Supporting Information

Exposing the Nucleation Site in α-Helix Folding: A Joint Experimental and Simulation Study

Arusha Acharyya,[1,‡] Yunhui Ge,[2,‡] Haifan Wu,[3,‡] William F. DeGrado,[3,*] Vincent A.Voelz,[2,*] and Feng Gai[1,*]

[1]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104

[2]Department of Chemistry, Temple University, Philadelphia, PA 19122

[3]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158

Peptide Synthesis

Linear peptide synthesis. The linear precursors of cross-linked peptides were synthesized using 9-fluorenylmethoxy-carbonyl (Fmoc) chemistry. All peptides were synthesized on a 0.1 mmol scale using a Biotage Initiator+ Alstra peptide synthesizer. A typical reaction cycle includes Fmoc deprotection, washing, and coupling steps. The deprotection was carried out for 5 min at 70 ºC with 4.5 mL 20% 4-methylpiperidine in DMF. A standard coupling step (for all amino acids except histidine and cysteine) was done for 5 min at 75 ºC with 5 equivalents Fmoc-protected amino acids, 4.98 equivalents HCTU, and 10 equivalents DIPEA (relative to the amino groups on resin) in DMF at a final concentration of 0.125 M amino acids. For histidine, a coupling reaction was done at room temperature for 30 min. Cysteine coupling was conducted at 50 ºC. Peptide cleavage was carried out in the presence of TFA/EDT/TIS (95:2.5:2.5, v/v) for 2 h at room temperature. The crude peptide was obtained after precipitation in cold diethyl ether.

Peptide cross-linking. Linear peptides (crude) were dissolved in a 1:1 (v/v) mixture of acetonitrile and 100 mM NH_4HCO_3 solution (pH 8) at a concentration of 0.5 mg/mL and then treated with TCEP (1.1 eq.) for 1 h at room temperature. α,α-dibromo-m-xylene (1.5 eq. in 100 µL DMF) was added. The reaction progress was monitored by analytical HPLC every 2 h. Extra alkylation agent can be added if necessary. Upon completion, TFA was added to adjust the pH to 2-3.

Peptide purification and characterization. Peptide purification was carried out on a Varian Prostar 210 HPLC system with a Higgins C4 semi-prep column (10 µm, 250 × 10 mm) using solvent A (0.1% TFA in water) and B (0.1% TFA in acetonitrile). After 5 min equilibration with 5% B at a flow rate of 5 mL/min, a gradient of 5–20% B in 15 min followed by 20–40% B in 100 min was used. The mass and purity of synthesized peptides were verified by a Shimazu AXIMA MALDI-TOF mass spectrometer and an HP 1100 analytical HPLC system, respectively (Figure S6).

Global Fitting of CD T-melts.

To extract folding-unfolding thermodynamics, the CD T-melts of C-Cap and N-Cap1 were globally fit to the following equation:

$$\theta(T) = \frac{\theta_F + K_{eq}(T) \times \theta_U}{1 + K_{eq}(T)}, \qquad (1)$$

where

$$K_{eq}(T) = \exp(-\Delta G(T) / RT) \qquad (2)$$

and

$$\Delta G(T) = \Delta H_{\mathrm{m}} + \Delta C_{\mathrm{p}} \cdot (T - T_{\mathrm{m}}) - T \cdot [\Delta S_{\mathrm{m}} + \Delta C_{\mathrm{p}} \cdot \ln(T / T_{\mathrm{m}})] \qquad (3)$$

Here, θ_F and θ_U are the folded and unfolded CD baseline, respectively, $K_{eq}(T)$ is the equilibrium constant for unfolding, $T_m = \Delta H_m/\Delta S_m$ is the thermal melting temperature, ΔH_m is the enthalpy change at T_m, ΔS_m is the entropy change at T_m, and ΔC_p is the heat capacity change, which has been assumed here to be zero. In the fit, both θ_F and θ_U were assumed to be temperature independent and θ_F was treated as a global fitting parameter.

Molecular Dynamics Simulations

Crosslinked peptides were constructed using the AmberTools tleap program. The non-natural amino acids were built using UCSF Chimera and parameterized using the Generalized AMBER force field (GAFF),[1] with partial charges computed using the AM1-BCC methods.[2] The ACPYPE program[3] was then used to convert the topology file format for use with the GROMACS.[4]

To generate initial structures of massive explicit solvent simulation, implicit solvent replica-exchange molecular dynamics (REMD) simulations were performed using GROMACS 4.5.4.[4] AMBER ff99sb-ildn-NMR force field[5] in conjunction with the OBC GBSA implicit solvation model[6] was used. Stochastic (Langevin) integration was used with a time step of 2 fs. Twenty-four replicas with temperatures exponentially spaced from 300 to 450 K were chosen to ensure broad conformational sampling. Each replica was running for 2 μs which yields 48 μs in total for each design. Exchanges were attempted every 10 ps. 10 conformations were taken from the lowest temperature replica (300K) using k-centers clustering algorithm in MSMBuilder3[7] based on root mean square deviation (rmsd). Cubic periodic boxes were filled with solvated protein and counterions (~100 mM NaCl) to neutralize the system. A full list of particle numbers and box sizes can be found in Table S1. Simulations were minimized and then pressure-equilibrated at 1 atm for 200 ps using constant-pressure molecular dynamics coupled to a Berendsen thermostat with time constant 1 ps and compressibility 4.5×10^{-5} bar^{-1}. Trajectory data was generated using constant-volume molecular dynamics at 300 K, a stochastic (Langevin) integration with a 2 fs time step, and friction constant 1 ps^{-1}, coupled to a Berendsen thermostat. Hydrogen bonds were constrained using the LINCS algorithm,[8] and Particle Mesh Ewald electrostatics was used with nonbonded cutoffs of 9 Å. Snapshots of protein atoms were recorded every 100 ps, and all atoms every 1 ns.

Table S1. Number (#) of particles, periodic box size (A), and simulation time for each peptide.

Peptide	# of Atoms	# of Na^+	# of Cl^-	# of Water	A (nm^3)	Time (μs)
C-Cap	9288	6	7	3028	95.38	235.8
N-Cap1	11207	7	8	3667	115.29	228.8
N-Cap2	9273	6	7	3018	95.13	203.5

Table S2. Folding thermodynamic parameters obtained from fitting the CD T-melts of C-Cap and N-Cap1 to a two-state model (i.e., Eq. 1). The reference temperature for both cases is the corresponding thermal melting temperature (T_m).

	C-Cap	N-Cap1
ΔH_m (kcal mol^{-1})	-11.9	-9.1
ΔS_m (cal mol^{-1} K^{-1})	-42	-34
ΔC_p (cal mol^{-1} K^{-1})	0	0
T_m (°C)	8.3	-6.4

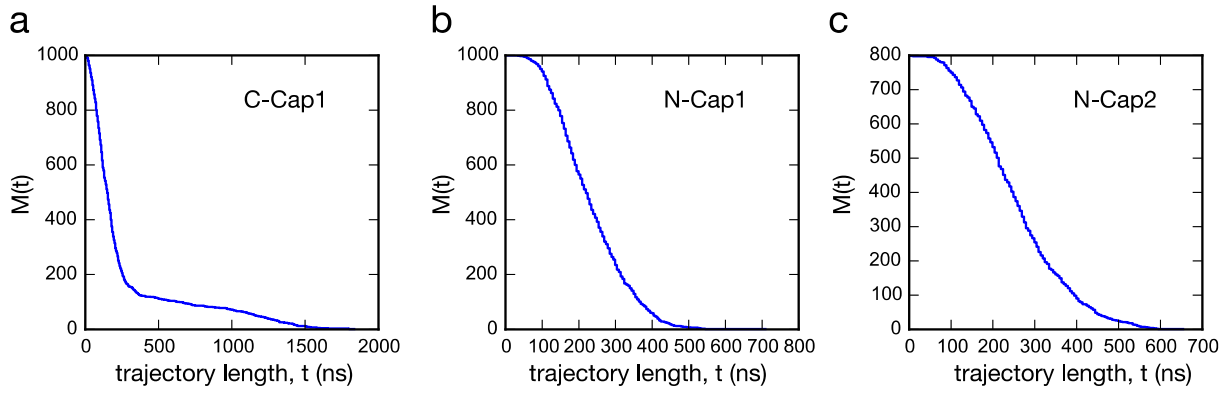

Figure S1. Distributions of trajectory lengths shown as M(t), the number of trajectories that reach a given length of time (t).

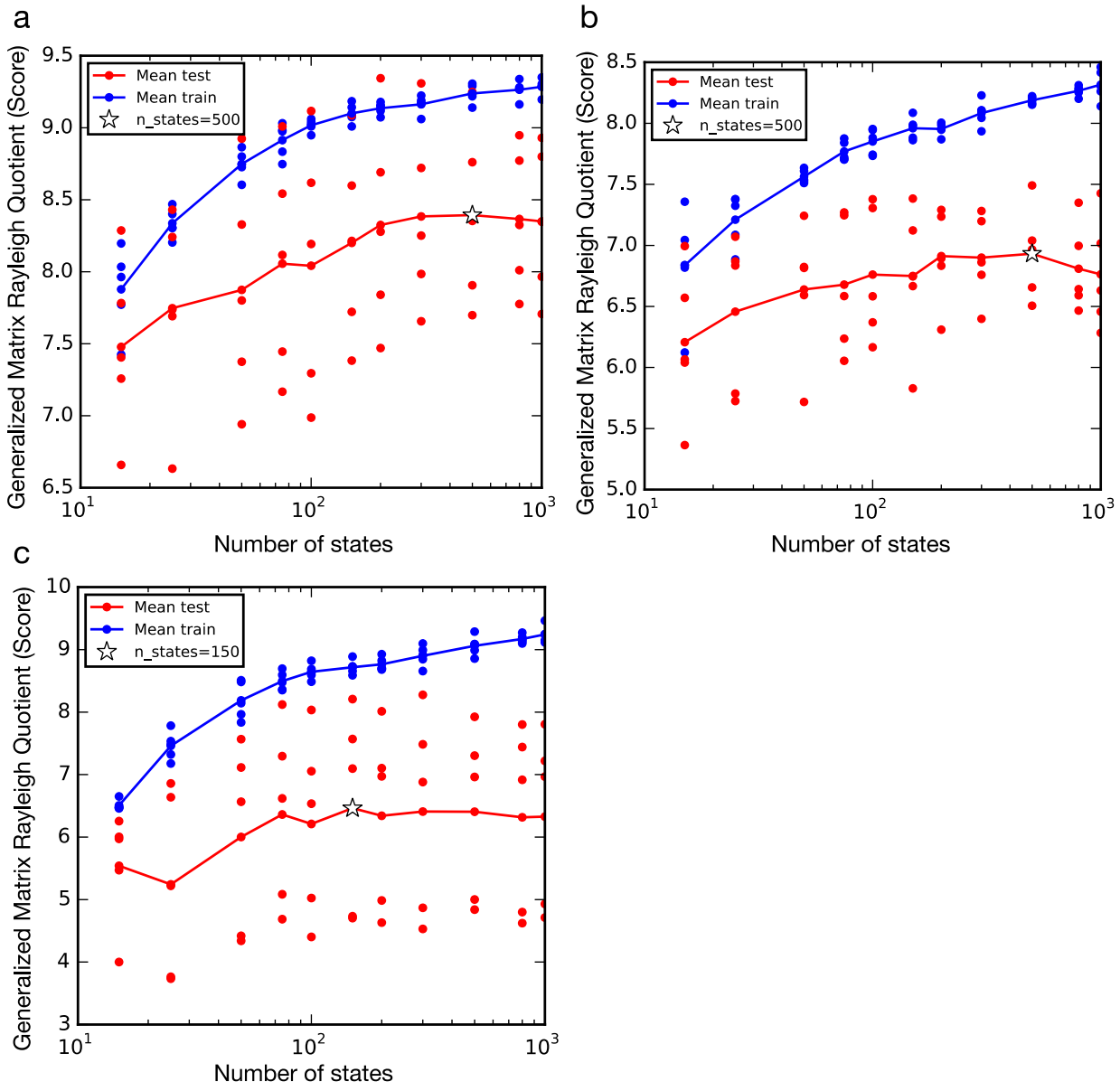

Figure S2. The generalized matrix Rayleigh quotient (GMRQ) method was used to optimize the number of states for constructing an MSM of the stapled peptide. Here, other model construction parameters are held fixed (i.e. 4 tICA components, tICA lag time of 5 ns).

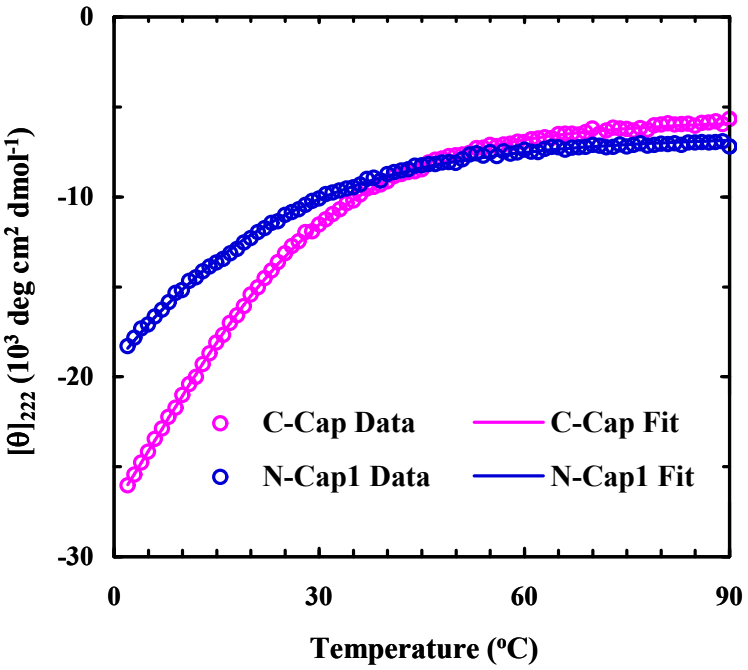

Figure S3. Global fitting results of the CD T-melts of C-Cap and N-Cap1, as indicated. The resultant thermodynamic parameters for folding are given in Table S2.

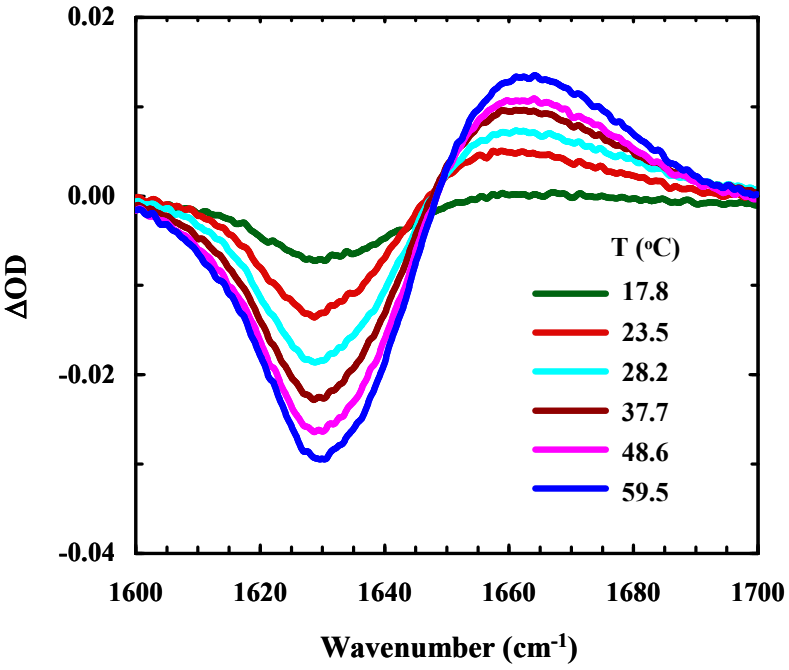

Figure S4. Difference FTIR spectra of the C-Cap peptide, which were generated by subtracting the spectrum collected at 12.9 °C from those collected at higher temperatures, as indicated.

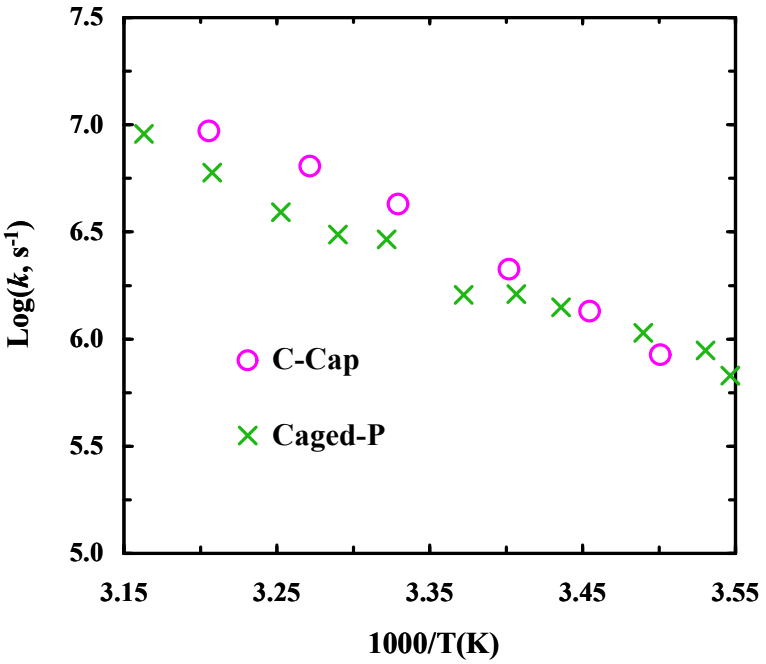

Figure S5. Temperature dependence of the *T*-jump induced relaxation rate constants of C-Cap and cyc-RKAAAD (Caged-P) peptides, as indicated.

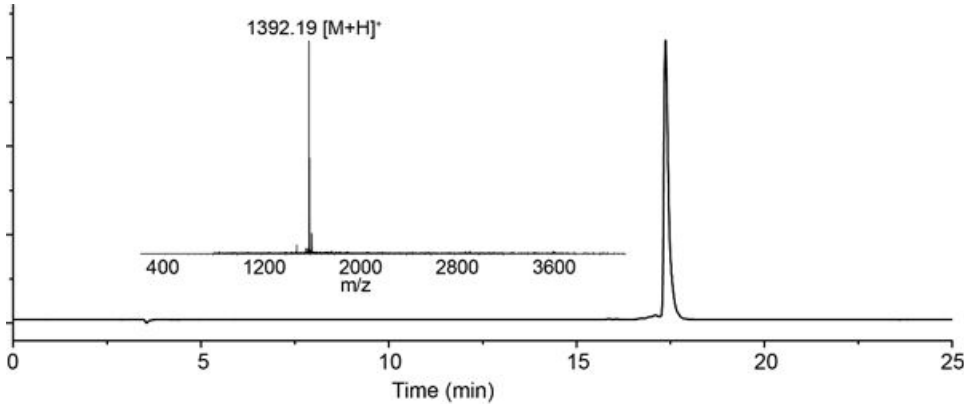

Figure S6. Analytical HPLC traces and mass spectrometry data of N-cap1 peptide

References

1. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. a., Development and testing of a general amber force field. In *Journal of Computational Chemistry*, Wiley-Blackwell: **2004**,*25*,1157-1174.

2. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. In *Journal of Computational Chemistry*, **2002**,*23*,1623-1641.

3. Sousa da Silva, A. W.; Vranken, W. F., ACPYPE - AnteChamber PYthon Parser interfacE. In *BMC Res Notes*, BioMed Central: **2012**,*5*,367.

4. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. In *SoftwareX*, **2015**,*1-2*,19-25.

5. Li, D.-W.; Brüschweiler, R., NMR-based protein potentials. In *Angew. Chem. Int. Ed. Engl.*, **2010**,*49*,6778-6780.

6. Onufriev, A.; Bashford, D.; Case, D. a., Exploring protein native states and large-scale conformational changes with a modified generalized born model. In *Proteins: Structure, Function, and Bioinformatics*, **2004**,*55*,383-394.

7. Harrigan, M. P.; Sultan, M. M.; Hernández, C. X.; Husic, B. E.; Eastman, P.; Schwantes, C. R.; Beauchamp, K. A.; Mcgibbon, R. T.; Pande, V. S., MSMBuilder: Statistical Models for Biomolecular Dynamics. In *Biophysj*, Biophysical Society: **2017**,*112*, 10-15.

8. Hess, B., P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. In *Journal of Chemical Theory and Computation*, **2008**,*4*, 116-122.