Supplementary Information

Structurally different chemical chaperones show similar mechanical roles with independent molecular mechanisms

Deep Chaudhuri[1†], Debojyoti Chowdhury[2†], Soham Chakraborty[3†], Madhu Bhatt[3], Rudranil Chowdhury[3], Aakashdeep Dutta[3], Ayush Mistry[3], Shubhasis Haldar[1,2,3*]

[1]Department of Chemistry, Ashoka University, Sonepat, Haryana, India

[2]Department of Chemical and Biological Sciences, S.N. Bose National Center for Basic Sciences, Kolkata, West Bengal

[3]Department of Biology, Trivedi School of Biosciences, Ashoka University, Sonepat, Haryana, India

†Authors contributed equally to this work

*To whom correspondence may be addressed. Email: shubhasis.haldar@bose.res.in

Protein L preparation:

The structure of the protein was obtained from PDB ID 1hz6 with a resolution of 1.7 Å[1]. This contains three chains, among which we used a chain having the protein L B1 sequence (chain B) that matches our in vitro experimental setup. The water elimination and other structural modifications mentioned above were performed using PyMOL version 2.5.2. The H++ server was used to determine the protonation states of the amino acid residues in protein L[2].

Solvation box preparation:

In our study, we performed SMD simulation to unfold the protein L; relaxation MD simulation and native protein-osmolyte simulation each with and without osmolytes (TMAO and trehalose). In all of these simulations, the CHARMM36 force field[3] was used for the protein, the Shea(m) force field[4] (Appendix 1) was used for TMAO, and the TIP3P model[5] was used for water molecules. The force field for trehalose was prepared from the PRODGR webserver (Appendix 2). The salt (NaCl) concentration was kept at 0.15 M using the distance method according to the respective box size in all of these simulations. The TMAO and trehalose concentrations were kept at 3 M and 1 M for each corresponding box size. The box detail is summarized in supplementary Table 1. Periodic boundary conditions were applied for each system. All the simulations were carried out using NAMD 3.0. Three replicates of each simulation were carried out, and the results are presented as averages with error analysis.

Equilibration and energy minimization:

Each system was energy minimized for 40 ps (20000 timesteps) and equilibrated for 250 ps (125000 timesteps). The systems were equilibrated under NPT conditions at 310 K with the damping coefficient set as 5 ps^{-1}. The pressure was regulated using a Nosé–Hoover Langevin piston with a decay period of 50 fs and a damping time constant of 10 fs. The system was then equilibrated under NVT conditions at 310 K.

Steered molecular dynamics (SMD) simulation:

The SMD simulations for all the systems were carried out in the explicit solvent under NPT conditions at 310 K with NAMD 3.0 and the CHARMM36 force field. For each system, the SMD simulation was started with the same configuration that was saved from the equilibration phase. The first residue, the C_α-terminus, was held fixed, and the other C_α-terminus was attached to the pseudoatom, which was pulled with a virtual spring using a spring constant of 415 pN/Å. The SMD-atom was pulled along a vector that links the fixed and the

SMD atoms. The pulling velocity was set to 0.01 Å/ps because. All other atoms are free to move during the simulations. Each protein was pulled up to 25 ns (2 fs/step), as that was sufficient to completely unfold the protein. For relaxation simulation, the final coordinates of each unfolding simulation were used. A time step of 2 fs was used for relaxing the systems using a harmonic potential decreasing from 7.0 to 5.0 to 1.0 kcal/mol $Å^2$. Relaxation was only performed for 25 ns, as it was enough to visualize the difference in refolding in the absence and presence of osmolytes.

Osmolyte-folded protein L simulation:

The simulation started from the native folded structure of protein L kept inside the box with a corresponding number of osmolyte molecules, 0.15 M NaCl, and water molecules. The CHARMM36 force field was used with the TIP3P water model, and simulations were carried out with the Gromacs Program of 2019.4 software version[6]. The systems were prepared as mentioned above, and the simulations were run for 100 ns both with and without TMAO (Supplementary Table 1). The same simulation protocol was used for trehalose. Each system was equilibrated at 310 K and 1 bar pressure and then simulated with a 2 fs step size for 100 ps. Energy minimization was carried out using the steepest descent algorithm[7]. Each system reached equilibrium at 0.1 ns. Subsequently, the systems were gradually heated using a Berendsen thermostat[8] with a coupling constant of 0.1 ps to reach a temperature of 310 K to perform equilibration in the NVT ensemble. Later, the solvent density was maintained using a Parrinello-Rahman barostat[9] with a coupling constant of 0.1 ps at 1 bar and 310 K to perform equilibration in the NPT ensemble by slowly releasing the restraint on heavy atoms in multiple stages. Simulations were run for 100 ns in each case (protein L; protein L + 3 M TMAO; protein L + 1 M trehalose).

Analysis of computational results:

Simulation result analysis:

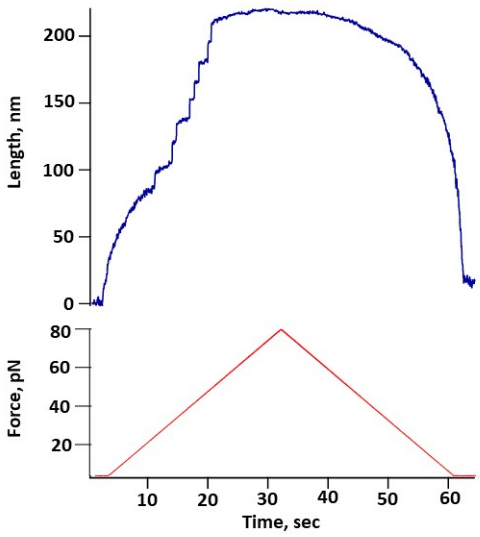
For H-bonding calculations, the cutoff length for conventional H-bonds was set at 3.5 Å. All the results from the simulations were visualized in VMD[10]. The graphs were generated using grace and Igor Pro 8.0. Simulation data were analyzed with BIOVIA Discovery studio visualizer (v21.1.0.20298, San Diego: Dassault Systèmes, 2021).

Protein L sequence:

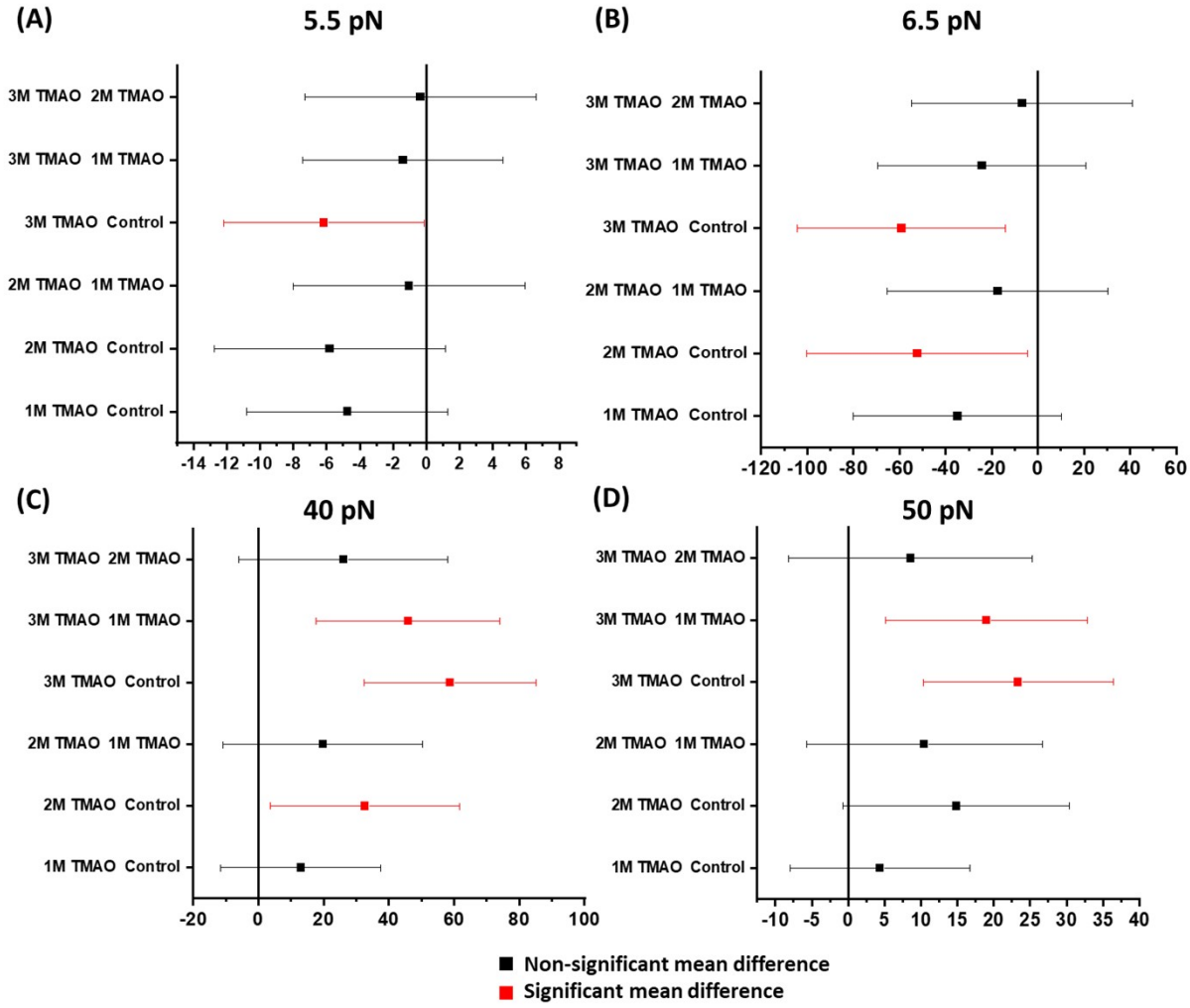
The detailed sequence of the construct is listed below, with the color labeled for the corresponding component in the construct. A short dipeptide linker between two adjacent components of the construct is underlined.

HaloTag-**(PL)$_8$**-His$_6$Tag-AviTag

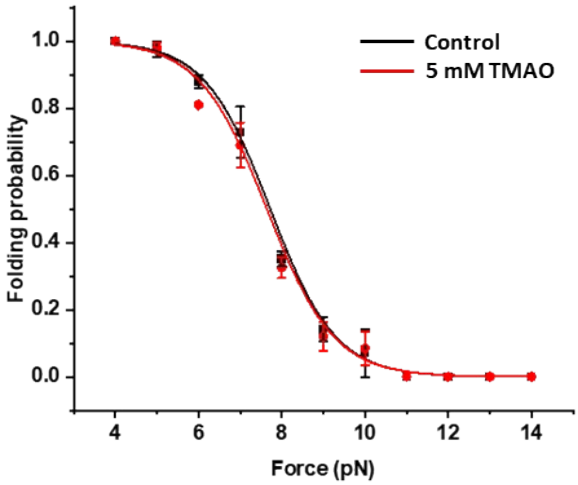
MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLFLHGNPTSSYVWRNIIPHVAP
THRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHDWGSALGFH
WAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGT
LPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDW
LHQSPVPKLLFWGTPGVLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEI
ARWLSTLEISGASEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKD
NGEWTVDVADKGYTLNIKFAGASEEVTIKANLIFANGSTQTAEFKGTFEKATSE
AYAYADTLKKDNGEWTVDVADKGYTLNIKFAGASEEVTIKANLIFANGSTQTAE
FKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKFAGASEEVTIKAN
LIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKF
AGASEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLKKDNGEWTVDV
ADKGYTLNIKFAGASEEVTIKANLIFANGSTQTAEFKGTFEKATSEAYAYADTLK
KDNGEWTVDVADKGYTLNIKFAGASEEVTIKANLIFANGSTQTAEFKGTFEKAT
SEAYAYADTLKKDNGEWTVDVADKGYTLNIKFAGASEEVTIKANLIFANGSTQT
AEFKGTFEKATSEAYAYADTLKKDNGEWTVDVADKGYTLNIKFAGHHHHHHG
GGLNDIFEAQKIEWHE



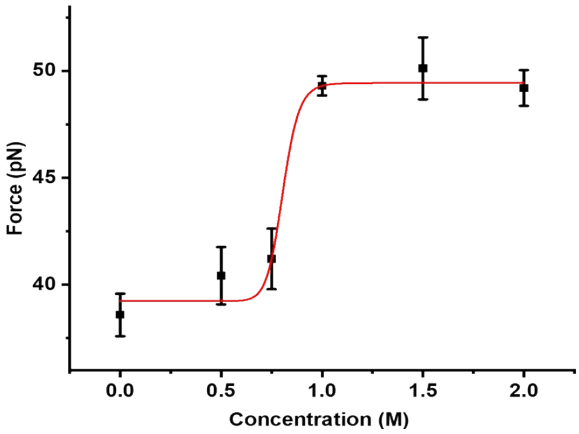
Supplementary Figure 1: Protein L unfolding obtained through force-ramp experiment: *Protein L unfolding was monitored by force-ramp technology by increasing the force from 4 to 80 pN at a loading rate of 2.53 pN/s. We observed eight unfolding steps unambiguously during the force-increase scan. However, during the force-decrease scan, refolding events seem to be faster and smoother due to the refolding of all domains within the same force range, and thus, no distinct refolding steps are observed.*



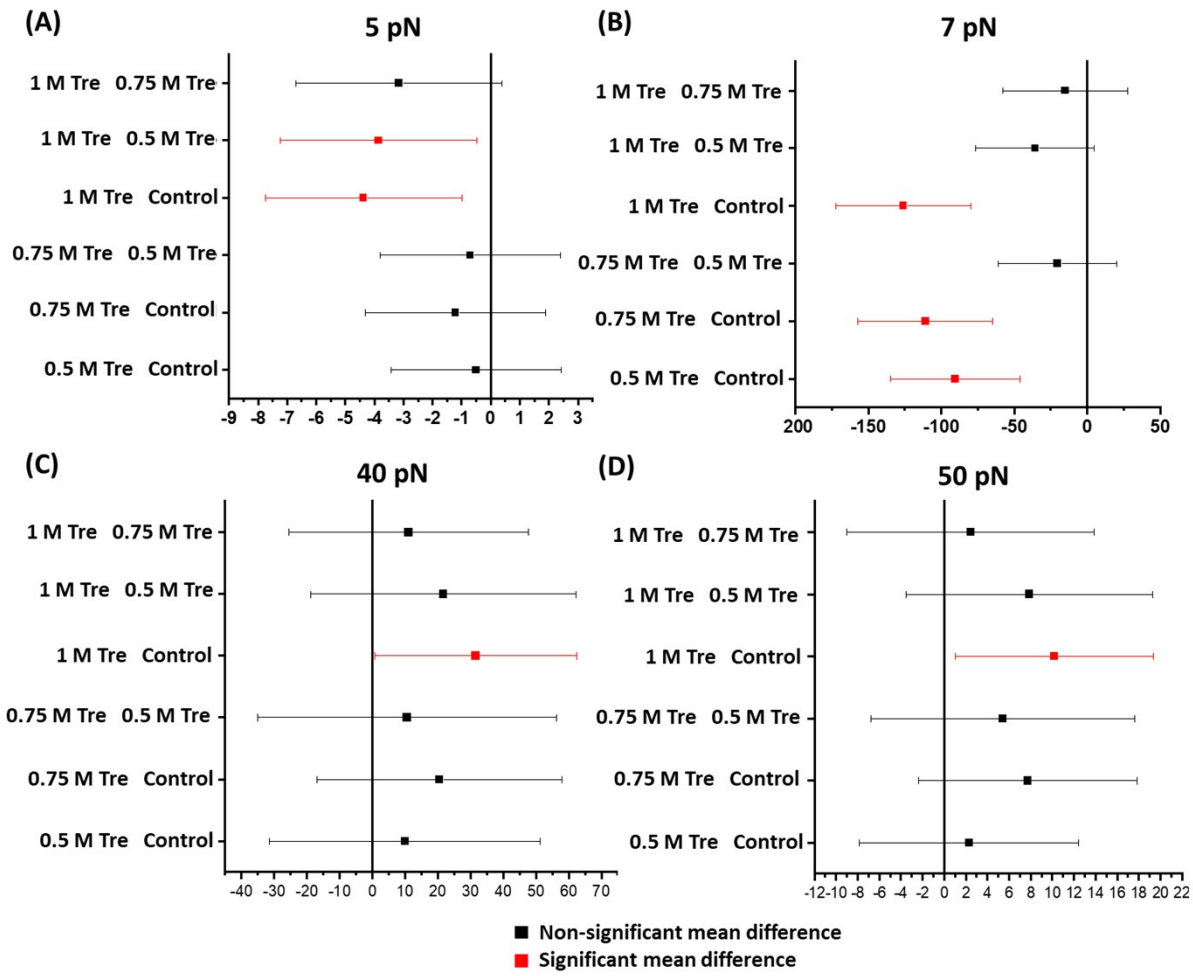
Supplementary Figure 2: Pair comparison analysis of the MFPT population at different TMAO concentrations: (A and B) *Bonferroni post hoc test ($^{*}p < 0.05$) was performed to analyze pairwise comparisons among refolding MFPT values (at 5.5 and 6.5 pN forces) with different TMAO concentration sets.* ***(C and D)*** *Similarly, unfolding MFPT pairs were checked with different TMAO concentrations at 40 and 50 pN forces. Although we found visible differences among these MFPT values during both refolding and unfolding, they are not significant in all cases.*



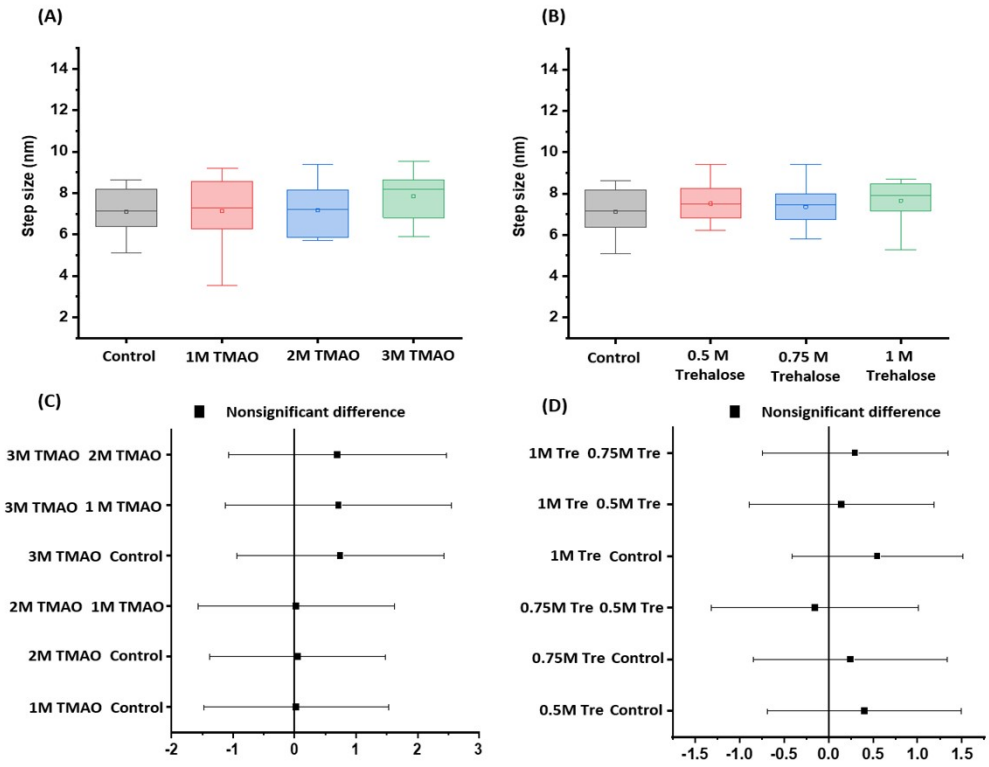
Supplementary Figure 3: Effect of 5 mM TMAO on protein L folding dynamics: *We observed that 5 mM TMAO does not affect the folding dynamics of protein L (red trace) and strongly overlaps with its absence (control, black trace).*



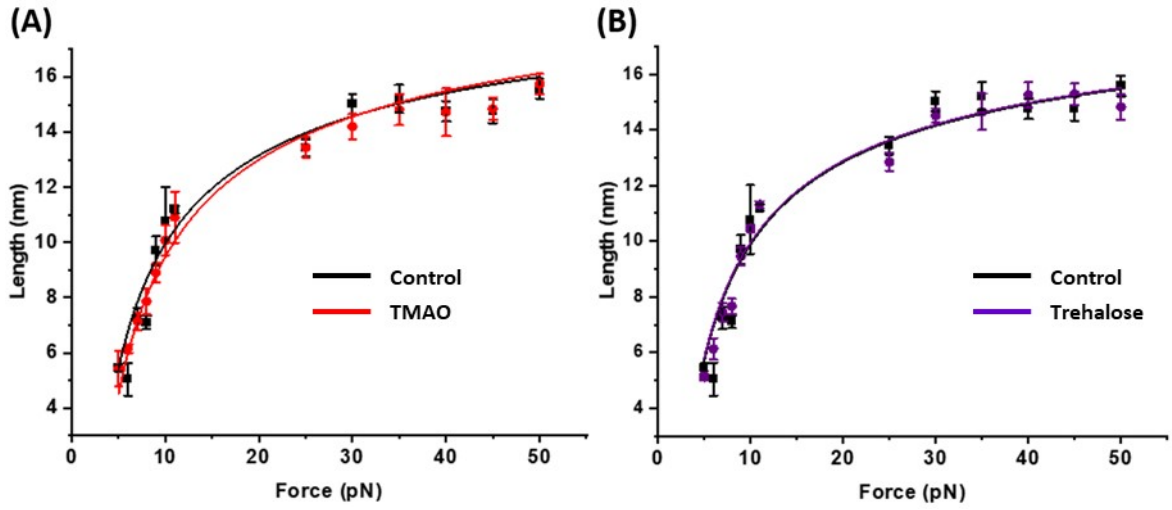
Supplementary Figure 4: Most probable unfolding force of protein L with trehalose concentration: *We plotted the most likely unfolding force as a function of increasing concentration of trehalose. Above 1 M, the unfolding force becomes flattened and reaches its maximum mechanical strength.*



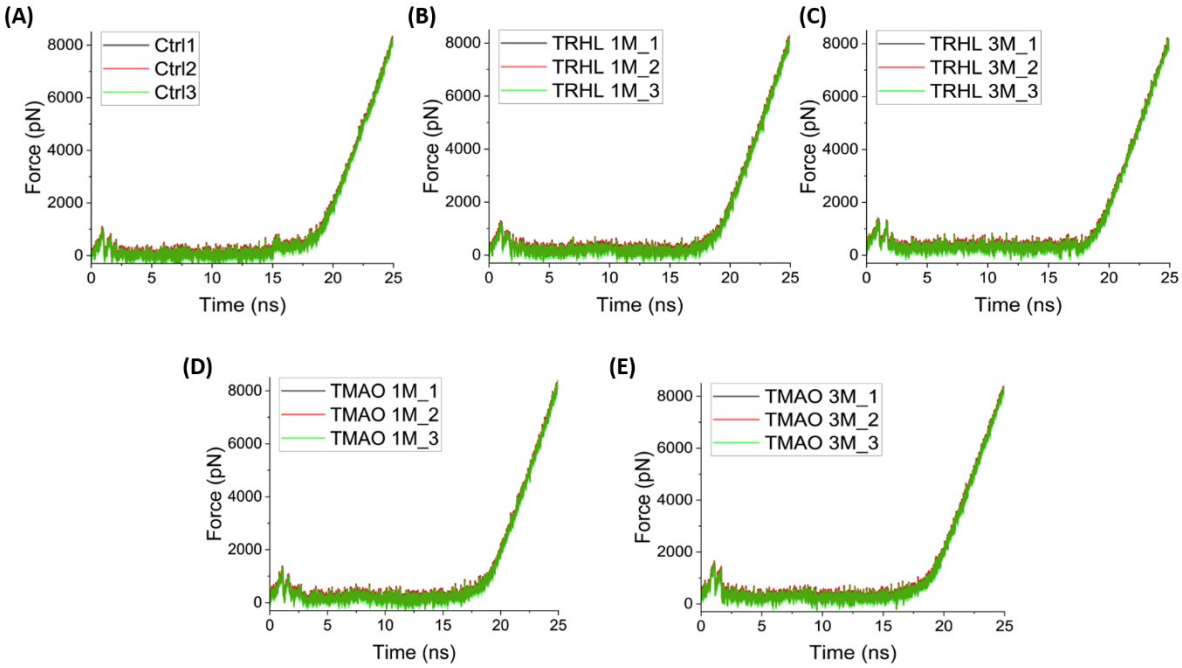
Supplementary Figure 5: Pair comparison analysis of the MFPT population at different trehalose concentrations: ***(A and B)*** *Bonferroni post hoc test ($^{*}p < 0.05$) was performed to analyze pair comparison among refolding MFPT values (at 5 and 7 pN forces) with different trehalose concentration sets.* ***(C and D)*** *Similarly, unfolding MFPT pairs were checked with different trehalose concentrations at 40 and 50 pN forces. Although we found visible differences among these MFPT values during both refolding and unfolding, 1 M trehalose and the control had only statistically significant differences.*



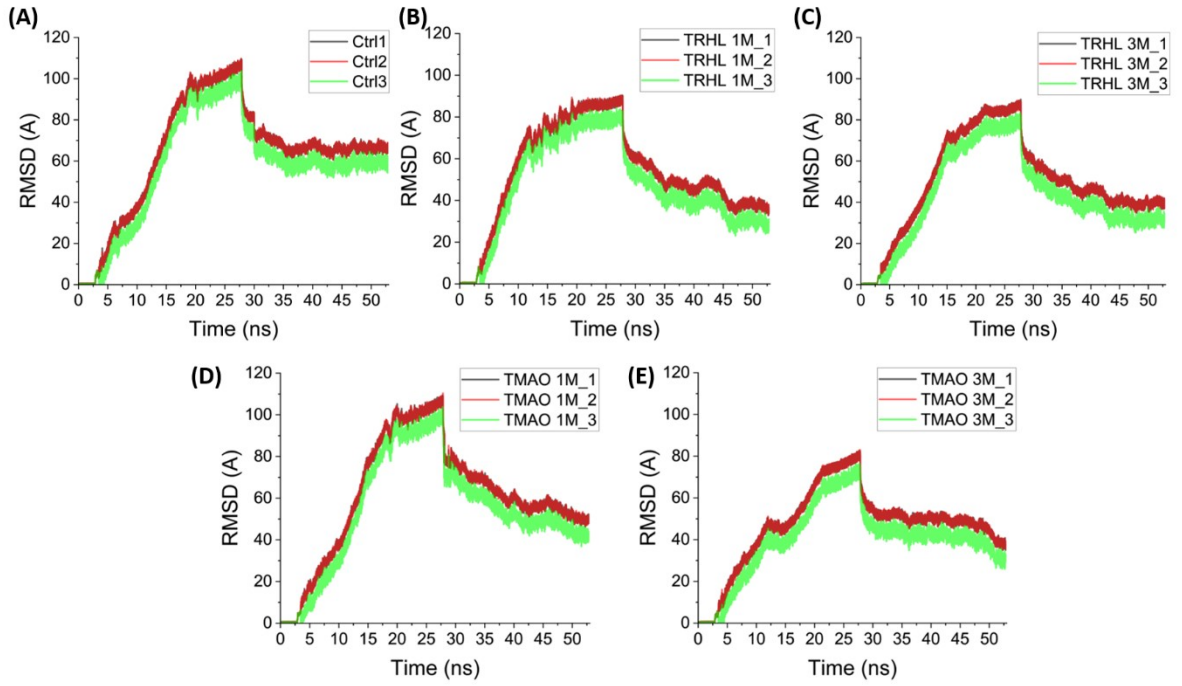
Supplementary Figure 6: Statistical analysis of the protein L extension with different osmolyte concentrations at 8 pN force: (A) TMAO: *We determined the extension both in the absence and presence of TMAO and observed that TMAO does not affect the protein L step size at 8 pN force.* ***(B) Trehalose:*** *Similar to TMAO, trehalose also does not change the protein L extension.* ***(C and D)*** *We performed one-way ANOVA to check the statistical significance of the step size with various concentrations of* ***(C)*** *TMAO and* ***(D)*** *trehalose at a particular refolding force of 8 pN. Then, we correlated the mean difference by Bonferroni post hoc (*P<0.05).*



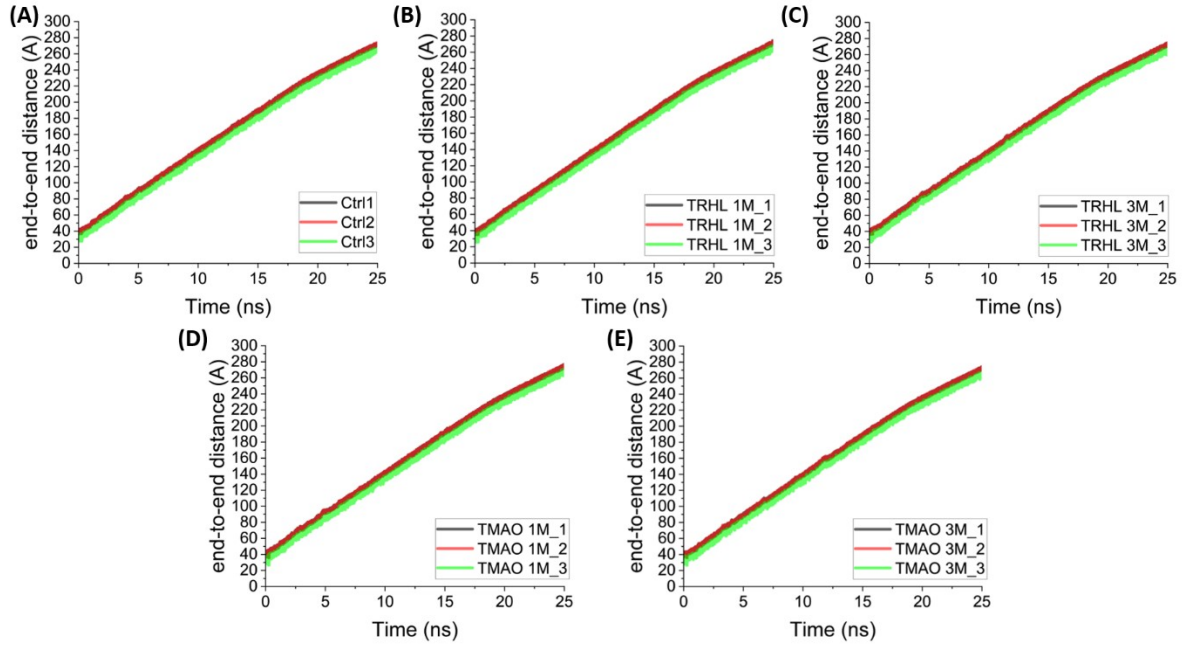
Supplementary Figure 7: Extension of protein L was measured in the presence and absence of osmolytes. *In the presence of both* ***(A)*** *TMAO and* ***(B)*** *trehalose, we observed that the unfolding extensions strongly overlapped with those in their absence (control, black) while fitting to the WLC model. Data were analyzed by taking n=>5 molecules. Error bars are the standard error of the mean.*



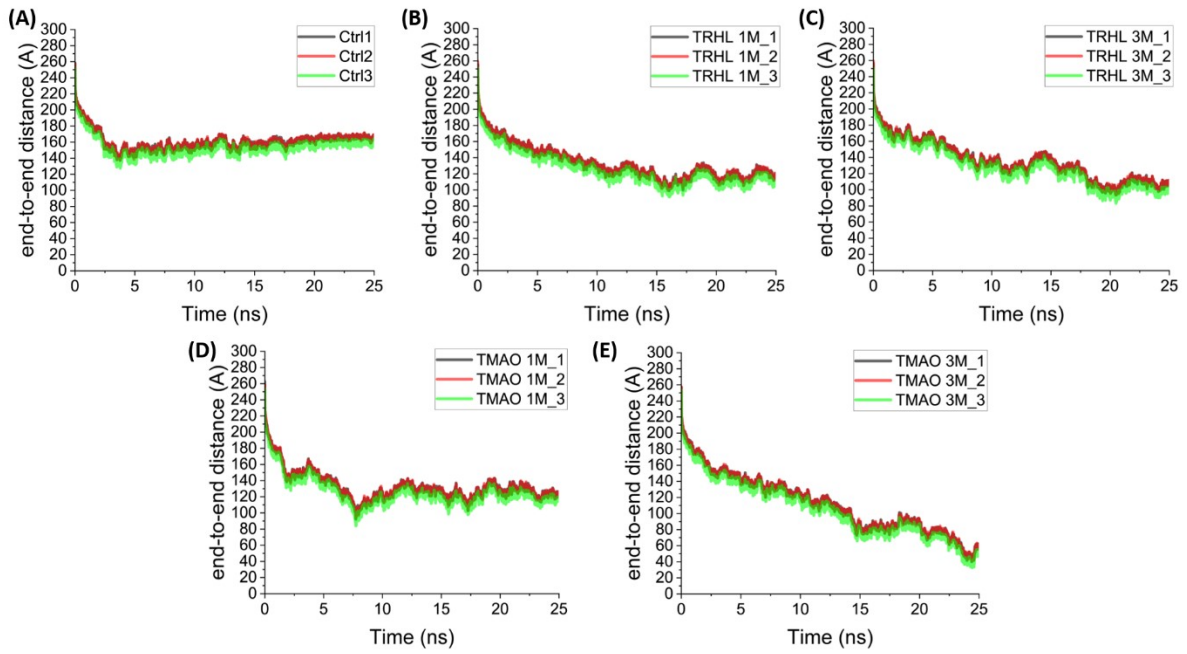
Supplementary Figure 8: *Traces of constant-velocity pulling force vs time triplicates for Protein L in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



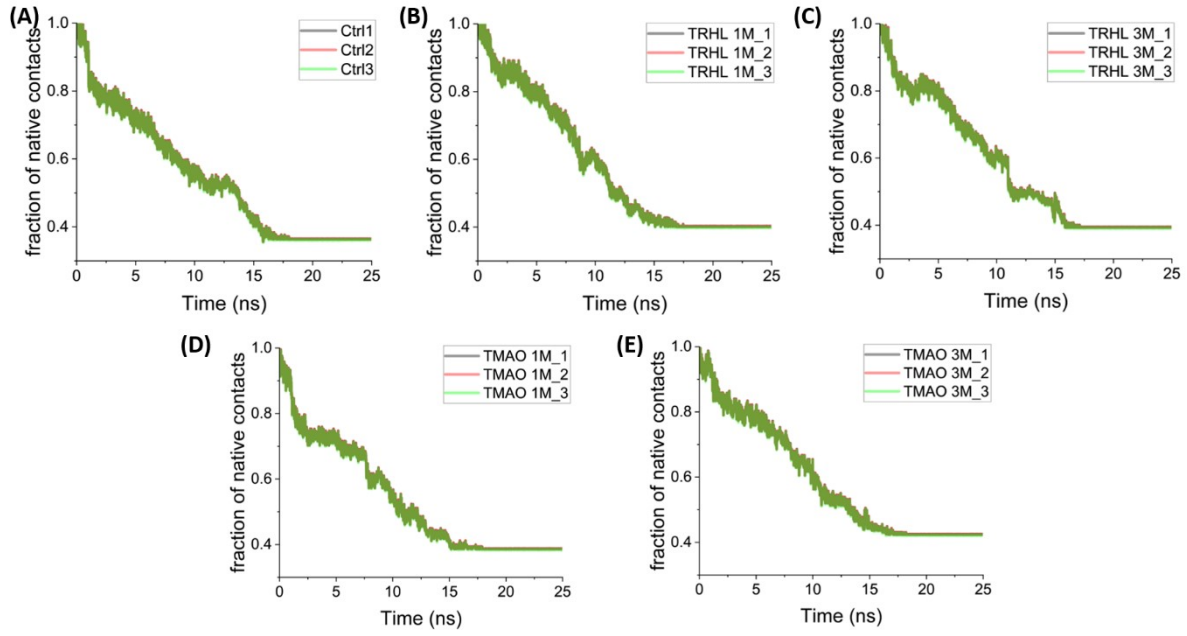
Supplementary Figure 9: *Traces of RMSD (root mean square deviations) vs time triplicates for Protein L in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



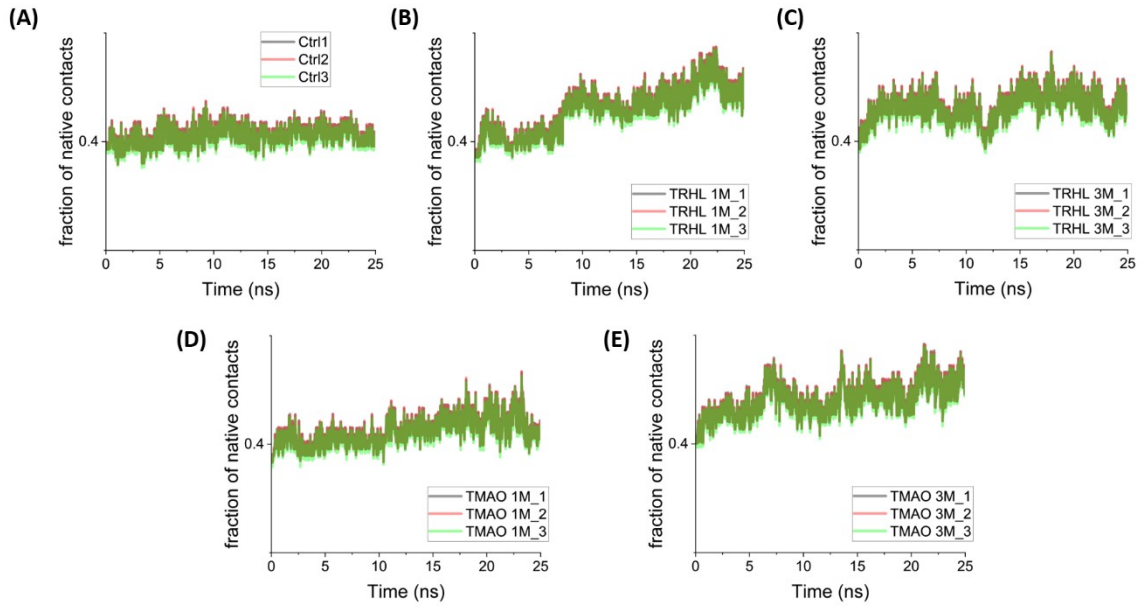
Supplementary Figure 10: *Traces of end-to-end distance vs time triplicates for Protein L unfolding in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



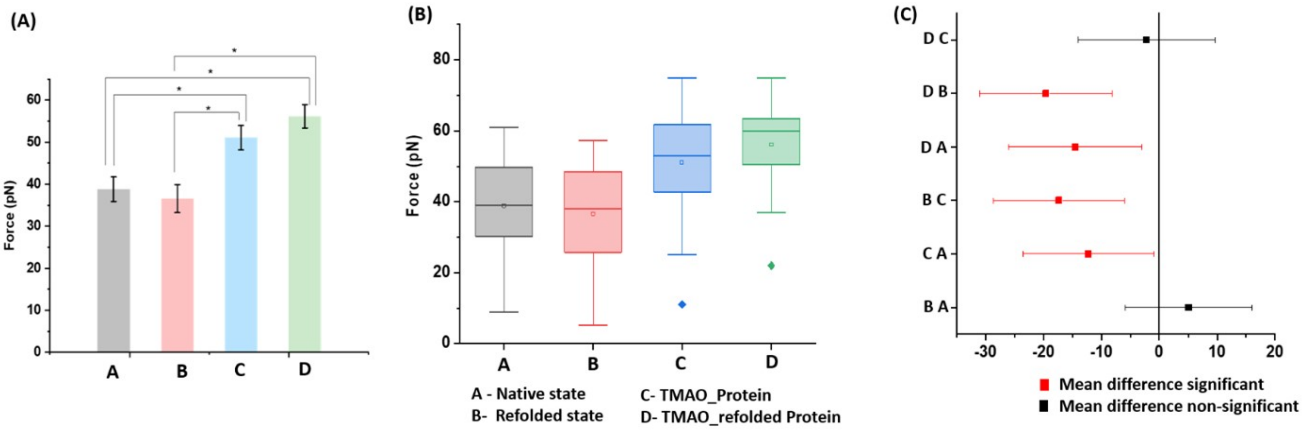
Supplementary Figure 11: *Traces of end-to-end distance vs time triplicates for Protein L relaxation in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



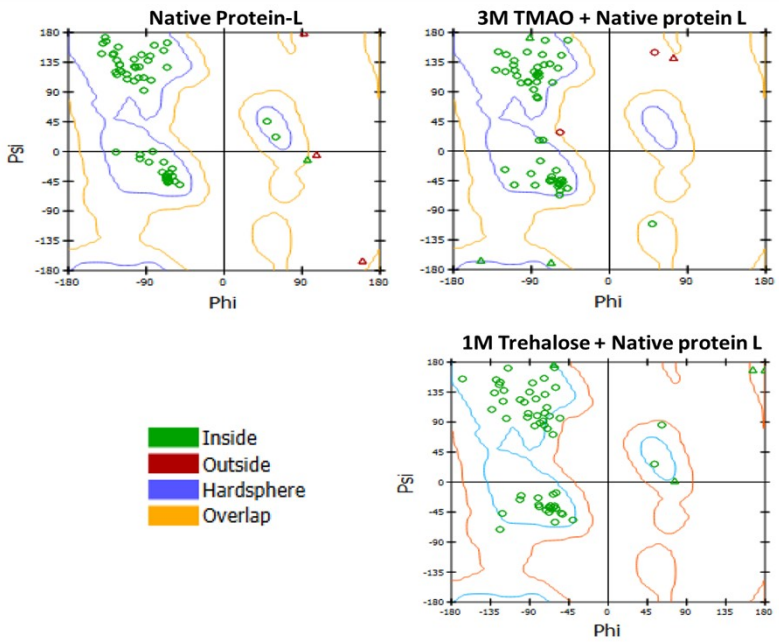
Supplementary Figure 12: *Traces of fraction of native contacts vs time triplicates for Protein L unfolding in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



Supplementary Figure 13: *Traces of fraction of native contacts vs time triplicates for Protein L relaxation in (A) Control (Ctrl), (B) 1 M trehalose (TRHL) (C) 3 M trehalose (D) 1 M TMAO (E) 3 M TMAO.*



Supplementary Figure 14: Mechanical stability of protein L: (A) *Comparison of the mechanical stability of native and refolded proteins both in the presence and absence of TMAO. Without TMAO, native and refolded protein L show unfolding forces of 38.8 ± 2.9 and 36.6 ± 3.3 pN with nonsignificant differences, whereas in the presence of 3 M TMAO, the native state and the refolded state require unfolding forces of 51.1 ± 2.9 pN and 56.2 ± 2.8 pN, respectively.* ***(B and C)*** *One-way ANOVA of protein L (native and refolded) unfolding force in the presence and absence of TMAO, as indicated in the figure. We performed one-way ANOVA of mechanical strength by measuring the unfolding force to check the statistical significance of variability in this property. We further performed a Bonferroni post hoc correction to check the significance of the mean difference in the various abovementioned conditions.*



***Supplementary Figure 15: Ramachandran plot analysis of native protein L**: TMAO destabilizes the left-handed helical region but stabilizes more of the right-handed helical region and beta sheet region, which overall stabilize the protein. Trehalose stabilizes the right-handed and left-handed helical region and stabilizes the beta sheet region.*

System	Box Type	Box Size (nm x nm x nm)	No. of TMAO molecules	No. of water molecules	Neutral. ions (NaCl)
Protein L	Cubic	10 x 10 x 10	0	2401	0.15 M
3 M TMAO + Protein L	Cubic	13 x 13 x 13	293	5078	0.15 M
1 M TMAO + Protein L	Cubic	13 x 13 x 13	91	5156	0.15 M
3 M trehalose + Protein L	Cubic	10 x 10 x 10	122	2259	0.15 M
1 M trehalose + Protein L	Cubic	10 x 10 x 10	41	2378	0.15 M

***Supplementary Table 1: Details of solvation and preparation of simulation systems for all the simulations performed**.*

References:

1 O. Jw, K. De, B. D and Z. Ky, Structures of the B1 Domain of Protein L From Peptostreptococcus Magnus With a Tyrosine to Tryptophan Substitution, https://pubmed.ncbi.nlm.nih.gov/11264576/, (accessed May 20, 2020).

2 R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, W537–W541.

3 J. Huang and A. D. MacKerell Jr, *J. Comput. Chem.*, 2013, **34**, 2135–2145.

4 P. Ganguly, J. Polák, N. F. A. van der Vegt, J. Heyda and J.-E. Shea, *J. Phys. Chem. B*, 2020, **124**, 6181–6197.

5 P. T. Kiss and A. Baranyai, *J. Chem. Phys.*, 2011, **134**, 054106.

6 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

7 W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, 1977, **34**, 1311–1327.

8 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

9 M. Parrinello and A. Rahman, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

10 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38, 27–28.

11 M. Kayikci, A. J. Venkatakrishnan, J. Scott-Brown, C. N. J. Ravarani, T. Flock and M. M. Babu, *Nat. Struct. Mol. Biol.*, 2018, **25**, 185–194.