doi:10.1006/jmbi.2000.3969 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **301**, 1019–1027

JMB

Two-state Expansion and Collapse of a Polypeptide

Stephen J. Hagen and William A. Eaton*

Laboratory of Chemical Physics Building 5, NIDDK, National Institutes of Health, Bethesda MD 20892-0520, USA

**Corresponding author*

The initial phase of folding for many proteins is presumed to be the collapse of the polypeptide chain from expanded to compact, but still denatured, conformations. Theory and simulations suggest that this collapse may be a two-state transition, characterized by barrier-crossing kinetics, while the collapse of homopolymers is continuous and multiphasic. We have used a laser temperature-jump with fluorescence spectroscopy to measure the complete time-course of the collapse of denatured cytochrome *c* with nanosecond time resolution. We find the process to be exponential in time and thermally activated, with an apparent activation energy ~9 k_BT (after correction for solvent viscosity). These results indicate that polypeptide collapse is kinetically a two-state transition. Because of the observed free energy barrier, the time scale of polypeptide collapse is dramatically slower than is predicted by Langevin models for homopolymer collapse.

Keywords: cytochrome *c*; molten globule; collapse; temperature-jump; denatured states

Introduction

When an unfolded polymer chain is transferred into a poor solvent it collapses to form a compact globule. Similarly, an unfolded polypeptide that is transferred from a denaturing solvent (e.g. guanidine hydrochloride solution) into water also collapses to a compact configuration. Collapse of the polypeptide may either coincide with, or precede, the formation of the native (folded) structure. If collapse and folding are not simultaneous, the polypeptide may possess both expanded and compact denatured states. The compact denatured configuration has been variously called the ''molten globule'', the ''compact globule'', or the ''folding intermediate''. Both the collapse transition and the nature of these compact denatured states are of great importance to the understanding of protein folding. Consequently, the dynamics of polymer collapse has been the subject of numerous theoretical investigations, starting with the work by de Gennes (1985).

Recently, novel experimental techniques have been developed for the study of polypeptide collapse. As a result it is now becoming possible to compare theoretical models for collapse of idealized chains with experimental observations on polypeptides. Nanosecond and microsecond-resolved triggering and observation of protein folding reactions (Eaton *et al*., 1997; Winkler & Gray, 1998) make it possible to ask: how does a fully unfolded polypeptide chain collapse to a compact state? Do non-specific, hydrophobic forces drive the polypeptide to a near-random ensemble of compact configurations, as in the collapse of a homopolymer, or do specific inter-residue interactions generate a two-state, or cooperative, transition to compact, native-like structures (Brooks, 1998; Chan & Dill, 1991; Gutin *et al*., 1995; Shastry & Roder, 1998; Sosnick *et al*., 1996)?

Chan *et al*. (1997) used a turbulent mixing device (dead time ~70 μs) to search for the collapse of cytochrome *c*. They monitored collapse through the fluorescence of the lone tryptophan residue (Trp59), which is quenched by intramolecular resonant energy transfer (Förster transfer) to the heme, which is covalently attached to the (104 amino acid residue) polypeptide at positions 14, 17 and 18. These experiments improved the time resolution of mixing experiments by nearly two orders of magnitude and completely resolved the previously unobserved, sub-millisecond folding of cytochrome *c*. However, substantial quenching of

Present address: S. J. Hagen, Department of Physics, University of Florida, Gainesville FL 32611, USA.

E-mail address of the corresponding author: eaton@helix.nih.gov

This paper was incorrectly published in J. Mol. Biol. (2000) 297, 781-789 without the authors' corrections. The publisher regrets any inconvenience caused to the authors.

0022-2836/00/041019–9 $35.00/0

the Trp59 fluorescence within the instrument dead time indicated the presence of an unresolved collapse process. Shastry & Roder (1998) repeated these mixing experiments with an improved signal:noise ratio and a dead time of ~45 μs. They observed the initial fluorescence quenching to be nearly exponential in time, with a decay time ~25-65 μs. They attributed this quenching to collapse of the unfolded chains to a compact ensemble, and then analyzed the rates and signal amplitudes in terms of a sequential three-state model. Here we have labeled these states as shown in Figure 1:

$$D_E \rightleftharpoons D_C \rightleftharpoons N$$

D_E and D_C respectively are the expanded and compact denatured ensembles, and N is the native (folded) state. Shastry & Roder's (1998) analysis indicated that the Trp59 fluorescence in state D_C is >95 % quenched. Pollack *et al*. (1999) used submillisecond-resolved small-angle X-ray scattering and the kinetic results of Shastry & Roder (1998) and Chan *et al*. (1997) to demonstrate that the radius of gyration of the compact denatured molecule is no more than 30 % greater than that of the native state, consistent with nearly complete quenching of Trp59 in state D_C.

Shastry & Roder (1998) could not demonstrate unambiguously that the observed collapse from D_E to D_C is exponential in time, since their instrument dead time was comparable to the observed collapse time. We have resolved the entire time course of this rapid process by triggering the re-equilibration between more and less expanded configurations of the cytochrome *c* chain with a nanosecond laser temperature-jump. We monitored the process through the fluorescence energy transfer between Trp59 and the heme. We find the time-course of the fluorescence change to be exponential, with an Arrhenius-like temperature dependence that exceeds the temperature dependence of the solvent viscosity. These results indicate that expansion and collapse of the denatured protein is a two-state process, with an energy barrier separating expanded and compact ensembles.

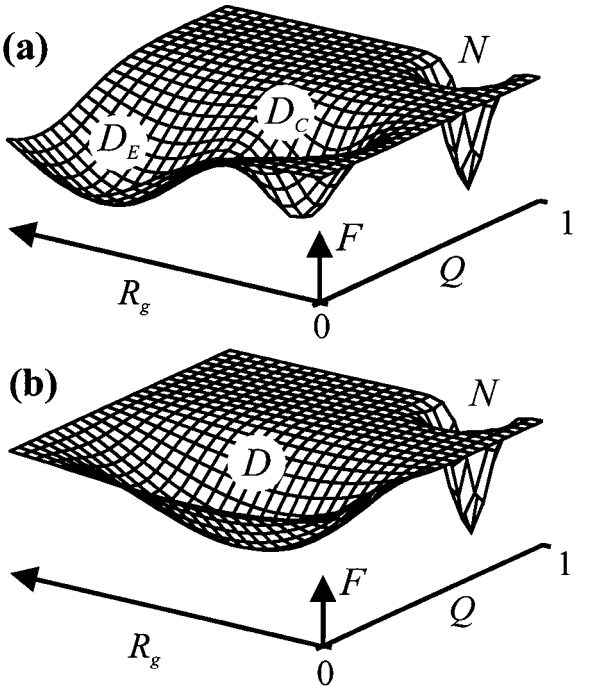

Figure 1. Two possible free energy surfaces for a denatured polypeptide. The free energy F is plotted as a function of the chain's radius of gyration, R_g, and a variable, Q, that represents the degree of native-like structure. For example, Q could be the fraction of native intra-chain contacts that are achieved in a particular configuration of the molecule. (a) System with two denatured states, D_E and D_C, and a fully native state (N). D_E is fully expanded and has very few native contacts, while D_C is compact and more native-like; (b) System with one denatured state (D) and a native state (N). In this case D contains both expanded and compact configurations of the polypeptide, but no barrier separates these ensembles. An interesting outstanding issue is whether the kinetics of protein collapse and folding can be predicted from models of diffusive motion on such surfaces (Eaton, 1999; Muñoz & Eaton, 1999; Onuchic *et al*., 1997).

Results

The nanosecond T-jump at $t = 0$ causes an abrupt drop in tryptophan fluorescence emission, due to the intrinsic temperature dependence of tryptophan's fluorescence yield (Thompson *et al*., 2000). Under weakly unfolding conditions, within the lower tail of the thermal unfolding curve (e.g. T ~20 °C in Figure 2), the temperature change also triggers detectable microsecond and millisecond relaxations in the Trp59 fluorescence (Figure 3). For $t \geqslant 10$ ns, this biphasic relaxation of the fluorescence signal $f(t,T)$ is well-described by:

$$f(t, T) = a_1 \exp(-t/\tau_1(T)) + a_2 \exp(-t/\tau_2(T)) + a_0(T) \quad (1)$$

Here τ_1 is ~100 μs, τ_2 is ~1-10 ms, T is the sample temperature immediately after the T-jump laser pulse (i.e. $t > 10$ ns), and $a_0(T)$ is the equilibrium fluorescence at T. (At longer times, $t \sim 100$ ms, thermal diffusion cools the probed volume and generates a third relaxation, not shown in Figure 3.) The faster process (τ_1) is the re-equilibration of the denatured molecules between different denatured states (Figure 1). Specifically, it reflects a net expansion of the denatured chains. In a two-state model for the denatured states (Figure 1(a)), the relaxation rate $1/\tau_1$ equals the sum of the rates for expansion ($D_C \Rightarrow D_E$) and collapse ($D_E \Rightarrow D_C$) of a denatured molecule. The slower ($\tau_2 > 1$ ms) observed phase represents re-equilibration between native and denatured states. The kinetics of the slower folding/unfolding process characterized by a_2 and τ_2

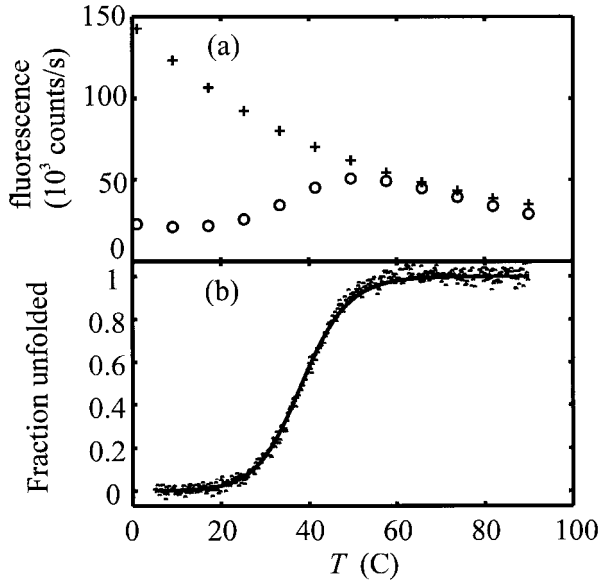

Figure 2. (a) 350 nm fluorescent emission of 1 μM horse ferric cytochrome *c* (o) or peptide ITWKE (+) with excitation at 264 nm. Solvent is 2.19 M GdnHCl, 0.5 M imidazole, 0.1 M phosphate (pH 7.0); (b) Fraction unfolded, as determined by circular dichroism at 230 nm.

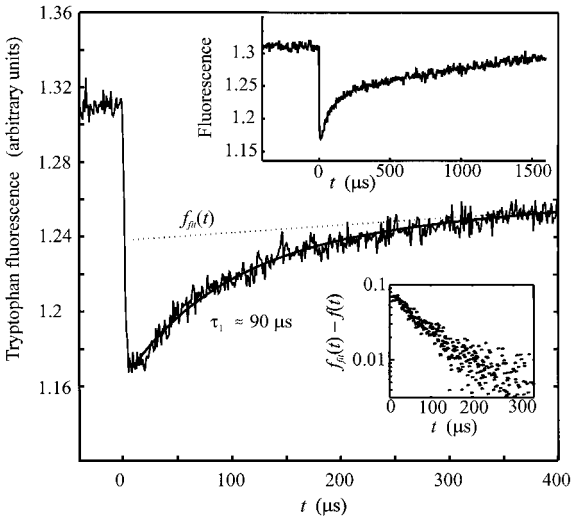

Figure 3. Trp-59 fluorescence $f(t,T)$ of cytochrome *c* detected after a nanosecond temperature jump at time $t = 0$ ($T(t > 0) = 30.3\,^{\circ}\mathrm{C}$, $\Delta T = 10\,^{\circ}\mathrm{C}$). The data for $0 < t < 1.5$ ms are fit to equation (1) (solid curve), although the main panel shows only the data for $t \leqslant 400$ μs. The upper inset shows the rest of the data. The dotted curve is $f_{\mathrm{fit}}(t) = a_2 \exp(-t/\tau_2) + a_0$, or the extrapolation to short times of the long-time ($t \gg \tau_1$) behavior of the fluorescence decay. The lower inset shows the difference $f_{\mathrm{fit}}(t) - f(t)$ for $t \leqslant 400$ μs, showing that the microsecond fluorescence relaxation closely approximates an exponential decay. Solvent is 1.53 M Gdn HCl, 0.5 M imidazole, 0.1 M phosphate (pH 7.0).

have been described in detail elsewhere (Chan *et al*., 1997; Shastry & Roder, 1998).

The microsecond τ_1 process is only observed at low GdnHCl concentrations (<~2 M), within a narrow range of temperatures below the midpoint of the thermal unfolding transition. Outside this range of conditions, a ten-degree temperature jump does not induce a detectable re-equilibration between expanded and compact configurations. This is a consequence of the fact that very few molecules occupy compact denatured states, even under experimental conditions that are optimized for these *T*-jump studies. As a result, the microsecond relaxation generates a small signal that is difficult to study in detail: in Figure 3 the relaxation amplitude is only ~5% of the baseline fluorescence, which is itself quite small below the thermal unfolding transition.

Because the observed relaxation amplitudes are small, it is important to establish that the microsecond relaxation in Figure 3 is the same expansion/collapse process observed when fully unfolded cytochrome *c* was diluted from high to low denaturant concentration in rapid mixing experiments (Shastry & Roder, 1998). In both experiments a sub-millisecond relaxation is observed in the fluorescence of Trp59 as the cytochrome *c* re-equilibrates between denatured ensembles. However, the relaxation in the rapid mixing experiment appears significantly faster: for example, we find a ~90 μs relaxation in 1.53 M Gdn HCl at $T = 30\,^{\circ}\mathrm{C}$ (Figure 3), while Shastry & Roder (1998) report a ~33 μs relaxation upon dilution of fully unfolded cytochrome *c* into the same solvent at $T = 22\,^{\circ}\mathrm{C}$. However, since the capillary mixing device used in the dilution studies had an intrinsic dead time of ~45 μs, most of the polypeptide collapse was not resolved experimentally. Because of this dead time, mixing experiments can determine rapid relaxation rates precisely only if the $t = 0$ signal amplitude is also known. Shastry & Roder (1998) assumed the $t = 0$ (fully unfolded) fluorescence of cytochrome *c* to be independent of GdnHCl concentration. However, Chan *et al*. (1997) observed that the fluorescent quantum yield of the nine-residue tryptophan-containing peptide decreased with decreasing GdnHCl concentration. This finding would imply that Shastry & Roder's (1998) assumption systematically underestimates collapse times for cytochrome *c*. Differences in protein concentration may also contribute to the difference between our results and those of Shastry & Roder. In any case, the greater time resolution of the temperature jump technique reveals the full time course of the relaxation and allows unambiguous determination of these microsecond rates.

Figure 3 demonstrates that a single-exponential time course describes the microsecond relaxation observed in the tryptophan fluorescence. This

suggests (although it does not prove) that collapse is a two-state (cooperative) process, with a barrier separating expanded and compact denatured ensembles on the multi-dimensional free energy surface of the molecule (Figure 1(a)). The temperature dependence of τ_1 more explicitly demonstrates the existence of a barrier. An Arrhenius plot (Figure 4), shows that $\tau_1 \approx A_0 \exp(E_a/k_BT)$, indicating that an activation energy $E_a \approx 40$ kJ/mol ≈ 16 k_BT controls the rate of re-equilibration between compact and expanded states, with $A_0 \sim 10$ ps. Figure 5 shows that both the relaxation time τ_1 and the energy barrier E_a vary only weakly over the range of denaturant concentrations for which the relaxation is detectable through T-jump studies. In general our observed τ_1 is three to five times slower than reported by Shastry & Roder (1998). However, these authors report a comparable activation energy, $E_a \approx 12$ k_BT, for the relaxation rate (although at pH 4.5), providing further evidence that we are in fact resolving the same physical process.

Not all of E_a can be attributed to the polypeptide chain, as the chain's diffusion rate D presumably governs expansion and collapse. D should vary inversely with the viscosity η of the solvent, i.e. $\tau_1 \propto 1/D \propto \eta(T)$, while near room temperature the viscosity of water is nearly Arrhenius-like, $\eta(T) \propto \exp(E_{solvent}/k_BT)$, where $E_{solvent} \approx 7$ $k_B \times 300$ K (Lide, 1998). Solvent viscosity can therefore account for roughly half of the observed barrier E_a, while $E_{chain} = E_a - E_{solvent} \approx 9$ k_BT of the observed barrier actually arises from the dynamics of the polypeptide chain itself.

The positive slope in the Arrhenius plot for τ_1 raises the question of whether E_{chain} is the effective energy barrier for the expansion ($D_C \Rightarrow D_E$ in Figure 1(a), at rate k_e) or the collapse ($D_E \Rightarrow D_C$, at rate k_c) of denatured cytochrome c. While the total relaxation is thermally activated over the range

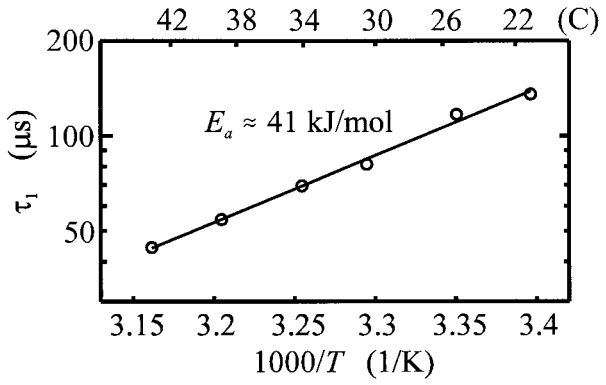

Figure 4. Arrhenius plot for the microsecond relaxation, representing re-equilibration of denatured cytochrome c between expanded and compact conformations. Abscissa indicates the sample temperature after the T-jump of $\Delta T \approx 10\,^\circ$C. Solvent conditions same as for Figure 3.

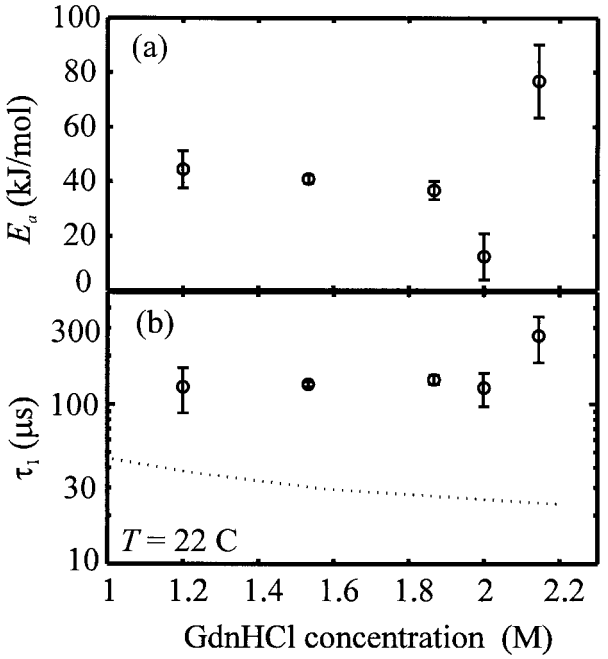

Figure 5. (a) Energy barrier $E_a = d\ln(\tau_1)/d(1/k_BT)$ *versus* denaturant concentration. (b) Relaxation time τ_1 at $T = 22\,^\circ$C. The broken line shows the fast relaxation times observed in the rapid-mixing experiment by Shastry & Roder. The sample is cytochrome c 150 μM in GdnHCl, 0.5 M imidazole, 0.1 M phosphate (pH 7.0).

$T = 21 - 43\,^\circ$C, the two individual rates presumably have different energetic barriers. Therefore it is of interest to determine whether one or both of these barriers are non-zero, i.e. whether there are energetic barriers to both expansion and collapse of the chain. For typical two-state systems one resolves this question by measuring the equilibrium constant, $K_{eq} = k_e/k_c$: once K_{eq} is known, the individual rates k_c and k_e are obtained from $k_c^{-1} = (1 + K_{eq})\tau_1$ and $k_e = K_{eq}$ k_c. Arrhenius plots for k_c and k_e then indicate whether both rates exhibit energy barriers. The present system is not so easily treated, however, since it has multiple states. N is heavily populated under the most favorable equilibrium conditions, while the states of interest (D_E and D_C in the three state model of Figure 1(a)) are present at levels no greater than a few per cent. Furthermore, two of the states (D_C and N) are strongly quenched (and therefore indistinguishable) in their fluorescence emission. Thus, the expansion-collapse transition is experimentally accessible only through kinetic studies.

We therefore used the Arrhenius data of Figure 4 to estimate the size of the energy barriers for k_c and k_e separately. We use the relation:

$$1/\tau_1 = k_e + k_c = A_e \exp(-(H^\ddagger - H_c)/k_BT) + A_c \exp(-(H^\ddagger - H_e)/k_BT)$$

where A_c and A_e are rate prefactors for collapse and expansion of the chain, and H_c, $H^\ddagger$ and H_e are the apparent enthalpies of the collapsed state, the transition state, and the expanded state, respectively. This is equivalent to:

$$\begin{aligned} 1/\tau_1 &= A_c \exp(-(H^\ddagger - H_e)/k_B T) \\ &\quad \times [1 + A_e/A_c \exp((H_c - H_e)/k_B T)] \\ &= A_c \exp(-\Delta/k_B T)[1 + K_{eq}] \end{aligned}$$

where $\Delta = H^\ddagger - H_e$. Since Shastry *et al*. (1998) have estimated $K_{eq} \approx 1$ for cytochrome *c* in 1.5 M GdnHCl at $T = T_0 \equiv 22\,^\circ$C, we can approximate K_{eq} at T near T_0 as:

$$K_{eq}(T) \approx \exp(H(T - T_0)/k_B T_0^2)$$

with $H = H_e - H_c$ and $k_B T_0 = 2.45$ kJ/mol K. Thus:

$$1/\tau_1 \approx A_c \exp(-\Delta/k_B T)[1 + \exp(H(T - T_0)/k_B T_0^2)]$$

We estimate A_c, Δ, and H by fitting the τ_1 (T) data to this expression: the linearity of the Arrhenius plot places a limit on H, with the assumption that these various parameters are themselves independent of T. The data for cytochrome *c* in 1.53 M GdnHCl (pH 7.0) are most consistent with $\Delta \geqslant 8\ k_B T$ and $H \leqslant 14\ k_B T$ (expansion with increasing T indicates $H > 0$). These produce a 50 % increase in the final sum-of-squares compared to the most probable values of $\Delta \approx 16\ k_B T$ and $H \approx 0$. Thus our data indicate a net positive energy barrier for both expansion and collapse of the chain, even after subtraction of the solvent viscosity's contribution ($\sim 7\ k_B T_0$) from these activation energies.

Discussion

At temperatures below the theta temperature, the net interaction between the residues of a polymer chain is attractive, driving the chain to collapse from a random coil to a compact globule (Williams *et al*., 1981). This non-specific collapse has long been of interest because of its potential relevance to early stages of protein folding. The equilibrium properties of homopolymer and heteropolymer collapse have been studied extensively. Homopolymers are generally expected to collapse in a continuous (i.e. second-order-like or barrierless) transition, although in the case of short, stiff chains, a two-state transition appears possible (Williams *et al*. 1981; Chan & Dill, 1991; Grosberg & Khokhlov, 1994). Heteropolymer collapse is naturally more complicated, owing to the possibility of specific interactions and the formation of structure (Chan & Dill, 1991; Garel *et al*., 1994). For heteropolypeptides that fold to proteins, Ptitsyn & Uversky (1994) have used a length-scaling argument as evidence that a compact denatured state with native-like overall topology, the so-called ''molten globule'', represents a distinct state separated from a more unfolded state by a free energy barrier; the collapse of a polypeptide to this molten globule is then thermodynamically first-order like. Much less is known about polymer collapse kinetics. These kinetics are difficult to study experimentally, and only limited laboratory data are available, either for proteins (Shastry & Roder, 1998; Ballew *et al*., 1996; Gilmanshin *et al*., 1998) or for other polymers (Chu *et al*., 1995; Zhu & Napper, 1997). Theoretical work has addressed collapse kinetics through the use of phenomenological models (de Gennes, 1985; Grosberg *et al*., 1988; Thirumalai, 1995; Buguin *et al*., 1996) and Langevin dynamics (Pitard, 1999; Pitard & Orland, 1999; Timoshenko *et al*., 1995), as well as Monte-Carlo and molecular dynamics simulations (Kavassalis & Sundararajan, 1993; Milchev & Binder, 1994; Ostrovsky & Bar-Yam, 1995; Kuznetsov *et al*., 1995, 1996). The collapse of polypeptides has been studied primarily through lattice models and other computer simulations (Camacho & Thirumalai, 1993; Dill & Chan, 1997; Dobson *et al*., 1998; Gutin *et al*., 1995; Klimov & Thirumalai, 1996; Onuchic *et al*., 1997; Shakhnovich, 1997). Such simulations have led to the suggestion (Klimov & Thirumalai, 1996) that the rate of protein folding varies inversely with the difference between the theta temperature and folding temperature of the molecule.

These theoretical studies indicate that homopolymer collapse is an extended process, occuring in several kinetic stages. In the early stage, clusters of residues form along the chain. These clusters then aggregate, increasing in size as they decrease in number. This ''coarsening'' stage may be followed by slower compaction, reptation, or other relaxations as the globule develops. The kinetics of these phases may exhibit power law, rather than exponential, time dependence. Multi-phasic, non-exponential collapse of homopolymers has been observed experimentally (Chu *et al*., 1995; Zhu & Napper, 1997).

One may imagine that the collapse of a small, single-domain protein ($N \leqslant 100$ residues), consisting of a relatively stiff chain, need not encompass all of the same phases characteristic of longer chains. The initial compaction of a single-domain protein may, for example, resemble only the formation of a single hydrophobic cluster on a long homopolymer. Therefore, we compare the observed collapse kinetics of cytochrome *c* with a recent prediction for the early time behavior of a collapsing homopolymer (Pitard, 1999; Pitard & Orland, 1998). Using a Langevin-equation approach, these authors predict that a chain initially at equilibrium at its theta temperature, then quenched to lower T, collapses according to:

$$\langle R_g^2 \rangle \approx (1 - (t/\tau_c)^{3/4}) \langle R_g^2(0) \rangle$$

for short times t. Here R_g is the radius of gyration and the characteristic collapse time is $\tau_c \propto a_0^2 (N a_0^3/v)^{4/3}/D_o$, where a_0 is the length of the monomer unit, D_o is the monomer diffusion constant, and v is the excluded volume per monomer. The same τ_c

also characterizes chain expansion. Using the values of the constants and equation 15 given by Pitard (1999), the prediction for the 104-residue cytochrome *c* in water is $\tau_c \approx 70$ ns, three orders of magnitude faster than the observed collapse. With the inclusion of hydrodynamic interactions, Pitard (1999) obtained:

$$\langle R_g^2 \rangle \approx (1 - t/\tau_{c,h})\langle R_g^2(0)\rangle$$

at short times, which is presumably equivalent to an exponential collapse $\langle R_g^2 \rangle \approx \langle R_g^2(0)\rangle \exp(-t/\tau_{c,h})$. Here the time scale is $\tau_{c,h} \propto \eta\, a_0^6 N/(k_B T v)$. For cytochrome *c*, $\tau_{c,h} \approx 0.6$ μs (equation 26 of pitard (1999)), which is still much shorter than our experimental time. Although the inclusion of hydrodynamic interactions slows the collapse time, one generally expects hydrodynamic interactions to accelerate collapse, since the fluid medium couples the diffusive motion of non-adjacent monomers. In general, theoretical studies of collapse have not successfully predicted the collapse time scales observed in experiments (Chu *et al.*, 1995; Zhu *et al.*, 1997). A more significant difference with our result is that the Langevin analysis does not predict thermal activation in the collapse times τ_c and $\tau_{c,h}$, except through the viscosity η. The "excess" activation energy E_{chain} discussed above is potentially a unique feature of heteropolypeptide collapse.

The exponential character of the microsecond fluorescence decay shown in Figure 3 should also have a significant bearing on theoretical models for collapse. It strongly suggests that the polypeptide collapses in a two-state process, unlike the continuous collapse (cluster formation) envisioned for homopolymers. Even if $\langle R_g^2 \rangle$ for the cytochrome *c* ensemble decreases exponentially with time, as in some homopolymer collapse theories, the decay of the fluorescence can still exhibit either exponential or non-exponential kinetics. This is because the efficiency of Förster energy transfer from the Trp59 to the heme varies as $R_0^6/(R_0^6 + r^6)$, where *r* is the Trp59-to-heme distance and $R_0 \approx 34$ Å. Figure 6 shows two hypothetical scenarios for the collapse of an ensemble of chain molecules (with $N = 41$ segments of length 3.8 Å and Flory ratio ~8, like the polypeptide connecting Trp59 with the heme at His18). In both cases the initial distribution of end-to-end distances is Gaussian, $P(r) \propto r^2 \exp(-3\, r^2/2\langle r^2(t=0)\rangle)$. In both cases also the mean-square end-to-end distance $\langle r^2 \rangle$ decays exponentially with time:

$$\langle r^2(t)\rangle = \langle r^2(\infty)\rangle(1 - e^{-t}) + \langle r^2(0)\rangle e^{-t}$$

Figure 6(a) presents a continuous collapse, in which the ensemble maintains a Gaussian distribution during collapse:

$$P(r, t) \propto r^2 \exp(-3r^2/2\langle r^2(t)\rangle)$$

Here $\langle r^2(t)\rangle$ decays exponentially from ≈69 Å to ≈10 Å. By contrast, Figure 6(b) models a two-state collapse:

$$P(r, t) = P_{expanded}(r)e^{-t} + P_{compact}(r)(1 - e^{-t})$$

Here $P_{expanded}(r)$ and $P_{compact}(r)$ are two different Gaussian distributions. The insets to Figure 6(a) and (b) show very different predictions for the fluorescence yield $f(t)$ in the two cases. If the ends of these chains are labeled with a fluorescence donor and acceptor pair, two-state collapse does generate an exponential fluorescence decay, while fluorescence decay for continuous collapse is clearly non-exponential in time. Our experimentally observed microsecond decay (Figure 3) is clearly more similar to the two-state case (Figure 6(b)).

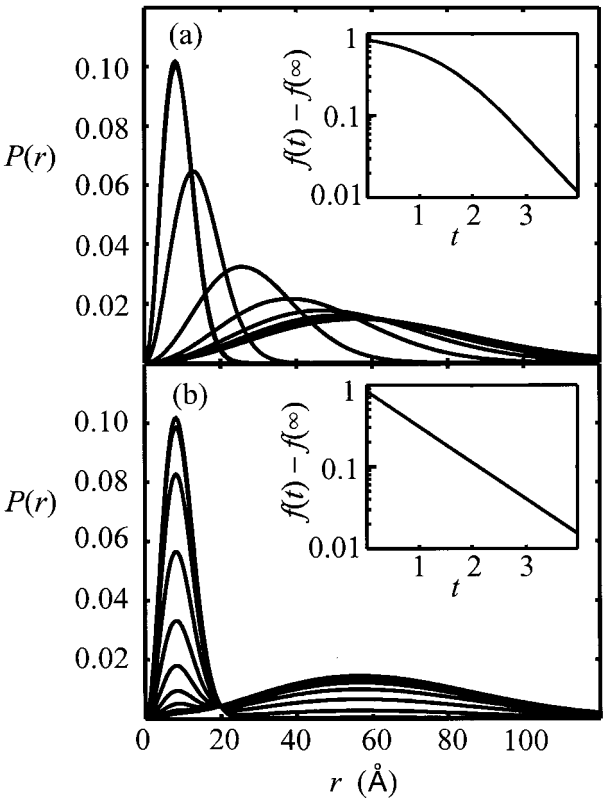

Figure 6. Simulated distribution $P(r)$ of end-to-end distances *r* for a Gaussian chain of 41 amino acid residues (e.g. the polypeptide linking His18 to Trp59 of cytochrome *c*) undergoing collapse to a compact state $\langle r^2 \rangle \approx (10\ \text{Å})^2$ from an initial distribution with $\langle r^2 \rangle \approx (69\ \text{Å})^2$. The collapse kinetics obey $\langle r^2(t)\rangle - \langle r^2(\infty)\rangle \propto e^{-t}$. Curves represent a succession of 12 different times logarithmically spaced from $t = 0.01$ to $t = 31$. (a) Continuous collapse, in which the chain maintains a Gaussian distribution $P(r)$ during collapse; (b) Two-state collapse, in which the chain interconverts between two separate Gaussian distributions. Insets (a) and (b) The predicted decay of the ensemble averaged fluorescence yield, $f(t)$, for tryptophan residues at one end of the chains quenched by Förster exchange with heme groups at the other end.

Our data show that the initial collapse of a polypeptide from an unfolded to a compact configuration can be a two-state process, requiring passage over a free energy barrier. This is unlike the more continuous, extended process envisioned for the collapse of long homopolymers. What is the origin of this free energy barrier to collapse? If the polypeptide's energy and entropy both decrease linearly as the chain collapses, no free energy barrier is encountered. However, from simulations and general theoretical considerations Pande & Rokhsar (1998) have argued that a free energy barrier does divide the unfolded and compact-denatured ensembles of ''designed'' (i.e. having a native state) heteropolymers. This barrier arises from the presence of both combinatorial entropy and loop entropy in partially compact molecules; the sum of these entropies exhibits a minimum as the chain collapses, giving rise to a net barrier in the free energy. This analysis does not consider possible energetic contributions to the free energy barrier, although we find in cytochrome *c* a significant activation energy for both expansion and collapse of the chain. This energy barrier could potentially arise from the net positive charge on the molecule. The cytochrome *c* molecule carries a net charge of roughly +10 at neutral pH. Electrostatic repulsion could impede collapse until the molecule becomes sufficiently compact to benefit from short range hydrophobic interactions, giving rise to an energetic barrier, as discussed by Stigter *et al*. (1991). This simple model suggests that the thermodynamics and kinetics of polypeptide collapse will depend on the net charge, salt concentration (electrostatic screening), etc. of the system.

Conclusions

Although studies of polymer collapse have long been at least partially motivated by the interest in protein folding as a special case, the comparison between homopolymer collapse as it is understood theoretically, and protein folding as it can be observed experimentally, is only now being explored. Our results confirm, with dramatically improved time resolution, that the unfolded cytochrome *c* chain collapses to a compact state in a two-state transition. The exponential time course of the relaxation, together with the Arrhenius-law observed for its rate, strongly argue that the expanded and compact states represent two distinct local minima in the free energy surface of the denatured polypeptide. This is significantly unlike the picture that has developed from theoretical studies of homopolymer collapse, although it may be compatible with the results of recent dynamics simulations of proteins and designed heteropolymers.

Finally, it is interesting that the collapse time for cytochrome *c* approaches the time scale for end-to-end diffusion of the chain. The two endpoint residues of a randomly coiled, freely diffusing polypeptide can be expected to make physical contact on time scales of order $N^{3/2}\ \tau_0$, where $\tau_0 \sim 0.1$ μs and N is the number of amino acid residues (Hagen *et al.* 1996, 1997). For cytochrome *c* ($N = 104$), this predicts end-to-end diffusion should occur in ~100 μs. Thus, although the presence of attractive intrachain interactions could be expected to accelerate chain collapse relative to simple end-to-end diffusion, the presence of a free energy barrier compensates for these interactions and slows the collapse process to longer time scales. The slow time scale for collapse of cytochrome *c*, and the existence of a free energy barrier, imply that specific intrachain interactions drive a cooperative transition to (and maintain the stability of) the collapsed states of cytochrome *c*. Further studies of this transition will provide the only rigorous test of current theoretical models for collapse and folding.

Materials and Methods

The 150 μM horse ferric cytochrome *c* (Sigma Chemical Co.) was prepared in 0.5 M imidazole (zone refined to reduce background fluorescence) and ~1-2 M guanidine hydrochloride (GdnHCl, from ICN Biomedicals Inc.), in 0.1 M potassium phosphate (pH 7.0). Figure 2 shows the temperature-dependence of the 350 nm fluorescence emission of cytochrome *c*, with 264 nm excitation of Trp59. Figure 2 also shows the fluorescence of the tryptophan-containing pentapeptide acetyl-ITWKE-amide (California Peptide Research, Inc.). The pentapeptide mimics Trp59 in its local peptide environment on the cytochrome *c* chain, but without the heme group. The fluorescence of the pentapeptide falls monotonically with T, as does that of the unfolded protein (at $T > \sim 50\,^{\circ}\mathrm{C}$). However, in the native protein the Trp59 fluorescence emission is quenched by resonant energy transfer with the heme at His18. Thus, the fluorescence of the protein increases with T as the protein thermally unfolds ($T_{\mathrm{midpoint}} \approx 39\,^{\circ}\mathrm{C}$). Figure 2 also shows the unfolding transition of cytochrome *c* as observed in far-UV circular dichroism.

The temperature-jump instrument has been described (Thompson *et al.,* 1997). A Nd:YAG laser (Continuum, Inc.) generates a ~10 ns pulse at 1064 nm that is Raman shifted by a CH_4-He mixture to 1.54 μm, which lies within a vibrational absorption (OH stretch overtone) of water. The ~20 mJ pulse is then split into two counterpropagating pulses that are focused to a ~1 mm diameter spot on the sample, which is contained in a quartz cuvette (0.5 mm path length). The infrared pulse generates a rapid rise of ~8-10 deg. C in the temperature of the solvent in the focal region within a time ~10 ns; the injected heat dissipates on a time scale ~0.1 seconds.

The probe source is a frequency-doubled Ar^+ CW laser (Coherent, Inc.) at 264 nm (4 mW) focused to a ~40 μm spot on the sample. This excites the fluorescence of Trp59 at the center of the heated region of the sample. A wide-aperture lens gathers the Trp59 emission at 90 degrees from the excitation beam and focuses it through a bandpass filter (350 nm) onto a photomultiplier tube. We typically signal-averaged the data from ~3500 Nd:YAG laser shots at 1.67 Hz to generate a single fluorescence relaxation curve. We determined the magnitude of the laser temperature jump by measuring its effect on the fluorescence of a solution of *N*-acetyl-tryptophana-

mide (NATA, from Sigma Chemical Co.) in 1.7 M guanidine hydrochloride and 0.5 M imidazole (pH 7). The fluorescence of this solution decreases by approximately 1.7 % with each deg. C of temperature rise.

Acknowledgments

We thank James Hofrichter and Peggy Thompson for helpful discussions.

References

Ballew, R. M., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA,* **93**, 5759-5764.

Brooks, C. L. (1998). Simulations of protein folding and unfolding. *Curr. Opin. Struc. Biol.* **8**, 222-226.

Buguin, A., Brochard-Wyart, F. & de Gennes, P. G. (1996). Collapse of a flexible coil in a poor solvent. *C.R. Acad. Sci. Paris,* **322**, 741-746.

Camacho, C. J. & Thirumalai, D. (1993). Kinetics and thermodynamics of folding in model proteins. *Proc. Natl Acad. Sci. USA,* **90**, 6369-6372.

Chan, C.-K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofrichter, J. (1997). Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl Acad. Sci. USA,* **94**, 1779-1784.

Chan, H. S. & Dill, K. A. (1991). Polymer principles in protein structure and stability. *Ann. Rev. Biophys. Biophys. Chem.* **20**, 447-490.

Chu, B., Ying, Q. & Grosberg, Y. A. (1995). Two-stage kinetics of single-chain collapse. Polystyrene in cyclohexane. *Macromolecules,* **28**, 180-189.

de Gennes, P. G. (1985). Kinetics of collapse for a flexible coil. *J. Phys. Letters,* **46**, L-639-L-642.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10-19.

Dobson, C. M., Sali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Edit.* **17**, 868-893.

Eaton, W. A. (1999). Searching for ''downhill scenarios'' in protein folding. *Proc. Natl Acad. Sci. USA,* **96**, 5897-5899.

Eaton, W. A., Muñoz, V., Thompson, P. A., Chan, C. K. & Hofrichter, J. (1997). Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* **7**, 10-14.

Garel, T., Leibler, L. & Orland, H. (1994). Random hydrophilic-hydrophobic copolymers. *J. Phys. II France,* **4**, 2139-2148.

Gilmanshin, R., Callender, R. H. & Dyer, R. B. (1998). The core of apomyoglobin E-form folds at the diffusion limit. *Nature Struct. Biol.* **5**, 363-365.

Grosberg, A. Y. & Khokhlov, A. R. (1994). *Statistical Physics of Macromolecules*, AIP Press, New York.

Grosberg, A. Y., Nechaev, S. K. & Shakhnovich, E. I. (1988). The role of topological contstraints in the kinetics of collapse of macromolecules. *J. Phys.* **49**, 2095-2100.

Gutin, A. M., Abkevich, V. I. & Shakhnovich, E. I. (1995). Is burst hydrophobic collapse necessary for protein folding? *Biochemistry,* **34**, 3066-3076.

Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. (1996). Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA,* **93**, 11615-11617.

Hagen, S. J., Hofrichter, J. & Eaton, W. A. (1997). Rate of intrachain diffusion of unfolded cytochrome *c*. *J. Phys. Chem. series B,* **101**, 2352-2365.

Kavassalis, T. A. & Sundararajan, P. R. (1993). A molecular dynamics study of polyethylene crystallization. *Macromolecules,* **26**, 4144-4150.

Klimov, D. K. & Thirumalai, D. (1996). Criterion that determines the foldability of proteins. *Phys. Rev. Letters,* **76**, 4070-4073.

Kuznetsov, Y. A., Timoshenko, E. G. & Dawson, K. A. (1995). Kinetics at the collapse transition of homopolymers and random copolymers. *J. Chem. Phys.* **103**, 4807-4818.

Kuznetsov, Y. A., Timoshenko, E. G. & Dawson, K. A. (1996). Kinetic laws at the collapse transition of a homopolymer. *J. Chem. Phys.* **104**, 3338-3347.

Lide, D. R. (Ed.) (1998). *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, Florida.

Milchev, A. & Binder, K. (1994). Anomalous diffusion and relaxation of collapsed polymer chains. *Europhys. Letters,* **26**, 671-676.

Muñoz, V. & Eaton, W. A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl Acad. Sci. USA,* **96**, 11311-11316.

Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.* **48**, 545-600.

Ostrovsky, B. & Bar-Yam, Y. (1995). Motion of polymer ends in homopolymer and heteropolymer collapse. *Biophys. J.* **68**, 1694-1698.

Pande, V. S. J. & Rokhsar, D. S. (1998). Is the molten globule a third phase of proteins?. *Proc. Natl Acad. Sci. USA,* **95**, 1490-1494.

Pitard, E. (1999). Influence of hydrodynamics on the dynamics of a homopolymer. *Eur. Phys. J. ser. B,* **7**, 665-673.

Pitard, E. & Orland, H. (1998). Dynamics of the swelling or collapse of a homopolymer. *Europhys. Letters,* **41**, 467-472.

Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999). Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA,* **96**, 10115-10117.

Ptitsyn, O. B. & Uversky, V. N. (1994). The molten globule is a third thermodynamical state of protein molecules. *FEBS Letters,* **341**, 15-18.

Shakhnovich, E. I. (1997). Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struct. Biol.* **7**, 29-40.

Shastry, M. C. R. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.* **5**, 385-392.

Shastry, M. C. R., Sauder, J. M. & Roder, H. (1998). Kinetic and structural analysis of submillisecond folding events in cytochrome *c*. *Acc. Chem. Res.* **31**, 717-725.

Sosnick, T. R., Mayne, L. & Englander, S. W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome *c* folding. *Proteins: Struct. Funct. Genet.* **24**, 413-426.

Stigter, D., Alonso, D. O. V. & Dill, K. A. (1991). Protein stability: electrostatics and compact denatured states. *Proc. Natl Acad. Sci. USA,* **88**, 4176-4180.

Thirumalai, D. (1995). From minimal models to real proteins: time scales for protein folding kinetics. *J. Phys.* **5**, 1457-1467.

Thompson, P. A., Eaton, W. A. & Hofrichter, J. (1997). Laser temperature-jump study of the helix-coil kinetics of an alanine peptide interpreted with a ''kinetic zipper'' model. *Biochemistry,* **36**, 9200-9210.

Thompson, P. A., Muñoz, V., Jas, G. S., Henry, E. R., Eaton, W. A. & Hofrichter, J. (2000). The helix-coil kinetics of a heteropeptide. *J. Phys. Chem.* **104**, 378-389.

Timoshenko, E. G., Kuznetsov, Y. A. & Dawson, K. A. (1995). Kinetics at the collapse transition: Gaussian self-consistent approach. *J. Chem. Phys.* **102**, 1816-1823.

Williams, C., Brochard, F. & Frisch, H. L. (1981). Polymer Collapse. *Ann. Rev. Phys. Chem.* **32**, 433-451.

Winkler, J. R. & Gray, H. B. (1998). Editors of *Protein Folding*. *Acc. Chem. Res.* **31**.

Zhu, P. W. & Napper, D. H. (1997). The longer time collapse kinetics of interfacial poly(*N*-isopropylacrylamide) in water. *J. Chem. Phys.* **106**, 6492-6498.

Edited by A. R. Fersht

(Received 31 August 1999; received in revised form 21 December 1999; accepted 10 January 2000)