Supporting Information

Ihalainen *et al.* 10.1073/pnas.0712099105

SI Text

Molecular System. Alanine 1-^{13}C and $H_2{}^{18}O$ were purchased from Cambridge Isotopes Laboratories. The ^{16}O to ^{18}O exchange in the Ala-carbonyl was performed as described in ref. 1, and the FMOC-protection for the ^{13}C = ^{18}O-labeled alanine was performed as described in ref. 2. IR, together with mass spectroscopic analysis, confirmed that the sample contained >90% FMOC-Ala-1-^{13}C = ^{18}O-^{18}OH. Ac-AACAK 5AAAAK^{10}AAACK^{15}A-NH_2 labeled at eight different positions were prepared by using FMOC-based solid-phase peptide synthesis (GL Biochem). The peptides were then cross-linked with the photoisomerizable linker (3) to obtain the photoswitchable peptide. After removing TFA, the sample was diluted in D_2O to a concentration of ≈1 mM.

Experimental Setup. For the time-resolved measurements, two electronically synchronized femtosecond laser systems were used (4). The output of laser system 1 was frequency doubled to generate pulses at 420 nm, which switches the azobenzene from the *cis* to the *trans* state. The output of laser system 2 was used to pump an OPA to obtain IR probe pulses (100 fs, center frequency 1,620 cm^{-1}, bandwidth 240 cm^{-1} FWHM). The probe and reference beams were frequency-dispersed in a spectrometer and imaged onto a 2 × 32-pixel HgCdTe detector array. The sample was switched back to the initial *cis*-state by using a cw-Argon-ion laser at 366 nm, (and a properly filtered Hg-lamp for the stationary FTIR data).

Force-Field and Implicit Solvation Model. All simulations and most of the analysis of the trajectories were performed with the program CHARMM (5); the rest of the analysis was done with the program WORDOM (6), which is particularly efficient in handling large sets of trajectories. The peptide Ac-AACAR 5AAAAR^{10}AAACR^{15}A-NH_2 was simulated with and without the linker. The MD simulations were performed (at slightly higher temperature values of 8°C instead of 6°C and 57°C instead of 45°C) to compare with a previous IR spectroscopy study of the Arg-peptide (7) before the experimental work on the isotope-labeled Lys-peptide had started. The Arg/Lys difference is not expected to affect the agreement between experimental and simulation results, because of the similar side-chain length and positively charged end groups. Also, similar folding kinetics are observed for the Arg- and Lys-peptides by IR spectroscopy analysis (7). All heavy storms were considered explicitely as well as the hydrogen atoms bound to nitrogen or oxygen atoms (PARAM19 force field). The default cutoff of 7.5 Å was used for the nonbonding interactions. A mean field approximation based on the solvent-accessible surface area was used to describe the main effects of the aqueous solvent (8).

Parametrization of the Cross-Linker. The atom types for the cross-linker atoms were derived from the PARAM19 amide backbone and phenyl ring of Phe. For the double bond of the two nitrogen atoms between the rings of the photoswitch, the parameters for the dihedral angles were taken from ref. 9. The parameters for bonds, angles, and impropers were derived per analogy with the corresponding proteinaceous fragments in PARAM19.

Dihedral Function. To emulate the photoswitching process (i.e., isomerization of the C-n = N-C dihedral) the energy term

$$E^{N=N}_{dihedral} = 21.6(1 + cos(2\theta - 180)) + 4.36(1 + cos4\theta) \quad [1]$$

(black curve in Fig. S1) was modified by removing the minimum at $\theta = 0°$ and using a force constant of 60 kcal/mol to preserve the shape of the function from 90° to 180°

$$E^{N=N}_{dihedral,trans} = 60.0(1 + \cos\,\theta) + 4.36(1 + \cos\,2\theta) \quad [2]$$

(red curve in Fig. S1). The Langevin dynamics simulations of folding were run with the modified dihedral term starting from snapshots saved during equilibrium *cis* runs. The isomerization took place within a few picoseconds.

MD Simulations of the Peptide with Cross-Linker in the *cis* and *trans* Conformations. The equilibrium ensemble of the peptide in the *cis* conformation of the cross-linker was sampled by two replica exchange MD (REMD) (10) simulations (9 μs and 15 μs) of six replica each at temperature values of 281 K, 304 K, 330 K, 358 K, 388 K, and 420 K. Upon merging the REMD simulation segments at 281 K (totalling 24 μs), 100 equally spaced snapshots were selected as starting structures for the folding runs at low temperature (see below). The same procedure was used to select, from simulations at 330 K, the 50 starting conformations of the folding runs at high temperature. The equilibrium ensemble of the peptide in the *trans* conformation of the cross-linker was sampled by four REMD simulations (3 μs each, for a total of 12 μs) of six replica each with the same temperature values used for the cis equilibrium REMD simulations.

MD Simulations of Folding. As in the time-resolved IR experiments, in the MD simulations, the folding process is triggered by an ultrafast isomerization of the cross-linker within a few picoseconds. Therefore, the MD runs closely mimic the photo-triggered α-helix formation. Fifty (at 330 K) or 100 (at 281 K) Langevin dynamics runs of 4 μs each were started from the equilibrium *cis* conformation of the peptide previously sampled by REMD (see above). Moreover, 50 Langevin dynamics runs at each of the two temperature values were performed for the peptide without the linker. A friction coefficient of 1 ps^{-1} was used. This value is much smaller than the one of water (43 ps^{-1} at 330 K) to allow for sufficient sampling within the microsecond time scale of the simulations. A time step of 2 fs was used and the coordinates were saved every 20 ps for a total of 2 × 10^5 snapshots for each 4-μs run. A series of 100 4-μs runs requires two months on a 100-CPU cluster (about 0.5 μs simulation time per week on a single CPU). Using explicit water simulations it would have been impossible to obtain the 0.4 milliseconds of simulation time required to sample a statistically significant number of folding transitions at low temperature, which is a necessary condition for the present analysis.

Hydrogen Bond Analysis. The number of hydrogen atoms (covalently bound to amide nitrogen or guanidinium nitrogen atoms) close to each carbonyl oxygen was calculated along the MD trajectories to compare with the experimental signal of the isotopically labeled peptide. Oxygen–hydrogen distance cutoffs of 2.6 Å or 2.8 Å (with or without O···H–N angle cutoff of 120°) yielded essentially identical kinetic traces. Therefore, the distance cutoff of 2.6 Å (without angle criterion) was used.

1. Murphy RC, Clay KL (1990) Preparation of labeled molecules by exchange with oxygen-18 water. *Methods Enzymol* 193:338–348.
2. Bolin DR, Sytwu II, Humiec F, Meienhofer JJ (1989) Preparation of oligomer-free N-alpha-FMOC and N-alpha-urethane amino acids. *Int J Peptide Protein Res* 33:353-359.
3. Kumita JR, Smart OS, Woolley GA (2000) Photo-control of helix content in a short peptide. *Proc Natl Acad Sci USA* 97:3803–3808.
4. Bredenbeck J, Helbing J, Hamm P (2004) Continuous scanning from picoseconds to microseconds in time resolved linear and nonlinear spectroscopy. *Rev Sci Instrum* 75:4462-4467.
5. Brooks BR, *et al.*(1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
6. Seeber M, Cecchini M, Rao F, Settanni G, Caflisch A (2007) Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23:2625–2627.
7. Ihalainen JA, *et al.* (2007) Folding and unfolding of a photoswitchable peptide from picoseconds to microseconds. *Proc Natl Acad Sci USA* 104:5383–5388.
8. Ferrara P, Apostolakis J, Caflisch A (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins Struct Funct Bioinf* 46:24–33.
9. Carstens H (2004) Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modellbildung. PhD thesis (Ludwig Maxmilians University, Munich).
10. Sugita V, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141-151.
11. Muff S, Caflisch A (2008) Kinetic analysis of molecular dynamics simulations reveals changes in the denatured state and switch of folding pathways upon single-point mutation of a β-sheet miniprotein. *Proteins Struct Funct Bioinf* 70:1185–1195.
12. Andersen CAF, Palmer AG, Brunak S, Rost B (2002) Continuum secondary structure captures protein flexibility. *Structure (London)* 10:174–184.

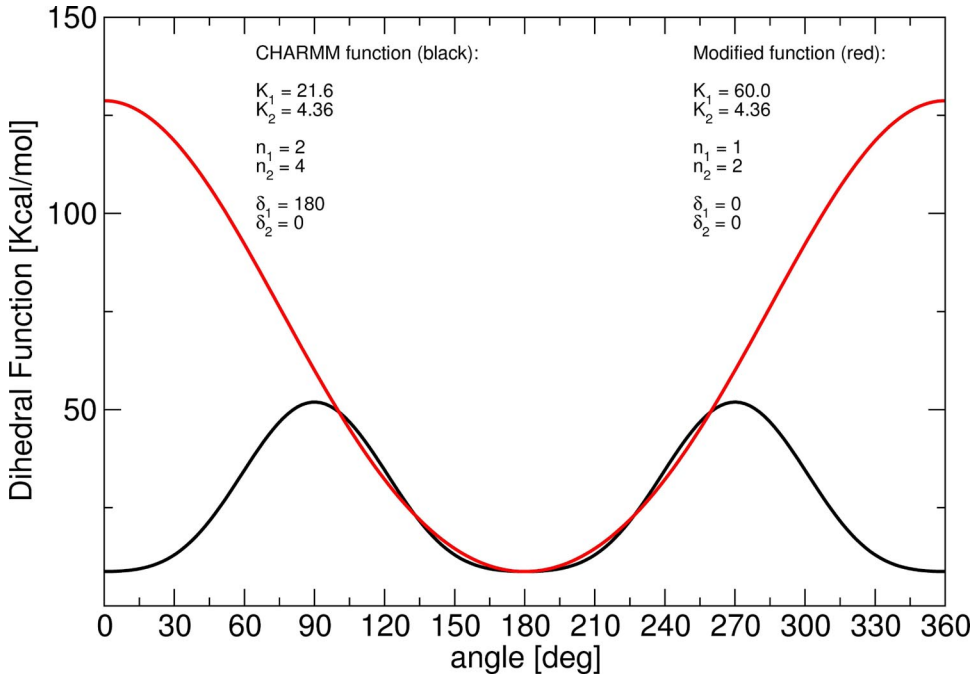

Fig. S1. Dihedral energy function: $E^{N=N}_{dihedral} = K_1(1 + \cos(n_1\theta - \delta_1)) + K_2(1 = \cos(n_2\theta - \delta_2))$.

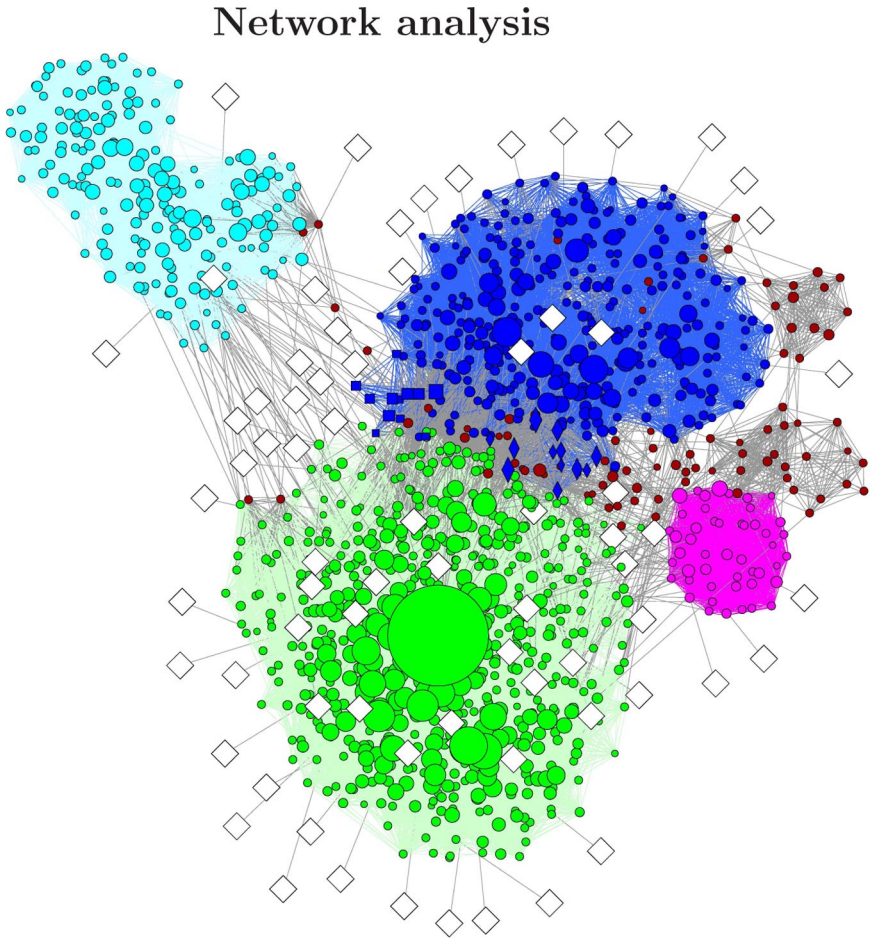

Fig. S2. See legend of Fig. 4 in main text and Table S1.

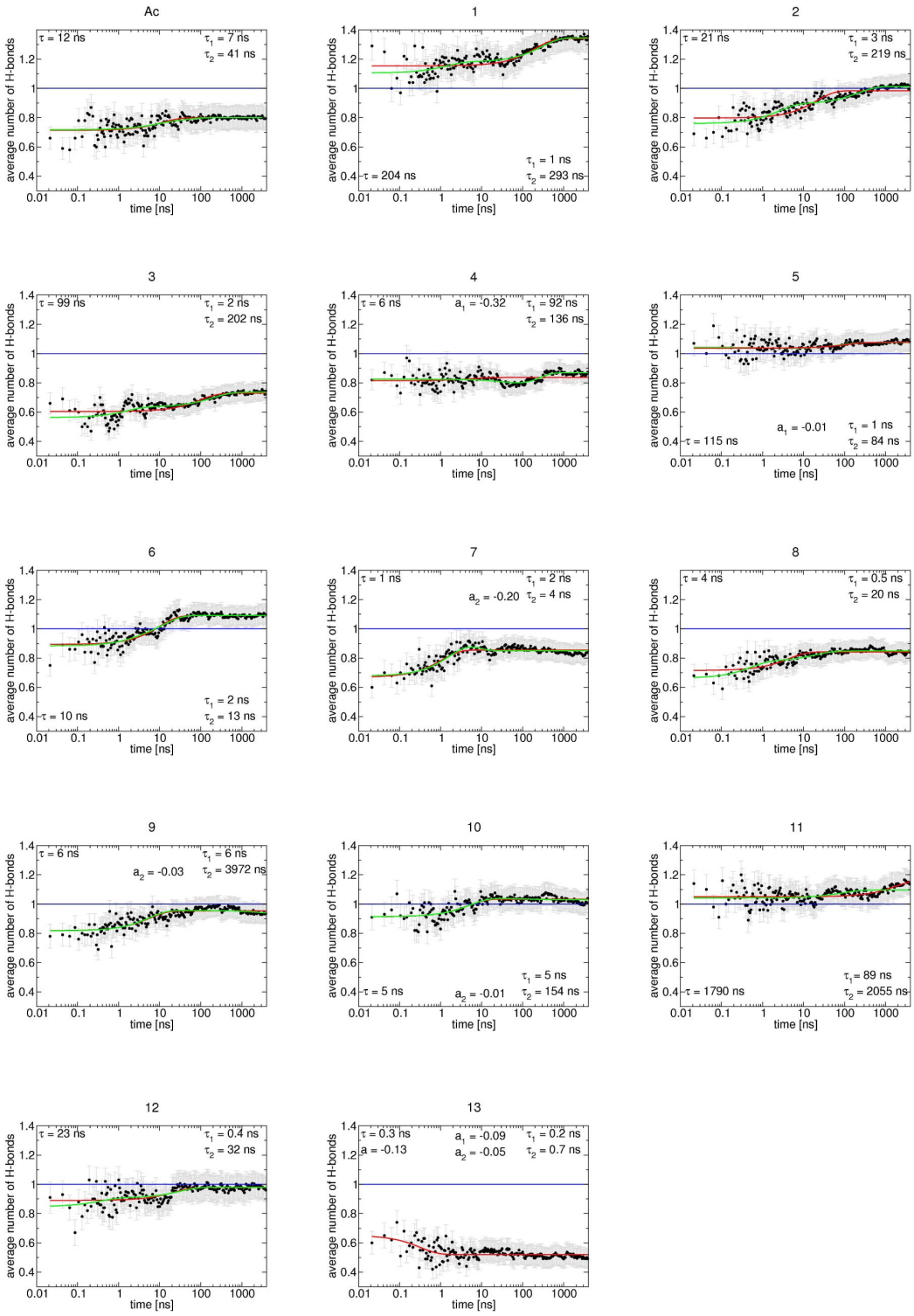

Fig. S3. Folding signal of individual carbonyl groups from Langevin dynamics simulations (first set, 281 K; second set, 330 K). The semilog plots show the time dependence of the average number of hydrogen atoms within 2.6 Å of the backbone carbonyl oxygen. Single- and double- (only at 281 K) exponential fitting curves (red and green, respectively) are shown with folding times and negative amplitudes given within each plot. All of the other values are shown in Tables S2 and S3. The blue horizontal line at $y = 1.0$ is drawn as a reference. Carbonyl groups in residues 14–16 are almost never involved in hydrogen bonds and are therefore not shown.

Site-selective traces: 330 K

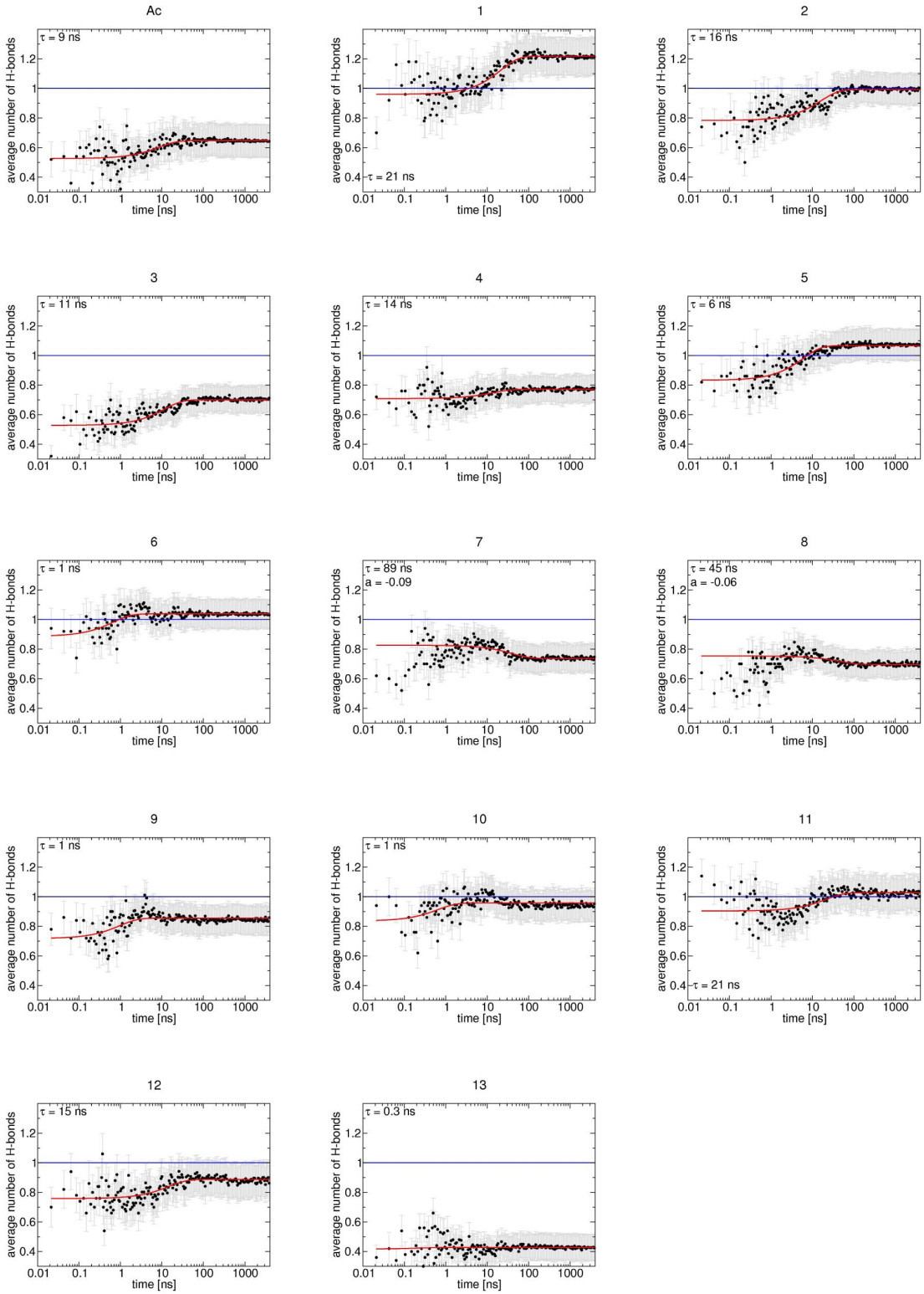

Fig. S3. Continued.

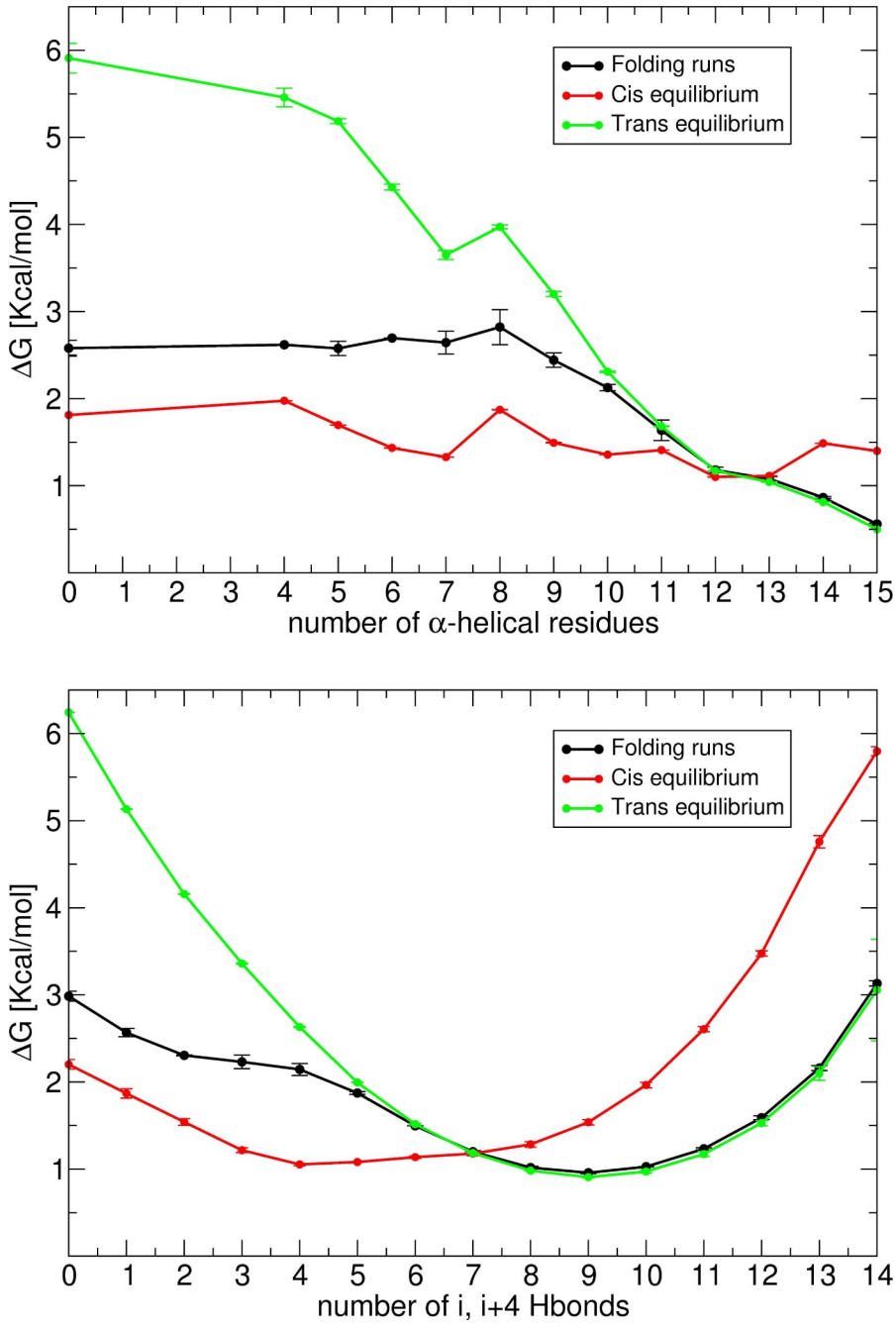

Fig. S4. Free-energy profiles at 281 K as a function of the number of α-helical residues (*Upper*) or number of α-helical backbone hydrogen bonds (*Lower*). A total of 4×10^6 conformations were used for the folding runs, 1.2×10^6 conformations for the *cis* equilibrium simulations, and 0.6×10^6 conformations for the *trans* equilibrium simulations. Values at equilibrium were extracted from REMD segments at 281 K: 24 μs for *cis* and 12 μs for *trans*. The number of α-helical residues was calculated by DSSP (12), whereas the number of α-helical hydrogen bonds was calculated by CHARMM using a cutoff of 2.6 Å for the distance between hydrogen and oxygen atoms. Error bars are calculated by block averages, where the dataset is divided into two equal subsets. In contrast to the network analysis, the complexity of the folding process is hidden in these projections along a one-dimensional progress variable.

Table S1. Free-energy basins (i.e. folded state and most populated metastable states) determined by kinetic grouping analysis (KGA) (11) for the folding runs at 281 K with τ_{commit} = 10 ns

			Helical content in basins		
Most populated node [conformation]	Weight, %	τ_f, ns	H	H + G + I	Color of KGA basin
-HHHHHHHHHHHHHHH-	28.3	5	13.6 ± 3.1	14.6 ± 3.2	Green
–TT–IHHHHHHHHH-	32.6	396	7.3 ± 2.8	8.9 ± 3.1	Blue circles
-BSSSB-HHHHHHHHH-	1.3	424	7.7 ± 2.7	8.7 ± 2.5	Blue squares
–TTS-SHHHHHHHHH-	0.1	395	7.1 ± 1.7	8.8 ± 1.5	Blue diamonds
–HHHHHII-S-TTS–	17.4	928	5.5 ± 2.4	7.9 ± 2.2	Cyan
–SSSSS–SHHHHHH-	7.0	398	4.2 ± 2.7	5.4 ± 2.9	Magenta
Others	7.0	–	–	–	Brown

Mean folding times (τ_f) are average values for snapshots in a basin. The colors indicated in the last column are those used in Fig. S2. The first column contains the string of secondary structure of native or metastable states (i.e., of KGA basins). There are eight possible ″letters″ in the secondary structure ″alphabet:″ ″H,″ ″G,″ ″I,″ ″E,″ ″B,″ ″T,″ ″S,″ and ″–″, standing for α-helix, 3_{10} helix, π-helix, extended, isolated β-bridge, hydrogen bonded turn, bend, and unstructured, respectively (12). Because the N- and C-terminal residues are always assigned an ″–″ (12), a 16-residue peptide can, in principle, assume $8^{16} \simeq 10^{14}$ conformations.

Table S2. Single- and double-exponential fit parameters for individual folding signals at 281 K from the Langevin dynamics runs

	Single exponential		Double exponential			
Residue	a	τ, ns	a_1	τ_1, ns	a_2	τ_2, ns
Ac		12	0.05	7	0.04	41
1	0.19	204	0.08	1	0.16	293
2	0.19	21	0.13	3	0.12	219
3	0.13	99	0.07	2	0.10	202
4	0.02	6	**−0.32**	92	0.36	136
5	0.04	115	**−0.01**	1	0.04	84
6	0.20	10	0.05	2	0.16	13
7	0.18	1	0.40	2	**−0.20**	4
8	0.13	4	0.11	0.5	0.07	20
9	0.14	6	0.14	6	**−0.03**	3,972
10	0.12	5	0.13	5	**−0.01**	154
11	0.11	1,790	0.05	89	0.01	2,055
12	0.09	23	0.06	0.4	0.07	32
13	**−0.13**	0.3	**−0.09**	0.2	**−0.05**	0.7
14	**−0.04**	9	-	-	-	-
15	**−0.04**	11	-	-	-	-
16	**−0.06**	2,650	-	-	-	-

Negative amplitudes are shown in bold type.

Table S3. Single-exponential fit parameters for individual folding signals at 330 K from the Langevin dynamics runs

	Single exponential	
Residue	*a*	τ, ns
Ac	0.12	9
1	0.26	21
2	0.21	16
3	0.17	11
4	0.06	14
5	0.23	6
6	0.15	1
7	**−0.09**	89
8	**−0.06**	45
9	0.14	1
10	0.12	1
11	0.12	21
12	0.13	15
13	0.01	0.3
14	**−0.11**	2
15	**−0.06**	4
16	**−0.03**	5

Negative amplitudes are shown in bold type.