Ultrafast Folding of a Computationally Designed Trp-Cage Mutant: Trp^2-Cage†

Michelle R. Bunagan, Xi Yang, Jeffery G. Saven,* and Feng Gai*

Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received: September 17, 2005

Miniproteins provide useful model systems for understanding the principles of protein folding and design. These proteins also serve as useful test cases for theories of protein folding, and their small size and ultrafast folding kinetics put them in a regime of size and time scales that is now becoming accessible to molecular dynamics simulations. Previous estimates have suggested the "speed limit" for folding is on the order of 1 μs. Here a computationally designed mutant of the 20-residue Trp-cage miniprotein, Trp^2-cage, is presented. The Trp^2-cage has greater stability than the parent and folds on the ultrafast time scale of 1 μs at room temperature, as determined from infrared temperature-jump experiments.

Introduction

There is growing interest in identifying small proteins that fold on microsecond time scales under physiological conditions.[1] Such proteins provide simple model systems for understanding the principles underlying protein folding and can be readily synthesized, which facilitates a wide variety of experiments. These proteins also serve as useful test cases for current theories of protein folding,[2] and their small size and ultrafast folding kinetics put them in a regime of size and time scales that is now becoming accessible to molecular dynamics simulations.[3] To the extent that such systems obey Arrhenius or Kramer type kinetics, identifying fast folders can put constraints on the preexponential term of the folding rate, i.e., the conformational reconfiguration time among structures in the transition-state ensemble.[4,5] Previous estimates have suggested this reconfiguration time or "folding speed limit" is on the order of 1 μs.[4,5,6] Here we present a computationally designed Trp-cage mutant that folds on this time scale.

The 20-residue Trp-cage (Figure 1) reported by Andersen and co-workers[7,8] folds cooperatively in a two-state manner with a melting temperature (T_m) of 42 °C at pH 7. These authors suggest that the four Pro residues are important for the stability of this small folding motif; the hydrophobic interaction between Trp and Pro is stabilizing, whereas the Pro residues also place restrictions on backbone conformations and help to decrease the conformational entropy of the unfolded state. Qiu et al.[9] have shown that Trp-cage folds in about 4 μs at room temperature, which renders it one of the fastest folding proteins known to date.[1] Consequently, many computational studies[10−17] have been published, aiming to understand the folding pathways and free energy landscape of Trp-cage.

While Pro plays a critical role in the folding of Trp-cage, it appears in this miniprotein much more frequently than in natural proteins. Therefore, it is not clear if every single Pro in Trp-cage is necessary for folding. For instance, additional intrachain interactions and increased flexibility of the chain resulting from a mutation of one or more of the Pro residues may stabilize the protein and affect the folding rate.[18] However, in light of the fact that Trp-cage is small and its stability is the result of several exquisitely balanced interactions, random mutations are likely to fail. Herein, we employed a computational design method to identify structurally consistent mutations that may increase the thermal stability as well as the folding rate of the resulting sequences. Other studies have found computationally designed proteins to be among the fastest folding proteins known, with folding times in the range of 3−30 μs.[19,20] Previously, we have used computational design to identify tailored mutations that increase hydrophobic content, resulting in a 47-residue three-helix-bundle protein with a $(1\ \mu s)^{-1}$ folding rate.[21,22] Herein we apply the method to the a much smaller miniprotein, the 20-residue Trp-cage, and investigate the folding stability and kinetics of a designed mutant.

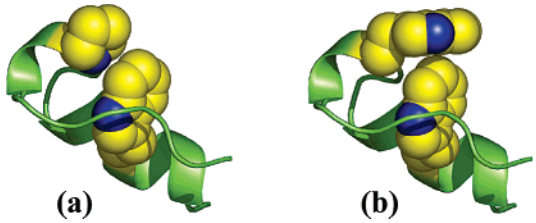

Figure 1. (a) Wild-type Trp-cage (1L2Y).[7] The sequence of the Trp-cage is NLYIQWLKDGGPSSGRPPPS. (b) Calculated Trp^2-cage (i.e., P12W mutant of Trp-cage). In each case, side chains 6 and 12 are rendered as space-filling.

Materials and Methods

Design Method. Identifying structurally consistent mutants of Trp-cage was accomplished using a statistical design method, which yields the site-specific probabilities of the amino acids at variable positions in a particular structure.[23,24] Specifically, the method involves a Monte Carlo sampling of sequences.[25,26] The energies of interatomic interactions were determined using a form of the AMBER potential with a modified hydrogen-bonding term and a distant-dependent dielectric constant ($\epsilon = 4r_{ij}$, where r_{ij} is the distance between atom i and atom j).[27,28] Interatomic pair interactions were truncated at a maximum of 30 kcal/mol. No solvation-related energies or scoring functions were used in the calculations. The backbone coordinates and the identities of unmutated residues were constrained, while at variable positions multiple amino acid identities and side-

† Part of the special issue "Michael L. Klein Festschrift".

* Authors to whom correspondence should be addressed. E-mail: saven@sas.upenn.edu; gai@sas.upenn.edu.

10.1021/jp055288z CCC: $33.50
Published on Web 12/01/2005

chain conformations (rotamers) were permitted. The Dunbrack rotamer library[29] with at most 10 rotamers per amino acid was used. The Monte Carlo sampling is guided by a configurational biasing and replica exchange algorithm (BMCREM) to ensure representative sampling of sequences and rotamers.[26] Thirty-three replicas were used, spanning temperatures from $T = 0.6$ kcal/mol to $T = 103$ kcal/mol, and the temperature distribution was chosen so as to achieve exchange ratios between neighboring temperatures of 0.2−0.3. At each Monte Carlo step, the algorithm attempted to vary the rotamer and/or identity of five distinct residues. Attempted exchanges between temperatures occurred every 1000 steps. After an initial equilibration of 10^5 steps, sequences were collected every other step, with the total trajectory (at each temperature) of 6×10^5 steps. Probabilities of the different rotamer states were determined from sequences sampled at $T = 1.5$ kcal/mol. The site-specific probabilities of individual amino acids were determined by summing over the probabilities of the site-specific rotamer states of the amino acids. The coordinates for the Trp-cage were taken from the Protein Data Bank (PDB code 1L2Y).

Experimental Methods. The proteins used in the current study were synthesized by employing the standard Fmoc-protocol and purified by reverse phase high-performance liquid chromatography (HPLC). The identities of the samples were verified by matrix-assisted laser desorption ionization mass spectroscopy. Residual trifluoroacetic acid (TFA) from the synthesis was removed by multiple lyophilizations against a 0.1 M DCl solution, which also allowed amide hydrogen−deuterium (H−D) exchange. A complete H−D exchange is important because the amide vibrational modes of deuterated protein samples are often different from those of hydrogenated protein samples. Static IR spectra were collected on a Nicolet Magna 860 spectrometer equipped with a mercury cadmium telluride (MCT) detector. A thermostated sample cell with an optical path length of 52 μm and two compartments allowed the measurements of both the sample and the buffer under the same conditions. The protein samples used in both equilibrium and time-resolved IR measurements were prepared by dissolving the lyophilized solids directly in 50 mM phosphate buffer solutions (D_2O, pH* = 7.0). The final concentration was 2−3 mM. Circular dichroism (CD) data were collected on an AVIV 62DS spectropolarimeter using a 1 mm quartz cell. The protein concentration used in the CD measurements was about 40−60 μM, which was determined optically by the tryptophan absorbance at 280 nm using $\epsilon_{280} = 6970$ cm^{-1} M^{-1} for the wild-type and $\epsilon_{280} = 12\,660$ cm^{-1} M^{-1} for the Trp^2-cage. The one-dimensional (1D) NMR spectra of Trp-cage and Trp^2-cage were collected on a Bruker DMX 600 MHz spectrometer at 30 °C. The concentration of the sample was 1 mg/mL in H_2O/D_2O phosphate buffer solution (20 mM, pH 7).

The temperature-jump IR setup has been described in detail previously.[30] Briefly, the 1.9 μm temperature-jump pulse was generated via Raman-shifting the fundamental output of a Nd:YAG laser in H_2. The temperature-jump-induced transient absorbance changes were measured by a continuous wave (CW) IR diode laser in conjunction with a 50 MHz MCT detector and a digital oscilloscope. Similar to that used in the equilibrium experiments, a thermostated two-compartment sample holder was employed to allow the separate measurements of the transient signals of both the sample and the buffer under identical conditions. The buffer measurements provide information for both background subtraction and temperature-jump amplitude determination.

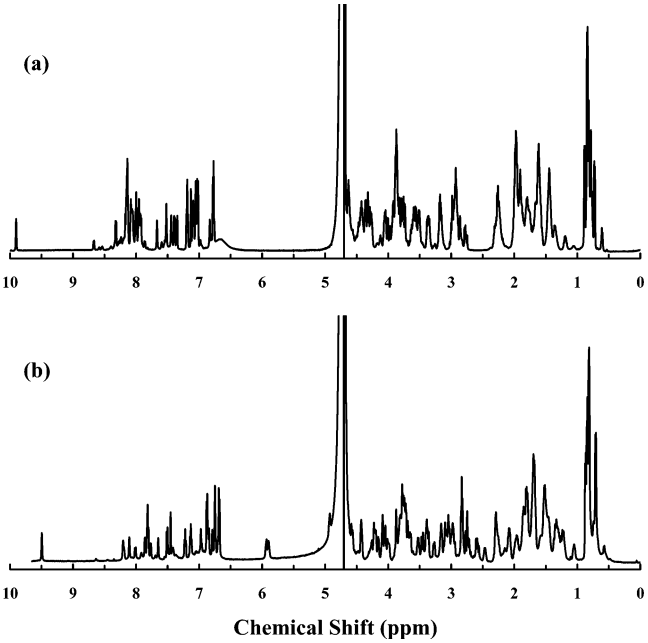

Figure 2. NMR spectra of (a) Trp-cage and (b) Trp^2-cage. The spectra of Trp-cage and Trp^2-cage at 30 °C show well-dispersed narrow lines, indicating that both peptides are folded at this temperature.

Results and Discussion

The biased Monte Carlo replica exchange method (BMCREM)[26] was applied to an atom-based model of the Trp-cage protein[7] (PDB code 1L2Y). The method yields the site-specific probabilities of the amino acids (and their side-chain rotamer states) at variable positions for a fixed set of backbone coordinates. The identities of several residues contacting the central Trp were fixed (N1, Y3, W6, G11, P17, P18, and P19). The remaining positions were subjected to mutation to any of 19 natural amino acids (Cys was not included). In particular, site 12 was varied, since an unexpected change in the chemical shift dispersion has been observed for P12 upon melting.[7] "Frustrated" positions were identified using $r = \ln(w_{mp}/w_{wt})$, where w_{mp} and w_{wt} are the probabilities of the most probable and wild-type amino acids at a variable residue position.[21] In a second calculation, only the most frustrated sites (those having the largest values of r), sites 9 and 12, were allowed to mutate with the remainder of the protein fixed at the original Trp-cage identities, yielding two calculated mutations: P12W (Trp^2-cage) and the conservative D9E. Given the very conservative nature of the D9E mutation on the exterior of the protein, which maintains the charge of the residue, we opted to study in detail the less intuitive P12W mutation and reserve study of D9E and other mutations to future work. In the resulting structural model, W12 has an edge-to-face stacking interaction with the central W6 (Figure 1). This interaction may stabilize the structure at the expense of losing a Pro that conformationally constrains the main chain. To examine whether the mutation to Trp is site-specific and provide a comparison, the mutation R16W was identified as unfavorable in a separate design calculation. The R16W substitution disrupts the salt bridge with Asp at site 9 in the wild-type structure.

The native Trp-cage and two of the computationally designed mutants (i.e., P12W and R16W) were characterized by several spectroscopic methods. The 1D proton NMR spectrum of Trp^2-cage collected at 30 °C shows well-dispersed narrow lines (Figure 2), indicating that it is folded at this temperature.

Following Neidigh et al.,[7] we have also characterized the thermal unfolding transition of Trp^2-cage using CD spectros-

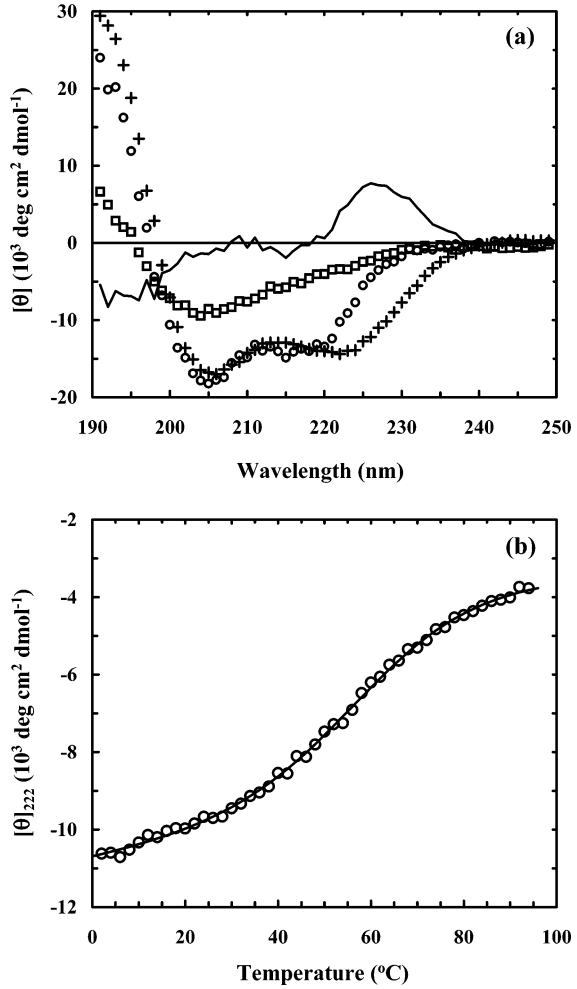

Figure 3. (a) CD spectra of Trp-cage (+), Trp^2-cage (○), and R16W (□) at 4 °C. The solid line corresponds to the difference between the CD signals of Trp-cage and Trp^2-cage. (b) Mean residue ellipticity (○) of Trp^2-cage as a function of temperature. Fitting these data to an apparent two-state model (solid line) yields the following thermodynamic parameters for folding: $T_m = 56.9$ °C, $\Delta H_m = -16.1$ kcal mol^{-1}, $\Delta S_m = -48.8$ cal mol^{-1} K^{-1}, and $\Delta C_p = -148$ cal mol^{-1} K^{-1}.

copy. As indicated (Figure 3a), the far-UV CD spectrum of Trp^2-cage collected at 4 °C is almost identical to that of Trp-cage, except in the wavelength region of 215−240 nm. This difference is presumably due to the positive CD signal originating from excitonic coupling[31] between W6 and W12. It has been shown that an edge-to-face stacking interaction between two aromatic residues can give rise to a positive CD band centered at about 229 nm.[31] Indeed, the difference CD spectrum generated by subtracting the CD spectrum of Trp-cage from that of Trp^2-cage exhibits a positive maximum at about 227 nm, similar to that observed for the Trpzips[32] which also contain paired Trp residues stacked in an edge-to-face manner. This edge-to-face interaction is the most common relative orientation of the two side chains,[33] and a recent computational study by Brooks and co-workers[34] indicated that the electrostatic multipole moments of the indole ring are particularly important in determining the edge-to-face orientation of Trp side chains in proteins.

The thermal melting temperature (T_m) of Trp^2-cage, obtained by fitting the temperature-dependent ellipticities at 222 nm to an apparent two-state model,[7] increases to ∼57 °C (Figure 3b). This increase in stability is fully consistent with the computational result and also suggests a strong interaction between the two Trp residues. The latter has been used recently to stabilize designed β-hairpins.[32] Also consistent with the computational

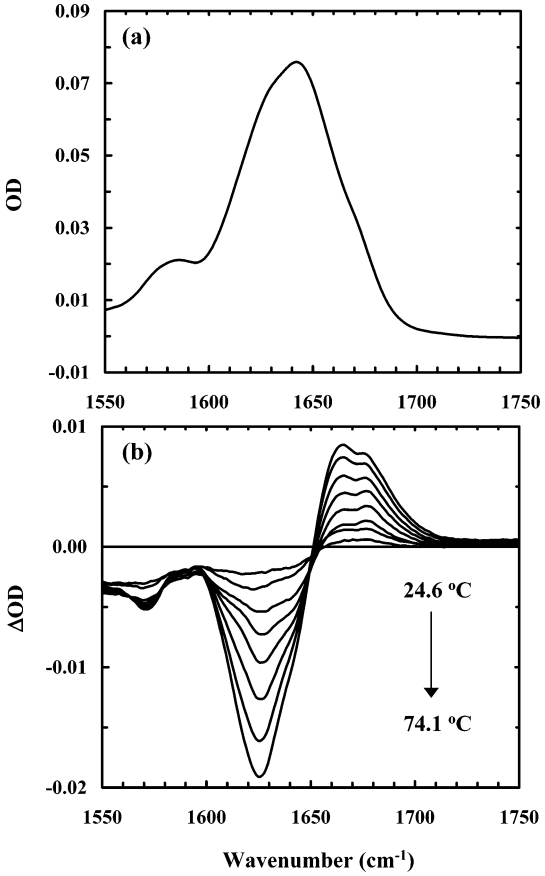

Figure 4. (a) A representative FTIR spectrum of Trp^2-cage in D_2O phosphate buffer solution (pH* = 7) measured at 17.4 °C. (b) Difference FTIR spectra of Trp^2-cage that were generated by subtracting the spectrum collected at 17.4 °C from those collected at higher temperatures, as indicated.

result, the CD spectrum of the negatively designed mutant, R16W, lacks the characteristic features of helical structures (Figure 3a), indicating that this mutant is not folded. Taken together, these results suggest that computations are especially useful for guiding the search for structurally consistent mutants, even for small folding motifs.

The thermal unfolding transition of Trp^2-cage was also studied by Fourier transform IR (FTIR) spectroscopy. Specifically, the amide I′ band of Trp^2-cage, which arises mostly from the stretching vibration of the backbone carbonyls and is sensitive to structural determinants,[35] was collected as a function of temperature. As shown (Figure 4), the amide I′ band of Trp^2-cage measured at 17.4 °C exhibits typical features of helical proteins, whereas its difference FTIR spectra show that the helical components of Trp^2-cage unfold with the increase of temperature. Globally analyzing[21] the temperature-dependent amide I′ bands of Trp^2-cage collected between 17.4 and 74.1 °C yielded a thermal melting temperature of 56.7 °C, suggesting that both CD and IR probe the same thermal unfolding process. More importantly, these FTIR spectra show that no aggregation occurs at the concentration used and to the highest temperature, monitored by the distinct spectral features at 1618 and 1680 cm^{-1} associated with aggregates.[36]

We studied the folding kinetics of the Trp^2-cage by employing an infrared laser-induced temperature-jump (T-jump) method.[30] As shown (Figure 5a), the T-jump-induced relaxation kinetics of the Trp^2-cage, probed at 1625 cm^{-1}, where helices absorb,[35] exhibit two distinct phases. The fast phase (∼15 ns) is instrumentation-limited and, as discussed previously,[21] likely

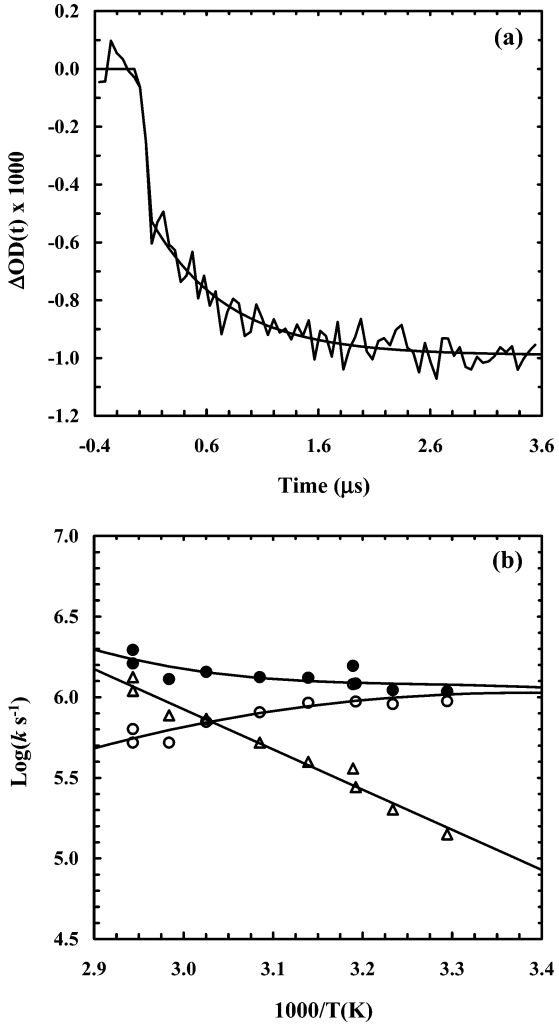

Figure 5. (a) Representative T-jump relaxation trace in response to a T-jump from 58.6 to 66.6 °C, probed at 1625 cm^{-1}. The smooth line is the fit to the following function, $\Delta OD(t) = A[1 - B \exp(-t/\tau_1) - C \exp(-t/\tau_2)]$, with $A = -0.00099$, $B = 0.48$, $C = 0.59$, $\tau_1 = 0.015$ μs, and $\tau_2 = 0.68$ μs. (b) Arrhenius plot of the observed (●), folding (○), and unfolding (△) rate constants of Trp^2-cage. Lines are fits to the Eyring equation, i.e., $\ln(k) = \ln(D) - \Delta G^{\ddagger}/RT$, where D is a constant and $\Delta G^{\ddagger}$ is the free energy of activation that is a function of temperature, as described by the following function: $\Delta G^{\ddagger} = \Delta H^{\ddagger}(T_m) + \Delta C_p^{\ddagger}(T - T_m) - T[\Delta S^{\ddagger}(T_m) + \Delta C_p^{\ddagger} \ln(T/T_m)]$, where T_m is the thermal melting temperature, 56.9 °C. Global fitting of the folding and unfolding kinetics yields the following thermodynamic parameters of activation for folding when $D = 10^{10}$ s^{-1} is used: $\Delta H^{\ddagger}(T_m) = -4.8$ kcal mol^{-1}, $\Delta S^{\ddagger}(T_m) = -33.6$ cal mol^{-1} K^{-1}, $\Delta C_p^{\ddagger} = -140$ cal mol^{-1} K^{-1}. It is worth pointing out that the value of $\Delta S^{\ddagger}(T_m)$ is not absolute because the prefactor D is arbitrarily set to 10^{10} s^{-1}.

due to temperature-induced spectral changes that are not related to the global folding/unfolding of the Trp^2-cage conformation. Within our experimental uncertainties, the slow relaxation phase can be modeled by a single-exponential function, suggesting that kinetically Trp^2-cage also folds in a two-state manner. This is in agreement with the results of Qiu et al,[9] who also observed first-order T-jump relaxation kinetics for Trp-cage. For a two-state folding scenario, the observed relaxation rate constant (k_R) is simply the sum of the folding rate constant (k_F) and unfolding rate constant (k_U). Therefore, k_F and k_U can be determined in two-state folding if the equilibrium constant K_{eq} is known, since $K_{eq} = k_F/k_U$. Hence, using the equilibrium constants obtained from CD and the measured relaxation rate constants, we were able to determine the folding and unfolding rate constants of Trp^2-cage. As shown (Figure 5b), the Trp^2-cage exhibits ultrafast folding behavior[1] with a maximum folding rate of $k_F = (0.93 \pm 0.15\ \mu s)^{-1}$ at 22.5 °C.

These findings suggest that the folding energy landscape of Trp^2-cage is probably very smooth and the free energy barrier for folding is small. Indeed, the enthalpy of activation for folding at the thermal melting temperature was determined to be negative (Figure 5b), suggesting that the folding process is downhill in energy. The folding of the Trp-cage structure brings side chains into tertiary contacts that are separated by up to 13 residues in sequence. Two residues separated by this distance in sequence would be expected to collide on the 0.3−1 μs time scale,[6] yet Trp^2-cage is able to achieve the fully folded state on the same time scale. This has in part been facilitated by "minimizing frustration"[37] via computational design.[21] In addition, the ultrafast folding rate observed for this protein suggests that the prefactor of transition-state folding models may be submicrosecond: $k_0 > (1\ \mu s)^{-1}$.

Microscopically, the stability of the folded state of a two-state folder is controlled by both the folding and the unfolding rates. Hagen and co-workers[9] have shown that the folding and unfolding rates of the Trp-cage at 22.7 °C are $(4.1\ \mu s)^{-1}$ and $(12.7\ \mu s)^{-1}$, respectively. Therefore, the increased stability observed for the Trp^2-cage around this temperature can be attributed to mostly an increased folding rate. As indicated (Figure 5b), the folding and unfolding rates of the Trp^2-cage at 22.7 °C are $(0.94\ \mu s)^{-1}$ and $(10.5\ \mu s)^{-1}$, respectively.

In summary, a computationally designed Trp-cage mutant, Trp^2-cage, shows increased thermal stability and an increased folding rate. Therefore, this study demonstrates the versatility of using the Trp−Trp interaction as a means to stabilize the folded conformation of designed sequences.[32] Furthermore, the Trp^2-cage exhibits one of the fastest protein folding rates known to date.[1] Because of its small size and fast folding kinetics, this miniprotein should provide a rich model system for further experimental and computer simulation studies.

Acknowledgment. We thank the National Institutes of Health (GM-065978, RR-01348, and GM-61267) for funding.

References and Notes

(1) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76−88.
(2) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70−75.
(3) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102−106.
(4) Portman, J. J.; Takada, S.; Wolynes, P. G. *J. Chem. Phys.* **2001**, *114*, 5082−5096.
(5) Yang, W. Y.; Gruebele, M. *Nature* **2003**, 423, 193−197.
(6) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−11617.
(7) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat Struct. Biol.* **2002**, *9*, 425−430.
(8) Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Biol.* **2002**, *9*, 408−410.
(9) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(10) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7587−7592.
(11) Chowdhury, S.; Lee, M. C.; Xiong, G.; Duan, Y. *J. Mol. Biol.* **2003**, *327*, 711−717.
(12) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.
(13) Ota, M.; Ikeguchi, M.; Kidera, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17658−17663.
(14) Chowdhury, S.; Lee, M. C.; Duan, Y. *J. Phys. Chem. B* **2004**, *108*, 13855−13865.
(15) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280−13285.
(16) Nikiforovich, G. V.; Andersen, N. H.; Fesinmeyer, R. M.; Frieden, C. *Proteins* **2003**, *52*, 292−302.

(17) Snow, C. D.; Zagrovic, B.; Pande, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 14548−14549.
(18) Eyles, S. J.; Gierasch, L. M. *J. Mol. Biol.* **2000**, *301*, 737−747.
(19) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15486−15491.
(20) Gillespie, B.; Vu, D. M.; Shah, P. S.; Marshall, S. A.; Dyer, R. B.; Mayo, S. L.; Plaxco, K. W. *J. Mol. Biol.* **2003**, *330*, 813−819.
(21) Zhu, Y.; Fu, X.; Wang, T.; Tamura, A.; Takada, S.; Saven, J. G.; Gai, F. *Chem. Phys.* **2004**, *307*, 99−109.
(22) Wang, T.; Zhu, Y.; Gai, F. *J. Phys. Chem. B* **2004**, *108*, 3694−3697.
(23) Zou, J.; Saven, J. G. *J. Mol. Biol.* **2000**, *296*, 281−294.
(24) Kono, H.; Saven, J. G. *J. Mol. Biol.* **2001**, *306*, 607−628.
(25) Zou, J.; Saven, J. G. *J. Chem. Phys.* **2003**, *118*, 3843−3854.
(26) Yang, X.; Saven, J. G. *Chem. Phys. Lett.* **2005**, *401*, 205−210.
(27) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(28) Kono, H.; Doi, J. *J. Comput. Chem.* **1996**, *17*, 1667−1683.
(29) Dunbrack, R. L. J.; Cohen, F. E. *Protein Sci.* **1997**, *6*, 1661−1681.
(30) Huang, C. Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2788−2793.
(31) Grishina, I. B.; Woody, R. W. *Faraday Discuss.* **1994**, *99*, 245−262.
(32) Cochran, A. G., Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 5578−5583.
(33) Samanta, U.; Pal, D. Chakrabarti, P. *Acta Crystallogr., Sect. D* **1999**, *55*, 1421−1427.
(34) Guvench, O.; Brooks, C. L., III *J. Am. Chem. Soc.* **2005**, *127*, 4668−4674.
(35) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181−364.
(36) Timasheff, S. N.; Susi, H.; Stevens, L. *J. Biol. Chem.* **1967**, *242*, 5467−5473.
(37) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.