state protein CI2.

CI2 Confirms That Transient Aggregation Can Arise from Unfolded Protein. CI2 normally displays three refolding phases. The first phase ($\sim$70 s^{-1}) is folding of protein with all *trans*-prolines, and the last phase ($\sim$0.05 s^{-1}) arises from isomerization in the remaining *cis*-proline population (*32*). The second phase ($\sim$10 s^{-1}), however, is shown here to be associated with transient aggregation. As with U1A, the rapid first phase of the refolding trace shows two-state folding of monomeric protein (*24*). This phase constitutes most of the refolding amplitude at low concentrations of protein. The slower, second phase which results from folding of transient aggregates grows larger at high concentrations of protein. The increase in amplitude of this second phase causes an overall retardation of the refolding reaction (Figure 4). Although the aggregation of CI2 is not as pronounced as with U1A, the effects on the folding amplitudes are still clearly resolved at moderately increased protein concentrations (10$-$100 μM). The corresponding refolding rate constants show a small decrease: between 1 and 400 μM CI2, the first phase changes from $\sim$70 to $\sim$40 s^{-1} and the second phase from $\sim$10 to $\sim$6 s^{-1}. We cannot determine whether this small change is real or just an artifact of the curve fitting as the amplitudes change; if the second phase is locked to 10 s^{-1}, the first phase remains at 70 s^{-1} throughout the concentration range with fits as good as those when both phases are free to vary in the equations.

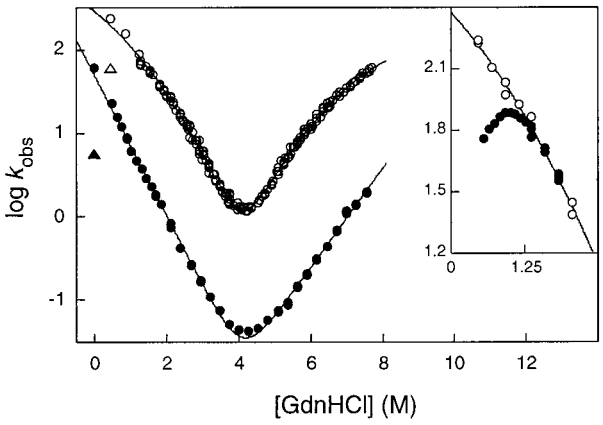

FIGURE 3: Chevron plots of U1A (○) and CI2 (●) showing the GdnHCl dependence of the refolding rate constant (k_f) and the unfolding rate constant (k_u). The units are in s^{-1}. The right limbs of the plots show log k_u and the left limbs log k_f. Two-state folding is indicated by the solid curves. At protein concentrations below 1 μM, U1A folds in a two-state process (○), whereas at higher protein concentrations refolding becomes slower than predicted from eq 1 (△). The deviation is here exemplified with the U1A mutant I84A which has been particularly well characterized with respect to the aggregation behavior (insert): log k_f at [U1A] = 0.1 μM (○) and at [U1A] = 3 μM (●). For CI2, the corresponding retardation of log k_f is clearly manifested first above 200 μM (▲). The deviation from two-state behavior (eq 1) is seen only at low [GdnHCl] and is caused by transient aggregation of denatured protein.

Thus, it appears that CI2 shows an aggregation behavior similar to that of U1A, albeit at higher concentrations of protein. That is, folding of CI2 may occur either directly and rapidly from monomeric protein or more slowly from reversible aggregates. Direct folding is two-state, whereas folding from the aggregate is slower than predicted from eq 1. At high protein concentrations this leads to retardation of the refolding kinetics and deviation from two-state behavior (Figure 3). Interestingly, the aggregate phase of CI2 ($\sim$10 s^{-1}) is present at a residual level also at very low protein concentrations (5$-$10%). The residual population of aggregates may be caused by protein trapped in *cis*-proline conformations. Unfortunately, it is not possible to compare the aggregation behavior of U1A and CI2 under identical conditions: U1A data refer to GdnHCl-jump experiments to final conditions of $\sim$0.5 M GdnHCl whereas CI2 data are based on acid-jump experiments to physiological conditions. This is because the refolding phases of CI2 at 0.5 M GdnHCl are too close to be accurately resolved and because U1A cannot be denatured by acid. Nevertheless, it is evident that, according to the two-state criterion in protein folding (eq 1), the aggregating species is the unfolded protein.

Characterization of the Aggregates. Fluorescence spectra of the dead-time species were constructed from individual refolding traces by plotting the fluorescence intensity level at time 0 versus emission wavelength. Excitation was at 280 nm. For U1A, the spectra of the native and the denatured protein are different, with the maximum of emission shifting from 335 nm for the native protein to 355 nm for the coil, with a concomitant decrease in intensity. The fluorescence

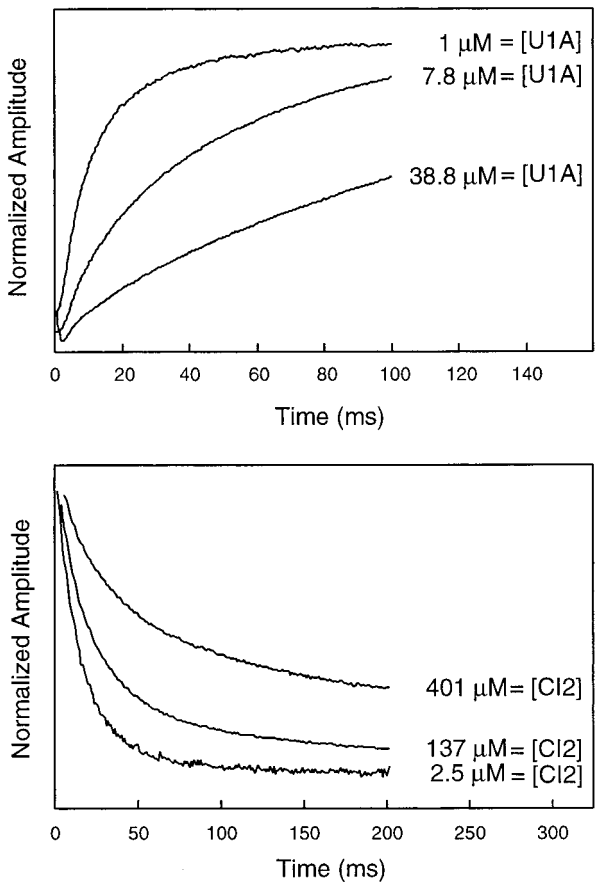

FIGURE 4: (Top) Time courses for refolding of U1A at different protein concentrations. The folding reaction is biphasic: the fast phase corresponds to two-state folding of monomeric protein and the slow phase to folding from reversibly formed aggregates. The amplitude of the slow phase increases at high protein concentrations, leading to an overall retardation of the refolding process. (Bottom) Corresponding data for CI2. In contrast to U1A, CI2 shows a fluorescence decrease upon refolding, and the onset of aggregation occurs at higher protein concentrations.

of CI2 is quenched in the native state, and denaturation produces an increase in intensity. The emission maxima are 340 and 356 nm for N and D, respectively. At low protein concentrations ($\sim$1 μM), the dead-time species of both U1A and CI2 have spectral characteristics consistent with unfolded protein, in agreement with previous results. The slight blue shift observed for the dead-time species of U1A is likely due to a general increase in the compactness of the coil under poor solvent conditions (*40*). Interestingly, the dead-time species show the same coil-like spectrum also at high protein concentrations (Figure 6). This suggests that the fluorescing tryptophans are as solvent exposed in the aggregate as in the unfolded protein; i.e., the tryptophans are not directly involved in the aggregate structure. Considering the high hydrophobicity of these residues, the results disfavor an aggregation mechanism driven by nonspecific hydrophobic collapse.

DISCUSSION

Could Aggregates Be Present Prior to Initiation of the Folding Reaction? A simple explanation for the aggregation behavior of U1A and CI2 would be that the aggregates are present already before initiation of the folding reaction, i.e., in the reservoir syringe containing the denatured protein. In

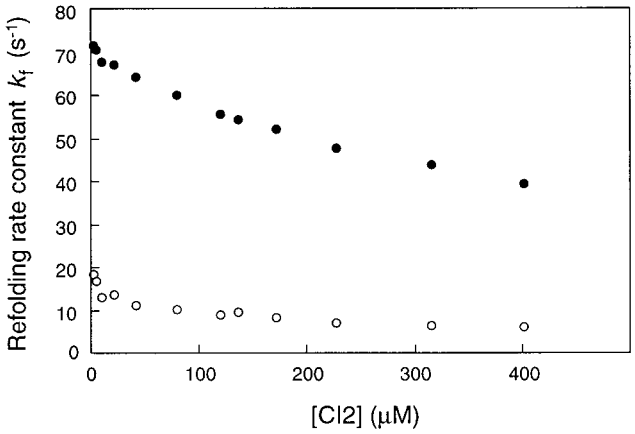

FIGURE 5: Protein concentration dependence of the rate constants of the fast (●) and slow (○) refolding phase of CI2. It is uncertain whether the small concomitant decrease in the rate constants is real or just an artifact of the curve fitting procedure: if the second phase is locked to 10 s^{-1} the fast phase remains constant without significantly affecting the quality of the fit.

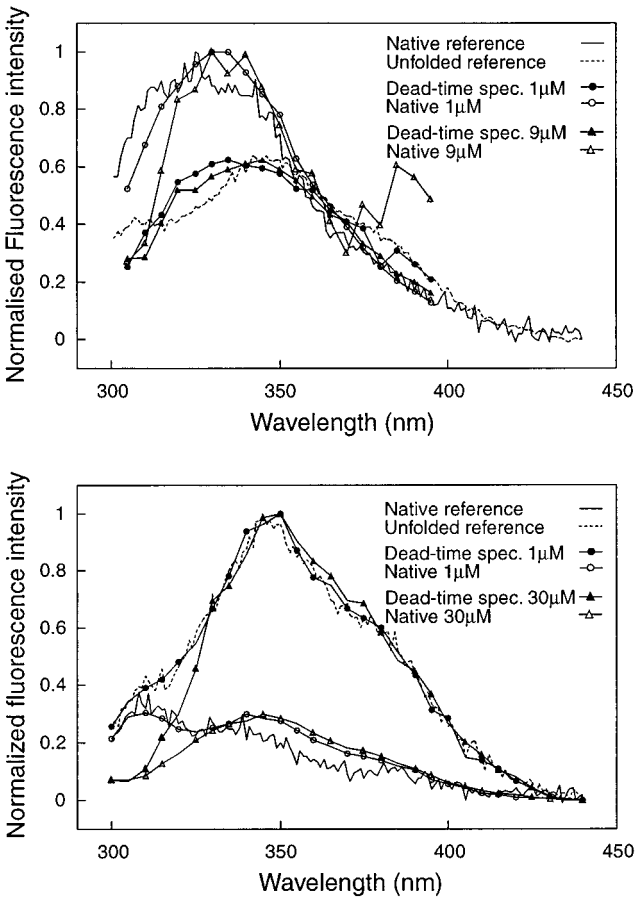

FIGURE 6: Normalized dead-time spectra of the refolding species of U1A and CI2 at different protein concentrations. The dead-time spectra of both U1A and CI2 resemble that of the unfolded protein at all protein concentrations, indicating that the tryptophans are not involved in the aggregation process. Reference spectra were obtained with the SX.18MV instrument in equilibrium mode and are averages of nine scans.

the case of U1A this would mean at 5.2 M GdnHCl and with CI2 at pH 1.5. This appears unlikely since 5 M GdnHCl dissolves readily any macroscopic aggregates of U1A. Also no signs of aggregation have been observed in NMR studies of the acid-denatured state of CI2 ([CI2] = 600 μM, pH 1.6, 40 °C) (*30*), and solubility studies suggest that aggregates

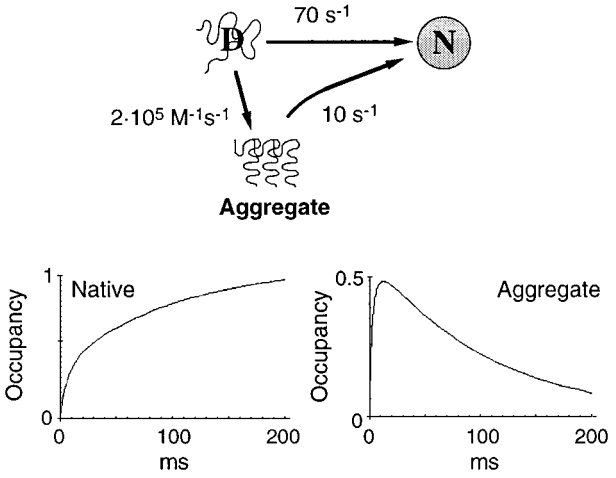

FIGURE 7: Reaction schemes of U1A and CI2, where aggregation and folding occur in parallel. The rate constants refer to refolding in the absence of denaturant. The lower panel show corresponding simulations of the occupancy of native protein and aggregates at [CI2] = 0.6 mM under the assumption that the aggregates are dimers (software: Mathematica). From the kinetic resemblance between transient aggregation fragment association, we speculate that the aggregates are specifically joined by intermolecularly formed folding nuclei. Note that the results do not exclude that folding of the aggregates proceeds via the denatured ensemble.

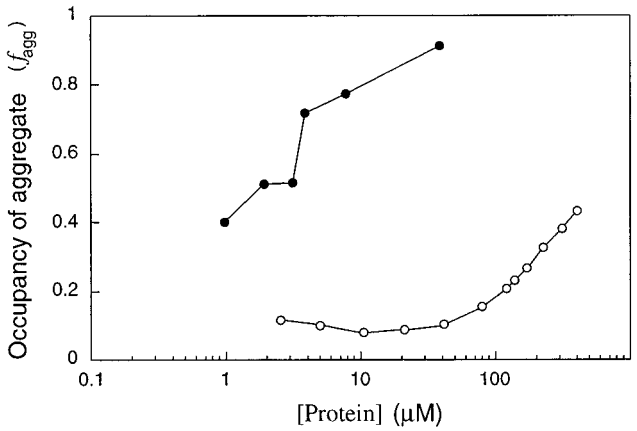

FIGURE 8: Occupancy of the transient aggregates of U1A (●) and CI2 (○) at different protein concentrations calculated from eq 2. In fits where [U1A] ≥ 3.1 μM, the rate constant of the fast phase was locked to 200 s^{-1}. The protein concentration is shown on a logarithmic scale.

are generally not stable at high concentrations of denaturant (*5*, *41*, *42*). In addition, an equilibrium between unfolded molecules and aggregates would "consume" unfolded protein and thereby lower the unfolding midpoint at high concentrations of protein. No such effects can be detected with either U1A (Table 1) or CI2 (*30*).

Kinetic Competition between Folding and Aggregation. After mixing the denatured protein into physiological conditions, aggregates could either form in a preequilibrium relaxation in the dead time of measurements (<2 ms) or in parallel with the folding reaction. Preequilibration assumes a fast and reversible exchange between monomers and aggregates. This gives rise to a single refolding phase in which monomers and aggregates disappear together. In contrast, U1A and CI2 show biphasic refolding which rather indicates that, once formed, the aggregates constitute traps which fold slower than the monomeric protein. More plausibly, folding and aggregation occur in parallel (*19*, *43*, *44*). Thus, the protein could either fold directly by the fast phase from the coil or become kinetically trapped in an aggregate (Figure 7). The probability of either event is simply ordained by protein concentration. The fast refolding phase then measures the direct two-state channel. Competing formation of aggregates takes place at the same time but is not resolved since the accompanying change in tryptophan fluorescence is negligible. At increased concentrations of protein, however, the aggregation is revealed by a decreased amplitude of the fast phase; the more rapid the trapping, the smaller the fraction of monomeric protein folding fast. The slow phase shows, accordingly, the conversion of aggregates into native monomers. In Figure 7, we illustrate this reaction as a parallel route, although it is equally possible that the aggregates are dead ends which fold via the unfolded ensemble.

The Extent of Aggregation Can Be Estimated from the Extent of Slow Refolding. The similar spectra of unfolded protein and the aggregates make it possible to estimate the occupancy of aggregates (f^{agg}) according to

$$f^{agg} = \frac{A_{slow}}{A_{slow} + A_{fast}} \quad (2)$$

where A_{slow} and A_{fast} are the amplitudes of the slow and fast refolding phases, respectively. With U1A, the aggregation is already significant at 1 μM of protein (Figure 8). With CI2, corresponding levels of aggregation are not reached below 200 μM.

The Aggregation Rate Is Near Diffusion Control. We assume in Figure 7 that direct folding and aggregation have similar rates when half of the protein population escapes by the fast channel and half is trapped in aggregates, i.e., when $f^{agg} = 0.5$. At 2 μM protein the "effective" rate of aggregation is thus ~200 s^{-1} for U1A, yielding an association rate constant of ~1×10^8 s^{-1} M^{-1} (Figure 7). This value is near the diffusion limit for a 100 residue protein (*45*, *46*). Similar rates have been found for the association step in the folding of the dimeric Arc repressor (*45*). For CI2 the association is considerably slower, at 3×10^5 s^{-1} M^{-1}. The reason for CI2's lower association rate is unclear but may be due to differences in charge, specificity, precursor conformation, or even the order of aggregation. For comparison, the rate constant for association of fragments of cleaved CI2 is also rather slow, 3.7×10^3 s^{-1} M^{-1} (47).

Aggregation Directly from the Coil. Aggregation is often connected to intermediates and molten globules (*5*, *6*, *8*−*10*), largely because such states often accumulate in connection to aggregation (*5*). Further evidence for the involvement of intermediates has been derived from the apparent lack of correlation between aggregation and protein stability (*48*), see also ref *49*. But why do intermediates aggregate? One explanation is that they provide large, contiguous hydrophobic patches which merge upon aggregation (*3*). Coils are less susceptible to associate because they display broken patterns of hydrophobic and polar/charged groups which are more difficult to match (*3*). Unfolded proteins may also be contracted with a bias of charged residues at the periphery (*30*, *40*). A refined version of the "associating intermediate" model postulates that aggregates are kept together by complementary, native-like surfaces (*50*−*52*). In principle,

this idea could be extended to include any stable motifs, also non-native, which are compatible with the primary sequence. The mechanism is epitomized in domain swapping where two or more native structures are interconnected by "overlapping" polypeptides (*53*).

A different conclusion is reached in a solubility study of apomyoglobin where unfolded protein, rather than an intermediate, seems to be the precursor for reversible aggregation in urea denaturation experiments (*41*). It is argued that solubility problems around the unfolding transition region are most simply related to the occupancy of unfolded protein under poor solvent conditions, in this case around 2.5 M urea. At lower urea concentrations, the soluble native conformation is predominant and at higher urea concentrations, the aggregates dissolve. The authors find no evidence for intermediates under these conditions and point out that several of the experimental features associated with intermediates can also arise from compact denatured states. The results in this study support this view. With U1A and CI2 the aggregates appear to form directly from unfolded proteins. The main difference is that by rapid mixing we are able to plunge the unfolded protein into solvent conditions which are much worse than in the transition region. This causes the protein to aggregate both faster and at lower concentrations. Another difference is that we do not resolve precipitation under stopped-flow conditions, probably because there is too little time for the unfolded protein to form large enough aggregates to pass the solubility limit.

The Mechanism of Association. Unspecific aggregation of hydrophobic moieties seems inconsistent with the seemingly ordered behavior of the transient aggregates: they form reversibly, are precisely reproducible from one experiment to another, and appear independent of "dirty" cells and even the presence of other denatured proteins. For example, refolding of 1 μM U1A in the presence of 10 μM slow folding protein does not give rise to increased levels of aggregation; i.e., U1A only aggregates with itself, indicating some degree of specificity (data not shown). In unspecific aggregation, one would also expect involvement of the tryptophans, but the dead-time spectra in Figure 6 clearly disfavor this. Thus, the aggregation is likely to be specific. Interfaces for aggregation of unfolded protein could then be high-energy states in the denatured ensemble, or they could be induced upon association. The latter may be exemplified by arc repressor (*45*, *54*). Native arc repressor is a homodimer which folds in one step from the two unfolded peptides without accumulating intermediates. A related example is the folding of complementary peptide fragments, i.e., fragments that result from cutting a protein in two pieces (*29*). Separated, fragments are unable to form any structure, but when mixed together they adopt the native structure in one highly concerted step without intermediates. In the case of CI2, the fragment results have been taken as evidence for a nucleation condensation mechanism where the fragments join by a folding nucleus which is only "flickeringly" present in the denatured ensemble (*29*). Interestingly, there is an underlying resemblance between the kinetics of fragment recombination and aggregation. Although the fragment recombination of CI2 appears 50 times slower than the aggregation, part of this difference is accounted for by the fragments' need for complementarity: while aggregates may form at every collision, collisions between identical fragments are unproductive for folding. Further, the intact chain is able to form native-like overlaps in more than one way: head to tail, tail to head, and "full domain swap". A conservative estimate is that this intrinsically makes aggregation around 12 times faster than fragment association, reducing the rate difference to less than a factor of 5. The difference can be reduced further if we assume larger aggregates than dimers. It is plausible that the kinetic correspondence reflects a similarity between the folding nuclei and the aggregation nuclei; the only difference may be that folding nuclei are formed intramolecularly and aggregation nuclei are formed between complementary regions of separate chains; cf. Broglia et al. (*52*). CI2's tendency to form such complementary overlaps is verified by crystallography where CI2 mutants with loop extensions in some cases are domain swapped (K. Stott and A. R. Fersht, unpublished). Also, the transient aggregation of CI2 is enhanced by helix-stabilizing mutations (A. Laudurner and A. R. Fersht, unpublished); the helix forms an integral part of CI2's folding nucleus, and its accumulation may increase the probability for productive association by a domain swap mechanism. In a broader perspective, the interpretation is consistent with a funnel-like energy landscape for folding where native contacts are more favorable than non-native (*55*); i.e., aggregation as well as folding follows the "principle of minimal frustration". Evolution allows only one choice of interaction: the native contacts.

The high association rate of the homodimeric arc repressor may thus be explained by the protein's extensive intramolecular interface which could form in a more flexible manner than complementary overlaps. It is possible that U1A's equally fast aggregation signals the involvement of similar quaternary contact patterns. In this case, the linking interface would however be non-native and be caused by a promiscuous stretch of the sequence which could be called "intermolecularly frustrated"; i.e., the stretch can form ordered structures with homologous stretches of other polypeptides although it is intramolecularly well behaved. To this end, it is tempting to draw a parallel with a long-lasting ambiguity in protein folding, namely, the role and nature of folding intermediates. Why are intermediates needed as precursors for aggregation when, as observed here for two-state proteins, aggregation occurs readily and even faster from the coil?

ACKNOWLEDGMENT

We thank Daniel Otzen for providing the dead-time spectra of U1A in Figure 6, Mikael Akke for help with Figure 1, Tomas Ternström for the data on the I84A mutant of U1A, and Jörgen Ström for help with anisotropy and CD measurements.

REFERENCES

1. Carrell, C. W., and Lomas, D. A. (1997) *Lancet 350*, 134−138.
2. Thomas, P. J., Qu, B. H., and Petersen, P. L. (1995) *Trends Biochem. Sci*., 456−459.
3. Fink, A. L. (1998) *Folding Des. 3*, R9−R23.
4. Klein, J., and Dhurjati, P. (1995) *Appl. Environ. Microbiol. 61*, 1220−1225.
5. Brems, D. N. (1988) *Biochemistry 27*, 4541−4546.
6. King, J., Haase-Pettingell, C., Robinson, A. S., Speed, M., and Mitraki, A. (1996) *Trends Biochem. Sci. 10*, 57−66.

7. Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O., and Kihara, H. (1993) *Biophys. J. 65*, 912−917.
8. Brems, D. N., and Havel, H. A. (1989) *Proteins: Struct., Funct., Genet. 5*, 93−95.
9. Cleland, J. L., and Wang, D. I. C. (1990) *Biochemistry 29*, 11072−11078.
10. Safar, J., Roller, P. P., Gadjusek, C., and Gibbs, C. J. (1994) *Biochemistry 33*, 8375−8383.
11. Speed, M. A., Wang, D. I., and King, J. (1996) *Nat. Biotechnol. 10*, 1283−1287.
12. Weinreb, P. H., Zhen, W., Poon, A. W., Conway, K. A., and Lansbury, P. T. J. (1996) *Biochemistry 35*, 13709−13715.
13. Walsh, D. M., Lomakin, A., Benedek, G. B., Condron, M. M., and Teplow, D. B. (1997) *J. Biol. Chem. 272*, 22364−22372.
14. Creighton, T. E. (1986) *Methods Enzymol. 131*, 156−172.
15. Kwon, K.-S., Lee, S., and Yu, M.-H. (1995) *Biochim. Biophys. Acta 1247*, 179−184.
16. van den Oetelaar, P. J. M., de Man, B. M., and Hoenders, H. J. (1989) *Biochim. Biophys. Acta 995*, 82−90.
17. Saxena, A. M. (1988) *J. Mol. Biol. 200*, 579−591.
18. Hiragi, Y., Inoue, H., Sano, Y., Kajiwara, K., Ueki, T., Kataoka, M., Tagawa, H., Izumi, Y., Muroga, Y., and Amemiya, Y. (1988) *J. Mol. Biol. 204*, 129−140.
19. Goldberg, M. E., Rudolph, R., and Jaenicke, R. (1991) *Biochemistry 30*, 2790−2797.
20. Pecorari, F., Minard, P., Desmadril, M., and Yon, J. M. (1996) *J. Biol. Chem. 271*, 5270−5276.
21. Speed, M. A., Wang, D. I., and King, J. (1995) *Protein Sci. 4*, 900−908.
22. Cleland, J. L., and Wang, D. I. (1992) *Biotechnol. Prog. 8*, 97−103.
23. Jamin, M., and Baldwin, R. L. (1998) *J. Mol. Biol. 276*, 491−504.
24. Silow, M., and Oliveberg, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084−6086.
25. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
26. Mitraki, A., and King, J. (1989) *Bio/Technology 7*, 690−697.
27. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
28. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
29. Neira, J. L., Davis, B., Ladurner, A. G., Buckle, A. M., de Prat Gay, G., and Fersht, A. R. (1996) *Folding Des. 1*, 189−208.
30. Tan, Y.-J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 980−992.
31. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−279.
32. Tan, Y. J., Oliveberg, M., Otzen, D. E., and Fersht, A. R. (1997) *J. Mol. Biol. 269*, 611−622.
33. Lu, J., and Hall, K. B. (1997) *Biophys. Chem.*, 111−119.
34. Oliveberg, M. (1997) *Acc. Chem. Res. 31*, 765−772.
35. Silow, M., and Oliveberg, M. (1997) *Biochemistry 36*, 7633−7637.
36. Oliveberg, M., Tan, Y.-J., Silow, M., and Fersht, A. (1998) *J. Mol. Biol. 277*, 933−943.
37. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) *Biochemistry 34*, 13656−13662.
38. Dalby, P. A., Oliveberg, M., and Fersht, A. R. (1998) *Biochemistry 37*, 4674−4679.
39. Otzen, D. E., Kristensen, O., Proctor, M., and Oliveberg, M. (1999) *Biochemistry 38*, 6499−6511.
40. Oliveberg, M., Vuilleumier, S., and Fersht, A. R. (1994) *Biochemistry 33*, 8826−8832.
41. De Young, L. R., Dill, K. A., and Fink, A. L. (1993) *Biochemistry 32*, 3877−3886.
42. Rudolph, R., Zettlmeissl, G., and Jaenicke, R. (1979) *Biochemistry 18*, 5572−5575.
43. Zettlmeissl, G., Rudolph, R., and Jaenicke, R. (1979) *Biochemistry 18*, 5567−5571.
44. Kiefhaber, T., Rudolph, R., Kohler, H. H., and Buchner, J. (1991) *Bio/Technology 9*, 825−829.
45. Milla, M. E., and Sauer, R. T. (1994) *Biochemistry 33*, 1125−1133.
46. Schreiber, G., and Fersht, A. R. (1996) *Nat. Struct. Biol. 3*, 427−431.
47. de Prat Gay, G., and Fersht, A. R. (1994) *Biochemistry 33*, 7957−7963.
48. Chrunyk, B. A., Evans, J., Lillquist, J., Young, P., and Wetzel, R. (1993) *J. Biol. Chem. 268*, 18053−18061.
49. Liemann, S., and Glockshuber, R. (1999) *Biochemistry 38*, 3258−3267.
50. Oberg, K., Chrunyk, B. A., Wetzel, R., and Fink, A. L. (1994) *Biochemistry 33*, 2628−2634.
51. Uversky, V. N., Segel, D. J., Doniach, S., and Fink, A. L. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 5480−5483.
52. Broglia, R. A., Tiana, G., Pasquali, S., Roman, H. E., and Vigezzi, E. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 12930−12933.
53. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) *Protein Sci. 4*, 2455−2468.
54. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry 34*, 13914−13919.
55. Bryngelson, J. D., and Wolynes, P. G. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 7524−7528.

BI9909997