Formation of a Compact Structured Ensemble without Fluorescence Signature Early during Ubiquitin Folding

Z. Qin,[†,‡] J. Ervin,[§,¶] E. Larios,[⊥] M. Gruebele,[*,§,⊥,∥] and H. Kihara[*,‡]

Department of Physics, Kansai Medical University, 18−89 Uyama-Higashi, Hirakata 573-1136, Japan, and Departments of Chemistry, Physics, and Biophysics and Computational Biology, University of Illinois at Urbana−Champaign, Illinois 61801

Received: June 14, 2002; In Final Form: September 13, 2002

The fluorescent Ub* mutant of the small globular protein ubiquitin is studied by multiple spectroscopic techniques under low temperature/high viscosity conditions to reveal unambiguously the formation of a compact structured intermediate preceding the barrier crossing to the native state. When detected by fluorescence intensity alone, ubiquitin appears to fold via a quasi two-state mechanism. In contrast, circular dichroism reveals the early formation of a structured ensemble, whose signal between 210 and 230 nm exceeds that of the native state, and whose unfolding curve is cooperative. The ensemble radius of gyration determined by small-angle X-ray scattering is only 1.2 times that of the native state, close to the classical molten globule value. Experiments performed in water/ethylene glycol mixtures at temperatures as low as −20 °C allow us to set an upper limit of $\leq 8k_BT$ on the barrier height for formation of the intermediate ensemble. The very small fluorescence burst phase indicates that the tryptophan residue is not yet packed into a nativelike conformation in the intermediate ensemble and that kinetics monitored by fluorescence intensity alone are not always a reliable indicator of two-state folding.

Introduction

Small globular proteins often fold over a single barrier, but evidence also has been uncovered for intermediate states G separated by barriers from both the native and unfolded states. In some cases, G has a high free energy, and subtle deviations from apparent two-state behavior occur.[1] In other examples, the evidence points to a large transient intermediate population. For example, energy transfer (FRET) measurements between the heme group and tryptophan in cytochrome *c* clearly show that the distance between these two groups rapidly shrinks before the barrier to the native state is crossed.[2−4] Small-angle X-ray scattering confirms a rapid collapse.[5] Circular dichroism (CD) measurements of secondary structure content reveal a small, fast "burst phase" in cytochrome *c*.[6,7] Although fast, the initial collapse is compatible with a viscosity-corrected activation energy of $\approx 7k_BT$ ($\approx$17 kJ/mol, assuming a 0.02 ns^{-1} prefactor), making the populated intermediate a thermodynamic state.[4] Stopped-flow folding kinetics of a fluorescent variant of ubiquitin monitored by fluorescence intensity have shown both two-state and three-state behavior.[8,9] For slower-folding mutants, no evidence for an intermediate is found. For faster-folding mutants, evidence for an unresolved ("burst phase") intermediate was found. Again, the intermediate is assigned to a thermodynamic state, separated by a barrier from the unfolded state. X-ray scattering studies of lysozyme have revealed "burst phase" formation of a collapsed state, followed by slower formation of the native state via branched kinetics.[10] In this case, the burst phase intermediate is interpreted as an unstructured collapsed polypeptide.

The central question is as follows: Are these intermediates G local minima on the free energy surface, having significant compactness, and secondary structure, or are they merely contracted polypeptide chains resulting from changes in solvent conditions, but without significant structure or barriers separating them from the unfolded state? The cytochrome *c* and ubiquitin results discussed above have also been interpreted in terms of an unstructured, barrier-free collapse.[11−13] For example, the small CD burst phase of cytochrome *c* does not differ from the CD burst phase of a mixture of protein fragments too small to fold individually and can be interpreted as an extrapolation of the unfolded baseline.[12] The larger burst phase of FRET experiments may be interpreted as aggregation of the large aromatic tryptophan and heme moieties, instead of a fully compact structured state, while X-ray scattering only measures overall compactness and not secondary structure content. The folding kinetics of ubiquitin have been reinvestigated by stopped- and continuous-flow kinetics, using fluorescence intensity detection.[14] In that work, extrapolation of kinetics with different initial denaturant concentrations and dead times, as well as extrapolation of the main folding phase, are compatible with the absence of a burst phase. Although these results cannot conclusively rule out a structured compact intermediate, they raise a note of caution.

Two issues are primarily responsible for the difficulties in resolving our central question. First, the time resolution of stopped-flow experiments, and sometimes even of continuous-flow experiments, limits our ability to distinguish burst phases from the main folding phase when the wild-type or mutant folds very rapidly. Second, X-ray and FRET experiments report only

* Corresponding authors.
† Current address: National Laboratory of Biomacromolecules, Institute of Biophysics, Academica Sinica, Beijing 100101, China.
‡ Kansai Medical University.
§ Department of Chemistry, University of Illinois at Urbana−Champaign.
¶ Current address: UltraPhotonics, 48611 Warm Springs Blvd., Fremont, CA 94539.
⊥ Department of Physics, University of Illinois at Urbana−Champaign.
∥ Department of Biophysics and Computational Biology, University of Illinois at Urbana−Champaign.

10.1021/jp021433m CCC: $22.00
Published on Web 11/22/2002

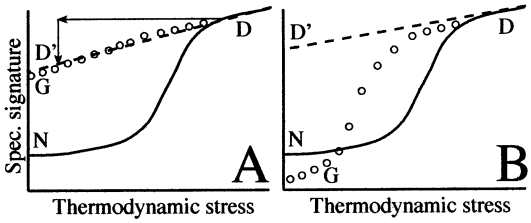

Figure 1. Signal changes during (un)folding as a function of thermodynamic stress (e.g., T or denaturant concentration): N = native, D = denatured, and G = globule state. Key: solid curve, steady-state signal; dashed curve, extrapolated unfolded signal D′; circles, potential structured compact non-native ensemble or "globule". A: Spectroscopic technique insensitive to the difference between D and G. The horizontal arrow indicates initiation of refolding by stopped-flow, the vertical arrow the small burst phase. A difference between D′ and G would be hard to detect. B: Spectroscopic technique sensitive to the difference between D and G. In the ideal case, G's spectroscopic signal would lie outside the D′−N range, and its denaturation curve would be highly nonlinear.

on overall or pairwise compactness, and tryptophan fluorescence intensity is sensitive only to the final stages of solvent exclusion and packing around the tryptophan side chain.[15,16]

Two changes in approach can solve these problems (albeit introducing one of their own, a change in solvent conditions). First, the use of a high viscosity solvent at lower temperature slows down the main folding phase, increasing the dynamic range to the burst phase. The burst amplitude can thus be determined without extrapolation. Second, simultaneous application of a technique such as CD or IR can provide secondary structural information lacking in X-ray, FRET, or fluorescence intensity measurements. A "best case scenario" is illustrated in Figure 1: although a technique such as fluorescence intensity may yield a fairly small burst phase (D → D′ in Figure 1A), another technique may actually provide a cooperative overshoot beyond the native state, which cannot be interpreted as a simple extrapolation. Such an overshoot has been observed during the folding of the large β-sheet protein β-lactoglobulin.[17−19]

Here, we study the fluorescent tryptophan-containing mutant Ub* of the small globular protein ubiquitin.[8,9,20] We use water−ethylene glycol mixtures at low temperature to increase the dynamic range between burst phase and main folding phase. We also study the folding by three techniques: fluorescence intensity (solvent exposure and quenching due to native packing), CD (secondary structure content) and small-angle X-ray scattering (radius of gyration). Our experimental results show unambiguously that during the burst phase, Ub* forms a state with near-native compactness and significant secondary structure, which we can distinguish from the main folding phase without extrapolation. The CD signal of the intermediate actually exceeds that of the native state, as in the "best case" scenario in Figure 1B. The fluorescence intensity of the tryptophan residue (monitored from 325 to 400 nm) does not change significantly in the burst phase, indicating that the tryptophan remains solvent exposed. An upper limit of $8k_BT$ is set on the unfolded-intermediate barrier. The intermediate thus satisfies the criteria for a molten globule (compact state with secondary structure, but lacking nativelike tertiary packing). Although our result applies only under low-temperature high viscosity solvent conditions, it indicates that Ub* is at best an apparent two-state folder even at higher temperature and in aqueous solution.

Materials and Methods

There are two general approaches to measuring speed limits and structural content for sub-millisecond burst phases. One can initiate and detect rapid refolding by methods such as cold denaturation/*T*-jump or continuous mixing.[21,22] Alternatively, one can slow the kinetics by low temperature/high viscosity.[17] Our previous observation of fast ubiquitin *T*-jump refolding at 8 °C required destabilizing mutants, whose low stability led to apparent two-state folding.[23] Here we observe the refolding kinetics of the more stable Ub* mutant over a wide range of temperatures (−20 to +4 °C) and viscosities (25−45% aqueous ethylene glycol solutions).

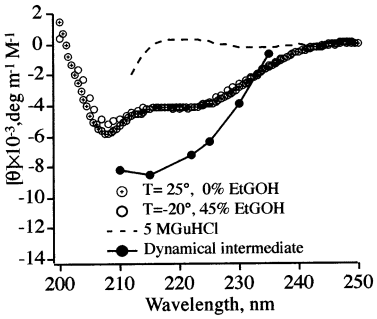

Figure 2. Ub* circular dichroism spectra under different solvent conditions. The native spectra are similar to one another over the whole −20 to +25 °C/0−45% EtGOH range, compared to the GuHCl denatured state and the burst phase intermediate. The latter has a larger magnitude of the CD spectrum than even the native state. GuHCl and intermediate data were acquired at −20 °C.

Ubiquitin. Ubiquitin contains five β-strands and nine turns, in addition to α-helical and short 3_{10} helical segments.[24] A Phe45Trp mutation (abbreviated as Ub*) has been introduced to allow fluorescence detection. Previous kinetic studies and an NMR structure refinement show that this mutation has only minor effects on the equilibrium and nonequilibrium behavior of the protein.[8,9,20] The Ub* mutant has a radius of gyration of 11.7 Å.[20] Ub* was produced by site-directed mutagenesis, starting with a wild-type plasmid provided by Tracy Handel.[25] The protein was overexpressed in *Escherichia coli* (BL21), and the soluble fraction was purified using a method similar to that given in the literature.[25] The lyophilized protein was verified to be Ub* by low resolution electrospray mass spectrometry and SDS−PAGE.

Solution Conditions. Ubiquitin folding was investigated at various temperatures and solvent compositions ranging from −20 to +25 °C and from 0% to 45% ethylene glycol (EtGOH), allowing us to access a wide range of solvent viscosities. These conditions will be abbreviated *X*°/*Y*% henceforth. Unless otherwise indicated, aqueous solvents were buffered with 40 mM, pH 5.9 phosphate. Guanidinium hydrochloride (GuHCl) was used in all denaturant measurements. As shown in Figure 2, native Ub* circular dichroism spectra are similar under all solvent conditions used, when compared to the GuHCl unfolded state or the rapidly formed intermediate ensemble described later. Moreover, −20°/45% and 25°/0% solutions have identical free energies of unfolding, connected by an isostability curve at intermediate temperatures and EtGOH concentrations (Figure 2, Table 1, and results section).

Unfolding Titrations. Unfolding titrations were made in the 0−5.5 M GuHCl range at various temperatures and EtGOH concentrations. Three probes were used to monitor the equilibrium configuration of Ub*. Circular dichroism spectra from 205 to 250 nm were taken with a spectropolarimeter (UNISOKU) for each sample, at 100 μM protein concentration, and backgrounds of protein-free solution were subtracted. The integrated fluorescence intensity was collected by a spectrophotometer (USP-539 UNISOKU) using a filtered 40 μM protein solution to reduce scatter. Fluorescence was excited at 295 nm to probe

TABLE 1: Thermodynamic Parameters Derived from Linear Free Energy Fits[a]

conditions	m, kJ mol^{-1} M^{-1}	C_m, M	R_g, Å	η, cP
25°/0%	8.5(1.0)	3.8(0.1)	11.7[b]	1
4°/45%	8.5(3.0)	3.1(0.2)	13.2(0.1)	5
−10°/25%	7.1(1.0)	3.3(0.1)		5
−20°/45%	8.5(2.0)	3.8(0.1)	14.5(0.1)	13
G, −20°/45%	4.3(2.0)	2.0(0.1)	17(2)	13
D, −20°/45%			26.5(0.1)	13

[a] Sample fits are shown in Figure 3. Except for G and D, the radius of gyration is for the native state at 0 M GuHCl. Uncertainties shown are two standard deviations rounded up to the nearest digit. [b] From the solution NMR structure; does not include contributions from solvent shells.

Trp only, and integrated at wavelengths above 325 nm. For steady-state X-ray scattering experiments, the concentration was 800 μM, and backgrounds from protein-free buffer solutions were subtracted (see below).

Refolding Kinetics. The same three probes were used to monitor refolding kinetics: circular dichroism from 210 to 255 nm, the integrated fluorescence intensity excited at 295 nm and collected above 325 nm, and small-angle X-ray scattering. All kinetic transients were measured with a stopped-flow apparatus with a maximum 6 ms dead time at the highest viscosities.[17,26] In all cases a protein solution containing 5 M GuHCl was diluted with a similarly buffered solution of identical temperature and EtGOH concentration. The final concentration for the circular dichroism measurements ranged from 34 to 130 μM to test for aggregation effects. For the fluorescence measurements, it was 80 μM. The X-ray scattering measurements reported here were taken at beamline 15A1 of the Photon Factory at KEK in Tsukuba, Japan.[27] The X-ray wavelength was $\lambda = 1.50$ Å with either the position-sensitive proportional counter for kinetics[27] or the newly installed CCD detector for steady-state measurements.[28] The final protein concentration was 800 μM for kinetic measurements. X-ray scattering was analyzed using the Guinier approximation,[29] assuming an exponential dependence of the scattering intensity on h^2 where $h = 4\pi(\sin\theta)/\lambda$, and θ is half the scattering angle. The CCD data were noise-subtracted and cylindrically averaged.[28,30] A nonlinear least-squares fit then yielded the radius of gyration (R_g) of the molecule.

Results

Steady-State Data. CD spectra were collected for ethylene glycol concentrations between 0 and 50% (15% intervals) and temperatures of +25 to −20 °C (5−10 °C intervals). They vary comparatively little over the entire solvent/temperature range, as illustrated by the +25°/0% and −20°/45% extremes shown in Figure 2. "Comparatively little" here refers to the strongly contrasting CD spectra of denatured protein in 5 M GuHCl, and of the kinetic intermediate ensemble discussed later, both also shown in Figure 2.

GuHCl titration reveals that the thermodynamic transition is well approximated by a two-state transition in two respects. First, all GuHCl titrations of Ub* from 0 to 5.5 M, in the −20°/45% to +25°/0% range, can be fitted individually by two-state models (Table 1). Figure 3A illustrates this for the two extremes, +25°/0%, and −20°/45%. Also, all unfolding titrations monitored by several different techniques under identical solvent conditions can be fitted simultaneously with the same midpoints and m values (Table 1). Figure 3B compares the 25°/0% scaled CD and fluorescence titrations. Values identical within experimental error are also obtained by CD, fluorescence and small-angle X-ray scattering at −20°/45%, by fluorescence and small-angle

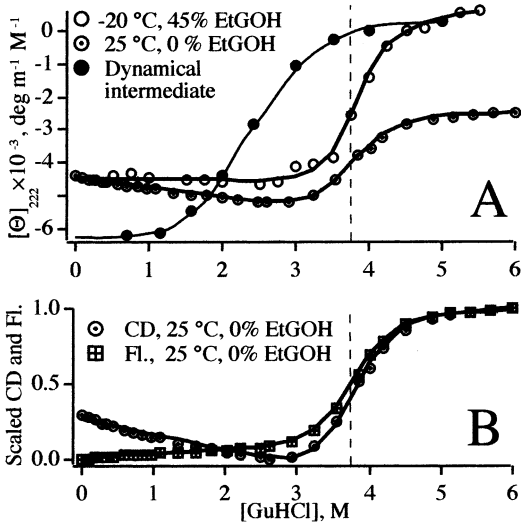

Figure 3. GuHCl unfolding titrations. Fits (solid lines) are two-state models with signals $\theta_i = \theta_i^{(0)} + \theta_i^{(1)}C$ for the denatured (i = D), native (i = N) and globule (i = G) states (C = denaturant concentration, $\theta^{(1)}$ = baseline slope). Using these, $\theta_{obs} = \theta_i[i] + \theta_j[j]$ and $\Delta G_{ij} = m(C - C_m)$, $K_{ij} = e^{-\Delta G_{ij}/kT} = [i]/[j]$, the data were fitted to the denaturation transition midpoint C_m, and to m, the sensitivity of the free energy to GuHCl concentration (Table 1). A: CD data for D→N at 25 °C/0% EtGOH (⊕) and D → N at −20 °C/45% EtGOH (o); the vertical dashed line and equal m values indicate that these two conditions have identical folding free energies. G → N at −20 °C/45% EtGOH was obtained from the burst phase CD data. G → N has a lower transition midpoint, but a clear S-shaped denaturation curve. B: Comparison of the CD and fluorescence data at 25 °C/0% EtGOH shows the same C_m and m values by different techniques, verifying two-state folding. The same is true of CD, fluorescence, and X-ray data at −20 °C/45% EtGOH (not shown).

X-ray scattering at +4°/45%, and by CD and fluorescence at −10°/25%. This agreement among multiple spectroscopic techniques over the whole range of conditions is a strong indication of cooperative two-state unfolding, at least near the transition midpoint ($C_m \pm 1$ M) of the steady-state unfolding titrations.

GuHCl titrations also reveal that tuning to lower temperatures, while simultaneously increasing EtGOH concentration, keeps ubiquitin on an isostable free energy curve given to linear order by the constraint [EtGOH, %] $= -1.0(T - T_0)$, where $T_0 = 25$ °C. As expected from this relation the −10°/25% and +4°/45% points lie slightly lower in stability than either extreme at +25°/0% and −20°/45% (Table 1). Although one would certainly not expect all details of the free energy surfaces at +25°/0% and −20°/45% to be identical, the average native bias (folded to unfolded free energy difference) is the same under these two conditions.

The native radii of gyration (0 M GuHCl) were found to lie between 13.2 and 14.5 Å by SAXS (Table 1). These values are about 2 Å larger than the previously reported literature values based on bare X-ray or NMR structures.[24] This is commonly observed in aqueous solution X-ray scattering, and probably due to solvent shell contributions to R_g. The 26.5 Å radius of gyration we observe in 5 M GuHCl is slightly smaller than the random coil value of 30 Å derived using literature procedures,.[31,32] The 5 M GuHCl state of Ub* is clearly a bit more compact than a random coil.

Kinetic Data. Examples of the nonequilibrium measurements are shown in Figures 2−4. Figure 4 shows 5 → 0.7 M GuHCl stopped-flow refolding experiments with $\tau_d \approx 6$ ms dead time and time resolution. At −20°/45%, −10°/25%, and +4°/45%,

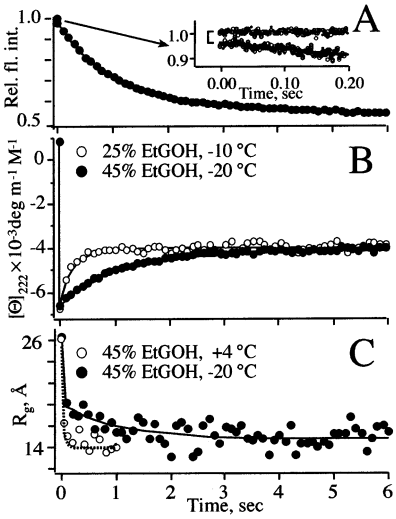

Figure 4. Ub* refolding kinetics slowed by low temperature/high solvent viscosity. A: Fluorescence signal at −20 °C/45% EtGOH. Only a very small (4%) burst phase is seen, bracketed in the insert at full time resolution (the top trace in the insert is the steady-state data before the GuHCl jump). No large burst phase is evident at higher temperatures or lower EtGOH concentrations either. B: Circular dichroism signal at −20 °C/45% EtGOH and −10 °C/25% EtGOH. The same G → N folding phase observed by fluorescence is preceded by a >160% burst phase. At higher temperature and lower EtGOH concentration the G → N rate rapidly increases, but the burst amplitude remains constant up to the highest temperatures (4 °C) where the G → N phase can be resolved. C: Small-angle X-ray scattering reveals a decrease of R_g to ≈1.2 times the native value during the burst phase, followed by a smaller decrease to the native value during the main folding phase also observed by fluorescence.

TABLE 2: Kinetic Fitting Parameters Derived from the Stopped-Flow Data (F = fluorescence, C = CD, and X = SAXS)[a]

conditions	A_{Burst}, %	t_{Fast}, s	A_{Fast}, %	t_{Slow}, s	A_{Slow}, %
4°/45%, F	7	0.043(1)	74	0.46(4)	19
4°/45%, X	80		20	-	-
−10°/25%, C	165	0.17(3)	−56	1.2(6)	−9
−20°/45%, C	163	0.9(1)	−51	4.2(4)	−12
−20°/45%, F	4		77		19
−20°/45%, X	64		32		4

[a] Exponential lifetime uncertainties (2σ) in parenthesis are in the last digit.

the main folding phase can be completely resolved because of the increased solvent viscosity afforded by these conditions ($\eta_{-20°/45\%} = 13\eta_{25°/0\%}$). The kinetic data for all techniques were fitted simultaneously to a double exponential decay with an initial offset to account for a potential burst phase (Table 2).

At −20°/45%, the fluorescence intensity monitors a main folding phase that accounts for nearly 80% of the amplitude and has an observed rate constant $k_{main} \approx 1.1$ s^{-1}. It speeds up to 5.9 s^{-1} at −10°/25% and 23 s^{-1} at +4°/45%. This corresponds to the partially resolved millisecond main folding phase previously observed at 25 °C. Most of the remaining signal is fitted by a slower phase, possibly due to proline or other isomerization events discussed elsewhere for Ub* (Table 2). The distorting effect of such phases can be mitigated by using double jump techniques,[14] but since our dead time is up to 150 times faster than the main folding phase, this was deemed unnecessary. A small (4%) burst phase remains. It is unambiguously resolved in the insert in Figure 4A because the product $\tau_d k_{main}$ is at least 20 times smaller under our conditions than in previous studies. Barring this small burst phase, our fluorescence data could be explained by a two-state model.

The picture is entirely different by circular dichroism (at −20°/45%, −10°/25%). Large burst phases are seen where the fluorescence hardly changes. The CD overshoots from +500 ± 400 to −6800 ± 400 deg m^{-1} M^{-1} (> 160%) at 222 nm, and the k_{main} phase accounts for about −50% of the signal. Figure 2 shows the CD spectrum of the burst phase at −20°/45%. It exceeds the native CD spectrum everywhere except at λ > 235 nm, indicating that the large CD spectrum of the intermediate is due to secondary structure formation, not aromatic side chain contributions. Because of the large signal-to-noise ratio of the CD burst phase, we can set an upper limit of $\tau_{burst} < 2$ ms: no tail of a decay is evident in the data from 6 ms onward.

The cooperativity of the CD burst phase is shown in Figure 3A. A titration of the burst phase, obtained by direct measurement of the burst phase amplitudes at different final GuHCl concentrations, shows an S-shaped transition with a midpoint of 2.5 M GuHCl and an end point at 0.7 M GuHCl outside the D−N range. Unlike the 4% fluorescence change, the denaturation curve of the CD burst phase cannot be explained by linear baseline extrapolation (Figure 1A). Because the main phase is much slower than our dead time, the data in Figure 3A were obtained without any extrapolation (they are the first point after mixing).

The CD burst phase amplitude remains constant at −6800±400 deg m M^{-1} over the full concentration range of 34−130 μM, and the rate k_{main} remains unchanged, ruling out concentration-dependent causes such as aggregation. The burst phase amplitude in Figure 4B also shows no sign of decreasing as the temperature is raised from −20 to −10 °C, and the EtGOH concentration is lowered from 45% to 25%. Above 4 °C, k_{main} increases by more than a factor of 30 compared to −20 °C, and we can no longer resolve the main phase reliably, but there are no changes in burst phase up to that point.

Small-angle X-ray scattering (at −10°/25% and +4°/45%) also reveals a very large change preceding the main folding phase observed by fluorescence. The radius of gyration changes during the burst phase from 26.5 to ≈17 Å, to within 20% of the native value, under both conditions. This is a classical molten globule radius. It then relaxes with rate constant k_{main} to the observed steady-state native value of ≈14.5 Å.

To summarize, the largest change in both the circular dichroism and X-ray scattering data at −20°/45%, −10°/25%, and +4°/45% occurs in the dead time of the experiment. The burst phase persists over a 4-fold protein concentration range and at the highest temperatures where it can be distinguished reliably from the main fluorescence phase, indicating the formation of a very compact ensemble with considerable secondary structure and a cooperative unfolding curve. The smallness of the same burst phase in the fluorescence data indicates that the tryptophan probe does not yet have nativelike packing.

Discussion

During the folding of many small proteins, some spectroscopic signatures change very rapidly, before subsequent slower kinetic phase(s) form the native state.[18,21,33,34] The temptation to assign that initial fast phase to the formation of a thermodynamic intermediate is strong. However, a bona fide thermodynamic intermediate must be separated by a barrier from the denatured ensemble as well as the native state and must be both more compact and more structured than the initial denatured state.

It has been pointed out that a rapidly changing observable following initiation of refolding does not by itself indicate that a folding event has occurred.[11,35] This idea is mainly based on experiments that show only a modest change in amide protection, fluorescence intensity, or CD signature of a protein at early times. Such small changes could be interpreted as an extrapolation of the unfolded state signal (Figure 1A).

The present work puts strict limits on both of these scenarios under our solvent conditions. The combination of low temperature and high viscosity solvent allows us to unambiguously resolve the burst phase amplitude and also to place an upper limit on the barrier separating the denatured state from any rapidly formed nonequilibrium folding ensemble. The near-native radii of gyration in Figure 4, a cooperative burst phase in Figure 3, and burst CD spectra outside the range bracketed by the D and N states in Figure 2 cannot be explained by a small extrapolation of the D state spectroscopic signature, or by a slight contraction of the unfolded polypeptide chain.

Time Scales and Free Energies. Thermodynamically, Ub* behaves like a two-state folder. Unfolding titrations under identical solvent conditions, but monitored by different spectroscopic probes, were all fitted with a single set of thermodynamic parameters (Figure 3 and Table 1; a limit of $< 10\%$ third state contribution can be set by our signal-to-noise-ratio). Additional local minima therefore lie higher in free energy than D and N by $\geq 3kT$ near the midpoint (± 1 M GuHCl) of the denaturation transitions. Well below the titration midpoint, only N contributes to the spectroscopic signature. Local minima must still lie $\geq 3kT$ above N to be unpopulated, but may lie below D.

Because our observed burst phase is much faster than the main folding phase, we can treat the final G $\rightarrow$ N process as quasi two-state, subject to a preequilibrium. To obtain the temperature-dependent free energy, we use Kramers' equation in the high viscosity limit:[36]

$$k_f = \nu^{\dagger}\frac{\eta_0}{\eta(T)}e^{-\Delta G^{\dagger}(T)/k_B T} \quad (1)$$

The most natural choice of prefactor $\nu^{\dagger}$ is a barrier-free diffusive contact time in the solvent rather than $k_B T/h$. We choose a value of (50 ns)$^{-1}$ at 25 °C, conservatively near the bottom end of the estimated range.[37−39] Under our conditions and those in ref.,[14] $k_{main} \approx k_f$. Our rates (Table 2), estimated viscosities,[40,41] and the 25 °C rate from ref 14 interpolated to 0.7 M GuHCl yield $\Delta G^{\dagger}(T) = 28-30$ kJ mol^{-1} at 0.7 M GuHCl concentration. The consistency of the free energies in the -20 to $+25$ °C range indicates that our k_{main} may indeed be assigned to the main folding phase previously observed at room temperature.[8,14,42] It accounts for nearly all the fluorescence signal observed by us. It would appear that no early folding intermediates with a fluorescence signature differing significantly from the unfolded state in the 325−400 nm range exist under our experimental conditions.

Yet what is barely visible in the fluorescence intensity data is very prominent by CD or SAXS. The burst phase is at least 450 times faster than k_{main}, setting an upper limit of $\leq 8k_B T$ (≤ 19 kJ/mol) on its activation energy, compared to 14 $k_B T$ (35 kJ/mol) for the main phase. A compact and structured ensemble accumulates rapidly over a low or absent barrier before the final barrier crossing to the native state.

Structure Formation in the Early Folding Ensemble. The early folding ensemble is clearly more than a slightly contracted denatured state: its CD spectrum in the secondary structure region exceeds that of the native state; its radius of gyration is only 1.2 times that of the native state; its titration curve by CD is cooperative. We have here a "best case" where a spectroscopic observable corresponds more nearly to Figure 1B than Figure 1A.

CD overshoots have been observed previously in two cases: excess formation of α-helical structure at short times[18,43] and aromatic side chain contributions during packing.[44] The present observation differs from both of these cases. The burst phase CD spectrum in Figure 2 is not a simple α-helical spectrum. Also, overshoots caused by aromatic side chains lead to increased CD at >235 nm and are resolvable by stopped-flow even at room temperature because the required side chain packing is a slow process.

The rapid changes in the structure of Ub* seen by CD and X-ray scattering are difficult to observe by fluorescence. Fluorescence is not sensitive to early solvated stages of protein compaction and secondary structure formation because interactions giving rise to quenching of native protein fluorescence are highly specific, short range, and orientation-dependent.[45,46] A recent analysis of ubiquitin showed that fluorescence intensity alone cannot be taken as a reliable indicator for thermodynamic three-state folding,[16] and the present work shows that observation of a single major phase by fluorescence alone cannot rule out accumulation of a structured and compact ensemble preceding this phase.

It would be very interesting to observe Ub* under our conditions by NMR hydrogen exchange. It may well turn out that no large protection factors are observed, indicating that the observed intermediate, albeit compact, has highly fluctuating and easily exchanged secondary structure. There are indications for this in bovine ubiquitin at higher temperature.[47] Hydrogen exchange and tryptophan fluorescence (except in the FRET case) may turn out not to be very sensitive indicators of compact but solvated structures that form before side chain packing and solvent exclusion occur during the final stages of folding. For example, recent work on apomyoglobin shows that the ^{13}C chemical shifts show propensity for nativelike fluctuating structure even at pH 2.3,[48] when apomyoglobin is normally considered highly unfolded, and no amide proton protection exists. Thorough solvation of intermediates may also complicate the interpretation of ϕ value analysis, which relies on changes in side-chain interactions: considerable structure may exist in intermediates even in cases where ϕ values are small.

Our experiments could resolve the main phase only to 4 °C, or to 25% EtGOH, because of the same dead-time limitations encountered by previous studies. The existence of a compact intermediate with secondary structure is therefore proven only under those conditions. However, tuning by temperature and ethylene glycol, especially considering that these two variables counteract each other (C_m is identical at $-20°/45\%$ and $+25°/0\%$ in Table 1), is expected to produce only small free energy changes between 4 °C (our highest T) and room temperature, and 25% EtGOH (our lowest EtGOH concentration) and aqueous solution. Even at 25°/0%, Ub* is thus at best an apparent two-state folder; the state G lies at most a few kJ/mol higher in free energy than D and N. Such high-energy intermediates have been studied, and tuning experiments can be designed to reveal indirect signatures of their existence.[1,49−51] As discussed in ref 52, such intermediate basins, even if not highly populated, are important for our understanding of the folding free energy surfaces because they have nonlocal effects on the folding kinetics.

One previous report of ubiquitin folding studied by CD seems incompatible with our data at first glance (assuming the $-10°/25\%$ results extrapolate to room-temperature aqueous solution).

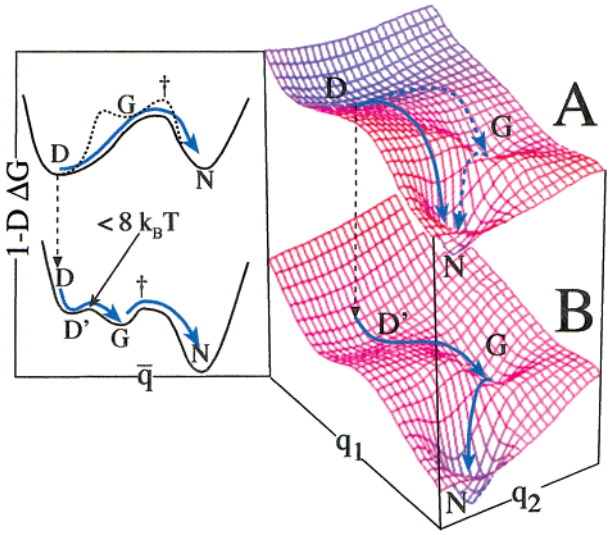

Figure 5. Free energy model for structure formation during burst phases (blue = lowest free energy; red = highest). The upper free energy surface is under conditions biased toward the denatured state (≈4 M GuHCl), the lower free energy surface is under conditions biased toward the native state (0 M GuHCl). The topography of both surfaces is very similar, but due to the exponential sensitivity of populations and rates to free energy, the apparent one-dimensional scenarios are very different. When conditions are biased against the native state, folding proceeds quasi two-state, perhaps involving a high free energy intermediate (dotted curve). When biased toward the native state, the protein rapidly collapses over a small barrier to a very compact ensemble with significant secondary structure, followed by two-state folding over a higher barrier.

A study of bovine ubiquitin in pH 5.5 and 8.5 aqueous buffers at 20 °C reports a monotonic decrease of CD_{225} from the unfolded to the folded value in 10 ± 3 ms, with a ≈3 ms dead-time.[47] However, that kinetic transient has a signal-to-noise ratio < 3:1. Our bandwidth-adjusted signal-to-noise ratio is at least 50 times greater because our solvent conditions allow $\tau_d k_{main} \ll 1$. By adding noise to our data and increasing $\tau_d k_{main} \approx 1$, we obtain an overdamped CD transient similar to the one previously reported. The same study also reports no amide proton protection in the dead time, in contrast with Briggs and Roder, who report substantial protection (50%) within 10 ms of mixing.[53]

A Consistent Picture. Ub* rapidly collapses to a compact ensemble with significant secondary structure before folding more slowly to the native state. Yet this is not evident by fluorescence analysis at $\lambda > 325$ nm, nor are any intermediates evident during steady-state unfolding titrations.

A qualitative picture that reconciles these facts and accounts for all the observed data is shown in Figure 5. The free energy is shown as a function of two reaction coordinates: q_1 leads to the native state, while q_2 accounts for nonnative secondary structure that may be present in compact nonnative states whose CD exceeds the native state's in magnitude. Surface A corresponds to denaturing conditions (e.g., 4 M GuHCl) and is biased toward D. Surface B corresponds to refolding conditions and is biased toward N (e.g 0.7 M GuHCl). The state G is always substantially higher in free energy than N.

Under denaturing conditions, the lowest free energy path may lead directly from D to N (solid arrow) or may proceed via G (dashed arrows). Current experimental data for Ub* under denaturing conditions are insufficient to distinguish a "rough" transition state region with high-energy intermediates[49,54] from a smooth transition state region with no intermediates.[50] Under folding conditions, the lowest free energy path leads from D′ via G to N. One-dimensional cuts through the 2-D surfaces along the minimum energy paths are shown to the left in Figure 5

Consider a steady-state unfolding titration first. Surface A smoothly morphs to surface B during titration. Although the mechanism may switch from the solid arrow to the dotted arrow in A, $\Delta G_N \ll \Delta G_G < \Delta G^\dagger$ throughout, yielding a cooperative two-state mechanism, as the protein evolves on a succession of surfaces between A and B.

In a stopped-flow experiment, the molecule starts on surface A at D, and is promoted to surface B. All subsequent time evolution occurs on surface B. There it first relaxes to D′, which is slightly more nativelike than D. (The unfolded slope of the denatured state in Figure 3A is toward the native CD signal) From D′, G forms over a barrier $\leq 8k_BT$ ($\leq$19 kJ/mol). Although G is compact and contains secondary structure, the tryptophan residue is still solvated and mobile, so the D′ to G transition is not evident when probed by fluorescence intensity (and perhaps not by hydrogen exchange either). Finally the native state forms over a more substantial $14k_BT$ (35 kJ/mol) barrier as the core packs and solvent is excluded, and this process can be monitored by fluorescence intensity, as well as by CD and X-ray scattering.

The fact that G has an S-shaped denaturation curve (Figure 3A) is by itself not proof of a substantial barrier separating D′ and G. A continuous distribution of microstates $\{i\}$ between D and G will also yield an S-shaped denaturation curve, if the microstates' relative free energies tune linearly as $\Delta\Delta G_i = m_i^{(2)}$-$[GuHCl]\delta\bar{q}$. (Here, $m^{(2)}$ is the tuning slope, and $\delta\bar{q}$ is the reaction coordinate in Figure 5.)[55] On the other hand, cooperativity such as in Figure 3A is often associated with at least a small barrier. Exactly where in the $(0-8)k_BT$ range the D to G barrier falls under our conditions remains to be seen.

Two important general conclusions are illustrated by Figure 5. First, critical structure formation during folding cannot be universally assigned to an "uphill" or a "downhill" process. The same structure that forms net uphill on A (between D and G) forms net downhill on B before the barrier is reached. The concepts "uphill" and "downhill" are solvent- and sequence-dependent. Second, it is not fruitful to insist upon a single folding mechanism on the free energy surface. Protein populations and rate constants are exponentially sensitive to the free energies ΔG and $\Delta G^\dagger$, respectively. Topographically very similar surfaces such as 5A and 5B can therefore result in apparently different folding mechanisms, or even parallel mechanisms when such paths come within $3k_BT$ of one another.[56] Characterization of the free energy surface itself is thus a more robust goal for folding studies. Another recent example of this includes an immunogloulin binding domain, whose mechanism can be switched drastically by a few mutations,[57] and the Pin WW domain, whose rate-limiting barrier can be moved by drastic mutations.[54]

Acknowledgment. X-ray scattering experiments were performed under approval of the Photon Factory (Proposal No. 2000G330 by M.G.). The authors thank Dr. Masaki Kojima of Tokyo University of Pharmacy and Life Science, and Dr. Kazuki Ito of the Foundation for Advancement of International Science (FAIS), for their treatment of the CCD data. M.G., J.E., and E.L. were supported by NIH Grant GM057175. The travel was supported by a joint grant of H.K. and M.G. from the JSPS and NSF-INT0089286.

References and Notes

(1) Bachmann, A.; Kiefhaber, T. *J. Mol. Biol.* **2001**, *306*, 375.
(2) Chan, C.; Hu, Y.; Takahashi, S.; Rousseau, D. L.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1779.

(3) Shastry, M. C. R.; Roder, H. *Nat. Struct. Biol.* **1998**, *5*, 385.
(4) Hagen, S.; Eaton, W. A. *J. Mol. Biol.* **2000**, *301*, 1019.
(5) Akiyama, S.; Takahashi, S.; Kimura, T.; Ishimori, K.; Morishima, I.; Nishikawa, Y.; Fujisawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1329.
(6) Chen, E.; Wood, M. J.; Fink, A. L.; Kliger, D. S. *Biochemistry* **1998**, *37*, 5589.
(7) Akiyama, S.; Takahashi, S.; Ishimori, K.; Morishima, I. *Nat. Struct. Biol.* **2000**, *7*, 514.
(8) Khorasanizadeh, S.; Peters, I.; Butt, T.; Roder, H. *Biochemistry* **1993**, *32*, 7054.
(9) Khorasanizadeh, S.; Peters, I.; Roder, H. *Nat. Struct. Biol.* **1996**, *3*, 193.
(10) Segel, D.; Bachmann, A.; Hofrichter, J.; Hodgson, K.; Doniach, S.; Kiefhaber, T. *J. Mol. Biol.* **1999**, *288*, 489.
(11) Sosnick, T. R.; Shtilerman, M. D.; Mayne, L.; Englander, S. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 8545.
(12) Englander, S. W.; Sosnick, T. R.; Mayne, L. C.; Shtilerman, M.; Qi, P. X.; Bai, Y. *Acc. Chem. Res.* **1998**, *31*, 767.
(13) Englander, S. W. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 213.
(14) Krantz, B. A.; Sosnick, T. R. *Biochemistry* **2000**, *39*, 11696.
(15) Lakowicz, J. R. *Principles of fluorescence spectroscopy*; Plenum Press: New York, 1983.
(16) Ervin, J.; Larios, E.; Osvath, S.; Schulten, K.; Gruebele, M. *Biophys. J.* **2002**, *83*, 473.
(17) Qin, Z.; Hu, D.; Shimada, L.; Nakagawa, T.; Arai, M.; Zhou, J.; Kihara, H. *FEBS Lett.* **2001**, *507*, 299.
(18) Arai, M.; Ikura, T.; Semisotnov, G. V.; Kihara, H.; Amemiya, V.; Kuwajima, K. *J. Mol. Biol.* **1998**, *275*, 149.
(19) Fujiwara, K.; Arai, M.; Shimizu, A.; Ikeguchi, M.; Kuwajima, K.; Sugai, S. *Biochemistry* **1999**, *38*, 4455.
(20) Laub, P.; Khorisanizadeh, S.; Roder, H. *Protein Sci.* **1995**, *4*, 973.
(21) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5759.
(22) Roder, H.; Shastry, M. C. R. *Curr. Opin. Struct. Biol.* **1999**, *9*, 620.
(23) Sabelko, J.; Ervin, J.; Gruebele, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6031.
(24) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
(25) Lazar, G. A.; Desjarlais, J. R.; Handel, T. M. *Protein Sci.* **1997**, *6*, 1167.
(26) Kihara, H. *J. Synchrotron Radiat.* **1994**, *1*, 74.
(27) Amemiya, Y.; Wakabayashi, K.; Hamanaka, T.; Wakabayashi, T.; Matsushita, T.; Hashizume, H. *Nucl. Instrum. Methods* **1983**, *208*, 471.
(28) Ito, K.; Kamkubo, H.; Arai, M.; Kuwajima, K.; Amemiya, Y.; Endo, T. Photon Factory Activity Report. Photon Factory: Tsukuba, Japan, 2001.
(29) Guinier, A.; Fournet, G. *Small-Angle Scattering of X-rays*; John Wiley & Sons: New York, 1955.
(30) Amemiya, Y.; Ito, K.; Yagi, N.; Asano, Y.; Wakabayashi, K.; Ueki, T.; Endo, T. *Rev. Sci. Instrum.* **1995**, *66*, 2290.
(31) *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman: New York, 1992; p 547.
(32) Miller, W. G.; Goebel, C. V. *Biochemistry* **1968**, *7*, 3925.
(33) Takahashi, S.; Yeh, S.; Das, T. K.; Chan, C.; Gottfried, D. S.; Rousseau, D. L. *Nat. Struct. Biol.* **1997**, *4*, 44.
(34) Jennings, P.; Wright, P. *Science* **1993**, *262*, 892.
(35) Qi, P. X.; Sosnick, T. R.; Englander, S. W. *Nat. Struct. Biol.* **1998**, *5*, 882.
(36) Kramers, H. A. *Physica* **1940**, *7*, 284.
(37) Hagen, S. J.; Hofrichter, J.; Eaton, W. A. *J. Phys. Chem. B* **1997**, *101*, 2352.
(38) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597.
(39) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220.
(40) *CRC Handbook of Tables for Applied Engineering Science*; 2nd ed.; Bolz, R. E., Tuve, G. L., Eds.; CRC Press: Boca Raton, FL, 1973.
(41) Weast, R. C. *CRC Handbook of Chemistry and Physics*; CRC Press, Inc.: Cleveland, OH, 1997.
(42) Khorasanizadeh, S.; Peters, I. D.; Roder, H. *Protein Eng.* **1995**, *Suppl. S 21*, 21.
(43) Hamada, D.; Segawa, S.; Goto, Y. *Nat. Struct. Biol.* **1996**, *3*, 868.
(44) Chaffotte, A. F.; Guillou, Y.; Goldberg, M. E. *Biochemistry* **1992**, *31*, 9693.
(45) Steiner, R. F.; Kirby, E. P. *J. Phys. Chem.* **1969**, *73*, 4130.
(46) Harris, D. L.; Hudson, B. S. *Biochemistry* **1990**, *29*, 5276.
(47) Gladwin, S. T.; Evans, P. A. *Folding Des.* **1996**, *1*, 407.
(48) Yao, J.; Chung, J.; Eliezer, D.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2001**, *40*, 3561.
(49) Pappenberger, G.; Saudan, C.; Becker, M.; Merbach, A. E.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 17.
(50) Ternstrom, T.; Mayor, U.; Akke, M.; Oliveberg, M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14854.
(51) Wagner, C.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6716.
(52) Gruebele, M. *Curr. Opin. Struct. Biol.* **2002**, *12*, 161.
(53) Briggs, M.; Roder, H. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2017.
(54) Jäger, M.; Nguyen, H.; Crane, J.; Kelly, J.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373.
(55) Parker, M. J.; Marqusee, S. *J. Mol. Biol.* **1999**, *293*, 1195.
(56) Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9029.
(57) Nauli, S.; Kuhlman, B.; Baker, D. *Nat. Struct. Biol.* **2001**, *8*, 602.