THE JOURNAL OF PHYSICAL CHEMISTRY B

ARTICLE

pubs.acs.org/JPCB

Reducing Lambda Repressor to the Core

Maxim B. Prigozhin,[†] Krishnarjun Sarkar,[†] Dennis Law,[‡] William C. Swope,[‡] Martin Gruebele,[*,†,§] and Jed Pitera[‡]

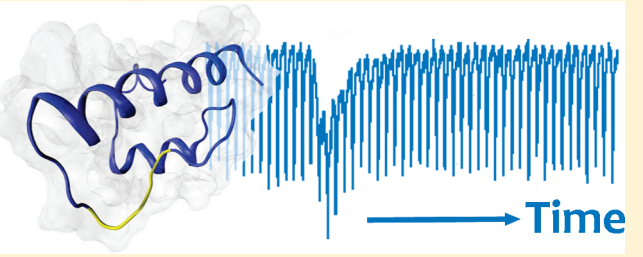
[†]Department of Chemistry, University of Illinois, Urbana, Illinois 61801, United States

[‡]IBM Research, IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120, United States

[§]Department of Physics and Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Lambda repressor fragment λ^*_{6-85} is one of the fastest folding small protein fragments known to date. We hypothesized that removal of three out of five helices of λ^*_{6-85} would further reduce this protein to its smallest folding core. Molecular dynamics simulations singled out two energetically stable reduced structures consisting of only helices 1 and 4 connected by a short glycine/serine linker, as well as a less stable control. We investigated these three polypeptides and their fragments experimentally by using circular dichroism, fluorescence spectroscopy, and temperature jump relaxation spectroscopy to gain insight into their thermodynamic and kinetic properties. Based on the thermal melts, the order of peptide stability was in correspondence with theoretical predictions. The most stable two-helix bundle, λ_{blue1}, is a cooperatively folding miniprotein with the same melting temperature and folding rate as the full-length λ^*_{6-85} pseudo wild type and a well-defined computed structure.



INTRODUCTION

Fragments of the lambda repressor protein containing amino acids 1−101 or 6−85 have long been subject to folding studies, including folding-function correlations,[1] remodeling of the hydrophobic core,[2] detection of fast two-state folding by NMR,[3,4] and detection of multiple folding mechanisms of mutants.[5,6] These fragments comprise as few as five of the helices from the complete protein, yet they have turned out to be thermodynamically quite stable in thermal and chemical denaturation studies.[7]

How small can a lambda repressor fragment be made while maintaining a high melting temperature and cooperativity? A previous experimental and computational analysis suggested that helices 1 and 4 constitute the minimal folding core and fold first.[8] Protein stability was particularly sensitive to mutations in this pair of helices, and molecular dynamics simulations showed that they are most likely to form secondary structure at room temperature. Explicit solvent replica-exchange thermodynamics simulations of a lambda repressor fragment (see Supporting Information) also showed higher stability of helices 1 and 4 relative to other secondary structure. These results are in line with the idea that native-like preorganized secondary structure can accelerate folding of lambda repressor fragments,[9] whereas non-native secondary structure (e.g., extended structure that appears upon addition of guanidine hydrochloride[10]) can greatly slow down folding of lambda repressor fragment from microseconds to milliseconds.[11]

On the basis of this knowledge, we computationally redesigned λ^*_{6-85} (the tyrosine 22 → tryptophan mutant of fragment 6−85 of bacteriophage lambda repressor[12]) to remove helices 2, 3, and 5, and connect helices 1 and 4 directly by a short flexible linker optimized to allow native-like contacts to form between these two helices. Two of the fragments, λ_{blue1} and λ_{blue2}, had a native-like modeled structure and proved most stable in molecular dynamics modeling, so we expressed them and investigated their folding experimentally. λ_{blue3} was also conformationally similar to λ^*_{6-85} but was energetically less favorable than the first two variants. It served as a control to reveal a correlation between the computed energies and the experimentally observed stability and folding rate.

λ_{blue1} turns out to have the same high melting point (62 °C) as the pseudo wild type λ^*_{6-85} fragment, and it folds at about the same rate (τ_{obs} = 15 ± 4 μs). Its circular dichroism (CD) spectrum reveals well-formed helical content. Its melting points, as measured by CD and fluorescence, are in close agreement. λ_{blue1} is an apparent two-state folder. In contrast, its individual N and C terminal fragments are much less helical and do not show a cooperative transition. Thus, interaction between helices 1 and 4 is critical for the folding of λ_{blue1}. λ_{blue2} also has some helical content, while λ_{blue3} has a cooperative transition at much lower

Received: October 24, 2010
Revised: December 23, 2010
Published: February 14, 2011

temperature, and a CD spectrum indicating more β sheet content.

No attempt was made to redesign hydrophobic side chains that interacted with helices 2, 3, and 5 in the original λ^*_{6-85}, to keep the new miniprotein as close as possible to the original λ^*_{6-85} sequence. As a result, λ_{blue1} is more prone to aggregation than λ^*_{6-85}. We do predict two mutations to make the protein more soluble. Likewise, the fluorescence signal could be enhanced by adding a quencher (e.g., histidine, cysteine) in contact with tryptophan 22. In its present form, λ_{blue1} will be useful for two applications: long time or replica exchange molecular dynamics simulations of fast folding that compare the reduced to the original molecule, and further redesign to create a stable two-helix bundle with low aggregation propensity and large fluorescence signal. We make PDB coordinates for the predicted structures available to facilitate such studies.

COMPUTATIONAL AND EXPERIMENTAL METHODS

Simulation Protocol. All simulations were carried out using the AMBER10 suite of programs. Protein structures were visualized using the PyMOL software package. The AMBER parm96 all-atom force field[13] was used in conjunction with a modified generalized Born/solvent accessible surface area implicit solvent model.[14] The interior dielectric was 1.2, the exterior dielectric was 80.0, and the surface tension was 0.005 kcal/mol/Å^2. This combination of simulation parameters has been shown to be effective at discriminating between folded and unfolded conformations of many different proteins.[15]

After construction, the folded and unfolded forms of each variant protein were subjected to a two-stage structural optimization protocol. In the first stage, 10 000 steps of steepest descent energy minimization were performed while the atoms in helices 1 and 4 were restrained to their initial positions by relatively strong (5 kcal/mol/Å^2) Cartesian positional restraints. This allowed relaxation of the linker regions without distortion of the helices. In the second stage, an additional 10 000 steps of steepest descent energy minimization were performed without the helical positional restraints, allowing all atoms to move.

Once optimized, the folded and unfolded structures of each variant were simulated using conventional molecular dynamics at four different temperatures (300, 350, 400, 450 K). The folded structures of each variant were simulated for 6.25 ns, while the unfolded structures were simulated for 5 ns. The final 1.25 ns of both the folded and unfolded state trajectories were used for subsequent analysis. Uncertainties were estimated using the observed variance and correlation time of the data and are reported as ± 1 standard deviation. The target temperatures were enforced using an Andersen thermostat[16] with collisions every 10 ps. All bonds to hydrogen atoms were constrained to their equilibrium lengths using the SHAKE algorithm,[17] and a 2 fs time step was employed. The aggregate simulation time required for the entire project was 1.3 μs. Additional details are provided in the Supporting Information.

Protein Expression. λ_{blue1} gene inserted in a pET15-b vector with a histidine tag coding sequence was obtained from GenScript Corp. (Piscataway, NJ). λ_{blue2} and λ_{blue3} were derived from λ_{blue1} by site-directed mutagenesis (Quickchange kit; Stratagene, La Jolla, CA). Genes were expressed in *Escherichia coli* BL21 cells (Stratagene Corp., La Jolla, CA). The cell cultures were grown at 37 °C in LB Lennox Broth (Fisher Scientific Inc., Waltham, MA) and 100 μg/mL ampicillin (Fisher Scientific Inc., Waltham, MA) with shaking at 220 rpm. Protein biosynthesis was induced when absorbance at 600 nm reached 0.6−0.8 OD using isopropyl β-D-thiogalactopyranoside (IPTG) (Inalco, Milano, Italy) with the final concentration of 1 mM. Cell growth was allowed to proceed for 10 h after induction at 25 °C with agitation. The cells were collected by centrifugation for 10 min at 5000 rpm. Cell pellets were resuspended in 150 mL of the lysis buffer consisting of 50 mM Na_2HPO_4, 500 mM NaCl, and 10 mM imidazole at pH 8.0. Lysis was achieved by sonication on ice and cell debris was removed by centrifugation. The supernatant was loaded onto a Ni-NTA His-Bind column (Novagen Inc., Madison, WI) that was washed with lysis buffer. The column was then washed with a buffer containing 50 mM Na_2HPO_4, 500 mM NaCl, and 20 mM imidazole at pH 8.0 and eluted with another buffer containing 50 mM Na_2HPO_4, 500 mM NaCl, and 250 mM imidazole at pH 8.0. After purification, the proteins were dialyzed at 4 °C for two 6-h periods against 0.1 M pH 7.0−7.5 phosphate buffer using 3500 MWCO dialysis tubing (Fisher Scientific Inc., Waltham, MA). The histidine tag was retained on all the mutants. Purity was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, and the concentration was determined by UV−visible absorption spectroscopy.

Equilibrium Folding Experiments. Thermal unfolding experiments were performed by incrementing the temperature by 3° degrees in the 20−89 °C range and measuring CD at 222 nm as well as the integrated fluorescence with the 280 nm excitation wavelength at each temperature using a Jasco spectrometer equipped with a Peltier temperature control (Jasco Inc., Easton, MD). Output fluorescence was passed through a long-pass filter with FWHM transmittance at 305 nm. CD spectra in the 200−250 nm region were recorded at 20 °C using the same spectrometer. The concentration of protein solutions used in CD measurements was approximately 10 μM.

Temperature Jump Experiment. Temperature jumps on the order of 8−10 °C were initiated with a Surelite Q-switched Nd:YAG laser (Continuum Inc., Santa Clara, CA) Raman-shifted to 1.9 μm. Fluorescence excitation was achieved by a Mira Ti:Sapphire laser (Coherent Inc., Santa Clara, CA) tripled to the wavelength of 281 $\pm$ 2 nm using a third harmonic generator (CSK Optronics Inc., Torrance, CA). The power of the laser at the sample cell was usually on the order of 3 mW at 14 ns time intervals between pulses. Peptide fluorescence was passed through a B370 band-pass filter from Hoya Corp. (Santa Clara, CA). The data were digitized using a 500 ps resolution oscilloscope (Tektronix RTD720A; Tektronix Inc., Beaverton, OR) and analyzed with a program written in LabWindows (National Instruments Inc., Austin, TX). Equilibrium temperature of the sample was maintained using an automated temperature controller (Lake Shore 330; Lake Shore Cryotronics Inc., Westerville, OH). The details of the T-jump instrument can be found elsewhere.[18] The jump temperature was calibrated using a 250 μM solution of L-tryptophan (Sigma-Aldrich Inc., St. Louis, MO). Sample concentrations were difficult to measure due to protein aggregation. On the basis of the T-jump signal and absorbance measurements, the samples were in the concentration range of 20−30 μM.

Data Analysis. All data were analyzed using an Igor Pro software package (Wavemetrics Inc., Lake Oswego, OR). Fluorescence decays were fitted to a linear combination of the lifetime right before the jump ($\chi = 1$) and a lifetime 0.5 ms after the jump ($\chi = 0$). This analysis allowed to plot relative lifetime shift as a function of time, $\chi(t)$. The zero time was estimated from the

Raman scattering peak of the solvent and the first five decays were not included in the analysis (Figure 4). The resulting exponential decay was fitted to obtain the folding-unfolding relaxation time constant τ.

Thermal denaturation curves from both fluorescence and CD experiments were fitted using a two-state thermodynamic model. Fluorescence signal base lines $F_i(T)$ were assumed to be linear functions of temperature for the native and denatured states i=N and D:

$$F_i = b_i + m_i(T - T_m) \quad (1)$$

b_i and m_i were obtained by fitting the first and the last 3−4 points in the CD and fluorescence melt profiles. The overall signal $F(T)$ was then fitted as a linear combination of the fraction that each population contributed at temperature T multiplied by the baseline signal of that population (eq 2).

$$F(T) = \frac{F_D}{1+K_{eq}} + \frac{F_N K_{eq}}{1+K_{eq}} \text{ with } K_{eq} = e^{-\Delta G(T)/RT} \quad (2)$$

The free enegy was expanded as a Taylor series about the transition temperature between folded and unfolded states (eq 3):

$$\Delta G(T) \approx \Delta G(T_m) + \Delta S(T_m)T_m - \Delta S(T_m)T \quad (3)$$

The $\Delta G(T_m)$ term is equal to zero because the difference in free energy between folded and unfolded states at the midpoint temperature T_m is zero. The linear coefficient $\Delta S(T_m) = -\partial\Delta G(T)/\partial T$ would correspond to the folding entropy at T_m in the limit where eq 3 is exact over the entire fitting range. A second order term in the Taylor expansion was not needed to fit the data within measurement uncertainty. An alternative model for ΔG in terms of a constant heating capacity for folding gave a similar quality fit and T_m.

RESULTS

Design of Truncated Protein Structures. The initial structure used for the construction of truncated lambda repressor variants was taken from the 1LMB crystal structure,[19,20] with substitution Y22W in helix 1 corresponding to a fluorescent pseudo wild type λ^*_{6-85}.[12,21,22] An additional mutation D14A slightly stabilizes (2.5 °C) the full λ^*_{6-85} fragment.[21] The truncated proteins were constructed using portions of the two largest helices, 1 and 4, connected by short glycine-serine linker segments.

Three variants (a−c) of each helix were constructed (Table 1). The sequences were chosen to minimize the size of helix 1 while preserving its interactions with helix 4. In h1a and h1b, the residues linking helix 1 to helix 2 (Gly 30 to Ser 32) were retained as part of the linker to help connect helix 1 to helix 4 in the modified protein. In h1b an additional tyrosine was inserted, which interacts strongly with Trp 22 in some λ^*_{6-85} mutants.[22] Three variants of helix 4 were constructed in a similar fashion (Table 1). The residues that connect helices 3 and 4 in λ^*_{6-85} (Phe 51/Gly 53 to Leu 57) were kept to form part of the linker between helix 1 and helix 4.

Visual inspection of the truncated helix pairs suggested that the C-terminus of the helix 1 variants would be separated from the N-terminus of the helix 4 variants by 10−20 Å, so a series of three extended Gly-Ser repeat linkers (GSG, GSGS, GSGSG) were tested. The 27 possible helix 1/linker/helix 4 combinations were then constructed in a putative folded and in a fully extended conformation (see Computational and Experimental Methods). As a reference, two variants with the wild type sequence linking helices 1 and 4 instead of the Ser-Gly linkers were also constructed.

Table 1. Residue Ranges and Sequences of Lambda Repressor Helix Variants

variant name	λ^*_{6-85} residue range	amino acid sequence
h1a	7−32	LTQEQLEAARRLKAIWEKKKNELGLS
h1b	7−33	LTQEQLEAARRLKAIWEKKKNELGLSY
h1c	7−30	LTQEQLEAARRLKAIWEKKKNELG
h4a	51−71	FNGINALNAYNAALLAKILKV
h4b	51−77	FNGINALNAYNAALLAKILKVSVEEFS
h4c	53−70	GINALNAYNAALLAKILK

Simulation to find low energy structures. All 27 variants and two controls were subjected to the same automated protocol of structural optimization by energy minimization using implicit solvent simulations to provide a coarse measure of the stability of each sequence in the folded state (see Computational and Experimental Methods and Supporting Information for details). During the first stage of optimization, the atoms in helices 1 and 4 were restrained to their initial positions. This allowed relaxation of the linker regions without distortion of the helices. In the second stage, all atoms were allowed to move.

Once structurally optimized, the folded and unfolded conformations of each variant were simulated at 300 K using conventional molecular dynamics. Additional simulations were carried out at 350, 400, and 450 K as controls but were not used to select sequences for synthesis, since most fragments unfolded at the higher temperatures. Ideally, large-scale thermodynamic sampling such as replica-exchange molecular dynamics would be performed for each variant to estimate its stable structures and melting temperature. However, our experience is that such simulations require extensive simulation to converge for proteins as large as lambda repressor, making them impractical for the screening approach taken here.[8] Though limited, our simulations were sufficient to allow local structural relaxation of the linker and interhelical angles in folded conformations and to allow the extended "unfolded" conformations to collapse to disordered structures with radii of gyration comparable to the folded conformations. The average radius of gyration of the folded simulations is 11.1 ± 0.1 Å compared to 11.3 ± 0.7 Å for the unfolded simulations. Two metrics were used to assess the stability of the folded state for each variant sequence (Table 2). First, the Cα-RMSD of the residues in helices 1 and 4 (relative to the original X-ray crystal structure) was calculated for each saved coordinate. Sequences were classified as structurally stable if the average Cα-RMSD for the helices was less than 2.5 Å. Second, the per-atom energy gap between folded and unfolded conformations at 300 K was used as a measure of energetic stability, allowing comparison of sequences with different lengths. The mutants with the wild type linker (bottom of Table 2) had energy gaps of −0.023 to −0.025 kcal/mol/atom. Thus sequences were classified as energetically stable if the observed energy gap was at least −0.020 kcal/mol/atom.

Protein Selection and Expression. Three variants from Table 2 with low Cα-RMSD values suggesting structural stability

Table 2. Twenty Nine Truncated Variants, Including Two Wild Type Linkers for Reference[a]

variant name	$\langle E_{folded} \rangle$, kcal/mol	$\langle E_{unfolded} \rangle$, kcal/mol	ΔE/atom, kcal/mol	$\langle$RMSD$\rangle$ (Å)
h1aGSGh4a	−1211.3 ± 1.7	−1159.3 ± 2.7	−0.065 ± 0.004	3.35 ± 0.14
h1aGSGh4b	−1322.6 ± 2.0	−1309.8 ± 3.2	−0.014 ± 0.004	4.93 ± 0.05
h1aGSGh4c	−1135.0 ± 1.9	−1134.7 ± 3.6	−0.000 ± 0.005	3.08 ± 0.07
h1aGSGSh4a	−1209.4 ± 3.2	−1186.7 ± 2.8	−0.028 ± 0.005	4.08 ± 0.08
h1aGSGSh4b	*−1324.4 ± 2.2*	*−1303.9 ± 2.1*	*−0.023 ± 0.003*	*1.80 ± 0.07*
h1aGSGSh4c	−1119.1 ± 2.5	−1097.6 ± 2.6	−0.028 ± 0.005	2.87 ± 0.14
h1aGSGSGh4a	−1198.1 ± 3.4	−1158.6 ± 3.7	−0.048 ± 0.006	3.17 ± 0.17
h1aGSGSGh4b	*−1325.5 ± 4.4*	*−1301.5 ± 3.3*	*−0.026 ± 0.006*	*2.38 ± 0.14*
h1aGSGSGh4c	−1125.7 ± 1.6	−1093.8 ± 4.0	−0.041 ± 0.006	3.40 ± 0.15
h1bGSGh4a	−1210.4 ± 2.4	−1186.8 ± 2.8	−0.029 ± 0.004	2.94 ± 0.07
h1bGSGh4b	−1318.2 ± 2.0	−1311.7 ± 2.8	−0.007 ± 0.004	3.43 ± 0.23
h1bGSGh4c	−1135.0 ± 1.7	−1124.9 ± 2.9	−0.013 ± 0.004	2.83 ± 0.19
h1bGSGSh4a	−1212.8 ± 4.2	−1195.1 ± 5.3	−0.021 ± 0.008	5.07 ± 0.10
h1bGSGSh4b	*−1343.2 ± 3.1*	*−1331.0 ± 2.0*	*−0.013 ± 0.004*	*1.95 ± 0.12*
h1bGSGSh4c	−1142.7 ± 1.7	−1110.6 ± 3.5	−0.041 ± 0.005	4.73 ± 0.07
h1bGSGSGh4a	−1211.2 ± 1.7	−1186.0 ± 2.7	−0.030 ± 0.004	6.11 ± 0.16
h1bGSGSGh4b	−1328.3 ± 3.3	−1325.8 ± 3.3	−0.003 ± 0.005	3.07 ± 0.03
h1bGSGSGh4c	−1137.5 ± 1.9	−1120.8 ± 2.0	−0.021 ± 0.003	3.28 ± 0.15
h1cGSGh4a	−1187.9 ± 2.8	−1174.4 ± 4.6	−0.017 ± 0.007	2.47 ± 0.05
h1cGSGh4b	−1311.2 ± 4.2	−1284.7 ± 4.9	−0.031 ± 0.006	2.06 ± 0.07
h1cGSGh4c	−1115.4 ± 3.1	−1096.9 ± 2.9	−0.026 ± 0.006	3.18 ± 0.07
h1cGSGSh4a	−1207.4 ± 2.6	−1160.7 ± 2.0	−0.060 ± 0.004	3.17 ± 0.09
h1cGSGSh4b	−1298.2 ± 2.8	−1296.2 ± 3.8	−0.002 ± 0.005	3.11 ± 0.13
h1cGSGSh4c	−1122.5 ± 2.7	−1114.6 ± 3.0	−0.011 ± 0.006	3.30 ± 0.06
h1cGSGSGh4a	−1174.0 ± 3.0	−1168.8 ± 1.9	−0.007 ± 0.005	4.17 ± 0.05
h1cGSGSGh4b	−1297.7 ± 2.4	−1274.3 ± 3.6	−0.027 ± 0.005	3.69 ± 0.46
h1cGSGSGh4c	−1114.0 ± 2.5	−1092.9 ± 2.3	−0.029 ± 0.004	5.21 ± 0.05
h1aWTh4a	−1401.4 ± 3.7	−1377.9 ± 3.9	−0.023 ± 0.005	3.15 ± 0.12
h1aWTh4b	−1536.6 ± 3.5	−1509.3 ± 2.6	−0.025 ± 0.004	1.98 ± 0.02

[a] Average folded and unfolded energies, energy difference per atom, and average Cα-RMSD for helices 1 and 4 at 300 K are shown. Reported uncertainty estimates are one standard deviation. Italic entries were expressed and studied experimentally.

were chosen for expression. The two sequences λ_{blue1} = h1a-(GSGSG)h4b and λ_{blue2} = h1a(GSGS)h4b had the lowest folded energy and also satisfied the stability criterion, having per-atom energy gaps of at least −0.020 kcal/mol/atom. The third sequence, λ_{blue3} = h1b(GSGS)h4b, had only half the energy gap per atom. We synthesized it as a control to see if computed stability gaps and experimental melting points would correlate. Figure 1 shows the computed structures of the three chosen variants.

For protein expression, the λ_{blue1} gene was inserted in a pET15-b vector, and the protein was expressed as described in Computational and Experimental Methods. The other two variants were made by site-directed mutagenesis and expressed similarly. A fourth variant (h1cGSGh4b) was also attempted but did not express well, possibly because the higher hydrophobicity of the longer fragment induced more severe proteolysis in the expression host. Expression of the designed protein thus becomes an additional practical selection criterion.

λ_{blue1} Has Helical Structure and Unfolds Cooperatively. The three variants we expressed were sufficiently soluble to carry out equilibrium experiments. CD spectra for each λ variant were collected at 20 °C from 200 to 250 nm (Figure 2). The CD spectrum of λ_{blue1} had the most negative ellipticity. The CD profile of λ_{blue1} showed characteristic α-helix peaks at 209 and 222 nm. The deeper peak at 222 nm is indicative of a high helix to coil ratio. In contrast, the CD profile of λ_{blue2} had a deeper peak at 210 nm, indicative of increased random coil. The small CD signal of λ_{blue3} with a single minimum at 218 nm had no α-helical signature. The mean residue ellipticity of λ_{blue1} was at least −8000° M^{-1} m^{-1} at 222 nm (Computational and Experimental Methods).

Thermal unfolding of the three constructs was measured by circular dichroism at 222 nm, as shown in Figure 2 (Computational and Experimental Methods). The CD melt of λ_{blue1} exhibited a cooperative melting transition with a melting temperature of T_m = 62 ± 2 °C and $\Delta S(T_m)$ = 0.41 ± 0.01 kJ/mol/K (Table 3 and Computational and Experimental Methods). The CD melt of λ_{blue3} yielded a T_m of 47 ± 3 °C but similar cooperativity. The CD melt for λ_{blue2} could not be fitted accurately because of the high temperature baseline, but visual inspection of Figure 2 indicates that it lies between the other two molecules.

Thermal unfolding was also measured by integrated fluorescence intensity, as shown in Figure 3 (Computational and Experimental Methods). Lambda repressor fragments without the Q33Y residue are known to have only a small change in tryptophan fluorescence upon folding. For λ_{blue1} the transition could be observed (Figure 2 and inset), with T_m = 65 ± 2 °C.

dx.doi.org/10.1021/jp110175x |J. Phys. Chem. B 2011, 115, 2090–2096

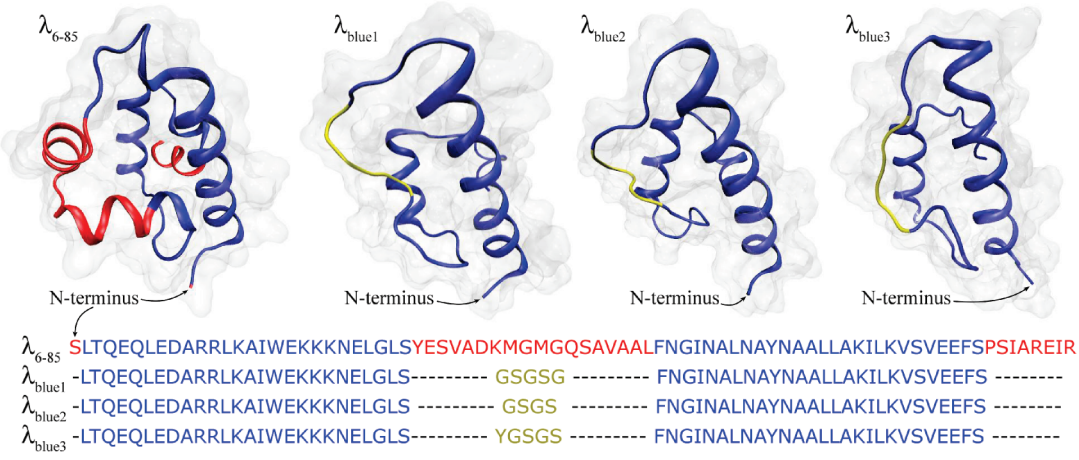


Figure 1. Model structure of λ^*_{6-85} based on the X-ray crystal structure of a closely related lambda repressor fragment PDB 3KZ3,[5,20] and molecular dynamics models of the structures for $\lambda_{blue1-3}$ computed in the present work. A van der Waals surface is also shown for each model, along with their sequences.

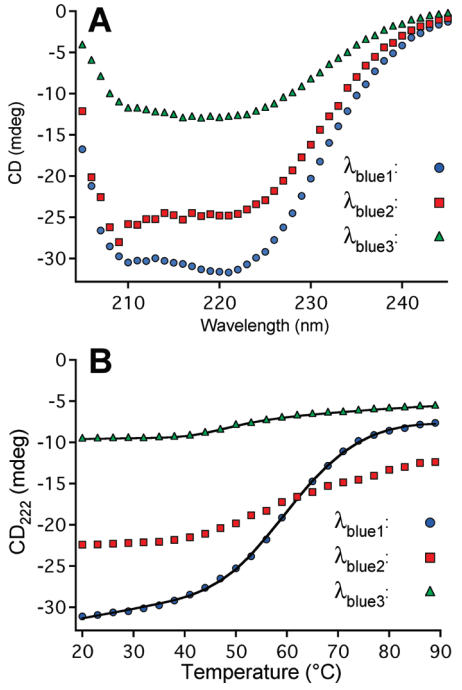


Figure 2. Circular dichroism spectra (A) and thermal denaturations for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 3.

Table 3. Fitting Results for the Thermodynamic Titration Experiments (± errors are two standard deviations of the mean)

		protein	
experiment	parameters	λ_{blue1}	λ_{blue3}
CD melt	T_m (°C)	62 ± 2	47 ± 3
	$\Delta S(T_m)$ (kJ/mol °C)	0.406 ± 0.012	0.57 ± 0.16
fluorescence melt	T_m (°C)	65 ± 2	45 ± 2
	$\Delta S(T_m)$ (J/mol °C)	0.406 (fixed)	0.47 ± 0.08

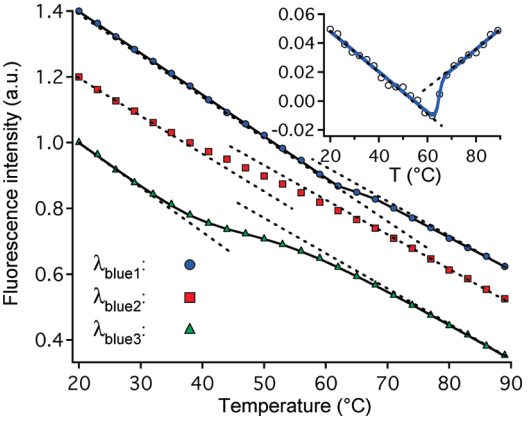


Figure 3. Fluorescence intensity-detected thermal denaturation traces for $\lambda_{blue1-3}$. Thermodynamic fits are shown as black lines in the bottom plot, with values for the fitting parameters in Table 3. The inset shows the λ_{blue1} transition with the average overall baseline subtracted to reveal the transition more clearly.

This melting temperature differs by 3 °C from the CD measurement, but considering the very small fluorescence signal and uncertainties, this difference is probably not significant. For λ_{blue2}, the transition lies at about 50 °C. The λ_{blue3} sequence contains the Q33Y mutation, and a much larger fluorescence change is observed, indicating that despite the nonhelical CD spectrum, an interaction between W22 and Y33 is formed upon folding. The melting temperature is 45 ± 2 °C, within measurement uncertainty of the CD transition midpoint.

λ_{blue1} Fragments Are Less Helical and Not Cooperative. The above results beg the question whether the apparent cooperativity and secondary structure is really a property of the whole miniprotein, or just independent structure and folding of helices 1 and 4. We purchased fragments 1−26 (helix 1) and 32−58 (helix 4) of λ_{blue1} and measured their CD spectra and thermal titration. Both fragments show much deeper peaks at 205−208 nm in the CD spectrum, indicative of more random coil content (Figure 4). Both fragments, as well as their 1:1 mixture, show no cooperative CD transition. Fragment 1−26 has a broad transition at much lower temperature than the full miniprotein. Most

dx.doi.org/10.1021/jp110175x |J. Phys. Chem. B 2011, 115, 2090–2096

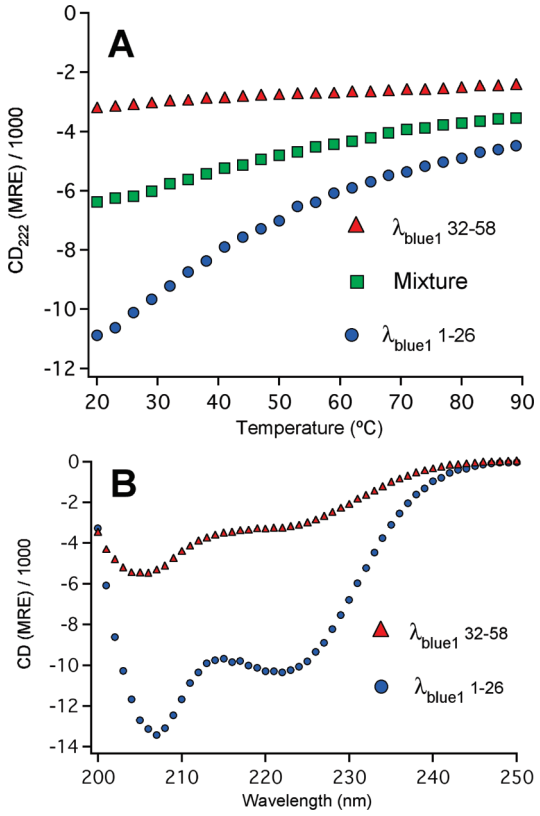


Figure 4. CD spectra and thermal melts of N terminal and C terminal fragments of λ_{blue1}. The mixture of both peptides in A simply produces an average of the two individual fragment melting curves. The spectra in B show much more random coil content (negative CD below 210 nm) than full λ_{blue1} in Figure 2.

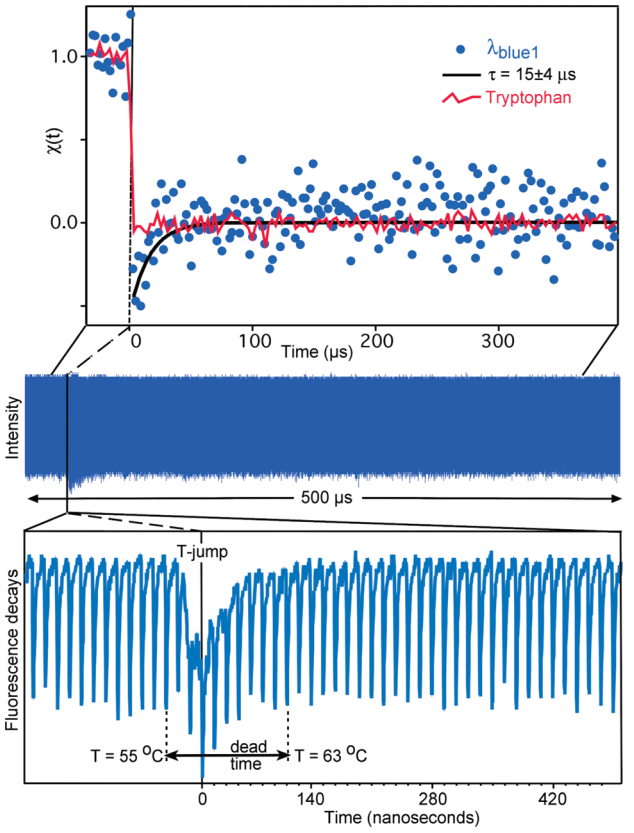


Figure 5. Temperature jump experiment: 0.5 ms of data containing over 30 000 tryptophan fluorescence decays were collected (the blue band shows fluorescence intensity), with a detail of the raw data near the T-jump shown below. Raman scatter from the infrared pump beam obscures a few fluorescence decays, setting the dead time of ≈50 ns. The top graph shows the fluorescence lifetime analysis $\chi(t)$ for λ_{blue1} (15 μs fitted response) and for tryptophan (instantaneous response). The 15 μs relaxation phase is also visible in the intensity data near $t = 0$.

notably, a mixture of the two fragments shows no enhanced cooperativity, indicating that the entropic constraint provided by the linker is important for folding.

Temperature Jump Experiments. Since λ_{blue1} had a CD spectrum consistent with helical structure, a melting point similar to the pseudo wild type, and a cooperative fluorescence transition, we performed a laser temperature jump experiment. Protein solutions (20−30 μM) were jumped by 8−10 °C, and relaxation kinetics were detected by the picosecond time-resolved change in fluorescence decay (see Computational and Experimental Methods). We observed a relaxation phase that could be fitted to τ_{obs} = 15 ± 4 μs at a final temperature of 63 ± 1 °C. No further relaxation was observed up to 0.5 ms (Figure 5). Kinetics of λ_{blue2} folding could not be resolved. λ_{blue3} showed a single fast phase (1.3 ± 0.4 μs) when jumped to a final temperature of 49 ± 1 °C.

DISCUSSION

As expected based on the computed per atom energy gap between the folded and unfolded states, the control λ_{blue3} turned out to be considerably less stable than λ_{blue1} and λ_{blue2}. Although it had one of the smallest computed Cα-RMSD values (Table 2), it has low helix content and the lowest melting point (45−47 °C) observed among the three constructs. The small difference in the computed energy gaps of λ_{blue1} and λ_{blue2} did not enable them to be ranked correctly, but λ_{blue2} had the larger computed RMSD and turned out to be less helical than λ_{blue1} by CD spectroscopy.

Of the three peptides studied, only λ_{blue1} had all of the features of a small helix bundle protein: cooperative CD and fluorescence melts, a CD spectrum with a minimum at 222 nm deeper than at 210 nm, a high T_m, and a large cooperativity parameter ΔG_1. The full λ^*_{6-85} has a melting temperature of 61 °C under the same solvent conditions, and a fitted $\Delta S(T_m)$ = 0.68 kJ/mol/K.[22] In Table 3, we find 62 °C and 0.41 kJ/mol/K for λ_{blue1}. The slightly smaller cooperativity is expected for a smaller molecule. Based on a simple model of linear scaling of protein heat capacity with chain length,[23] one would predict $\Delta S(\lambda_{blue1}) = \Delta S(\lambda^*_{6-85})N(\lambda_{blue1})/N(\lambda^*_{6-85}) \approx 0.49$ kJ/mol/K. Thus, λ_{blue1} is nearly as stable and cooperative a folder as λ^*_{6-85} when correcting for the smaller protein size. Moreover, the results for isolated helices shown in Figure 4 prove that helical structure and cooperative folding are really properties of the miniprotein, not of the N- and C-terminal helices alone.

The folding relaxation time of λ_{blue1} near its melting temperature (15 μs at 63 °C) is also very close to λ^*_{6-85} (17 μs at 61 °C). The linker connecting helices 1 and 4 in λ_{blue1} is shorter than in the pseudo wild type (5 vs 18 residues). On the basis of entropic considerations alone (random chain linkers), one would have expected λ_{blue1} to fold slightly faster than λ^*_{6-85}. The comparable relaxation times indicate one of two things: either the contact energy driving term of λ_{blue1} is smaller than that of the pseudo

wild type, or helices 2 and 3 in the wild type form a structured and compact linker during helix 1−4 folding.

The nearly identical stability and fast folding of λ_{blue1} relative to λ^*_{6-85} suggest that the formation and docking of helices 1 and 4 is both the principal thermodynamic stabilizing interaction and kinetic docking step during folding. This obervation confirms the mutation and molecular dynamics results that led to the present work,[8] which suggested that helices 1 and 4 are the earliest to form and make contacts. Oas and co-workers suggested a transition state without helix 3 in the rate-limiting step,[24] and our result is consistent with that analysis. However, it appears that helix 2 is not as important in the early kinetics and contact formation as suggested by their collision-diffusion model.

Helices 1 and 4 lie close to the N and C termini of the full λ^*_{6-85} fragment. A number of small proteins form initial contacts between two remote helices separated by long loops (e.g., apomyoglobin[25] and cytochrome c[26]). The rapid folding kinetics of λ^*_{6-85}, and identification of helices 1 and 4 as its folding core, together suggest that early contact formation between two remote helices is also predominant there. As long as the greater reduction of loop entropy during collision of remote helices is compensated by optimal contact energies between these helices, the free energy penalty associated with this mode of structure formation can be ameliorated. It remains unclear why the 'remote' mechanism should be preferred in vitro over a mechanism where adjacent helices initiate folding. We speculate that there could be a reason in vivo. A small contact energy between 'center helices' and a large contact energy between 'outer' helices could prevent early collapse and non-native contact formation until an entire protein subunit is available for folding. There is certainly evidence in some cases for sequence-control during cotranslational folding to improve in vivo foldability of proteins.[27] Such an explanation is appealing for whole proteins (e.g., apomyoglobin, apocytchrome c) but could also apply to internal subunits of proteins (e.g., lambda repressor) if they fold independenly, which λ^*_{6-85} and λ_{1-101} clearly do.

With a relaxation time of $\tau \geq 15\ \mu s$, the reduced lambda repressor is still an approximate two-state folder by the criterion of fast/slow kinetic phase amplitudes from ref 28. However, it is not far from being an incipient downhill folder with a barrier of $<3k_BT$ based on a survey of many lambda repressor mutants.[5] In the present work, we left hydrophobic residues in helices 1 and 4 unmodified even if they were no longer required for native contact formation. It may be possible to increase the folding rate further by reducing non-native hydrophobic contacts. It certainly should be possible to improve solubility by eliminating some of the most solvent-exposed hydrophobic side chains. We predict Y60 and L64 on helix 4 as prime candidates for replacement by similarly sized polar residues. However, for comparison with molecular dynamics simulations contrasting a reduced folding core with a full small protein, λ_{blue1} is the best target as it most closely resembles the sequence of the original protein.

■ ASSOCIATED CONTENT

Supporting Information. A description of the explicit solvent simulations of wild-type λ^*_{6-85} along with a figure showing helical stability as a function of temperature, tables showing the average radius of gyration of each simulated variant, sequence listings for each variant, and files containing representative coordinates of λ_{blue1}, λ_{blue2}, and λ_{blue3}. This material is available free of charge via the Internet at http://pubs.acs.org

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: gruebele@scs.uiuc.edu.

■ ACKNOWLEDGMENT

This work was supported by the IBM Blue Gene Project (J.P., D.L., and W.S.), and by National Science Foundation grant MCB-1019958 (M.G., M.B.P., and K.S.).

■ REFERENCES

(1) Hecht, M. H.; Nelson, H. C. M.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 2676.
(2) Lim, W. A.; Hodel, A.; Sauer, R. T.; Richards, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 423.
(3) Huang, G. S.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6878.
(4) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305.
(5) Liu, F.; Gao, Y.-G.; Gruebele, M. *J. Mol. Biol.* **2009**, *397*, 789.
(6) Dumont, C.; Matsumura, Y.; Kim, S. J.; Li, J.; Kondrashkina, E.; Kihara, H.; Gruebele, M. *Protein Sci.* **2006**, *15*, 2596.
(7) Hecht, M. H.; Sturtevant, J. M.; Sauer, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 5685.
(8) Larios, E.; Pitera, J. W.; Swope, W. C.; Gruebele, M. *Chem. Phys.* **2006**, *323*, 45.
(9) Myers, J. K.; Oas, T. G. *Nat. Struct. BIol.* **2001**, *8*, 552.
(10) Yang, W.; Larios, E.; Gruebele, M. *J. Am. Chem. Soc.* **2003**, *125*, 16220.
(11) Yang, W. Y.; Gruebele, M. *Phil. Trans. R. Soc. London B* **2005**, *43*, 13018.
(12) Ghaemmaghami, S.; Word, J. M.; Burton, R. E.; Richardson, J. S.; Oas, T. G. *Biochemistry* **1998**, *37*, 9179.
(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R., Jr.; K. M., M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(14) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.
(15) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878.
(16) Andersen, H. C. *J. Chem. Phys.* **1980**, *72*, 2384.
(17) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(18) Ervin, J.; Sabelko, J.; Gruebele, M. *J. Photochem. Photobiol.* **2000**, *B54*, 1.
(19) Beamer, L. J.; Pabo, C. O. *J. Mol. Biol.* **1992**, *227*, 177.
(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(21) Myers, J. K.; Oas, T. G. *Biochemistry* **1999**, *38*, 6761.
(22) Yang, W. Y.; Gruebele, M. *Biochemistry* **2004**, *43*, 13018.
(23) Ghosh, K.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10649.
(24) Burton, R. E.; Myers, J. K.; Oas, T. G. *Biochemistry* **1998**, *37*, 5337.
(25) Jennings, P.; Wright, P. *Science* **1993**, *262*, 892.
(26) Xu, Y.; Mayne, L.; Englander, S. W. *Nat. Struct. Biol.* **1998**, *5*, 774.
(27) M.S., E.; I.M., S.; P.L., C. *J. Mol. Biol.* **2008**, *383*, 683.
(28) Yang, W.; Gruebele, M. *Biophys. J.* **2004**, *87*, 596.