Accelerated Publications

An Early Intermediate in the Folding Reaction of the B1 Domain of Protein G Contains a Native-like Core†

Soon-Ho Park,‡ Karyn T. O'Neil,§,|| and Heinrich Roder*,‡,||

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, The DuPont Merck Pharmaceutical Company, Wilmington, Delaware 19880, and Department of Biochemistry & Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received August 4, 1997; Revised Manuscript Received October 10, 1997⊗

ABSTRACT: The folding kinetics of a 57-residue IgG binding domain of streptococcal protein G has been studied under varying solvent conditions, using stopped-flow fluorescence methods. Although GB1 has been cited as an example of a protein that obeys a two-state folding mechanism, the following kinetic observations suggest the presence of an early folding intermediate. Under stabilizing conditions (low denaturant concentrations, especially in the presence of sodium sulfate), the kinetics of folding shows evidence of a major unresolved fluorescence change during the 1.5 ms dead time of the stopped-flow experiment (burst phase). Together with some curvature in the rate profile for the single observable folding phase, this provides clear evidence of the rapid formation of compact states with native-like fluorescence for the single tryptophan at position 43. In refolding experiments at increasing denaturant concentrations, the amplitude of the sub-millisecond phase decreases sharply and the corresponding slope (*m* value) is only about 30% lower than that of the equilibrium unfolding curve indicative of a pre-equilibrium transition involving cooperative unfolding of an ensemble of compact intermediates. The dependence on guanidine hydrochloride concentration of both rates and amplitudes (including the equilibrium transition) is described quantitatively by a sequential three-state mechanism, U ⇌ I ⇌ N, where an intermediate (I) in rapid equilibrium with the unfolded state (U) precedes the rate-limiting formation of the native state (N). A 66-residue fragment of GB1 with an N-terminal extension containing five apolar side chains exhibits three-state kinetic behavior virtually identical to that of the 57-residue fragment. This is consistent with the presence of a well-shielded native-like core excluding the N-terminal tail in the early folding intermediate and argues against a mechanism involving random hydrophobic collapse, which would predict a correlation between overall hydrophobicity and stability of compact states.

It has long been thought that, for a protein to acquire its unique native structure within a reasonable time span, it has to go through a series of progressively more structured intermediate states (*1*−*3*). Indeed, partially folded kinetic intermediates have been well documented for a large number of proteins, using methods such as optically detected stopped-flow or quenched-flow hydrogen exchange (*4*−*6*). Kinetic complexities in folding can sometimes be attributed to intrinsically slow steps, such as peptide bond isomerization, disulfide exchange, metal coordination, or dimerization events. However, even in the absence of such rate-limiting factors, many proteins exhibit well-populated intermediates during early stages of folding. In particular, time-resolved circular dichroism or fluorescence measurements often show evidence of conformational changes occurring within the first few milliseconds of refolding (the so-called burst phase or missing amplitude effect), indicating that some secondary and tertiary structural features can appear long before the rate-limiting step in folding (e.g., refs 7 and 8). The effect of denaturant concentration on the rate of folding can provide additional evidence of the accumulation of early intermediates (*9*, *10*). However, the structural properties of these early intermediates and their mechanistic role are still poorly understood (*11*−*13*).

Two competing models currently discussed in the literature are a mechanism involving nonspecific hydrophobic collapse (model A) and a sequential three-state folding mechanism (model B). Model A attributes the burst-phase effect in protein folding to a largely random hydrophobic collapse of the unfolded chain into a compact denatured state, D, caused by the change in solvent conditions used to induce folding (*12*, *14*). D has been described as the denatured state of the protein under physiological conditions, which may be relatively compact due to nonspecific hydrophobic interactions (e.g., ref 15). Since such a collapse is likely to result in some non-native contacts, D is not necessarily an obligatory intermediate, and the formation of N may require partial unfolding. Structurally, model A is essentially a two-state process. In contrast, model B assumes that refolding

† This work was supported by NIH Grant GM35926 to H.R. and by NIH Grant CA06927 and an appropriation from the Commonwealth of Pennsylvania to the Institute for Cancer Research.

* Correspondence should be addressed to this author at the Fox Chase Cancer Center. Phone: (215) 728-3123. Fax: (215) 728-3574. E-mail: H_Roder@fccc.edu.

‡ Fox Chase Cancer Center.

§ The DuPont Merck Pharmaceutical Company.

|| University of Pennsylvania.

⊗ Abstract published in *Advance ACS Abstracts,* November 1, 1997.

S0006-2960(97)01914-4 CCC: $14.00

from the unfolded state, U, proceeds along a sequential pathway, U ⇌ I → N, with at least one obligatory intermediate, I, which lies along a direct path toward the native state, N (for a recent review, see ref 11). In this scheme, U and I interconvert rapidly compared to the I → N step, which limits the rate of folding. In using such empirical kinetic schemes to describe protein folding, it is important to realize that by "state" we generally refer to a statistical ensemble of conformations with some common attributes rather than a unique structure. In particular, I is thought to represent an ensemble of compact states with some native-like aspects, at both the secondary and tertiary structural level.

The discovery of several small proteins that fold rapidly and efficiently without detectable intermediates in an apparent two-state folding reaction (e.g., refs 16 and 17) appears to support model A. Because of their small size, it has been suggested that these proteins can fold directly to the native state without going through an off-pathway random hydrophobic compaction that is expected for some larger proteins (*18*). However, recent systematic folding studies of ubiquitin (*10*) as well as several other proteins (*9*, *10*, *19*−*24*) appear to favor model B. Ubiquitin variants containing a tryptophan at position 45 and Val, Leu, or Ile at position 26 showed a significant burst phase and curved rate profiles (semilog plot of rate *vs* denaturant concentration), which was attributed to an ensemble of early intermediates with a well-developed core structure (*10*), stabilized primarily by hydrophobic interactions (*19*). Quantitative kinetic modeling showed that the rates of folding and unfolding, as well as the corresponding amplitudes, for ubiquitin and all of its variants investigated are fully consistent with a sequential three-state mechanism featuring an obligatory intermediate (model B). For example, mutation of a hydrophobic core residue, Val 26 to Ala (V26A), which destabilized both I and N, slowed the overall rate of folding and resulted in apparent two-state kinetics. However, as predicted by model B, the intermediate became observable again when V26A[1] was investigated under more stabilizing conditions in the presence of 0.4 M Na_2SO_4. On the basis of these observations, Khorasanizadeh et al. (*10*) predicted that some of the proteins with apparent two-state folding reactions may exhibit detectable intermediates under sufficiently stabilizing conditions.

Here, we report kinetic evidence indicating the presence of an early intermediate in the folding reaction of the B1 domain of streptococcal protein G (GB1), a bacterial cell surface protein found in *Streptococcus* containing two or three homologous domains that recognize the Fc region of IgG (*25*, *26*). These IgG binding domains have identical three-dimensional structures consisting of a four-stranded β-sheet with a flanking α-helix (*27*−*30*). GB1 contains no cysteines or prolines, and there are no known metal binding sites or cofactors. Thus, it is an ideal model for studying elementary conformational events in protein folding. Despite its small size, GB1 has a well-developed hydrophobic core (*27*) and is remarkably stable (*31*). Previous equilibrium (*31*) and kinetic (*32*) studies suggested that GB1 obeys a two-state folding mechanism. However, rigorous kinetic criteria for a two-state mechanism have not been applied. In fact, using pulsed hydrogen exchange and two-dimensional NMR, Kuszewski et al. (*33*) found evidence of amide protection prior to the rate-limiting step in refolding. In this study, we measured the kinetics of folding and unfolding of GB1 as a function of denaturant concentration, using the stopped-flow method, which revealed a well-populated early folding intermediate with native-like fluorescence properties. In addition to this 57-residue fragment of protein G (GB1−57) corresponding to the structured region of the B1 domain (*27*), we also investigated a longer fragment (GB1−66) containing a 9-residue N-terminal extension. Our observation that addition of this rather hydrophobic segment has no effect on the stability of the intermediate and the kinetics of folding is consistent with the rapid formation of a compact state with a protected hydrophobic core.

MATERIALS AND METHODS

Materials. Recombinant GB1−57 and GB1−66 were prepared as described by O'Neil et al. (*34*). GB1−57 and GB1−66 were at least 95% pure on the basis of Coomassie brilliant blue staining after SDS−PAGE. Purified GB1−57 and GB1−66 had the expected mass per charge as measured by a Voyager DE matrix-assisted laser desorption mass spectrometer (PerSeptive Biosystems Inc., Farmingham, MA). Ultrapure grade GuHCl was purchased from ICN Biomedicals Inc. (Costa Mesa, CA). The GuHCl concentration was determined by refractive index measurement (*35*) using a Reichert-Jung Mark II Abbe refractometer (Cambridge Instruments Inc., Buffalo, NY). All other chemicals were reagent grade or better.

Spectral Measurements. Fluorescence emission spectra were measured using a AMINCO-Bowman Series 2 Luminescence Spectrometer (SLM-AMINCO, Urbana, IL). For measurements of emission spectra, a single tryptophan (Trp 43) was excited at 295 nm and the emission spectrum from 310 to 450 nm was recorded. CD spectra were recorded at 20 °C on an Aviv model 62A DS spectropolarimeter (Aviv, Lakewood, NJ). Far-UV (200−260 nm) and near-UV (250−350 nm) CD spectra of 40 μM GB1 in 20 mM acetate buffer at pH 4.0 were recorded using 1 and 10 mm path length cells, respectively.

Equilibrium Unfolding Measurements. For equilibrium unfolding measurements, Trp 43 was excited at 295 nm (4 nm band width), and the fluorescence emission at 340 nm (8 nm band width) was measured as a function of GuHCl concentration. Equilibrium unfolding experiments were analyzed assuming a linear dependence of the free energy of unfolding, ΔG_U, on the GuHCl concentration, C, described by eq 1 (*36*)

$$\Delta G_U = \Delta G_U^\circ - m_{eq}C = m_{eq}(C_m - C) \qquad (1)$$

where ΔG_U° represents the free energy of unfolding at 0 M GuHCl, m_{eq} is the slope of the transition curve, and C_m is the midpoint concentration.

Stopped-Flow Kinetic Measurements. The time course of refolding and unfolding of GB1 in various GuHCl concentrations was followed by monitoring changes in the intrinsic tryptophan fluorescence emission as a function of time. All measurements were made using a SFM-4/QS stopped-flow

[1] Abbreviations: CD, circular dichroism; DSC, differential scanning calorimetry; GB1, IgG binding domain B1 of streptococcal protein G; GB1−57, 57-residue fragment of GB1; GB1−66, 66-residue fragment of GB1; GuHCl, guanidine hydrochloride; LB1, IgG binding domain of peptostreptococcal protein L; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; V26A, ubiquitin variant with Phe 45 to Trp and Val 26 to Ala mutations.

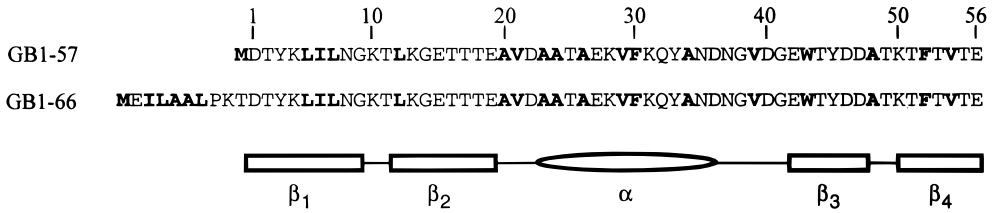

FIGURE 1: Amino acid sequence of GB1−57 and GB1−66, using the residue numbering of Gronenborn et al. (*27*). Hydrophobic residues are shown in bold. In the secondary structure diagram, the four β-strands are shown as rectangles and the single α-helix is indicated by an ellipse.

instrument (BioLogic, Claix, France). The fluorescence of Trp 43 was excited at 295 nm, and the change in fluorescence emission above 324 nm was measured using a high-pass glass filter with a 324 nm cutoff. The refolding reaction was measured after 11-fold dilution of a GB1 solution unfolded in 5 M GuHCl with buffer (20 mM sodium acetate at pH 4.0) and 1−4 M GuHCl solutions to the desired final GuHCl concentration. The unfolding reaction was induced by 11-fold dilution of native GB1 in 1.5 M GuHCl with 4−6 M GuHCl solutions to the desired final GuHCl concentration. Unless indicated otherwise, final protein concentrations were 4 μM for GB1−57 and 2 μM for GB1−66. All the GuHCl solutions contained 20 mM sodium acetate at pH 4.0. Various final GuHCl concentrations were generated by changing the volume ratio of buffer and GuHCl solutions. The final GuHCl concentrations were checked by measuring the refractive index of the solutions collected in each experiment, which were within 0.05 M of the expected GuHCl concentration. The dead time of the stopped-flow experiment was about 1.5 ms. The temperature of the stopped-flow cell compartment was maintained at 20 °C using an RCS-6D water bath (Brinkmann Instrument Co., Westbury, NY). All solutions were filtered and degassed prior to stopped-flow experiments. For the analysis of kinetic traces, at least five kinetic traces were averaged for each final GuHCl concentration. The averaged kinetic traces were then analyzed by single-exponential nonlinear least-squares fitting.

Quantitative Kinetic Modeling. To model the kinetic data, we describe the folding reaction as a series of coupled elementary reactions with microscopic rate constants, k_{IJ}. Each elementary reaction is considered to be a reversible two-state process. It is widely assumed that the logarithm of the elementary rate constant is linearly dependent on GuHCl concentration as in eq 2:

$$\ln k_{IJ} = \ln k_{IJ}^{\circ} + (m_{IJ}^{\ddagger}/RT)C \qquad (2)$$

where k_{IJ}° represents the microscopic rate constant at 0 M GuHCl, C denotes the concentration of GuHCl, and $m_{IJ}^{\ddagger}/RT$ represents the slope of ln k_{IJ} *vs* C. Observable rates and fractional amplitudes were calculated by solving the associated rate matrix (*37*, *38*) using standard numeric methods. Calculations were performed on a personal computer using a custom program (cf. ref 22) written in the ASYST language (Keithly Metrobyte, Rochester, NY).

RESULTS

GB1−57 and GB1−66. The amino acid sequences of the two fragments of GB1 studied are shown in Figure 1. GB1−57 corresponds to residues 2−56 of the B1 domain studied by Gronenborn et al. (*27*) with an additional Asp of the native protein G sequence at position 1 and an unprocessed N-terminal Met. GB1−66 has a nine-residue extension at the N terminus corresponding to the protein G sequence that connects the B1 domain with the preceding domain (*25*, *26*). Since five of these residues are apolar, inclusion of the N-terminal extension in GB1−66 results in an ∼30% increase in overall hydrophobicity compared to that of GB1−57.

Far- and near-UV CD spectra at pH 4.0 and 20 °C for GB1−57 and GB1−66 are identical to each other and to the previously reported spectra (*32*, *34*, *39*) (data not shown). GB1−57 and GB1−66 are also indistinguishable in terms of their tryptophan fluorescence emission spectra, under both native and denaturing conditions (data not shown). In 5 M GuHCl, the emission maximum lies near 350 nm, consistent with an unfolded conformation in which Trp 43 is fully exposed to solvent. The blue-shifted emission spectrum observed in the native state with a maximum near 340 nm and an ∼4-fold increase in intensity is consistent with the partial burial of Trp 43 in the hydrophobic interior, as seen in the NMR and X-ray structures (*27*−*30*). These observations confirm that the single tryptophan residue can serve as a sensitive probe for measuring conformational changes of GB1.

Unfolding Equilibrium. The GuHCl-induced unfolding transitions for GB1−57 and GB1−66 were monitored by tryptophan fluorescence, and the resulting equilibrium parameters are listed in Table 1. The free energy of unfolding is, within error, identical for GB1−57 and GB1−66, indicating that the N-terminal extension of GB1−66 is disordered and solvent-exposed, under both native and denaturing conditions. Both fragments are stabilized by about 2 kcal/mol on addition of 0.4 M Na_2SO_4. The salt-dependent stabilization is mainly reflected in an increase in C_m, while the addition of sodium sulfate has little effect on m_{eq}, which is a measure of the increase in exposed surface area when the protein unfolds (*40*). The m_{eq} values observed are consistent with those reported by O'Neil et al. (*34*) and Kuszewski et al. (*33*) measured under comparable conditions of pH and temperature.

Folding/Unfolding Kinetics. Representative kinetic traces for the fluorescence-detected folding and unfolding reaction of GB1−57 at pH 4.0 and 20 °C in the presence of 0.4 M Na_2SO_4 are shown in Figure 2. In contrast to most other proteins, the increase in fluorescence observed on refolding from 5 M GuHCl to various lower GuHCl concentrations

Table 1: Thermodynamic Parameters for Two GB1 Fragments from Equilibrium Unfolding Measurements and Kinetic Modeling[a]

		GB1−57			GB1−66		
method	$[Na_2SO_4]$ (M)	m_{eq}	C_m	ΔG_U°	m_{eq}	C_m	ΔG_U°
equilibrium unfolding	0.0	1.7 ± 0.1	2.7 ± 0.1	4.6 ± 0.4	1.7 ± 0.1	2.6 ± 0.1	4.4 ± 0.4
	0.4	1.9 ± 0.1	3.6 ± 0.1	6.8 ± 0.5	1.8 ± 0.1	3.4 ± 0.1	6.1 ± 0.5
kinetic modeling	0.0	1.9	2.6	4.9	1.9	2.5	4.7
	0.4	2.1	3.3	7.0	2.0	3.2	6.5

[a] The units for m_{eq}, C_m, and ΔG_U° are kilocalories per mole per molar, molar, and kilocalories per mole, respectively. All the measurements were done at 20 °C and pH 4.0. The m_{eq}, C_m, and ΔG_U° from kinetic modeling are defined as $m_{eq} = m_{UI} - m_{IN}^\ddagger + m_{NI}^\ddagger$, $\Delta G_U^\circ = \Delta G_{IU}^\circ + \Delta G_{NI}^\circ$, and $C_m = \Delta G_U^\circ / m_{eq}$, respectively, where $\Delta G_{IU}^\circ = -RT \ln(1/K_{UI}^\circ)$ and $\Delta G_{NI}^\circ = -RT \ln(k_{NI}^\circ / k_{IN}^\circ)$. K_{UI}°, k_{NI}°, k_{IN}°, m_{UI}, $m_{IN}^\ddagger$, and $m_{NI}^\ddagger$ are listed in Table 2.

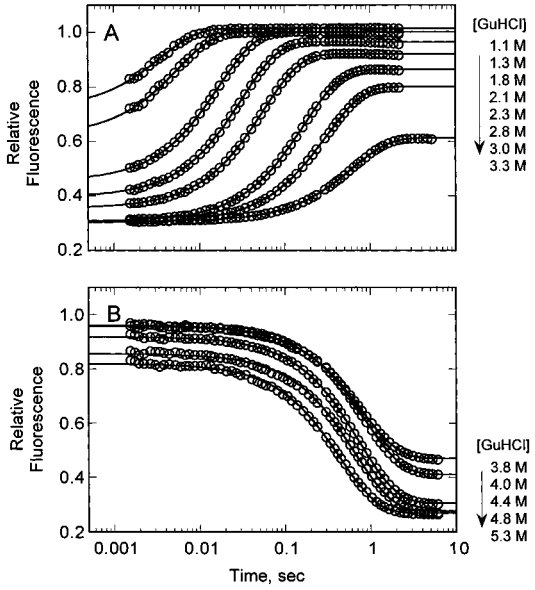

FIGURE 2: Kinetic traces for GB1−57 refolding (A) and unfolding (B) in varying concentrations of GuHCl in the presence of 0.4 M Na_2SO_4. The observed relative fluorescence (scaled with respect to the native state) is plotted *vs* time on a log scale (○). The lines represent single-exponential fits obtained by nonlinear least-squares analysis.

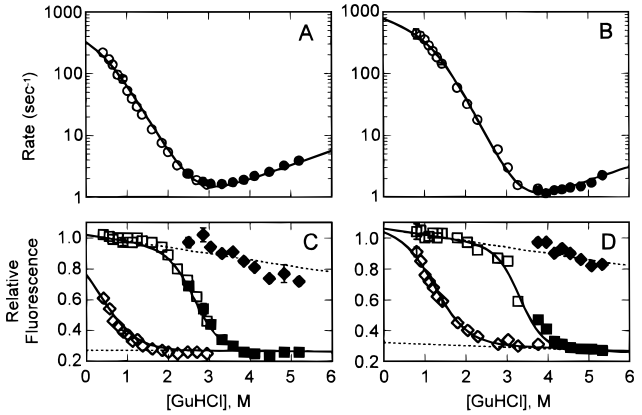

FIGURE 3: Denaturant-dependent kinetics of folding and unfolding of GB1−57 in the absence (A and C) and presence (B and D) of 0.4 M Na_2SO_4. Panels A and B show the rates of refolding (○) and unfolding (●) on a log scale *vs* GuHCl concentration. Panels C and D show the initial fluorescence intensities obtained by extrapolation to time zero (◇) and the equilibrium intensities at long times (□). Open symbols are from refolding experiments, and filled symbols are from unfolding experiments. The solid lines were obtained by kinetic modeling, based on Scheme 1 (Table 2). Baselines for the unfolded and the native states are shown as dotted lines.

(panel A) is well represented by a single-exponential phase (solid lines), which can be attributed to the lack of prolines in the sequence of GB1. Interestingly, it appears that a significant fraction of the total fluorescence change occurs within the 1.5 ms dead time of the stopped-flow apparatus (burst phase) in the refolding experiment under stabilizing conditions (low GuHCl concentration), indicating rapid formation of a state with native-like Trp 43 fluorescence. The increase in fluorescence observed on unfolding at several GuHCl concentrations above C_m (panel B) was also fitted to a single-exponential phase. However, in contrast to refolding, there is no evidence of an unresolved fluorescence change during unfolding; the small decrease in the initial signal with increasing GuHCl concentration is explained by the denaturant dependence of the native baseline value (see Figure 3). In addition to these measurements at a final protein concentration of 4 μM, we also repeated one set of refolding measurements (at 0.8 M GuHCl in the presence of 0.4 M Na_2SO_4) at both lower (0.5 and 0.9 μM) and higher (9 μM) concentrations of GB1−57. The kinetic traces obtained (after normalization with respect to a GuHCl-denatured control sample at each protein concentration) were identical within error, indicating that the folding reaction is unimolecular under the conditions used.

Figure 3 summarizes the kinetic results for GB1−57. The observed folding/unfolding rates on a log scale (rate profiles in the two upper panels) and the relative fluorescence amplitudes (lower panels) are plotted *vs* GuHCl concentration, both in the absence (panels A and C) and in the presence (panels B and D) of 0.4 M Na_2SO_4. In the absence of sulfate (Figure 3A), a nearly linear rate profile for refolding is observed, as expected for a two-state process (e.g., ref 17). Nevertheless, a significant burst phase is observed (Figure 3C). Under more stabilizing conditions in 0.4 M Na_2SO_4 (Figure 3B,D), the rate profiles are slightly curved at low GuHCl concentrations and the missing amplitude effect becomes more prominent. These observations are inconsistent with a simple two-state process. Above 1 M GuHCl, the initial fluorescence signal approaches the unfolded baseline (dotted lines at the bottom of panels C and D) and the rate profile becomes linear, as expected for a two-state reaction, indicating that the intermediate is no longer populated. The denaturant dependence of the relative fluorescence at long times (squares in panels C and D of Figure 3) is equivalent to the equilibrium GuHCl unfolding measurements and results in very similar m_{eq} and C_m values (Table 1). The initial fluorescence observed in unfolding experiments in the presence and absence of sodium sulfate agrees reasonably well with the expected fluorescence of the native state *vs* GuHCl concentration (upper dotted line in Figure 3C,D), and the logarithm of the unfolding rate varies linearly with GuHCl concentration. These observations

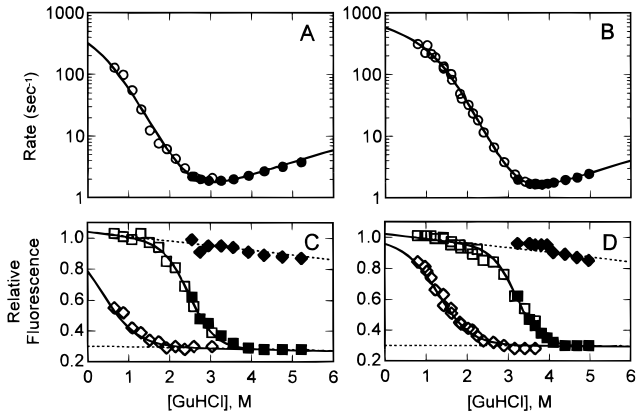

FIGURE 4: Denaturant-dependent kinetics of folding and unfolding reaction for GB1−66. Symbols and line types are the same as those in Figure 3.

indicate that GB1 unfolds in a single cooperative step without detectable intermediates.

The kinetics of refolding and unfolding for GB1−66 are summarized in Figure 4. Again, a pronounced burst phase is observed, especially in the presence of sodium sulfate (Figure 4D), where as much as 80% of the total fluorescence change occurs within the dead time (below 1 M GuHCl). Moreover, the rate profile clearly levels off toward low GuHCl concentrations. Comparison with Figure 3 reveals that the kinetic behavior of GB1−66 is very similar to that of GB1−57.

Quantitative Kinetic Modeling. Since the observation of a burst phase indicates the presence of at least one intermediate in the GB1-folding reaction, we used a sequential three-state model as a minimal kinetic mechanism (Scheme 1):

Scheme 1

$$U \underset{k_{IU}}{\overset{k_{UI}}{\rightleftharpoons}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N$$

where U and N represent the unfolded and native states, respectively, and I represents an ensemble of early intermediates with native-like fluorescence. In this scheme, I is an obligatory intermediate on a direct path from U to N. Interconversion between U and I is fast compared to the rate-limiting I to N folding step. Scheme 1 predicts two folding phases, a kinetically unresolved fast phase and an observable slow phase. In the pre-equilibrium limit ($k_{UI} + k_{IU} \gg k_{IN}$), the fast phase has a rate given by $\lambda_1 = k_{UI} + k_{IU}$, and the corresponding amplitude accounts for the burst-phase phenomenon (see, e.g., refs 10 and 22). The rate of the slow observable phase can be approximated by eq 3

$$\lambda_2 = f_I k_{IN} + k_{NI} \quad (3)$$

where f_I represents the fractional concentration of the intermediate, which can be expressed as

$$f_I = k_{UI}/(k_{IU} + k_{UI}) = K_{UI}/(1 + K_{UI}) \quad (4)$$

$K_{UI} = k_{UI}/k_{IU}$ is the pre-equilibrium constant of the U ⇌ I transition. Under stabilizing conditions (low [GuHCl]), f_I becomes nearly unity, and the observable rate approaches k_{IN}, which limits the rate of formation of N. If k_{IN} has little denaturant dependence, the observable rate will level off, resulting in a curved or kinked rate profile. As the concentration of GuHCl increases, the intermediate is destabilized and f_I becomes smaller, resulting in a decrease in the observable folding rate.

Table 2: Kinetic Parameters for a Sequential Three-State (U ⇌ I ⇌ N) Model Describing GB1 Folding

protein	[Na_2SO_4] (M)	$K_{UI}°$	m_{UI}	$k_{IN}°$	$m_{IN}^{‡}$	$k_{NI}°$	$m_{NI}^{‡}$	α_I[a]
GB1−57	0.0	2.5	1.35	450	−0.20	0.23	0.31	0.73
	0.4	28.6	1.50	800	−0.30	0.14	0.30	0.71
GB1−66	0.0	2.5	1.35	450	−0.25	0.37	0.27	0.72
	0.4	28.6	1.45	600	−0.30	0.25	0.27	0.72

[a] Relative compactness of I defined as $\alpha_I = m_{UI}/m_{eq}$, $m_{eq} = m_{UI} - m_{IN}^{‡} + m_{NI}^{‡}$.

In Figures 3 and 4, the calculated rate for the observable phase, λ_2, and the relative fluorescence at equilibrium and at time zero are plotted as solid lines. Since the U to I transition occurs within the dead time of the stopped-flow measurement, we cannot resolve individual rate constants, but the pre-equilibrium constant, K_{UI}, can be measured. To calculate the kinetic fluorescence amplitude, the relative fluorescence emission for I was assumed to be around 95% of the fluorescence intensity of N, which fits the observable burst-phase behavior well.

The kinetic parameters corresponding to the best simultaneous fit for all rates, kinetic amplitudes, and equilibrium data (Figures 3 and 4) are listed in Table 2. As a quantitative measure of the compactness of I, it is useful to consider α_I, the ratio of the *m* value for the pre-equilibrium transition, m_{UI}, relative to the equilibrium *m* value, m_{eq} ($m_{eq} = m_{UI} - m_{IN}^{‡} + m_{NI}^{‡}$), which is included in Table 2. Both GB1−57 and GB1−66 in the presence and absence of sodium sulfate exhibit α_I values of 0.71−0.73, indicating that over 70% of the solvent-accessible surface area of U is buried upon formation of I (i.e., I is fairly compact). The small negative values of $m_{IN}^{‡}$ (−0.2 to −0.3) indicate that the transition state for folding is only slightly more compact than I. The rate of unfolding also depends only mildly on [GuHCl] ($m_{NI}^{‡} \sim 0.3$), indicating that the transition state of unfolding is similar to the native state in terms of solvent-accessible surface area. Addition of sodium sulfate stabilizes I relative to U (~10-fold increase in K_{UI}) but has little effect on the rate-limiting step, k_{IN}, which is consistent with the small $m_{IN}^{‡}$ values observed (Table 2). Moreover, the fact that α_I remains essentially the same in the presence of 0.4 M Na_2SO_4 indicates that the stabilizing salt has no significant effect on the solvent-accessible surface area of I.

Interestingly, GB1−66 exhibits virtually the same kinetic behavior as GB1−57. As shown in Table 2, slight adjustment of a few kinetic parameters is sufficient to reproduce the GB1−66 folding kinetics. In particular, K_{UI} is identical under both solvent conditions studied. Despite its substantial hydrophobicity, the N-terminal extension in GB1−66 does not contribute to the stability of I, indicating that it does not contribute to chain collapse and formation of a hydrophobic core.

DISCUSSION

Early Intermediate in GB1 Folding. Recent kinetic folding studies on a number of proteins, including the N-terminal domain of phosphoglycerate kinase (*21*), ribonuclease A (*20*), ubiquitin variants (*10*), and ribonuclease H (*24*), were

interpreted in terms of a minimal three-state mechanism with an obligatory early folding intermediate. The key features in kinetic folding experiments indicating accumulation of an early folding intermediate are (a) a downward curvature or kink in the rate profile and (b) significant loss of signal in the dead time of the kinetic measurement (burst phase). The stopped-flow fluorescence measurements for the GB1-folding reaction showed both of these effects, and thus demonstrate the existence of an early folding intermediate. This is especially remarkable in view of the small size of GB1, since a number of small proteins (typically less than 100 residues) were well described by a two-state folding mechanism (see, e.g., ref 18). With only 56 residues in its globular region, GB1 is the smallest protein with a detectable folding intermediate. Thus, kinetic complexities in folding are apparently not directly related to the size of a protein. Our observation of a well-populated intermediate under favorable conditions (e.g., in 0.4 M Na_2SO_4 at low denaturant concentrations) confirms our earlier prediction (*10*) that some of the proteins with apparent two-state folding behavior may conform to a three-state mechanism under sufficiently stabilizing conditions.

Kinetic Mechanism of GB1 Folding. A major problem in using empirical kinetic schemes to describe folding reactions concerns the connectivity of states. In addition to the sequential folding mechanism with an obligatory intermediate (Scheme 1), one can consider two alternative three-state mechanisms, an off-pathway mechanism and a triangular mechanism. In the off-pathway mechanism ($I \rightleftharpoons U \rightleftharpoons N$), the native state cannot be reached directly from I without going through U. The triangular mechanism is composed of two parallel processes, a direct transition from U to N and a sequential pathway from U to N via I (*41*). Although formal kinetic analysis cannot readily discriminate between these alternative schemes, they represent very different physical situations (*11*). The triangular mechanism suggests that there are two distinct populations of unfolded species that can fold along either of two pathways. This appears unlikely in the case of GB1 because of its lack of prolines and disulfides as a potential source of heterogeneity. Furthermore, NMR analysis of GB1 in 8 M urea showed no evidence of conformational heterogeneity (*42*). The off-pathway mechanism implies that the intermediate contains non-native interactions, which have to be disrupted before folding can proceed. The random collapse of the chain discussed above (model A) is a possible example of such a nonproductive event. Two observations argue against such a mechanism. First, using two-dimensional NMR measurements coupled with H/D exchange labeling, Kuszewski et al. (*33*) observed weak but measurable protection (protection factors of 3−5) for many amide protons in GB1 and more substantial protection (protection factors of 5−8) for a subset (K4, A20, T25, A26, T51, K50, and F52). This observation is a clear indication of rather specific conformational preferences within an ensemble of compact states. In a nonspecifically collapsed state, one would expect a more uniform distribution of protection factors, in contrast to the observed pattern of preferential protection of certain core residues (*33*). Second, our observation of very similar kinetic behavior for GB1−57 and GB1−66 and, in particular, identical pre-equilibrium constants for formation of I (Table 2) indicates that the hydrophobic residues in the N-terminal extension are excluded from the core of I. Thus, the first detectable intermediate already has a solvent-shielded hydrophobic core, which is clearly inconsistent with a disordered collapsed state formed as a result of nonspecific hydrophobic interactions. In contrast, a general hydrophobic collapse would predict a correlation between overall hydrophobicity and the stability of compact intermediates, which is clearly not observed.

Previous equilibrium (*31*) and kinetic (*32*) measurements of GB1 folding, including differential scanning calorimetric (DSC) measurements, were consistent with a two-state folding mechanism without any detectable intermediates. The ratio of van't Hoff enthalpy (ΔH_{vH}) and calorimetric enthalpy (ΔH_{cal}) was close to unity which represents a fairly stringent criterion for a two-state mechanism (*43*). Stopped-flow measurements of the kinetics of refolding at pH 11.2 from the alkaline-denatured state (monitored by absorbance changes of aromatic side chains) showed no indications of populated intermediates and were interpreted as a two-state process (*32*). However, these observations do not contradict the three-state model proposed in this study. First, the kinetic folding intermediate accumulates transiently with significant population only at low denaturant concentrations. At the midpoint of the equilibrium unfolding transition, its population is less than 0.05%, far below detectable levels in equilibrium DSC measurements. Second, marginally stable intermediates may not give rise to a significant burst phase or curvature in the rate profile, and would be difficult to detect. In fact, the kinetic measurements by Alexander et al. (*32*) were performed at pH values above 11 where the native state of GB1 is significantly destabilized, and intermediates are also expected to be unstable, thus explaining the apparent two-state behavior.

Comparison of the Folding Mechanisms for GB1 and LB1. It is interesting to compare our results on GB1 with a kinetic folding study on the B1 domain of peptostreptococcal protein L, LB1 (*44*), given the fact that these two proteins have very similar three-dimensional structures but show very little sequence homology (~15% identity). Under conditions where the two proteins have similar stabilities, the rates of folding and unfolding for LB1 are 10-fold slower than those of GB1. Scalley et al. (*44*) found no clear evidence of a populated intermediate in LB1 (although a minor burst phase was observed in the presence of NaI as a quenching reagent), and they concluded that LB1 follows a two-state folding mechanism. These differences in folding behavior between GB1 and LB1 are perhaps not surprising, considering their diverse sequences. However, as mentioned above, a three-state folding mechanism can be reduced to a two-state mechanism when the intermediate is not stable enough to accumulate. Thus, it will be important to identify the interactions missing in LB1 that stabilize the compact intermediate in GB1.

Conclusions. Although previous equilibrium and kinetic folding studies of GB1 suggested a two-state mechanism without populated intermediates (*31*, *32*), our kinetic results clearly demonstrate rapid formation of an ensemble of early folding intermediates with a largely solvent-shielded tryptophan. With only 57 amino acids, GB1 is the smallest protein with a prominent early folding intermediate described so far, indicating that kinetic complexities are not directly related to the size of a protein. The GuHCl-dependent kinetics of folding and unfolding can be modeled quantitatively, using a three-state mechanism with an obligatory

intermediate. In terms of its fluorescence properties and denaturant dependence (*m* value analysis), the intermediate appears highly compact with over 70% of the initial solvent-exposed surface area buried. Our observation that the pre-equilibrium is unaffected by addition of an N-terminal extension with five apolar residues implies that the core residues of the native structure are already sequestered during the initial stages of folding, which argues against a mechanism involving a nonspecific hydrophobic collapse.

ACKNOWLEDGMENT

We thank W. Colón and M. C. R. Shastry for critical reading of the manuscript. We also thank J. M. Sauder for helpful advice on kinetic modeling and S. Seeholzer for mass spectrometry.

REFERENCES

1. Kim, P. S., and Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.
2. Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653−683.
3. Ptitsyn, O. B. (1995) *Adv. Protein Chem. 47*, 83−229.
4. Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol. 3*, 84−91.
5. Roder, H., and Elöve, G. A. (1994) in *Mechanisms of Protein Folding: Frontiers in Molecular Biology* (Pain, R. H., Ed.) pp 26−55, Oxford University Press, New York.
6. Evans, P. A., and Radford, S. E. (1994) *Curr. Opin. Struct. Biol. 4*, 100−106.
7. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., and Nagamura, T. (1987) *FEBS Lett. 221*, 115−118.
8. Elöve, G. A., Chaffotte, A. F., Roder, H., and Goldberg, M. E. (1992) *Biochemistry 31*, 6876−6883.
9. Matouschek, A., Kellis, J. T. J., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
10. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) *Nat. Struct. Biol. 3*, 193−205.
11. Roder, H., and Colón, W. (1997) *Curr. Opin. Struct. Biol. 7*, 15−28.
12. Creighton, T. E. (1997) *Trends Biochem. Sci. 22*, 6−11.
13. Ptitsyn, O. B. (1994) *Protein Eng. 7*, 593−596.
14. Sosnick, T. R., Mayne, L., and Englander, S. W. (1996) *Proteins 24*, 413−426.
15. Oliveberg, M., and Fersht, A. R. (1996) *Biochemistry 35*, 2738−2749.
16. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
17. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
18. Villegas, V., Azuaga, A., Catasús, L., Reverter, D., Mateo, P. L., Avilés, F. X., and Serrano, L. (1995) *Biochemistry 34*, 15105−15110.
19. Khorasanizadeh, S., Peters, I. D., Butt, T. R., and Roder, H. (1993) *Biochemistry 32*, 7054−7063.
20. Houry, W. A., Rothwarf, D. M., and Scheraga, H. A. (1995) *Nat. Struct. Biol. 2*, 495−503.
21. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.
22. Sauder, J. M., MacKenzie, N. E., and Roder, H. (1996) *Biochemistry 35*, 16852−16862.
23. Walkenhorst, W. F., Green, S. M., and Roder, H. (1997) *Biochemistry 36*, 5795−5805.
24. Raschke, T. M., and Marqusee, S. (1997) *Nat. Struct. Biol. 4*, 298−304.
25. Fahnestock, S. R., Alexander, P., Nagle, J., and Filpula, D. (1986) *J. Bacteriol. 167*, 870−880.
26. Guss, B., Eliasson, M., Olsson, A., Uhlen, M., Frej, A.-K., Jornvall, H., Flock, I., and Lindberg, M. (1986) *EMBO J. 5*, 1567−1575.
27. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) *Science 253*, 657−661.
28. Achari, A., Hale, S. P., Howard, A. J., Clore, G. M., Gronenborn, A. M., Hardman, K. D., and Whitlow, M. (1992) *Biochemistry 31*, 10449−10457.
29. Lian, L.-Y., Derrick, J. P., Sutcliffe, M. J., Yang, J. C., and Roberts, G. C. K. (1992) *J. Mol. Biol. 228*, 1219−1234.
30. Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G. L. (1994) *Biochemistry 33*, 4721−4729.
31. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry 31*, 3597−3603.
32. Alexander, P., Orban, J., and Bryan, P. (1992) *Biochemistry 31*, 7243−7248.
33. Kuszewski, J., Clore, G. M., and Gronenborn, A. M. (1994) *Protein Sci. 3*, 1945−1952.
34. O'Neil, K. L., Hoess, R. H., Raleigh, D. P., and DeGrado, W. F. (1995) *Proteins 21*, 11−21.
35. Pace, C. N. (1986) *Methods Enzymol. 131*, 266−280.
36. Pace, C. N. (1975) *CRC Crit. Rev. Biochem. 2*, 1−43.
37. Berberan-Santos, M. N., and Martinho, J. M. G. (1990) *J. Chem. Educ. 67*, 375−379.
38. Pogliani, L., and Terenzi, M. (1992) *J. Chem. Educ. 69*, 278−280.
39. Goward, C. R., Irons, L. I., Murphy, J. P., and Atkinson, T. (1991) *Biochem. J. 274*, 503−507.
40. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci. 4*, 2138−2148.
41. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9029−9033.
42. Frank, M. K., Clore, G. M., and Gronenborn, A. M. (1995) *Protein Sci. 4*, 2605−2615.
43. Privalov, P. L. (1979) *Adv. Protein Sci. 33*, 167.
44. Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R. I., and Baker, D. (1997) *Biochemistry 36*, 3373−3382.

BI971914+