

RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY | 

# Ultrafast folding of $\alpha_3D$ : A *de novo* designed three-helix bundle protein

Yongjin Zhu, Darwin O. V. Alonso, Kosuke Maki, , and Feng Gai [Authors Info & Affiliations](#)December 11, 2003 100 (26) 15486-15491 <https://doi.org/10.1073/pnas.2136623100> 3,981 | 136

## Abstract

Here, we describe the folding/unfolding kinetics of  $\alpha_3D$ , a small designed three-helix bundle. Both IR temperature jump and ultrafast fluorescence mixing methods reveal a single-exponential process consistent with a minimal folding time of  $3.2 \pm 1.2 \mu\text{s}$  (at  $\approx 50^\circ\text{C}$ ), indicating that a protein can fold on the 1- to 5- $\mu\text{s}$  time scale. Furthermore, the single-exponential nature of the relaxation indicates that the prefactor for transition state (TS)-folding models is probably  $\geq 1 (\mu\text{s})^{-1}$  for a protein of this size and topology. Molecular dynamics simulations and IR spectroscopy provide a molecular rationale for the rapid, single-exponential folding of this protein.  $\alpha_3D$  shows a significant bias toward local helical structure in the thermally denatured state. The molecular dynamics-simulated TS ensemble is highly heterogeneous and dynamic, allowing access to the TS via multiple pathways.

### Sign up for PNAS alerts.

Get alerts for new articles, or get an alert when an article is cited.

[MANAGE ALERTS](#)

Proteins fold along a conformational landscape that begins with a largely disorganized, unfolded structure and proceeds to a highly compacted native state with greatly reduced conformational freedom. One model for protein folding posits a conformational search that proceeds via an initial hydrophobic collapse with further folding occurring within a restricted volume (1). Although many proteins exhibit multiple stages of

folding, some small, topologically simple proteins fold without detectable intermediates, indicating that the conformational search and collapse are sometimes essentially synchronous processes, which is described by the nucleation–condensation mechanism (2). This folding process can shift to the framework model with increasing secondary structure propensities (2). The folding of small proteins by either mechanism can occur on the millisecond time scale or faster (3–14). *De novo* peptide and protein design provides an attractive approach for probing the kinetics of protein folding because it allows one to define both the sequence and its resulting interactions. Unstructured peptides have been used to determine the rate of formation of the first contact between amino acid side chains in the random coil state (15–17). Also, studies on monomeric helices revealed that the formation of isolated helices occurs with a time constant of 100–300 ns (depending on the sequence) near room temperature (18, 19) and can be characterized by a relatively rough landscape (20). In contrast,  $\beta$ -hairpin formation in a model peptide occurs in  $\approx 6 \mu\text{s}$  (21). Finally, the role of electrostatic complementarity has been probed by examining the second-order rate of folding of designed coiled-coil peptides (22).

Here, we report folding/unfolding studies of  $\alpha_3D$ , a *de novo* designed three-helix bundle protein (23).  $\alpha_3D$  is an excellent system for probing the maximal rate of folding in a small globular protein. It was designed to be stabilized only by the packing of hydrophobic side chains and lacks buried polar residues or structured loops that might introduce significant kinetic barriers to folding (24). This protein adopts a well defined tertiary structure in which most of its solvent-inaccessible side chains adopt predominantly a single conformation (25). Also, the observed rates of amide hydrogen–deuterium exchange and the thermodynamic parameters for folding are within the range expected for a protein of its size, indicating that its core is relatively well packed (26).

## Materials and Methods

The amino acid sequence of  $\alpha_3D$  is given in ref. 25. The protein was expressed and purified as described (23). CD data were collected on an AVIV 62DS spectropolarimeter by using a 1-mm quartz cell. Fourier transform IR spectra were collected on a Nicolet Magna-IR 860 spectrometer by using  $2\text{-cm}^{-1}$  resolutions and a thermostated two-compartment  $\text{CaF}_2$  sample cell ( $52 \mu\text{m}$ ) (20).

The 3-ns, 10-mJ, 1.9- $\mu\text{m}$ , and 10-Hz temperature jump (*T*-jump) pulse was generated by Raman shifting the Nd:YAG fundamental, 1,064 nm (Infinity, Coherent, Santa Clara, CA), in a mixture of  $\text{H}_2$  and Ar pressurized at 750 psi. A *T*-jump of 10–15°C can be obtained routinely in an  $\approx 40\text{-nl}$  laser interaction volume, and the final temperature in the current study ranges from 45 to 84°C. A continuous wave lead salts IR diode laser (Laser Components, Wilmington, MA) was used as the probe. Transient absorbance changes of the probe induced by the *T*-jump pulses were detected by a 50-MHz mercury cadmium telluride detector (Kolmar Technologies, Newburyport, MA). Digitization of the signal was accomplished by using a Tektronix TDS 3052

digital oscilloscope (20).

For continuous-flow fluorescence measurements, we used a capillary mixer similar to that described (27). Mixing efficiency and dead time ( $103 \pm 10 \mu\text{s}$  for experiments at or above room temperature and  $168 \pm 10 \mu\text{s}$  for those at  $12^\circ\text{C}$ ) were determined by monitoring the quenching of *N*-acetyltryptophanamide fluorescence by *N*-bromosuccinimide in the presence of urea, which results in somewhat longer dead times compared to aqueous buffer. Folded  $\alpha_3D$  (in 2 M urea, pH 2.6) was mixed with a 10-fold excess of concentrated urea solution in 200 mM phosphoric acid at pH 2.6 to final urea concentrations between  $\approx 5.5$  and 9 M. Final protein concentration was 40–50  $\mu\text{M}$ .

Molecular dynamics (MD) simulations were performed by using ENCAD (28) and established protein (29) and water (30) potential functions. All simulations began with the NMR solution structure 2A3D (25). Two simulations each were performed at 100 and  $225^\circ\text{C}$  at low pH (Glu, Asp, and His protonated), as well as a  $25^\circ\text{C}$  control at neutral pH (Glu, Asp, and His ionized). The protein was solvated with water molecules extending at least 8 Å from any protein atom, yielding 2,893–2,994 waters. The box volume was adjusted to reproduce the experimental density appropriate to that temperature: 0.997, 0.958, and 0.829 g/ml for 25, 100, and  $225^\circ\text{C}$ , respectively (31). Several steps were performed to further prepare the system for MD. First, water alone was minimized for 5,001 steps, followed by 5,001 steps of MD and another 1,001 steps of minimization. The protein was then subjected to 501 steps of minimization. Finally, the entire system was minimized for 501 steps. Production simulations were then performed. A nonbonded cutoff of 8 Å with smooth, force-shifted truncation was used (29), and the nonbonded list was updated every two to five steps. Slightly different setups were used for the duplicate trajectories, which is equivalent to changing the random number seed for MD. Periodic boundary conditions and the minimum image convention were used to reduce edge effects within the microcanonical ensemble. The simulations were then performed by using a 2-fs time step, for a total of 409 ns or 0.409  $\mu\text{s}$ . Structures were saved every 0.2 ps for analysis, resulting in 250,000 structures for each 50 ns. The actual simulation times were: neutral pH, 298 K, 10 ns; low pH, 373\_1, 50 ns; 373\_2, 182 ns; 373\_3, 72 ns; and 498\_1, 25 ns; 498\_2, 50 ns; 498\_3, 20 ns (10 Å instead of 8 Å nonbonded cutoff was used for this simulation). Putative transition state (TS) ensembles were identified by using a conformational clustering procedure (32).

## Results

Ultrafast folding of  $\alpha_3D$ : A *de novo* designed three-helix bundle

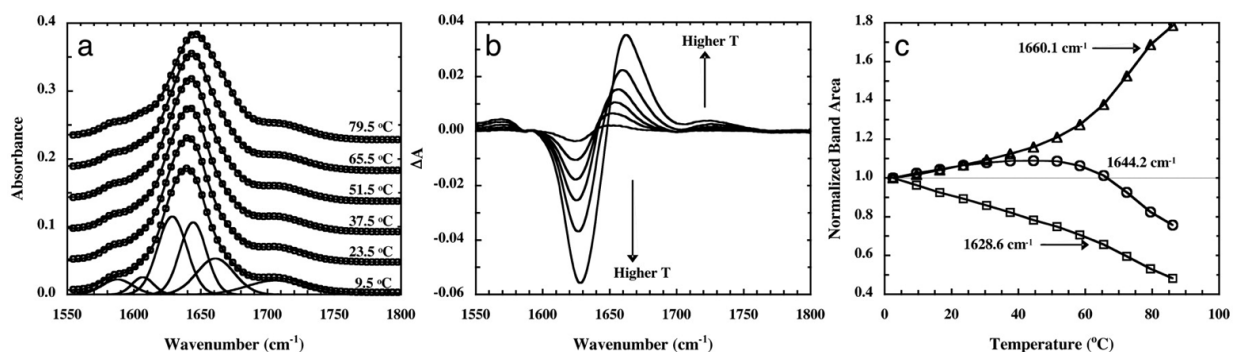


thermodynamic parameters for thermal unfolding of  $\alpha_3D$  at pH\* 2.2 ( $T_m = 73.2^\circ\text{C}$ ,  $\Delta H_m = -36 \text{ kcal}\cdot\text{mol}^{-1}$ , and

$\Delta S_m = -104 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ ), measured as described (26), are characteristic of small globular proteins (Fig. 6, which is published as supporting information on the PNAS web site). The computed value of  $\Delta C_p$ , approximately  $-650 \text{ cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ , is similar to that observed at higher pH (Fig. 6, which is published as supporting information on the PNAS web site).

IR spectroscopy provides an excellent probe of not only the secondary structure of a protein but also its degree of hydration (33). As a prelude to the kinetic measurements, we examined the equilibrium thermal unfolding of  $\alpha_3D$  at pH\* 2.2 by monitoring the amide I' band (Fig. 1a), which arises mainly from the amide C=O stretching vibration. The IR difference spectra (Fig. 1b) in this region indicate that increasing temperature results in a major decrease in IR absorbance at  $\approx 1,630 \text{ cm}^{-1}$  with a concomitant increase of  $\approx 1,660 \text{ cm}^{-1}$ . By using a global analysis technique, the IR spectra for the amide I' region were decomposed into six components (Fig. 1a), three of which are related to significantly populated backbone conformations and three of which are associated with side-chain vibrations. All three main-chain components show a pretransition baseline, followed by a thermal unfolding transition, which occurs at the same  $T_m$  as when the transition is followed by CD (Fig. 1c).

**Fig. 1.**



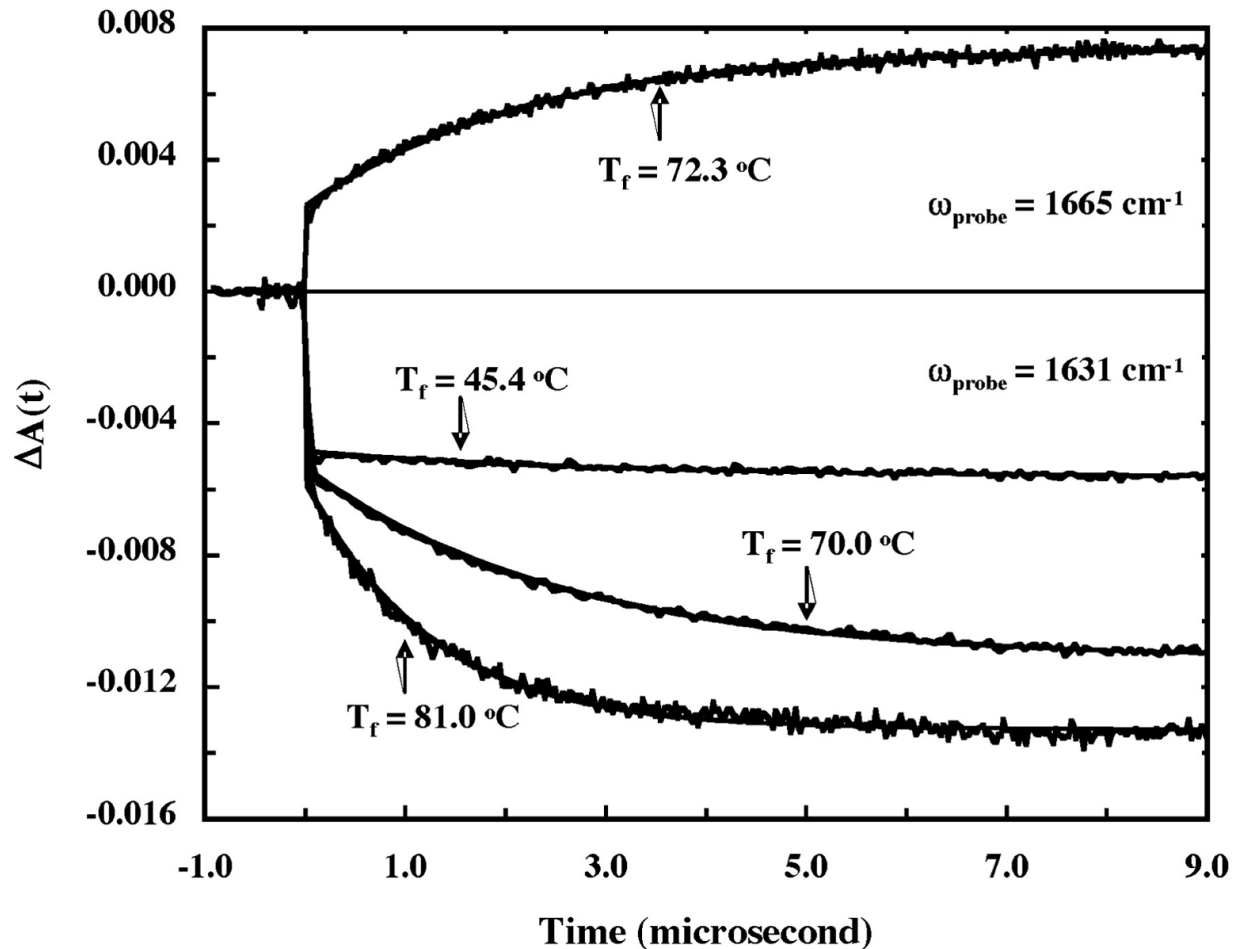
(a) IR spectra of  $\alpha_3D$  in  $D_2O$  (50 mM phosphate buffer, pH\* 2.2 uncorrected pH reading in  $D_2O$ ) were collected every 7 °C, from 2.5 to 86.5 °C. Representative IR spectra in the amide I' region ( $\circ$ ) are shown. These spectra (total 13) were modeled (solid lines) globally by 6 G with the following parameters:  $\nu_1 = 1,587.5 \text{ cm}^{-1}$ ,  $\Delta\nu_1 = 26.8 \text{ cm}^{-1}$ ;  $\nu_2 = 1,606.6 \text{ cm}^{-1}$ ,  $\Delta\nu_2 = 18.6 \text{ cm}^{-1}$ ;  $\nu_3 = 1,628.6 \text{ cm}^{-1}$ ,  $\Delta\nu_3 = 25.7 \text{ cm}^{-1}$ ;  $\nu_4 = 1,644.2 \text{ cm}^{-1}$ ,  $\Delta\nu_4 = 23.4 \text{ cm}^{-1}$ ;  $\nu_5 = 1,660.1 \text{ cm}^{-1}$ ,  $\Delta\nu_5 = 32.8 \text{ cm}^{-1}$ ; and  $\nu_6 = 1,705.7 \text{ cm}^{-1}$ ,  $\Delta\nu_6 = 49.0 \text{ cm}^{-1}$ ; where  $\nu$  is band position and  $\Delta\nu$  is full width at half maximum. The bands that make up the fit for the 9.5 °C spectrum are shown. The bands at 1,705 and 1,587  $\text{cm}^{-1}$  are due to protonated and deprotonated carboxylates, respectively, and the band at 1,606  $\text{cm}^{-1}$  arises from amino acid side chains. The remaining three bands are due to amide C=Os. (b) Difference IR spectra generated by subtracting the spectrum collected at 2.5 °C from the spectra in a. Arrows indicate the direction of changes when temperature is increased. (c) Relative band areas of the three amide bands as a function of temperature.

Based on isotopic labeling, the bands at 1,628 and 1,644  $\text{cm}^{-1}$  have been assigned to fully and/or partially hydrated and dehydrated amides (33), respectively. In the pretransition region, the hydrated amide peak decreases in intensity in a linear manner as a function of temperature. These changes appear to be related to the weakening of the hydrogen bonds with solvent molecules and fraying of the helices (33) (see also MD

*Simulations*). The peak at  $1,660\text{ cm}^{-1}$ , which is associated with nonhelical amides, also increases linearly with temperature up to the onset of the thermal unfolding transition after which it shows a sigmoidal increase typical of thermal unfolding curves. We assign this peak of the thermally unfolded state to amides that are not involved in stable secondary structures. Its peak position ( $1,660\text{ cm}^{-1}$ ) differs from the value of  $\approx 1,645\text{ cm}^{-1}$  reported for fully unstructured peptides but is close to that expected for short  $3_{10}$  helices and type I turns (34), which are frequently found near the ends of well structured helices and in noncooperatively formed “nascent” helices (35). These residual structural preferences in the denatured state of  $\alpha_3\text{D}$  may contribute to its rapid rate of folding.

**Kinetics of Unfolding of  $\alpha_3\text{D}$ .** The relaxation kinetics of  $\alpha_3\text{D}$  was measured by time-resolved IR spectroscopy after a laser-induced nanosecond  $T$ -jump (20). When monitored at  $1,631\text{ cm}^{-1}$ , the relaxation kinetics exhibit an initial instantaneous component followed by a first-order relaxation (Fig. 2). The instantaneous component is faster than the 10-ns rise time of our IR detection system and is attributed to very rapid changes in solvation and fraying of the ends of the helices. The slower phase is well resolved with a relaxation time constant of a few microseconds around the thermal  $T_m$  ( $\approx 73^\circ\text{C}$ ). Because the longest delay time used in this study was  $\approx 20\text{ }\mu\text{s}$ , we cannot rule out the possibility that additional kinetic events occur on a longer time scale. However, the kinetic amplitudes at all temperatures are within  $>90\%$  of the expected equilibrium values, suggesting that there are no major relaxation processes taking place on a longer time scale. Thus, we attributed the microsecond relaxation phase to the folding–unfolding transition of the  $\alpha_3\text{D}$  tertiary structure.

**Fig. 2.**

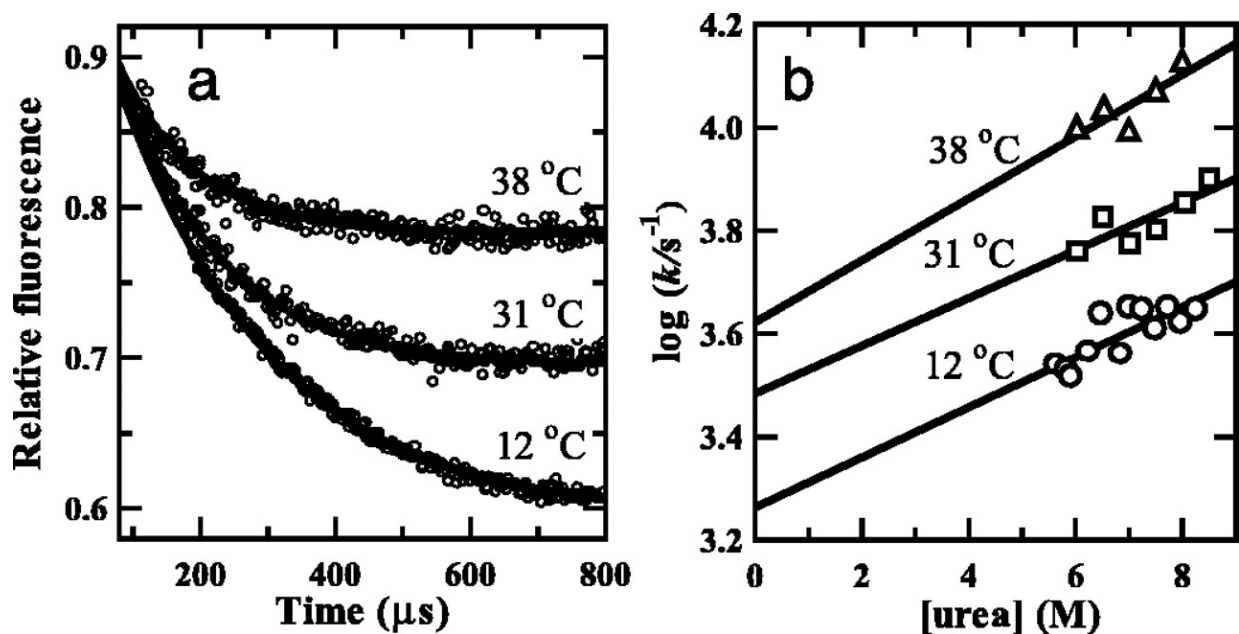


*T-jump-induced relaxation kinetics of  $\alpha_3D$  measured by time-resolved IR spectroscopy at 1,631 and 1,665  $\text{cm}^{-1}$  as well as different final temperatures. The fast phase rose instantaneously and was not resolvable with our instrument setup, whereas the slow phase was modeled by a single-exponential function. The relaxation time constants corresponding to different final temperatures are  $\tau$  (45.4 $^\circ\text{C}$ ) = 2.8  $\mu\text{s}$ ,  $\tau$  (70.0 $^\circ\text{C}$ ) = 2.6  $\mu\text{s}$ ,  $\tau$  (81.0 $^\circ\text{C}$ ) = 1.2  $\mu\text{s}$ , and  $\tau$  (72.3 $^\circ\text{C}$ ) = 2.2  $\mu\text{s}$ . The T-jump amplitude was  $\approx 10^\circ\text{C}$  for each case, and the final temperature is indicated.*

The kinetic measurements were extended to lower temperature in a series of experiments in which the intrinsic Trp fluorescence was monitored to report the overall tertiary structure of the protein. These experiments were conducted by using an efficient capillary mixer with a dead time of  $\approx 100 \mu\text{s}$  (27) for continuous-flow measurements of the rate of unfolding as a function of both temperature and urea concentration (Fig. 3). Equilibrium measurements of the fluorescence of the single Trp vs. urea concentration (pH 2.6) showed an unfolding transition characterized by a sigmoidal decrease in fluorescence between 3 and 6 M urea with a midpoint,  $C_m$ , of 4.6 M (Fig. 7). The fluorescence changes associated with unfolding, triggered by a rapid jump in urea concentration from 2 to 6 M, or above, exhibit a single-exponential decay over the 100- to 800- $\mu\text{s}$  time range (Fig. 3a). The rate of unfolding increases as a function of both urea concentration and temperature (Fig. 3b). At each temperature measured (12–38 $^\circ\text{C}$ ), the unfolding rate in the absence of urea was obtained by linear extrapolation (Fig. 3b). Similarly, refolding from the urea-denatured state (7 M) to final concentrations between 2.5 and 4.5 M (12 $^\circ\text{C}$ ) showed single-

exponential kinetics consistent with a two-state mechanism [linear dependence of  $\log(k_f)$  vs. urea concentration joining the rate of unfolding near  $C_m$ ]. However, at lower urea concentrations, deviations from strict two-state behavior are observed at 12°C. This may be due to accumulation of partially folded intermediates or increased ruggedness of the energy landscape. Transient association is unlikely to be responsible for the complex folding behavior because recent continuous-flow fluorescence experiments (data not shown) indicate that the folding kinetics at 12°C is independent of protein concentration over the 10- to 100- $\mu$ M range.

Fig. 3.

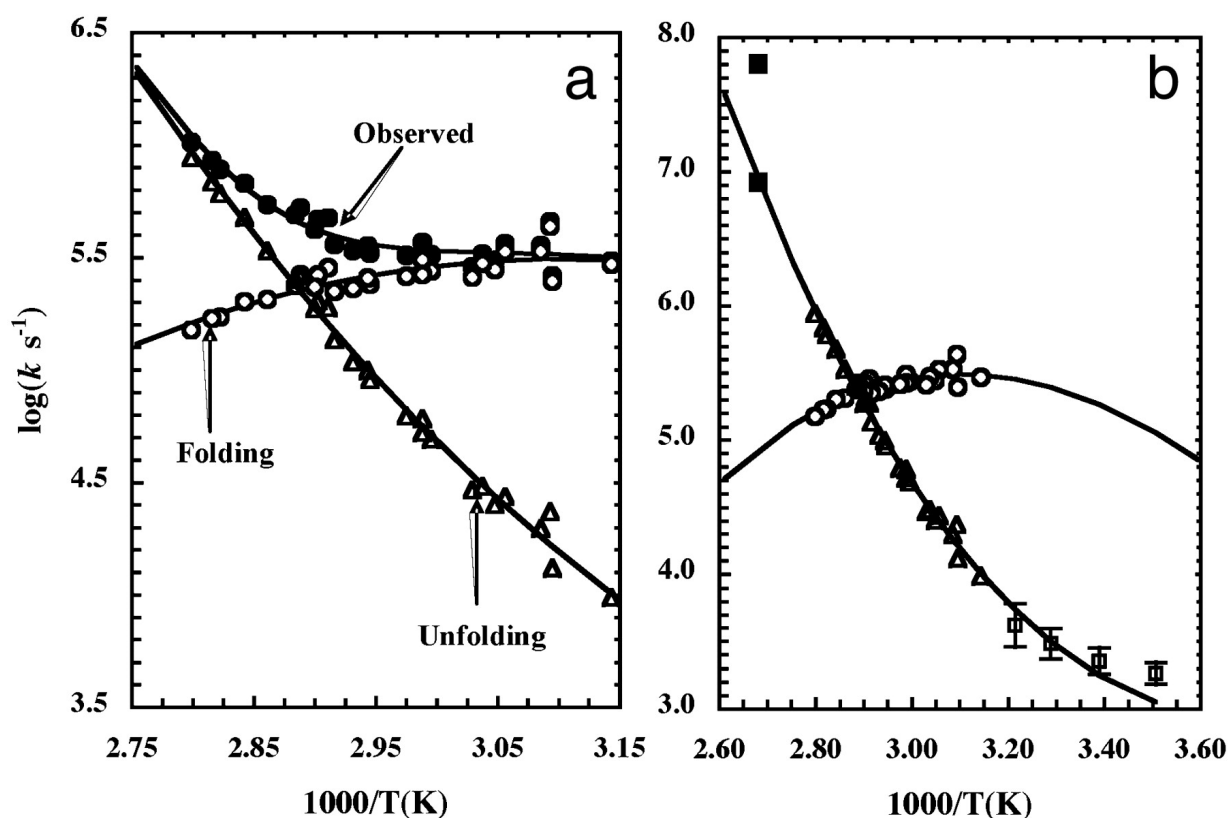


Unfolding kinetics of  $\alpha_3D$  measured by continuous-flow fluorescence. (a) Representative continuous-flow fluorescence traces, including unfolding experiments at 12, 31, and 38°C and a final urea concentration of 6.5 M. Solid lines represent single-exponential fits. (b) Semilogarithmic plot of the rate constant of unfolding vs. urea concentration; representative data at three temperatures are shown. The rates in the absence of urea, obtained by linear regression (lines), are included in Fig. 4b.

Fig. 4a illustrates an Arrhenius plot of the logarithm of the rate constants vs.  $1/T$  for the T-jump data. The folding and unfolding rate constants ( $k_f$  and  $k_u$ , respectively) were calculated from the observed rate constant ( $k_{obs}$ ) based on the relationships  $k_{obs} = k_f + k_u$  and  $K_{eq} = k_f/k_u$ . The plots show strong deviations from linearity; the plot for the folding rate is curved downward with a maximum near 49°C and a folding time of  $3.2 \pm 1.2 \mu$ s. Conversely, the unfolding curve shows upward curvature (Fig. 4a). Such nonlinear behavior has also been observed with other proteins (36–38) and in HP lattice models (39). Non-Arrhenius temperature dependence of folding may be the result of local traps on the energy landscape that retard folding under conditions where the driving force for folding is small (40). Another possibility is that the curvature originates from the difference in heat capacity between the folded or unfolded states and the TS

(36–38, 41). Our results appear to be consistent with this interpretation:  $\Delta G^\ddagger$  for both the folding and unfolding of  $\alpha_3D$  (Fig. 4) was calculated by using an equation similar to that commonly used to characterize the equilibrium thermal folding–unfolding transition. The extrapolated rates obtained from the continuous-flow measurements are in excellent agreement with the predicted values from  $T$ -jump down to room temperature (Fig. 4*b*). This agreement suggests that secondary and tertiary structure, which are probed by the  $T$ -jump IR and continuous-flow fluorescence methods, respectively, are essentially formed in parallel (kinetically indistinguishable). Furthermore, the equivalence of the rates of folding from the thermally unfolded and urea-unfolded states suggests that folding is occurring from a relatively unstructured state. As in the folding measurements, some divergence is observed at 12°C, possibly due to deviations from strict two-state behavior.

Fig. 4.



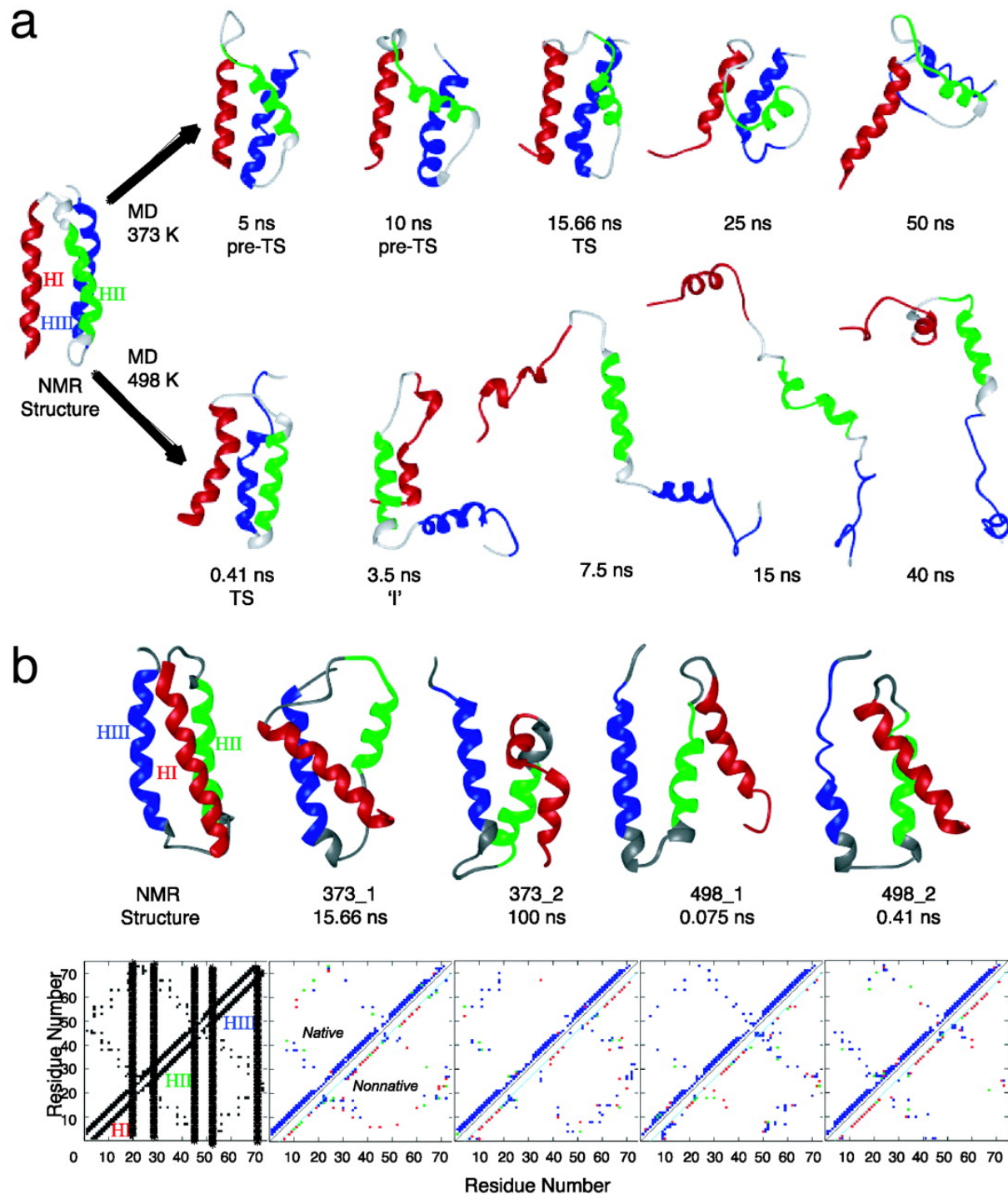
(a) Arrhenius plots of the observed ( $\bullet$ ), folding ( $\circ$ ), and unfolding ( $\Delta$ ) rate constants. Lines are fits to the Eyring equation, i.e.,  $\ln(k) = \ln(D) - \Delta G^\ddagger/RT$ , where  $D$  is a constant and  $\Delta G^\ddagger$  is the free energy of activation that is a function 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 73.2°C. Global fitting of the folding and unfolding kinetics yields the following thermodynamic parameters of activation when  $D = 10^6 \text{ s}^{-1}$  is used: (i) for folding,  $\Delta H^\ddagger(T_m) = -6.1 \text{ kcal}\cdot\text{mol}^{-1}$ ,  $\Delta S^\ddagger(T_m) = -20.5 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ,  $\Delta C_p^\ddagger = -255 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ; (ii) for unfolding,  $\Delta H^\ddagger(T_m) = 29.8 \text{ kcal}\cdot\text{mol}^{-1}$ ,  $\Delta S^\ddagger(T_m) = 83.1 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ,  $\Delta C_p^\ddagger = 392 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ . (b) Same as a but with an extended x axis. For comparison, unfolding rate constants ( $\square$ ) measured in mixing experiments are also shown. These data show that both folding and unfolding rates exhibit nonlinear temperature dependence. The folding time at 25°C is predicted to be  $\approx 4.8 \mu\text{s}$  according to the fit in a, and the unfolding time at 100°C is predicted to be  $\approx 100 \text{ ns}$ . The times required to reach the TS in two

*independent simulations at 100°C are shown as ■.*

Folding involves the burial of hydrophobic residues, which leads to a change in the heat capacity of the system. Therefore, the difference in heat capacity between the unfolded and the TSs, i.e.,  $\Delta C_p^\ddagger$ , provides a measure of how similar the TS structure is to the folded state. For  $\alpha_3D$ ,  $\Delta C_p^\ddagger$  was  $-255 (\pm 50) \text{ cal}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$  for folding. This value corresponds to  $\approx 40\%$  of the full heat capacity change between the folded and unfolded states, which is smaller than the values observed for many other small proteins, suggesting that the TS ensemble encountered during folding of  $\alpha_3D$  is relatively disordered or at least has a high solvent accessible surface area. In addition, the net enthalpy gained (approximately  $-6 \text{ kcal}\cdot\text{mol}^{-1}$ ) at the TS is only  $\approx 16\%$  of that gained on formation of the native state, which is also consistent with a relatively open, poorly packed TS.

**MD Simulations.** MD simulations of thermal unfolding were performed at 100 and 225°C to provide a structural framework for interpretation of the experimental results ([Fig. 5](#)). The protein reached the TS for folding within 1 ns at 225°C (75 and 410 ps, in two independent simulations) and later at 100°C (15.7 and 100.1 ns for two different simulations). The times to reach the TSs are in agreement with experiment ([Fig. 4b](#)). Extrapolation of the *T*-jump data to 100°C yields a time constant for unfolding (ensemble kinetics)  $\approx 100$  ns ([Fig. 4b](#)), which predicts probabilities  $(1 - e^{-t/\tau})$  of 15% and 63% for single-molecule trajectories to reach the TS within 16 and 100 ns, respectively.

**Fig. 5.**



Representative structures from the MD simulations. (a) Loss of helical and tertiary structure as a function of time for two of the simulations. The regions of the native helix are red (HI), green (HII), and blue (HIII). Thickened ribbons denote helical structure, as determined by using the *DSSP* algorithm (58). (b) A representative structure from each TS ensemble, identified by using a conformational clustering procedure (32), is displayed along with a side chain contact pay with native contacts shown above the diagonal and nonnative below. The map shows the average population of the contacts over a 5-ps TS ensemble, moving from blue to green to red for the most to least populated contacts.

Consistent with our interpretation of the IR data, the number of main-chain water hydrogen bonds

dropped (48 and 35 at 100°C and 225°C, respectively) and both the intra- and inter-protein hydrogen bonds were weakened with increasing temperature in the pre-TS region. These effects are due to the lower solvent density and increased water and peptide motion/entropy.

The TS structures show increases in total solvent-accessible surface area of 30–50% of the total change on unfolding, a value somewhat smaller than the  $\approx 60\%$  increase seen in  $\Delta C_p$  measurements, which are more closely related to changes in nonpolar surface area. The significance of this discrepancy is not clear at this time. The overall properties of the four independently derived TS ensembles are similar, despite the large spread in temperature, as observed for other proteins (8, 13, 42, 43). At a more detailed level, dynamic contacts between hydrophobic side chains stabilize the TS, but because of the nonspecific nature of these interactions, the nucleus is fairly delocalized, involving different residues at different times in various TS ensembles, as reflected in the helix packing angles and side-chain packing (Fig. 5b).

After the TS, the protein unfolds further, particularly at the higher temperature, at which improved sampling is achieved (Fig. 5a). The pathway of unfolding was quite variable with respect to loss of the helices. For example, in 373\_1, HI was retained much longer than the other two helices and HIII was more stable than HII ( $HI \gg HIII > HII$ ). A similar pattern was found for the 498\_1 simulation ( $HI \approx HIII \gg HII$ ), whereas the other two simulations showed a different order (373\_2,  $HIII \approx HII > HI$  and 498\_2,  $HII > HI > HIII$ ). The protein unfolds through an intermediate in some simulations; e.g., in 498\_2 an intermediate with altered helix packing was populated from 0.88 to 4.8 ns (see the 3.5-ns snapshot in Fig. 5a). Further unfolding led to a denatured state with residual helical structure [ $\approx 38\%$  at 225°C by consideration of ( $\varphi$ ,  $\phi$ ) angles (44) and 8% if we require that segments form  $\geq 6$  hydrogen bonds]. Thus, much of the sequence remains within the helical region of conformational space, but little well ordered and extended helical structure is formed. In addition, both  $3_{10}$ - and  $\pi$ -helical structure are populated in the denatured state, as reflected by the average number of  $i \rightarrow i + 3$  (6.7) and  $i \rightarrow i + 5$  (7.7) hydrogen bonds, respectively. This type and degree of residual structure is in agreement with the IR data.

## Discussion and Conclusions

The folding of  $\alpha_3D$  brings side chains into tertiary contacts that are separated by up to 50 residues in sequence. Two residues separated by this distance in sequence would be expected to collide on the 1- to 3- $\mu s$  time scale (15–17, 45–47), yet  $\alpha_3D$  is able to achieve the fully folded state in which all tertiary contacts are in place on this same time scale. This finding demonstrates the importance of the sequence in guiding and accelerating the folding process.  $\alpha_3D$  folds faster than any globular protein directly measured to date, forming its native structure at a rate that is roughly equivalent to the time scale of formation of a highly helical intermediate in the natural three-helix bundle protein (1.5  $\mu s$ ), the engrailed homeodomain (En-HD) (13). The acquisition of native structure in En-HD occurs on a somewhat longer time scale at room

temperature:  $\approx 15 \mu\text{s}$  (8, 13). Furthermore, mutants of  $\lambda$ -repressor headpiece fold with double exponential kinetics; only the phase with the lower amplitude is as fast as  $\alpha_3$ D (48).

TS theory for two-state protein folding predicts that the observed folding rate constant ( $k_{\text{fold}}$ ) is given by  $k_{\text{fold}} = k_{\text{mol}} \exp(-\Delta G^\ddagger/k_B T)$ , where  $\Delta G^\ddagger$  is the activation-free energy and  $k_{\text{mol}}$  is a molecular correlation time (48). This equation breaks down at very low activation free energies, leading to kinetic curves that are no longer well described by single exponentials and instead require multiple or stretched exponential to adequately describe the data. Thus, the finding that  $\alpha_3$ D displays single-exponential kinetics would predict that there is significant free energy of activation, thereby indicating that the effective molecular correlation time is  $< 3 \mu\text{s}$ . What then is the limiting value of  $k_{\text{mol}}$  for a protein of this size and topology? Recently, we demonstrated that monomeric helix formation can be best described by a stretched exponential with a correlation time of  $\approx 100\text{--}500 \text{ ns}$  (19), depending on the sequence and temperature. Furthermore, the observed rate constant depends rather weakly on the chain length. Thus, we expect  $k_{\text{mol}}$  to be similar to this value for small helical proteins, i.e.,  $10^6$  to  $10^7 \text{ s}^{-1}$ . Recently, Li *et al.* (49) also argued that  $\approx 1 (\mu\text{s})^{-1}$  should be a reasonable estimate for the prefactor of the TS theory expression of the folding rate.

Finally, we ask what features contribute to the exceptionally fast folding of  $\alpha_3$ D. Its relatively simple topology is one important factor, although its rate of folding is under-predicted by approximately a factor of 10 by consideration of contact order (50) alone. More specifically, the extreme speed of folding can be attributed to the design of the  $\alpha_3$ D sequence; the helical regions are already predisposed toward this conformation in the unfolded state by virtue of their intrinsic helical potential as well as numerous medium-range interactions between residues separated by a single helical turn. Native-like contacts in unfolded states are likely to speed up the search for the final conformation because preexisting hydrophobic clusters would help to reduce the entropic cost associated with folding, as well as to narrow the conformational space that must be searched along the folding pathway (51–53), provided they are not too strong (54). Our IR thermal melting results support such a scenario in the case of  $\alpha_3$ D. Furthermore, the helices can be increasingly stabilized as their apolar side chains become buried during the consolidation of the hydrophobic core. In addition, the core of  $\alpha_3$ D lacks interactions between buried polar side chains, which require more precise and hence less probable orientations of the interacting groups (24). Consequently, the packing in the TS ensemble is dynamic and variable, leading to a heterogeneous but topologically native-like TS, as illustrated in Fig. 5*b*. Interestingly, a heterogeneous TS has also been experimentally elucidated in a fast-folding, two-stranded coiled coil peptide (55).

It is also interesting to compare the folding process observed for  $\alpha_3$ D with that of En-HD, which folds by means of a highly structured intermediate in a manner predicted by the framework model of protein folding (13). En-HD folds rapidly because its starting point contains a high degree of helical structure, such that folding involves the correct docking of essentially preformed helices. With  $\alpha_3$ D, however, rapid folding

appears to be due to the heterogeneity of the packing interactions in a loose TS ensemble, which is easily approached from many avenues. By considering the MD simulations in reverse, the path leading from the denatured to the TS is highly diverse with respect to the order of helix formation. In addition,  $\alpha_3$ D is poised for folding through the bias for helical conformations (including  $\alpha$ -,  $3_{10}$ -, and  $\pi$ -helical structures) in the denatured state. So, unlike En-HD, folding of  $\alpha_3$ D proceeds from a disordered, but biased, denatured state, thereby covering more of the folding reaction coordinates. Moreover, the TS is easier to achieve because it is stabilized by relatively nonspecific and very rapidly interchangeable packing interactions. This dynamic flexibility is maintained to some small degree even in the native state, as determined by the order parameters of the internal side chains in the protein core. The  $S^2$  order parameters for the side-chain methyl groups in  $\alpha_3$ D are somewhat lower than several other native proteins, indicating that its core is on the more mobile end of the spectrum, but far more ordered than a molten globule (26).

In summary, En-HD and  $\alpha_3$ D seem to represent variations on a common folding mechanism (2).  $\alpha_3$ D appears to fold by means of the nucleation–condensation mechanism in which secondary and tertiary structure formation are linked, although there is support for a high energy intermediate from MD that is invisible to experiment except at low temperature. In contrast, En-HD contains a higher helical propensity and slides from nucleation–condensation to the framework model in which folding involves the docking of essentially preformed helices. Unlike  $\alpha_3$ D, En-HD populates a folding intermediate, which might contribute to its 5- to 10-fold slower folding rate. However, c-Myb, a structural homolog of En-HD, has a lower helical propensity, like that of  $\alpha_3$ D (data not shown), and it also folds by means of the nucleation–condensation model with an experimentally silent, high-energy intermediate (this intermediate becomes visible with a minor mutation) (56). Despite the lack of a well populated intermediate, c-Myb folds slower than both En-HD and  $\alpha_3$ D. Finally, the recent work of Plaxco and coworkers (57) also shows that other *de novo* designed proteins can fold as fast as or faster than their closest natural analog. Therefore, it will now be of considerable interest to characterize the determinants of the folding of these natural and designed proteins and the roles played by helical propensity, core hydrophobicity, and the rigidity of the native and TSs.

## Note

Abbreviations: T-jump, temperature jump; TS, transition state; MD, molecular dynamics; En-HD, engrailed homeodomain.

---

## Acknowledgments

We thank D. Thirumalai for helpful discussion. This work was supported by the National Science Foundation and the National Institutes of Health (GM-54616 to W.F.D.; CHE-0094077, GM-065978, and

5P41-RR01348 to F.G.; GM-056250 and CA06927 to H.R.; and GM-058079 to V.D.).

## Supporting Information

### Supporting Figure 6 Legend

Supporting Figure 6 Legend

DOWNLOAD

.79 KB

### Adobe PDF - 6623Fig6.pdf

Adobe PDF - 6623Fig6.pdf

DOWNLOAD

8.95 KB

### Supporting Figure 7 Legend

Supporting Figure 7 Legend

DOWNLOAD

.67 KB

### Adobe PDF - 6623Fig7.pdf

Adobe PDF - 6623Fig7.pdf

DOWNLOAD

7.33 KB

## References

1

Baldwin, R. L. (1989) *Trends Biochem. Sci.* **14**, 291–294.

↶

[Go to reference](#) | [Crossref](#) | [PubMed](#) | [Google Scholar](#)

2

Daggett, V. & Fersht, A. R. (2003) *Trends Biochem. Sci.* **28**, 18–25.

See all references ▾

 | [Crossref](#) | [PubMed](#) | [Google Scholar](#)

SHOW ALL REFERENCES

[VIEW FULL TEXT](#) | [DOWNLOAD PDF](#)

## Further reading in this issue

RESEARCH ARTICLE | DECEMBER 9, 2003 | ✓

### **Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments**

Marcus Fändrich, Vincent Forge, [...] Stephan Diekmann

RESEARCH ARTICLE | DECEMBER 8, 2003 | ✓

### **Markov chain Monte Carlo without likelihoods**

Paul Marjoram, John Molitor, [...] Simon Tavaré

RESEARCH ARTICLE | DECEMBER 15, 2003 | ✓

### **Activation of peroxisome proliferator-activated receptor $\delta$ induces fatty acid $\beta$ -oxidation in skeletal muscle and attenuates metabolic syndrome**

Toshiya Tanaka, Joji Yamamoto, [...] Juro Sakai

# Trending ⓘ

---

[RESEARCH ARTICLE](#) | NOVEMBER 25, 2025 | ⓘ

## **Tattoo ink induces inflammation in the draining lymph node and alters the immune response to vaccination**

In this study, we characterized the immune responses to the tattoo ink accumulating in the lymph nodes...

[Arianna Capucetti](#), [Juliana Falivene](#), [...] [Santiago F. Gonzalez](#)

[RESEARCH ARTICLE](#) | NOVEMBER 24, 2025 | ⓘ

## **A legacy of genetic entanglement with wolves shapes modern dogs**

Although wolves and dogs can interbreed and produce fertile offspring, hybridization between the two is far...

[Audrey T. Lin](#), [Regina A. Fairbanks](#), [...] [Logan Kistler](#)

[RESEARCH ARTICLE](#) | NOVEMBER 20, 2025 | ⓘ

## **The potential existential threat of large language models to online survey research**

Surveys are a primary source of data across the sciences, from medicine to economics. I demonstrate...

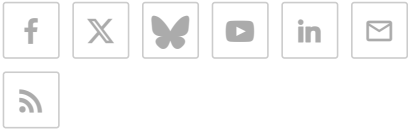
[Sean J. Westwood](#)

## Sign up for the PNAS Highlights newsletter

*Get in-depth science stories sent to your inbox twice a month.*

*name@example.com*

SUBSCRIBE >



BROWSE

- CURRENT ISSUE
- PNAS NEXUS*
- SPECIAL FEATURES
- LIST OF ISSUES
- TOPICS, COLLECTIONS, AND ARTICLE TYPES
- PNAS IN THE NEWS
- FRONT MATTER
- JOURNAL CLUB
- MULTIMEDIA
- PODCASTS

INFORMATION

- ABOUT
- SUSTAINABLE DEVELOPMENT GOALS
- EDITORIAL BOARD
- AUTHORS
- REVIEWERS
- SUBSCRIBERS
- LIBRARIANS
- PRESS
- COZZARELLI PRIZE
- PNAS UPDATES