Supplementary Information

Exceptional mechanical stability of the spider silk C-terminal domain

Yuelong Xiao*, Senmiao Li, and Peng Zheng

State Key Laboratory of Coordination Chemistry, Chemistry and Biomedicine Innovation Center (ChemBIC), School of Chemistry and Chemical Engineering, Nanjing University, Nanjing, Jiangsu, 210023, China

*Corresponding author: Correspondence and requests for materials should be addressed to Yuelong Xiao (Email: yuelongxiao@smail.nju.edu.cn)

This Supplementary Information Includes:

Supplementary Methods

Protein engineering. The genes were synthesized by GenScript Inc. Most plasmids

were constructed in pQE-80L/pET-28a vector, and all proteins were expressed in *E. coli* BL21(DE3). *Oa*AEP1(C247A) refers to the cysteine 247 to alanine mutant of asparaginyl endoprotease 1 from *Oldenlandia affinis*, abbreviated as *Oa*AEP1. The expression and purification of *Oa*AEP1 follow the protocol from the referenced study[1]. ELP is an elastin-like polypeptide.[2] To create fusion proteins and and introduce a protease recognition site, a three-restriction enzyme system(*BamH* I-*Bgl* II-*Kpn* I) was employed. Site-directed mutagenesis was used to generate C-terminal domain (CTD) mutants.

After transformation, all proteins were expressed in *E. coli* BL21(DE3). Transformed *E. coli* was grown in 15 mL LB (Luria Bertani) medium at 37 °C for 16 h. After centrifugation to remove the supernatant, the pellet was transferred into 800 mL of LB medium and grown at 37 °C for 3 h. When OD_{600} reached approximately 0.6, protein expression was induced by adding 1 mM IPTG overnight at 18 °C to obtain the target protein. Bacteria were then harvested by centrifugation, and the supernatant was discarded. The cell pellets were resuspended in 40 mL lysis buffer (50 mM Tris, pH 7.4, 1 mM PMSF) with the addition of 0.1 mg DNase and 0.1 mg RNase. The suspension was then lysed under 900 bar pressure using a homogenizer with a 4 °C cooling water cycle. The lysate was centrifuged (18,000 rpm, 18 min), and the clarified lysate was mixed with Ni-NTA affinity beads (TALON) for 40 min. The beads were equilibrated with 50 mL washing buffer (50 mM Tris, pH 7.4, 2 mM imidazole, 400 mM NaCl). The bound protein was eluted using elution buffer (50 mM Tris, pH 7.4, 250 mM imidazole, 400 mM NaCl). Finally, the purified proteins were transferred to buffer (50 mM Tris, 100 mM NaCl, pH 7.4) through desalting column and stored at -80 ℃ with 20% glycerol.

Surface functionalization and protein immobilization. The AFM probes (MLCT-BIO, Bruker Corp.) and silicon-based glass coverslip were used for surface modification following the method described previously. Briefly, to introduce the NH2 group, the probes and glass coverslips were cleaned by plasma treatment and immersed in a 2% (v/v) APTES toluene solution for 1 h. Then NH_2-functionalized probes and glass coverslip were then reacted with a solution mixture of 2 mM $ImSO_2N_3$, 4 mM K_2CO_3, and 20 M $CuSO_4$ to convert the amino group to the N_3 group. After flushing, they were further incubated with the hetero-crosslinker DBCO-PEG_4-maleimide (4 mM in DMSO) at 37 °C for 1 h to add the maleimide group. Finally, the peptide GL-ELP_{20}-Cys/Cys-ELP_{20}-NGL was reacted onto the surface of glass coverslip and probes to serve as a spacer, preventing non-specific interactions during subsequent AFM measurements. The probes and glass coverslips were cleaned with water and stored at 4 °C for up to several weeks. For AFM-SMFS measurements, the target protein was pipetted onto the functionalized glass surface in a volume of 50 μL with approximately 1 μM *Oa*AEP1, while the probes were incubated with 60 μL of storage buffer containing 60 μM GL-GB1-XDoc and 3 μM *Oa*AEP1 for 30 min. After incubation, the surface was washed with AFM measurement buffer. The experiments were conducted at two different pH values using AFM buffer A (50 mM Tris, 100 mM NaCl, pH 7.4) or AFM buffer B (50 mM MES, 100 mM NaCl, pH 5.7).

AFM-SMFS experiment. Single-molecule AFM experiments were performed with the commercial JPK ForceRobot AFM. The D tip of the Si_3N_4 cantilever was used. The cantilevers were functionalized and immobilized with target proteins covalently as described above. After calibration, its accurate spring constant was determined by a thermally induced fluctuation method, the probe was contacted with the surface for a brief period (20~50 ms) at 300~500 pN, and the target protein was caught through the specific Coh/Xmod-Doc protein-protein interaction, then the probe retracted. As a result, a force-extension curve, possibly including the interaction/unfolding event, was obtained.

SMFS data analysis. The data were first filtered by JPK data processing and then analyzed by Igor Pro 6.12. The worm-like chain model (Eq. 1) was used to fit curves with a persistence length of ~0.4 nm.[3]

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right] \quad (1)$$

where *F(x)* is the force applied to the polymer (polypeptide chain) under a polymer extension *x*. *p* is the persistence length of the polymer. L_c is the contour length. k_B is the Boltzmann constant, and *T* is the temperature in kelvin.

Gaussian function (Eq. 2) was used to fit the histogram.

$$f(x) = W_0 + W_1 * e^{-\left(\frac{x - w_2}{w_3}\right)^2} \quad (2)$$

where W_0, W_1, W_2 and W_3 are arbitrary real constants ($W_3 \neq 0$).

Bell-Evans model was used to calculate loading rate data.

$$F = \frac{k_B T}{\Delta x}\ln\left(\frac{\Delta x}{k_{off} k_B T}\right) + \frac{k_B T}{\Delta x} ln(av)$$

where *F* is the unfolding force, k_B is the Boltzmann constant, *T* is the temperature in kelvin, Δx is the distance between the bonded state and the transition state, k_{off} is the unfolding rate constant at zero force, *a* is the slope of the force-extension curves immediately (~3 nm) before the rupture event and *v* is the velocity in the rupture event.

Supplementary Note

Protein sequence of Coh-(GB1)$_2$-**CTD**-NGL (MW: ~50 kDa):
MRGSHHHHHHRS**MGTALTDRGMTYDLDPKDGSSAATKPVLEVTKKVFD**
TAADAAGQTVTVEFKVSGAEGKYATTGYHIYWDERLEVVATKTGAYAK
KGAALEDSSLAKAENNGNGVFVASGADDDFGADGVMWTVELKVPADAK
AGDVYPIDVAYQWDPSKGDLFTDNKDSAQGKLMQAYFFTQGIKSSSNPS
TDEYLVKANATYADGYIAIKAGEPRSMDTYKLILNGKTLKGETTTEAVD
AATAEKVFKQYANDNGVDGEWTYDDATKTFTVTERSMDTYKLILNGKT
LKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTERS**G**
AASAAVSVGGYGPQSSSAPVASAAASRLSSPAASSRVSSAVSSLVSSGPTNQ
AALSNTISSVVSQVSASNPGLSGCDVLVQALLEVVSALVSILGSSSIGQINY
GASAQYTQMVGQSVAQALAGRSGSHHHHHHGSNGL

Protein sequence of GL-GB1-**Xmod**-Doc (MW: ~38 kDa)
MGLHHHHHHGSMDTYKLILNGKTLKGETTTEAVDAATAEKVFKQYAND
NGVDGEWTYDDATKTFTVTETSGSVVPSTQPVTTPPATTKPPATTIPPSDDP
NAVVPTSGGSRSGG**NTVTSAVKTQYVEIESVDGFYFNTEDKFDTAQIKKA**
VLHTVYNEGYTGDDGVAVVLREYESEPVDITAELTFGDATPANTYKAVE
NKFDYEIPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATAT
LTYYAATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELT
RFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDARS

Protein sequence of GL-**(ELP)$_{20}$**-Cys (MW: ~10 kDa):
MGLHHHHHHGS**VPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVP**
GGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGA
GVPGGGVPGGGVPGEGRSC

Protein sequence of Cys-**(ELP)$_{20}$**-NGL (MW: ~10 kDa):
MGHHHHHHCGS**VPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGAGVP**
GGGVPGGGVPGEGVPGEGVPGVGVPGVGVPGVGVPGVGVPGAGVPGA
GVPG GGVPGGGVPGEGRSNGL

Supplementary Figures

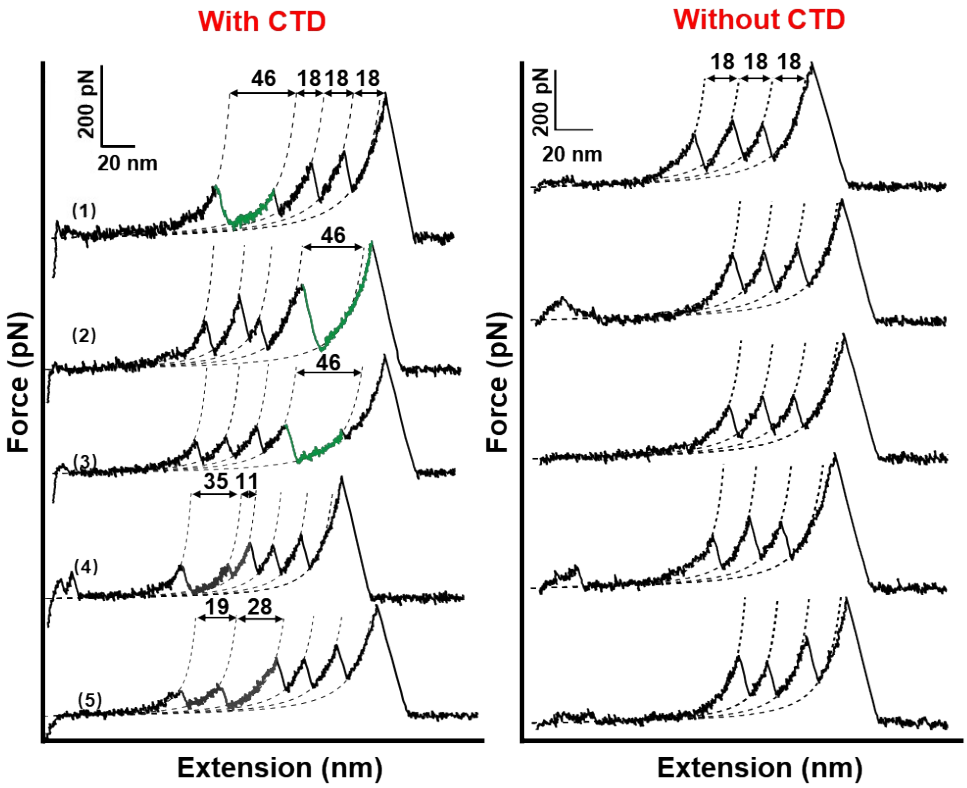

Fig. S1 Representative force-extension curves showing polyprotein unfolding events with (left) and without (right) the CTD.

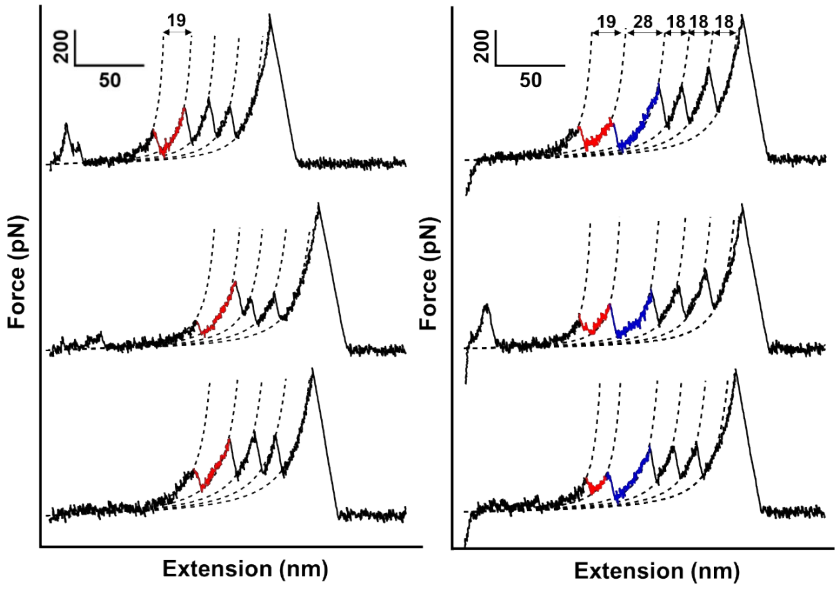

Fig. S2 Representative force-extension curves of the polyprotein unfolding containing the 19 nm unfolding event.

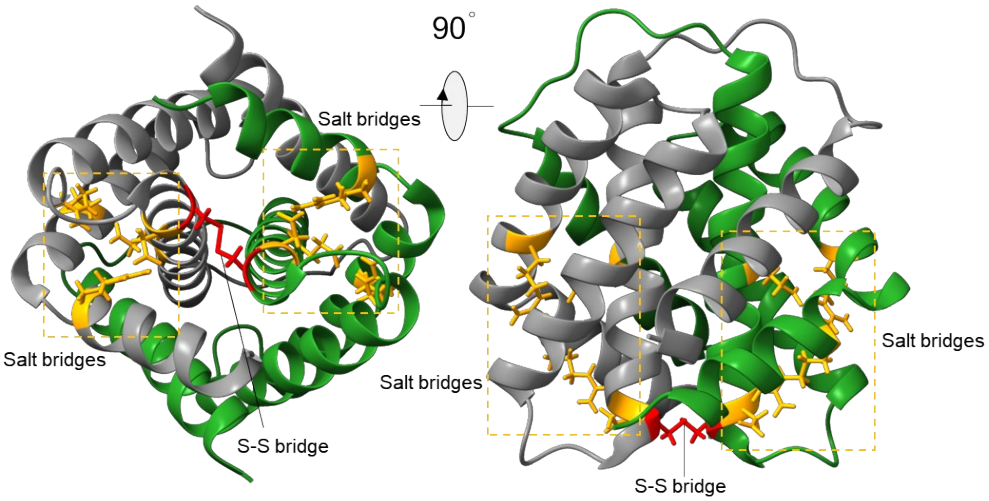

Fig. S3 3D structure of CTD with the salt bridges prominently highlighted.

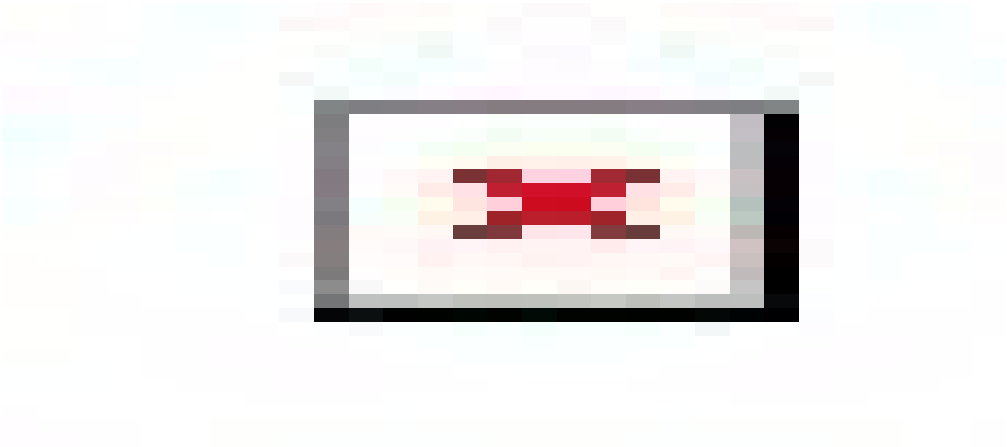

Fig. S4 SMD trajectories of stretching CTD.

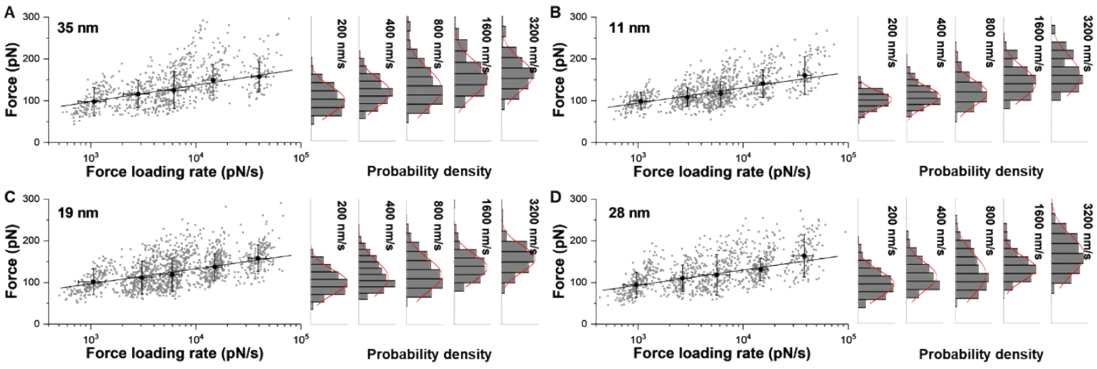

Fig. S5 The unfolding force of the CTD partial unfolding show a linear relationship with the logarithm of the loading rate. (A) 35 nm, (B) 11 nm, (C) 19 nm, (D) 28 nm.

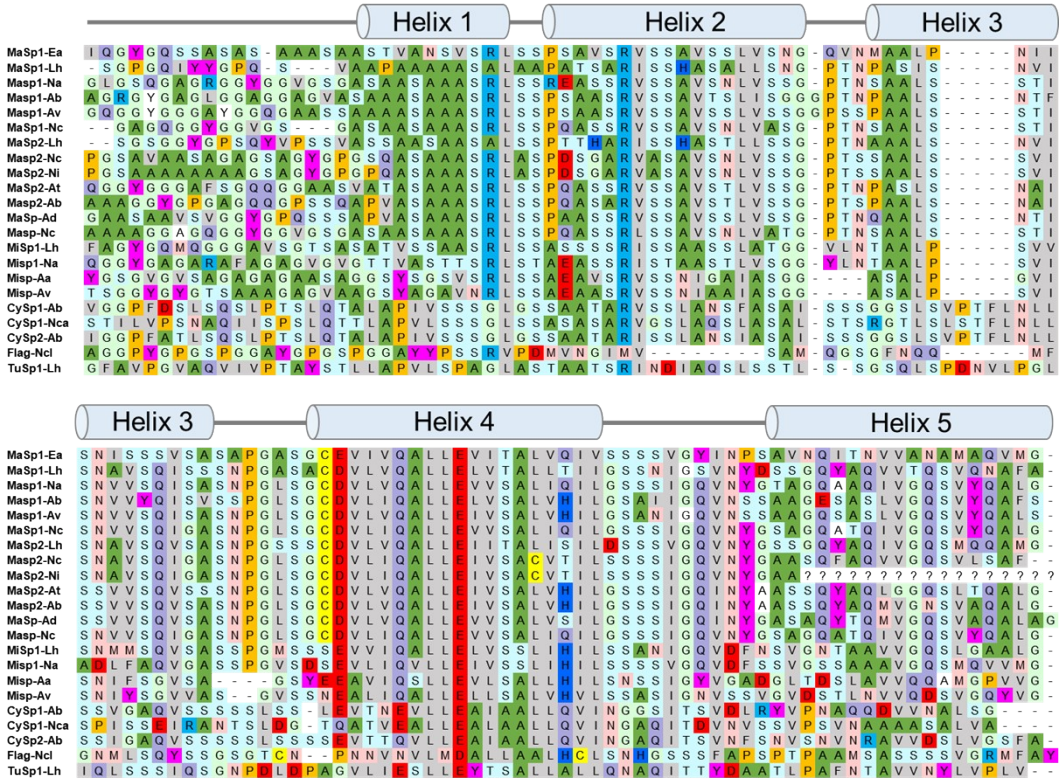

Fig. S6 Sequence alignment of CTDs from different spider species and various spidroin types.

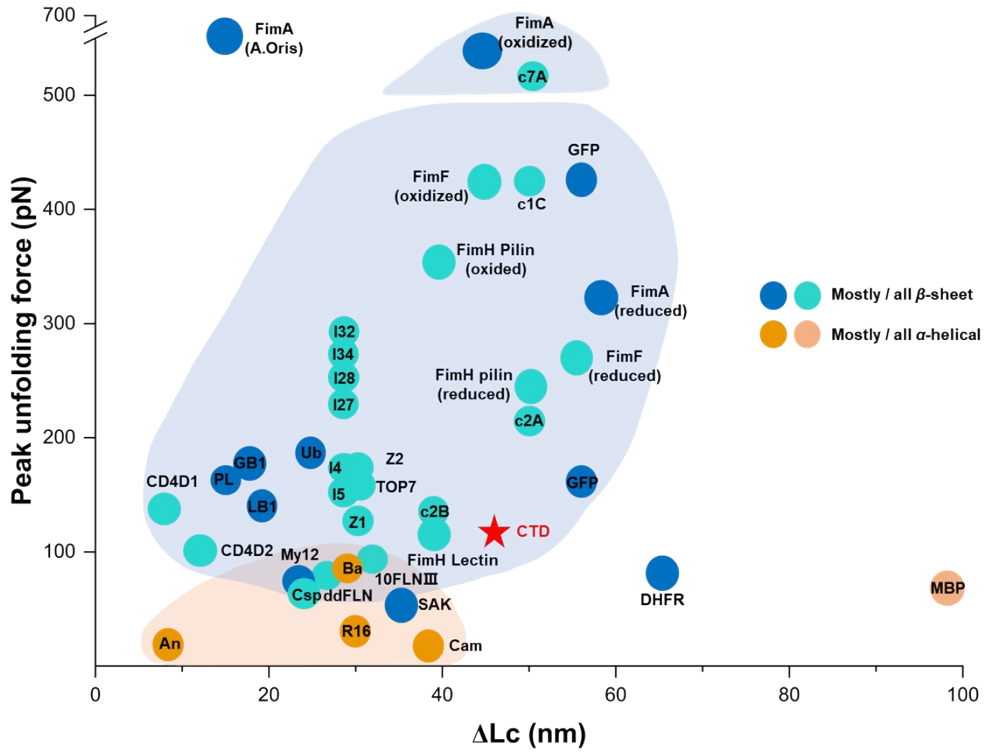

Fig. S7 A selection of studied proteins comparing mechanical stability, structure and ΔLc.

Supplementary Table

Table S1 AFM-SMFS results of the CTD.

	46 nm	35 nm	11 nm	19 nm	28 nm
k_{off} (s^{-1})	1.03±0.37	0.21±0.13	0.24±0.21	0.13±0.11	0.33±0.29
Δx (nm)	0.20±0.01	0.24±0.02	0.24±0.03	0.26±0.03	0.23±0.03

Supplementary References：

1. R. Yang, Y. H. Wong, G. K. T. Nguyen, J. P. Tam, J. Lescar and B. Wu, *J. Am. Chem. Soc*, 2017, 139, 5351-5358.
2. Y. Deng, T. Wu, M. Wang, S. Shi, G. Yuan, X. Li, H. Chong, B. Wu and P. Zheng, *Nat. Commun*, 2019, 10, 2775.
3. W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash and H. E. Gaub, *ACS Nano*, 2017, 11, 6346-6354.