

# Characterization of the Folding and Unfolding Reactions of Single-Chain Monellin: Evidence for Multiple Intermediates and Competing Pathways<sup>†</sup>

Ashish K. Patra and Jayant B. Udgaonkar\*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065, India

Received June 11, 2007; Revised Manuscript Received August 7, 2007

**ABSTRACT:** The mechanisms of folding and unfolding of the small plant protein monellin have been delineated in detail. For this study, a single-chain variant of the natively two-chain monellin, MNEI, was used, in which the C terminus of chain B was connected to the N terminus of chain A by a Gly-Phe linker. Equilibrium guanidine hydrochloride (GdnHCl)-induced unfolding experiments failed to detect any partially folded intermediate that is stable enough to be populated at equilibrium to a significant extent. Kinetic experiments in which the refolding of GdnHCl-unfolded protein was monitored by measurement of the change in the intrinsic tryptophan fluorescence of the protein indicated the accumulation of three transient partially structured folding intermediates. The fluorescence change occurred in three kinetic phases: very fast, fast, and slow. It appears that the fast and slow changes in fluorescence occur on competing folding pathways originating from one unfolded form and that the very fast change in fluorescence occurs on a third parallel pathway originating from a second unfolded form of the protein. Kinetic experiments in which the refolding of alkali-unfolded protein was monitored by the change in the fluorescence of the hydrophobic dye 8-anilino-1-naphthalenesulfonic acid (ANS), consequent to the dye binding to the refolding protein, as well as by the change in intrinsic tryptophan fluorescence, not only confirmed the presence of the three kinetic intermediates but also indicated the accumulation of one or more early intermediates at a few milliseconds of refolding. These experiments also exposed a very slow kinetic phase of refolding, which was silent to any change in the intrinsic tryptophan fluorescence of the protein. Hence, the spectroscopic studies indicated that refolding of single-chain monellin occurs in five distinct kinetic phases. Double-jump, interrupted-folding experiments, in which the accumulation of folding intermediates and native protein during the folding process could be determined quantitatively by an unfolding assay, indicated that the fast phase of fluorescence change corresponds to the accumulation of two intermediates of differing stabilities on competing folding pathways. They also indicated that the very slow kinetic phase of refolding, identified by ANS binding, corresponds to the formation of native protein. Kinetic experiments in which the unfolding of native protein in GdnHCl was monitored by the change in intrinsic tryptophan fluorescence indicated that this change occurs in two kinetic phases. Double-jump, interrupted-unfolding experiments, in which the accumulation of unfolding intermediates and native protein during the unfolding process could be determined quantitatively by a refolding assay, indicated that the fast unfolding phase corresponds to the formation of fully unfolded protein via one unfolding pathway and that the slow unfolding phase corresponds to a separate unfolding pathway populated by partially unfolded intermediates. It is shown that the unfolded form produced by the fast unfolding pathway is the one which gives rise to the very fast folding pathway and that the unfolded form produced by the slower unfolding pathway is the one which gives rise to the slow and fast folding pathways.

Much of the thrust of recent research in protein folding has been directed toward understanding the folding of proteins that fold apparently in a two-state manner without any detectable accumulation of folding intermediates (1–3). As a consequence, much has been learned about the nature of the kinetic barrier that separates unfolded and folded proteins and how it is determined, at least in part, by structural parameters of the folded protein such as topology (4, 5). However, another consequence, because it seemed

that these proteins may not require intermediates to fold, has been an erosion of the importance ascribed previously to folding intermediates in determining a protein folding pathway. This has happened even though for many of the apparent two-state folders, it is likely that intermediates are indeed populated transiently but that they cannot be detected because either they are too unstable or they follow the rate-limiting step in folding. Indeed, in the case of several two-state folders, a change in folding conditions can be used to stabilize unstable folding intermediates sufficiently so that they are populated to detectable extent (6); for other proteins it is possible to destabilize folding intermediates sufficiently so that they can no longer be detected and folding consequently appears to be two-state (7). Examining the roles

<sup>†</sup> This work was funded by the Tata Institute of Fundamental Research and by the Department of Science and Technology, Government of India.

\* Corresponding author (e-mail jayant@ncbs.res.in; telephone 91-80-3666150; fax 91-80-23666150).

played by partially folded intermediates in determining the development of structure during folding has become more important because folding intermediates may not be discrete states but may be ensembles of different structural conformations of similar energy, and the structural component that predominates in the intermediate ensemble may depend on the folding conditions (8–11). Experimental observations that some aspects of folding reactions such as the initial polypeptide chain collapse may be gradual structural transitions (12) and may also be nonspecific in nature (13, 14) and that initial intermediates observed during folding may be misfolded (15–17) have questioned the relevance of folding intermediates. The experimental demonstration that folding, as well as unfolding, may occur via multiple pathways populated by structurally distinct intermediates on an energy landscape (18–24) and the observations that the pathway utilized may depend on the starting conditions (10, 25) as well as the conditions during folding have also brought into focus the necessity for understanding the roles played by intermediates. This necessity has been reinforced by the view derived from statistical models of protein folding (26–28) that intermediates are not only unnecessary for folding to occur but may actually serve as kinetic traps that impede folding.

Although delineating the roles of partly folded intermediate structures in specifying the utilization of competing protein folding and unfolding pathways is crucial, determining their role in protein aggregation is turning out to be equally important. Proteins very often aggregate transiently during folding (29–31) as well as unfolding (32) because of the self-association of folding or unfolding intermediates, and in some cases the aggregation is specific, leading to the formation of amyloid protofibrils (33, 34). In the case of some proteins, amyloid formation is linked directly to disease (35). It has hence become important to delineate the roles of intermediates not only during folding and unfolding but also as the starting points for aggregation leading to amyloid formation. For this, it is essential to study a protein that is well suited to serve as a model protein for studying the mechanism of protein folding as well as protein aggregation. One such protein is the small protein monellin.

Monellin is an intensely sweet protein, isolated originally from the berry of an African plant, *Dioscoreophyllum cuminisii* (36). It is well suited as a model protein for studying the mechanism of protein folding because it unfolds and refolds reversibly in both its naturally occurring heterodimeric form and in artificially created monomeric variants. Hence, it is attractive as a model protein not only for the study of the folding of a monomeric protein (37) but also for the study of how multimeric proteins fold (38, 39), for the study of the folding and unfolding of a protein in the absence of the potentially complicating presence of chemical denaturants, and for the study of how chaperones facilitate the complementation of polypeptide chains when assisting the folding of a multimeric protein. It is also well suited as a model protein for studying the mechanism of protein aggregation (40, 41).

Naturally occurring heterodimeric monellin consists of a chain A of 45 amino acids, comprising three antiparallel  $\beta$ -strands, and a chain B of 50 amino acids, comprising one long  $\alpha$ -helix intersecting with two antiparallel  $\beta$ -strands. The five  $\beta$ -strands form a single sheet (42). The two chains are

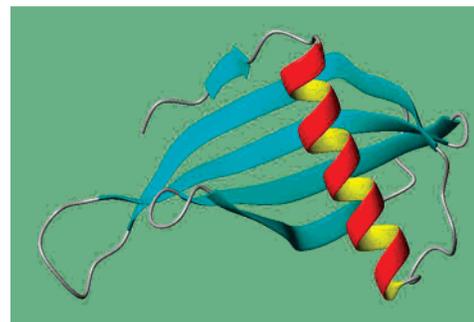


FIGURE 1: Structure of single-chain monellin (MNEI). The structure has been drawn using entry **1FA3** in the PDB, using the program Molmol.

held together mainly by interchain hydrophobic interactions, H-bonds and salt bridges (43). The sweet-taste determinant of monellin appears to be located in the loop containing  $\beta$ -strands 4 and 5, corresponding to residues 16–35 of the A chain (42, 44). The structure of monellin is similar to that of the cystatin family of thiol proteinase inhibitors (45). Its fold, the so-called  $\beta$ -grasp fold, is one of the 10 most common protein folds, which is shared not only by the cystatins but also by ubiquitin, the B1 domain of protein G, and chymotrypsin inhibitor II.

Two single-chain variants of monellin have been shown to be similar structurally to the naturally occurring dimeric protein (43) (Figure 1). In one variant, the C terminus of the B chain has been linked directly to the N terminus of the A chain (46), whereas in the other variant (MNEI), a Gly-Phe dipeptide has been used to link the two termini (44). Both single-chain variants appear to be functionally as sweet as the naturally occurring two-chain protein. The single-chain variants of monellin unfold reversibly (37) and are more stable than the naturally occurring two-chain protein (44, 46). They can be unfolded completely in GdnHCl, and they refold completely back to the native state upon removal of denaturant. The mechanism of folding remains, however, unknown.

In this study, the mechanism of refolding of the GdnHCl-unfolded single-chain monellin (MNEI), as well as the mechanism of unfolding by GdnHCl, has been characterized in detail. The change in fluorescence accompanying the refolding of the protein is shown to occur in three kinetic phases—very fast, fast, and slow. Double-jump, interrupted refolding experiments indicate that four parallel folding pathways are operative from two unfolded forms of the protein and that native protein forms in a single, very slow kinetic phase that is silent to fluorescence change. It is shown that some of this heterogeneity arises due to the occurrence of proline isomerization in the unfolded protein. Interestingly, the heterogeneity seen in the fast phase of folding cannot be ascribable to proline isomerization, but arises, instead, because two folding pathways operate in competition with each other. It is shown to arise because of kinetic partitioning during folding from a very early folding intermediate. The change in fluorescence accompanying the unfolding of the protein is shown to occur in two kinetic phases, and double-jump, interrupted unfolding experiments indicate the existence of two parallel unfolding pathways leading to the two unfolded forms of the protein. Unfolding intermediates can be detected on the slower of the two unfolding pathways.

## MATERIALS AND METHODS

**Materials.** All chemicals and reagents were obtained from Sigma (unless otherwise mentioned) and were of the highest purity grade.

**Purification of Single-Chain Monellin.** The gene for MNEI was subcloned into the pET22b+ vector between the sites of *NdeI* and *HindIII*. The amino acid sequence of the protein expressed in the *Escherichia coli* BL 21 DE3\* strain is MGEWEIIDIGPFTQNLGKFAVDEEN-KIGQYGRLTFNKVIRPCMKKTIYENEG-FREIKGYEYQLVYVYASDKLFRADISEDYK-TRGRKLLRFNGPVPPP. The underlined residues are found in chain B of two-chain monellin. The C terminus of chain B is attached to the N terminus of chain A through a Gly-Phe dipeptide linker. Cells were grown overnight after induction with 0.25 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). The purification protocol used has been described previously (47). Very briefly, the harvested cells were broken by sonication. This was followed by ion exchange chromatography on an SP-Sepharose column and then by gel filtration on a G-75 column.

The purity of the protein was checked on SDS-PAGE and was found to be >99% pure. ESI mass spectroscopy indicated a pure protein of mass 11403 Da, which corresponds to the mass of the above sequence. In addition, a second minor (~10%) mass of 11271 Da was observed, which corresponds to the mass of the above sequence with the N-terminal methionine cleaved out. Typically, 50–100 mg of protein was obtained from 1 dm<sup>3</sup> of *E. coli* growth. In all studies, the monellin concentration was determined using the extinction coefficient,  $\epsilon^{277\text{nm}} = 1.46 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$  (36).

**Buffers and Solutions.** The native buffer used for all equilibrium and kinetic experiments was 50 mM sodium phosphate, 250  $\mu\text{M}$  EDTA, and 1 mM DTT at pH 7. The presence of DTT in all buffers ensured that the monellin molecules did not exist as dimers formed through intermolecular disulfides. The unfolding buffer was native buffer containing 4–8 M GdnHCl (ultrapure, 99.9% from USB). The concentrations of stock solutions of GdnHCl were determined by measuring the refractive index using an Abbe 3L refractometer from Milton Roy. All measurements were carried out at room temperature (25 °C). All buffer solutions were passed through 0.22  $\mu\text{m}$  filters and degassed before use.

**Fluorescence Spectra.** Fluorescence spectra were collected on a Spex Fluoromax 3 spectrofluorometer. For intrinsic fluorescence measurements, the protein samples were excited at either 280 or 295 nm and emission spectra were collected from 300 to 400 nm, with a response time of 1 s, an excitation bandwidth of 0.3 nm, and an emission bandwidth of 5 nm. Measurements were made using protein concentrations of 4  $\mu\text{M}$ , and each spectrum was recorded as the average of five fluorescence emission wavelength scans.

**Circular Dichroism (CD) Spectra.** CD spectrum measurements were carried out on a Jasco J 720 spectropolarimeter. Far-UV CD spectra were collected with a bandwidth of 1 nm, a response time of 1 s, and a scan speed of 20 nm min<sup>-1</sup>. The protein concentration used was 20  $\mu\text{M}$  in a 0.1 cm path length cuvette. Thirty wavelength scans were averaged for each spectrum. Near-UV CD spectra were collected with

a bandwidth of 1 nm, a response time of 1 s, and a scan speed of 20 nm min<sup>-1</sup>. Measurements were made using protein concentrations of 100  $\mu\text{M}$  in a 1 cm path length cuvette. Fifty wavelength scans were averaged for each spectrum.

**Equilibrium Unfolding Experiments.** Equilibrium unfolding of monellin was monitored using fluorescence, far-UV CD, and near-UV CD as probes. Protein (8–10  $\mu\text{M}$  for fluorescence, 8–15  $\mu\text{M}$  for far-UV CD, and 65–70  $\mu\text{M}$  for near-UV CD) was incubated in different concentrations of GdnHCl ranging from 0 to 6 M, for 6 h. The equilibrium fluorescence signals were measured on the stopped-flow module (SFM-4) from Biologic, with which the kinetic experiments were also done. The protein sample was excited at 280 nm, and fluorescence emission was monitored using a 340 nm band-pass filter. For each point, the signal was averaged for 120 s. During the monitoring of equilibrium unfolding using far-UV CD at 222 nm as the probe, a 0.2 cm path length cuvette was used, and the signal was averaged for 30 s. During the monitoring of equilibrium unfolding using near-UV CD at 270 nm as the probe, a 1 cm path length cuvette was used, and the signal was averaged for 60 s.

**Kinetic Experiments.** Rapid (milliseconds) mixing of solutions and the observation of kinetic processes during protein folding or unfolding reactions were done using a Biologic SFM-4 machine. Mixing dead-times of the order of 6.2 ms were obtained, using a cuvette of 0.15 cm path length, with flow rates of 5 mL s<sup>-1</sup>. The final protein concentration in the cuvette ranged from 8 to 12  $\mu\text{M}$ . Excitation was at 280 nm, and fluorescence emission was measured using a 340 nm band-pass filter. Data were acquired in three time domains on different channels, with different sampling times for each domain. For refolding experiments, monellin was first unfolded for 6 h in unfolding buffer containing 4 M GdnHCl. Thirty microliters of equilibrium-unfolded protein was diluted into 270  $\mu\text{L}$  of premixed native and unfolding buffers, so that the final GdnHCl concentration was in the range of 0.3–1.5 M. For unfolding experiments, monellin was equilibrated in native buffer prior to use. Thirty microliters of equilibrated native protein at pH 7 was diluted into 270  $\mu\text{L}$  of premixed unfolding and native buffers, so that the final GdnHCl concentration was in the range of 3.5–6 M.

**pH Jump-Induced Refolding in the Absence of Denaturant.** Monellin was first unfolded by adding 1 volume of 0.1 M NaOH to 2 volumes of monellin in 50 mM phosphate buffer, pH 7, and adjusting the final pH to 12. Refolding was initiated by mixing 1 volume of unfolded protein at pH 12 with 6 volumes of refolding buffer (50 mM sodium phosphate, pH 6.8) in the Biologic SFM-4 module, so that the final pH was 7. A mixing dead-time of 1.5 ms was achieved by using a cuvette of 0.08 cm path length and a flow rate of 6 mL s<sup>-1</sup>. The final protein concentration in the cuvette was ~10  $\mu\text{M}$ . When refolding was monitored by measurement of intrinsic tryptophan fluorescence, the fluorescence was excited at 280 nm and emission was measured using a 340 nm band-pass filter.

When refolding was monitored by measurement of ANS fluorescence, the refolding buffer at pH 6.8 contained ~45  $\mu\text{M}$  ANS. The final concentration of the protein during refolding was 4  $\mu\text{M}$ , and that of ANS was 40  $\mu\text{M}$ . ANS

fluorescence was excited at 350 nm, and emission was measured using a 450 nm band-pass filter.

**Double-Jump Assay for the Formation of the N State during Folding in 0.5 M GdnHCl.** A Biologic SFM-4 mixing module was used to mix 30  $\mu\text{L}$  of equilibrium-unfolded protein in 4 M GdnHCl buffer with 220  $\mu\text{L}$  of native buffer, such that the final GdnHCl concentration was 0.5 M. The refolding mixture was aged for different lengths of time in a delay loop of 190  $\mu\text{L}$  volume (the intermixer volume was 220  $\mu\text{L}$ ). After different refolding times (from 300 ms to 3600 s), 50  $\mu\text{L}$  of the solution in the delay loop was mixed with 305  $\mu\text{L}$  of unfolding buffer containing 6 M GdnHCl so that the final GdnHCl concentration during unfolding was 5 M. The final protein concentration was 5  $\mu\text{M}$  in a 0.15 cm path length cuvette. A dead-time of 25 ms was obtained using flow rates of 1.4  $\text{mL s}^{-1}$ . The unfolding reaction after each time of refolding was monitored by measurement of fluorescence emission at 340 nm. The amount of native protein formed at each time of refolding was determined from measurement of the amplitude of the slowest phase of unfolding, which occurred at a rate of  $0.011 \pm 0.002 \text{ s}^{-1}$ . The amounts of two folding intermediates formed at each time of refolding were determined from measurement of the amplitudes of the two faster kinetic phases of unfolding, which had observed rates of  $4.5 \pm 1$  and  $0.2 \pm 0.04 \text{ s}^{-1}$ . The fractions of N and the two folding intermediates formed at any time were determined as the amplitudes of the corresponding kinetic phases of unfolding, relative to the amplitude of the total amplitude of the unfolding reaction initiated when the refolding reaction was allowed to proceed for 3600 s.

**Double-Jump Assay for the Formation of U during Unfolding in 5 M GdnHCl.** A Biologic SFM-400 mixing module was used to mix 80  $\mu\text{L}$  of equilibrated native protein solution with 140  $\mu\text{L}$  of unfolding buffer containing 7.85 M GdnHCl such that the final GdnHCl concentration was 5 M. The unfolding protein solution was aged in a delay loop of 190  $\mu\text{L}$  (the intermixer volume was 220  $\mu\text{L}$ ). After different times of unfolding (from 5 to 800 s), 50  $\mu\text{L}$  of the unfolding protein in the delay loop was mixed with 450  $\mu\text{L}$  native buffer so that the final GdnHCl concentration was 0.5 M. The final protein concentration was 5  $\mu\text{M}$  in a 0.15 cm path length cuvette. The mixing dead-time was 11.9 ms with a flow rate of 6.25  $\text{mL s}^{-1}$ . The refolding reaction after each time of unfolding was monitored by measurement of the fluorescence emission at 340 nm. The amplitudes of the three observable kinetic phases of refolding in 0.5 M GdnHCl, which had observed rate constants of  $7 \pm 1$ ,  $1 \pm 0.3$ , and  $0.1 \pm 0.02 \text{ s}^{-1}$ , were determined at each time of unfolding. Their relative amplitudes were determined by dividing the individual refolding amplitudes by the total amplitude of the refolding reaction initiated after the unfolding had been allowed to proceed for 800 s.

The double-jump, interrupted unfolding experiment was repeated by carrying out the refolding assay in 1.5 M GdnHCl rather than in 0.5 M GdnHCl. In 1.5 M GdnHCl, the refolding kinetics are simpler, and hence it is possible to identify unfolding intermediates that refold with kinetics that are different from the kinetics starting from U. Both manual mixing and stopped-flow mixing experiments were carried out. The manual mixing experiments were done by mixing 200  $\mu\text{L}$  of protein in native buffer with 475  $\mu\text{L}$  of

7.1 M GdnHCl so that the GdnHCl concentration became 5 M. After different times of unfolding, 1580  $\mu\text{L}$  of native buffer was added so that the final concentration of GdnHCl became 1.5 M. The final protein concentration was 2–4  $\mu\text{M}$  in a 1 cm path length cuvette, in which rapid mixing was ensured by a magnetic stirrer. The stopped-flow mixing experiments were done on an SFM-400 module, for the initial time points of unfolding (from 200 ms to 10 s). For this, 70  $\mu\text{L}$  of protein was mixed with 150  $\mu\text{L}$  of unfolding buffer containing 7.35 M GdnHCl in a 190  $\mu\text{L}$  loop (the intermixer volume was 220  $\mu\text{L}$ ) so that the final GdnHCl was 5 M. After different times of unfolding (from 200 ms to 10 s), 70  $\mu\text{L}$  of the solution from the delay loop was mixed with 162  $\mu\text{L}$  of native buffer so that the final GdnHCl concentration became 1.5 M. The final protein concentration was 6–10  $\mu\text{M}$  in a 0.15 cm path length cuvette. The mixing dead-time was 8 ms with a flow rate of 6.44  $\text{mL s}^{-1}$ . Refolding was observed after different times of unfolding. At the initial time points of unfolding (from 200 ms to 10 s), only a decrease in the fluorescence intensity with a rate of  $0.014 \pm 0.001 \text{ s}^{-1}$  was observed. At intermediate time points of unfolding (from 20 to 50 s), there was first a decrease and then an increase in the fluorescence intensity, with rates of  $0.014 \pm 0.001$  and  $0.0024 \pm 0.0002 \text{ s}^{-1}$ , respectively. At later time points (75–500 s), two kinetic phases of increasing fluorescence intensity at rates of  $0.025 \pm 0.001$  and  $0.0024 \pm 0.0002 \text{ s}^{-1}$  were observed. The fractions of U and the intermediate formed at any time was determined from the amplitudes of the increasing and decreasing phases of fluorescence change, respectively, relative to the amplitude of refolding phase initiated when the unfolding was first allowed to proceed for 500 s.

**Data Analysis.** (a) *Equilibrium Unfolding Data.* Equilibrium unfolding transitions were analyzed using a two-state  $\text{N} \leftrightarrow \text{U}$  model (13) to obtain the values for the free energy of unfolding in water  $\Delta G_{\text{U}}$  and the midpoint of transition ( $C_{\text{m}}$ ).

(b) *Kinetic Refolding and Unfolding Data.* The kinetic traces for refolding monitored by the change in fluorescence were fit to two- or three-exponential equations, as necessary. The kinetic traces for unfolding were fit to a single-exponential equation.

(c) *Interrupted Unfolding Assays.* The kinetic traces for refolding in 0.5 M GdnHCl, after first unfolding in 5 M GdnHCl, were fit to three-exponential equations.

The kinetic traces for refolding in 1.5 M GdnHCl, after first unfolding in 5 M GdnHCl for <20 s, were fit to a single-exponential equation. The kinetic traces for refolding in 1.5 M GdnHCl, after first unfolding in 5 M GdnHCl for times beyond 20 s, were fit to a two-exponential equation.

The formation and decay of intermediates were fit to the equation

$$A_{(t)} = A \left( \frac{\lambda_2}{\lambda_1 - \lambda_2} \right) (e^{-\lambda_1 t} - e^{-\lambda_2 t}) \quad (1)$$

where  $A_{(t)}$  is amplitude at any time  $t$ ,  $A$  is the maximum amplitude of the intermediate phase, and  $\lambda_1$  and  $\lambda_2$  are observed rates of formation and decay of intermediates.

(d) *Interrupted Refolding Assay.* The kinetic traces for unfolding in 5 M GdnHCl, after first refolding in 0.5 M

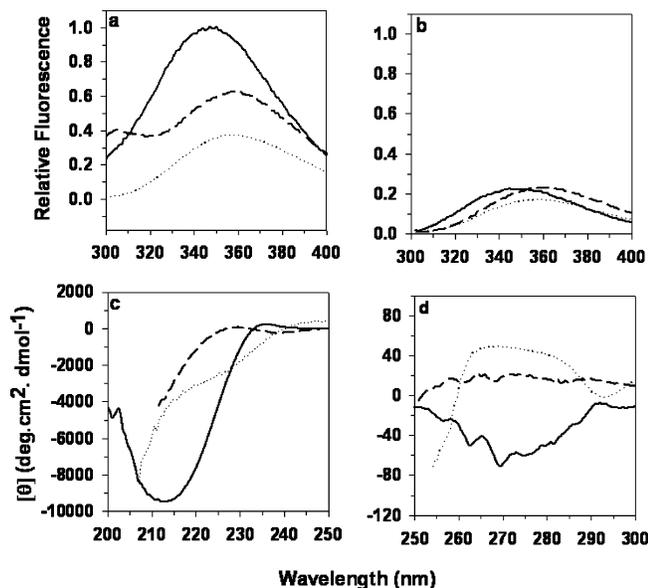


FIGURE 2: Spectroscopic characterization of monellin at pH 7 and 25 °C: (a) fluorescence emission spectra with excitation at 280 nm; (b) fluorescence emission spectra with excitation at 295 nm; (c) far-UV CD spectra; (d) near-UV CD spectra. In all panels, the solid lines indicate the spectra of native protein, the dashed lines indicate spectra of protein unfolded in 6 M GdnHCl, and the dotted lines indicate spectra of protein unfolded at pH 12.

GdnHCl for different periods of time, were fit to a three-exponential equation.

## RESULTS

**Characterization of Single-Chain Monellin, MNEI.** The wavelength of maximum fluorescence emission of native single-chain monellin is 348 nm, whereas that of the unfolded protein is 355 nm (Figure 2a,b). The fluorescence of monellin, which contains seven tyrosine residues and one tryptophan residue in its sequence, is strongly quenched: the relative fluorescence of native single-chain monellin at 310 nm upon excitation at 280 nm is only 60% of that of an equimolar concentration of *N*-acetyltyrosine amide, whereas the relative fluorescence at 348 nm upon excitation at 295 nm is only 20% of that of an equimolar concentration of *N*-acetyltryptophan amide. Upon excitation at 280 nm, the fluorescence decreases upon unfolding in 6 M GdnHCl, because the tyrosine residues are excited principally, and the fluorescence at 355 nm in unfolded protein is low because little energy transfer occurs from the excited tyrosine residues to the sole tryptophan residue: the distances between these residues are larger in the unfolded protein. The poor efficiency of energy transfer in the unfolded state, but not in the folded state, is also evident in the observation that a distinct peak of tyrosine fluorescence is observed only for the unfolded protein and not for the folded protein. Upon excitation at 295 nm, at which wavelength the sole tryptophan residue is excited directly, the fluorescence spectrum undergoes a red shift upon unfolding in GdnHCl, such that the wavelength of maximum emission is shifted from 348 nm for the native protein to 355 nm for the unfolded protein. Figure 2c shows that the secondary structure of monellin, evident as principally a  $\beta$ -sheet in the far-UV CD spectrum, is lost completely upon unfolding in 6 M GdnHCl. Figure 2d shows that the specific tertiary structure of

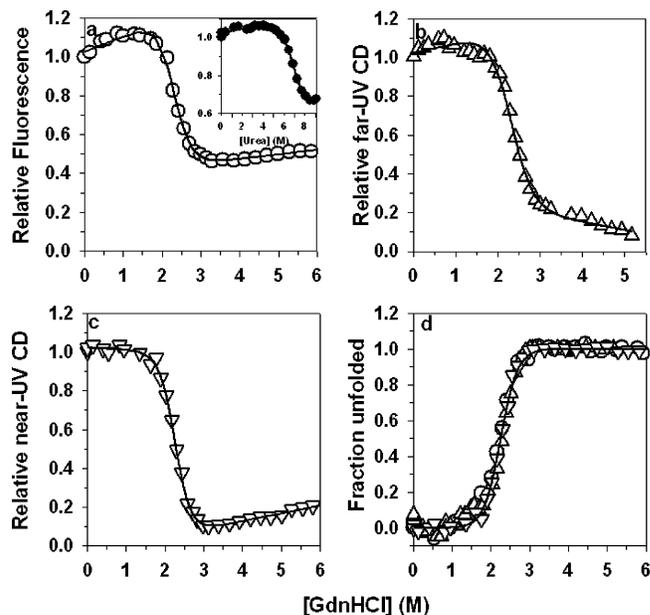


FIGURE 3: Stability of monellin at pH 7, 25 °C: (a) equilibrium GdnHCl-induced unfolding curve determined by monitoring fluorescence emission at 340 nm (the inset shows the equilibrium urea-induced unfolding curve); (b) equilibrium GdnHCl-induced unfolding curve determined by ellipticity at 222 nm; (c) equilibrium GdnHCl-induced unfolding curve determined by monitoring ellipticity at 270 nm; (d) fraction unfolded versus GdnHCl concentration; (O) fluorescence; ( $\Delta$ ) far-UV CD; ( $\nabla$ ) near-UV CD. The solid lines through the data are nonlinear least-squares fit to a two-state  $N \rightleftharpoons U$  model, and yield values for  $\Delta G_{UN}$  and  $m_{UN}$  were 6.2 kcal mol<sup>-1</sup> and 2.7 kcal mol<sup>-1</sup> M<sup>-1</sup>, respectively.

monellin, as seen in the near-UV CD spectrum, is also completely lost upon unfolding in 6 M GdnHCl.

Figure 2a shows the fluorescence spectrum of monellin unfolded at pH 12 upon excitation at 280 nm: no distinct peak of tyrosine fluorescence is seen because the tyrosine residue has ionized to tyrosinate, which is not fluorescent. Monellin also appears to be completely unfolded at pH 12 in the absence of any denaturant: the maximum fluorescence emission at pH 12 occurs at 355 nm upon excitation at 295 nm as it does in 6 M GdnHCl. However, the intensity of the fluorescence of the protein at pH 12 is only 70% of that of the protein in 6 M GdnHCl (Figure 2b). Figure 2c indicates that the extent of secondary structure loss at pH 12, which is the same as that at pH 13 (data not shown), is nearly the same as that seen in 6 M GdnHCl. Similarly, Figure 2d shows that the near-UV CD spectrum obtained at pH 12 is not very different from that obtained in 6 M GdnHCl. Other studies (data not shown) indicated that monellin is not deamidated when incubated at pH 12 for up to 4 h and that the protein at pH 12 is as unfolded as it is at pH 13, as judged from the CD spectra (data not shown).

**Stability of Single-Chain Monellin.** Figure 3a shows the fluorescence-monitored, GdnHCl-induced equilibrium unfolding curve for monellin at pH 7, 25 °C. Fluorescence was excited at 280 nm and monitored at 340 nm, the wavelength at which the native and unfolded states differ most in their intrinsic fluorescence. The equilibrium unfolding curve displays a marked upward curvature at GdnHCl concentrations below those corresponding to the unfolding transition zone. Panels b and c of Figure 3 show the far-UV CD-monitored and near-UV CD-monitored GdnHCl-induced

equilibrium unfolding curve. The fluorescence and CD-monitored equilibrium curves indicate that the protein is half unfolded at a GdnHCl concentration of 2.3 M. The inset of Figure 3a shows the urea-induced equilibrium unfolding curve. Although the data indicate that the protein is not fully unfolded even at the highest concentration of urea used (9.5 M), the urea-induced fluorescence-monitored equilibrium unfolding curve also shows the 15% increase in fluorescence at low urea concentrations, which is seen at low GdnHCl concentrations. Hence, this increase is unlikely to be an effect of ionic strength. In Figure 3d, it is seen that when the data of Figure 3a–c are converted to plots of fraction of protein unfolded versus GdnHCl concentration, they overlap. An analysis of the data in Figure 3d according to a two-state  $N \rightleftharpoons U$  model indicates that the free energy of unfolding is  $6.2 \pm 0.5$  kcal mol<sup>-1</sup>.

**Kinetics of Refolding and Unfolding.** Figure 4a shows the time courses of the increase in fluorescence during refolding in different concentrations of GdnHCl. It is seen that there is no sub-millisecond burst phase change in fluorescence. Each kinetic trace commences from the same point, corresponding to the fluorescence of monellin in 4 M GdnHCl. Each kinetic trace for refolding in low concentrations of GdnHCl is well-described by a three-exponential equation, whereas refolding in concentrations of GdnHCl > 1.25 M is well-described by a two-exponential equation. As expected from the equilibrium unfolding curve of Figure 3a, the final fluorescence value for folding in low concentrations of GdnHCl exceeds the fluorescence of the native state.

Figure 4b shows the time course of the decrease in fluorescence during unfolding in different concentrations of GdnHCl. All kinetic traces are well-described by single-exponential equations, and when extrapolated back to  $t = 0$ , they do so to a fluorescence value that exceeds the fluorescence of the native protein in zero denaturant, by ~20%.

To determine whether the protein equilibrated in 1.5 M GdnHCl was indeed native, albeit with a higher fluorescence value, its unfolding kinetics was determined at different final concentrations of GdnHCl (data not shown). At each concentration of GdnHCl between 3.5 and 5.5 M, the kinetic trace of unfolding commenced at the value expected for protein equilibrated in 1.5 M GdnHCl, and the observed rate constant for unfolding was identical to that observed when folding was commenced from native buffer not containing any GdnHCl.

**pH-Jump-Induced Refolding.** Figure 4c shows the time course of the increase in intrinsic tryptophan fluorescence that accompanies refolding consequent to a jump in pH from 12 to 7. Fluorescence was excited at 280 nm. A sub-millisecond burst phase change in fluorescence is observed, but this can be accounted for by the increase in fluorescence expected when the deprotonated tyrosine residues (tyrosinates) in the protein at pH 12 are protonated to tyrosine at pH 7, immediately upon the jump in pH (see Figure 2a). The subsequent observable refolding reaction occurs in three kinetic phases characterized by apparent rate constants of  $108 \pm 7$ ,  $8 \pm 0.6$ , and  $0.6 \pm 0.2$  s<sup>-1</sup>.

The hydrophobic dye 8-anilino-1-naphthalenesulfonic acid (ANS) binds to exposed hydrophobic patches of partially structured folding intermediates. Upon binding, its fluorescence increases substantially over that of unbound ANS.

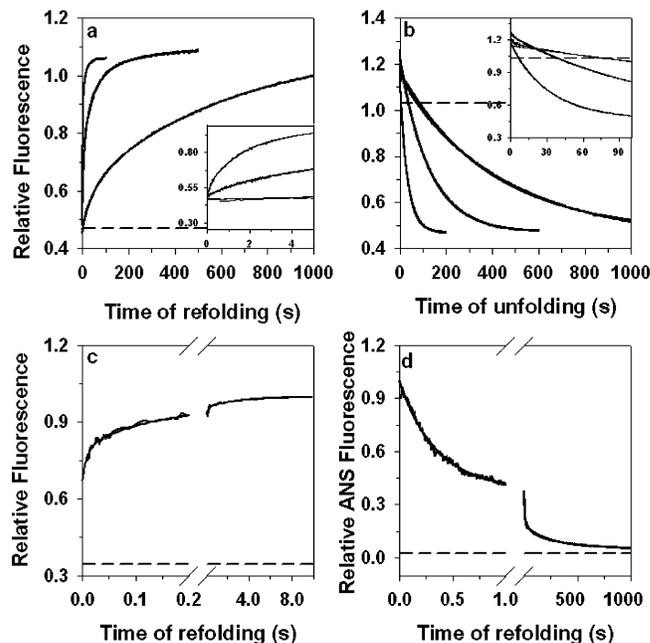


FIGURE 4: Folding and unfolding kinetics of monellin at pH 7, 25 °C. (a) Intrinsic tryptophan fluorescence-monitored kinetic traces of refolding in 0.6, 1, and 1.5 M GdnHCl, shown top to bottom. The solid lines through 0.6 and 1 M data traces represent nonlinear least-squares fits to a three-exponential equation, and the solid line through the 1.5 M data trace represents nonlinear least-squares fits to a two-exponential equation. The dashed line indicates the fluorescence value of protein in 4 M GdnHCl from which refolding was commenced. (b) Intrinsic tryptophan fluorescence-monitored kinetic traces of unfolding in 4, 5, and 6 M GdnHCl, shown top to bottom. The solid lines through the data traces represent nonlinear least-squares fits to a single-exponential equation. The dashed line represents the fluorescence of protein in native buffer in the absence of any denaturant. In both panels, the insets show the initial parts of the kinetic traces. In both panels, and for all traces; fluorescence values have been normalized to a value of 1 for the fluorescence of fully folded protein in 0 M GdnHCl. (c) Intrinsic tryptophan fluorescence-monitored kinetic trace of refolding induced by a pH jump from 12 to 7 in the absence of denaturant. The solid line through the data trace represents a nonlinear least-squares fit to a three-exponential equation. The values obtained for the apparent rate constants  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are  $108 \pm 7$ ,  $8 \pm 1.6$ , and  $0.6 \pm 0.2$  s<sup>-1</sup>, respectively, and those for the relative amplitudes of the observable kinetic phases  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are  $0.55 \pm 0.09$ ,  $0.30 \pm 0.06$ ,  $0.15 \pm 0.05$ , respectively. The dashed line indicates the fluorescence value of the protein at pH 12, from which refolding was commenced. (d) ANS fluorescence-monitored kinetic trace of refolding induced by a jump in pH from 12 to 7, in the absence of any denaturant. The solid line through the data trace represents a nonlinear, least-squares fit to a three-exponential equation. The values obtained for the apparent rate constants  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are  $5.2 \pm 1$ ,  $0.4 \pm 0.1$ , and  $0.003 \pm 0.0002$  s<sup>-1</sup>, respectively, and those for the relative amplitudes  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are  $0.7 \pm 0.09$ ,  $0.2 \pm 0.06$ ,  $0.1 \pm 0.05$ , respectively. The dashed line indicates the values of the fluorescence of ANS in the presence of the unfolded protein at pH 12 and of the native protein at pH 7.

Typically, ANS does not bind to unfolded protein or native protein, which makes it an ideal probe to monitor transiently populated intermediates during folding, which have exposed hydrophobic patches. Figure 4d shows that the ANS fluorescence is not enhanced in the presence of completely unfolded monellin at pH 12 or folded monellin at pH 7. It is seen, however, that the ANS fluorescence increases and decreases during the course of refolding at pH 7. The increase in fluorescence occurs in a sub-millisecond burst phase, during the mixing dead-time, and the subsequent decrease

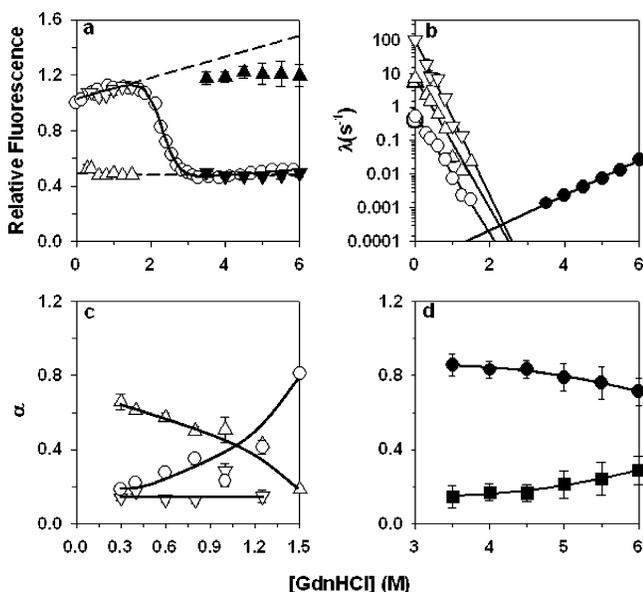


FIGURE 5: Dependence of folding and unfolding kinetics on GdnHCl concentration. Folding and unfolding kinetics were studied by monitoring fluorescence at 340 nm. (a) Kinetic versus equilibrium amplitudes: (○) equilibrium unfolding curve; (△)  $t = 0$  points of kinetic refolding traces; (▽)  $t = \infty$  points of kinetic refolding traces; (▲)  $t = 0$  points of kinetic unfolding traces; (▼)  $t = \infty$  points of kinetic unfolding traces. The solid line through the equilibrium unfolding data is a nonlinear least-squares fit to a two-state  $N \leftrightarrow U$  model. (b) Dependence of the observed rates for folding and unfolding on GdnHCl concentration: (○) slow phase of refolding; (△) fast phase of refolding; (▽) very fast phase of refolding; (●) slow phase of unfolding. The rates of folding in the absence of denaturant were obtained from intrinsic tryptophan fluorescence-monitored refolding induced by a pH jump from 12 to 7 (Figure 4c). The larger symbols at zero denaturant represent the apparent rate constants obtained from ANS fluorescence-monitored refolding induced by a pH jump from 12 to 7 (Figure 4d). The solid lines through the three sets of apparent folding rate constants are least-squares fits to the equation  $\log \lambda = \log \lambda_{\text{H}_2\text{O}} + m_f[\text{GdnHCl}]$  and yield values for  $\lambda_{\text{H}_2\text{O}}^{\text{v}}$ ,  $\lambda_{\text{H}_2\text{O}}^{\text{f}}$ , and  $\lambda_{\text{H}_2\text{O}}^{\text{s}}$  of 108, 8, and  $0.6 \text{ s}^{-1}$ , respectively, and for  $m_f^{\text{v}}$ ,  $m_f^{\text{f}}$ , and  $m_f^{\text{s}}$  of  $-2.3$ ,  $-2.2$ , and  $-1.8 \text{ M}^{-1}$ , respectively. The solid line through the apparent rate constants for unfolding is a least-squares fit to the equation  $\log \lambda^{\text{s}} = \log \lambda_{\text{H}_2\text{O}}^{\text{s}} + m_u[\text{GdnHCl}]$  and yields values for  $\lambda_{\text{H}_2\text{O}}^{\text{s}}$  and  $m_u^{\text{s}}$  of  $0.00002 \text{ s}^{-1}$  and  $0.52 \text{ M}^{-1}$ , respectively. (c) Dependence of the relative amplitudes of the very fast (▽), fast (△), and slow (○) phases of refolding on GdnHCl concentration. (d) Dependence of the relative amplitude of the slow phase (●) and burst phase (■) of unfolding on GdnHCl concentration. The relative amplitudes shown in panels c and d were calculated using linearly extrapolated native protein and unfolded protein baselines shown as dashed lines in panel a. The solid lines through the relative amplitude data have been drawn by inspection only. Error bars on the points represent the standard deviations of three separate measurements.

occurs in three kinetic phases with apparent rates of  $5.2 \pm 1$ ,  $0.4 \pm 0.1$ , and  $0.003 \pm 0.0002 \text{ s}^{-1}$  and relative amplitudes of 0.7, 0.2, and 0.1, respectively. Thus, the two faster apparent rate constants, determined by measurement of ANS fluorescence, match the two slower apparent rate constants, determined by measurement of intrinsic tryptophan fluorescence, for folding in zero denaturant.

**Dependence of Folding and Unfolding Kinetics on GdnHCl Concentration.** Figure 5 brings out the dependence of the folding and unfolding kinetics on GdnHCl concentration. In Figure 5a, it is seen that both refolding and unfolding kinetic traces extrapolate at  $t = \infty$  to the value of fluorescence expected from the equilibrium unfolding curve, indicating

that both refolding and unfolding reactions reached completion. Each refolding trace extrapolates back, at  $t = 0$ , to approximately the value expected by linear extrapolation of the unfolded protein baseline to low GdnHCl concentrations, indicating the absence of a sub-millisecond burst phase change during refolding. On the other hand, it is seen that each unfolding trace extrapolates back at  $t = 0$  to a fluorescence value which is 20% larger than the fluorescence of the native state in the absence of GdnHCl.

Figure 5a also shows that the  $t = 0$  values of the kinetic unfolding traces do not fall on the linearly extrapolated native protein baseline, which has a steep positive slope. It therefore appears that a fraction of the fluorescence change during unfolding occurs in a sub-millisecond burst phase. For example, the fluorescence of the native protein in 5 M GdnHCl is expected, from linear extrapolation of the native protein baseline, to be 140% of that of the native protein in native buffer, and 20% of the total change in fluorescence that occurs during unfolding in 5 M GdnHCl appears to occur during a sub-millisecond burst phase.

Figure 5b shows the dependence of the three observed rate constants of refolding and the single observed rate constant of unfolding on GdnHCl concentration. The observed rate constants are seen to have similar exponential dependences on GdnHCl concentration. For none of the three dependences of the observed rate constants on GdnHCl concentration is a rollover seen at low denaturant concentrations, indicating that the rate-limiting steps remain the same over the range of GdnHCl concentrations studied. When extrapolated to zero denaturant, the three observed rate constants are  $100 \text{ s}^{-1}$  (very fast),  $10 \text{ s}^{-1}$  (fast), and  $0.8 \text{ s}^{-1}$  (slow). These values match the observed rate constants in zero denaturant, which were determined by initiating refolding by a pH jump from 12 to 7. When extrapolated to zero denaturant, the observed slow rate constant of unfolding is  $0.00002 \text{ s}^{-1}$ .

Figure 5c shows how the relative amplitudes of the three kinetic phases, very fast, fast, and slow, vary when refolding and unfolding are carried out in different GdnHCl concentrations. Whereas the relative amplitude of the very fast kinetic phase appears to be invariant at 15%, the relative amplitude of the fast kinetic phase decreases with an increase in GdnHCl concentration. For refolding at GdnHCl concentrations  $\geq 1.5 \text{ M}$ , only the fast and slow phases of refolding are observed. The relative amplitude of the slow kinetic phase increases with an increase in GdnHCl concentration. The relative amplitudes of the two faster phases detected by ANS fluorescence measurement of refolding in zero denaturant match the relative amplitudes of the fast and slow phase of refolding in the lowest concentrations of GdnHCl used.

Figure 5d shows the relative amplitudes of slow phase and burst phase in unfolding kinetics. The relative amplitude of the slow phase decreases with GdnHCl concentration, whereas that of the burst phase increases.

**Double-Jump Unfolding Assays.** To determine the amounts of native protein and folding intermediates at different times during refolding, interrupted refolding experiments were carried out. At different times of refolding in 0.5 M GdnHCl, refolding was interrupted and the protein transferred to 5 M GdnHCl. Figure 6a shows the kinetics of unfolding when refolding was interrupted at different times. It is seen that the total amplitude of the unfolding reaction increases with the time at which refolding was interrupted, indicating the

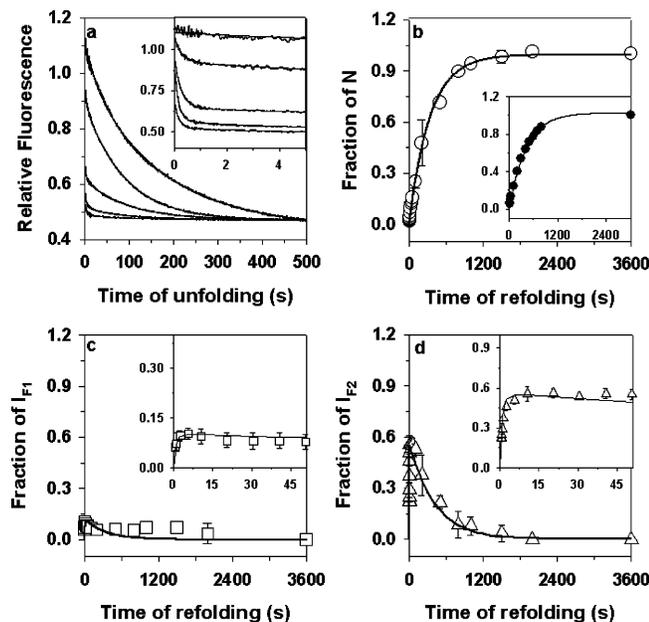


FIGURE 6: Kinetics of formation of the N state and two intermediate forms from the equilibrium-unfolded U form. Refolding was initiated by effecting a jump in GdnHCl concentration from 5 to 0.5 M. Panel a shows the kinetic unfolding traces obtained when the refolding was interrupted at different times (from bottom to top: 0.2, 10, 100, 500, and 3600 s) with a jump in GdnHCl concentration from 0.5 to 5 M. The inset shows the initial 5 s of the kinetic unfolding traces. The fraction of N, the fraction of intermediate  $I_{F1}$ , and the fraction of intermediate  $I_{F2}$  are plotted against time of refolding, in panels b, c, and d, respectively. At any time of refolding, N,  $I_{F1}$ , and  $I_{F2}$  were identified as the forms that unfold with rates of  $0.011 \pm 0.003$ ,  $0.2 \pm 0.04$ , and  $4.5 \pm 1$   $s^{-1}$ , respectively. The amounts of N,  $I_{F1}$ , and  $I_{F2}$  formed at each time of refolding, from 50 ms to 3600 s, were determined from a double-jump unfolding assay, as described under Materials and Methods. The fractions of N,  $I_{F1}$ , and  $I_{F2}$  formed at any time of refolding were determined as the amounts of N,  $I_{F1}$ , and  $I_{F2}$  formed at that time divided by the amount of N formed after folding was complete. The solid line in panel b describes the formation of N according to a single-exponential process, with an apparent rate constant of  $0.0024$   $s^{-1}$ . The inset in panel b shows the time course of formation of N during refolding in 1 M GdnHCl, determined by a manual mixing double-jump experiment. The solid line through the data in the inset is a fit of the data to a single-exponential equation and yields an apparent rate constant for the formation of N of  $0.0025$   $s^{-1}$ . Each data point represents the average of three independent determinations from different experiments, and the error bars represent the standard deviations.

increasing amount of partially and fully folded protein with time of refolding. Each kinetic unfolding trace is described by a three-exponential equation, and the relative amplitude of each kinetic phase of unfolding depends on the time at which refolding was interrupted. Of the three observed rate constants determined for each kinetic unfolding trace, one rate constant corresponds to the observed rate constant for the unfolding of native protein. At the longest time at which refolding was interrupted, which was 3600 s, unfolding occurs in one kinetic phase corresponding to this rate constant, indicating that all protein molecules have become fully folded at 3600 s.

Figure 6b shows that the time course of the increase in the relative amplitude of the kinetic phase corresponding to the unfolding of native protein is well-described as a single-exponential process. This represents the kinetics of formation of native protein. It is seen that the apparent rate constant

for the formation of native protein is  $0.0024$   $s^{-1}$ , which is 40-fold slower than the slowest rate constant of  $0.1$   $s^{-1}$  observed for the change in fluorescence that occurs during refolding in 0.5 M GdnHCl (Figure 5b). This result clearly indicates that the slowest phase of fluorescence change during refolding does not represent the formation of native protein but the formation of one or more intermediates which possess the same fluorescence properties as native protein. Native protein forms very slowly from these late intermediates in steps, which are silent to fluorescence change. It should be noted that the apparent rate constant for the formation of native protein,  $0.0024$   $s^{-1}$ , matches the slowest apparent constant measured by the change in ANS fluorescence during unfolding initiated by a pH jump from 12 to 7 (Figure 4d).

The other two observed rate constants of unfolding in the interrupted refolding experiments are much faster than the rate constant of unfolding of native protein. These two faster kinetic unfolding phases therefore represent the unfolding of two partially folded forms of the protein,  $I_{F1}$  and  $I_{F2}$ , which are more unstable than the native protein. For each kinetic unfolding trace, the relative amplitudes of these two kinetic phases can be taken to represent the relative amounts of  $I_{F1}$  and  $I_{F2}$  that were present at the time folding was interrupted, if it is assumed that the quantum yields of fluorescence of  $I_{F1}$ ,  $I_{F2}$ , and N are all the same. Panels c and d of Figure 6 show that the populations of  $I_{F1}$  and  $I_{F2}$  increase at a rate of  $1$   $s^{-1}$ , reach a maximum, and then decrease to zero at a rate of  $0.0024$   $s^{-1}$ . Thus, the rates at which  $I_{F1}$  and  $I_{F2}$  disappear are the same as the rate of appearance of native protein.

In a separate, manual mixing, experiment, refolding was commenced in 1 M GdnHCl, and the fraction of molecules present as N at any time of refolding was determined by carrying out the unfolding assay as described above (inset of Figure 6b). It was seen that the apparent rate of formation of N during refolding in 1 M GdnHCl is also  $0.0024$   $s^{-1}$ , as it is for refolding in 0.5 M GdnHCl. Hence, it appears that the rate of formation of N from the late intermediate(s) is independent of the GdnHCl concentration in which the protein is refolded.

**Double-Jump Refolding Assays.** To determine the amount of unfolded protein at different times during unfolding, as well as to detect any heterogeneity in the unfolded form, interrupted unfolding experiments were carried out. At different times during unfolding in 5 M GdnHCl, unfolding was interrupted by effecting a drop in GdnHCl concentration from 5 to 0.5 M. Figure 7a shows the kinetic traces of refolding in 0.5 M GdnHCl of protein that had been unfolded in 5 M GdnHCl for different times. It is seen that the total amplitude of the refolding reaction increases with the time for which the protein was unfolded prior to refolding. In general, the change in fluorescence accompanying the refolding reaction is described as a three-exponential process, with the three kinetic phases having rate constants similar in value to those measured for the refolding of equilibrium unfolded protein (Figure 5b). The relative amplitudes of the three kinetic phases of refolding vary with the time the protein was unfolded prior to the refolding jump. After the longest time for which unfolding was allowed to proceed prior to being interrupted by the refolding jump (800 s), the relative amplitudes of the three kinetic phases match the relative amplitudes observed for the refolding of equilibrium-unfolded protein, which are 15% for the very fast phase,

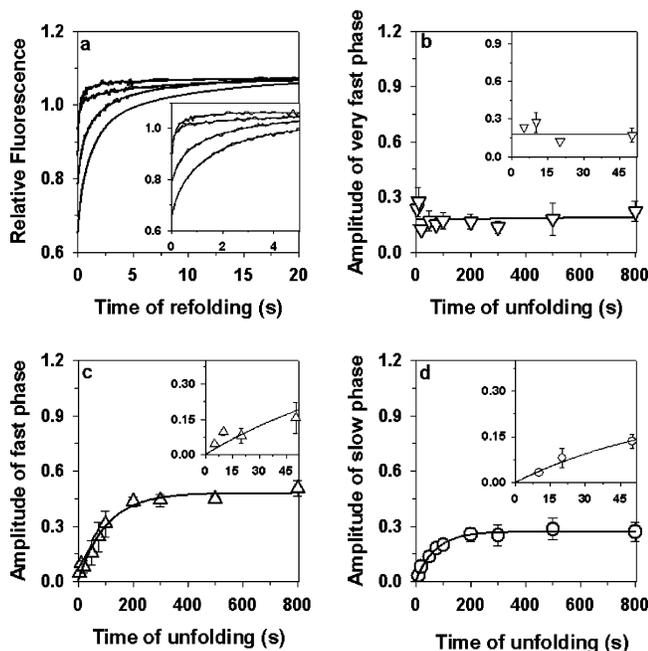


FIGURE 7: Kinetics of formation of the U form from the N state. Unfolding was initiated by effecting a jump in GdnHCl concentration from 0 to 5 M. Panel a shows the kinetic refolding traces obtained when unfolding in 5 M GdnHCl was interrupted at different times (from bottom to top: 10, 20, 100, and 200 s) by a jump in GdnHCl concentration from 5 to 0.5 M. The kinetic traces have been normalized to a value of 1 for fully folded protein in 0 M GdnHCl. The solid lines through the data represent fits to the sum of three exponentials. All refolding traces yielded the same three rate constants:  $7 \pm 2$ ,  $1 \pm 0.3$ , and  $0.1 \pm 0.03$   $s^{-1}$ . The inset shows the initial 5 s of the refolding traces. The relative amplitudes of the very fast, fast, and slow kinetic refolding phases are plotted against the time at which unfolding is interrupted by the jump in GdnHCl concentration from 5 to 0.5 M, in panels b, c, and d, respectively. The insets in panels b, c, and d show how these relative amplitudes vary during the first 50 s of unfolding. The solid lines through the data in panel b represent a linear fit of the data, and the solid lines through the data in panels c and d represent fits of the data to single-exponential equations and yield apparent rate constants of  $0.011$   $s^{-1}$  (panel c) and  $0.011$   $s^{-1}$  (panel d). Each data point in panels b, c, and d represents the average of three independent determinations from different experiments, and the error bars represent the standard deviations.

60% for the fast phase, and 25% for the slow phase (Figure 5c,d).

Panels b–d of Figure 7 show how the relative amplitudes of the three kinetic phases of refolding change with the time the protein was unfolded in 5 M GdnHCl prior to refolding. It is seen that the relative amplitude of the very fast phase is already at its final value (15%) even when the protein is unfolded for only 5 s. On the other hand, the relative amplitudes of the fast and slow phases develop to the final values seen for equilibrium unfolded protein at the same rate of  $0.011$   $s^{-1}$ , which matches the observed rate of unfolding of the protein in 5 M GdnHCl. Hence, the unfolded form that gives rise to the fast and slow phases of folding is formed at the same rate at which the protein is observed to unfold.

The double-jump refolding assays were also carried out by interrupting unfolding in 5 M GdnHCl by a refolding assay in 1.5 M GdnHCl. The change in fluorescence accompanying the refolding process in 1.5 M GdnHCl occurs in two kinetic phases:  $\sim 80$ – $85\%$  (slow) and  $\sim 15$ – $20\%$  (fast) (Figure 5c). The results of the double-jump refolding

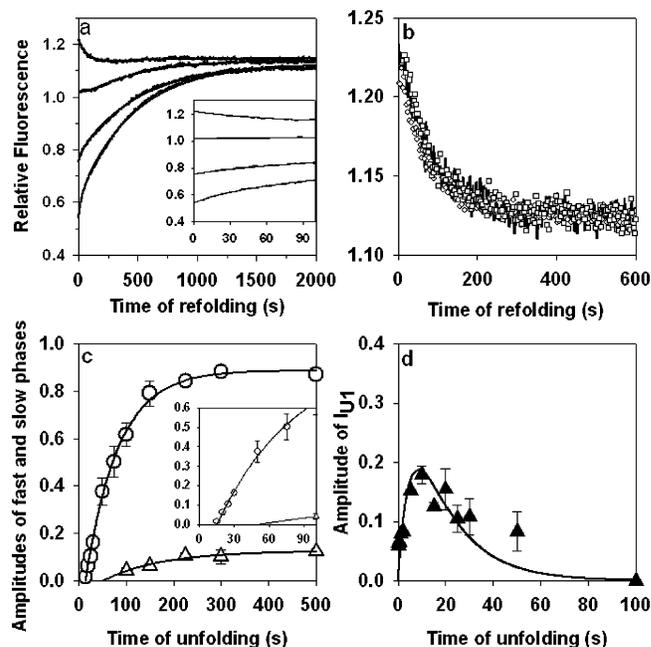


FIGURE 8: Kinetics of formation of the U form from the N state. Unfolding was initiated by effecting a jump in GdnHCl concentration from 0 to 5 M. Panel a shows the fluorescence-monitored kinetic refolding traces obtained when unfolding in 5 M GdnHCl was interrupted at different times (from top to bottom: 10, 50, 100, and 200 s) by a jump in GdnHCl concentration from 5 to 1.5 M. The kinetic traces have been normalized to a value of 1 for fully folded protein in 0 M GdnHCl. The solid lines through the data represent fits to the sum of exponentials. All refolding traces yielded the same rate constants:  $0.014 \pm 0.001$  and  $0.0024 \pm 0.0002$   $s^{-1}$ . The inset shows the initial 100 s of the refolding traces. Panel b shows the fluorescence-monitored kinetic traces of refolding in 1.5 M GdnHCl, obtained after unfolding in 5 M GdnHCl for 10 s, at different final protein concentrations:  $0.5$   $\mu M$  (solid line),  $1$   $\mu M$  ( $\square$ ), and  $8.15$   $\mu M$  ( $\circ$ ). Panel c shows the relative amplitudes of the slow ( $\circ$ ) and fast phases ( $\Delta$ ) of fluorescence increase during refolding, plotted against the time at which unfolding is interrupted by the jump in GdnHCl concentration from 5 to 0.5 M. The inset shows the initial 100 s of the same. The solid lines through the data in panel c represent fits of the data to a single-exponential equation. In both cases, the fits yield an apparent rate constant of  $0.011$   $s^{-1}$ . Panel d shows the relative amplitude of the decrease in fluorescence ( $\blacktriangle$ ), which is observed when unfolding is interrupted at any time within the first 50 s, plotted against the time at which unfolding is interrupted. The solid line through the data in panel d represents a fit of the data to a two-exponential equation (eq 1) and yields apparent rate constants of  $0.2$  and  $0.06$   $s^{-1}$ . Each data point in panels c and d represents the average of three independent determinations from different experiments, and the error bars represent the standard deviations.

assay are therefore less complex (Figure 8). Figure 8a shows kinetic traces of refolding of protein that had been unfolded for different times. The refolding kinetic trace of protein that was unfolded for any time longer than 75 s is well-described by a two-exponential increase in fluorescence. The apparent rate constants of this increase in fluorescence are  $0.025$  and  $0.0024$   $s^{-1}$ , which are the same as the fast and slow kinetic phases of refolding in 1.5 M GdnHCl (Figure 5c) of equilibrium unfolded protein. The relative amplitudes of these two kinetic phases of refolding are seen to increase with the time at which unfolding in 5 M GdnHCl was interrupted and to reach values of 80–85 and 15–20% at the longest time of prior unfolding.

For protein that was unfolded for a short time (from 200 ms to 10 s) in 5 M GdnHCl before being transferred to

1.5 M GdnHCl, the refolding trace is well-described as a single-exponential decrease in fluorescence. The apparent rate constant for this decrease in fluorescence is  $0.014\text{ s}^{-1}$  and is independent of the time (from 200 ms to 10 s) the protein was unfolded prior to the refolding jump. The refolding traces of protein that was unfolded for times between 10 and 50 s show first a decrease and then an increase in fluorescence, with rates of  $0.014$  and  $0.0024\text{ s}^{-1}$ , respectively. The fast kinetic phase of the increase in fluorescence, which occurs at a rate of  $0.025\text{ s}^{-1}$  during refolding in 1.5 M GdnHCl, cannot be discerned when the time of prior unfolding is  $<75\text{ s}$ . Figure 8b shows that the kinetic phase of decrease in fluorescence, which occurs at the rate of  $0.014\text{ s}^{-1}$ , is independent of protein concentration, indicating that it does not arise from transient aggregation during unfolding or refolding (32).

Figure 8c shows how the relative amplitudes of the slow and fast kinetic phases of the increase in fluorescence that accompanies refolding in 1.5 M GdnHCl increase with the time of prior unfolding in 5 M GdnHCl. The kinetics of the increase in the relative amplitude of the slow refolding phase ( $0.0024\text{ s}^{-1}$ ) display an initial lag phase of 15 s duration, after which its relative amplitude increases exponentially at a rate of  $0.011\text{ s}^{-1}$ , which is the same as the apparent rate constant of fluorescence change when native protein is unfolded in 5 M GdnHCl (Figure 5b). Similarly, the kinetics of the increase in the relative amplitude of the fast refolding phase ( $0.025\text{ s}^{-1}$ ) display an initial lag phase. This lag appears to be longer in duration than that seen for the slow refolding phase, but this is most likely because of the inaccuracies in the measurement of the fast refolding phase at shorter times of prior unfolding.

The kinetic phase of the decrease in fluorescence during refolding, which occurs at a rate of  $0.014\text{ s}^{-1}$ , is not seen in the case of protein that is unfolded to equilibrium (Figures 4 and 5), that is, when unfolding is commenced from U. It must therefore arise from the refolding of a partially unfolded intermediate that is populated transiently during the first 50 s of refolding. Figure 8d shows that the amplitude of this kinetic phase of the decrease in fluorescence ( $0.014\text{ s}^{-1}$ ) increases as the time of prior unfolding is increased from 0.2 to 10 s and then decreases to zero as the time of prior unfolding is increased further from 10 to 100 s. The time course of change of this amplitude of this kinetic phase is well-described as a two-exponential process with apparent rate constants  $0.2$  and  $0.06\text{ s}^{-1}$  for the increase and decrease in amplitude, respectively. Figure 8d therefore shows that the partially unfolded intermediate,  $I_{UE}$ , is populated maximally at about 10 s of unfolding in 5 M GdnHCl. The fluorescence of  $I_{UE}$  appears to exceed that of N by  $\sim 10\%$ .

## DISCUSSION

*Equilibrium Unfolding of Single-Chain Monellin Appears To Be Two-State.* The observation that the equilibrium unfolding transition curves obtained using different probes [fluorescence, far-UV CD (222 nm), and near-UV CD (280 nm)] are co-incident (Figure 3) suggests that intermediates are not populated to any significant extent at equilibrium at any concentration of denaturant. Although the fluorescence-monitored equilibrium unfolding curve shows a significant upward curvature in the native protein baseline region, the

possibility that a partially unfolded structure might be populated in 1.5 M GdnHCl could be ruled out by the observation that protein that had been equilibrated at this concentration of GdnHCl, in which its fluorescence was maximum, unfolded at the same rate as protein equilibrated in native buffer in zero denaturant.

*Refolding Kinetics of Single-Chain Monellin Are Complex in Strongly Stabilizing Conditions.* In this study, the refolding of single-chain monellin in strongly stabilizing conditions such as in 0.5 M GdnHCl is shown to occur in five kinetic phases. The fastest kinetic phase is apparent when refolding is monitored by measurement of the change in ANS fluorescence and is complete in the sub-millisecond time domain. Hence, one or more early intermediates,  $I_E$ , are formed in the sub-millisecond time domain. It is likely that  $I_E$ , at 1 ms of refolding, consists of more than one species because, subsequently, ANS dissociates from the folding protein molecules in three kinetic phases. Previous sub-millisecond measurements of the folding of single-chain monellin at pH 9.4 have identified  $I_E$  to be the product of a chain collapse reaction (48), which is complete in  $300\text{ }\mu\text{s}$ .  $I_E$  was reported to contain an insignificant amount of secondary structure. It was observed that the tryptophan fluorescence of  $I_E$  is the same as that of the fully unfolded protein. This could mean that  $I_E$  is only loosely compact in a manner that the tryptophan residue remains solvent accessible, or it could mean that the tryptophan residue (Trp4) is at too N-terminal a position on the sequence for it to report on the folding of the entire protein molecule. However,  $I_E$  obviously has sufficient solvent-accessible hydrophobic surface for ANS to bind.

Of the five kinetic phases, three are detected when intrinsic tryptophan fluorescence is monitored (Figures 3 and 4). The very fast phase, with an apparent rate constant, in zero denaturant, of  $100\text{ s}^{-1}$ , accounts for 15% of the total fluorescence change during refolding. The observation that its amplitude is independent of denaturant concentration suggests that the very fast phase arises from a separate population of unfolded molecules. The relative amplitude of the very fast phase indicates that this unfolded state ( $U_1$ ) comprises 15% of the unfolded protein at equilibrium. In a previous study of folding at pH 9.4 (48), a significant change in far-UV CD was seen to occur in this very fast phase. The fast and slow phases, with apparent rate constants in zero denaturant of  $8$  and  $0.6\text{ s}^{-1}$ , respectively, account for the remaining 85% of the fluorescence change. Surprisingly, the fast phase was not seen in an earlier study of the folding of alkali unfolded single-chain monellin at pH 9.4 in zero denaturant, by either fluorescence or CD measurement (48). The observation that with increasing GdnHCl concentration the relative amplitude of the slow phase decreases at the expense of that of the fast phase suggests that the fast and slow phases arise from the same population of unfolded molecules and that this population of unfolded molecules ( $U_2$ ) is distinct from that giving rise to the very fast phase of refolding.  $U_2$  is populated to  $\sim 85\%$  at equilibrium.

Finally, there is also a very slow phase of folding, with an apparent rate constant of  $0.0024 \pm 0.0003\text{ s}^{-1}$ . This kinetic phase is silent to fluorescence change and, hence, has not been reported in the earlier study of the refolding of single-chain monellin (48), where it was assumed that the slow phase leads to the formation of native protein. The very slow

phase can be detected in double-jump, interrupted folding experiments in which the amount of native protein can be determined quantitatively (Figure 6). These experiments indicate that the formation of native protein proceeds from intermediates possessing native-protein fluorescence, in the very slow phase of folding. The very slow phase of refolding could also be detected when folding was monitored by measurement of ANS fluorescence, indicating that some ANS dissociates from the refolding protein molecules in the very last step of folding. In the earlier study of the refolding of single-chain monellin (48), the very slow phase was not detected by measurement of ANS fluorescence, because the measurement was not done beyond 10 s of refolding.

*Fast and Slow Phases of Fluorescence Change Occur on Competing Fast and Slow Folding Pathways.* As discussed above, the fast and slow folding pathways originate from the same unfolded form,  $U_2$ . The simplest explanation for the origin of the fast and slow phases of refolding is that they represent two competing pathways of folding and that utilization of the slow pathway at the expense of the fast pathway dominates folding at higher concentrations of GdnHCl. It would be very unlikely that the fast and slow phases represent consecutive steps via an intermediate on a single pathway, because that would mean that the fluorescence properties (and, hence, structure) of the intermediate are changing drastically with a change in denaturant concentration.

Because the relative amplitudes of the fast and slow phases of fluorescence change (60 and 25%, respectively, in 0.5 M GdnHCl) are not proportional to the observed fast and slow rate constants (1 and  $0.1 \text{ s}^{-1}$  in 0.5 M GdnHCl), the kinetic partitioning leading to the observed relative utilization of the fast and slow folding pathways must occur very early after commencement of folding from refolding  $U_2$ . Because it is known that at least one burst phase intermediate,  $I_E$ , is populated within a few milliseconds of refolding, it would appear that the kinetic partitioning occurs during the  $U_2 \rightarrow I_E$  transition. The observation that the dissociation of ANS from  $I_E$  as it folds occurs in fast as well as slow kinetic phases (Figure 4d) supports the assumption that  $I_E$  feeds into both the fast and slow pathways of folding.  $I_E$  must therefore consist of two forms,  $I_{E1}$  and  $I_{E2}$ , and the relative populations of  $I_{E1}$  and  $I_{E2}$  would determine the relative utilization of the fast and slow refolding pathways.

*Fast Folding Phase Leads to the Formation of Two Intermediates on Competing Pathways.* It is observed from the interrupted refolding experiments in 0.5 M GdnHCl that the rate ( $1 \text{ s}^{-1}$ ) at which the two intermediates  $I_{F1}$  and  $I_{F2}$  build up in population is the same as the rate of the fast fluorescence change during refolding in 0.5 M GdnHCl. This suggests that the fast phase of fluorescence change represents the formation of  $I_{F1}$  and  $I_{F2}$ . If this is true, then the relative amplitude of this fast kinetic phase, which is 60% (Figure 5c), must reflect the maximum extent to which  $I_{F1}$  and  $I_{F2}$  are populated during refolding. Panels c and d of Figure 6 indicate that this is indeed so: at 10 s of refolding, when the fast phase of fluorescence change is complete, ~60% of the molecules are present as  $I_{F1}$  and  $I_{F2}$ . The agreement also supports the assumption (see Results) that the quantum yields of fluorescence of  $I_{F1}$ ,  $I_{F2}$ , and N are all the same.

$I_{F1}$  must be substantially more stable than  $I_{F2}$ :  $I_{F1}$  is seen to unfold 20-fold more slowly than  $I_{F2}$  (see Figure 6 legend).

Nevertheless, they form at the same rate and also transform into native protein at the same rate (Figure 6). The most likely explanation is that  $I_{F1}$  and  $I_{F2}$  form on two parallel pathways,  $I_{E1} \rightarrow I_{F1} \rightarrow N$  and  $I_{E1} \rightarrow I_{F2} \rightarrow N$ , because it is very unlikely that two intermediates of vastly different stabilities will form on a single pathway at the same time.

There could be an alternative explanation. There may be only one sequential pathway:  $I_{E1} \rightarrow I_{F1} \rightarrow N$ , and  $I_{F2}$  may be a dead-end, probably, misfolded intermediate that forms from  $I_{F2}$ . If this were true, then the  $I_{F1} \rightleftharpoons I_{F2}$  equilibrium would have to be rapid (~10-fold faster) compared to the formation of  $I_{F1}$ , to account for the observation that both  $I_{F1}$  and  $I_{F2}$  are seen to form essentially together (Figure 6). It has been suggested that such a dead-end, misfolded intermediate might be responsible for the observation, in pulsed-HX studies of the folding of several proteins (49, 50), that on-pathway intermediates are not populated fully at the time they have formed to their maximum extent. However, it is unlikely that such an explanation is correct in the present case: in 5 M GdnHCl, the unfolding of  $I_{F2}$  to  $I_{F1}$  would be much faster than it would be in 0.5 M GdnHCl, and both  $I_{F1}$  and  $I_{F2}$  would be expected to unfold in one observable kinetic phase, corresponding to the unfolding of  $I_{F1}$ . Because the unfolding of  $I_{F1}$  and  $I_{F2}$  in 5 M GdnHCl is seen to occur in two separate kinetic phases, it is very unlikely that one forms directly from the other. Hence, it appears that  $I_{F1}$  and  $I_{F2}$  must form separately on different competing pathways. It should be noted that in the case of the pulsed-HX experiments referred to above, as well as in the case of other experiments in which a dead-end misfolded intermediate has been postulated as an explanation for an on-pathway intermediate not being populated fully (51), competing pathways are a viable alternative explanation (52, 53).

*Intermediates Are Also Populated on the Pathways Giving Rise to the Very Fast and Slow Folding Phases of Fluorescence Change.* During refolding in 0.5 M GdnHCl, 25% of the fluorescence change occurs in the slow kinetic phase at a rate of  $0.1 \text{ s}^{-1}$  and 15% in the very fast kinetic phase at a rate of  $7 \text{ s}^{-1}$ . These kinetic phases of change in fluorescence appear to occur on two different folding pathways, one very fast and one slow (see above). Native protein is, however, formed in only one very slow kinetic phase with an apparent rate constant of  $0.0024 \text{ s}^{-1}$ . It therefore appears that intermediates must be populated on both the slow and very fast folding pathways. The very fast phase of fluorescence change would correspond to the formation of the intermediate  $I_{VF}$  from unfolded protein, on the very fast folding pathway. The slow phase of fluorescence change would correspond to the formation of the intermediate  $I_S$  from unfolded protein, on the slow folding pathway. Both  $I_{VF}$  and  $I_S$  must transform to N at the same very slow rate. The fluorescence properties of  $I_{VF}$  and  $I_S$  would be the same as that of N, because the very slow folding phase is silent to fluorescence change (see above). The double-jump, interrupted refolding experiment is, however, unable to detect the presence of  $I_{VF}$  and  $I_S$ . It seems that  $I_{VF}$  and  $I_S$  must be so unstable in 5 M GdnHCl that they unfold too quickly to be seen.

*Native Protein Is Formed via Multiple Pathways but in a Single Very Slow Kinetic Phase.* Each of the four parallel folding pathways (see above) is distinguished by the accumulation of a unique intermediate,  $I_{F1}$ ,  $I_{F2}$ ,  $I_{VF}$ , or  $I_S$ .  $I_S$

forms the most slowly, at an apparent rate of  $0.1 \text{ s}^{-1}$  in  $0.5 \text{ M GdnHCl}$ . Because only a single very slow kinetic phase with an apparent rate constant of  $0.0024 \text{ s}^{-1}$  is observed for the formation of N, it seems that all four intermediates,  $I_{F1}$ ,  $I_{F2}$ ,  $I_{VF}$ , and  $I_S$ , on the four different folding pathways transform to N at this very slow apparent rate. It is because N is formed on four different parallel pathways, from intermediates which themselves form at rates 40–4000-fold faster, that effectively no lag phase can be observed during the formation of N (Figure 6d).

*Folding Appears To Be Simpler in Marginally Stabilizing Conditions.* When refolding is carried out in marginally stabilizing conditions ( $1.5 \text{ M GdnHCl}$ ), only the fast and slow phases of refolding are observed. The value of the faster observed rate constant of refolding ( $0.025 \text{ s}^{-1}$ ) is such that this apparent rate constant could be attributed to either the fast or very fast kinetic phases of refolding (see Figure 5c). However, the observation, in interrupted unfolding experiments, that the relative amplitude of this kinetic phase increases concomitantly with that of the slow phase of refolding, as the time of prior unfolding is increased (Figure 8c), suggests that the observed rate constant of  $0.025 \text{ s}^{-1}$  corresponds to that of the fast refolding phase. The absence of the very fast phase of fluorescence change suggests that the intermediate  $I_{VF}$  is strongly destabilized and not populated. It is likely that the early intermediate,  $I_E$ , and the later intermediates,  $I_S$ ,  $I_{F1}$ , and  $I_{F2}$ , are also destabilized in  $1.5 \text{ M GdnHCl}$ . In fact, the observation that the apparent rate of the slow phase of fluorescence change during refolding in  $1.5 \text{ M GdnHCl}$  is the same as the apparent rate of formation of N ( $0.0024 \text{ s}^{-1}$ ) suggests that  $I_S$ , like  $I_{VF}$ , may not be populated in  $1.5 \text{ M GdnHCl}$ . In this context, it is to be noted that  $I_S$  and  $I_{VF}$  are more unstable than  $I_{F1}$  and  $I_{F2}$ : the unfolding reactions of the former in  $5 \text{ M GdnHCl}$  are too fast to be measured in the double-jump, interrupted refolding experiments (Figure 6), whereas those of the latter are measurable.

In summary, the refolding experiments, single-jump as well as double-jump, suggest the following. (1) Two populations of unfolded molecules,  $U_1$  and  $U_2$ , are present.  $U_1$  is populated to  $\sim 15\%$  and  $U_2$  to  $\sim 85\%$  at equilibrium. (2) A loosely compact folding intermediate  $I_E$  is populated at  $1 \text{ ms}$  of refolding.  $I_E$  is likely to consist of more than one species having intrinsic tryptophan fluorescence properties identical to those of  $U_1$  or  $U_2$ .  $I_E$  can bind ANS. (3) The very fast phase of fluorescence change during refolding originates from  $U_1$  and occurs on a separate folding pathway. (4) The fast and slow phases of fluorescence change during refolding originate from  $U_2$ , and they represent two folding pathways that compete with each other. (5) The fast phase of fluorescence change during refolding itself represents the formation of two intermediates,  $I_{F1}$  and  $I_{F2}$ , of differing stabilities on two competing pathways. (6) An intermediate  $I_{VF}$  is populated on the pathway of folding on which the very fast change in fluorescence occurs. Another intermediate,  $I_S$ , is populated on the pathway of folding on which the slow change in fluorescence occurs. (7) The fluorescence properties of  $I_{F1}$ ,  $I_{F2}$ ,  $I_{VF}$ , and  $I_S$  are identical to those of N. (8) Native protein is formed in a single very slow kinetic phase. (9)  $I_{F1}$  and  $I_{F2}$  have been shown to transform directly to N, and there is no necessity to postulate the involvement of additional intermediates in the transformation of  $I_{VF}$  or  $I_S$  to

N. (10) When folding is carried out in marginally stable conditions ( $1.5 \text{ M GdnHCl}$ ),  $I_{VF}$  does not appear to populate to any significant extent.

*Unfolding Kinetics of Single-Chain Monellin Is Also Complex.* Although the observable kinetics of the unfolding of single-chain monellin are single exponential when monitored by measurement of intrinsic tryptophan fluorescence, there is also an unobservable sub-millisecond burst phase change in fluorescence (Figure 5). The observations that unfolding occurs in two kinetic phases and that the relative amplitude of the burst phase increases, albeit only by  $15\%$ , at the expense of the observable phase suggest two unfolding pathways. It is unlikely that the two unfolding pathways originate from two populations of native monellin molecules because NMR studies of single-chain monellin have failed to detect heterogeneity in the native state (47, 54). It therefore appears that kinetic partitioning occurs early during unfolding, leading to two unfolding pathways, one fast and the other slow.

*Fast Unfolding Pathway Leads to the Formation of  $U_1$ .* The very fast phase of fluorescence change during refolding has been suggested to originate from the unfolded form  $U_1$ , as it transforms into the folding intermediate  $I_{VF}$ . The observation, in a double-jump, interrupted unfolding experiment in  $5 \text{ M GdnHCl}$  (Figure 7), that the relative amplitude of the very fast phase of fluorescence change reaches its final equilibrium value of  $15\%$  within  $5 \text{ s}$  of the commencement of unfolding (Figure 7) suggests that  $U_1$  is fully populated within  $5 \text{ s}$  of unfolding. In the direct unfolding experiment in  $5 \text{ M GdnHCl}$  (Figures 4b and 5a), the fast,  $\sim 15\%$  change in fluorescence, from the extrapolated native baseline at  $5 \text{ M GdnHCl}$  to the  $t = 0$  value of the observed kinetic trace of unfolding, occurs within the first few milliseconds of unfolding. This  $15\%$  change in fluorescence corresponds to the final population of  $U_1$  at equilibrium and indicates that  $U_1$  forms completely within a few milliseconds. There is no evidence for the population of an unfolding intermediate on the  $N \rightarrow U_1$  pathway.

*Slow Unfolding Pathway Leads to the Formation of the Unfolded Form  $U_2$ .* As described above, both the fast and slow phases of the change in fluorescence during refolding appear to originate from events taking place during the folding of only one of the two unfolded forms, namely,  $U_2$ . The observation, in a double-jump, interrupted unfolding experiment (Figure 7), that the relative amplitudes of the fast and slow phases of the change in fluorescence during refolding increase to their final equilibrium values with the time of unfolding, at the same apparent rate, indicates that this is indeed so. Moreover, this apparent rate, which corresponds to the rate of formation of  $U_2$  from N, occurs at the same rate as that of the slow observable change in fluorescence in a direct unfolding experiment. Thus, the slow unfolding pathway leads to the formation of  $U_2$ .

*Intermediates Are Populated on the  $N \rightarrow U_2$  Pathway.* The double-jump experiments in which unfolding in  $5 \text{ M GdnHCl}$  was interrupted at different times by refolding in  $0.5 \text{ M GdnHCl}$  were useful because they allowed the identification of two pathways of unfolding, as described above. But because the kinetics of refolding in  $0.5 \text{ M GdnHCl}$  is complex, it was not possible to identify unfolding intermediates that might be populated on the  $N \rightarrow U_2$  pathway. In principle, such unfolding intermediates can be identified on

the basis of the expectation that they would fold to N at rates faster than the rate of folding of  $U_2$  to N.

Hence, double-jump experiments in which unfolding in 5 M GdnHCl was interrupted by refolding in 1.5 M GdnHCl, were also carried out. In 1.5 M GdnHCl,  $U_1$  and  $U_2$  refold with the change in fluorescence occurring in only the fast (15–20%) and slow (80–85%) kinetic phases, which leads directly to the formation of N. In other words, the intermediate  $I_{VF}$  is too strongly destabilized in 1.5 M GdnHCl to accumulate to any significant extent (see above). The refolding assay for unfolding intermediates becomes simpler to analyze and interpret.

The 15 s lag observed in the formation of  $U_2$ , in the double-jump experiments in which unfolding in 5 M GdnHCl was interrupted by switching to refolding in 1.5 M GdnHCl (Figure 8c), clearly indicates that at least one intermediate is populated on the  $N \rightarrow U_2$  pathway. The double-jump experiment also indicates the population of an intermediate,  $I_{U1}$ , the fluorescence of which is more than that of U and which can fold back to N at a rate of  $0.014 \text{ s}^{-1}$ , which is 6-fold faster than the rate at which N is formed during refolding from U. The amount of  $I_{U1}$  at different times of unfolding could be quantified from the relative amplitude associated with its refolding rate. It was found that  $I_{U1}$  is formed at an apparent rate of  $0.2 \text{ s}^{-1}$  and disappears at a rate of  $0.06 \text{ s}^{-1}$ . These rates account for the 15 s lag phase seen in the formation of  $U_2$  from N, but they also indicate that  $I_{U1}$  cannot be transforming directly to  $U_2$ , because  $U_2$  forms at an apparent rate of  $0.011 \text{ s}^{-1}$  (see above). It becomes necessary to include an additional intermediate,  $I_{U2}$ , on the unfolding pathway, which forms from  $I_{U1}$  at a rate of  $0.06 \text{ s}^{-1}$  and which transforms to  $U_2$  at a rate of  $0.011 \text{ s}^{-1}$ .

*Rapidly Formed Unfolding Intermediate,  $I_{UE}$ , Allows Kinetic Partitioning along the Two Unfolding Pathways.*  $U_1$  is formed very rapidly from N, within a few milliseconds. Nevertheless, most of the protein molecules unfold via the slower  $N \rightarrow U_2$  pathway: for unfolding in 5 M GdnHCl, 85% of the unfolding molecules utilize this pathway and only 15% utilize the competing  $N \rightarrow U_1$  pathway. Such substantive utilization of the slower  $N \rightarrow U_2$  pathway will occur only if the first step on this pathway is faster than the rate of the  $N \rightarrow U_1$  reaction. Hence, it is necessary to postulate that an intermediate,  $I_{UE}$ , is formed so rapidly on the  $N \rightarrow U_2$  pathway that a majority of protein molecules are channeled into this pathway. The optical properties of  $I_{UE}$  would necessarily have to be the same as those of N; otherwise, a much larger burst phase change in fluorescence would be observed.

*Origin of Heterogeneity in the Unfolded Protein.* In the consideration of both the folding and unfolding mechanisms, it has been necessary to propose the existence of two unfolded forms,  $U_1$  and  $U_2$ . It is necessary to understand how this heterogeneity might arise. For many proteins, isomerization of the X-Pro bond is thought to be responsible for similar heterogeneity. An X-Pro bond that is cis in the native protein might become trans in the unfolded state, and unfolded protein molecules with a cis X-Pro bond might fold differently from those with a non-native trans X-Pro bond (55–58). Two of the six proline residues in monellin are in the cis conformation: Pro 41, in the middle of the chain, participates in the  $\beta$  sheet of monellin; and Pro 93, present toward the end of the C terminus, is in a flexible

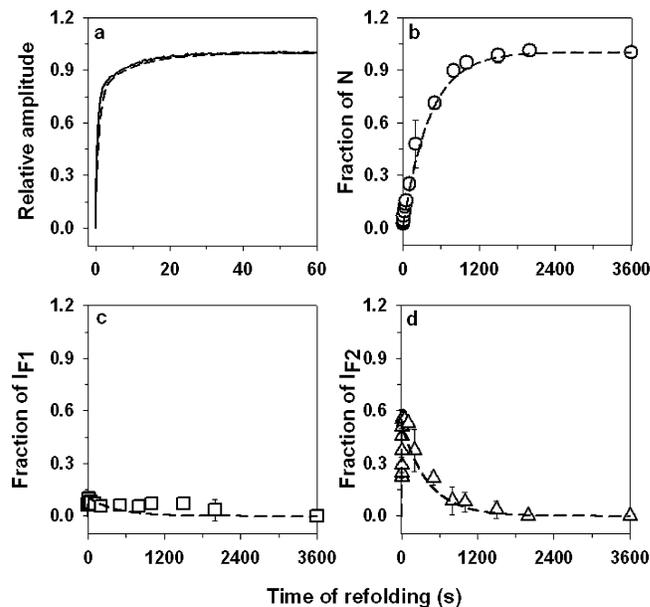


FIGURE 9: Kinetic simulations supporting the refolding mechanism of single-chain monellin. In each panel, experimental data from Figures 4 and 6 are shown along with the kinetic simulation (dashed line) to the refolding mechanism, carried out as described in the text: (a) fluorescence-monitored kinetic trace of refolding in 0.5 M (solid line), normalized between values of 0 for unfolded protein and 1 for native protein; (b) formation of N (O); (c) formation of intermediate,  $I_{F1}$  (□); (d) formation of  $I_{F2}$  (Δ).

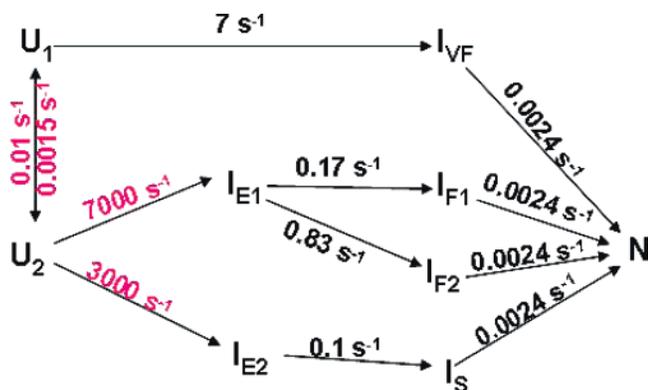
region. When a protein having two cis Pro residues is unfolded, four unfolded forms will be present in equilibrium:  $U_{CC}$ ,  $U_{CT}$ ,  $U_{TC}$ , and  $U_{TT}$ .  $U_{CC}$ , the unfolded form with both proline residues in the cis conformation, is expected to comprise only ~3% of the unfolded molecules (57, 59).  $U_{TT}$ , with both proline residues in the trans conformation, is expected to comprise ~69%, whereas  $U_{CT}$  and  $U_{TC}$  are expected to comprise ~28% (14% each) of the unfolded molecules. From the experiments reported in this study, it appears that it is necessary to postulate only two unfolded forms of single-chain monellin,  $U_1$  (15%) and  $U_2$  (85%). It therefore seems that  $U_2$  comprises  $U_{TT}$  and either  $U_{TC}$  or  $U_{CT}$  ( $69 + 14 = 83\%$ ).  $U_1$  would then comprise  $U_{CC}$  and either  $U_{CT}$  or  $U_{TC}$ . At present, it is not known whether it is Pro 41 or Pro 93 which is cis in  $U_{TC}$  and which is cis in  $U_{CT}$ , but in either case it would mean that these two X-Pro bonds are not equivalent.

*Proline Isomerization during Refolding.* If the above assignment of  $U_{CC}$ ,  $U_{CT}$ ,  $U_{TC}$ , and  $U_{TT}$  to the two observed unfolded forms,  $U_1$  and  $U_2$ , is correct, then it is expected that a very slow folding phase incorporating trans to cis proline isomerization should occur during the folding of both  $U_1$  and  $U_2$ . As discussed above, all four kinetic folding intermediates appear to transform to N in a very slow folding phase with a rate characteristic of proline isomerization. Proline isomerization reactions are characterized by high enthalpies of activation, but direct measurement of the temperature dependence of the very slow phase is not easy because it is silent to any optical change. Nevertheless, the observation that its rate is independent of the concentration of GdnHCl in which refolding is carried out (see Results), as is that of proline isomerization, suggests that no major structure formation reaction occurs during the very slow phase of folding. It has not been possible to detect an

expected rapid phase in the formation of N from the  $\sim 3\%$  of the unfolded molecules present as  $U_{CC}$ , given the error of measurement in the double-jump, interrupted refolding experiments, which report directly on the formation of N.

It is possible that  $I_{E1}$  and  $I_{E2}$ , both of which originate from  $U_2$ , differ in having two and one proline residues, respectively, in the non-native trans configuration. Then, the two pathways originating from  $I_{E1}$  might differ in which of the two proline residues undergoes trans to cis isomerization first and which second. Such an explanation has been suggested in the case of ribonuclease T1 (60), but in that case, the fastest steps on the two pathways had rates corresponding to those typically seen for proline isomerization. This explanation for two competing pathways defined on the basis of which proline residue undergoes isomerization first is very unlikely in the case of monellin because the observed fast rates of formation of  $I_{F1}$  and  $I_{F2}$  ( $\sim 1 \text{ s}^{-1}$ ) are much too fast for a proline isomerization reaction. Hence, it appears that while proline isomerization does give rise to heterogeneity in unfolded monellin, it does not give rise to the heterogeneity seen during the fast phase of folding.

**Mechanism of Refolding.** On the basis of the above discussion the following mechanism is proposed to describe the folding of single-chain monellin in strongly stabilizing conditions:



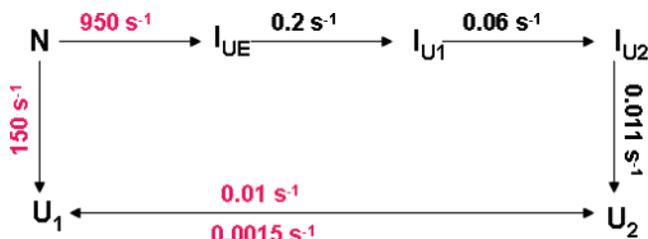
All of the kinetic data on refolding in 0.5 M GdnHCl were simulated numerically using the program KINSIM (61). Although nearly all of the apparent rate constants in the scheme have been determined experimentally, it was still necessary to make the following minimal assumptions: (1) It is assumed that  $I_E$  consists of two species,  $I_{E1}$  and  $I_{E2}$ , because this allows kinetic partitioning of the folding molecules along either of the two competing pathways, slow and fast. Values for the apparent rate constants of the  $U_2 \rightarrow I_{E1}$  and  $U_2 \rightarrow I_{E2}$  reactions have been assumed on the basis of the observation (Figure 3d) that  $I_E$  is populated at 1 ms of refolding. Previous sub-millisecond measurements have indicated that  $I_E$  is formed within 300  $\mu\text{s}$  of refolding in the absence of denaturant (48). (2) It is assumed that the  $U_1 \rightleftharpoons U_2$  reaction is a proline isomerization reaction. Hence, the forward and backward rate constants for this reaction have been assigned values typically seen for proline isomerization reactions (62, 63) such that the equilibrium population of  $U_1$  and  $U_2$  are 15 and 85%, respectively. (3) It is assumed that the fluorescence properties of  $I_{VF}$ ,  $I_S$ ,  $I_{F1}$ , and  $I_{F2}$  are the same as those of N and that the fluorescence properties of  $I_{E1}$  and  $I_{E2}$  are the same as those of either  $U_1$  or  $U_2$ . The

bases for these assumptions have been discussed above. (4) Although all steps in the mechanism are completely reversible, it is assumed that in strongly stabilizing conditions for folding, such as in 0.5 M GdnHCl, the reverse rate constants are very small in comparison to the forward rate constant. Hence, the reverse (unfolding) apparent rate constants have been ignored except for the step involving proline isomerization in the unfolded protein. (5) It is assumed that there are only two unfolded forms,  $U_1$  and  $U_2$  (see the unfolding mechanism below).

It is seen in Figure 9 that the refolding mechanism above, incorporating competing pathways, accounts well for all of the experimental data for refolding in 0.5 M GdnHCl, from single-jump as well as double-jump experiments.

When refolding is carried out in 1.5 M GdnHCl, the intermediate  $I_{VF}$  appears not to be populated to any significant extent, and as a result refolding occurs only in the slow and fast kinetic phases with apparent rate constants of 0.0024 and 0.025  $\text{s}^{-1}$ , respectively. The mechanism of refolding in 1.5 M GdnHCl appears to have simplified to  $U_1 \rightleftharpoons U_2 \rightarrow N$ , with  $U_2$  refolding as described in the refolding mechanism above. The direct route from  $U_1$  to N is not expected to be operative to any significant extent because the apparent rate constant for the  $U_1 \rightarrow N$  reaction (0.0024  $\text{s}^{-1}$ ) is slower than the apparent rate constant of the  $U_1 \rightarrow U_2$  reaction (0.01  $\text{s}^{-1}$ ). As a result, the 15% unfolded molecules present as  $U_1$  are channeled into the  $U_2 \rightarrow N$  folding pathway(s).

**Mechanism of Unfolding.** The following mechanism is proposed to describe the unfolding of single-chain monellin in strongly unfolding conditions.



All of the kinetic data on the unfolding of single-chain monellin in 5 M GdnHCl have been simulated numerically, using the program KINSIM (61). Values for nearly all of the steps in this mechanism have been determined experimentally in this study, but it was still necessary to make the following assumptions. (1) The value of the rate constant of the  $N \rightarrow I_{UE}$  transition was assumed to be sufficiently faster than the  $N \rightarrow U_1$  reaction so that 85% of the unfolding protein molecules utilized the  $N \rightarrow U_2$  pathway. It is known experimentally that the  $N \rightarrow U_2$  is complete within a few milliseconds (Figure 4). (2) It is assumed that the fluorescence properties of  $I_{UE}$ ,  $I_{U1}$ , and  $I_{U2}$  are the same as those of N, as discussed above. (3) The rate constants for the  $U_1 \rightleftharpoons U_2$  reaction have been assumed to have values representative of proline isomerization reactions (see above). (4) Although all steps in the above mechanism are reversible, it is assumed that in strongly unfolding conditions, such as in 5 M GdnHCl, the reverse reactions can be ignored.

It is clear from Figure 10 that the unfolding mechanism above, incorporating multiple pathways, accounts well for all of the experimental refolding data, from both single-jump and double-jump experiments.

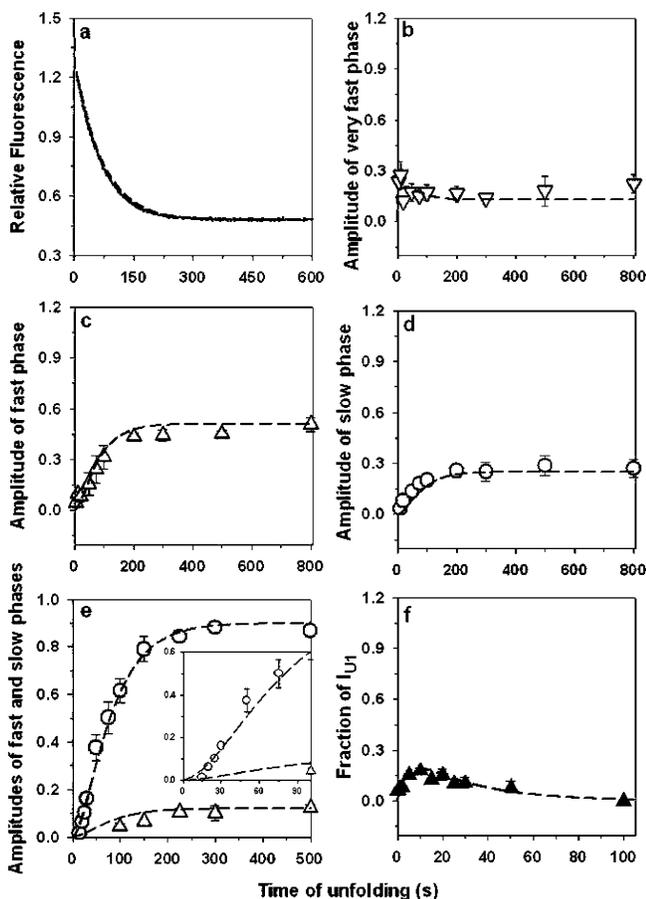


FIGURE 10: Kinetic simulations supporting the unfolding mechanism of single-chain monellin. In each panel, experimental data from Figures 4, 7, and 8 are shown along with the kinetic simulation (dashed line) to the unfolding mechanism, carried out as described in the text: (a) fluorescence-monitored kinetic trace of unfolding in 5 M GdnHCl (solid line); (b) relative amplitude of the very fast phase ( $\nabla$ ) (the simulation shows how the amount of  $U_1$  changes with time of unfolding in 5 M GdnHCl); (c) relative amplitude of the fast phase of refolding in 0.5 M GdnHCl ( $\Delta$ ), as a function of the time of prior unfolding in 5 M GdnHCl; (d) relative amplitude of the slow phase of refolding in 0.5 M GdnHCl ( $\circ$ ), as a function of the time of prior unfolding in 5 M GdnHCl (the simulations in panels c and d show how the amount of  $U_2$ , which gives rise to the fast and slow phases of refolding in 0.5 M GdnHCl, changes with time of unfolding); (e) relative amplitude of the fast ( $\Delta$ ) and slow ( $\circ$ ) phases of refolding in 1.5 M GdnHCl, as a function of the time of prior unfolding in 5 M GdnHCl (the simulations show how the amount of  $U_2$ , which gives rise to the fast and slow phases of refolding in 1.5 M GdnHCl, changes with the time of unfolding in 5 M GdnHCl); (f) relative amount of  $I_{U1}$  ( $\blacktriangle$ ) as a function of the time of unfolding in 5 M GdnHCl.

**Kinetic Partitioning Leads to Competing Folding Pathways.** In protein folding studies, it has been relatively easy to detect parallel folding pathways arising from multiple unfolded forms that originate from proline isomerization in the unfolded protein. As a consequence, it has been common to attribute any heterogeneity observed in folding reactions to proline isomerization reactions, even though studies with several proteins show that proline isomerization is not the sole cause of such heterogeneity in folding reactions (19, 22, 59). Nevertheless, it has been far more difficult in folding studies to detect competing folding pathways that originate from a single unfolded form or intermediate. In this study, it has been shown directly, by the use of a double-jump assay for quantifying two intermediates which form in parallel, that

the fast phase of fluorescence change, representing the major phase of folding in stabilizing conditions, occurs on two competing pathways. These competing pathways arise from kinetic partitioning from the same early folding intermediate. The kinetic partitioning occurs too quickly to be attributed to proline isomerization (see above). It has also been suggested that kinetic partitioning from the major unfolded protein leads to the fast and slow phases of folding. This kinetic partitioning occurs even more quickly, again far too quickly to be attributed to proline isomerization. It also appears that kinetic partitioning leads to two pathways for unfolding. The direct demonstration of kinetic partitioning leading to competing folding pathways is the major result of the present study. Understanding the structural origin of such kinetic partitioning is an important future goal.

**Significance of Multiple Pathways for Folding and Unfolding.** The major result of this study is the demonstration that multiple pathways are available to the protein for folding and unfolding. The possibility of multiple pathways for folding was brought out elegantly in a model for folding in which the folding of a protein was likened to the assembly of a jigsaw puzzle: assembly can start anywhere (multiple pathways), but the end product is always the completed puzzle (native state) (64). Since then, multiple folding and unfolding pathways have been reported for several proteins. It appears that for several proteins, only one of a few available pathways is operative under a given set of experimental conditions. For example, in the case of barstar, as conditions become more stabilizing, one pathway dominates at the expense of the other (19). In the case of thioredoxin, folding is channeled along one out of several available folding pathways, in the presence of the chaperone GroEL (65). In the case of the monomeric  $\lambda$  repressor, the utilization of different folding pathways has been reported (66) to be highly sensitive to a change in sequence (mutation). Here, it has been shown, in the case of the single-chain monellin, MNEI, that as conditions become less stabilizing, the slow folding pathway predominates over the other available folding pathways. In making multiple pathways available for the folding or unfolding of a protein, evolution has conferred robustness on the folding process: if mutation or a change in folding conditions disables one folding pathway, another pathway can take over. It appears that evolution has achieved such robustness by allowing the folding reactions along any one pathway to occur in steps defined by folding intermediates. Any one pathway can then be selected over another by a change in folding conditions, because such a change in folding conditions would change the relative stabilities of the folding intermediates on the different pathways.

**Conclusion.** In summary, it is shown that monellin utilizes four parallel folding pathways originating from two unfolded states. Although part of the heterogeneity seen can be attributed to the presence of more than one unfolded form, the major result of the current study is the demonstration of kinetic partitioning leading to competing pathways for folding. One of the pathways originates from one unfolded form, and the other pathways originate from the second unfolded form. We show that the pathways originating from the second unfolded form all compete with one another and that a change in folding conditions leads to greater utilization of the slow pathway at the expense of the other pathways.

The results bring out the roles played by folding intermediates in directing the utilization of alternative folding pathways when many pathways are available.

## ACKNOWLEDGMENT

We thank P. A. Temussi and H. Iijima for sending to us the gene of the single-chain variant of monellin, MNEI, used in this study. We thank members of our laboratory for discussions and comments on the manuscript.

## REFERENCES

- Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry* 30, 10428–10435.
- Jackson, S. E. (1998) How do small single-domain proteins fold? *Fold. Des.* 3, R81–91.
- Bhuyan, A. K., and Udgaonkar, J. B. (2001) Folding of horse cytochrome *c* in the reduced state, *J. Mol. Biol.* 312, 1135–1160.
- Baker, D. (2000) A surprising simplicity to protein folding, *Nature* 405, 39–42.
- Fersht, A. R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism, *Proc. Natl. Acad. Sci. U.S.A.* 97, 1525–1529.
- Inaba, K., Kobayashi, N., and Fersht, A. R. (2000) Conversion of two-state to multi-state folding kinetics on fusion of two protein foldons, *J. Mol. Biol.* 302, 219–233.
- Spudich, G. M., Miller, E. J., and Marqusee, S. (2004) Destabilization of the *Escherichia coli* RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism, *J. Mol. Biol.* 335, 609–618.
- Pradeep, L., and Udgaonkar, J. B. (2002) Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar, *J. Mol. Biol.* 324, 331–347.
- Pradeep, L., and Udgaonkar, J. B. (2004) Osmolytes induce structure in an early intermediate on the folding pathway of barstar, *J. Biol. Chem.* 279, 40303–40313.
- Gruebele, M. (2002) Protein folding: the free energy surface, *Curr. Opin. Struct. Biol.* 12, 161–168.
- Sridevi, K., Lakshmikanth, G. S., Krishnamoorthy, G., and Udgaonkar, J. B. (2004) Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble, *J. Mol. Biol.* 337, 699–711.
- Sinha, K. K., and Udgaonkar, J. B. (2005) Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition, *J. Mol. Biol.* 353, 704–718.
- Agashe, V. R., and Udgaonkar, J. B. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride, *Biochemistry* 34, 3286–3299.
- Sosnick, T. R., Shtilerman, M. D., Mayne, L., and Englander, S. W. (1997) Ultrafast signals in protein folding and the polypeptide contracted state, *Proc. Natl. Acad. Sci. U.S.A.* 94, 8545–8550.
- Hamada, D., Segawa, S., and Goto, Y. (1996) Non-native  $\alpha$ -helical intermediate in the refolding of  $\beta$ -lactoglobulin, a predominantly  $\beta$ -sheet protein, *Nat. Struct. Biol.* 3, 868–873.
- Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002) Im7 folding mechanism: misfolding on a path to the native state, *Nat. Struct. Biol.* 9, 209–216.
- Bollen, Y. J., Kamphuis, M. B., and van Mierlo, C. P. (2006) The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates, *Proc. Natl. Acad. Sci. U.S.A.* 103, 4095–4100.
- Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature* 358, 302–307.
- Shastri, M. C., and Udgaonkar, J. B. (1995) The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates, *J. Mol. Biol.* 247, 1013–1027.
- Zaidi, F. N., Nath, U., and Udgaonkar, J. B. (1997) Multiple intermediates and transition states during protein unfolding, *Nat. Struct. Biol.* 4, 1016–1024.
- Wu, Y., and Matthews, C. R. (2002) Parallel channels and rate-limiting steps in complex protein folding reactions: prolyl isomerization and the  $\alpha$  subunit of Trp synthase, a TIM barrel protein, *J. Mol. Biol.* 323, 309–325.
- Kamagata, K., Sawano, Y., Tanokura, M., and Kuwajima, K. (2003) Multiple parallel-pathway folding of proline-free Staphylococcal nuclease, *J. Mol. Biol.* 332, 1143–1153.
- Wright, C. F., Steward, A., and Clarke, J. (2004) Thermodynamic characterisation of two transition states along parallel protein folding pathways, *J. Mol. Biol.* 338, 445–451.
- Butler, J. S., and Loh, S. N. (2005) Kinetic partitioning during folding of the p53 DNA binding domain, *J. Mol. Biol.* 350, 906–918.
- Leeson, D. T., Gai, F., Rodriguez, H. M., Gregoret, L. M., and Dyer, R. B. (2000) Protein folding and unfolding on a complex energy landscape, *Proc. Natl. Acad. Sci. U.S.A.* 97, 2527–2532.
- Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol.* 4, 10–19.
- Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: the energy landscape perspective, *Annu. Rev. Phys. Chem.* 48, 545–600.
- Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., and Karplus, M. (2000) Understanding protein folding via free-energy surfaces from theory and experiment, *Trends Biochem. Sci.* 25, 331–339.
- Silow, M., and Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates, *Proc. Natl. Acad. Sci. U.S.A.* 94, 6084–6086.
- Ganesh, C., Zaidi, F. N., Udgaonkar, J. B., and Varadarajan, R. (2001) Reversible formation of on-pathway macroscopic aggregates during the folding of maltose binding protein, *Protein Sci.* 10, 1635–1644.
- Galani, D., Fersht, A. R., and Perrett, S. (2002) Folding of the yeast prion protein Ure2: kinetic evidence for folding and unfolding intermediates, *J. Mol. Biol.* 315, 213–227.
- Hamid Wani, A., and Udgaonkar, J. B. (2006) HX-ESI-MS and optical studies of the unfolding of thioredoxin indicate stabilization of a partially unfolded, aggregation-competent intermediate at low pH, *Biochemistry* 45, 11226–11238.
- Uversky, V. N., Li, J., and Fink, A. L. (2001) Evidence for a partially folded intermediate in  $\alpha$ -synuclein fibril formation, *J. Biol. Chem.* 276, 10737–10744.
- Jahn, T. R., Parker, M. J., Homans, S. W., and Radford, S. E. (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate, *Nat. Struct. Mol. Biol.* 13, 195–201.
- Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem.* 75, 333–366.
- Morris, J. A., and Cagan, R. H. (1972) Purification of monellin, the sweet principle of *Dioscoreophyllum cumminsii*, *Biochim. Biophys. Acta* 261, 114–122.
- Sung, Y. H., Hong, H. D., Cheong, C., Kim, J. H., Cho, J. M., Kim, Y. R., and Lee, W. (2001) Folding and stability of sweet protein single-chain monellin. An insight to protein engineering, *J. Biol. Chem.* 276, 44229–44238.
- Xue, W. F., Carey, J., and Linse, S. (2004) Multi-method global analysis of thermodynamics and kinetics in reconstitution of monellin, *Proteins* 57, 586–595.
- Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Thulin, E., and Linse, S. (2006) Intra- versus intermolecular interactions in monellin: contribution of surface charges to protein assembly, *J. Mol. Biol.* 358, 1244–1255.
- Konno, T., Murata, K., and Nagayama, K. (1999) Amyloid-like aggregates of a plant protein: a case of a sweet-tasting protein, monellin, *FEBS Lett.* 454, 122–126.
- Konno, T. (2001) Multistep nucleus formation and a separate subunit contribution of the amyloidogenesis of heat-denatured monellin, *Protein Sci.* 10, 2093–2101.
- Ogata, C., Hatada, M., Tomlinson, G., Shin, W. C., and Kim, S. H. (1987) Crystal structure of the intensely sweet protein monellin, *Nature* 328, 739–742.
- Somoza, J. R., Jiang, F., Tong, L., Kang, C. H., Cho, J. M., and Kim, S. H. (1993) Two crystal structures of a potentially sweet protein. Natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution, *J. Mol. Biol.* 234, 390–404.
- Tancredi, T., Iijima, H., Saviano, G., Amodeo, P., and Temussi, P. A. (1992) Structural determination of the active site of a sweet protein. A 1H NMR investigation of pMNEI, *FEBS Lett.* 310, 27–30.
- Murzin, A. G. (1993) Sweet-tasting protein monellin is related to the cystatin family of thiol proteinase inhibitors, *J. Mol. Biol.* 230, 689–694.

46. Kim, S. H., Kang, C. H., Kim, R., Cho, J. M., Lee, Y. B., and Lee, T. K. (1989) Redesigning a sweet protein: increased stability and renaturability, *Protein Eng.* 2, 571–575.
47. Spadaccini, R., Crescenzi, O., Tancredi, T., De, Casamassimi, N., Saviano, G., Scognamiglio, R., Di Donato, A., and Temussi, P. A. (2001) Solution structure of a sweet protein: NMR study of MNEI, a single chain monellin, *J. Mol. Biol.* 305, 505–514.
48. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T., Akiyama, S., and Fujisawa, T. (2005) Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin, *Proc. Natl. Acad. Sci. U.S.A.* 102, 2748–2753.
49. Udgaonkar, J. B., and Baldwin, R. L. (1995) Nature of the early folding intermediate of ribonuclease A, *Biochemistry* 34, 4088–4096.
50. Baldwin, R. L. (1995) The nature of protein folding pathways: the classical versus the new view, *J. Biomol. NMR* 5, 103–109.
51. Lyubovitsky, J. G., Gray, H. B., and Winkler, J. R. (2002) Mapping the cytochrome C folding landscape, *J. Am. Chem. Soc.* 124, 5481–5485.
52. Juneja, J., and Udgaonkar, J. B. (2002) Characterization of the unfolding of ribonuclease a by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding, *Biochemistry* 41, 2641–2654.
53. Krishna, M. M., and Englander, S. W. (2007) A unified mechanism for protein folding: predetermined pathways with optional errors, *Protein Sci.* 16, 449–464.
54. Spadaccini, R., Trabucco, F., Saviano, G., Picone, D., Crescenzi, O., Tancredi, T., and Temussi, P. A. (2003) The mechanism of interaction of sweet proteins with the T1R2-T1R3 receptor: evidence from the solution structure of G16A-MNEI, *J. Mol. Biol.* 328, 683–692.
55. Garell, R. L., and Baldwin, R. L. (1973) Both the fast and slow refolding reactions of ribonuclease A yield native enzyme, *Proc. Natl. Acad. Sci. U.S.A.* 70, 3347–3351.
56. Schmid, F. X. (1986) Fast-folding and slow-folding forms of unfolded proteins, *Methods Enzymol.* 131, 70–82.
57. Kiefhaber, T., Kohler, H. H., and Schmid, F. X. (1992) Kinetic coupling between protein folding and prolyl isomerization. I. Theoretical models, *J. Mol. Biol.* 224, 217–229.
58. Shastry, M. C., Agashe, V. R., and Udgaonkar, J. B. (1994) Quantitative analysis of the kinetics of denaturation and renaturation of barstar in the folding transition zone, *Protein Sci.* 3, 1409–1417.
59. Bradley, C. M., and Barrick, D. (2005) Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory, *J. Mol. Biol.* 352, 253–265.
60. Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X. (1990) Folding of ribonuclease T1. 2. Kinetic models for the folding and unfolding reactions, *Biochemistry* 29, 3061–3070.
61. Barshop, B. A., Wrenn, R. F., and Frieden, C. (1983) Analysis of numerical methods for computer simulation of kinetic processes: development of KINSIM—a flexible, portable system, *Anal. Biochem.* 130, 134–145.
62. Wedemeyer, W. J., Welker, E., and Scheraga, H. A. (2002) Proline cis–trans isomerization and protein folding, *Biochemistry* 41, 14637–14644.
63. Wu, Y., and Matthews, C. R. (2003) Proline replacements and the simplification of the complex, parallel channel folding mechanism for the  $\alpha$  subunit of Trp synthase, a TIM barrel protein, *J. Mol. Biol.* 330, 1131–1144.
64. Harrison, S. C., and Durbin, R. (1985) Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. U.S.A.* 82, 4028–4030.
65. Bhutani, N., and Udgaonkar, J. B. (2001) GroEL channels the folding of thioredoxin along one kinetic route, *J. Mol. Biol.* 314, 1167–1179.
66. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) Protein folding dynamics: quantitative comparison between theory and experiment, *Biochemistry* 37, 5337–5343.

BI701142A