Supporting information

This document contains more detailed derivations of several key equations of the main manuscript and a comparison of the 1s and 2s approximations for two different polypeptides.

Detailed derivations of equations 9, 11, and 22

Equation 9: $\Delta G_{ij}(f)$ Free energy cost of forming an ij helical window relative to the entirely random-coil conformation when both are subject to fixed tension f.

Taking the natural logarithm of both sides of Equation 8, we get:

Equation 1-S

$$\Delta G_{ij}(f) = \Delta G_{ij}(f=0) - k_B T \, log \left\{ \frac{\int_{ij} dr \quad e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{ij} dr \quad e^{-\frac{u(r)}{k_B T}}} \right\} + k_B T \, log \left\{ \frac{\int_{rc} dr \quad e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{rc} dr \quad e^{-\frac{u(r)}{k_B T}}} \right\}$$

By defining the Gibbs free energy of the polypeptide containing the ij helical window as:

$$G_{ij}(f) \equiv -k_B T \, log \left\{ \int_{ij} dr \quad e^{-\frac{[u(r)-f\xi(r)]}{k_B T}} \right\} \qquad \text{Equation 2-S}$$

we have:

$$-k_B T \, log \left\{ \frac{\int_{ij} dr \quad e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{ij} dr \quad e^{-\frac{u(r)}{k_B T}}} \right\} = G_{ij}(f) - G_{ij}(f=0) = \int_0^f df' \left(\frac{\partial G_{ij}}{\partial f'} \right) \qquad \text{Equation 3-S}$$

Since the derivative of the Gibbs free energy with respect to force is the average extension:

$$\frac{\partial\, G_{ij}}{\partial f} = -\frac{\int_{ij}\, dr\, \xi(\boldsymbol{r})\;\; e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{ij}\, dr\;\; e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}} = -\langle\xi\rangle_{ij,f} \qquad \text{Equation 4-S}$$

we have:

$$-k_B T\, log\left\{\frac{\int_{ij}\, dr\;\; e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{ij}\, dr\;\; e^{-\frac{u(r)}{k_B T}}}\right\} = -\int_0^f df'\, \langle\xi\rangle_{ij,f'} \qquad \text{Equation 5-S}$$

Similarly for the polypeptide in the entirely random-coil conformation we have:

$$-k_B T\, log\left\{\frac{\int_{rc}\, dr\;\; e^{-\frac{[u(r)-f\xi(r)]}{k_B T}}}{\int_{rc}\, dr\;\; e^{-\frac{u(r)}{k_B T}}}\right\} = -\int_0^f df'\, \langle\xi\rangle_{rc,f'} \qquad \text{Equation 6-S}$$

Substituting Equations 5-S and 6-S into Equation 1-S we get Equation 9.

Equation 11: Average projection of a rigid rod on the direction of force

In Equation 11, $\left\langle \xi_{hel,j} \right\rangle_f$ is the average projection of a rigid rod of length $l_{hel,j}$ onto the direction of the applied force.

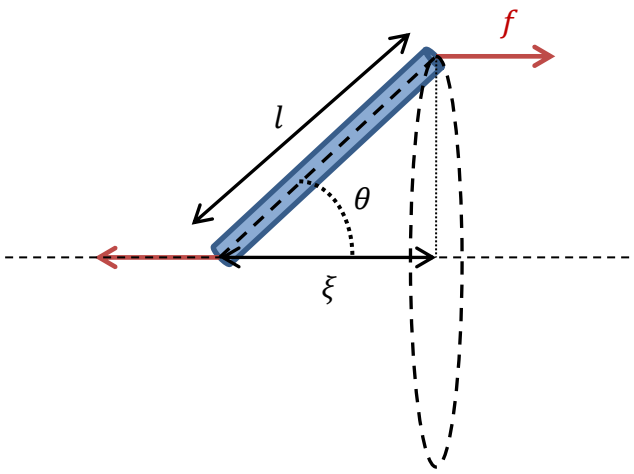

Figure 1-S schematic of a rigid rod under constant tension f.

$$\langle \xi \rangle_f = l \, \langle cos\theta \rangle$$

$$\langle cos\theta \rangle = \frac{\int_0^{\pi} d\theta \; cos\theta \; sin\theta \; e^{\frac{f \, l \, cos\theta}{k_B T}}}{\int_0^{\pi} d\theta \; sin\theta \; e^{\frac{f \, l \, cos\theta}{k_B T}}} = \frac{\int_{-1}^{1} du \; u \; e^{\frac{f \, l \, u}{k_B T}}}{\int_{-1}^{1} du \; e^{\frac{f \, l \, u}{k_B T}}} = \coth\left(\frac{f \, l}{k_B T}\right) - \left(\frac{f \, l}{k_B T}\right)^{-1}$$

Equation 22: $\Delta G_{ij}(L)$ Free energy cost of forming an *ij* helical window relative to the entirely random-coil conformation for the AFM cantilever displacement L .

Taking the natural logarithm of both sides of Equation 21, we get:

Equation 7-S

$$\Delta G_{ij}(L) =$$

$$\frac{k}{2}\left(L - \xi_{ij,L}\right)^2 - \frac{k}{2}\left(L - \xi_{rc,L}\right)^2 - k_B T \, log\left\{\int_{ij} dr \quad e^{-\frac{u(r)}{k_B T}} \, \delta\left(\xi(\boldsymbol{r}) - \xi_{ij}(L)\right)\right\} +$$

$$k_B T \, log\left\{\int_{rc} dr \quad e^{-\frac{u(r)}{k_B T}} \, \delta\left(\xi(\boldsymbol{r}) - \xi_{rc}(L)\right)\right\}$$

Defining the Helmholtz free energy of the molecule containing the ij helical window as:

$$A_{ij}(\xi) \equiv -k_B T \, log\left\{\int_{ij} dr \quad e^{-\frac{u(r)}{k_B T}} \, \delta(\xi(\boldsymbol{r}) - \xi)\right\} \qquad \text{Equation 8-S}$$

and assuming that for a given helical conformation ij the extension fluctuations are negligible at a given force (assuming the force-extension f vs. ξ curves are the same within a constant-force and constant-extension ensembles) we have:

$$A_{ij} = G_{ij} + \xi f \qquad \text{Equation 9-S}$$

where G_{ij} is the Gibbs free energy in the constant force ensemble as defined in Equation 2-S.

Using Equation 4-S we have:

$$A_{ij}(\xi) = G_{ij}\left(f(\xi)\right) + \xi \, f(\xi) = G_{ij}(f = 0) - \int_0^{f(\xi)} df' \, \langle \xi \rangle_{ij,f'} + \xi \, f(\xi) \qquad \text{Equation 10-S}$$

Similarly for the peptide in the entirely random-coil conformation:

$$A_{rc}(\xi) = G_{rc}\left(f(\xi)\right) + \xi \, f(\xi) = G_{rc}(f = 0) - \int_0^{f(\xi)} df' \, \langle \xi \rangle_{rc,f'} + \xi \, f(\xi) \qquad \text{Equation 11-S}$$

ΔG_{ij} is defined in AGADIR as:

$$\Delta G_{ij} = G_{ij}(f = 0) - G_{rc}(f = 0)$$ Equation 12-S

Substituting Equations 10-S to 12-S into Equation 7-S we arrive at Equation 22.

Force-extension prediction: 1-segment vs. 2-segments approximations

For polypeptides shorter than about 60 residues, the predictions of the single-segment (1s) approximation of AGADIR coincide with those of the multiple-segment (ms) model within 0.3% error[1]. This means that for most cases, ignoring the possibility of forming more than one helical segment does not introduce a substantial error. As explained in Section II.C, this is due to the large energy barrier associated with nucleating a new helical segment along the chain. To check the validity of the single-segment approximation for helical polypeptides under tension, we generated the force-extension data for a 60 residue long polypeptide $(AEAAKA)_{10}$ at 295K, pH=7.0 and ionic strength of 0.1M with both 1s and 2s approximations. Figure 2-S of the supporting information contains the results of average extensions along with the average helical contents as a function of tension using both the 1s and 2s approximations. As shown in this figure, the 1s and 2s results closely coincide with each other. This confirms the validity of the 1s approximation for force-extension predictions of $(AEAAKA)_{10}$. Meanwhile, the effect of protecting the polypeptide termini by acetylation and amidation has a considerable impact on average helical content. Figure 2-S of the supporting information shows that the protected chain has on average to about 10% more helical content than the unprotected one for the same tension.

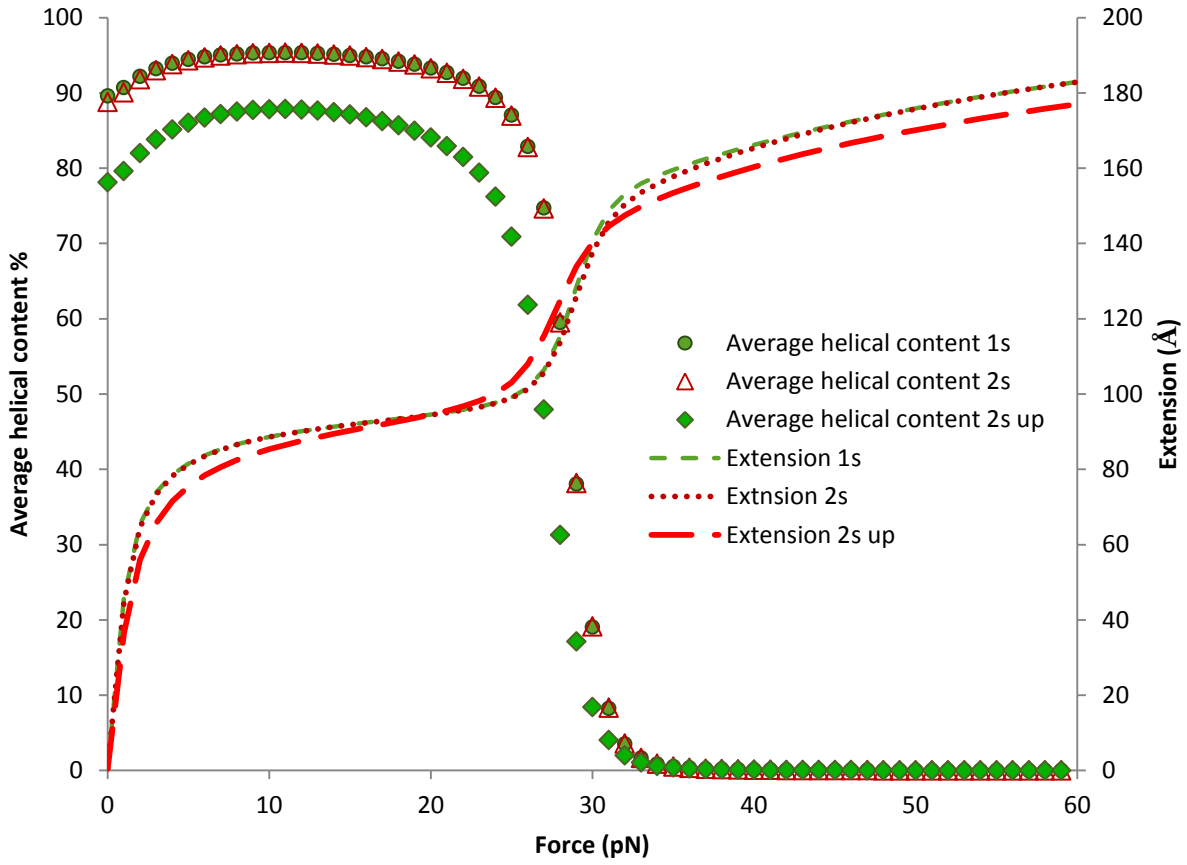

Figure 2-S Comparison of 1s and 2s approximations of average helical contents and force-extension curve predictions for $(AEAAKA)_{10}$ with both protected and unprotected (up) termini.

The 1s is a good approximation for relatively short polypeptides unless the chain is designed to form more than one separate helical segment. Figure 3-S of the supporting information presents force-extension data for a polypeptide of the following sequence: $G_7\,SA_{12}KRA_9\,G_8\,SA_{12}KRA_9$ at a temperature of 295K, pH of 7.0 and ionic strength of 0.1 M with acetylated N-terminus and amidated C-terminus. The two $SA_{12}KRA_9$ segments are the highly helical parts of the chain. Alanine (A) is very good helix former, Serine (S) is known as a good helix nucleating residue[2] and the electrically charged arginine (R) and lysine (K) are chosen for water solubility in anticipation in case future experimental investigations are done. The two helical segments are separated with eight Glycine (G) residues. Glycine is the only amino acid with no side-chain and therefore large freedom for its

dihedral angles. This makes it entropically unfavorable for glycine to lock into a helical conformation. Figure 3-S of the supporting information presents the force-extension predictions for the above polypeptide with both 1s and 2s approximation models. It is clear that for this polypeptide sequence, the 1s approximation leads to a large underestimation of the average helical content and consequently a less pronounced helix unraveling plateau within the force-extension curve.

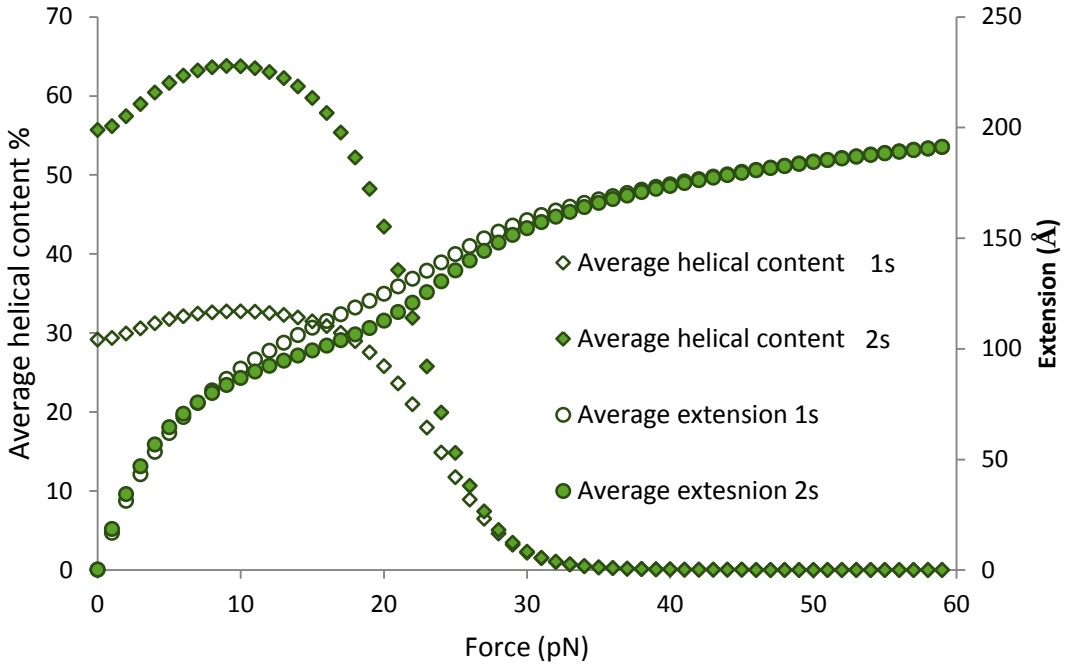

Figure 3-S Comparison of 1s and 2s model predictions of average helical content and force-extension curve of $G_7\,SA_{12}KRA_9\,G_8\,SA_{12}KRA_9$.

Figure 4-S and Figure 5-S of the supporting information present the residue level helicity and helical window probabilities for $G_7\,SA_{12}KRA_9\,G_8\,SA_{12}KRA_9$, respectively. As opposed to the $(AEAAKA)_{10}$ for which the 1s approximation does not result in any substantial change in the results, it is necessary to use the 2s approximation in order to accurately model $G_7\,SA_{12}KRA_9\,G_8\,SA_{12}KRA_9$.

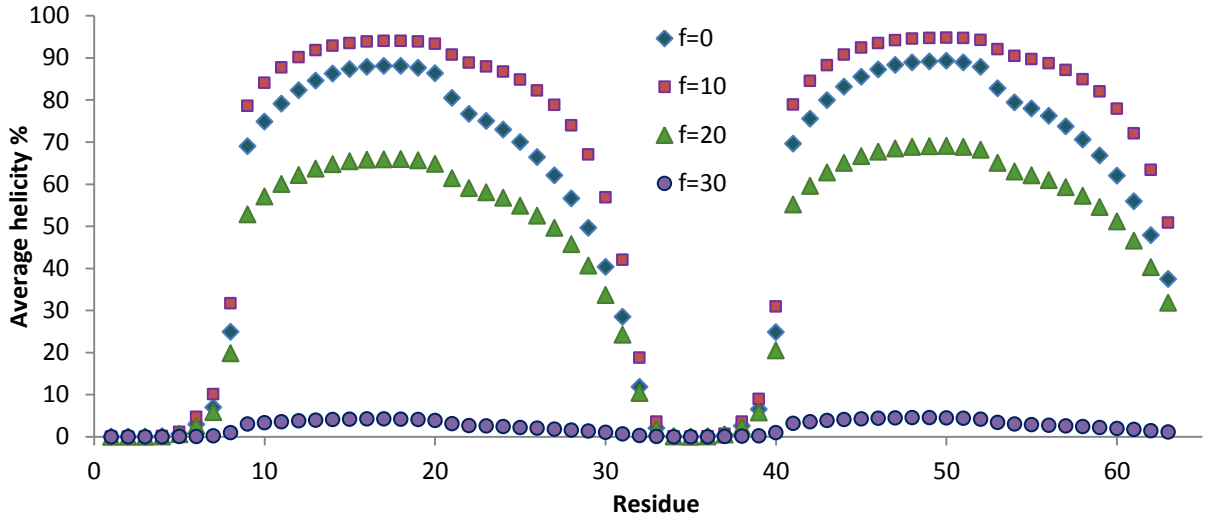

Figure 4-S residue level helicity prediction for G_7 $SA_{12}KRA_9$ G_8 $SA_{12}KRA_9$ for several different tensions.

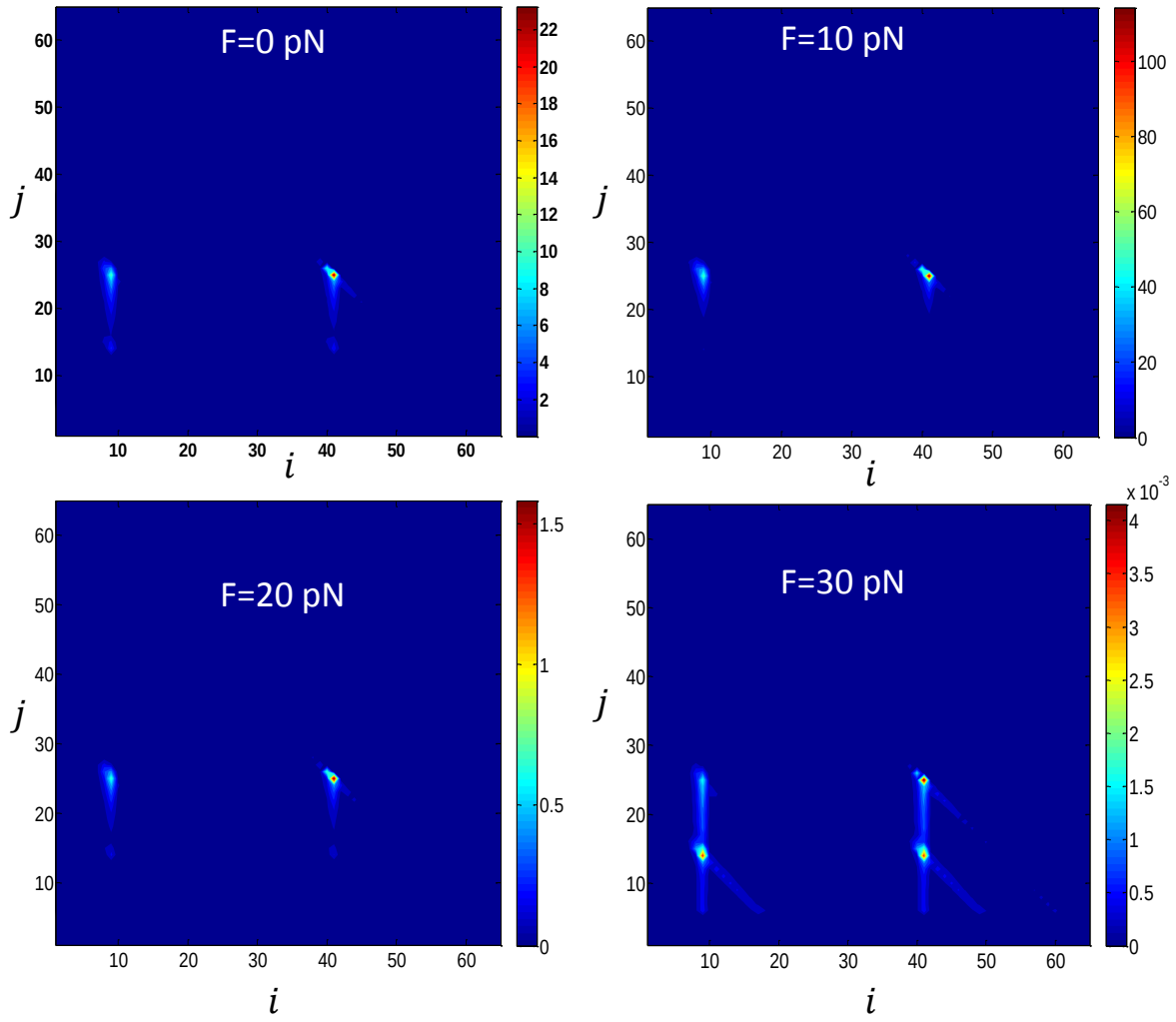

Figure 5-S Probability of all the possible helical windows K_{ij}'s within $G_7\,SA_{12}KRA_9\,G_8\,SA_{12}KRA_9$ under different tensions. K_{ij} is the probability of forming a j residue long helical window initiated at the i^{th} residue.

Bibliography

(1) Munoz, V.; Serrano, L. *Biopolymers* **1997**, *41*, 495–509.
(2) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173–191.