Supporting Information

A small molecule chemical chaperone optimizes its unfolded state contraction and denaturant like properties.

Sunny Sharma[¥], Suparna Sarkar[¥], Simanta Sarani Paul[¥], Syamal Roy[£] and Krishnananda Chattopadhyay[¥],*

[¥]Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Kolkata 700032

[£]Infectious diseases and Immunology Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Kolkata 700032

*Corresponding author: Krishnananda Chattopadhyay, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S. C. Mullick Road, Kolkata 700032, krish@iicb.res.in

Running title: Biophysical studies on small molecule chaperones

Supporting Materials and Methods:

Reagents and chemicals

L-Arginine, DL-Arginine (for the far-UV CD experiments) and Sucrose were purchased from Sigma Chemical Co. (St. Louis, MO). Acrylamide was obtained from USB Corporation (Cleveland, OH, USA). All other chemicals used were of the highest purity.

Molecular dynamic simulation.

All molecular dynamic simulations were performed in Gromacs version 4.5.3 [1,2] using GROMOS96 53A6 force field [3,4] at constant temperature and pressure (NTP) ensemble. The calculations were carried out using an Apple cluster equipped with Darwin operating system (OS release 9.6.0). Eight simultaneous processors were used for parallel processing of mdrun program of GROMACS using Open MPI module (version 1.4.3). The V-rescale [5] coupling was employed to maintain a constant temperature of 308 K with coupling constant of 0.1 psec for both protein and non-protein molecules in the system. Parrinello-Rahman coupling [6] was used to maintain constant semi isotopic pressure of 1 bar with coupling constant of 2 psec within a fixed volume of a dodecahedron box (diameter =0.75 nm) filled with water molecules (spc 216 water model). Periodic boundary condition was employed for defining perfect 3D tiling of the system. The particle mesh Ewald method (PME)[7] with grid spacing of 0.16 nm, was used for electrostatic calculations. A non-bonded cutoff of 1 nm for Lennard-Jones potential was used. The net charge of the protein was compensated by adding 0.1 M sodium chloride into the system. Steepest descent algorithm with the maximum of five thousand steps was applied for the energy minimization of the protein. Leap-frog integrator[8] was utilized for the molecular dynamics runs.

In the first simulation (simulation #1) we simulated the predicted model for 100 nsec to search for the global minimum with defined energy minimized structure. For the second simulation (simulation #2), we took the protein with the energy minimized structure obtained from the first simulation and induced single amino acid tryptophan mutations at several positions to generate a total of 8 different starting structures. We have virtually mutated the positions 5, 19, 30, 48, 51, 73, 77 and 89 with tryptophan residues using the Discovery studio v2.5.0.9164 software. The mutants were further optimized using molecular dynamics simulations for 30 nsec to obtain energy minimized configuration optimized for the mutant

proteins. In the third simulation (simulation #3), we simulated all the eight optimized structures obtained from the previous simulations for another 20 ns for further refinement of the structures around the inserted tryptophans. In these simulations we restrained the positions of all the atoms except the atoms present within spheres of 10 A around the inserted tryptophan residues. Atoms were restrained at a fixed reference point by applying a force constant in each dimension (x, y, z) having magnitude of 1000 for each atom. Solvent accessible surface area (SASA) of all the inserted tryptophan residues in the eight different structures were calculated from the trajectory file of simulation #2 and simulation #3 using g_sas command of GROMACS program. The initial un-equilibrated period in each trajectory was not taken into consideration while calculating solvent accessible surface area (SASA). All the SASA calculations were showing standard deviation in the range of 0.09 to 0.17.

PyMOL (version 0.99) was used as the visualization software.

Sequence analysis of KMP-11 using NCBI protein-protein BLAST [9] does not show any close homolog with available solved structures. Profile based search performed using PSI BLAST[10], however, indicates the presence of few remote homologs whose crystal structures are available. Since the sequence identities of the remote homologs are low, we have used a composite approach from I-TASSER (iterative threading assembly refinement) [11,12] of Zhanglab server which combines various techniques such as threading, ab-initio modeling and atomic-level structure refinement methods. From the I-TASSER analyses,, we have chosen a model structure which provides the best confidence score [11].

Validation of the predicted structure of KMP-11:

The generated model of KMP-11 has been verified using different computational protocols. Brief descriptions of these protocols and the results have been provided below:

(a) The φ and ψ torsion angles of the model were subjected to the Ramachandran plot analysis using Procheck program version 3.6.2 available at EMBL-EBI. The Ramachandran plot analysis (Figure S1) show that 94.3% of the residues lie within the most favored region, 5.7% of the residues within additional allowed region and no residues with in generously allowed region and disallowed region. G-factor of the overall average has been found to be -0.01. Our predicted model has the value of 94.3% with low G-Factor which suggests that the model structure is valid and of good stereochemical quality.

(b) In order to facilitate the overall quality of our model further, ProSA program (https://prosa.services.came.sbg.ac.at/prosa.php) was employed to study the Z-score of our model against the Z-score of all experimentally determined structures in current PDB. Z-score represents the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The Z-scores of different groups of structures obtained using a number of sources are shown in Figure S2. The NMR and X-ray crystal structures have been shown using dark blue and light blue colors respectively. The black dot in the following figures represents the predicted model. It can be observed that the Z-score value of the predicted model structure of KMP-11 is located well within the space of protein structures determined By NMR and X-ray methods. The present analysis suggests that the predicted model structure of KMP-11 is reasonably reliable and close to experimentally determined structures.

(c) The plot of local model quality predicted by ProSA program by plotting energies as a function of amino acid sequence position *i* over the window of 40 residues (dark green) and 10 residues (light green) shows that the energy values of all residues in our predicted model are well under zero, when observed in a window of 40 amino acid residues (Figure S3). In general, positive values correspond to problematic or erroneous parts of the input structure and hence another validation of good quality model.

(d) The packing quality of the model was verified with the help of Coarse and Fine Packing Quality Control programs of WHAT IF server (http://swift.cmbi.ru.nl/servers/html/index.html). The average score of our model analyzed in Coarse Packing Quality Control programs is 0.263. The predicted fine packing quality of our model shows that the model is made up of near perfect helices with good quality packing. This result is acceptable as the model was not the equilibrated structure and the model itself was obtained from the geometrically constrained crystal structure. To analyze it further, we have performed a 100 nsec molecular dynamics simulation to relax the geometrical constraints and to generate a well equilibrated structure in aqueous environment. The equilibrated structure has the overall average value of -0.687 and 0.53 for Coarse and Fine Packing Quality Control respectively.

(e) The far UV CD analyses of KMP-11 using K2D2 server (http://www.ogic.ca/projects/k2d2/) indicates that the protein contains 77% helix. The predicted structure of KMP-11 contains 80% helix.

(f) The predicted surface accessibility of the tryptophan residues in different tryptophan mutants (SASA values) shows good correlation with the experimentally observed Stern-Volmer constants (Figure 1e).

A 100nsec molecular dynamic simulation with the predicted model as the starting structure has been carried out in aqueous solution to obtain the final model structure of KMP-11 which is shown in Figure 1b in the manuscript text.

Site Directed Mutagenesis

Wild type KMP-11 gene cloned previously in pCMV-LIC vector was further sub-cloned in pET28a vector between NcoI and BamHI using the following set of primers: Fwd: 5'-CCCCATGGCCACCACGTACGAGGAG-3' and Rev: 5'-AAGGATCCCTCCTGATGATGATGATGATGCTTGGAACGGGTACTGCGCAGC-3'.
Site directed mutagenesis was performed with the help of Quick-change^{Tm} Site-Directed Mutagenesis kit (Stratagene, Agilent Technologies) to generate several single tryptophan (W) mutants, using the wild type KMP-11 construct as a template. Table S2 shows the single tryptophan mutants and the primers which were used to generate them. All the mutations were confirmed by DNA sequencing.

Expression and purification of the wild type and mutant KMP-11 proteins

Recombinant KMP-11 constructs (both wild type and mutants) were expressed and purified using Ni-NTA affinity chromatography using the Qiagen supplied protocol (Qiaexpressionist^{tm}, Qiagen, Germany), with slight modifications. The modifications included using 20mM imidazole in both lysis and wash buffers and ignoring addition of lysozyme during cell lysis. The purified protein fractions were checked on 15% SDS-PAGE gel and the amount of protein was estimated using BCA Protein Assay Kit (Pierce, Thermo Scientific, USA) in each case.

Steady state tryptophan fluorescence and acrylamide quenching measurements.

Steady state fluorescence spectroscopy and acrylamide quenching measurements were carried out using a PTI fluorimeter (Photon Technology International, USA). The machine was kept in analog mode. A cuvette with 1cm path length was used for the fluorescence measurements. An excitation wavelength of 295 nm was used to eliminate any contributions from the tyrosine fluorescence. Step size was 1nm; with an integration of 1sec. Excitation and

emission slits were 2.5nm in each case. Emission spectra between 305nm and 400nm were recorded in triplicate for each measurement. Typical protein concentration of 3μM was used. For the acrylamide quenching experiments, the protein solutions were titrated with a stock of 5M acrylamide. Necessary background corrections were made in each experiment.

Acrylamide Quenching Data Analysis: Assuming F_0 and F to be the intensity values of tryptophan fluorescence of a mutant protein in the absence and presence of acrylamide concentration [Q], the Stern-Volmer Equation [13] can be represented as follows:.

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \qquad [S1]$$

K_{SV} is the Stern-Volmer constant which can be determined using a linear plot of F_0/F vs. acrylamide concentrations [Q].

It is important to point out here that the K_{SV} values obtained from the steady state fluorescence experiments may have complications arising from the components of static quenching. This is because, steady state fluorescence intensity may contain both static and dynamic components. Fluorescence lifetime of tryptophan residue does not change by static quenching; and hence time resolved fluorescence experiments can un-ambiguously determine the dynamic quenching constant. We have carried out time resolved fluorescence measurements with representative tryptophan mutants (Y48W, F73W and F77W) and the dynamic quenching constants measured by the lifetime data show identical trend (Figure S4).

Fluorescence lifetimes were measured with a time-correlated-single-photon-counting (TCSPC) apparatus (Fluorocube Fluorescence Lifetime System by Horiba JovinYvon, Japan) using a picosecond pulsed diode laserwithexcitation at λ_{ex}= 295 nm. The total intensity decay curves, I(t), were fitted to a bi-exponential function:

$$I(t) = A_1\exp(-t/\tau_1) + A_2\exp(-t/\tau_2) \qquad [S2]$$

where A_i, τ_i denote the amplitude and lifetime of the i-th species. The fitting was done by an iterative procedure using the DAS 6.2 data analysis software supplied by IBH. Reduced χ^2 and weighted residuals served as parameters for ensuring goodness of fit.

The average lifetime of the tryptophan residue has been calculated using the following equation:

$$\tau_{av} = \frac{A_1\tau_1^2 + A_2\tau_2^2}{A_1\tau_1 + A_2\tau_2} \qquad [S3]$$

The average lifetimes in the absence and presence of acrylamide quencher have been denoted as t_{av0} and t_{av} respectively.

Far and near-UV CD experiments.

Far and near-UV CD experiments were carried out using a JASCO J720 spectropolarimeter (Japan Spectroscopic Ltd.). Far-UV CD measurements (between 200nm and 250 nm) were performed using a cuvette of 1mm path length. Protein concentration was 3μM. Scan speed was 50nm/min; with response being 2 sec. Bandwidth was set at 1nm. Three to five CD spectra were recorded in continuous mode and averaged.

For the near-UV CD experiments (between 250nm and 350 nm) a cuvette of 10mm path length was used. Scan speed was 10nm/min; with response being 2 sec. Bandwidth was set at 1nm. Protein concentration was 30μM in this case. Three to five CD spectra were recorded in continuous mode and averaged for each CD experiment.

Thermal Unfolding Experiments of the KMP-11 mutants using far-UV CD:

Temperature induced unfolding experiments of WT and mutant proteins were carried out using a JASCO J720 spectropolarimeter (Japan Spectroscopic Ltd.). These measurements were performed using a cuvette of 1 mm path length which was placed inside a temperature controlled cell holder whose temperature could be programmed using a computer interface. Changes in the secondary structure were monitored by measuring far-UV CD of the protein samples at 222nm. A heating rate of 1^0 C/min was used and the temperatures were varied between 20^0 C and 90^0 C. Bandwidth was 1nm and the response time was 1 sec. The reversibility of thermal transitions was checked and the extents of refolding were determined for all unfolding experiments.

Data analysis of the Thermal Unfolding Experiments: Thermal unfolding transitions were fit using the following equation (Equation S4). Equation S4 assumes the reversible unfolding between the native and heat induced unfolded protein:

$$\theta = \frac{\theta_N + m_N T + (\theta_D + m_D T)\exp[\frac{\Delta H}{R(\frac{1}{Tm} - \frac{1}{T})}]}{1 + \exp[\frac{\Delta H}{R(\frac{1}{Tm} - \frac{1}{T})}]} \qquad [S4]$$

Where θ is the observed ellipticity at 222nm; θ_N and θ_D are intercepts of the native and unfolded baselines; m_N and m_D are the slopes of the native and unfolded baselines; ΔH is the Vant Hoff's enthalpy; T_m is the mid-point of an unfolding transition. Since, the thermal unfolding transitions were not completely reversible; scan rate dependence of the mid-point of unfolding transition was expected. Hence, all unfolding experiments were performed at a constant scan rate of 1^0C/min.

Next, the refolding efficiency of the thermal unfolded protein was calculated for each case. For this purpose, far-UV spectra were measured both at 20^0 C and 90^0 C before and after the unfolding transitions, respectively. A far-UV CD spectrum was also measured, after the temperature was lowered back to 20^0 C, and the refolding percentages were calculated, using the following formula:

$$\%Refolded = \frac{(\theta_{Refolded} - \theta_{Unfolded})}{(\theta_{Folded} - \theta_{Unfolded})} \times 100 \qquad [S5]$$

Where $\theta_{Folded,}$ $\theta_{Refolded}$ and $\theta_{Unfolded}$ are the ellipticities at 222nm of native, refolded and unfolded baselines, respectively.

Fluorescence correlation spectroscopy (FCS): Fluorescence correlation spectroscopy has been carried out using a Confocor 3 LSM Meta system (Carl Zeiss, Evotec, Jena, Germany) as described before [14]. In a FCS experiment, the fluorescence fluctuations of fluorescently labeled molecules diffusing in and out of a small observation volume are measured and analyzed using suitable correlation function analyses. The measured diffusion time of the molecule can be used to calculate the hydrodynamic radius using Stokes' Einstein formalism. For these measurements, three double mutants (Y48W/S9C, F73W/S9C, and F77W/S9C) have been made. The single cysteine residue in each mutant has been labeled with

Alexa488Maleimide. The labeling procedure, the FCS setup and analyses protocols are discussed in detail in our previous publication [14].

All data analyses and curve fittings were carried out using Origin 8 software.

Figures:

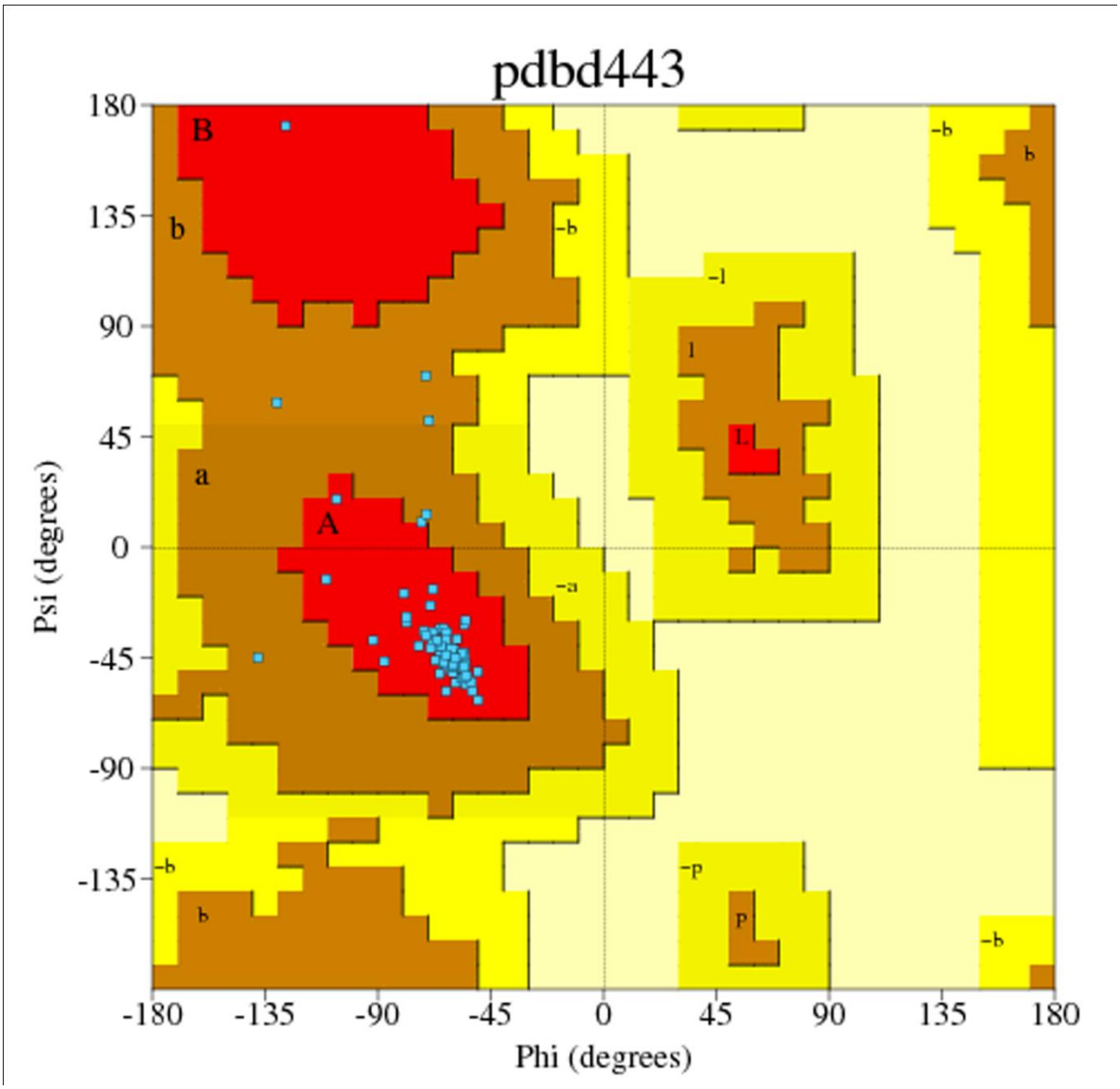

Figure S1:Most favoured regions in the Ramachandran plot are shown in red; additional allowed regions are shown in brown; generously allowed regions are shown in yellow and disallowed regions are shown in light yellowcolor. Blue color dots represent (φ, ψ) angles for each residue of the predicted structure. Maximum points are within the most favoured region for right handed alpha helix.

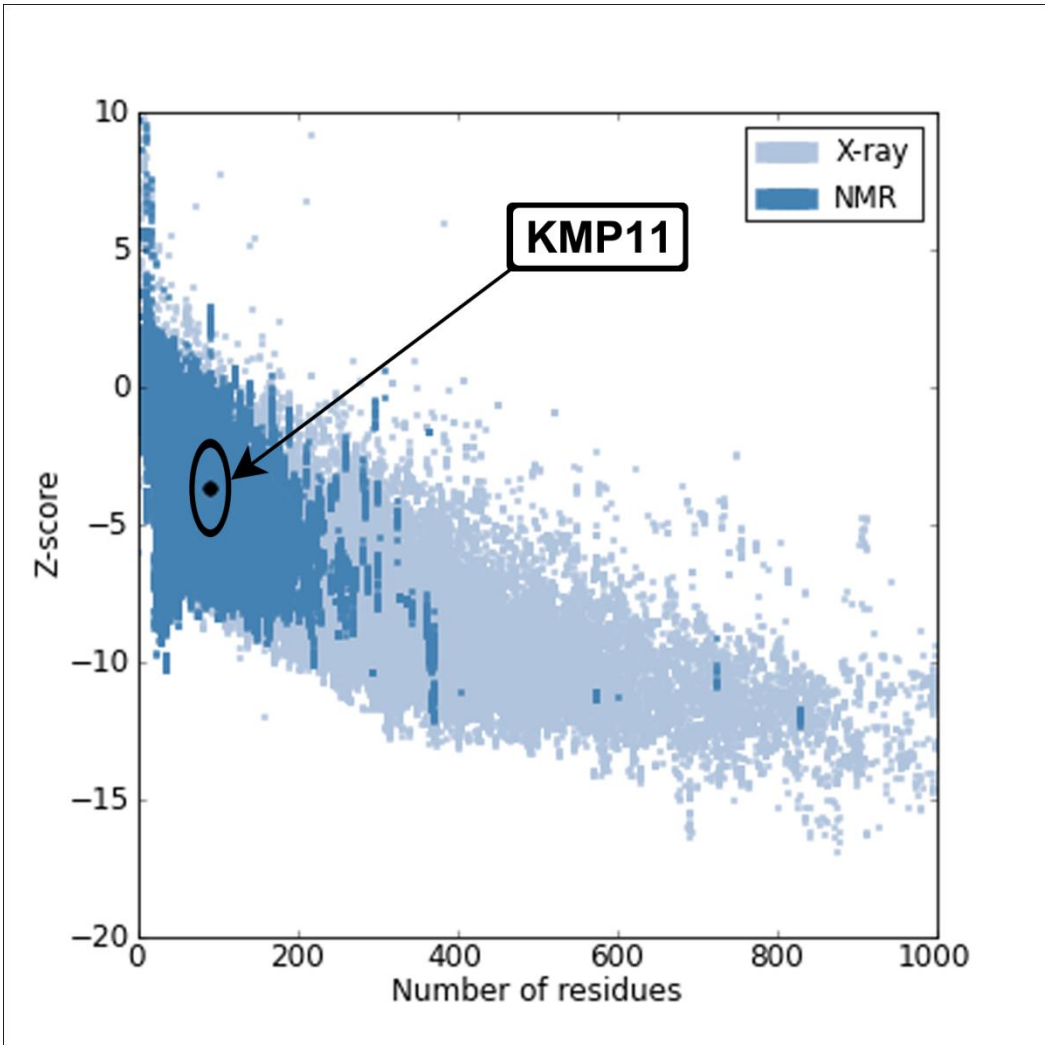

Figure S2: Plot showing the spread of Z-score (indicating overall model quality) range for all known protein having solved structure. X-axis indicates protein sequence length. Light blue points represent Z-score for different length protein whose structure is obtained by X-ray diffraction technique and dark blue for NMR techniques. Black dot represent the result of model structure for KMP11 which is well within the range of Z-score for all protein having solved structure and a length of 92 residues.

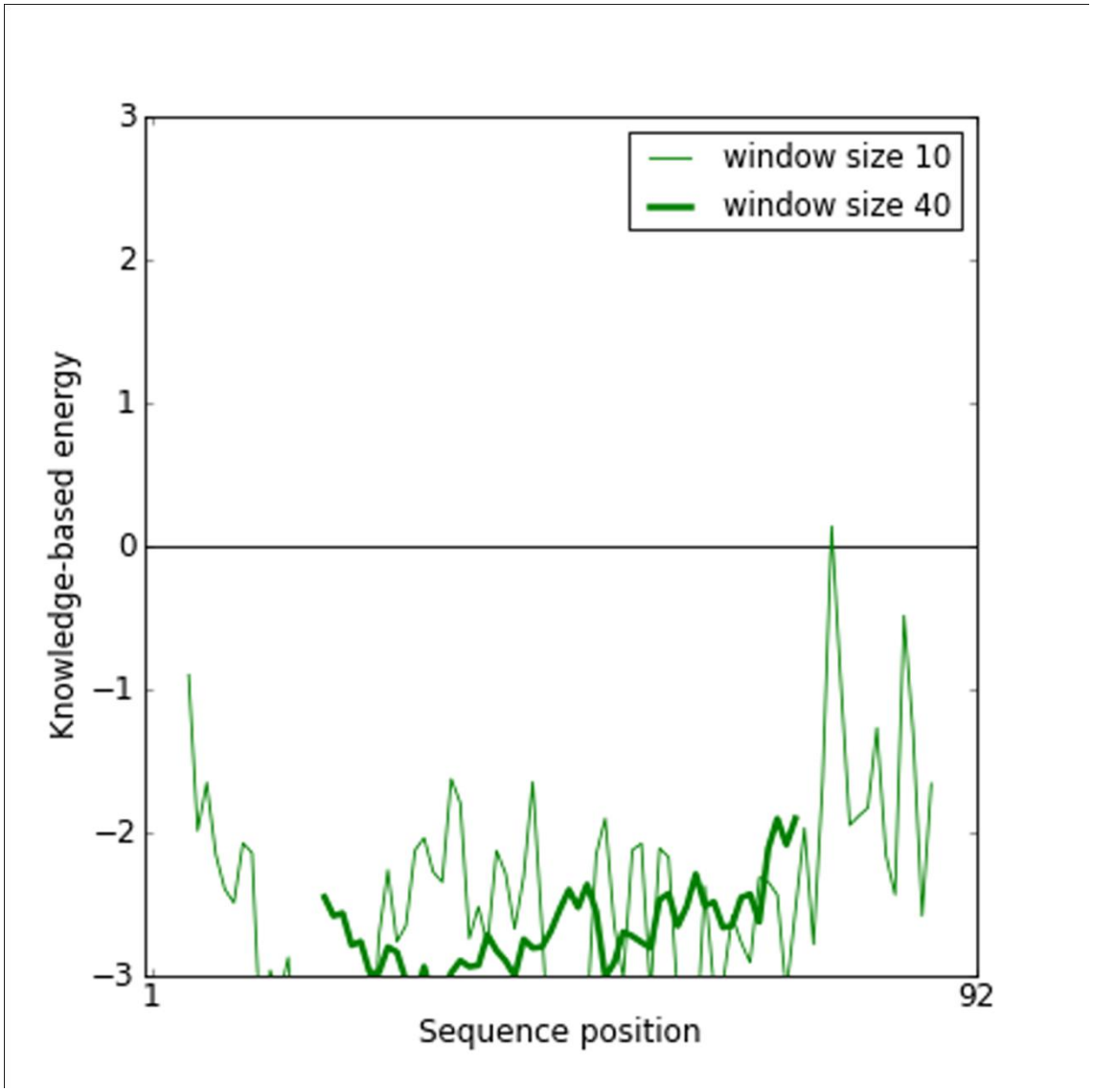

Figure S3: Plot showing the change in energy as a function of amino acid sequence of a particular position over a window of 10 residues shown in light green and a window of 40 residues shown in dark green.

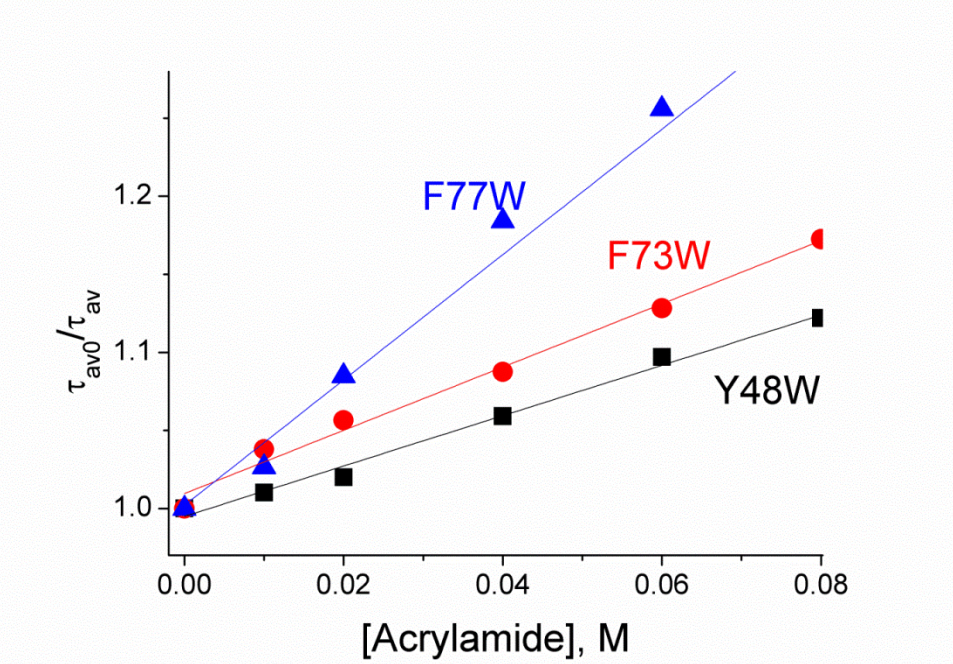

Figure S4: Acrylamide quenching experiments using time resolved fluorescence spectroscopy: The trend (determined using three independent experiments with different samples) observed with the time resolved data is similar to that observed with the steady state quenching measurements, which is described in detail in the manuscript text.

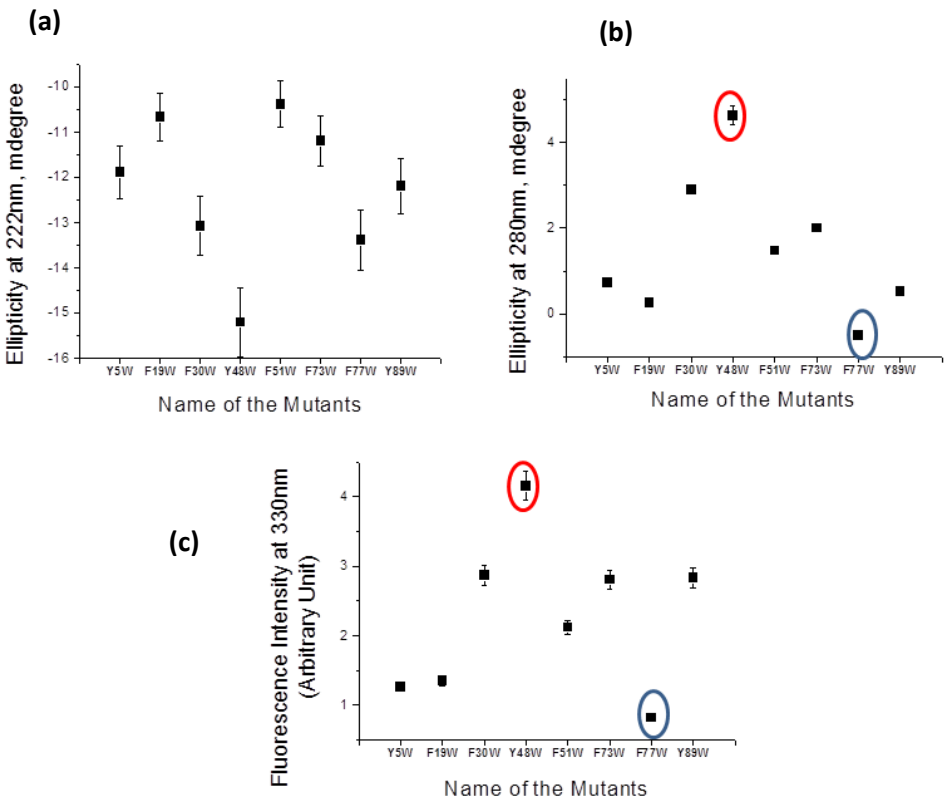

Figure S5: (a) Far, (b) near-UV CD and (c) steady state tryptophan fluorescence data for all KMP-11 mutants: The near-UV CD and steady state fluorescence intensity of Y48W and F77W are marked by red and blue circles respectively. The error bards were calculated using three independent experiments.

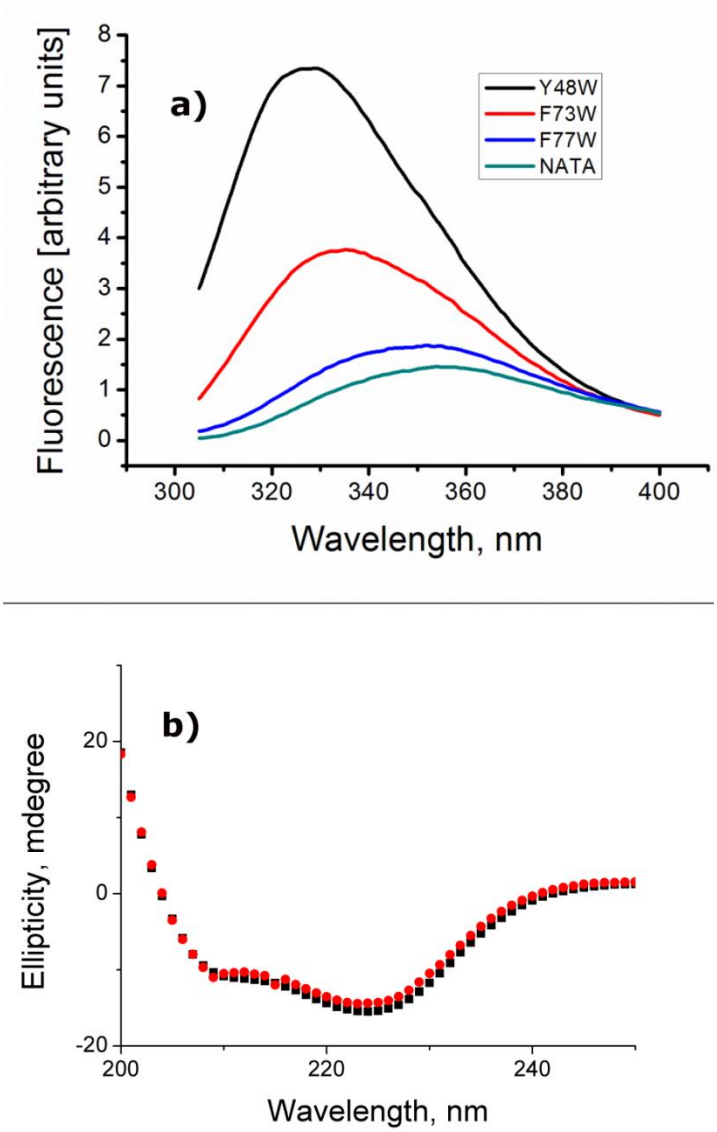

Figure S6: (a) Steady state fluorescence experiments with Y48W, F73W, F77W and n-acetyl tryptophanamide (NATA). The excitation wavelength of 295nm has been used. The emission maxima and the fluorescence intensity values of the F77W mutant and NATA are almost identical. (b) Far UV CD experiments suggest that the labeling by the fluorophore (red corresponds to Alexa488Maleimide labeled Y48W while the black curve shows the Y48W mutant) does not lead to any significant change in the secondary structure. All the other mutants also showed similar behaviors. The experiments are performed in 20mM sodium phosphate buffer at pH 7.4.

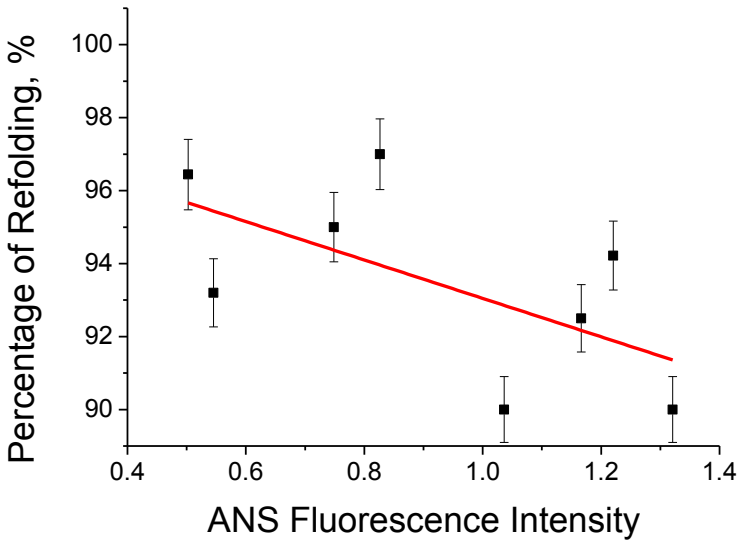

Figure S7: The inverse correlation between the ANS fluorescence intensity and the percentage of refolding efficiency establishes that the mutant proteins which bind ANS more efficiently have relatively higher aggregation propensity. The error bars of the percentage of refolding were calculated using three measurements.

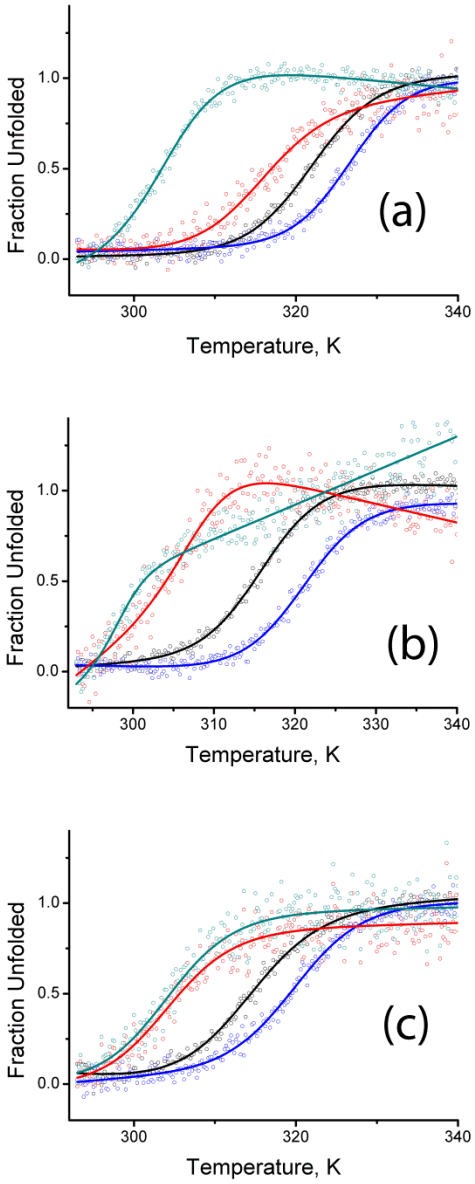

Figure S8: Thermal unfolding transitions of (a) Y48W (b) F77W and (c) F73W mutants in the absence (black) and presence of 500mM arginine (red), 500mM gdn.HCl (dark green) and 1M sucrose (blue). The lines drawn through the unfolding data correspond to their fits using a two-state transitions between the native (N) and unfolded (U) proteins. In the presence of gdn.HCl, particularly with F77W (b) and F73W (c) mutants, the unfolding transitions lack defined pre-transition baselines and hence they are difficult to fit using the two-state assumptions.

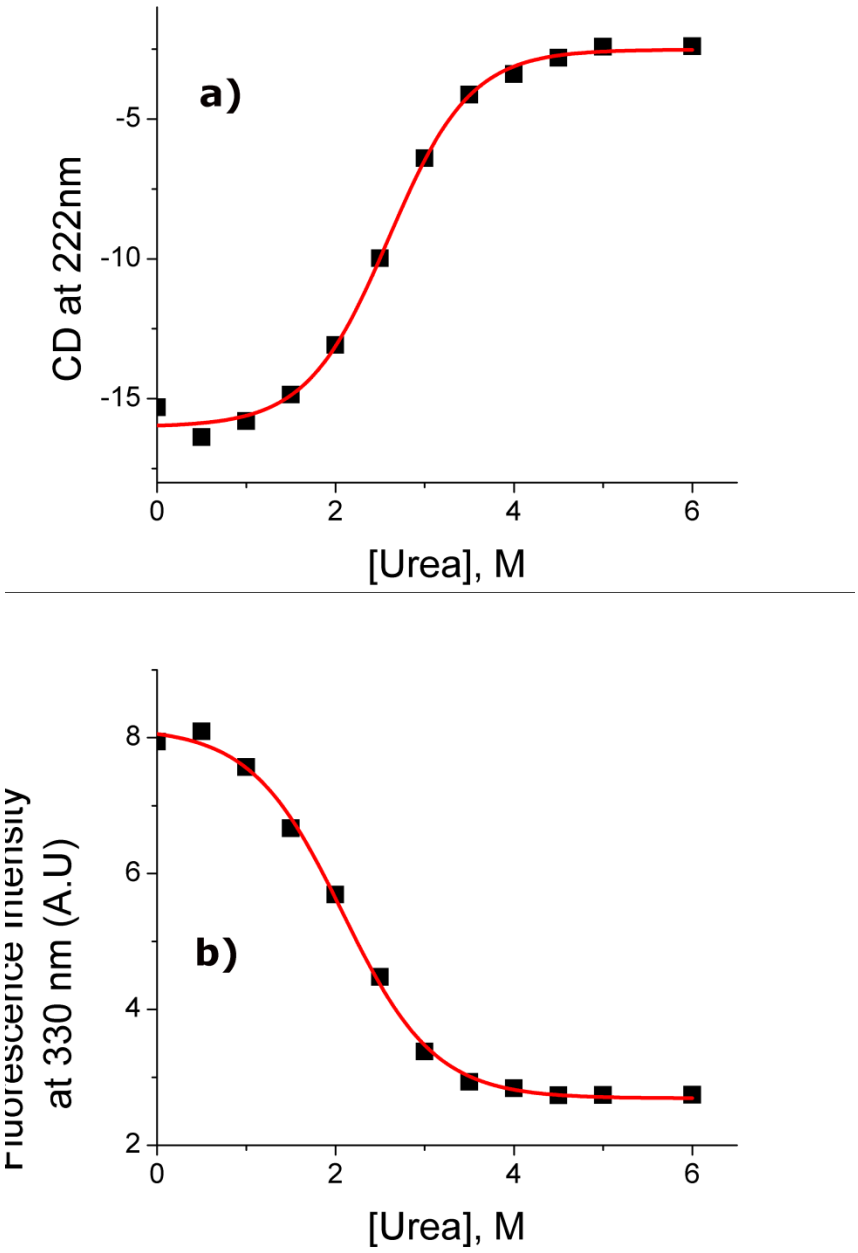

Figure S9: Urea induced unfolding transitions of Y48W monitored by a) CD at 222nm and b) steady state fluorescence intensity at 330nm. The red lines drawn through the data indicates the fit using a simple two state model.

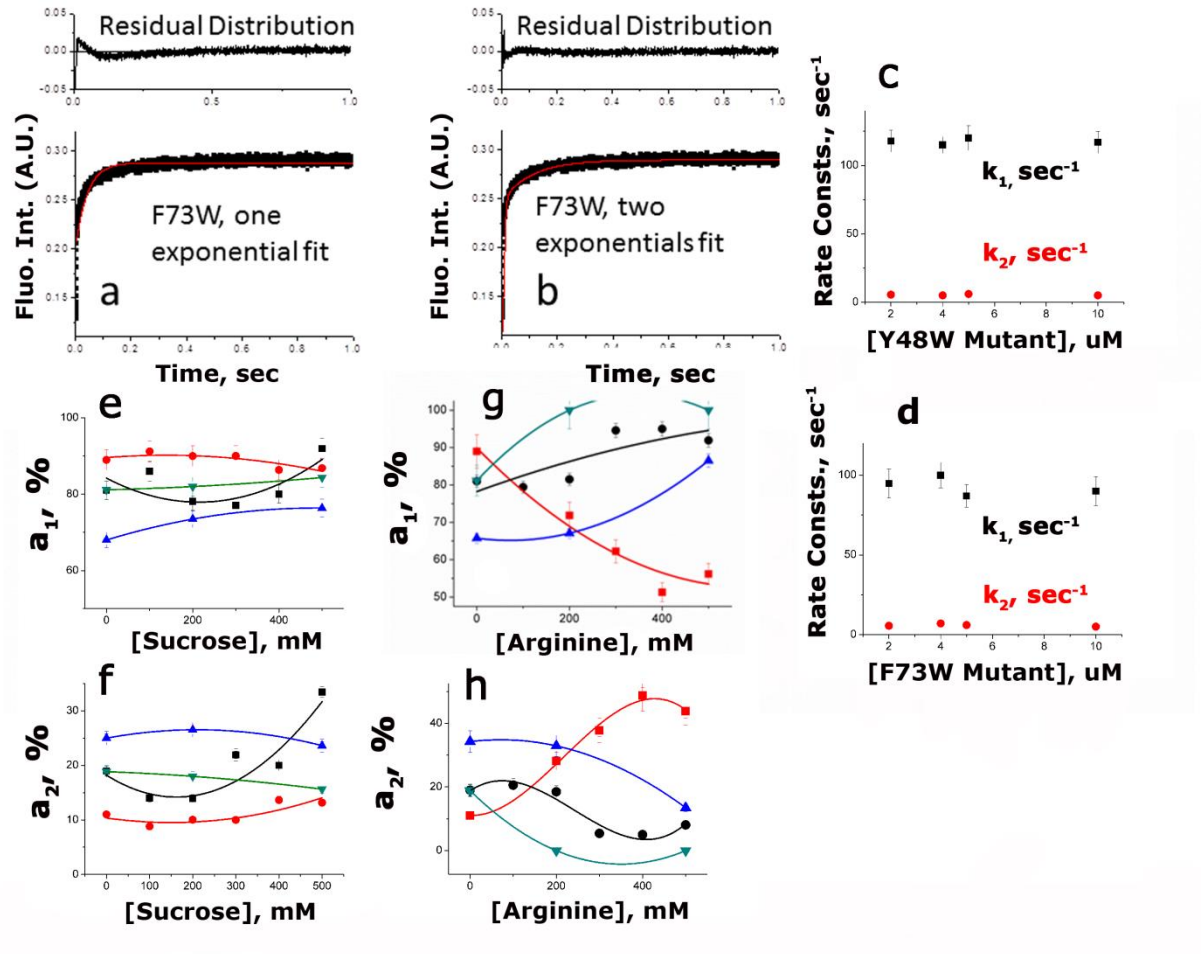

Figure S10: Refolding kinetics of the KMP-11 mutants using stopped flow fluorescence: (a) a single exponential fit to the F73W kinetics data did not yield satisfactory results (the red line drawn through the data shows the fit and the residual distribution is shown above the kinetic trace). (b) A two-exponential fit improved the kinetics fit significantly. The rate constants (k_1 and k_2) did not change with protein concentrations for any of the mutants and the representative data for Y48W (c) and F73W (d) mutants are shown. The amplitudes of the (e) fast, a_1 and (f) slow, a_2 kinetics component did not vary with the concentration of sucrose. In the presence of arginine, the amplitude of the fast component (a_1) increased (g) for the partially folded proteins. The amplitude of the second component (a_2) showed a proportional decrease with the increase in arginine concentration for these mutants (h).

In figures e-h red trace represents Y48W (native like control), whereas black, green and blue traces represent F73W, F51W and F30W respectively (all three partially folded). All error bars were calculated using three independent measurements.

Table S1: Tryptophan mutants of KMP-11, calculated solvent accessible surface area (SASA, %) and the experimental values of K_{SV} (M^{-1}):

Mutant class	Name of the mutant	Solvent surface accessibility (SASA) of tryptophan (%)[!]	The values of K_{SV}, M^{-1}[!!]
Buried Tryptophan	Y48W	12	3.13±0.2
Intermediate solvent exposure	Y5W	52.5	5.68±0.4
	F19W	36.3	4.68±0.3
	F30W	31	5.14±0.4
	F51W	34.3	4.27±0.3
	F73W	43	4.4±0.3
	Y89W	56.2	5.79±0.4
Completely exposed tryptophan	F77W	80.3	9.7±0.7

[!]**Solvent surface accessibility values were determined using computational methods as described in the materials and methods section. Percentile values of solvent accessible surface were calculated by comparing the surface accessibility value of tryptophan in the fully exposed conformation, in a short gly_gly_trp_gly_gly peptide.**

[!!]**K_{SV} values were measured using acrylamide quenching experiments.**

Table S2**:** Tryptophan mutants of KMP-11 and the sequences of the primers used:

Tryptophan mutations	*Primers (5' → 3')*
Y5W	Fwd: CCATGGCCACCACGTGGGAGGAGTTTTCGGCG
	Rev: CGCCGAAAACTCCTCCCACGTGGTGGCCATGG
F19W	Fwd: CCGCCTGGATGAGGAGTGGAACCGGAAGATGCAGG
	Rev: CCTGCATCTTCCGGTTCCACTCCTCATCCAGGCGG
F30W	Fwd: GAGCAGAACGCCAAGTGGTTTGCGGACAAGCCGG
	Rev: CCGGCTTGTCCGCAAACCACTTGGCGTTCTGCTC
Y48W	Fwd: CGAGATGAAGGAGCACTGGCAGAAGTTCGAGCGCAT
	Rev: ATGCGCTCGAACTTCTCCCAGTGCTCCTTCATCTCG
F51W	Fwd: GGAGCACTACGAGAAGTGGGAGCGCATGATCAAGGA
	Rev: TCCTTGATCATGCGCTCCCACTTCTCGTAGTGCTCC
F73W	Fwd: GAGCACTCGGAGCACTGGAAGCAGAAGTTCGCCG
	Rev: CGGCGAACTTCTGCTTCCAGTGCTCCGAGTGCTC
F77W	Fwd: CGGAGCACTTCAAGCAGAAGTGGGCCGAGCTGCT
	Rev: AGCAGCTCGGCCCACTTCTGCTTGAAGTGCTCCG
Y89W	Fwd: AGCAGAAGGCTGCGCAGTGGCCGTCCAAGC
	Rev: GCTTGGACGGCCACTGCGCAGCCTTCTGCT

Table S3: Experimentally measured K_{SV} values obtained by the acrylamide quenching experiments with the tryptophan mutants of KMP-11 in the presence of different concentrations of arginine, sucrose and gdn.HCl.

Name of the Mutants	K_{SV}, M^{-1}			
	No additive	500mM arginine	500mM sucrose	500mM gdn.HCl
Y48W	3.1±0.2	3.4±0.2	3±0.2	--
Y5W	5.7±0.3	7.6±0.4	5.3±0.3	9±0.6
F19W	4.7±0.2	6.7±0.5	4.2±0.2	8.9±0.7
F30W	5.1±0.3	7.1±0.4	4.9±0.2	9±0.6
F51W	4.3±0.2	6.4±0.3	4±0.2	9±0.8
F73W	4.4±0.2	6±0.3	4.4±0.2	9±0.7
F89W	5.8±0.3	7.5±0.4	5.6±0.3	8.9±0.7
F77W	9.7±0.5	9±0.5	9±0.5	8.8±0.6

The error bars were determined using three independent measurements. Interestingly, the K_{SV} values decrease in the case of many mutants in the presence of sucrose and for F77W in the presence of arginine. Although this decrease in K_{SV} is not completely understood, acrylamide quenching experiments with n-acetyl tryptophanamide (NATA) show decrease in the values of K_{SV} induced by the addition of both sucrose and arginine. It can be speculated from these data that the changes in solution physical behavior (for example, the increase in the viscosity of the solution) in the presence of sucrose or arginine may contribute to the decrease in the K_{SV} values.

Table S4: Thermodynamic parameters and refolding efficiencies of all the mutants in the absence and presence of different additives[!]

Name of the mutants	No Additives			500mM Arginine			500mM Sucrose			500mM Gdn.HCl
	[!!]ΔH, kcalM^{-1}	Mid-point, K	[$]RE, %	ΔH, kcalM^{-1}	Mid-point, K	RE, %	ΔH, kcalM^{-1}	Mid-point, K	RE,%	RE, %
Y48W	51.6	322	78	45.8	316	85	55.7	325	85	70
Y5W	51.5	318	85	41	309	99	55	321	92	
F19W	27	311	79	38.2	300	99	39.6	314	88	
F30W	53.9	316	75	42	308	85	59	319	81	
F51W	49.7	313	80	32	288	92	55	315	87	
F73W	53.6	315	75	42	305	99	57.5	318	80	68
Y89W	44.9	316	81	40.8	304	100	50	318	87	
F77W	53.5	316	82	51.9	308	90	57.1	319	87	65

!All these experiments were repeated at least three times and the average of three measurements for the mid-points and RE has been shown. The error bars for values of mid-points and RE are less than 5%.

[$]RE stands for the refolding efficiency.

!!ΔH values obtained from the fit has significant fluctuation between three measurements and hence the best values are shown. Because of these fluctuations, ΔH values are not discussed in the paper.

References:

1. van der Spoel, D., *et al.* Gromacs User Manual version 4.5.4. (2010). www.gromacs.org.

2. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory. Comp.* **4**, 435-477 (2008).

3. Oostenbrink, C., Soares, T.A., van der Vegt, N.F., and van Gunsteren, W.F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* ***34***, 273-284 (2005).

4. van Gunsteren, W.F., et al. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Vdf Hochschulverlag AG an der ETH Zu¨rich: Zu¨rich. (1996).

5. Bussi, G., Donadio, D., and Parrinello, M. Canonical sampling through velocity rescaling. J. Chem. Phys. ***126***, 014101 (2007).

6. Parrinello, M., and Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* ***52***, 7182-7190 (1981).

7. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. A smooth particle mesh ewald potential. *J Chem Phys* ***103***, 8577-8592 (1995).

8. Hockney, R.W., Goel, S.P., and Eastwood, J. Quiet High Resolution Computer Models of a Plasma. *J Comp Phys* ***14***, 148-158 (1974).

9. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. Basic local alignment search tool. *J Mol Biol* ***215***, 403-410 (1990).

10. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. acids res.* ***25***, 3389-3402 (1997).

11. Roy, A., Kucukural, A., and Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* ***5***, 725-738 (2010).

12. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* ***9***, 40 (2008).

13. Lakowicz, J.R. Principles of fluorescence spectroscopy. Kluwer Academic/Plenum Publishers, New York (1999).

14. Haldar, S., and Chattopadhyay, K. Effects of arginine and other solution additives on the self-association of different surfactants: an investigation at single-molecule resolution. *Langmuir* ***27***, 5842-5849 (2011).