Transient aggregates in protein folding are easily mistaken for folding intermediates

(U1A spliceosomal protein/aggregation artefacts/off-pathway intermediates)

MARIA SILOW AND MIKAEL OLIVEBERG*

Biochemistry, Chemical Centre, P.O. Box 124, S-221 00 Lund, Sweden

ABSTRACT It has been questioned recently whether populated intermediates are important for the protein folding process or are artefacts trapped in nonproductive pathways. We report here that the rapidly formed intermediate of the spliceosomal protein U1A is an off-pathway artefact caused by transient aggregation of denatured protein under native conditions. Transient aggregates are easily mistaken for structured monomers and could be a general problem in time-resolved folding studies.

In time-resolved refolding experiments, some unfolded proteins (**D**) collapse rapidly (<ms) into compact and partly structured denatured species prior to the formation of the native structure (**N**) (1–6). The nature of these folding intermediates[†] has attracted much attention because they are believed to reveal key information about the folding process. However, populated intermediates are not necessary for folding. Several small proteins (<90 residues) fold rapidly in a two-state process directly from the unfolded state (7–12). Two-state folding is seen also with larger proteins which may or may not accumulate intermediates depending on the conditions: intermediates that are populated under physiological conditions become destabilized and disappear in the presence of denaturant (3, 13) or at elevated temperatures (14). The findings have triggered a lively debate about the role of populated intermediates, and it has been argued that they could even be misfolds trapped in nonproductive pathways (15–19). Here we report that folding intermediates can also be artefacts of rapid aggregation of denatured protein.

Materials and Methods

The protein examined in this study is human spliceosomal protein U1A (20) (102 residues and no cysteines) in which the semiburied phenylalanine-56 was replaced with a tryptophan, a variant that occurs naturally in U1A from potato (21). The substitution produces large fluorescence changes upon denaturation and thereby facilitates time-resolved studies at very low protein concentrations.

Protein stability was obtained by standard linear free-energy assumptions and titration with guanidine·hydrochloride (Gdn·HCl) (22) using a Perkin–Elmer LS 50 luminescence spectrometer.

Unfolding and refolding kinetics were monitored using a SX.18MV stopped-flow instrument from Applied Physics set up for 1:10 volumes mixing. Excitation was at 280 nm and detection was with a 320-nm cut-off filter. All experiments were done at 25°C and the buffer was 50 mM Mes at pH 6.3.

The Folding Intermediate of U1A Is an Artefact Caused by Rapid (ms) Aggregation

A common way to classify folding behavior is to test if the experimental data obey the two-state relation, $N \rightleftarrows D$ (3, 7, 17),

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

where K_{D-N} is the equilibrium constant for unfolding, determined here by Gdn·HCl denaturation (22) (Table 1), and k_u and k_f are the rate constants from the unfolding and refolding kinetics, respectively. With U1A, the observed refolding kinetics, log k_f, deviates from the nearly linear Gdn·HCl dependence predicted by Eq. **1** at [Gdn·HCl] lower than 1.5 M (Fig. 1). The behavior is seen also with other proteins, where it has been suggested to result from a rapid collapse of the unfolded protein into an intermediate at low concentrations of denaturant (3, 23). Since this collapse occurs in the dead-time of the stopped-flow instrument, the observed reaction takes place from a stabilized form of the polypeptide which folds slower (3). Closer analysis of the U1A time course, however, reveals also a faster reaction preceding the predominant refolding phase (Fig. 2). Interestingly, the rate constant for this fast reaction (k_f^{fast}) agrees precisely with that expected for two-state folding according to Eq. **1** (Fig. 1). We conclude, therefore, that this fast reaction represents two-state folding directly from the denatured state.

Similar results were obtained recently with lysozyme, where a small population of unfolded protein folds fast in parallel with the slower conversions of various intermediates (17). The author points out that kinetic partitioning into fast and slow pathways (compare Fig. 2) is inconsistent with rapid interconversion between the unfolded state and other species in the pre-equilibrium, and he concludes that the intermediate, despite its rapid formation, is trapped in a nonproductive conformation that equilibrates slowly with the unfolded protein.

With U1A, the extent of fast refolding varies with protein concentration (Figs. 2 and 3). At low concentrations (<1 μM) U1A folds mainly by the fast pathway, but at higher concentrations the slow reaction becomes predominant. The behavior is typical of high-order reactions such as complexation and shows that the slow conversion into the native state takes place from aggregates which form in the dead-time of the stop-flow instrument (<5 ms). It remains to establish whether the fraction of monomer folding is determined by a rapid pre-equilibration of denatured monomers and aggregates or results from kinetic competition between monomer folding and ag-

Abbreviation: Gdn·HCl, guanidine·hydrochloride.
*To whom reprint requests should be addressed. e-mail: mikael.oliveberg@biokem.lu.se.
[†]The intermediates discussed in this study are those that accumulate transiently in the refolding reaction of small proteins without being restricted by nonnative disulfide links or nonnative proline isomerizations.

Biochemistry: Silow and Oliveberg

Proc. Natl. Acad. Sci. USA 94 (1997)   6085

Table 1. Comparison of stability and kinetic data shows that the fast folding reaction of U1A takes place directly from the denatured state.

Measurement	$\Delta G_{D-N}^{H_2O}$, kcal/mol	Gdn·HCl$_{1/2}$, M	m_{D-N}
Equilibrium unfolding	9.3 ± 0.2*	4.07 ± 0.02*	-2.3 ± 0.1*
Kinetics	9.1 ± 0.2†	4.08§	-2.2 ± 0.1¶

As expected from a two-state process, the free energy of unfolding in pure water ($\Delta G_{D-N}^{H_2O}$) derived from the kinetics is the same as from equilibrium denaturation experiments (Eq. **1**). Gdn·HCl$_{1/2}$ is the midpoint for the unfolding transition and m_{D-N} is the Gdn·HCl dependence of ΔG_{D-N} (22).

*The value is determined by standard Gdn·HCl denaturation experiments (22).

†$\Delta G_{D-N}^{H_2O} = -2.3RT\log K_{D-N} = -2.3RT(\log k_u - \log k_f^{fast})$ at [Gdn·HCl] = 0 M, Eq. **1**. The derivation does not take into account the *cis–trans* equilibria of the prolines in the denatured state, which contribute to a small underestimate of $\Delta G_{D-N}^{H_2O}$ (7). The difference is within the experimental error.

§Obtained from the intersect between the fits in Fig. 1, where $\log k_f = \log k_u$.

¶From Eq. **1** it follows that $m_{D-N} = m_{\ddagger-N} - m_{D-\ddagger}$ (7), where $m_{\ddagger-N}$ and $m_{D-\ddagger}$ are derived from the slopes of the polynomial fits to $\log k_u$ and $\log k_f$, respectively (Fig. 1) and $\ddagger$ represents the transition state.

gregation. The kinetic partitioning favors the latter explanation. If poor solubility under native conditions is a general property of denatured proteins (24), the competition scenario predicts that slowly folding proteins will fold with substantial formation of transient aggregates, whereas more rapidly folding proteins will tend to be two-state. At the Gdn·HCl concentration where aggregation occurs, $\log k_f$ vs. [Gdn·HCl] bends down and displays a positive slope (Fig. 1). From mass-action, this indicates that the protein expands upon
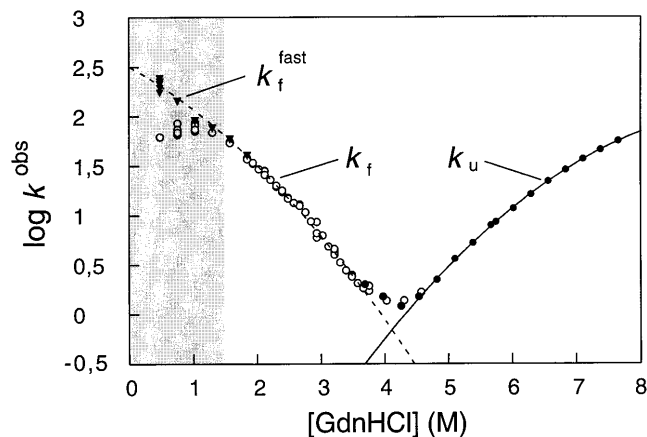

FIG. 1. Gdn·HCl dependence of the rate constants for folding and unfolding of U1A. The rate constants are in units of s^{-1}. The left arm of the V-shaped plot shows the refolding rate constant ($\bigcirc$, k_f, [U1A] = 3.1 μM; and $\blacktriangledown$, k_f^{fast}, [U1A] = 1 μM) following 1:10 dilution (stopped-flow) of denatured U1A (in 5.1 M Gdn·HCl) into lower [Gdn·HCl], and the right arm shows the unfolding rate constant ($\bullet$, k_u, [U1A] = 3.1 μM) upon 1:10 mixing of native protein (in water) into high [Gdn·HCl]. The curves are polynomial fits which precisely obey Eq. **1** and, hence, represent two-state folding directly from the denatured state—i.e., $\log k_f = \log k_u - \log K_{D-N}$. The deviation from two-state folding observed at low [Gdn·HCl] ($\bigcirc$) is found also for other proteins and is usually believed to result from accumulation of an intermediate. With U1A, the deviation is caused by transient aggregation of denatured protein under refolding conditions. At low protein concentrations the denatured protein remains monomeric during the refolding process and the rate constant (k_f^{fast}) follows Eq. **1**, but at higher protein concentrations the denatured protein aggregates in the dead-time of the stopped-flow instrument, giving rise to a retardation of the refolding rate. Conditions where aggregation occurs are marked gray.
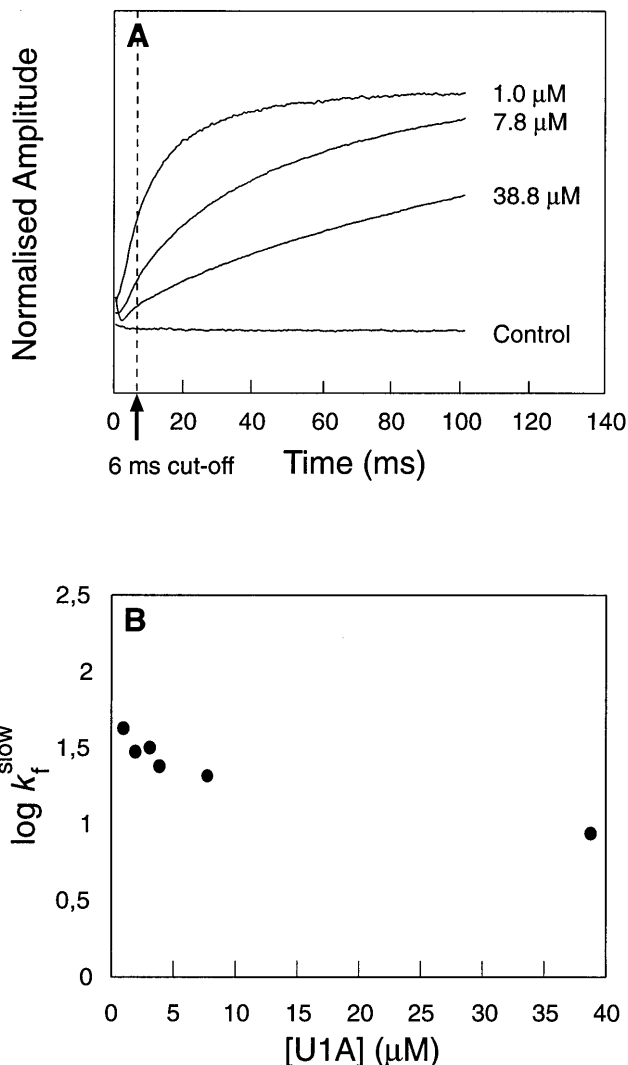

FIG. 2. (*A*) Time course for refolding of U1A at different protein concentrations. Final [Gdn·HCl] = 0.46 M. At moderate to high protein concentrations (>5 μM), the time course is dominated by the slow phase, but at low protein concentrations folding occurs mainly by the fast reaction. (*B*) The rate constant of the slow phase decreases slightly with increasing protein concentration, whereas the fast reaction appears independent of protein concentration. The negative concentration dependence of the slow phase is inconsistent with formation of aggregates, since this process would become faster at high protein concentrations. Hence, it is likely that the slow phase represents a dissociation process—i.e., folding from an aggregate. Data from the first 6 ms were excluded from the fits. Control experiments were conducted with free tryptophan and with U1A contained in the dilution buffer.

activation—i.e. the transition state exhibits more Gdn·HCl-binding sites than the dead-time species. This is another clear indication of the dead-time species being off-pathway, since it has to unfold (or dissociate) before it adopts the native conformation.

Accordingly, U1A may fold either rapidly in a two-state process (**D → N**) or more slowly from aggregates, depending on solvent conditions and protein concentration. Since the aggregation artefact is revealed only at very low protein concentrations, it may well have been missed in earlier rapid-mixing studies done at much higher protein concentrations (compare Figs. 2 and 3).

The curved plots of $\log k_f$ and $\log k_u$ under two-state conditions suggest that U1A features a nonlinear Gdn·HCl dependence of the activation energy. The behavior has been
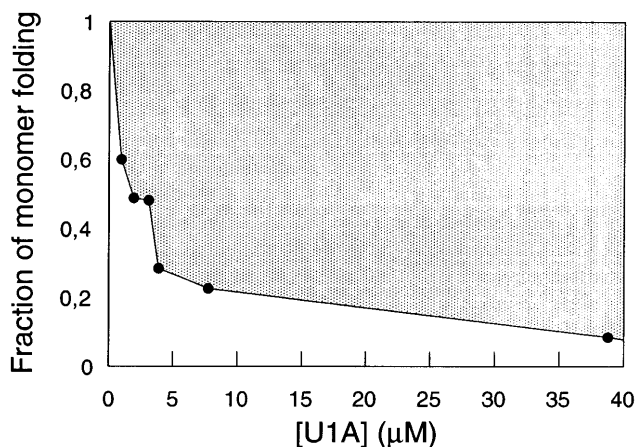
FIG. 3.   Fraction of monomer folding at different concentrations of U1A, expressed as the ratio of the amplitudes of the fast and slow refolding phase (compare Fig. 2*A*). In fits where [U1A] $\geq$ 3.1 μM the rate constant for the fast phase was locked to 200 s^{-1}. Since refolding is usually monitored at relatively high concentrations of protein, the proportion of monomer folding may be very small and undetected. For example, standard stopped-flow ($\approx$10 μM), stopped-flow CD (10–50 μM), and quench-flow NMR (>100 μM). Hence, tests of concentration dependence in these regions may not reveal aggregation artefacts.

examined in detail and is suggested to be caused by movements of the transition state along the top of a very broad and flat activation barrier upon destabilization by Gdn·HCl (unpublished results).

Could Concealed Aggregation Be a General Problem in Folding Studies?

Dead-time aggregation is found also with the small and rapidly folding chymotrypsin inhibitor 2 (CI2), but only at protein concentrations above $\approx$200 μM (compared with $\approx$1 μM for U1A) (M.O., Y.-J. Tan, and A. R. Fersht, unpublished results). It must be noticed, however, that previous results with CI2 were obtained at protein concentrations well below 200 μM, where aggregation can be excluded. We cannot detect any aggregation during folding of the ribonuclease barnase. The difference may be that barnase collapses rapidly into a folding intermediate (3) which is more soluble than the unfolded protein. Aggregation is observed also during refolding of larger proteins—e.g., bovine growth hormone (25), reduced lysozyme (26), and phosphoglycerate kinase (27). Since these proteins fold much more slowly than U1A there is ample of time for aggregation to occur, and the implications of the results for rapidly folding proteins are uncertain. It is now important to find out if artefacts of transient aggregation are a widespread problem in time-resolved folding studies.

1. Udgaonkar, J. B. & Baldwin R. L. (1988) *Nature (London)* **335,** 700–704.
2. Roder, H., Elöve, G. A. & Englander S. W. (1988) *Nature (London)* **335,** 694–699.
3. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M. & Fersht, A. R. (1990) *Nature (London)* **346,** 440–445.
4. Kim, P. S. & Baldwin R. L. (1990) *Annu. Rev. Biochem.* **59,** 631–660.
5. Radford, S. E, Dobson, C. M. & Evans, P. A. (1992) *Nature (London)* **358,** 302–307.
6. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995) *Science* **269,** 192–197.
7. Jackson, S. E. & Fersht A. R. (1991) *Biochemistry* **30,** 10428–10435.
8. Alexander, P., Orban, J. & Bryan, P. (1992) *Biochemistry* **31,** 7243–7248.
9. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L. & Serrano, L. (1994) *Biochemistry* **33,** 2142–2150.
10. Kuszewski, J., Clore, G. M. & Gronenborn, A. M. (1994) *Protein Sci.* **3,** 1945–1952.
11. Kragelund, B. B., Robinson, C. V., Knudsen, V., Dobson, C. M. & Poulsen, F. M. (1995) *Biochemistry* **34,** 7217–7224.
12. Schindler, T. S., Herrier, M., Marahiel, M. A. & Schmid, F. X. (1995) *Nat. Struct. Biol.* **2,** 663–673.
13. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993) *Biochemistry* **32,** 7054–7063.
14. Oliveberg, M., Tan, Y. J. & Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8926–8929.
15. Shakhnovich, E. I. & Finkelstein A. V. (1989) *Biopolymers* **28,** 1667–1680.
16. Chan, H. S. & Dill K. A. (1994) *J. Chem. Phys.* **100,** 9238–9275.
17. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9029–9033.
18. Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 10869–10873.
19. Mirny, L. A., Abkevich, V. & Shakhnovich, E. I. (1996) *Folding and Design* **1,** 103–116.
20. Nagai, K., Oubridge, C., Jessen, T.-H., Li, J. & Evans, P. R. (1990) *Nature (London)* **348,** 515–520.
21. Simpson, G. G., Clark, G. P., Rothnie, H. M., Boelens, W., van Venrooij, W. & Brown, J. W. S. (1995) *EMBO J.* **14,** 4540–4550.
22. Pace, C. N. (1986) *Methods Enzymol.* **131,** 266–279.
23. Muñoz, V., Lopez, E. M., Jager, M. & Serrano, L.(1994) *Biochemistry* **33,** 5858–5866.
24. Dill, K. A. & Shortle, D. (1991) *Annu. Rev. Biochem.* **60,** 795–825.
25. Brems, D. N. (1988) *Biochemistry* **27,** 4541–4546.
26. Goldberg, M. E., Rudolph, R. & Jaenicke, R. (1991) *Biochemistry* **30,** 2790–2797.
27. Pecorari, F., Minard, P., Desmadril, M. & Yon, J. M. (1996) *J. Biol. Chem.* **271,** 5270–5276.