BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

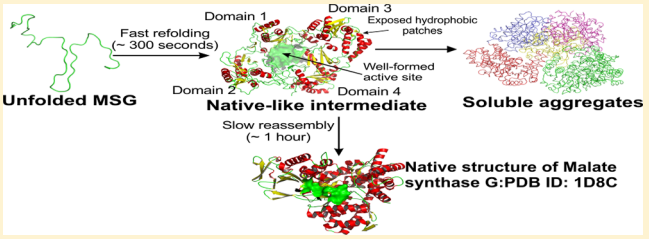
pubs.acs.org/biochemistry

Functional Intermediate in the Refolding Pathway of a Large and Multidomain Protein Malate Synthase G

Vinay Dahiya and Tapan K. Chaudhuri*

Kusuma School of Biological Sciences, Indian Institute of Technology Delhi, Hauz Khas, New Delhi − 110016, India

ABSTRACT: Despite their prevalence in biological systems, information about the folding pathways of large and multidomain proteins is meager, as they often unfold irreversibly under in vitro conditions which make their folding studies difficult or even impossible. The folding mechanism of a large (82 kDa) and multidomain protein Malate synthase G (MSG) has been demonstrated in the present study using intrinsic tryptophan fluorescence, enzymatic activity, and extrinsic fluorophore ANS as probes for monitoring the refolding process. Refolding of MSG is found to occur in three kinetic phases. Denatured MSG forms a collapsed state in the burst phase of refolding, which then gives rise to an active intermediate having the same tryptophan fluorescence and enzymatic activity as native MSG in the slow phase. Native topology of MSG is formed from the active intermediate in the very slow phase of refolding which is silent to tryptophan fluorescence change and is susceptible to aggregation at higher protein concentrations. Dependence of rates of very slow phase on GdnHCl concentration suggests that it is not solely a cis/trans proline isomerization limited process but might involve an additional folding event of the domains, not forming the active site of the protein. In light of the above findings, the appearance of a functional intermediate during refolding of MSG was predicted to be an instance of weak interdomain cooperativity. This work has significant implications in the characterization of the refolding intermediates of multidomain proteins in general and MSG in particular, where weak interdomain cooperativity might contribute toward generation of a functional intermediate during its refolding.



Protein folding studies to date are mainly focused on small, single domain proteins which follow a reversible two-state mechanism N ↔ U.[1−7] However, a majority of proteins in the cell belong to the class of large multidomain family[8] which often unfold irreversibly in vitro due to the presence of competing side reactions such as aggregation and misfolding. A prerequisite in elucidating the protein folding pathways is the possibility of unfolding the protein reversibly, which in the case of large and multidomain proteins is extremely difficult because of the formation of partially folded intermediates (nonnative conformers) that tend to self-associate (aggregate) into disordered complexes.[9] It is assumed that stronger interdomain interactions in multidomain proteins result in protein misfolding and aggregation, which ultimately leads to low protein refolding yields in vitro.[10] Thus, knowledge of in vitro conditions that lead to efficient refolding of large and multidomain proteins is still of interest as these provide the experimental basis for studying the pathways of folding for such proteins. One such large and multidomain protein is *Escherichia coli* Malate synthase G or MSG, an 82 kDa monomeric globular enzyme[11] that catalyzes the condensation of glyoxylate and acetyl coenzyme A, producing malate, an intermediate in the citric acid cycle.[12] Malate synthase G comprises of four domains;[13] its structure is based on $\beta 8/\alpha 8$ (TIM barrel) fold, which is centrally located with an inserted β-sheet domain within the fold of the barrel and an N-terminal and the α-helical C-terminal domain flanking from it on both ends. The structure of MSG is known and has been elucidated both by X-ray crystallography[13] and NMR spectroscopy.[14] The study of its folding will yield many useful insights into the folding mechanism of other large and multidomain proteins which are often not studied because of various difficulties associated with them such as irreversible unfolding, formation of aggregates, requirement of extremely low concentration during refolding due to which sometimes it becomes difficult to monitor the process spectroscopically, complex thermodynamic and kinetic analysis, etc. Since MSG has 12 tryptophan residues, it is possible for us to do spectroscopic studies at extremely low concentrations of MSG. Besides, it has a well-developed functional assay through which we could monitor refolding using recovery of enzymatic activity as a probe. From our in vivo experiments, it was known that despite being large and multidomain in nature, almost the whole over-expressed MSG remains soluble in the *E. coli* cytosol during its overexpression (unpublished observations, V.D., and T.K.C.). This observation set the stage for trying in vitro refolding of MSG as chances appeared promising for folding it reversibly in vitro which was a precondition to understand its folding pathway.

The process of protein folding is characterized by a high degree of cooperativity at least for small (<20 kDa) and single domain proteins where the population of partially folded intermediates never exceeds 5%.[15] However, folding/unfolding

Received: March 13, 2013
Revised: May 9, 2013
Published: May 29, 2013

   dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

transitions of multidomain proteins are often characterized by the presence of a few partially folded equilibrium intermediates that have conformations in which some domains are folded and other ones are unfolded.[16,17] Individual domains fold independently when the interaction between domains is weak;[18] however, cooperative folding is observed when there is extensive interaction among domains.[19,20] Hence, in a multidomain protein there is always a possibility that the domains could be arranged either in a closer fashion to provide cooperativity in the conformation, or in a scattered way so that the interdomain cooperativity could be lost or weakened. Keeping this view in mind, in the present case of investigating the refolding mechanism of large multidomain protein MSG, we have attempted to demonstrate the issue of its interdomain cooperativity. It is very essential to have data related to domain wise folding for the assignment of the cooperativity factor in the whole multidomain protein. However, in the absence of folding data of individual domains in MSG, we have attempted to explore the overall cooperativity issue using some of the events such as recovery of activity, intrinsic fluorescence, native-like topology, etc. The basis of deducing cooperativity linked information in multidomain proteins using the mentioned properties may stand on the fact that the appearance of activity during refolding may not coincide with the recovery of nativelike topology, or the recovery of the intrinsic fluorescence signal may not match with the time frame of recovery of activity, or even the monitoring parameters for unfolding processes (secondary and tertiary structure loss) may not overlap each other. Hence, we demonstrated the interdomain cooperativity of MSG with the help of various physicochemical parameters obtained while monitoring the equilibrium unfolding and refolding processes. Identification and characterization of unfolding and refolding intermediates in a protein folding process always provide meaningful information to deduce the finer routes for the said events. Our experimental findings have enabled us to characterize different intermediate species accumulating during various steps of the refolding process of MSG. We have finally demonstrated the overall mechanistic aspects of MSG folding and proposed a model that fits with the events taking place in the pathways.

MATERIALS AND METHODS

All buffers and reagents used in experiments were of ultrapure grade. GdnHCl purchased from USB Corporation (USA). All the experiments were performed in MSG refolding buffer comprised of 20 mM Tris, 300 mM NaCl, 10 mM $MgSO_4$, 1 mM TCEP-HCl, and 10% glycerol at pH 7.9.

Methods. *MSG Cloning and Overexpression.* A 5.2-kb fragment clone in pCS19 vector (generous gift from Prof. Laura Baldoma, University of Barcelona, Spain) containing glcB (2.1 kb), gene encoding MSG, was amplified by PCR using the following oligonucleotide primers: 5′-AGC CCA TGG AGC ATG AGT CAA ACC ATA ACC-3′ and 5′-AAA CTC GAG CTT ATG ACT TTC TTT TTC GCG-3′. The primers were designed using the glcB gene sequence from NCBI (accession number X74547). The resulting PCR fragment was ligated into a pET-28b expression vector (Novagen) at Nco I-Xho I site to link the gene to the C-terminal hexahistidine tag. The recombinant vector was used to transform BL21-DE3 cells for MSG overexpression. The cells were grown in LB plus kanamycin (30 μg/mL) at 37 °C and induced with 1 mM IPTG when the O.D. reached ∼0.6.

Purification of MSG. The expression of histidine tagged MSG was induced with 1 mM IPTG for 6 h at 37 °C. The cells were harvested by centrifugation, and the pellet was resuspended in lysis buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 1 mM DTT, 1 mg/mL lysozyme and 1 mM PMSF). After 1 h incubation at 4 °C, cells were lysed by ultrasonication using Branson sonifier 250 (Netherlands). The lysate was centrifuged at 48000*g* for 1 h. After filtering through a 0.2 μm Millipore filter, the supernatant was passed through a Ni-NTA chelating column, HiTrap HP (Pharmacia, Piscataway, NJ, USA) equilibrated in 20 mM sodium phosphate buffer pH 7.4, 500 mM NaCl, 1 mM DTT in AKTA FPLC unit. The his-tagged MSG was eluted using a linear gradient of 0−500 mM imidazole where it elutes at 300 mM imidazole. Fractions containing MSG were identified by SDS-PAGE, and those having greater than 99% purity were pooled and dialyzed against 20 mM Tris buffer, pH 7.4, 100 mM NaCl, 1 mM TCEP-HCl. Dialyzed protein was concentrated using Amicon tubes (from Millipore, Billerica, MA, U.S.A.) with 30 kDa molecular weight cut off membrane. Concentration of purified protein was determined using an extinction coefficient of 92 820 M^{-1} cm^{-1} at 280 nm. The extinction coefficient of MSG was computed using the ProtParam tool of ExPASy (http://web.expasy.org/protparam/).

Equilibrium Unfolding Experiments. For equilibrium unfolding experiments, MSG to a final concentration of 0.25 μM was incubated in refolding buffer containing different concentrations of GdnHCl (0−6 M) or urea (0−8 M) for 6 h at 25 °C. Equilibrium unfolding of MSG was monitored by both intrinsic tryptophan fluorescence and far-UV CD. For tryptophan fluorescence, a Biologic MOS-450 optical system with a 1 cm path length cuvette was used. Protein was excited at 295 nm, and tryptophan fluorescence was recorded at 340 nm using band-pass filter (Asahi spectra). Far-UV CD experiments were performed on JASCO J-815 CD polarimeter (JASCO, Japan) using an optical cuvette of path length 0.2 cm.

In Vitro Unfolding and Refolding of MSG. The protein was unfolded in 20 mM Tris, 300 mM NaCl, 10 mM $MgSO_4$, 1 mM TCEP HCl, 10% glycerol buffer, pH 7.9 containing 3 M GdnHCl for 2 h. Unfolding was monitored by loss of enzymatic activity, far UV-CD, tryptophan and ANS fluorescence. Refolding was initiated by rapid dilution of GdnHCl denatured protein in refolding buffer, which was the same as unfolding buffer without GdnHCl. Refolding mixture was then incubated at 25 °C for 1 h, and the refolding yield was calculated from enzymatic activity recovered by the refolded protein in 1 h as a percentage of native protein. Final protein concentration in the refolding buffer was always kept less than 0.5 μM.

Intrinsic Tryptophan and ANS Fluorescence Monitored Refolding Kinetics Experiments. Refolding kinetics of MSG was monitored using intrinsic tryptophan and ANS-bound fluorescence of protein as probes. Refolding was initiated by rapid dilution of denatured protein in refolding buffer to different final GdnHCl concentrations while maintaining the final protein concentration at 0.25 μM, using either SFM-400 (Biologic) stopped flow machine or manual mixing. Manual mixing experiments were done in a Biologic MOS-450 optical system (NCBS, Bangalore) using a 1 cm path length cuvette. Dead time of 6 ms could be achieved in stopped flow using 0.15 cm path length cuvette while in manual mixing dead time was 10 s. Refolding kinetic traces were monitored by measurement of the change in tryptophan fluorescence at 340 ± 10 nm, using a band-pass filter (Asahi spectra). The excitation wavelength

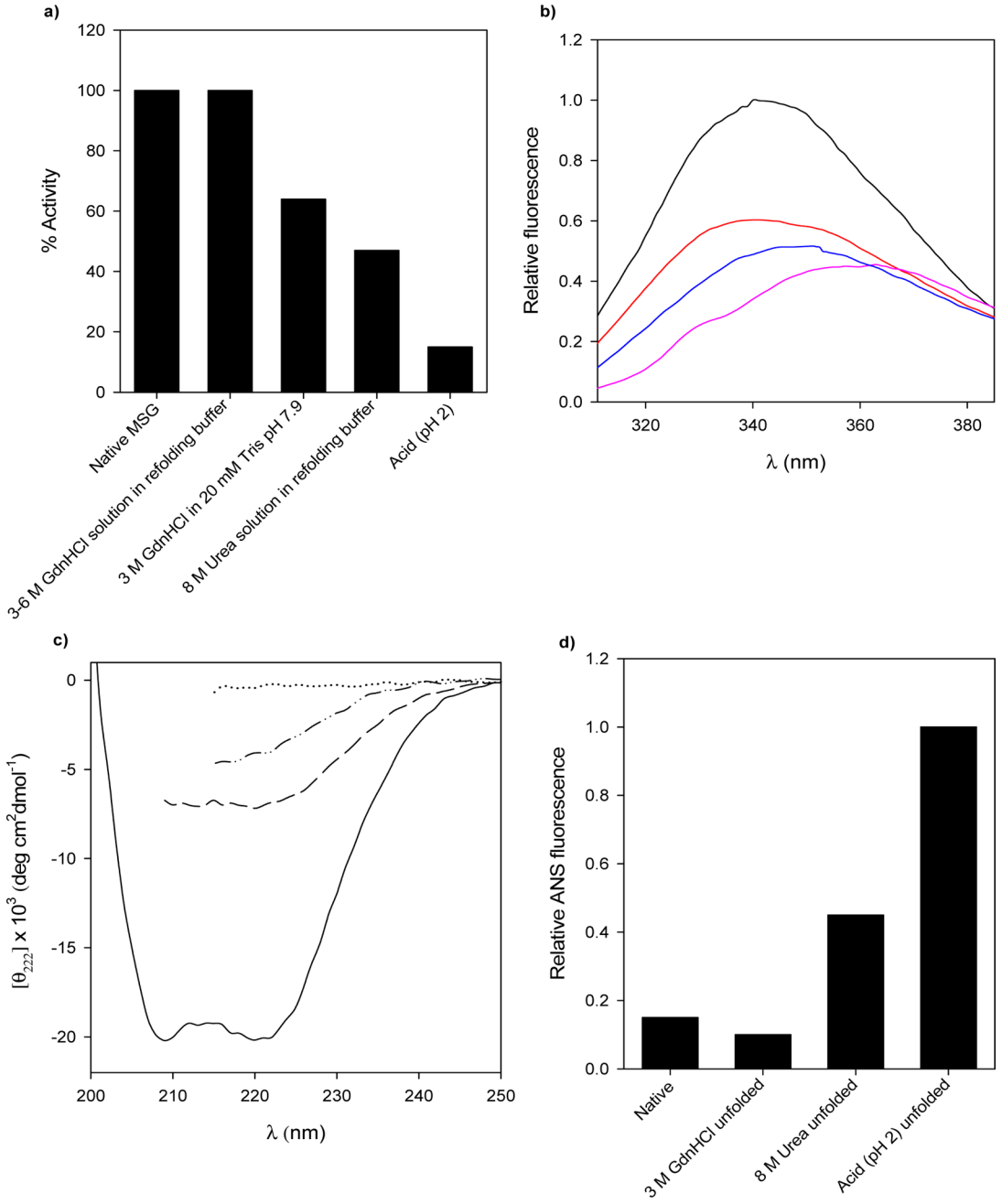


Figure 1. Optimization of in vitro unfolding of MSG. (a) Histogram showing the effect of different unfolding conditions on final refolding yield of MSG. MSG was unfolded in different mentioned denaturants. Refolding was monitored at 25 °C by enzymatic activity assay of MSG. Refolding yield was calculated from activity recovered by refolded protein in 1 h as a percentage of the same amount of native protein activity. Refolding buffer used was 20 mM Tris pH 7.9, 300 mM NaCl, 10 mM $MgSO_4$, 1 mM TCEP HCl, 10% glycerol. For acid unfolding, MSG was incubated in 25 mM glycine phosphate (pH 2) at 25 °C for 2 h. Final MSG concentration during refolding was always kept less than 0.5 μM. (b) Characterization of different unfolded states by tryptophan fluorescence. Denatured states present in acid (red curve), 8 M urea (blue curve), 3 M GdnHCl (pink curve) differ from each other as is evident from the difference in λ_{max} for their tryptophan fluorescence emission spectra. Native MSG curve is shown in black. (c) Far UV-CD spectra of (——) native state; (—— —— ——) acid unfolded; (——·· ——··) 8 M urea; (·····) 3 M GdnHCl. (d) ANS fluorescence of different unfolded states. After 2 h of denaturation, unfolded mixtures were incubated with 50 μM ANS for 5 min at 25 °C before recording their fluorescence. Protein solutions were excited at 350 nm, and emission was recorded at 450 nm. All the fluorescence values were normalized with respect to ANS fluorescence of the acid unfolded state.

was 295 nm, and the excitation slit width was 2 nm. Protein concentration-dependent refolding kinetic studies were performed by diluting unfolded protein to different final protein concentrations and a fixed final GdnHCl concentration. For ANS-bound fluorescence monitored kinetics, refolding kinetic traces were monitored by measurement of decrease in ANS fluorescence at 450 nm, using a band-pass filter (Asahi spectra). The excitation wavelength was 350 nm. GdnHCl concentration was varied between 0.1 and 0.325 M by first mixing unfolding buffer (3 M GdnHCl) containing 55 μM ANS with refolding buffer also containing 55 μM ANS in different proportions. Then a fixed volume of unfolded protein was added to the

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

above mixture such that the final protein concentration was always 0.25 μM. This procedure ensured that the concentration of protein (0.25 μM) and ANS (50 μM) did not change during the refolding reactions in different concentrations of GdnHCl.

Refolding Kinetics Monitored by Enzymatic Activity. MSG catalyzes the following reaction:

$$\text{acetyl coenzyme A} + \text{glyoxylate} \rightarrow \text{malate} + \text{CoASH}$$

Each assay mixture contained 100 μL of 1 M Tris buffer (pH 7.9), 10 μL of 1 M $MgCl_2$, 10 μL of 0.1 M glyoxylate, 40 μL of 10 mM acetyl coenzyme A, and 12 μg of protein in a total volume of 900 μL. The reaction was initiated by adding 750 μL of 0.25 μM refolded protein solution into the assay mixture at indicated time intervals. After 5 min, reaction was stopped by adding 2 mL of 6 M GdnHCl, and color was developed by addition of 100 μL of 10 mM DTNB. Enzymatic activity was measured by recording absorbance at 412 nm in Beckman coulter DU800 spectrophotometer (USA). All assays were carried out at 25 °C in the refolding buffer at pH 7.9 and in triplicates. Percentage MSG activity recovered upon refolding, taking activity of native protein as 100%, was plotted against time of refolding.

Size Exclusion Chromatography. Size exclusion chromatography was carried out in a HPLC system (Waters, USA) using $TSKgelG3000SW_{XL}$ (Tosoh, Japan) prepacked column with molecular weight cut off 10−500 kDa. Refolded protein solutions, after 300 s and 1 h of refolding, were injected into the column using 20 μL loop. Refolding buffer (20 mM Tris, 300 mM NaCl, 10 mM $MgSO_4$, 1 mM TCEP HCl, 10% glycerol, pH 7.9) was used as the mobile phase with 1 mL/min flow rate. Native MSG in refolding buffer was run as control. Final protein concentration used was 1 μM.

Dynamic Light Scattering. Dynamic laser light scattering has been employed in order to detect the formation of soluble aggregates during the refolding process of MSG. We used a Malvern Zeta sizer APS system from Malvern instruments (UK) with 384-well plate (NII, New Delhi, India). Samples with protein concentration of 1 μM were filtered through 0.2 μm Millipore filters before DLS measurements. DLS scans were taken for MSG refolded for different time intervals. Data of 13−26 acquisitions were averaged and analyzed with the software provided with the instrument. Using this software, autocorrelation functions were calculated, and a regularization fit was performed in order to obtain the size (hydrodynamic diameter) of protein. Depending on the number of peaks (each peak represents a distinct population of particles with a defined size), we were able to determine the state of aggregation of protein.

Data Analysis. *Kinetic Refolding Traces.* Tryptophan monitored kinetic refolding traces for 0.1−0.3 M GdnHCl were analyzed using single exponential equation, whereas those for 0.4−0.5 M GdnHCl were analyzed using double exponential equation. ANS monitored refolding kinetic traces for all GdnHCl concentrations were fitted to double exponential equation. Each kinetic trace at each GdnHCl concentration was individually fitted to obtain the observed rate constant (k) and relative amplitude (α) of each of the observed kinetic phases of refolding.

Chevron Plot. Downward curvature in the observed rate constants of refolding kinetics was fitted to following equation

$$k_{obs} = k_f \exp(-m[D] - n[D^2]) \quad (1)$$

where k_{obs} is the observed rate constant at each GdnHCl concentration, k_f is the rate constant of folding in water, m is the surface area change of the protein from unfolded state (U) to transition state (TS), and n is the parameter describing curvature of refolding chevron arm. [D] is the denaturant concentration.

RESULTS

Unfolding Condition Is Crucial in Determining Final Refolding Yield of MSG. In vitro refolding of large multidomain proteins is known to be difficult without the use of molecular chaperones.[21] However, the detailed mechanistic study of protein folding becomes much simpler if the process is spontaneous, because no other molecular participations are to be considered. In the case of MSG, several of the commonly used methods of refolding, such as fast dilution, pulsed dilution, or slowly dialyzing out denaturant as well as partial denaturation by acid followed by refolding, all failed to provide substantial refolding yields. In contrast, unfolding of MSG in 3 M GdnHCl followed by rapid dilution in the refolding buffer as described in Materials and Methods provided 100% refolding yield for concentrations below 0.5 μM (i.e.,40 μg/mL). Being large and multidomain in nature, aggregation is probably a major hindrance in the refolding of this protein, so the refolding conditions were optimized in order to minimize protein losses due to aggregation. The major finding from the in vitro refolding studies of MSG was that, apart from the conditions for the mentioned process, conditions used for unfolding of the protein were found to be equally important in order to achieve complete refolding of the same. Figure 1a shows percentage of refolded MSG recovered upon refolding from different unfolded states. We observed that keeping the refolding conditions the same and varying only the unfolding conditions, different refolding yields were obtained. The GdnHCl unfolded state is best suited for MSG refolding as compared to acid or urea unfolded state (Figure 1a). Refolding yield is 100% when MSG is denatured in refolding buffer containing GdnHCl anywhere from 3 to 6 M. However, when unfolding was carried out in refolding buffer containing 8 M urea, refolding yield decreased to 45% (Figure 1a). The acid unfolded state also showed quite poor refolding yield (Figure 1a). This is most likely because of varying unfolded states present in different unfolding conditions.

Unfolding of MSG was monitored by measuring the loss of its enzymatic activity and changes in the tertiary and secondary structural elements. Loss of tertiary structure of the protein was monitored by tryptophan fluorescence which shows that MSG is only partially unfolded in acidic condition at pH 2 (red curve, Figure 1b). Maximum change in the intrinsic fluorescence intensity upon unfolding of the protein, at 340 nm (λ_{max} of native MSG), occurs in 3 M GdnHCl (pink curve). Tryptophan residues are maximum exposed to a polar environment in 3 M GdnHCl as is evident by its λ_{max} of tryptophan fluorescence emission (λ_{max} is 363 nm in the case of 3 M GdnHCl as compared to 351 and 341 nm, in 8 M urea and in acid, respectively). In 3 M GdnHCl, the enzymatic activity of MSG is completely lost after 2 h of denaturation. However, during denaturation in acid (pH 2) and with 8 M urea, 30% and 20% of activity still remains for the protein solution after the same time of incubation. Figure 1c shows the presence of some residual secondary structure in 8 M urea and in the acidic condition (pH 2), in contrast to the 3 M GdnHCl incubated solution, where the secondary structure is completely lost.

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

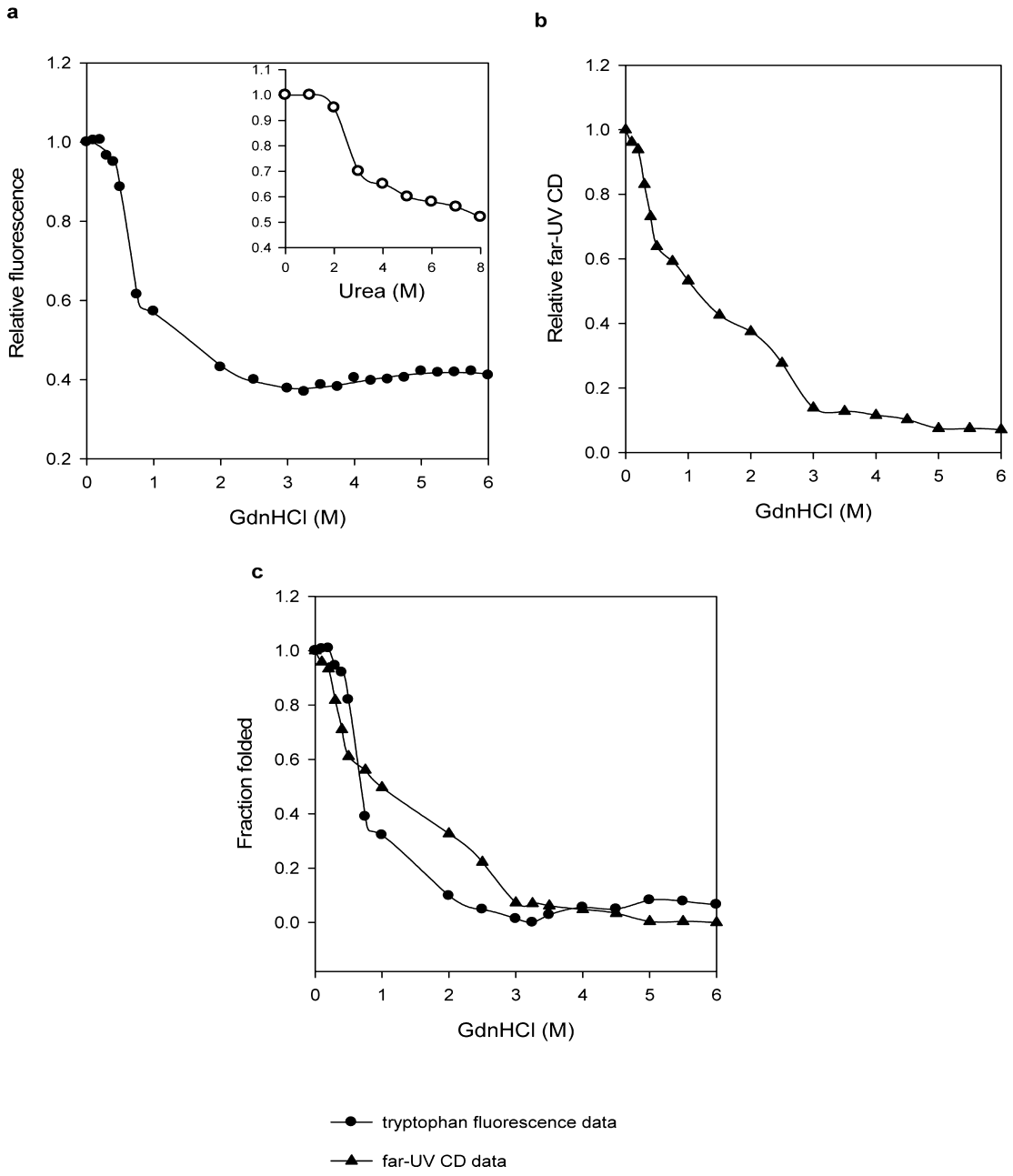


Figure 2. Equilibrium unfolding of MSG. Equilibrium unfolding of MSG was monitored by tryptophan fluorescence and far-UV CD. MSG was unfolded in refolding buffer containing different GdnHCl concentrations ranging from 0 to 6 M. Final concentration of MSG was 0.25 μM during the experiment. Solid lines were drawn by inspection only. (a) GdnHCl-induced equilibrium unfolding curve determined by monitoring fluorescence emission at 340 nm (the inset shows the equilibrium urea induced unfolding curve); (b) GdnHCl induced equilibrium unfolding curve determined by ellipticity at 222 nm. (c) Fraction folded versus GdnHCl concentration calculated from relative tryptophan fluorescence and relative far-UV CD data respectively; (●) fluorescence; (▲) far-UV CD.

Considering the fact that ANS can still bind to some extent to the urea and acid unfolded states of MSG, it may be speculated that the protein is not fully unfolded in both of these denaturants (Figure 1d). Whether MSG is completely unfolded (exists as random coil) in 3 M GdnHCl cannot be judged based only on these probes, but in the light of above findings it can be derived that MSG at least exists in a maximum unfolded state in 3 M GdnHCl.

The findings of poor refolding yields of MSG from both acid and urea denatured states (Figure 1a) suggest that partially unfolded forms of MSG are not good precursor for refolding, in contrast to another large protein, the maltose binding protein (MBP) which was refolded efficiently from partially denatured state.[22] Even from the 3 M GdnHCl incubated state of MSG, where the protein was found to be maximally unfolded, complete refolding was not accomplished when the unfolding buffer used was not the same as the refolding buffer (Figure 1a). Since the refolding buffer contained reducing agent (i.e., TCEP-HCl, Tris (2-carboxyethyl) phosphine hydrochloride) and glycerol, this implies that to achieve complete reversibility of the unfolding transition, it was important to minimize intermolecular associations between MSG molecules during unfolding as well as refolding. MSG contains six cysteine residues, out of which cys617 is in the flexible active site loop of

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

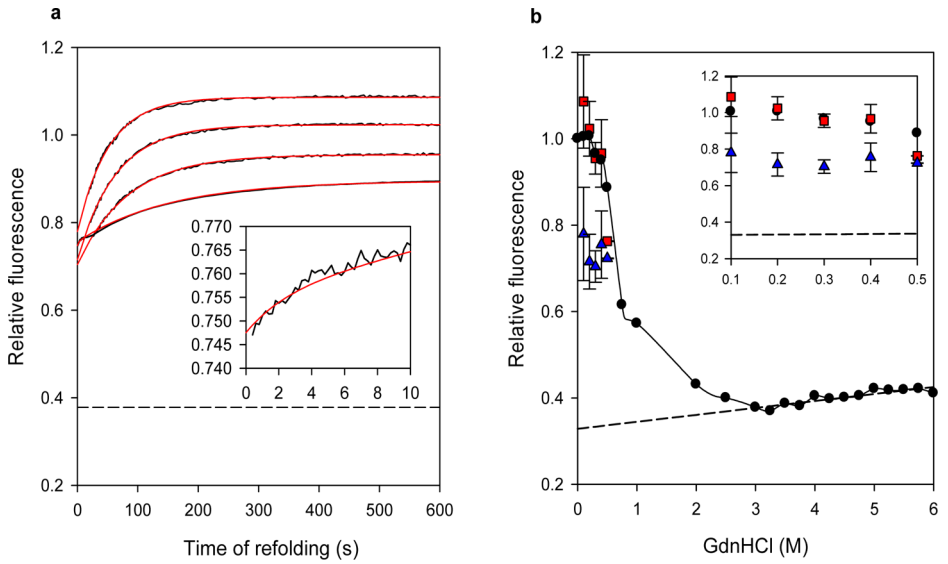


Figure 3. Effect of GdnHCl concentration on the refolding kinetics of MSG. Refolding kinetics of MSG was monitored by a change in intrinsic tryptophan fluorescence at 340 nm at pH 7.9, 25 °C. Protein was excited at 295 nm. (a) Refolding kinetic traces (continuous black line) when unfolded protein in 3 M GdnHCl was refolded by diluting GdnHCl in the refolding buffer to 0.1, 0.2, 0.3, and 0.4 M GdnHCl (top to bottom). The continuous red lines are the least-squares fits of the data to single exponential equation for 0.1, 0.2, and 0.3 M GdnHCl and double exponential equation to 0.4, 0.45, and 0.5 M GdnHCl. The inset shows the initial part of the kinetic trace of 0.4 M GdnHCl. (b) Kinetic versus equilibrium amplitudes of refolding. ●, equilibrium unfolding curve; red squares, $t = \infty$ points of the refolding kinetic traces; blue triangles, $t = 0$ points of the refolding kinetic traces. The black continuous line through the equilibrium unfolding data has been drawn by inspection only. The inset shows the data for the lowest GdnHCl concentrations. In both panels, the broken line is an extrapolation of the unfolded protein baseline from equilibrium unfolding studies. The error bars, wherever shown, represent the spreads of measurements from at least three separate experiments. In all the experiments, the final protein concentration used was 0.25 μM. Dead time for manual mixing was 10 s while for stopped flow mixing was 6 ms.

the C-terminal domain, and if it forms an intermolecular disulfide bond during refolding, the protein will no longer remain active.[13] Hence, the absence of TCEP-HCl during refolding most likely would not allow the formation of an active site in refolded MSG. In practice, when TCEP-HCl was absent in the refolding buffer, almost no recovery of functional refolded MSG was achieved through refolding. Reactivation and aggregation compete with each other during refolding; their relative proportions depend on both denaturing and renaturing conditions.[23] In particular, denatured states of larger multidomain proteins quite often show up as more expanded structures with a strong tendency to form aggregates.[24−26] Thus, for large and multidomain proteins, unfolding conditions become crucial in determining the fate of refolding process.

Equilibrium Unfolding of MSG Is Not a Two-State Process. Figure 2a shows the GdnHCl-induced equilibrium unfolding curve for MSG at pH 7.9, 25 °C monitored by intrinsic fluorescence spectroscopy. Solutions of MSG were excited at 295 nm, and tryptophan fluorescence emission was monitored at 340 nm, the wavelength at which native and unfolded states differ most in their intrinsic fluorescence. The equilibrium unfolding curve clearly displays two transitions with two separate m and C_m values. This signifies the presence of at least one stable intermediate at equilibrium. The protein is completely unfolded at 3 M GdnHCl, after which no further change in the tryptophan fluorescence of unfolded protein was observed. The inset of Figure 2a shows the urea-induced equilibrium unfolding transition curve of MSG. The protein is not completely denatured even at the highest concentration of urea (8 M) used in this experiment. Figure 2b shows that far-UV CD monitored equilibrium unfolding curve also does not show a single transition. In Figure 2c, plots of fraction of folded protein, calculated from the data in Figure 2a,b, versus GdnHCl concentration have been presented, and there was no overlap between the curves.

MSG Refolding Consists of a Burst Phase. Refolding of MSG was performed both by manual and stopped flow mixing by denaturant dilution in the refolding buffer such that final concentration of protein was always less than 0.5 μM. Refolding occurs in two kinetic phases, one fast and one slow (Figure 3a). The slow phase is found at all final GdnHCl concentrations in the refolding buffer, i.e., 0.1−0.5 M, while the fast phase was observed only for GdnHCl concentrations from 0.4 to 0.5 M (Figure 3a). The fast phase could be observed only in stopped flow mixing. The $t = 0$ points of kinetic refolding traces do not fall on the extrapolated unfolding baseline even in stopped flow mixing which had a dead time of 6 ms (Figure 3b). Surprisingly, $t = 0$ points were found to be same for both manual and stopped flow mixing. The total amplitude of fluorescence change for both kinetic phases (Figure 4c,d) is much less than the amplitude observed in equilibrium unfolding study (Figure 3b) indicating that the observed fast and slow kinetic phases do not account for the entire refolding reaction and there is a burst phase. The burst phase intermediate (C) displays 60−70% of native tryptophan fluorescence (Figure 4e). Notably, the relative amplitude of burst phase is weakly dependent on GdnHCl concentration, an indication of the stable nature of C (Figure 4e).

Quite remarkably, for slow phase of refolding, the dependence on GdnHCl concentration of the logarithm of the observed rate constant shows downward curvature (Figure 4a). It should be mentioned that refolding rate constants obtained from both manual and stopped flow mixing for the slow phase were found to be same for all the GdnHCl concentrations studied, ruling out any possibility that the downward curvature is an experimental artifact. The downward curvature is

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

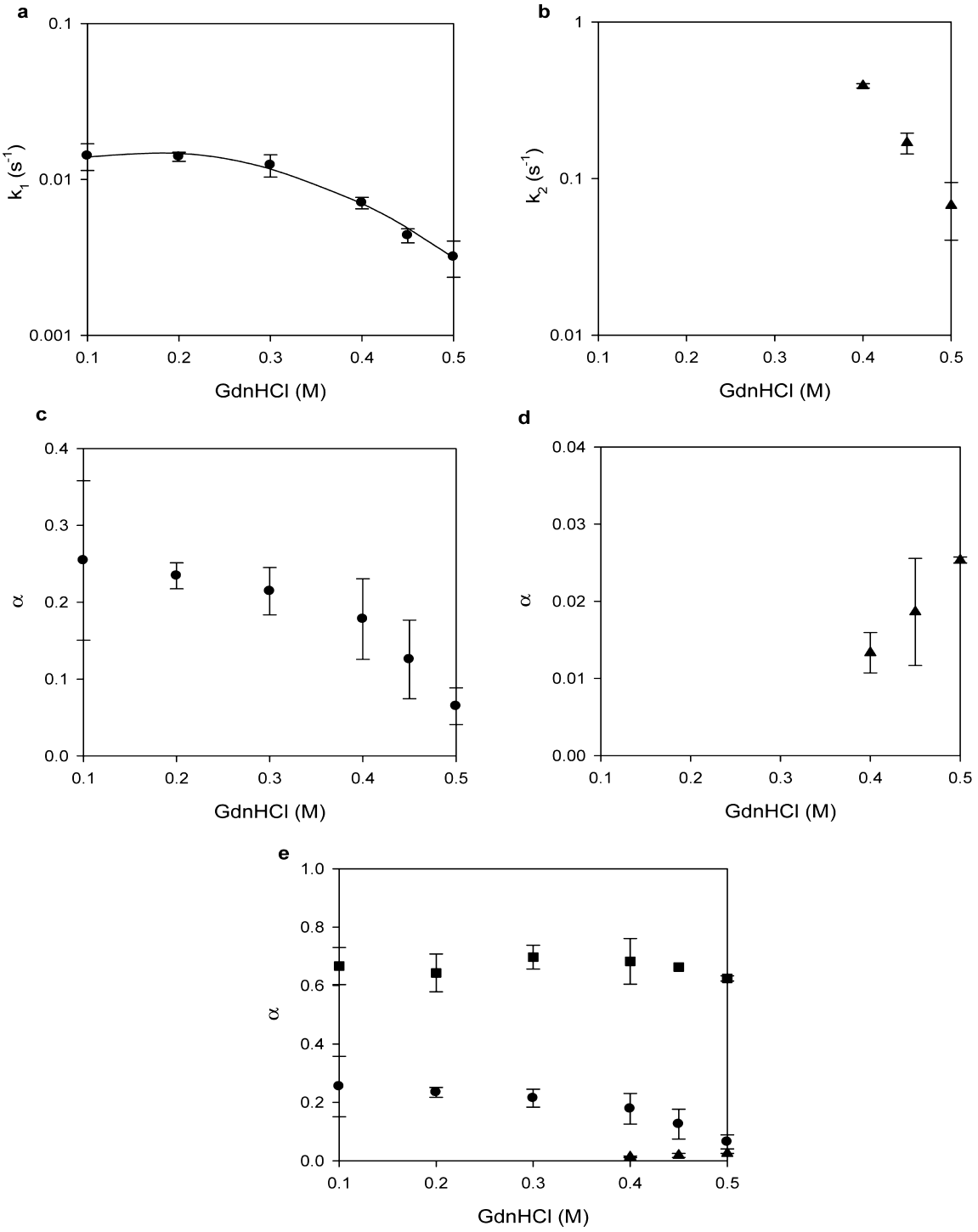


Figure 4. Dependence of the observed rate constants and relative amplitudes on GdnHCl concentration. Denaturant-dependent refolding kinetic studies were performed at 0.25 μM MSG concentration and were monitored by intrinsic tryptophan fluorescence change at 340 nm. Refolding rate constants and relative amplitudes were obtained from single exponential fits of kinetic traces for 0.1−0.3 M GdnHCl and from double exponential fits for 0.4−0.5 M GdnHCl respectively. (a, b) The observed rate constants for the slow (●) and fast phases (▲) of refolding respectively. (c, d) The relative amplitudes of the slow and fast phases of refolding respectively. (e) The relative amplitude of slow (●) and fast phases (▲) of refolding compared with the relative amplitude of burst phase (■). The black continuous line in (a) was drawn using eq 1. The fit yielded values of k_f = 0.01 s^{-1} and m = 4.78 M^{-1}. Relative amplitudes in (c−e) were plotted using linearly extrapolated unfolding baseline shown as a dashed line in Figure 3b. In all panels, the error bars represent the spreads of measurements from at least three separate experiments.

indicative of kinetic intermediate (I) getting formed during refolding. Figure 4c,d shows relative amplitudes of slow and fast phases of refolding. The relative amplitude of slow phase decreases with an increase in GdnHCl concentration from 0.1 to 0.5 M while that of fast phase decreases with decrease in GdnHCl concentration from 0.5 to 0.4 M. Since the relative amplitude of fast phase is very small and decreases with decrease in GdnHCl concentration, the phase was not observed at residual concentrations of GdnHCl < 0.4 M.

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

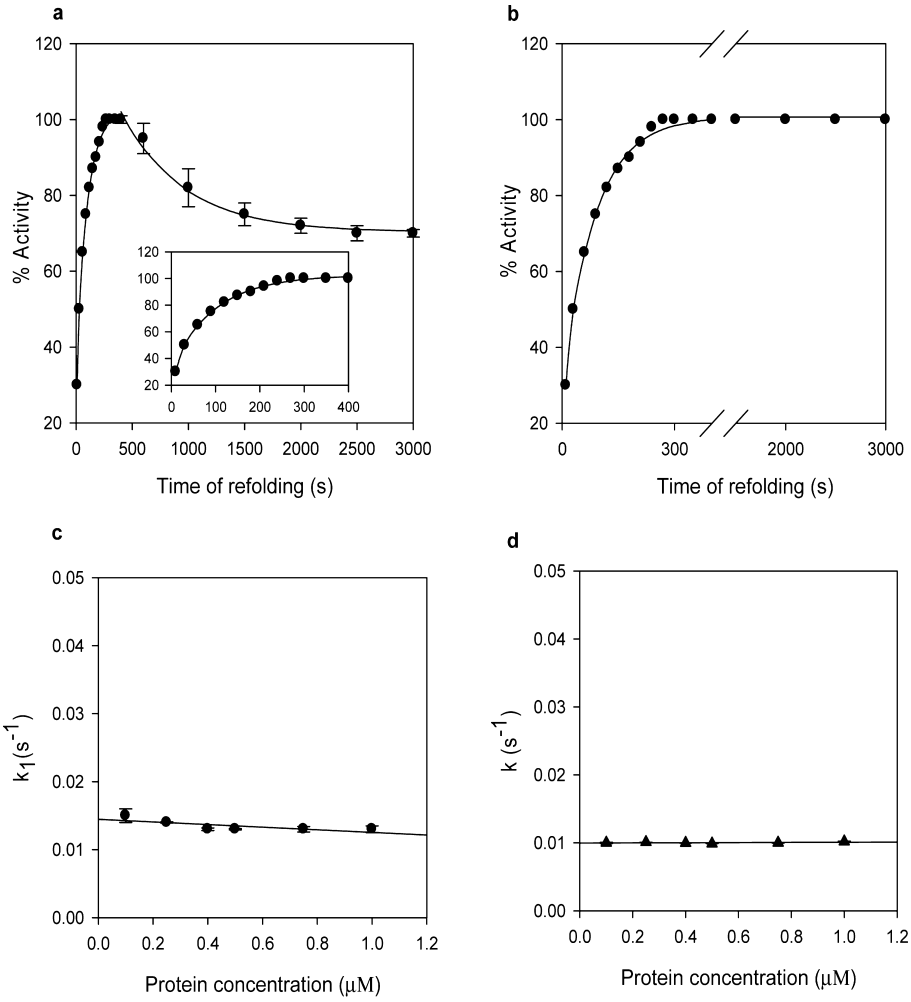


Figure 5. Dependence of refolding kinetics of MSG on protein concentration. Refolding of MSG was carried out at four different concentrations of protein (0.1, 0.25, 0.4, and 1 μM) and monitored by enzymatic activity as described in Materials and Methods. Final GdnHCl concentration in refolding buffer was kept at 0.1 M. (a) Representative refolding kinetic trace of 1 μM MSG. Indicated time points correspond to the time after which aliquots from the refolding mixture were added to the activity assay mixture. (b) Representative refolding kinetic trace of 0.25 μM protein. Black continuous line is the single exponential fit to data points. (c, d) Dependence of observed refolding rate constants obtained from tryptophan fluorescence (●) and enzymatic activity (▲) data on protein concentration at pH 7.9, 25 °C. Rate constants shown here correspond to the slow phase of refolding monitored by tryptophan fluorescence and enzymatic activity respectively. In all panels, the error bars represent the spreads of measurements from at least three separate experiments.

Nonnative State of MSG Has Same Functionality As Native State. Refolding of MSG was performed at different concentrations, and the process was monitored by measurement of tryptophan fluorescence and enzymatic activity. GdnHCl was diluted to 0.1 M where we observe only a slow phase of refolding in tryptophan fluorescence monitored refolding kinetics. While monitoring the time course of refolding using different concentrations of denatured MSG, it was observed that for concentrations above 0.5 μM (1 μM here) enzymatic activity decreased by 20−30% (Figure 5a) with a rate of 0.0018 ± 0.0003 s^{-1}. This decrease was found to be concentration dependent. For all concentrations tested below 0.5 μM (0.1, 0.25, 0.4), no decrease in refolding signals was observed with time (Figure 5b, only 0.25 μM trace is shown). Since the recovery of enzymatic activity at the initial stage of refolding was the same irrespective of the concentration of protein used (Figure 5a,b), the question arises why this decrease was happening. It could not be because of unfolding or misfolding of the folded protein, as in that case the decrease should occur at all protein concentrations used. To test whether the subsequent decrease of the initially regained enzymatic activity was due to aggregation, the refolding process of MSG was monitored through dynamic light scattering (Figure 6). Figure 6a shows the size distribution by mass of native protein solution where native MSG comprises 99.8% by mass of the solution. The plot clearly shows that the native protein solution does not contain any aggregates as only a single peak is obtained. The mass % is calculated from area under the curve and is a measure of homogeneity of the sample. Hence, the native MSG solution used in this experiment is homogeneous. However, two molecular distributions were obtained after 1 h of refolding when the final protein concentration in the refolding buffer was 1 μM (Figure 6b). Besides a nativelike distribution, there was another distribution centered at 45 nm diameter with a broad width. Mass percentage of aggregates during refolding increased with a rate of 0.0018 s^{-1} (Figure 6c) which was same as the rate of loss of enzymatic activity, indicating that these aggregates are directly responsible for the decrease observed in activity of MSG over time, during refolding at higher concentrations. The

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

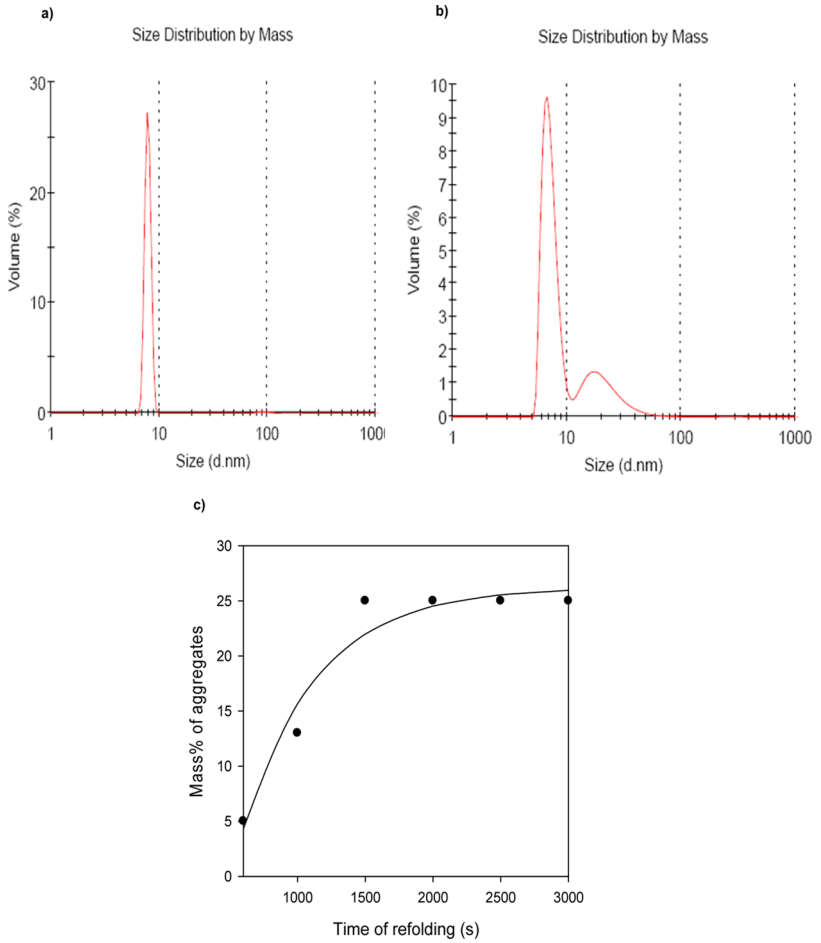


Figure 6. Monitoring soluble aggregates formation during refolding by dynamic light scattering. (a) Size distribution of native protein solution by mass where 99.8% of the total sample mass comprises of protein in the monomeric state, i.e., without any aggregation. In this state, the distribution is dominated by a monomer peak with mean diameter 7.9 nm and a rather small peak width. Such a distribution is characteristic of native protein. Mass % is given by area under the peak. (b) Size distribution by mass of refolded protein solution after 1 h of refolding. Two distributions are obtained, a nativelike distribution with a mean size of 7.9 nm which comprises 75% of the total sample mass and a broad distribution with a mean size of 45 nm. This distribution represents 25% of total sample mass and indicates formation of small aggregates during the refolding process. (c) Plot of mass % of aggregates in the solution formed during refolding vs time of refolding. Dead time of the instrument was 10 min, so earlier time points could not be taken.

justification for the fact that MSG is not aggregating in its slow refolding phase, leading to regain of nativelike tryptophan fluorescence (Figure 3a) and enzymatic activity (Figure 5a,b), comes from the observation that the rate of slow phase does not depend on protein concentration (Figure 5c,d).

After 1500 s, mass percentage of aggregates in the solution was 25% (Figure 6c). These results imply that recovery of entire tryptophan fluorescence and enzymatic activity does not correspond to the formation of native MSG. Had native MSG been formed, aggregation should not have taken place. Figure 7 shows size exclusion chromatogram of MSG at different times of refolding. Retention time of native protein was compared with the protein refolded for 300 s and 1 h. It was found that retention time of MSG refolded for 300 s (blue line) was slightly less than the protein refolded for 1 h (red line), which exhibited same the retention time as that of the native MSG (black line). Thus, conformation of MSG present at 300 s of refolding is not same as native protein despite it having nativelike enzymatic activity. The slight difference in the absorption between MSG refolded for 300 s vs 1 h (blue and red peaks in Figure 7) may be attributed to the fact that at the early stage of refolding there were more protein molecules

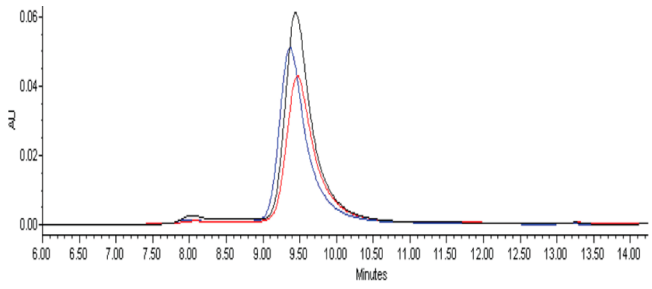


Figure 7. Size exclusion chromatograms. Gel filtration profiles of native and refolded MSG monitored by absorbance measurement at 280 nm are shown. Black line represents elution profile of 1 μM native MSG with a retention time of 9.44 min. Blue line represents elution profile of 1 μM MSG after 5 min of refolding with a retention time of 9.34 min. Red line represents elution profile of 1 μM MSG after 1 h of refolding with a retention time of 9.44 min. Its absorbance at 280 nm is less than that of native protein due to conglomeration of protein molecules during refolding process to form small aggregates. The peak corresponding to the soluble oligomers cannot be seen perhaps due to their very less absorption.

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

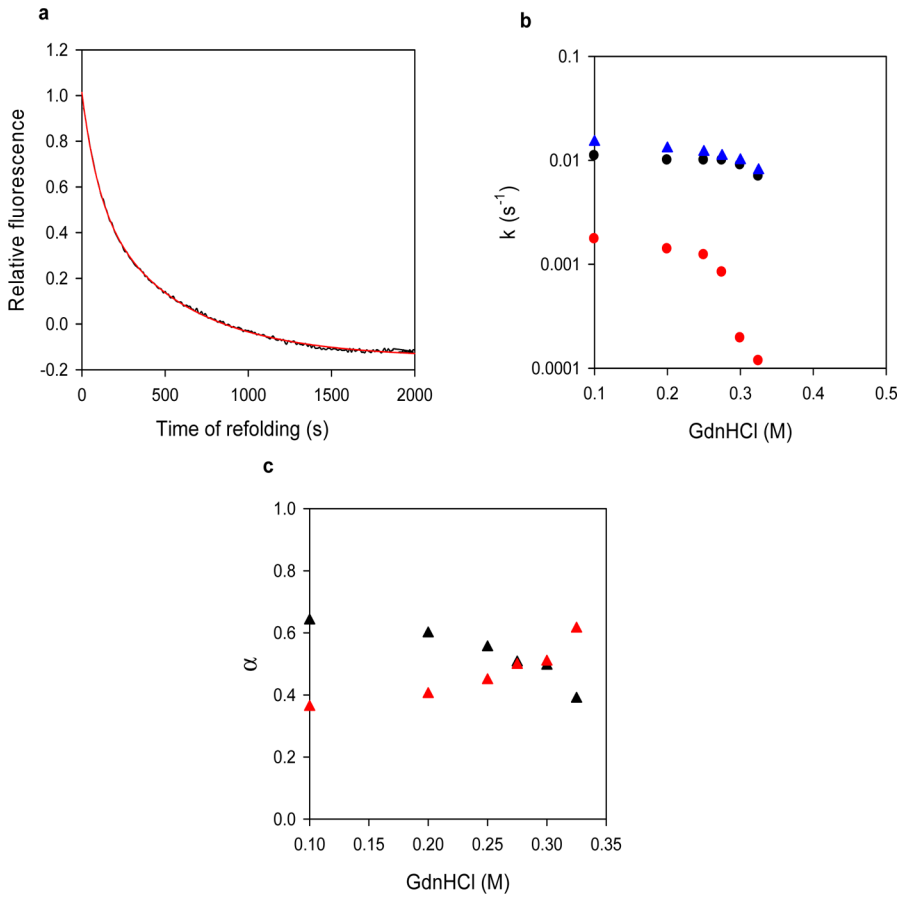


Figure 8. ANS fluorescence monitored kinetics of MSG in refolding buffer at 25 °C. MSG was unfolded to equilibrium in 3 M GdnHCl and was refolded in 0.1−0.325 M GdnHCl in the presence of 50 μM ANS. Final protein concentration in the refolding buffer was 0.25 μM. Protein was excited at 350 nm, and kinetics of folding was followed by monitoring total ANS fluorescence at 450 nm. Refolding kinetic traces obtained at all GdnHCl concentrations used were fitted to double exponential equation. (a) Representative refolding kinetic trace (black continuous line) when unfolded protein was diluted in refolding buffer containing 0.1 M GdnHCl and 50 μM ANS. The continuous red line is least-squares fits of the data to two-exponential equation. (b) Dependence of observed rate constants of slow (●) and very slow (red circles) phases on GdnHCl concentration; (blue triangles) depicts rate constant of tryptophan fluorescence monitored slow phase of refolding. (c) Dependence of relative amplitudes of slow (▲) and very slow (red triangles) phases on GdnHCl concentration. The fluorescence amplitudes are shown relative to the total change in ANS fluorescence observed (i.e., the sum of the fluorescence amplitudes of both the slow and very slow phases) when the refolding reaction was initiated by jumping from 3 M to 0.1−0.325 M GdnHCl.

having enzymatic activity and other physicochemical properties which eluted at 9.34 min. However, during 1 h incubation time some lower order aggregates were formed (~25%), which should not elute along with the monomeric form of the proteins, perhaps nativelike molecules eluted at 9.44 min. Aggregates corresponding to 25% of the protein molecule could not be traced, perhaps due to very low absorption.

Formation of Native MSG Occurs in Very Slow Phase of Refolding. To validate that formation of native topology of MSG is a very slow process, refolding kinetics was monitored by ANS binding method. The hydrophobic molecule ANS binds to hydrophobic clusters on proteins that are hydrated and therefore accessible,[27] and such binding is accompanied by a large change in fluorescence emission of ANS-bound protein. Since, ANS does not bind to native or unfolded MSG (data not shown), it was used as a probe to investigate native MSG formation along the refolding time scale. Figure 8a shows refolding kinetic trace of MSG in 0.1 M GdnHCl monitored by measuring ANS-bound fluorescence. The decrease in fluorescence intensity of ANS−protein complex that occurs during refolding takes place in two kinetic phases, slow and very slow, and hence can be easily monitored. Enhancement in fluorescence intensity of protein−ANS complex, which originates from ANS binding to hydrophobic segments on the protein molecules, could not be monitored in stopped-flow system since it occurred within 6 ms dead time of the instrument. Further, the decrease in protein−ANS conjugate fluorescence intensity that occurs due to folding of protein was monitored by manual mixing only as it was observed to be a very slow process. ANS fluorescence monitored refolding kinetics was performed for 0.1−0.325 M final GdnHCl concentrations as after that ANS fluorescence was not found to decay. Figure 8b,c shows the dependence of the rate constants and the relative amplitudes of the two phases on the residual concentration of GdnHCl during refolding. As the refolding rates of slow phase, monitored by ANS-bound fluorescence measurement, coincide with the rates of refolding monitored by measurement of tryptophan fluorescence (Figure 8b), this indicates that both the probes are monitoring the same kinetic processes. These rates were found to be independent of GdnHCl concentration, with a slight decrease observed at 0.3 and 0.325 M GdnHCl. However, a very slow phase was also observed, when the refolding kinetics of MSG was monitored by ANS-bound fluorescence measurement, which was not

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

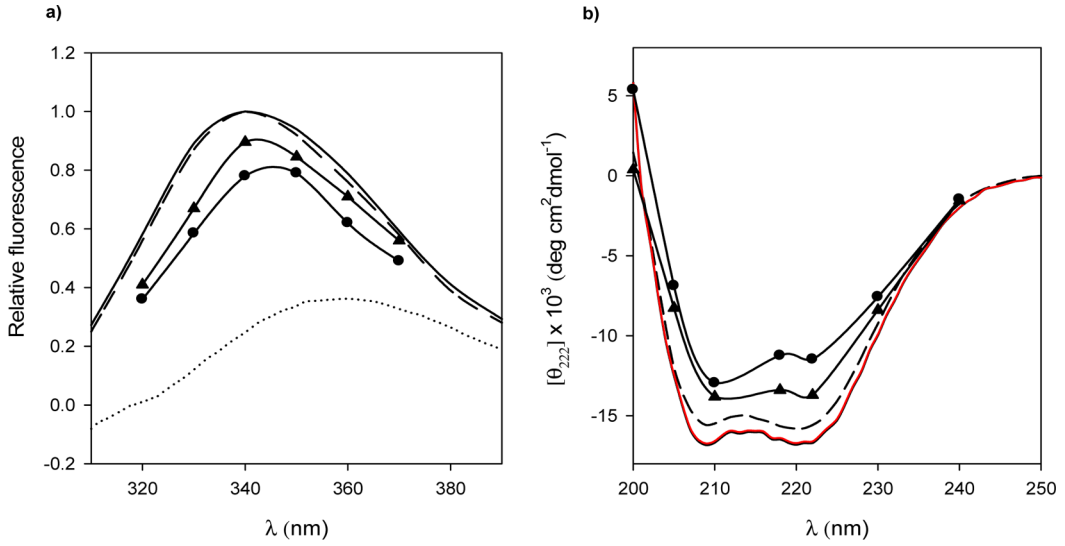


Figure 9. Tryptophan fluorescence and far UV-CD spectra of MSG. MSG was unfolded in 3 M GdnHCl and diluted in refolding buffer such that final protein and GdnHCl concentration were 0.25 μM and 0.1 M, respectively. All spectra were recorded in refolding buffer of MSG with necessary background subtractions. (a) (——) Native state of MSG; (-----) MSG after 300 s of refolding i.e., I_N; (●) and (▲) are the t = 0 and t = 30 s values obtained from refolding kinetic curves of MSG at different wavelengths and represent C and I intermediates, respectively; (·····) unfolded state in 3 M GdnHCl. (b) Far UV-CD spectra obtained at different times of refolding to monitor secondary structure formation in transient intermediates. (●) and (▲) represent C and I, respectively; (-----) represents I_N; (red line) MSG after 1 h of refolding; (——) native MSG. Spectra were recorded on a JASCO J-815 CD polarimeter using 2 mm path length cuvette.

present during monitoring the process by measuring intrinsic tryptophan fluorescence. The rates of this phase do not depend on the concentration of GdnHCl until 0.25 M after which they show a dependence on GdnHCl concentration with rates becoming extremely low at 0.3 and 0.325 M GdnHCl. On the basis of these observations, it may therefore be proposed that the slow phase observed during monitoring the refolding kinetics of MSG through measurement of tryptophan fluorescence (Figure 3a) represents the formation of functional intermediate (I_N) which converts to native MSG in the very slow phase that is silent to the change in tryptophan fluorescence emission.

Characterization of Kinetic Intermediates in the Refolding of MSG. In order to get a more detailed insight into nonnative nature of kinetic intermediates formed during the refolding pathway of MSG, they were further characterized by following their far-UV CD and tryptophan fluorescence spectra. Near-UV CD spectroscopy could not be performed because of concentration constraints of MSG during refolding. Maximum concentration of MSG that could be used during refolding, without any possibility of aggregation, was 0.5 μM which was considerably low to generate good near-UV CD spectra.

Figure 9a shows tryptophan fluorescence spectra of kinetic intermediates formed during refolding of MSG. Tryptophan fluorescence spectrum of burst phase intermediate (C) and transient intermediate (I) in Figure 9a was generated by measuring the refolding kinetics of MSG in 0.1 M GdnHCl as described in Materials and Methods at various wavelengths between 320 to 370 nm in Perkin Elmer LS55 fluorimeter. All the kinetic curves at these wavelengths fitted well in a single exponential equation. The t = 0 and t = 30 s fluorescence values thus obtained from the kinetic curves plotted as a function of wavelength represent fluorescence spectrum of C and I respectively. Since rate of formation of I is not known, t = 30 s values were chosen arbitrarily as fluorescence at 30 s was midway of the overall fluorescence change. As the nativelike intermediate (I_N) is quite stable whose fluorescence does not change with time (Figure 3a), its spectrum was generated by scanning its tryptophan fluorescence over the whole wavelength range in the same way as native protein. Although the number of data points for the spectrum of C and I is limited, it provides a good comparison with native protein conformation. It appears that most of the tertiary structural elements are formed in C and I, and as far as tryptophan fluorescence is concerned I_N is structurally similar to native protein. The conformational transitions from C to I to I_N are apparent in their λ_{max} emission which shifts from around 345 to 342 to 340 nm (Figure 9a). Figure 9b shows far UV-CD spectra of transient folding intermediates which give an estimation of secondary structural transitions taking place during refolding. Refolding was performed by manual mixing with a dead time of 10 s, and spectra of C and I were generated in the same way as fluorescence spectrum from CD kinetic curves at various wavelengths from 200 to 240 nm. The intermediates have reduced ellipticity and are not same as native. It appears that they have a different composition of secondary structural elements. The fractions of secondary structure in transient folding intermediates of MSG were estimated from the spectra in Figure 9b by the method of Yang et al.[28] and are shown in Table 1. It is evident that the burst phase intermediate has

Table 1. Secondary Structure Fractions of MSG Calculated from far UV-CD Spectra by Yang Reference Spectra

fraction secondary structure content	native (%)	burst phase intermediate (C) (%)	I (30 s) (%)	I_N (300 s) (%)	complete refolding (1 h) (%)
α-helix	33	37	32	32	33
β-sheet	27	8	12	20	27
turn	10	30	23	16	10
coil	30	25	33	32	30

dx.doi.org/10.1021/bi400328a | *Biochemistry* 2013, 52, 4517−4530

more of α-helix or β-turn and less of β-sheet. As refolding progresses, more coils and β-sheets are formed (I and I_N). Spectra of I_N were taken after 300 s of refolding and are quite similar to native MSG spectra. Finally, the CD spectrum of refolded protein became the same as native protein, which reports the completion of the refolding process of MSG protein (Figure 9b).

DISCUSSION

The refolding process of malate synthase G is characterized by the presence of a burst phase intermediate (Figure 3) with significant secondary and tertiary structural elements (Figure 9). A considerable blue shift in λ_{max} emission of tryptophan fluorescence was observed in the burst phase intermediate upon refolding (363 nm for unfolded protein to 345 nm for burst phase intermediate, Figure 9a), suggesting the collapse of the unfolded MSG to C state leading to burial of most of the tryptophan residues in the latter. A fast pre-equilibrium can be assumed for the U ↔ C transition during early stage of MSG refolding. Although the C state consists of some non-native secondary structure, significant native secondary structure is also present in it (Table 1). These observations are consistent with the existing folding model which reveals that both local and nonlocal interactions are dominant forces in early folding intermediates. Such intermediates having a compact collapsed structure, stabilized by hydrophobic interactions and local hydrogen bonds, are useful for guiding the folding process of large proteins, as it would become increasingly difficult to form specific nonlocal interactions unless the molecule assumes the compact conformation. Monitoring the refolding process of MSG through changes in far-UV CD spectra with time reveals that the formation of secondary structure of the protein starts as soon as denaturant is removed from the environment and continues until completion of refolding (Table 1). Predicting the fractional secondary structure of refolded protein from far-UV CD spectra by the reference method used here (Yang et al. 1986)[28] may not be very accurate because it requires extension of CD data to vacuum-UV region which was limited in the present case by the presence of 0.1 M GdnHCl in the refolding mixture in which the far-UV CD spectra of the protein could be recorded up to 200 nm only. Still a qualitative assessment of the order of secondary structure formation in the folding intermediates can be carried out; however, limitation caused by truncation of data at 200 nm should be kept in mind while judging the accuracy in the quantitative assessment of the calculated secondary structural elements from recorded ellipticity values.

Nativelike α-helices are formed at very early stage of MSG refolding as compared to β-sheets that are formed slowly and appears only in I_N to a considerable amount, which has similar content of α-helices and coils as native MSG (Table 1). The burst phase intermediate in the MSG refolding process is characterized by a small fraction of non-native α-helices and majority of non-native β-turns. Since an average 10% of β-sheet is formed in the early stages of MSG refolding, this suggests that 20% of β-sheet was in the form of non-native α-helices and turns in the early folding intermediates (C and I, Table 1). Formation of I_N from C involves rearrangement of non-native α-helices and turns to native β-sheet and formation of coils. Transition of I_N to native state involves formation of nativelike β-sheet (Table 1). As mentioned earlier from the experimental evidence that the conformation of I_N is not same as native MSG (Figures 7 and 8), these results suggest that final stages of refolding involve slow acquisition of both tertiary and secondary structural elements.

Rearrangements of C to I_N during MSG refolding occurs via the formation of another folding intermediate I, which manifests itself by causing a downward curvature in the refolding arm of chevron (Figure 4a). Such downward curvatures are observed on the refolding pathway of several proteins.[29−31] Since it has been observed that the rate of slow phase of MSG refolding, monitored by measuring the changes of tryptophan fluorescence, is independent of the protein concentration (Figure 5c), it is an indication that I be an on pathway intermediate. Whether the slow and fast phases observed during tryptophan fluorescence monitored refolding process (Figure 4a,b) belong to competing or parallel pathways could not be deduced from their plots of relative amplitudes as majority of the signal changes associated with the refolding process of MSG is lost in the burst phase (Figure 4e).

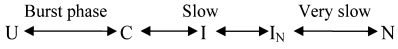
It is known from the earlier findings that the active site of MSG is located in the cleft between TIM barrel and C-terminal domain which are found to be crucial for its activity.[13] The observation that I_N has same catalytic activity as the native protein suggests that at least two domains constituting the active site have been correctly folded and docked with each other during slow phase of tryptophan fluorescence monitored refolding, thereby making the enzyme functional. Moreover, the domain-wise analysis of amino acid sequence of MSG suggests that 8 out of 12 tryptophan residues in MSG are present in the two adjacent domains forming the active site. When the structure of MSG was analyzed by PyMOL software, it was found that out of the four tryptophans present in the remaining two domains, not forming the active site, two of the tryptophans are present on the surface (data not shown). These results suggest that major contribution to tryptophan fluorescence change observed during refolding of MSG comes from the domains forming the active site and the rest two domains do not contribute substantially. Thus, recovery of native tryptophan fluorescence and enzymatic activity may indicate the presence of folded TIM barrel and C-terminal domain in the I_N, but whether the other two domains are folded or not in the intermediate state cannot be derived from these observations alone.

As MSG has 31 prolines, the very slow phase of native MSG formation from functional intermediate might just be a result of slow cis/trans isomerization of proline residues which are known to be responsible for slow conformational changes detected during protein folding.[32] However, the rates of the very slow phase are not entirely independent of GdnHCl concentration and show a downward curvature (Figure 8b). This evidence rules out the possibility that this slow conformational change simply results from, and is rate-limited by, proline isomerization events. The dependence of the rates of refolding on the concentration of GdnHCl is indicative of an additional folding event occurring during very slow phase of transition of I_N to N. This observation could be related to weak interdomain cooperativity in MSG, leading to better organization of the two domains forming the active site in I_N than the other two. Since transition of I_N to N comprises some β-sheet formation (Table 1), the other two domains (among which one is a β-sheet domain) might attain their final three-dimensional structure in the very slow phase of refolding. The fact that GdnHCl induced equilibrium unfolding of MSG is not a two state process, and the transition curves obtained for the unfolding process monitored by far-UV CD and tryptophan

fluorescence spectroscopy do not overlap (Figure 2), also suggests noncooperative transition between N to U states. The extremely slow conversion of I_N to N is consistent with the folding model of multidomain proteins where the fully native structure is only acquired when all the nativelike inter- and intradomain interactions[33,34] have been formed, making the process slower.

The results discussed so far are well accounted by the minimal kinetic mechanism depicted in Scheme 1.

Scheme 1



Detailed investigation of the cooperativity of structure formation and order of formation of the domains in MSG requires very sophisticated experimentation. More specific methods, sensitive to the interdomain interaction, such as putting fluorescent probes at strategic positions and monitoring FRET, or carrying out pulsed H-D exchange combined with NMR or mass spectrometry with various intermediate species, may provide the information on the refolding event in a residue-specific manner. Besides, small angle X-ray scattering (SAXS) can provide information about changes in the size and shape of folding intermediates during refolding,[35,36] which will be helpful in the present case where the native state is formed from compaction of the functional intermediate. However, these methodologies are worth being considered for further characterization of folding intermediates.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tkchaudhuri@bioschool.iitd.ac.in, tapanchaudhuri@hotmail.com; tel: 91-11-2659-1012; fax: +91-11-2659-7530.

Funding

V.D. is the recipient of CSIR, Govt. of India Senior Research Fellowship, and T.K.C. acknowledges the financial assistance from DBT, Government of India.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the generous gifts of pcs19MSG plasmid from Prof. Laura Baldoma, University of Barcelona, Spain. We are really grateful to Prof. Jayant B. Udgaonkar, National Centre for Biological Sciences, Bangalore (NCBS), and his Ph.D. student Mr. Nilesh Aghera for helping us to carry out fluorescence kinetics experiments on the SFM-400 (Biologic) stopped flow machine and Biologic MOS-450 optical system and for their invaluable assistance and advice. We thank Dr. Amulya Panda and the members of his laboratory in the National Institute of Immunology, New Delhi, for helping us with the dynamic light scattering instrument and providing valuable suggestions to interpret the light scattering data. The authors finally thank Prof. Arthur L. Horwich, Yale University, USA, and Mr. Jon Tally, Department of Biochemistry and Molecular Biology, Kansas University Medical Center, USA, for reading the manuscript carefully and providing valuable suggestions for improving the quality. Dr. Srayanta Mukherjee, Bioinformatics Division of Kansas University is acknowledged for his help in giving the beautiful shape to the graphical abstract.

ABBREVIATIONS

MSG, malate synthase G; ANS, 1-anilino-8-naphthalenesulfonate; GdnHCl, guanidine hydrochloride; TCEP HCl, Tris(2-carboxyethyl) phosphine hydrochloride; DTNB, 5,5′-dithiobis-(2-nitrobenzoic acid)

REFERENCES

(1) Jamin, M., and Baldwin, R. L. (1996) Refolding and unfolding kinetics of the equilibrium unfolding intermediate of apomyoglobin. *Nature 3*, 613−619.

(2) Chaudhuri, T. K., Arai, M., Terada, T. P., Ikura, T., and Kuwajima, K. (2000) Equilibriums and kinetic studies on folding of the authentic and recombinant forms of human α- lactalbumin by circular dichroism spectroscopy. *Biochemistry 39*, 15643−15651.

(3) Baryshnikova, E. N., Melnik, B. S., Finkelsteinx, A. V., Semisotnov, G. V., and Bychkova, V. E. (2005) Three-state protein folding: experimental determination of free energy profile. *Protein Sci. 14*, 2658−2667.

(4) Arai, M., and Kuwajima, K. (1996) Rapid formation of a molten globule intermediate in refolding of α lactalbumin. *Folding Des. 1*, 275−285.

(5) Matthews, B. W. (1993) Structural and genetic analysis of protein stability. *Annu. Rev. Biochem. 62*, 139−160.

(6) Schindler, T., Herder, M., Marahiel, M. A., and Schmid, F. X. (1995) Extremely rapid protein folding in the absence of intermediates. *Nat. Struct. Biol. 2*, 663−673.

(7) Teilum, K., Maki, K., Kragelund, B. B., Poulsen, F. M., and Roder, H. (2002) Early kinetic intermediate in the folding of Acyl-Coenzyme A binding protein detected by fluorescence labeling and ultra rapid mixing. *Proc. Natl. Acad. Sci. U. S. A. 99*, 9807−9812.

(8) Orengo, C. A., and Thornton, J. M. (2005) Protein families and their evolution − a structural perspective. *Annu. Rev. Biochem. 74*, 867−900.

(9) Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F. U. (1999) Identification of in vivo substrates of the chaperonin GroEL. *Nature 402*, 147−154.

(10) Frydman, J. (2001) Folding of newly translated proteins in vivo: The role of molecular chaperones. *Annu. Rev. Biochem. 70*, 603−647.

(11) Molina, I., Pellicer, M., Badia, J., Aguilar, J., and Baldoma, L. (1994) Molecular characterization of *Escherichia coli* malate synthase G. Differentiation with the malate synthase A isoenzyme. *Eur. J. Biochem. 224*, 541−548.

(12) Berg, J. M., Tymoczko, J. L., and Stryer, L. (2002) The citric acid cycle, in *Biochemistry*, 5th ed., p 484, WH Freeman & Co., New York.

(13) Howard, B. R., Endrizzi, J. A., and Remington, S. J. (2000) Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: mechanistic implications. *Biochemistry 39*, 3156−3168.

(14) Tugarinov, V., Choy, W. Y., Orekhov, V. Y., and Kay, L. E. (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc. Natl. Acad. Sci. U. S. A. 102*, 622−627.

(15) Privalov, P. L., and Khechinashvilli, N. N. (1974) A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol. 86*, 665−684.

(16) Privalov, P. L. (1982) Stability of proteins: proteins which do not present a single co-operative system. *Adv. Protein Chem. 35*, 1−104.

(17) Ramsay, G., and Freire, E. (1990) Linked thermal and solute perturbation analysis of co-operative domain interactions in proteins. Structural stability of diphtheria toxin. *Biochemistry 29*, 8677−8683.

(18) Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007) The folding and evolution of multi-domain proteins. *Nat. Rev. Mol. Cell Biol. 8*, 319−330.

(19) Privalov, P. L. (1996) Intermediate states in protein folding. *J. Mol. Biol. 258*, 707−725.

(20) Sudharshan, E., and Rao, A. G. (1999) Involvement of cysteine residues and domain interactions in the reversible unfolding of lipoxygenase-I. *J. Biol. Chem. 274*, 35351−35358.
(21) Lilie, H., Schwarz, E., and Rudolph, R. (1998) Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol. 9*, 497−501.
(22) Gardner, K. H., Zhang, X., Gehring, K., and Kay, L. E. (1998) Solution NMR studies of a 42 kDa *Escherichia coli* maltose binding protein/β-cyclodextrin complex: chemical shift assignments and analysis. *J. Am. Chem. Soc. 120*, 11738−11748.
(23) Zettlmeissl, G., Rudolph, R., and Jaenicke, R. (1979) Reconstitution of lactic dehydrogenase noncovalent aggregation vs reactivation: physical properties and kinetics of aggregation. *Biochemistry 18*, 5567−5571.
(24) Duy, C., and Fitter, J. (2006) How aggregation and conformational scrambling of unfolded states govern fluorescence emission spectra. *Biophys. J. 90*, 3704−3711.
(25) Tanford, C. (1968) Protein denaturation. *Adv. Protein Chem. 23*, 121−282.
(26) Buchner, J., and Kiefhaber, T. (2005) *Protein Folding Handbook*, Vol. *1*−*5*, Wiley-VCH Verlag, Weinheim, Germany.
(27) Stryer, L. (1965) The interaction of naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol. 13*, 482−495.
(28) Yang, J. T., Wu, C. S., and Martinez, H. M. (1986) Calculation of protein conformation from circular dichroism. *Methods Enzymol. 130*, 208−269.
(29) Shastry, 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.
(30) Park, S. H., Shastry, M. C., and Roder, H. (1999) Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat. Struct. Biol. 6*, 943−947.
(31) 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.
(32) Nall, B. T. (1994) Proline isomerisation as a rate-limiting step. *Mechanisms of Protein Folding* (Pain, R. H., Ed.) pp 80−103, IRL Press at Oxford University Press, Oxford, UK.
(33) Khan, F., Chuang, J. I., Gianni, S., and Fersht, A. R. (2003) The kinetic pathway of folding of barnase. *J. Mol. Biol. 333*, 169−186.
(34) Vendruscolo, M., Paci, E., Karplus, M., and Dobson, C. M. (2003) Structures and relative free energies of partially folded states of proteins. *Proc. Natl. Acad. Sci. U. S. A. 100*, 14817−14821.
(35) Fink, A. L. (1995) Compact intermediate states in protein folding. *Annu. Rev. Biophys. Biomol. Struct. 24*, 495−522.
(36) Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y., and Kuwajima, K. (1998) Kinetic refolding of β-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy. *J. Mol. Biol. 275*, 149−162.