SUPPORTING INFORMATION

Spontaneous refolding of the large multidomain protein malate synthase G proceeds through misfolding traps

Vipul Kumar[§], Tapan K. Chaudhuri[§]

[§]Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, New Delhi 110016, India.

Table of contents

Supporting Figure S1 to S8
Supporting Table S1 to S2

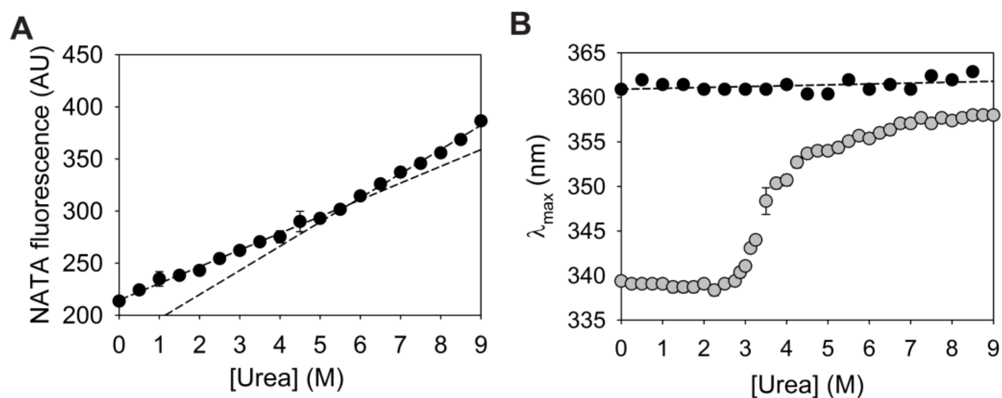

Figure S1. Determination of unfolding baseline for MSG. (A) A non-linear increase in fluorescence intensity of N-acetyl-L-tryptophanamide (NATA) at high urea concentrations, shows intrinsic nature of Trp residues. The observation can be mistaken for transition near 6.5 M urea. (B) The plot of wavelengths corresponding to maximum fluorescence emission against urea concentration for MSG (grey) represents a red shift in peak maxima till ~7 M urea. The phenomenon clearly establishes the incomplete unfolding of the protein in 3.5-6.5 M urea region. A similar plot of wavelengths for maximum fluorescence intensity has also been shown for NATA (black) to exclude the possibility of artifact due to solvent behaviour. The higher values as compared to unfolded protein can be correlated to highly exposed nature of NATA in the solution as compared to Trp residues in the proteins. The error bars represent standard deviation from three individual readings.

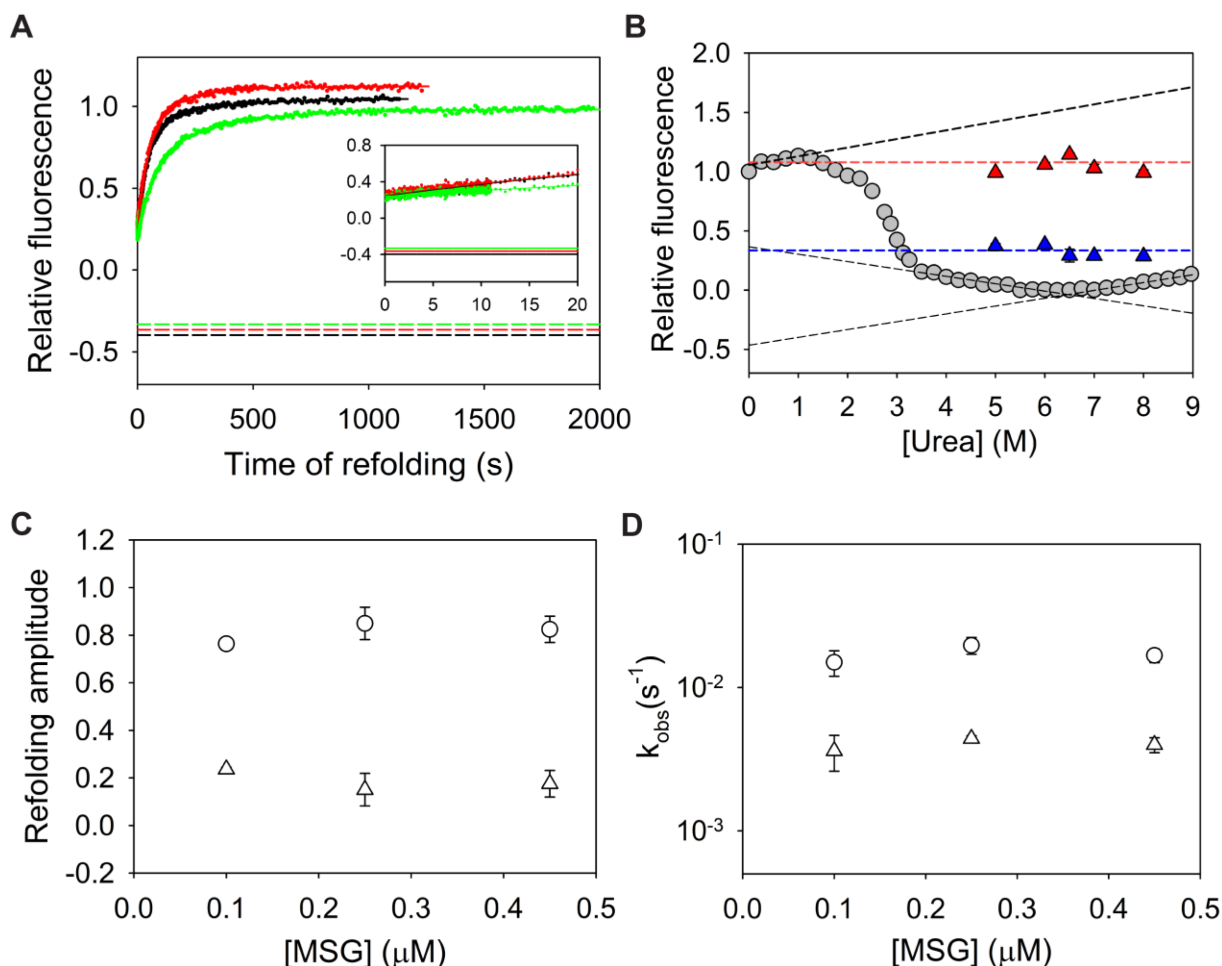

Figure S2. The refolding behavior of MSG via Trp fluorescence probe. (A) The representative refolding traces at 1 M (black), 1.5 M (red) and 2 M (green) urea with Trp fluorescence probe. Inset represents the absence of any fast refolding phase within manual mixing dead time (10 s) collected through the stopped-flow module. The dashed lines in inset represent fluorescence signals corresponding to unfolded baseline extended to refolding urea concentrations. (B) The initial Trp fluorescence obtained from extrapolation of refolding traces to zero time of mixing are denoted as blue symbols while final fluorescence after completion of refolding are shown as the red symbols. The blue and red dashed lines represent fluorescence corresponding to the equilibrium intermediate (in 4-6 M urea) and the native protein at 0.5 M urea respectively. It is found that the initial signals of refolding correlate well with the signals corresponding to equilibrium intermediate in 4-6 M urea (not the unfolded baseline), and do not depend on the unfolding urea concentrations. The protein was equilibrated in a number of unfolding conditions (5, 6, 6.5, 7 and 8 M urea) before collection of refolding traces via dilution to 0.5 M of the denaturant. The equilibrium intermediate (in 4-6 M urea) appears to be populated kinetically during refolding and probably gets trapped in local minima while searching for native conformation under strongly native conditions. (C, D) The relative amplitudes and apparent rates of refolding (k_{obs}) for the protein are independent of its concentration. As the aggregation is a concentration-dependent phenomenon, the plots exclude the possibility of aggregation during refolding (1). The error bars represent standard deviations from three individual experiments.

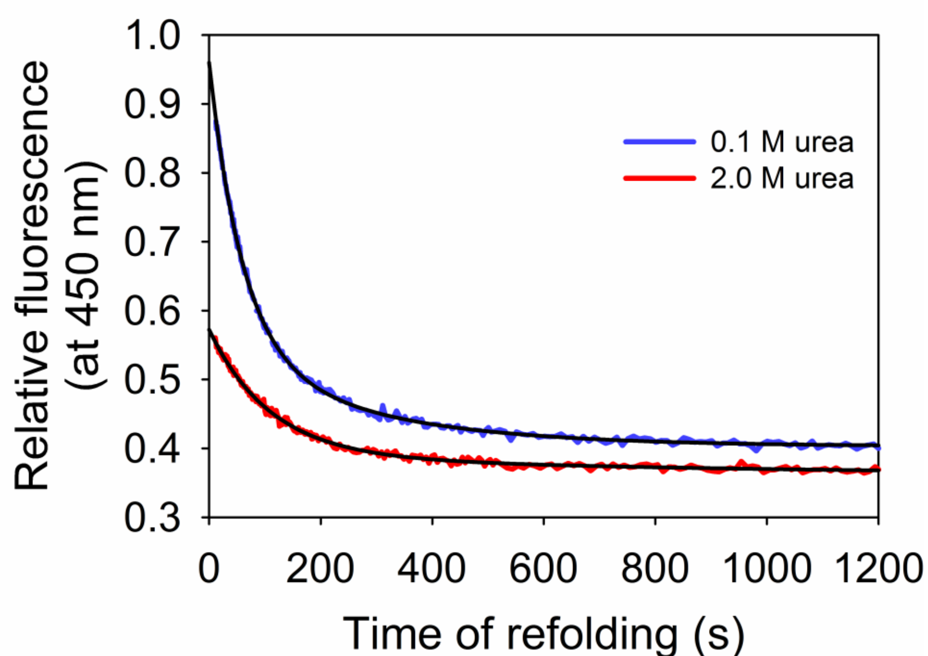

Figure S3. Refolding kinetics of MSG probed by ANS dye. 8-anilino-1-naphthalenesulfonic acid (ANS) molecules bind to the exposed hydrophobic patches of the initial intermediates of the protein and exhibit high fluorescence emission. The release of the dye molecules from the structure due to burying of hydrophobic patches during refolding is traced by fluorescence at 450 nm (excitation at 350 nm). The decrease in refolding amplitude at higher urea represent lesser population of the binding intermediate at higher urea concentration.

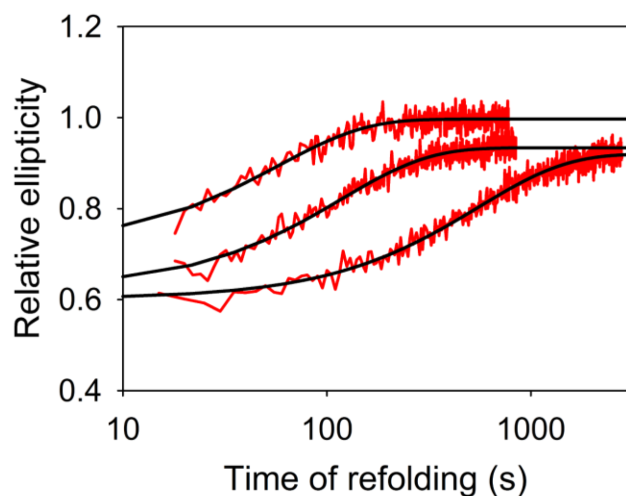

Figure S4. Refolding kinetics of MSG probed by secondary structure content. Representative refolding traces for refolding at 1, 2 and 2.5 M (top to bottom) of urea probed by ellipticity at 222 nm. The thick black lines show single exponential fit of the data, as reliable double exponential fit could not be achieved. Each kinetic trace exhibits average of three kinetics experiments.

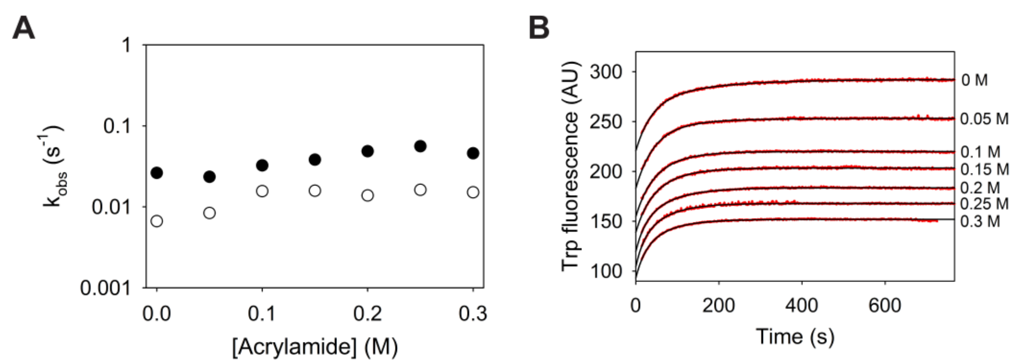

Figure S5. MSG refolding in different acrylamide concentrations. (A) The effect of acrylamide concentration on refolding kinetics of MSG. (B) The extrapolation of refolding traces of MSG, probed by fluorescence at 340 nm (excitation at 295 nm), under different acrylamide concentrations. Each kinetic trace represents average of three traces to reduce noise in the fluorescence data.

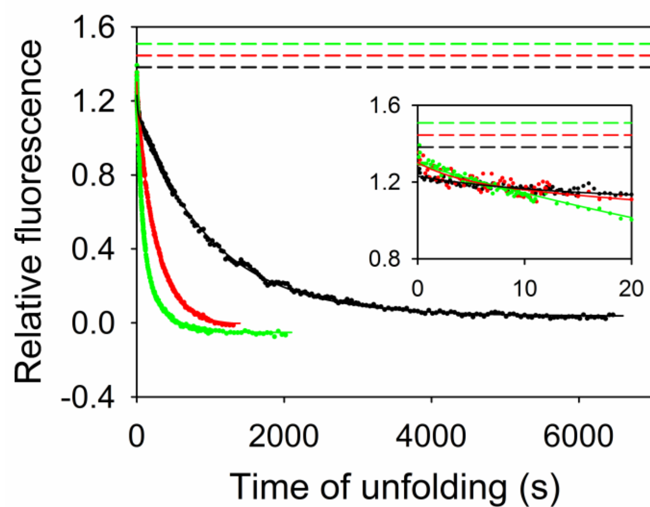

Figure S6. Ensemble unfolding kinetics of MSG. The representative kinetic traces of MSG for unfolding to 5 M (black), 6 M (red) and 7 M (green) of urea with Trp fluorescence probe are shown. The solid lines show exponential fit of the raw data (dots). Inset represents the absence of any fast unfolding phase that could be missed during manual mixing (10 s) method, collected through stopped flow module.

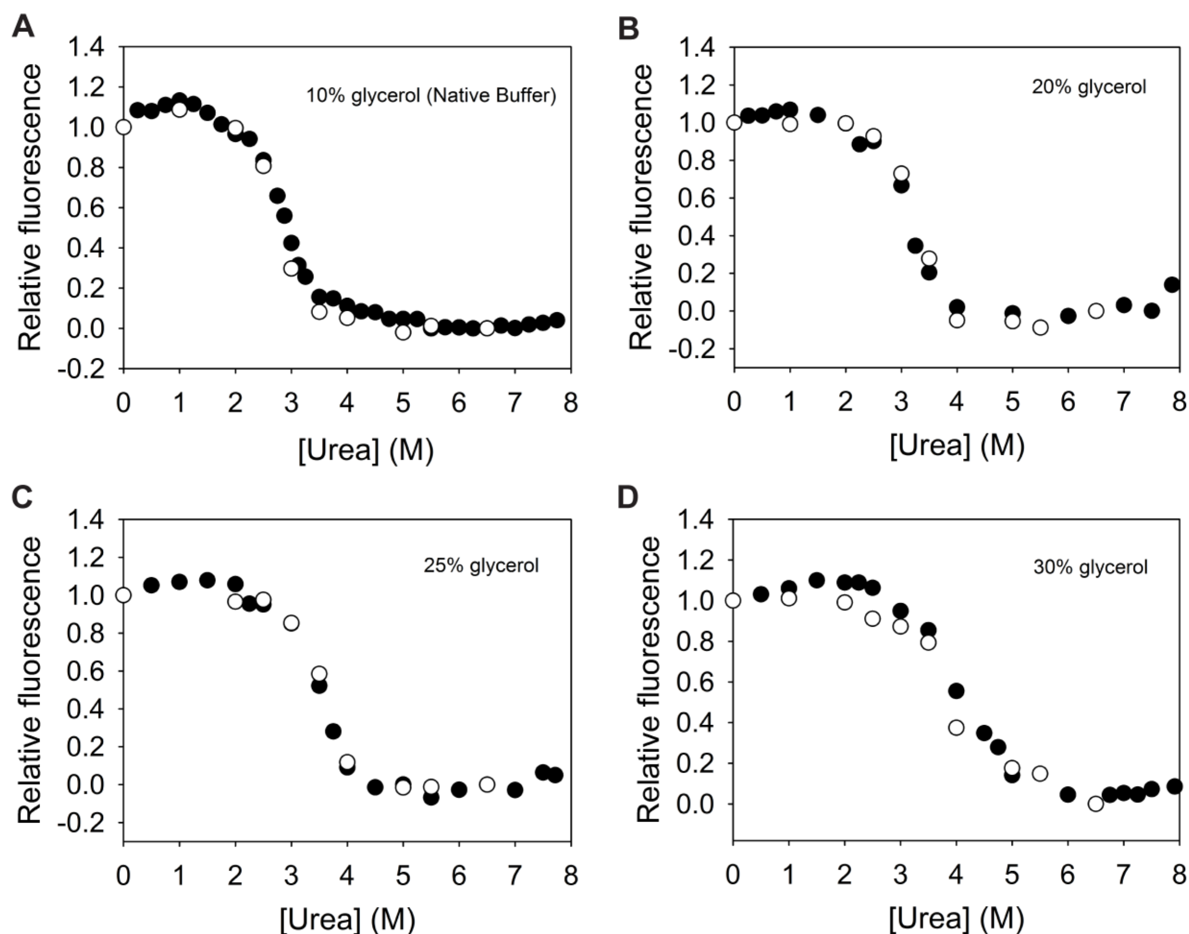

Figure S7. Reversibility of MSG unfolding in presence of glycerol. (A, B, C) The equilibrium unfolding curves for MSG are completely reversible until 25% glycerol in the buffer. The addition of the viscogen stabilizes the protein and shifts the mid-point of denaturation to higher urea indicating a uniform alteration in free energy landscape (2). (D) The equilibrium denaturation curve in presence of 30% glycerol shows hysteresis near 2 M of the denaturant. The cooperativity index for the protein is significantly affected, pointing towards non-uniform alteration in the free energies of native and unfolded states.

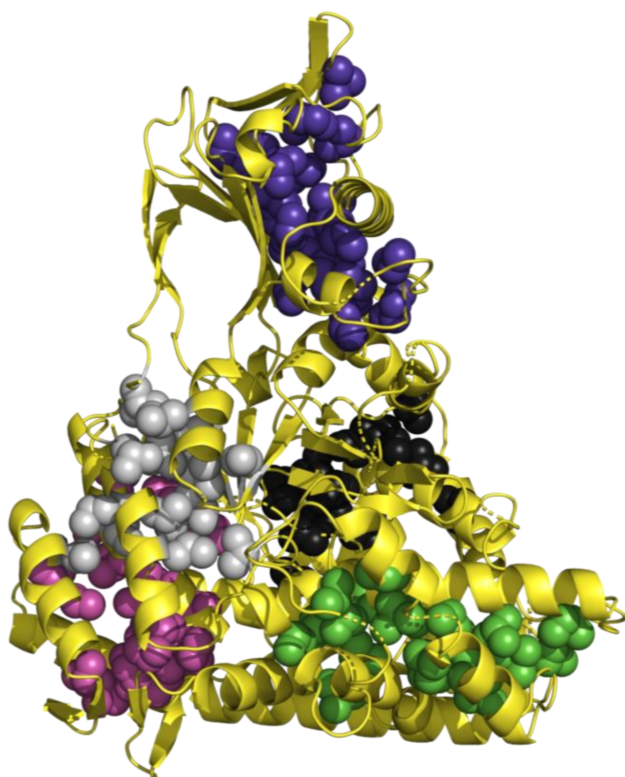

Figure S8. The Branched Aliphatic amino acid Side Chain (BASiC) clusters in MSG. The presence of four major hydrophobic clusters in MSG is shown (Table 2). Where cluster-1 (purple), cluster-2 (green), cluster-3 (magenta), cluster-6 (grey), and cluster-7 (black) are shown. The lower absolute contact order of the ILV residues in the cluster ensures minimum conformational search and points towards fast and stable core formation on sub-millisecond scale.

Table S1. Stern-Volmer (K_{SV}) constants for fluorescence quenching in MSG by acrylamide

Conformational state	K_{SV} (M^{-1})
Native	2.01 ± 0.07
Kinetic misfolded state (I_M) at 1 M urea	4.39 ± 0.09
Equilibrium state at 3.5 M urea	7.85 ± 0.17
Equilibrium state at 4 M urea	7.78 ± 0.15
Equilibrium state at 4.5 M urea	7.68 ± 0.17
Equilibrium state at 5 M urea	9.02 ± 0.27
Equilibrium state at 5.5 M urea	8.49 ± 0.33
Unfolded state at 6.5 M urea	9.10 ± 0.29

Table S2. Parameters from the fitting of interrupted refolding kinetics.

Urea concentration jumps (M)	Probe	Refolding k_{obs} for fitting (s^{-1})	Fractional amplitudes
6.5→0.1→7.0	Fast unfolding phase (phase-1)	1.97e-2	-0.678 ± 0.046
		4.41e-3	0.092 ± 0.029
	Slow unfolding phase (phase-2)	4.41e-3	-0.060 ± 0.010
6.5→2.0→7.0	Fast unfolding phase (phase-1)	1.12e-2	-0.729 ± 0.053
		2.89e-3	0.222 ± 0.044
	Slow unfolding phase (phase-1)	2.89e-3	-0.115 ± 0.009

The fast and slow phases of unfolding at 7 M urea (phases 1 and 2 respectively) probe the population of (I_N+N) and N respectively. The errors represent standard errors of fitting with single and double exponential equations.

References

1. Silow, M., and Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci.* **94**, 6084–6086
2. Hurle, M. R., Michelotti, G. A., Crisanti, M. M., and Matthews, C. R. (1987) Characterization of a slow folding reaction for the α subunit of tryptophan synthase. *Proteins Struct. Funct. Genet.* **2**, 54–63