Biophysical Journal, Volume *110*

Supplemental Information

Single-Molecule Chemo-Mechanical Spectroscopy Provides Structural Identity of Folding Intermediates

Hesam N. Motlagh, Dmitri Toptygin, Christian M. Kaiser, and Vincent J. Hilser

Supplementary Text for

Single-molecule chemo-mechanical spectroscopy provides structural resolution of protein folding intermediates

Hesam N. Motlagh*[a], Dmitri Toptigyn[b], Christian M. Kaiser[b], and Vincent J. Hilser*[a,b]

[a]T.C. Jenkins Department of Biophysics, [b]Department of Biology,

The Johns Hopkins University, Baltimore, Maryland 21218

*Correspondence: VJH (Hilser@jhu.edu) or HNM (Hnekoor1@jhu.edu)

Tel: 410-516-6072; Fax: 410-516-5213

Supplementary Text S1 – Unfolding rates are marginally decelerated in the presence of osmolytes

The mean unfolding forces ($<F_{unf}>$) obtained in the presence and absence of osmolytes overlap significantly: $F_{unf,buffer}$ = 17.6±1.9pN (N=224), $F_{unf,1M\ Sorbitol}$ = 18.2±2.1pN (N=288), and $F_{unf,1M\ TMAO}$ = 18.2±2.2pN (N=340) which initially suggests no significant change in the unfolding rates. To quantitatively evaluate the unfolding, several force-ramp data sets were collected and analyzed using a model that converts the force rupture distribution to an intrinsic lifetime and distance to transition state (1). The model reveals that the distance to the transition state does not change appreciably in the presence of either osmolyte: $\Delta x^{\ddagger}_{Buffer}$ = 2.7±0.1nm, $\Delta x^{\ddagger}_{1M\ Sorbitol}$ = 2.8±0.1nm, and $\Delta x^{\ddagger}_{1M\ TMAO}$ = 2.3±0.1nm. Although the value for TMAO differs somewhat from the sorbitol and buffer values, this difference is likely an artifact of the model being sensitive to the shape of the distribution. Indeed, constant-force experiments in the main text reveal that $\Delta x^{\ddagger}$ is not significantly different under the three different conditions, consistent with previously reported findings (2). Taken together, these results suggest that neither TMAO nor sorbitol change the unfolding pathway of T4*.

We note that the distance to the transitions state determined in our experiments, applying force to the termini of T4*, is unusually large. Most globular proteins exhibit distances to the transition state of less than 1 nm, reflecting the brittle nature of stably folded proteins (3). Native T4* has a radius of gyration (R_g) of ~2nm (4). The $\Delta x^{\ddagger}$ values determined here and previously (5) suggest that the molecule can be extended by approximately this length before crossing the barrier to unfolding. We believe that the origin of the large $\Delta x^{\ddagger}$ values is likely the unstable N-terminal A-helix region (6) that may deform easily under mechanical load before the barrier to unfolding is crossed (Fig. 1A). Regardless of the origin of the large absolute values, $\Delta x^{\ddagger}$ appears to be the same in all cases, indicating that osmolytes do not appreciably affect the position of the barrier to unfolding.

The unfolding force distributions, analyzed as described above (1), suggest a change in the folded state lifetimes extrapolated to zero force (τ_o) when $\Delta x^{\ddagger}$ is fixed at 2.7nm. $\tau_{o,Buffer}$ = 46,101±188s, $\tau_{o,1M\ Sorbitol}$ = 69,916±261s, and $\tau_{o,1M\ TMAO}$ = 71,698±562s. Determining τ_o required extrapolation over a relatively large force range and the exclusion of transition state sliding, imparting some uncertainty onto the determined values. Nevertheless, given that osmolytes stabilize the native states of proteins, our results are not unexpected, as osmolytes have been shown to decrease unfolding rates both in bulk (7-9) and at the single-molecule level (2, 10-12). Taken together, the data are consistent with a slight decrease in the unfolding rate and no significant change in the pathway. These conclusions are supported by the constant force experiments conducted in the main text.

Supplementary Text S2 – Derivation of transfer free energy for intermediate states

The single-molecule folding traces yield direct access to the probability of the intermediate (P_I) relative to the unfolded state. All calculations below only consider the change in probability from Buffer to 1M TMAO. Since we have measured $P_{I,Buffer}$ and $P_{I,1M\ TMAO}$, we can calculate the free energy change of the chemo-mechanical perturbation using a Boltzmann distribution as follows:

$$\frac{P_{I,1M\,TMAO}}{P_{I,Buffer}} = \frac{\exp\left(\frac{\Delta G_{I,1M\,TMAO}}{R*T}\right)}{\exp\left(\frac{\Delta G_{I,Buffer}}{R*T}\right)} = \exp\left(\frac{-\left[\Delta G_{I,1M\,TMAO} - \Delta G_{I,Buffer}\right]}{R*T}\right) \qquad \text{(Eq. S1)}$$

where $\Delta G_{I,1M\ TMAO}$ and $\Delta G_{I,Buffer}$ are the free energies of the intermediate state relative to the unfolded state (i.e. reference state), R is the universal gas constant, and T is the absolute

temperature. Implicit in this formalism is the assumption of equilibrium. We consider this a justified assumption given the low force regime (i.e. near equilibrium) and the subsequent predictive capabilities of the model generate. Should the assumption of equilibrium be incorrect, the predictive capabilities would be compromised.

The value in the numerator on the right side of Eq. S1 can be re-written as a function of the transfer free energy of the intermediate and unfolded states:

$$\Delta G_{I,1M\ TMAO} - \Delta G_{I,Buffer} = G_{I,1M\ TMAO} - G_{U,1M\ TMAO} - G_{I,Buffer} + G_{U,Buffer} = \Delta G_{I,tr} - \Delta G_{U,tr} \quad \text{(Eq. S2)}$$

Where $\Delta G_{I,tr}$ and $\Delta G_{U,tr}$ are the transfer free energies of the intermediate and unfolded state to 1M TMAO respectively. Both of these transfer free energies depend on the accessible surface area (ASA) of the states in question. As mentioned in the main text, we treat the intermediate state as a contiguously folded portion of the protein that has the same ASA as the crystallographic structure. This is justified since the intermediate is on-pathway and is presumably native-like.

Let us consider a general intermediate within the context of the derivation above: the intermediate has a contiguously folded portion of amino acids N_{tr} through C_{tr}, where these values are integers corresponding to the amino acid numbers that are the boundaries for the folded portion of the molecule (Note: $1 \leq N_{tr} < C_{tr} \leq 164$). Several immediate relationships become apparent from this formalism. For instance, the unfolded portion of the molecule is amino acids 1 through N_{tr-1} and C_{tr+1} through 164 (the number of amino acids in T4*). Since these residues are also unfolded in the unfolded state their transfer free energies in Eq. S2 cancels out. The transfer free energy in Eq. S2 actually corresponds to the transfer free energy of the folded portion (i.e. amino acids N_{tr} through C_{tr}). This can be appreciated by calculating the free energy of both states:

$$\Delta G_{I,tr} = \sum_{i=1}^{N_{tr}-1}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) + \sum_{j=C_{tr}+1}^{164}(\Delta\alpha_j^{SC}\Delta g_{tr,j,SC}^{o} + \Delta\alpha_j^{BB}\Delta g_{tr,j,BB}^{o}) \quad \text{(Eq. S3)}$$

and

$$\Delta G_{U,tr} = \sum_{i=1}^{164}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) \quad \text{(Eq. S4)}$$

where i is the amino acid type from the primary sequence, Δg^{o}_{tr} is the free energy of transfer for the side-chain (SC) or backbone (BB) to 1M TMAO, and $\Delta\alpha_i$ is the fractional change in solvent ASA from the unfolded to the intermediate state (13, 14).

The $\Delta\alpha_i$ values in Eqs. S3 and S4 require estimates of the ASA of the unfolded and intermediate states:

$$\Delta\alpha_i^{bb\ or\ sc} = \frac{\sum_{j=1}^{n_i}(ASA_{i,j,U} - ASA_{i,j,I})}{n_i ASA_{i,Gly\cdot X\cdot Gly}} \quad \text{(Eq. S5)}$$

Where the numerator is summed over all amino acids j of type i, $ASA_{i,jU}$ is the ASA of the unfolded state, $ASA_{i,j,I}$ is the ASA of the intermediate state, n_i is the total number of groups of amino acid (AA) type i, and $ASA_{i,Gly.X.Gly}$ is the standard side-chain or backbone solvent accessibility of Gly-X-Gly tripeptides presenting the maximally exposed surface area (X it the amino acid of type i) (15). The values for α are calculated based on the steered molecular dynamics simulations to represent the unfolded state (see SI text 3 below).

Substituting Eqs. S3 and S4 into Eq. S2 results in a canceling out of terms such that now Eq. S2 reduces to:

$$\Delta G_{I,1M\ TMAO} - \Delta G_{I,Buffer} = -\sum_{N_{tr}}^{C_{tr}}(\Delta\alpha_i^{SC}\Delta g_{tr,i,SC}^{o} + \Delta\alpha_i^{BB}\Delta g_{tr,i,BB}^{o}) \quad \text{(Eq. S6)}$$

where all the terms are defined identically to Eqs. S3 and S4. The right hand side of Eq. S6 is actually a calculation of the transfer free energy of the folded portion of the intermediate to 1M TMAO. In particular, Eq. S6 is what was used to calculate the heat map in Figure 6B and to relate the experimental probability changes to the contour plot (i.e. by substituting Eq. S6 into Eq. S1). In all subsequent calculations, each amino acid was treated as either folded or not based on the intermediate boundaries defined by N_{tr} and C_{tr}. This approach was used for 1M Sorbitol as well which generated figure 6A.

References:

1. Dudko OK, Hummer G, & Szabo A (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc Natl Acad Sci U S A* 105(41):15755-15760.
2. Aioanei D, Brucale M, Tessari I, Bubacco L, & Samori B (2012) Worm-like Ising model for protein mechanical unfolding under the effect of osmolytes. *Biophysical journal* 102(2):342-350.
3. Jagannathan B, Elms PJ, Bustamante C, & Marqusee S (2012) Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proc Natl Acad Sci U S A* 109(44):17820-17825.
4. Ando N*, et al.* (2008) Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation. *Biochemistry* 47(42):11097-11109.
5. Kaiser CM, Goldman DH, Chodera JD, Tinoco I, & Bustamante C (2011) The Ribosome Modulates Nascent Protein Folding. *Science* 334(6063):1723-1727.
6. Hilser VJ, Townsend BD, & Freire E (1997) Structure-based statistical thermodynamic analysis of T4 lysozyme mutants: Structural mapping of cooperative interactions. *Biophys Chem* 64(1-3):69-79.
7. Russo AT, Rosgen J, & Bolen DW (2003) Osmolyte effects on kinetics of FKBP12 C22A folding coupled with prolyl isomerization. *Journal of molecular biology* 330(4):851-866.
8. Mukaiyama A, Koga Y, Takano K, & Kanaya S (2008) Osmolyte effect on the stability and folding of a hyperthermophilic protein. *Proteins* 71(1):110-118.
9. Pradeep L & Udgaonkar JB (2004) Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *The Journal of biological chemistry* 279(39):40303-40313.
10. Aioanei D*, et al.* (2011) Single-molecule-level evidence for the osmophobic effect. *Angewandte Chemie* 50(19):4394-4397.
11. Aioanei D, Tessari I, Bubacco L, Samori B, & Brucale M (2011) Observing the osmophobic effect in action at the single molecule level. *Proteins* 79(7):2214-2223.
12. Garcia-Manyes S, Dougan L, & Fernandez JM (2009) Osmolyte-induced separation of the mechanical folding phases of ubiquitin. *Proc Natl Acad Sci U S A* 106(26):10540-10545.
13. Auton M & Bolen DW (2007) Application of the Transfer Model to Understanding How Naturally Occuring Osmolytes Affect Protein Stability. *Methods in Enzymology* 428:397-418.
14. Auton M & Bolen DW (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proceedings of the National Academy of Sciences of the United States of America* 102(42):15065-15068.
15. Lee B & Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology* 55(3):379-400.

Supplementary Table S1 – Constant-force unfolding rate constants and statistics

Table 1 – Unfolding Kinetics

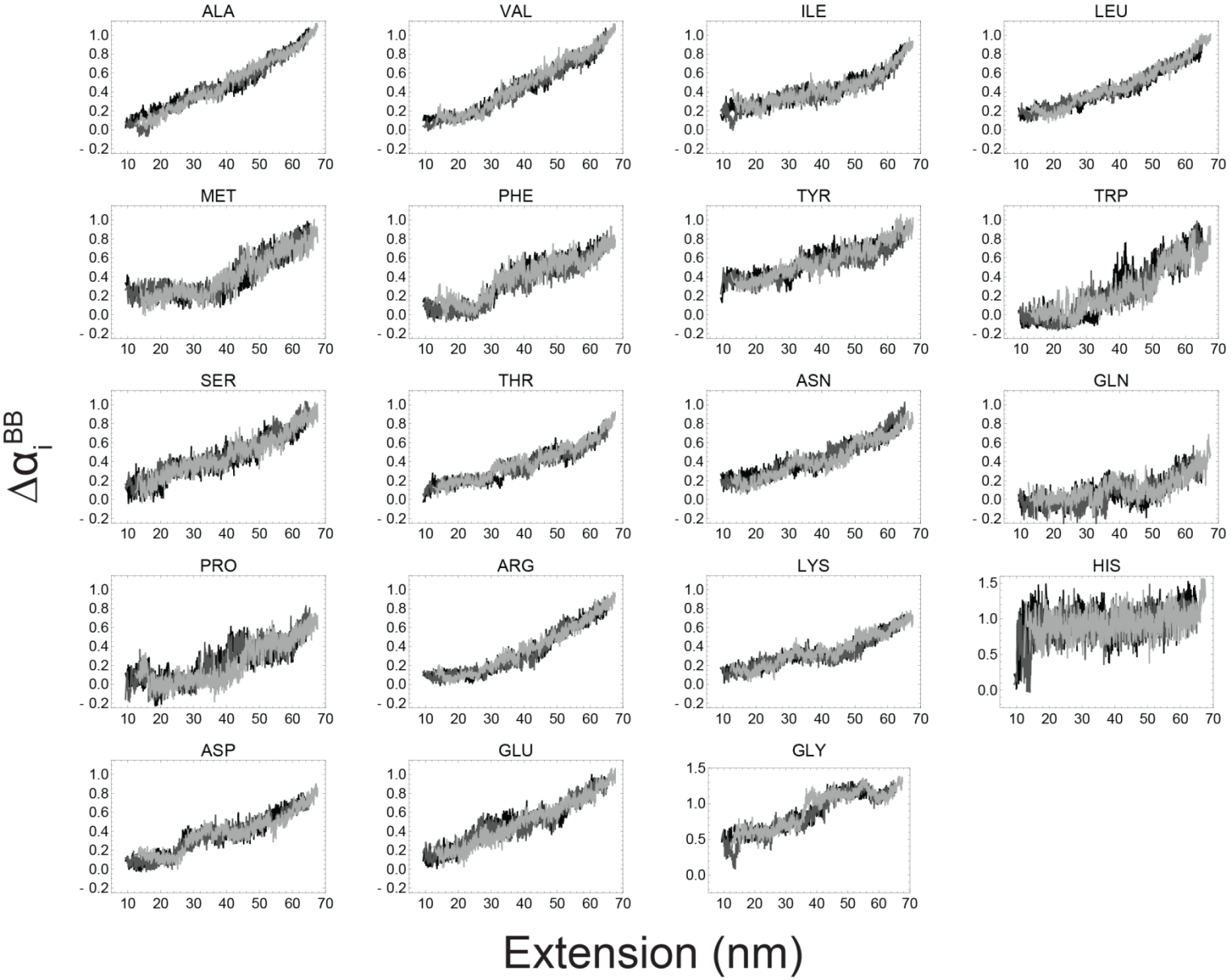
Condition	HKM		Sorbitol				TMAO			
Force (pN)	k_{app} [a]	N [b]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]
13	0.0854 ± 0.0008	67	N/A				N/A			
14	0.1781 ± 0.0025	100	0.1383 ± 0.0035	59	1.29	-0.25	0.0628 ± 0.0006	124	2.84	-1.04
15	0.4945 ± 0.0041	140	0.3527 ± 0.0067	73	1.40	-0.34	0.1796 ± 0.0010	186	2.75	-1.01
16	0.9898 ± 0.0068	138	0.7290 ± 0.0071	81	1.36	-0.31	0.5123 ± 0.0033	132	1.93	-0.66
17	1.8131 ± 0.0125	157	1.1872 ± 0.0132	79	1.53	-0.42	1.0446 ± 0.0150	131	1.74	-0.55
18	N/A		2.7291 ± 0.0965	21		N/A	1.1807 ± 0.0192	74		N/A
			Average		1.39	-0.33	Average		2.31	-0.82
			StdDev		0.10	0.07	StdDev		0.56	0.25

Supplementary Table S2 – Constant-force folding rate constants and statistics

Condition	HKM		Sorbitol				TMAO			
Force (pN)	k_{app} [a]	N	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]	k_{app} [a]	N [b]	Change [c]	ΔΔG(kb*T) [d]
3.5	6.6049 ± 0.1382	29	N/A				N/A			
4	3.0043 ± 0.0316	92	12.3369 ± 0.3832	59	4.11	1.41	N/A			
4.5	1.2467 ± 0.0263	154	3.5842 ± 0.0848	55	2.87	1.06	13.6907 ± 0.1844	106	10.98	2.40
5	0.8686 ± 0.0060	236	1.9445 ± 0.0533	112	2.24	0.81	10.4996 ± 0.1715	264	12.09	2.49
5.5	0.1177 ± 0.0062	60	0.2073 ± 0.0118	66	1.76	0.57	2.1226 ± 0.0468	84	18.04	2.89
6	N/A		N/A				0.7654 ± 0.0091	136		N/A
			Average		2.75	0.96	Average		13.70	2.59
			StdDev		1.02	0.36	StdDev		3.80	0.26

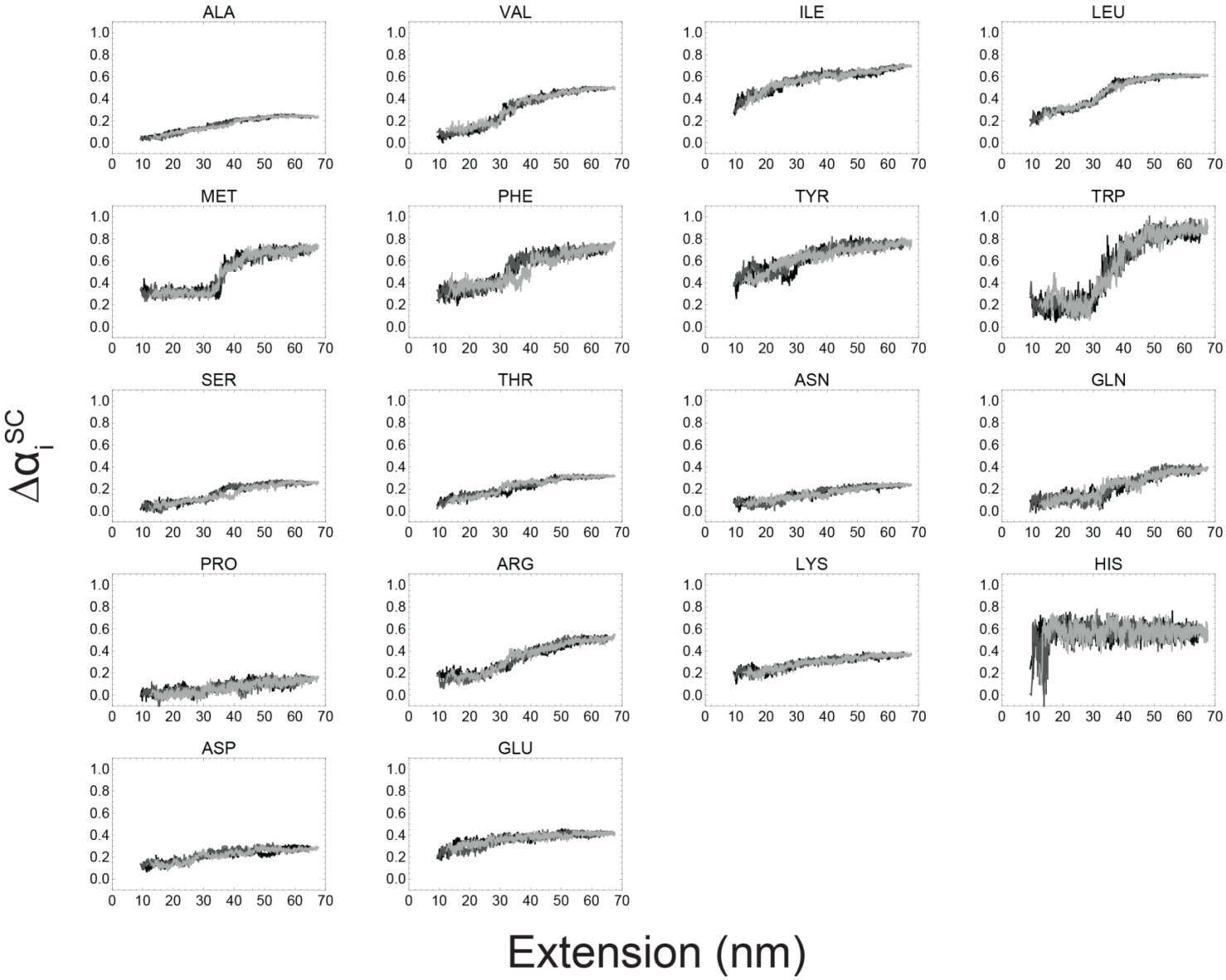
[a]The reported values for k_{app} are the apparent rates that describe the single-exponential lifetime distributions at each force for unfolding (Table 1) and folding (Table 2).
[b]N is the number of transitions observed at that force across all molecules. The approximate number of transitions from each molecule was approximately the same and thus the statistics are robustly determined.
[c]The reported change in the rate constant is calculated relative to HKM (Buffer) conditions.
[d]The change in activation free energy is calculated based on the relative change of the apparent rate constants.

Supplementary Figure S1 – Alpha values from steered molecular dynamics simulations



Supplementary Figure S1 – Shown are the calculated Δα values for all amino acids in T4* from the steered molecular dynamics simulations. These values are for the backbone (BB) and all three simulations are overlaid showing excellent reproducibility. Note that the majority of amino acids are simply a line between the initial Δα from the crystal structure to 1.0 when fully extended. Average values corresponding to the dimensions of the unfolded state from the BHMM were used in transfer free energy calculations of the intermediate (SI Text S2, and Figure 6A).

Supplementary Figure S2 – Alpha values from steered molecular dynamics simulations



Supplementary Figure S2 – Shown are the calculated Δα values for all amino acids in T4* from the steered molecular dynamics simulations. These values are for the side chains (SC) and all three simulations are overlaid showing excellent reproducibility. Note that the majority of amino acids are simply a line between the initial Δα from the crystal structure to the maximum accessibility when fully extended. Average values corresponding to the dimensions of the unfolded state from the BHMM were used in transfer free energy calculations of the intermediate (SI Text S2, and Figure 6A).