J|A|C|S
COMMUNICATIONS

Published on Web 10/12/2002

Smaller and Faster: The 20-Residue Trp-Cage Protein Folds in 4 μs

Linlin Qiu,[†] Suzette A. Pabit,[†] Adrian E. Roitberg,[‡,§] and Stephen J. Hagen*[,†]

Physics Department, Chemistry Department, and Quantum Theory Project, University of Florida, P.O. Box 118440, Gainesville Florida 32611-8440

Received July 29, 2002

The increasing numbers of small proteins currently known[1] to fold at rates exceeding 10^4 s^{-1} will provide the testing ground for all-atom molecular dynamic simulations of folding, which have just recently reached from the nanosecond to the microsecond domain.[2] However, while computation will continue extending to longer times, proteins must eventually encounter physical limitations to their folding speed. The finite speed of Brownian diffusion, for example, suggests[3] that folding speeds probably cannot much exceed $\sim 10^6$ s^{-1}. The exact limits, and the conditions under which a molecule may approach them, remain open questions. Here we report that a small, designed protein folds at a nearly diffusion-limited rate, despite a tertiary structure that might predict far slower folding.

Neidigh et al.[4] have designed a 20-residue sequence (NLYIQ WLKDG GPSSG RPPPS) that folds spontaneously and cooperatively to a "Trp-cage", a globular fold with a combination of secondary structure and tertiary contacts more typical of a larger, more complex molecule. The NMR structure (1L2Y) reported by Neidigh et al. reveals a compact hydrophobic core where three proline residues (Pro-12, Pro-18, Pro-19) and a glycine (Gly-11) pack against the aromatic side chains of Tyr-3 and Trp-6. The secondary structure elements include an α-helix extending from residues 2 through 8, followed by a 3_{10}-helix (residues 11−14), and a polyproline II helix at the C-terminus. Stability data also suggest that a salt bridge connects Asp-9 with Arg-16 in the folded state. The folding appears highly cooperative, with circular dichroism (CD), fluorescence, and chemical shift deviations generating virtually identical, sigmoidal thermal denaturation profiles. The Trp-cage therefore represents the smallest known cooperatively folding protein-like molecule.[4]

We used laser temperature-jump spectroscopy to measure the folding rate of the Trp-cage. Figure 1 shows the equilibrium unfolding curve for the Trp-cage obtained from far-UV CD and from the fluorescence of the single tryptophan residue. The weakly increasing Trp fluorescence below the unfolding midpoint ($T_f \approx$ 41 °C) shows that thermal unfolding enhances the Trp emission and tends to compensate its intrinsic, negative *T*-dependence. Our laser *T*-jump instrument time-resolves the Trp fluorescence changes that follow a rapid *T* increase. It uses a 1-m H_2 cell (Light Age) to Raman shift the 1.064 μm, 5−7 ns pulses of a Nd:YAG laser (Continuum Surelite) to 1.89 μm. Beam splitting and focusing optics separate the shifted pulses into pairs of counterpropagating pulses and focus these onto the aqueous sample, which flows within a flat-wall, fused silica capillary (100 μm i.d.). The resulting *T*-jump of 5−20 °C occurs within ~20−30 ns, and the sample temperature remains uniform and elevated for several milliseconds. A 5 ns pulse

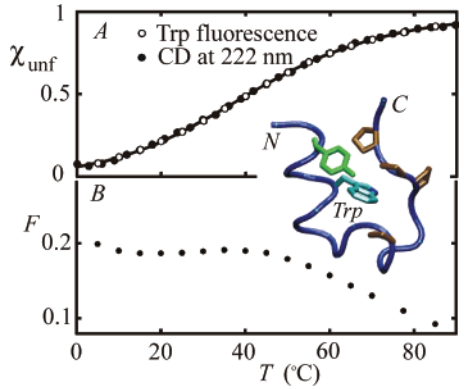

Figure 1. Equilibrium unfolding of the Trp-cage [inset]. (*A*) Fraction unfolded χ_{unf} in 15 mM phosphate, pH 7, as derived from 222-nm circular dichroism data and by Trp fluorescence data (266 nm excitation), with fit to two-state transition (solid curve) indicating $\Delta H \approx 48.6$ kJ $(mol)^{-1}$ and $\Delta S \approx 155$ J $(mol\ K)^{-1}$ for unfolding. (*B*) Wavelength-integrated fluorescence emission, with data normalized to reference (*N*-acetyl-Trp-amide) at 5 °C. Trp-cage synthesis by solid-phase method (FMOC) with reverse-phase HPLC purification (C-18 column).

from the 266 nm fourth harmonic of a Nd:YAG laser (Continuum Minilite) excites the Trp fluorescence; a microscope objective oriented at 90° projects this emission onto a photomultiplier. We assemble a complete kinetic profile from pulses measured at ~100 different pump/probe time delays in random order.

The temperature jump causes an initial drop (within ~30 ns) in the fluorescence signal, owing to the intrinsic *T* dependence of tryptophan emission (Figure 2A). The signal then increases sharply as the protein unfolds with a time constant $\tau \approx 1-10$ μs that varies according to the final temperature (but not the initial temperature). Thus, shortly after the *T*-jump, the net fluorescence has returned almost to its initial level, or even to a higher level, consistent with the generally flat or weakly rising equilibrium fluorescence curve of Figure 1B. (Solutions of free Trp and nonfolding "control" peptides exhibit no such relaxation.)

If an intermediate state lies between the folded and unfolded configurations of the Trp-cage, we expect it to generate an additional fluorescence relaxation after the microsecond process. We find no evidence for further relaxations (other than the millisecond thermal recovery), and—together with the fact that the observed relaxation amplitude matches expectations from equilibrium fluorescence data—this implies that the Trp-cage folds with two-state kinetics. Most small proteins (<100 residues) do fold with two-state kinetics. Although we cannot rule out an intermediate that reveals itself through a small-amplitude relaxation on the slow (~ms) time scale of thermal recovery, this would require that the two eigenvalues of the three-state kinetic system differ by a factor $\geq 10^3$; kinetic modeling suggests that this would conflict with the highly cooperative equilibrium unfolding of the Trp-cage[4] (i.e., the high free energy

* To whom correspondence should be addressed. E-mail: sjhagen@ufl.edu.
† Physics Department.
‡ Chemistry Department.
§ Quantum Theory Project.

10.1021/ja0279141 CCC: $22.00

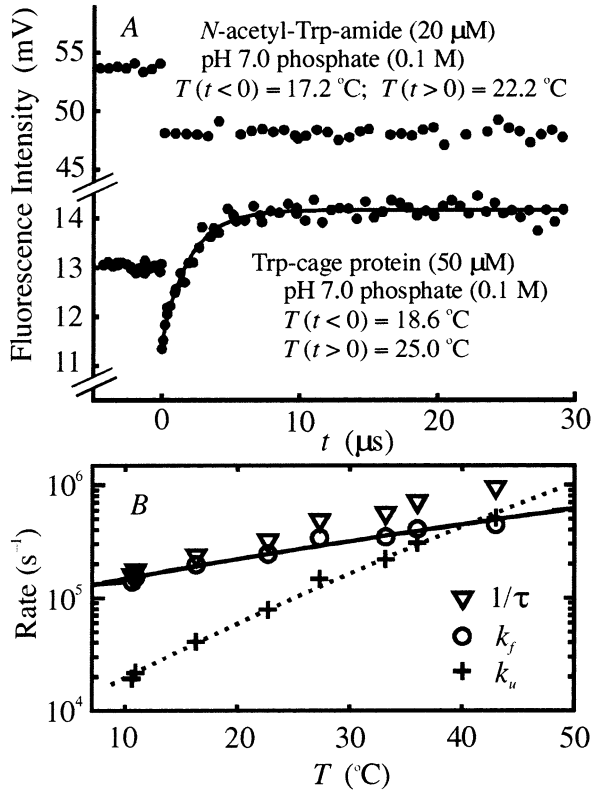

Figure 2. (*A*) Fluorescence response to 6.4 °C *T*-jump of Trp-cage (lower trace), and response of reference solution (free Trp) after *T*-jump at $t = 0$. (*B*) Relaxation rate $1/\tau$ from exponential fits to fluorescence relaxations in Trp-cage and rates of folding (k_f) and unfolding (k_u) calculated from τ and χ_{unf} (Figure 1). Curves indicate Arrhenius fits with E_a(folding) $= 27 \pm 1$ kJ/mol, E_a (unfolding) $= 76 \pm 5$ kJ/mol.

of any intermediate), the observed equilibrium constant, and reasonable values for the fluorescence of each state.

Figure 2B shows the rates of folding (k_f) and unfolding (k_u) derived from a two-state model for the measured relaxation rate $\tau^{-1} = k_f + k_u$ and the unfolded fraction $\chi_{unf} = k_u/(k_f + k_u)$ obtained from CD. In pH 7 phosphate at $T = 22.7$ °C, the observed relaxation time $\tau \approx 3.1$ μs gives a folding rate $k_f \approx (4.1\ \mu s)^{-1} = 240\,000\ s^{-1}$. Although slower than the rate of formation of simple isolated helices, this rate exceeds the fastest folding rates previously observed[1] for complete proteins, including the engrailed homeodomain (61 aa, $k_f \approx (27\ \mu s)^{-1}$ at 25 °C in water), the WW domains (38−44 aa, $k_f \leq (24\ \mu s)^{-1}$, and even the 16-residue β-hairpin fragment (~ 6 μs). It also surpasses the fastest rates previously inferred by extrapolation, such as Fe(II) cytochrome b_{562} (106 aa), whose folding rate in denaturant extrapolates to $k_f \approx (5\ \mu s)^{-1}$ in water, and the B-domain of protein A (58 aa), which extrapolates to $k_f \approx (8\ \mu s)^{-1}$.

Why does the Trp-cage fold so quickly? Previous studies have not found shorter chains to fold more rapidly.[5] Folding rates for two-state proteins correlate far better with relative contact order, a measure of intrachain connectivity in the tertiary structure.[5] Yet contact order does not explain the rate either, since the Trp-cage has a large relative contact order (0.19) more typical of far slower folding proteins ($k_f \approx 10\ s^{-1}$). Ignoring the many indole contacts in the hydrophobic core of the Trp-cage only reduces the contact order to 0.17, which predicts a still sluggish $k_f \approx 100\ s^{-1}$. A few other complete proteins also exceed the predictions of the contact order correlation, although by less than 50-fold.[6] We expect that the contact order correlation severely underestimates the folding speed for the Trp-cage because it requires the statistical independence of most long-range tertiary contacts in the folding transition state, an improbable scenario for a 20-residue chain with a persistence length of 4−5 residues. Contact order models must ultimately take account of chain-length and flexibility.[7]

The folding speed of the Trp-cage in fact approaches the rate of diffusional loop formation within the unfolded chain, which likely poses a physical limit to folding speed. Laser-spectroscopic studies[3] of polypeptides of different lengths show that the two endpoints of a 20-residue chain should diffuse into contact on a time scale ~ 0.2 μs, or $\sim 20\times$ faster than Trp-cage folding. This implies a nearly perfectly optimized free energy landscape: the Arrhenius fit to k_f (Figure 2B) confirms that the molecule encounters only a weak barrier in folding: the activation enthalpy of folding $E_a = 27 \pm 1$ kJ/mol only slightly exceeds that associated with the viscosity of the solvent (~ 17 kJ/mol), for a net barrier of ~ 10 kJ/mol $\approx 4\ k_B T$. Furthermore, residual interactions in the unfolded state may reduce the entropic cost of folding; the unfolded state does not exhibit the chemical shifts of a random coil, perhaps owing to the formation of hydrophobic clusters.[4]

The small size and exceedingly fast folding of the Trp-cage will certainly make it an important benchmark for MD simulation: we note that Simmerling et al. have already reported an all-atom structure prediction and folding simulation of this molecule.[8] This scrupulously optimized protein may also provide new insight into the relationship of tertiary structure to folding dynamics.

Acknowledgment. We thank Alfred Chung of the UF ICBR for synthesizing the Trp-cage. The National Science Foundation (MCB No. 0077907) and the National Institutes of Health (No. 5R01 AI049063-02) provided funding support.

References

(1) (a) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522. (b) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98*, 13002−13007. (c) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305−310. (d) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6587−6590. (e) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552−558. (f) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199; Jager, M.; Nguyen, H.; Crane J. C.; Kelly J. W.; Gruebele M. *J. Mol. Biol.* **2001** *311* 373−393. (g) Callender R. H.; Dyer R. B.; Gilmanshin R.; Woodruff W. H. *Annu. Rev. Phys. Chem.* **1998**, *49*, 173−202.

(2) (a) Shea, J. E.; Brooks, C. L. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499−535. (b) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.

(3) (a) Hagen, S. J.; Hofrichter, J.; Szabo, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11615−11617. (b) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220−7225. (c) Hagen, S. J.; Carswell, C. W.; Sjolander, E. M. *J. Mol. Biol.* **2001**, *305*, 1161−1171. (d) Bieri O.; Wirz J.; Hellrung B.; Schutkowski M.; Drewello M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597−9601.

(4) (a) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430. (b) Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Biol.* **2002**, *9*, 408−410.

(5) (a) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985−994. (b) Plaxco, K. W.; Simons, K. T.; Ruczinski, I.; Baker, D. *Biochemistry* **2000**, *39*, 11177−11183.

(6) (a) Islam, S. A.; Karplus, M.; Weaver, D. L. *J. Mol. Biol.* **2002**, *318*, 199−215. (b) Makarov, D. E.; Plaxco, K. W. *Protein Sci.*, in press.

(7) Koga, N.; Takada S. *J. Mol. Biol.* **2001**, *313*, 171−180.

(8) Simmerling, C. L.; Strockbine, B.; Roitberg A. E.; *J. Am. Chem. Soc.* **2002**, *124*, 11258−11259.

JA0279141