Binary and Ternary Aggregation within Tethered Protein Constructs

Wei Yuan Yang* and Martin Gruebele*†‡
*Center for Biophysics and Computational Biology, and †Department of Chemistry and ‡Department of Physics, University of Illinois at Urbana-Champaign, Champaign, Illinois

ABSTRACT The free energy per monomer of a protein aggregate varies with the number of participating monomers n. The change of this free energy with aggregate size, $\Delta\Delta G(n)$, is difficult to determine by sedimentation or concentration studies. We introduce a kinetic approach to quantitate the free energy of aggregates in the presence of tethers. By linking the protein U1A into dimers and trimers, a high effective concentration of the monomers is achieved, together with exact size control of the aggregates. We found that the free energy of the aggregate relative to the native monomer reached a maximum for $n = 2$, and decreased by $\Delta\Delta G(2) = -3.1\ kT$ between dimer and trimer.

INTRODUCTION

Protein folding, misfolding, and aggregation are intimately connected processes in the crowded environment of the cell. A number of natively unfolded proteins or destabilized mutants form aggregates that eventually progress to protofibrils and fibrils, leading to disease states (1). With the aid of denaturants and other means of destabilization, essentially any protein can be made to form aggregates (2).

As protein aggregates build up from monomers, the free energy initially increases as a function of n ($\Delta\Delta G(n) > 0$), until a thermodynamic nucleus is reached (3). This creates an effective barrier to aggregation if the nucleus is reached for $n > 1$. Then the free energy decreases as further monomers are added ($\Delta\Delta G(n) < 0$), until a value of n is reached where the aggregate becomes more stable than the native state (a stable ''structural'' nucleus). These nuclei involve at most a few copies of the monomer assembled (4).

At low protein concentrations, transient subcritical aggregates prevail, but they can compete with direct folding from the denatured state to the native structure (5). It has been postulated that protein folding barriers have evolved in part to prevent partially folded structures that favor transient association (6). Indeed, proteins that fold ''downhill'' over minimal free energy barriers are found to transiently aggregate more than slower folders (7).

Understanding how aggregate stability depends on aggregate size is fundamental to understanding the transition from transient aggregates (including the thermodynamic nucleus) to stable aggregates (sizes beyond the stable structural nucleus). Unfortunately, information about the high-energy (and thus scarce) thermodynamic nuclei cannot be obtained easily from concentration or sedimentation studies, where two-, three- and higher n-mer protein-protein interactions all compete simultaneously. Thus, $\Delta\Delta G(n)$ is difficult to determine by conventional thermodynamic techniques.

Here we develop a kinetic approach to address this question directly, and illustrate it experimentally with the 102-residue N-terminus of protein U1A. U1A is a spliceosomal RNA binding protein with an $\alpha + \beta$ sandwich structure (8). It has been shown by Oliveberg and co-workers that the U1A N-terminus is a millisecond two-state folder at submicromolar concentration. Above the micromolar range, its folding is preceded by two concentration-dependent slow phases due to transient aggregation into at least two aggregate forms (5,9,10). This makes U1A a useful prototype for our studies.

By tethering U1A repeats with flexible glycine linkers of varying length, we systematically study the thermodynamics and kinetics of the monomer, dimer, and trimer at low (micromolar or below) concentrations of the n-mer. The effective concentration of the monomers in these linked repeats is very high, leading to transient aggregation of the linked proteins. At the same time, the number of interacting units and their effective concentration (via tether length) can be controlled precisely. We determine how the free energy of the aggregate compared to the native state tunes with aggregate size. We also show that the unfolding thermodynamics of the construct are not highly sensitive to tether length or composition, opening the possibility for future tether length-dependent studies.

We hope that these results will stimulate statistical mechanical modeling of tethered transient protein association. Tethering automatically confines the monomers to a small volume, which can be useful for simulations. The effects of flexible linkers can be included in computational models such as off-lattice simulations with simplified interaction potentials. In a recent modeling study of protein-protein binding by Levy et al., glycine tethers, with a length designed to accommodate the N- to C-terminal distance between subunits in the dimer, were used to hold the monomers in proximity during the binding dynamics (11). This approach can be generalized for larger values of n along the lines of the experiments described here.

Submitted October 12, 2005, and accepted for publication January 6, 2006.

Address reprint requests to M. Gruebele, Dept. of Chemistry, University of Illinois at Urbana-Champaign, Champaign, IL 61801. Fax: 217-244-3186; E-mail: gruebele@scs.uiuc.edu.

0006-3495/06/04/2930/08 $2.00

doi: 10.1529/biophysj.105.075846

EXPERIMENTAL METHODS AND KINETIC MODELING

Proteins

Genes expressing the tethered proteins were constructed using the single U1A F56W (U1A* henceforth) gene obtained from the Baranger lab (12). Polymerase chain reaction was done with DNA primers containing the glycine linker sequences to expand the gene, followed by ligation into the original PET-28b U1A* vector (or in a second iteration into the vector containing the dimer, to make the trimer).

This was used to create five versions of the protein: the monomer U1A; a dimer with 10 glycines as the domain linker, DU1A*10; a dimer with six glycines as the domain linker, DU1A*6; a dimer with 14 residues in the linker, DU1A*14 (G_{12}AL); and a trimer with seven glycines as the first domain linker and 10 glycines as the second domain linker, TU1A*7,10). For brevity, the DU1A*10 and TU1A*7,10 constructs, which we investigated most extensively, are also referred to as DU1A* and TU1A* in the text. A model for the dimer with secondary structure elements labeled is shown in Fig. 1.

Proteins were expressed in Rosetta(DE3)pLysS cells (Novagen, Madison, WI) using lysogeny broth at pH 7.4, growing at 30°C after 2 mM isopropyl-β-D-thiogalactopyranoside induction overnight before harvesting. Although the monomers were isolated from the soluble fraction of the cell lysate, dimers and trimers were isolated from the inclusion bodies of the cell lysate. Purification was done using a Ni-NTA column (Qiagen, Valencia, CA; native conditions for the monomer and denaturing conditions for the dimer and trimer). His-tags were cleaved using Thrombin (Novagen) and removed using dialysis. The end protein product was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis and electrospray ionization mass spectrometry. Protein solutions were flash-frozen in 50 mM phosphate buffers at pH 7 and stored at –20°C for further usage.

Protein concentration is critical in the experiments to avoid interactions among n-mers and to observe only interactions among the construct repeats. After concentration-dependent checks as a function of pH, temperature, and guanidine hydrochloride (GuHCl) concentration by ultraviolet-visible, circular dichroism, and fluorescence spectroscopy, the following final protein concentrations were used in all the experiments: monomer, 0.5 μM; dimer, 0.2 μM; and trimer, 0.2 μM.

Denaturant melts

Protein stabilities were assessed by using GuHCl titrations on a JASCO J-715 CD spectrometer equipped with a two-syringe autotitrator. Protein solution temperatures were kept constant at 20°C during titrations by a Peltier temperature controller.

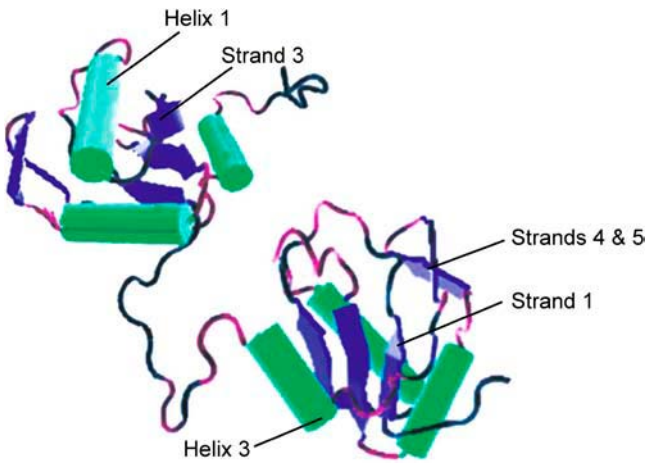

FIGURE 1 Model for the structure of folded DU1A*14. Helices (numbered 1–3) are shown in green; β-strands (numbered 1–5) are in blue.

Stopped-flow kinetics of aggregation and folding

Folding kinetics were measured using an Applied Photophysics SX.18MV system with 1:10 asymmetric mixing at 20°C. Proteins were situated in 6 M GuHCl solutions and mixed down to different low GuHCl concentrations. Folding was monitored by tryptophan fluorescence excited at 280 nm. Integrated fluorescence was collected at a 90° angle using a 320-nm cutoff filter (Schott WG 320). A minimal amount of excitation light was used to prevent any photobleaching during the measurement (slit width <1 mm). Data were collected using a logarithmically spaced time base, allowing for similarly weighted sampling of the slow and fast phases. Transients were fitted to double- or triple-exponential functions over the 10 ms to 25 s time range using the IGOR software package (WaveMetrics, Portland, OR).

Thermodynamic and kinetic model

The observed kinetics as a function of denaturant concentration minimally require a three-state model. Here we performed a detailed analysis using the general three-state model

$$\begin{array}{ccc} & \mathrm{U} & \\ \nearrow & & \searrow \\ \mathrm{A} & \leftrightarrow & \mathrm{N} \end{array} \qquad (1)$$

This model provides an average characterization of any aggregate ensembles that form during folding. It distinguishes between direct formation of the native state from U (i.e., individual domains fold by themselves) and indirect formation via the intramolecular domain aggregate(s) A. The model reveals how these contributions depend on n-mer size and denaturant concentration, the two parameters we tune.

Oliveberg and co-workers showed that U1A variants can form differentiated subensembles A_1 and A_2 (10), and our results support this finding for tethered proteins (see below). However, our data cannot yet distinguish nativelike/nativelike from nonnative/nonnative, mixed, or domain-swapped aggregate phases, and so we mainly restrict ourselves to a three-state analysis.

We used the simplest linear free energy model for the denaturant dependence. For each state i = U, A, or N,

$$\begin{aligned} G_\mathrm{i} &= G_{\mathrm{i}0} + G_{\mathrm{i}1}[GuHCl] \\ \Delta G^\dagger_{\mathrm{ij}} &= G^\dagger_{\mathrm{ij}} - (G_\mathrm{i} - G_\mathrm{j})/2 \\ k_{\mathrm{ij}} &= \nu_0 e^{-\Delta G^\dagger_{\mathrm{ij}}/\mathrm{RT}} \end{aligned} \qquad (2)$$

was used to calculate thermodynamic and activation free energies, with $G_{\mathrm{U}0} = G_{\mathrm{U}1} = 0$ making the unfolded state the reference state. All free energies are in units of kT_0, where T_0 is the temperature of the experiments (293 K). The simplicity of this model requires three caveats. Over the 4 M guanidine hydrochloride range most important for our experiments, the free energy can become nonlinear with denaturant concentration (13); barrier heights can shift by amounts differing from the average free energy difference of the surrounding states; and the prefactors ν_0 will not be exactly the same for the three transitions in Eq. 1. Analogous approximations are introduced in kinetic analyses of protein folding (for example, the prefactor is generally assumed to be independent of the reaction coordinate). More accurate free energies can be extracted by applying a detailed microscopic simulation model such as the one from Levy et al. (11). Fortunately, the approximations made in our kinetic model tend to cancel when we consider relative free energies of the different constructs, because the same approximations are made for all constructs.

The rate coefficients for the model in Eq. 1 are obtained by diagonalizing the rate matrix in the kinetic master equation

$$\begin{pmatrix}[\dot{A}]\\ [\dot{U}]\\ [\dot{N}]\end{pmatrix} = \begin{pmatrix}-k_{AU}-k_{AN} & k_{UA} & k_{NA}\\ k_{AU} & -k_{UA}-k_{UN} & k_{NU}\\ k_{AN} & k_{UN} & -k_{NU}-k_{NA}\end{pmatrix} \times \begin{pmatrix}[A]\\ [U]\\ [N]\end{pmatrix}, \quad (3)$$

yielding k_{fast}([GuHCl]) and k_{slow}([GuHCl]) as the two largest eigenvalues (the third eigenvalue being zero). For DU1A* and TU1A*, this produces a χ-shaped plot, as shown in Fig. 2. The two observed rate coefficients undergo an avoided crossing where the direct U-N folding rate (roughly equal to k_f of the monomer below the denaturation midpoint) drops below the aggregation rate (k_{agg}). Strictly speaking, the observed rates cannot be assigned to folding or aggregation near this avoided crossing because both contribute to each rate coefficient. At higher denaturant concentration, the slower rate coefficient then turns around and increases in normal chevron-plot fashion: the observed rate approaches the monomer unfolding rate above the denaturation transition midpoint. At the same time, the aggregation process (k_{agg}) becomes unobservably small in amplitude. Here, we are not interested in the usual chevron plot at high denaturant concentration, but in the χ-plot that occurs at lower denaturant concentration.

Tether and effective concentration modeling

In a concentration or sedimentation measurement, the number of monomers per aggregate (n) is allowed to vary freely, resulting in an equilibrium distribution of aggregate sizes.

In addition, it may be difficult to define precisely in a concentrated solution what the size of a given aggregate is. In our experiments, n is precisely controlled, and were it not for the tether, our free energies would be Gibbs free energies constrained so that all concentrations of aggregates above size n are zero.

The tether introduces additional largely entropic terms into the free energy. For a random coil tether, there is a contribution $\Delta G_T \approx \Delta G_{T0} - {}^3/_2 RT \ln[m_T]$ from the restriction of translational motion, thus creating an effective concentration of monomers dependent on tether length m_T. This term will be modified by corrections for persistence length dependent on tether length ($m_T < 10$) or amino acid composition (14–18). Stiff proline tethers would be an extreme example of amino acid composition effects, as demonstrated, for example, in single-molecule FRET studies (19), but even the alanine-leucine sequence in our DU1A*14 tether would likely alter the kinetics, although its effect on denaturation thermodynamics turns out to be small (see below).

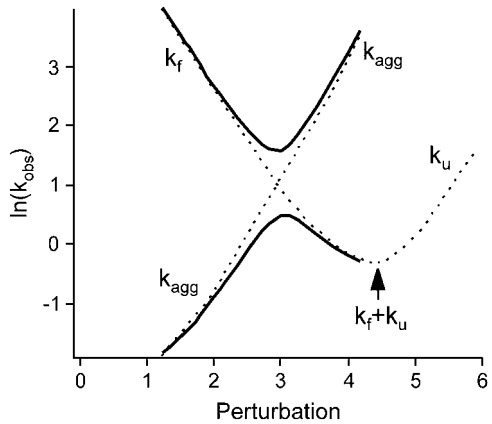

FIGURE 2 Behavior of the predicted observed rate coefficients (*solid lines*) and model rate coefficients (tuning according to Eq. 3, *dashed lines*) from the general three-state model. The observed rate coefficients of the three-state process undergo an avoided crossing with a characteristic χ-shape as k_{agg} (the aggregation rate from A to N in our experiment) overtakes the folding rate k_f. At higher perturbation (denaturant in our experiments), the observed rate switches from the folding rate to the unfolding rate in the conventional chevron plot manner.

Another important and largely entropic contribution involves the orientational freedom of the subunits, which is restricted by the tether, contributing a term $\Delta G_R \approx \Delta G_{Ri} - RT \ln[c_i m_P]^{\delta_i}$. In this equation, m_P is the length of the monomer sequence, c_i is a coefficient depending on the state (i = N, A, U), and δ_i is an exponent between 0 and ${}^2/_3$. As shown by Leitner and co-workers (20), the radius of a folded protein always increases faster with sequence length m_P than an exponent of 1/3 (0.4 is typical), and the orientational entropy scales maximally as R^2. These two effects combined produce the 0–${}^2/_3$ limit on δ. c_i decreases toward $1/m_P$ as the protein unfolds, because the tether has a minimal effect on restricting orientation of the unfolded protein, compared to restricting the orientation of the folded protein. Finally, ΔG_{Ri} has to take into account that the number of conformational microstates depends on the position of the aggregate interface with respect to the linker. This effect would be more important for aggregates with specific structure, as opposed to nonspecifically interacting monomers. Aggregate substates such as A_1 and A_2 discussed above would differ significantly in this parameter.

These free energy terms and smaller enthalpic corrections will require an off-lattice model of the protein-protein interaction, such as that described in Levy et al. (11), or a microscopic master equation model as in Muñoz et al. (21) for a complete description. Here we confine ourselves to the simplified kinetic analysis outlined above, illustrating what types of results such models have to account for. The effective concentration calculated below should be taken as a first approximation to a full model. What is important for the present purpose is that the effective concentration can be engineered to lie deeply within the limit ($\gg 1$ μM) where monomers strongly interact and transiently aggregate.

RESULTS

Monomeric U1A is a two-state folder

As in prior studies at pH 5.3 by Oliveberg and co-workers (5), the U1A* apparent folding rates at pH 7 and below 1 μM can be fitted by single exponentials. A chevron plot shows the typical ''V'' shape. The rate at pH 7, extrapolated to 0 M GuHCl, is very close to the pH 5.3 value reported in the literature (5) (Fig. 3).

Native states and denatured states of repeats are similar to the monomer

The CD spectra of the linked constructs are very similar to that of monomeric U1A* under native conditions (pH 7, 0 M GuHCl). Fig. 4 illustrates this for DU1A*10. The CD spectra of U1A* and linked constructs become random-coil-like at high denaturant concentration and low temperature (<30°C), but they increase in magnitude and change shape at high temperature. This effect has been observed previously for the U1A monomer. It was explained in terms of denatured structures that are random-coil-like on a long repeat scale (more than five residues), but look like extended structure (the type found in the β-sheet and polyproline II regions of the Ramachandran plot) on a short scale (≤5 residues) (22).

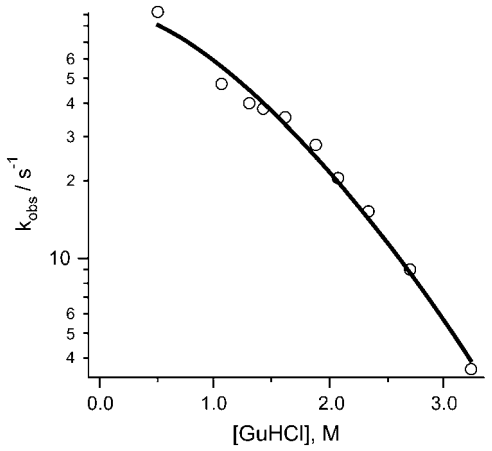

FIGURE 3 Relaxation rate of monomeric U1A* at 20°C, pH 7 in 50 mM potassium phosphate buffer, as a function of guanidine hydrochloride concentration. The rate data closely match those of Silow et al. (9).

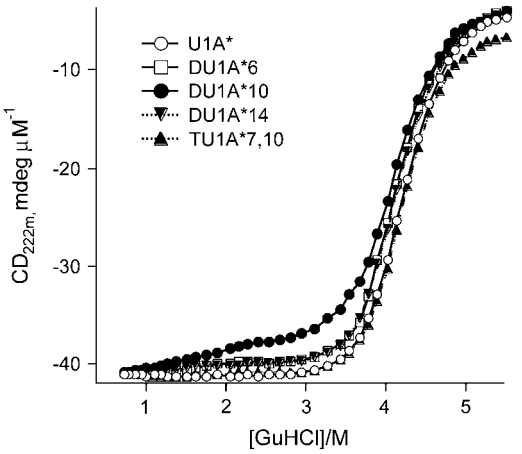

FIGURE 5 Guanidine hydrochloride denaturation curves of U1A*, U1A* dimers, and U1A* trimer (20°C, pH 7, 50 mM potassium phosphate buffer). The domain stabilities are essentially unaffected by the linkers, although DU1A*10 has a significantly larger native baseline. Concentrations are per monomer unit.

Tethered proteins have individually folded domains

If the domains were to strongly interact in thermodynamic equilibrium, one would expect a size-dependent change in cooperativity as the number of repeats increases. Instead, we find that sigmoidal GuHCl denaturation curves are retained with cooperativities and transition midpoints nearly identical to the monomeric U1A* (Fig. 5). Also, the total ellipticities are nearly multiples of the monomer value (Figs. 4 and 5). There are, however, subtle differences between the curves: DU1A*10 has a larger native baseline than DU1A*6 or 14; the unfolding transitions for all constructs except the trimer occur slightly lower than the monomeric protein. The magnitude of the denatured trimer CD spectrum is slightly larger than the monomer.

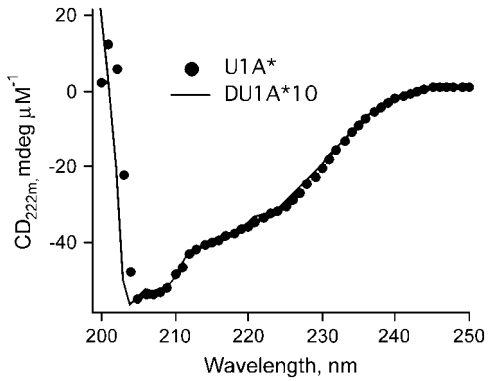

FIGURE 4 Smoothed CD spectra of U1A* and DU1A*10. Similar spectra are obtained for the other DU1A* and TU1A* variants (data not shown). The folded secondary structure of the linked proteins is not significantly perturbed by the linker. Concentrations are per monomer unit.

Kinetics are approximated by two phases

The *n*-mer kinetics contain two major characteristic timescales; double-exponential fits provide reasonably good empirical descriptions for these (but see below). The fast phase at low denaturant concentration of both DU1A* and TU1A* is nearly identical to the two-state folding phase of U1A* at $\leq$1-μM protein concentrations (Figs. 2, 6, and 7). Therefore we assign it to the individual folding of each domain to the native state. The slow phases observed for the *n*-mers do not exist in the monomeric U1A*, so we assign them to folding accompanied by interdomain interactions. As explained in the modeling section, these assignments are nominal because each phase contains contributions from N-A, A-U, and U-N kinetics in Eq. 1.

As expected from a chevron plot, the fast folding phase slows down when denaturant is added. At the same time, the slower interdomain aggregation phase speeds up. At $\approx$3 M guanidine hydrochloride in Figs. 6 and 7, the two phases cross and the ''folding phase'' becomes slower than the transient aggregation phase(s). This gives rise to the χ-shape illustrated in Fig. 2. (The chevron turnaround at $\approx$4.5 M GuHCl, where $k_f \approx k_u$ is not shown in Figs. 6 and 7.)

The three-state model was used to fit the kinetic data (bold curves in Figs. 6 and 7). The free energy parameters resulting from the fit are shown in Table 1, along with a correlation table (correlations near 1 mean that two parameters cannot be reliably determined independently from one another). All but G_{A0} and G_{A1} could be independently determined for TU1A*, but since measurements were performed down to 1 M GuHCl, G_{A0} is reasonably accurate ($\pm$1/2 *kT*). For DU1A*, G_{A0} and G_{A1} were also correlated, but in addition, the aggregate-state free energy at 0 M GuHCl (G_{N0}) was correlated with its activation energy. By fixing the native free energy at the value obtained from the GuHCl titrations (Fig. 5), very

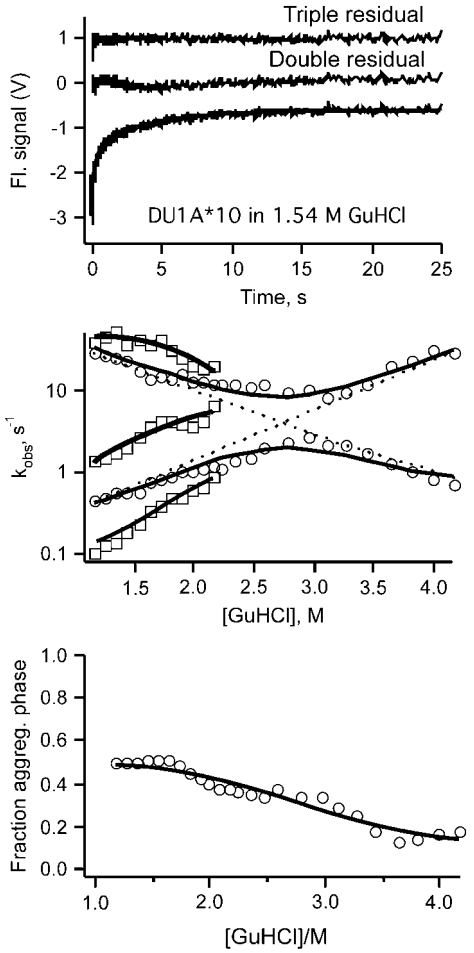

FIGURE 6 Kinetics of U1A* dimer DU1A*10. (*Top*) Representative relaxation with residuals. (*Middle*) Observed rate coefficients as a function of guanidine hydrochloride concentration from double- (*circles*) and triple-exponential (*squares*) fits. The double-exponential model curves (*solid lines*) are computed by fitting the general three-state model discussed in the text to the data (*circles*). The dashed lines follow the folding phase (descending) and aggregation phase (ascending) rate coefficients. (*Bottom*) Aggregation phase amplitude for the double-exponential fit; a triple exponential fit could account for two inflection points that may be present in the data.

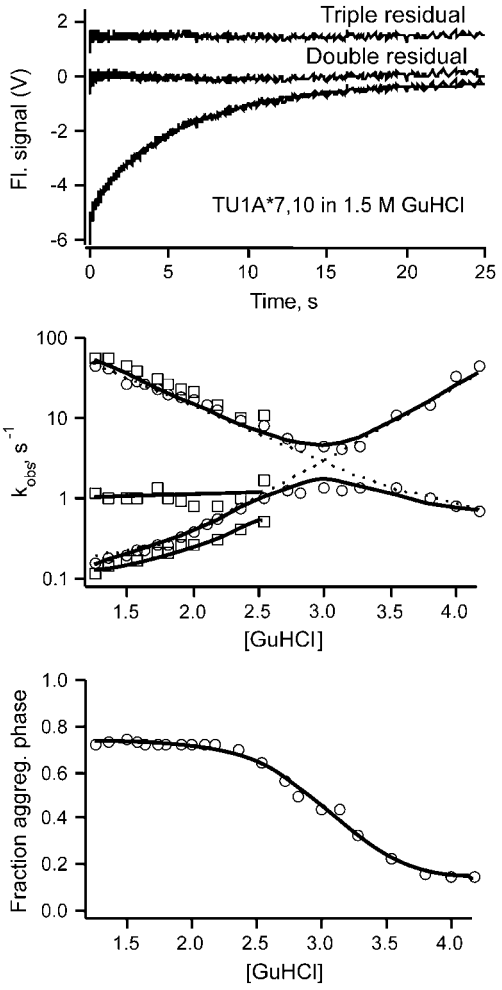

FIGURE 7 Kinetics of U1A* trimer TU1A*7,10. (*Top*) Representative relaxation with residuals. (*Middle*) Observed rate coefficients as a function of guanidine hydrochloride concentration. (*Bottom*) Aggregation phase amplitude for the double-exponential fit. All symbols and lines are analogous to those in Fig. 6.

similar results were obtained without the correlations. Thus, all parameters were determined with an accuracy of $\sim\pm 1/2$ *kT*. The resulting free energies extrapolated to 0 M GuHCl are shown in Fig. 8.

Kinetics of tethered U1As deviate from double exponentials at low GuHCl concentrations

Below 3 M GuHCl, the folding relaxations for the tethered U1As are not quite fitted within experimental uncertainty by double exponentials. Under these conditions we also fitted the relaxations to triple exponentials (Figs. 6 and 7). We again assign the fast phase to individual domain folding. The two slow phases are assigned to two or more distinct aggregation events. For DU1A*, both of the slower rate coefficients speed up as GuHCl is added. The middle rate coefficient approaches the fast folding rate coefficient near 3 M GuHCl, and ceases to make a distinct contribution. For TU1A*, the middle rate coefficient approaches the slowest rate coefficient near 3 M GuHCl, a qualitative difference between the dimer and trimer.

The slow phase disappears at high GuHCl concentrations

Although the aggregation phase(s) speed up relative to the folding phase when GuHCl is added, their amplitude decreases. The transition midpoint of the amplitude occurs near the avoided crossing of the refolding and aggregation rate coefficients (Figs. 6 and 7). The trimer has a slightly higher transition midpoint (3.0 M) than the dimer (2.8 M), as well as a markedly higher maximum percentage of relative slow-phase amplitude (0.75) than the dimer (0.5).

TABLE 1 DU1A*10 and TU1A*7,10 fitted free energy parameters (relative to the unfolded state), uncertainties (in parentheses), and correlation coefficients (lower half of the table)

	DU1A*10	TU1A*7,10
G_{A0} (kT_0)	−5.9 (0.10)	−11.1 (0.12)
G_{A1} (kT_0/M)	2.51 (0.04)	4.15 (0.05)
G_{N0} (kT_0)	−10.3 (0.16)	−12.4 (0.16)
G_{N1} (kT_0/M)	2.24 (0.05)	2.91 (0.07)
$G^{\dagger}_{AU}$ (kT_0)	12.51 (0.08)	13.34 (0.09)
$G^{\dagger}_{UN}$ (kT_0)	14.66 (0.08)	14.80 (0.08)
$G^{\dagger}_{NA}$ (kT_0)	17.5 (f)	17.45 (0.16)

(Lower triangle, DU1A*10; upper triangle, TU1A*7,10)

	$G_{A,0}$	$G_{A,1}$	$G_{N,0}$	$G_{N,1}$	$G^{\dagger}_{AU}$	$G^{\dagger}_{UN}$	$G^{\dagger}_{NA}$
G_{A0}	1	−0.96	−0.42	0.42	−0.45	0.25	0.86
G_{A1}	−0.97	1	0.23	−0.24	0.63	−0.14	−0.72
G_{N0}	−0.21	0.10	1	−0.97	−0.40	−0.76	−0.78
G_{N1}	0.20	−0.09	−0.93	1	0.36	0.67	0.74
$G^{\dagger}_{AU}$	−0.52	0.65	−0.36	0.34	1	0.20	−0.06
$G^{\dagger}_{UN}$	0.05	0.01	−0.73	0.54	0.21	1	0.52
$G^{\dagger}_{NA}$	−0.93	−0.86	−0.39	0.37	−0.40	0.15	1

Numbers in parentheses represent standard deviations.

DISCUSSION

Validity of the tethered construct concept

It has been shown by Oliveberg and co-workers that monomeric U1A* transiently aggregates during folding above 1 μM (5). To ensure extensive domain interactions and demonstrate the concept of tethered transient aggregation, we produced DU1A* and TU1A* constructs with short domain linkers. The effective concentration for two $r \approx 1.5$ nm diameter proteins tethered by a 10-residue linker (random coil length $d \approx 1.1$ nm) is approximately $^1/_6(r + d)^{-3}A^{-1} \approx 16$ mM. Here, A is Avogadro's constant, and the factor of $^1/_6$ assumes that a full shell around each protein would consist of six proteins. The tethered constructs are thus deeply in the limit where their domains interact, by a concentration factor of 10^4 or more with regard to the lower limit set by Oliveberg and co-workers for the monomer. To avoid complications arising from intermolecular interactions, folding of DU1A* and TU1A* was monitored at 0.2 μM construct concentration, limiting the interaction to exactly two and three monomers, respectively.

The secondary structure and thermodynamic stability of the dimers and trimers are very similar to those of the monomer, as judged by their CD spectrum and GuHCl titrations (Figs. 4 and 5). The individual domains fold correctly within these tethered constructs, so transient aggregation does not lead to serious misfolding of the individual domains. In particular, tether length and composition do not have a large effect on thermodynamics and structure for the cases we studied in Fig. 4.

Tether length and composition may not have a large effect on stability, but for a quantitative comparison with aggregation models, the tether entropy and effective concentration must be taken into account in a more sophisticated way than discussed above. For example, the free energies of different aggregate structures that can be formed by subunits may depend on tether length and stiffness.

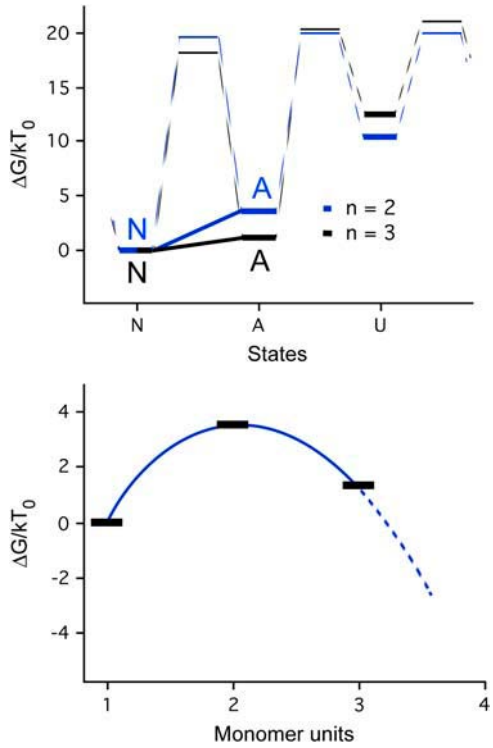

FIGURE 8 Free energies and activation barriers of DU1A*10 and TU1A*7,10 relative to the native state extrapolated to zero denaturant. (*Top*) The aggregate free energies are kinetic averages over the two or more aggregate substates that contribute to the kinetics in Figs. 6 and 7. (*Bottom*) Alternate view inspired by Ferrone (3) of the aggregate-state free energies $\Delta G(n)$ as a function of construct size. The $\Delta G(n)$ are referenced to their respective construct native free energies (for $n = 1$, there is of course no aggregate). As discussed in the text, extrapolation (*dotted line*) indicates that aggregates could become more stable than independently folded monomers within the constructs between $n = 3$ and 4.

Average aggregate size increases with the number of repeats

The rate coefficient of the slow phase increases with GuHCl concentration because GuHCl disrupts the interdomain oligomerization interface of DU1A* and TU1A* (Figs. 6 and 7). The slow rate coefficient has a much steeper [GuHCl] dependence for TU1A* than for DU1A*. This indicates that the transient domain aggregate of TU1A* on average buries a larger surface area compared to DU1A*. Indeed, Table 1 shows that the average amount of buried surface area for the transient domain aggregates formed (proportional to G_{A1}) is 1.65 times greater in TU1A* than in DU1A*, if we assume Tanford's linear free energy relationship (23). Note that this relative relationship holds even with the approximations made in our kinetic model (constant prefactors, linear free energy), inasmuch as they introduce similar errors for the two constructs.

Average aggregate free energy as a function of aggregate size

Formation of protein aggregates requires passage through a thermodynamic nucleation step (3), although the nucleus can be as small as $n = 1$ (no thermodynamic barrier). A concentration-dependent number of proteins must assemble until the addition of further monomers lowers the free energy, allowing the aggregate to propagate. The size is, however, difficult to measure: a quantitative measurement requires knowledge of the aggregate free energy as a function of size n. As we discuss below, at tether lengths of 7–10 glycines, the U1A* thermodynamic nucleus is very small ($n = 2$).

We experimentally characterized the dependence of the aggregate free energy on size at a specific effective concentration through the tethering of individual U1A* monomers. The number of repeats in the tethered construct allowed us to assess aggregates of fixed maximum size in an intramolecular fashion. For example, the DU1A* contains two domains, and can maximally only form dimeric intramolecular aggregates. Correspondingly, for TU1A*, up to three domains can participate in the intramolecular aggregate.

We utilized the folding kinetics of the tethered U1A* to measure the stability of intramolecular aggregates relative to the native state as a function of size. Below 3 M GuHCl, the rate coefficient and the denaturant dependence of the fastest phase indicate that it corresponds to monomeric folding. The slow phase therefore represents folding through the two- and three-domain transient aggregates of DU1A* and TU1A*, respectively.

By fitting the three-state model to the experimental data, we determined two fundamental parameters of the aggregation process, averaged over all aggregate substates. As shown in Fig. 8, the thermodynamic nucleus occurs at a size n_{TN}, where the aggregate free energy is maximized. Defining

$$\Delta\Delta G(n) = \Delta G_{AN}(n+1) - \Delta G_{AN}(n), \qquad (4)$$

where $\Delta G_{AN}(n)$ is the relative free energy between aggregate and native state. The aggregate of maximal free energy with respect to the native state ($\Delta\Delta G(1) = +4.4\ kT$) already occurs at $n_{TN} = 2$. This is the thermodynamic nucleus. Then aggregate stability relative to the native monomer increases by $\Delta\Delta G(2) = -3.1\ kT$ per monomer subunit added (Fig. 8 and Table 1). The aggregate state of TU1A* in Fig. 8 is nearly as stable as the native state per monomer unit. Extrapolating the free energy changes of U1A*, DU1A*, and TU1A* with a quadratic fit to the $n = 1,2,3$ data (3), stable aggregates are expected above 3.4 monomer units at the effective concentration (linker length) of our experiment.

Anticorrelation of the aggregation phase rate coefficient and amplitude

The aggregation phase speeds up when denaturant is added, indicating that GuHCI facilitates folding through interdomain aggregates. At the same time, the amplitude of the aggregation phase decreases. Eventually (>4 M GuHCl), direct folding/unfolding from U to N dominates the observed kinetics. Although denaturation accelerates folding through aggregates, the aggregates are at the same time more easily denatured than the monomeric native state, contributing a negligible amplitude by the time the N to U equilibrium constant reaches 1.

Multiple aggregate states

At low GuHCl concentration, tethered U1As exhibit increasing folding heterogeneity: the folding/aggregation relaxation deviates from the three-state model in Eq. 1. As a result, the aggregate free energies in Fig. 8 should be viewed as kinetic averages over potentially two or more subensembles of aggregates.

Strong nonnative interactions in the absence of denaturant enhance the free energy ruggedness within the broad unfolded conformational space (U + A), creating a distribution of distinguishable unfolded or aggregate states. All data could be fitted satisfactorily by triple-exponential functions, so we can subdivide the A state minimally into A_1 and A_2 substates. More substates cannot be excluded. This is consistent with work by Oliveberg and co-workers (10). We assign the label A_2 to the slowest phase observed.

A number of candidates exist for aggregate states: aggregates of native-like structures, aggregates of nonnative structures, domain-swapped aggregates, and combinations of these scenarios. These could give rise to a large number of phases underlying the observed heterogeneous aggregation kinetics. Although the current experiments do not provide direct structural information about the two (or more) aggregate states, they point toward a hypothesis worth further investigation.

With a triple exponential fit, the slowest rate in DU1A* takes over the domain-aggregation kinetics at high [GuHCl], and the middle rate merges with the fast folding rate. Thus the more rapidly interconverting A_1 state becomes very unstable at higher denaturant concentrations, whereas the slower interconverting A_2 state is more stable against denaturation. DU1A* may even have a double-sigmoidal amplitude (Fig. 6). In TU1A*, the middle rate remains at least as close to the slow rate as to the fast rate, so A_2 and A_1 are more comparable in stability.

A possible explanation compatible with all of these observations is in terms of a strongly interacting domain-swapped aggregate A_2, and a less strongly interacting aggregate A_1 (interaction of native-like and/or nonnative monomers within the constructs). In the dimer, the slowest A_2 state is more strongly bound than A_1, and persists at higher GuHCl. The trimer does not enhance the slowest phase any further at the expense of the medium phase: the third protein module does not further help the pairwise process of domain swapping, but does enhance the probability of forming a less spe-

cific aggregate A_1. Domain swapped aggregates such as A_2 have been implicated in Rop folding (24), and may turn out to be a common feature of transient aggregation of units within a repeat construct. It has also been identified by lattice models as a generic process in the phase diagram of transiently associating identical proteins, which is why we choose it as a likely explanation (25) for the most stable aggregate phase. Such models could be used to study the kinetics of the process in detail (currently, only the thermodynamics have been studied computationally), for direct comparison with the experimentally observed trends reported here. If such studies identify certain regions of the protein as critical, this would open the door to coupling experiments of the type reported here with mutant studies to identify the effect of specific side chains (26).

CONCLUSIONS

Tethering of protein monomers in conjunction with a study of their folding kinetics can be used to determine average aggregate stability as a function of aggregate size relative to the native state, revealing the elusive thermodynamic nucleus. Tether length and composition have no pronounced effect on the folding thermodynamics and secondary structure of the subunits. Tether length can thus be used in the future to investigate systematically the dependence of $\Delta\Delta G$ on the effective monomer concentration. Mutation studies can lead to further differentiation of the aggregate into structural subensembles. It is hoped that these experiments will stimulate the development of quantitative microscopic models: the proximity of monomers imposed by tethering is particularly useful for efficient implementation of computational models.

This work was supported by grant MCB 0316925 from the National Science Foundation.

REFERENCES

1. Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA.* 99:11884–11889.
2. Dobson, C. M. 2003. Protein folding and misfolding. *Nature.* 426: 884–890.
3. Ferrone, F. 1999. Analysis of protein aggregation kinetics. *Methods Enzymol.* 309:256–274.
4. Come, J. H., P. E. Fraser, and P. T. Lansbury. 1993. A kinetic model for amyloid formation in the prion diseases: importance of seeding. *Proc. Natl. Acad. Sci. USA.* 90:5959–5963.
5. Silow, M., and M. Oliveberg. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. USA.* 94:6084–6086.
6. Jacob, M., T. Schindler, J. Balbach, and F. X. Schmid. 1997. Diffusion control in an elementary protein folding reaction. *Proc. Natl. Acad. Sci. USA.* 94:5622–5627.
7. Yang, W. Y., and M. Gruebele. 2003. Folding at the speed limit. *Nature.* 423:193–197.
8. Lu, J., and K. B. Hall. 1997. Thermal unfolding of the N-terminal RNA binding domain of the human U1A protein studied by differential scanning calorimetry. *Biophys. Chem.* 64:111–119.
9. Silow, M., Y. Tan, A. R. Fersht, and M. Oliveberg. 1999. Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry.* 38:13006–13012.
10. Otzen, D. E., S. Miron, M. Akke, and M. Oliveberg. 2004. Transient aggregation and stable dimerization induced by introducing an Alzheimer sequence into a water-soluble protein. *Biochemistry.* 43:12964–12978.
11. Levy, Y., P. G. Wolynes, and J. N. Onuchic. 2004. Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. USA.* 101:511–516.
12. Nolan, S. J., J. C. Shiels, J. B. Tuite, K. L. Cecere, and A. M. Baranger. 1999. Recognition of an essential adenine at a protein-RNA interface: comparison of the contributions of hydrogen bonds and a stacking interaction. *J. Am. Chem. Soc.* 121:8951–8952.
13. Oliveberg, M. 1998. Alternative explanations for ''multistate'' kinetics in protein folding: transient aggregation and changing transition-state ensembles. *Acc. Chem. Res.* 31:765–772.
14. Krieger, F., B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber. 2003. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* 332:265–274.
15. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* 14:76–88.
16. Chang, I. J., J. C. Lee, J. R. Winkler, and H. B. Gray. 2003. The protein- folding speed limit: intrachain diffusion times set by electron-transfer rates in denatured Ru(NH3)(5)(His-33)-Zn-cytochrome *c*. *Proc. Natl. Acad. Sci. USA.* 100:3838–3840.
17. Camacho, C. J., and D. Thirumalai. 1995. From minimal models to real proteins: time scales for folding kinetics. *J. de Physique I.* 5:1457–1467.
18. Szabo, A., K. Schulten, and Z. Schulten. 1980. The first passage time approach to diffusion controlled reactions. *J. Chem. Phys.* 72:4350–4357.
19. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
20. Xu, X., and D. M. Leitner. 2003. Anomalous diffusion of vibrational energy in proteins. *J. Chem. Phys.* 119:12673–12679.
21. Muñoz, V., E. R. Henry, J. Hofrichter, and W. A. Eaton. 1998. A statistical mechanical model for β-hairpin kinetics. *Proc. Natl. Acad. Sci. USA.* 95:5872–5879.
22. Yang, W., E. Larios, and M. Gruebele. 2003. On the extended β-conformational propensity of polypeptides at high temperature. *J. Am. Chem. Soc.* 125:16220–16227.
23. Tanford, C. 1970. Protein denaturation. *Adv. Protein Chem.* 24:1–95.
24. Levy, Y., S. S. Cho, T. Shen, J. N. Onuchic, and P. G. Wolynes. 2005. Symmetry and frustration in protein energy landscapes: a near degeneracy resolves the Rop dimer-folding mystery. *Proc. Natl. Acad. Sci. USA.* 102:2373–2378.
25. Dima, R. I., and D. Thirumalai. 2002. Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics. *Protein Sci.* 11:1036–1049.
26. Gsponer, J., U. Haberthuer, and A. Caflisch. 2003. The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *Proc. Natl. Acad. Sci. USA.* 100:5154–5159.