1.5 mM $MgCl_2$, 420 mM KCl, 25% glycerol and 0.2 mM EDTA) Both buffers contained 1 mM dithiothreitol, 1 mM phenylmethylsulphonyl fluoride, 0.5 μg ml^{-1} leupeptin and 1 μM pepstatin. Each brain was homogenised in 5 ml of buffer A; nuclear proteins were eluted in 1 ml of buffer C. Extracts were not dialyzed. We carried out western blot transfers of CBP and p300 in cold transfer buffer containing 0.1% SDS for 2 h at 400 mA. HAT assays were done as described, with the nonspecific control antibody GST Z-5 (ref. 13). Western blot analysis of immunoprecipitates following the HAT assays showed that equivalent amounts of CBP or p300 precipitated from mutant and wild-type extracts (data not shown). We used equal amounts of wild-type and mutant total nuclear protein for western blot analysis and HAT assays.

Received 13 May; accepted 27 July 2002; doi:10.1038/nature01062.

1. Vo, N. K. & Goodman, R. H. CREB-binding protein and p300 in transcriptional regulation. *J. Biol. Chem.* **276,** 13505–13508 (2001).
2. Oike, Y. *et al.* Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. *Blood* **93,** 2771–2779 (1999).
3. Yao, T. P. *et al.* Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93,** 361–372 (1998).
4. Kung, A. L. *et al.* Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes Dev.* **14,** 272–277 (2000).
5. Gayther, S. A. *et al.* Mutations truncating the EP300 acetylase in human cancers. *Nature Genet.* **24,** 300–303 (2000).
6. Goodman, R. H. & Smolik, S. CBP/p300 in cell growth, transformation, and development. *Genes Dev.* **14,** 1553–1577 (2000).
7. Parker, D. *et al.* Analysis of an activator:coactivator complex reveals an essential role for secondary structure in transcriptional activation. *Mol. Cell* **2,** 353–359 (1998).
8. Parker, D. *et al.* Role of secondary structure in discrimination between constitutive and inducible activators. *Mol. Cell. Biol.* **19,** 5601–5607 (1999).
9. Radhakrishnan, I. *et al.* Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* **91,** 741–752 (1997).
10. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).
11. Munoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245,** 275–296 (1995).
12. Kraus, W. L., Manning, E. T. & Kadonaga, J. T. Biochemical analysis of distinct activation functions in p300 that enhance transcription initiation with chromatin templates. *Mol. Cell. Biol.* **19,** 8123–8135 (1999).
13. Yang, C., Shapiro, L. H., Rivera, M., Kumar, A. & Brindle, P. K. A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. *Mol. Cell. Biol.* **18,** 2218–2229 (1998).
14. Jayaraman, G. *et al.* p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. *J. Biol. Chem.* **274,** 17342–17352 (1999).
15. Mucenski, M. L. *et al.* A functional c-*myb* gene is required for normal murine fetal hepatic hematopoiesis. *Cell* **65,** 677–689 (1991).
16. Allen, R. D., Bender, T. P. & Siu, G. c-Myb is essential for early T cell development. *Genes Dev.* **13,** 1073–1078 (1999).
17. Herschlag, D. & Johnson, F. B. Synergism in transcriptional activation: a kinetic view. *Genes Dev.* **7,** 173–179 (1993).
18. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291,** 1001–1004 (2001).
19. Jacobsson, S. *et al.* Flow cytometric analysis of megakaryocyte ploidy in chronic myeloproliferative disorders and reactive thrombocytosis. *Eur. J. Haematol.* **56,** 287–292 (1996).
20. Jackson, C. W., Steward, S. A., Chenaille, P. J., Ashmun, R. A. & McDonald, T. P. An analysis of megakaryocytopoiesis in the C3H mouse: an animal model whose megakaryocytes have 32N as the modal DNA class. *Blood* **76,** 690–696 (1990).
21. Sumner, R., Crawford, A., Mucenski, M. & Frampton, J. Initiation of adult myelopoiesis can occur in the absence of c-Myb whereas subsequent development is strictly dependent on the transcription factor. *Oncogene* **19,** 3335–3342 (2000).
22. Rudolph, D. *et al.* Impaired fetal T cell development and perinatal lethality in mice lacking the cAMP response element binding protein. *Proc. Natl Acad. Sci. USA* **95,** 4481–4486 (1998).
23. Arnold, J. T. *et al.* A single injection of pegylated murine megakaryocyte growth and development factor (MGDF) into mice is sufficient to produce a profound stimulation of megakaryocyte frequency, size, and ploidization. *Blood* **89,** 823–833 (1997).
24. Kasper, L. H. *et al.* CREB binding protein interacts with nucleoporin-specific FG repeats that activate transcription and mediate NUP98-HOXA9 oncogenicity. *Mol. Cell. Biol.* **19,** 764–776 (1999).
25. Shapiro, L. H. Myb and Ets proteins cooperate to transactivate an early myeloid gene. *J. Biol. Chem.* **270,** 8763–8771 (1995).
26. Eckner, R. *et al.* Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev.* **8,** 869–884 (1994).
27. Lecine, P., Blank, V. & Shivdasani, R. Characterization of the hematopoietic transcription factor NF-E2 in primary murine megakaryocytes. *J. Biol. Chem.* **273,** 7572–7578 (1998).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements We thank M. Castor, D. Bush, S. Kelly and M. Harper for technical help; J. Gatewood, M. Paktinat, R. Cross, R. Ashmun and A. Hoffmeyer for FACS analyses; S. Steward and T. Pestina for help with the platelet counts; R. Piekorz for assistance with the bone marrow transplants; X. Xiong and J. Schroeder for advice on statistics; A. Hoffmeyer for discussions; R. Shivdasani for advice on purifying megakaryocytes; H. Singh for c-Myb mice; and the Transgenic/Gene Knockout Shared Resource for technical assistance. This work was supported by a grant from the NIH and from the National Cancer Institute Cancer Center Support (CORE) program, and by the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.B. (e-mail: paul.brindle@stjude.org).

Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy

Benjamin Schuler*†, Everett A. Lipman*† & William A. Eaton*

** Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA*
† These authors contributed equally to this work

Protein folding is inherently a heterogeneous process because of the very large number of microscopic pathways that connect the myriad unfolded conformations to the unique conformation of the native structure. In a first step towards the long-range goal of describing the distribution of pathways experimentally, Förster resonance energy transfer[1] (FRET) has been measured on single, freely diffusing molecules[2–4]. Here we use this method to determine properties of the free-energy surface for folding that have not been obtained from ensemble experiments. We show that single-molecule FRET measurements of a small cold-shock protein expose equilibrium collapse of the unfolded polypeptide and allow us to calculate limits on the polypeptide reconfiguration time. From these results, limits on the height of the free-energy barrier to folding are obtained that are consistent with a simple statistical mechanical model, but not with the barriers derived from simulations using molecular dynamics. Unlike the activation energy, the free-energy barrier includes the activation entropy and thus has been elusive to experimental determination for any kinetic process in solution.

The basic concepts of our FRET experiment are shown in Fig. 1. A green fluorescent donor dye and a red fluorescent acceptor dye were attached to cysteine residues introduced at the amino and carboxy termini of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima* (Csp*Tm*). This protein was selected for study because of its high stability, which makes it tolerant to structural perturbations, and the simplicity of its thermodynamic and kinetic behaviour in ensemble measurements[5]. If a folded Csp*Tm* molecule diffuses into the volume illuminated by a focused laser beam, then excitation of the donor dye results in rapid energy transfer to the acceptor dye because the termini are separated by only 1 nm, and most of the fluorescence photons are emitted by the acceptor. On addition of a chemical denaturant the protein unfolds, which results in a larger average distance between the donor and acceptor dyes. Consequently, the energy transfer rate is decreased and the fraction of photons emitted by the acceptor is lower. To help us interpret the results quantitatively, we used a control system consisting of two different lengths of type II polypro-

line helices labelled with the same dyes (Fig. 1). Polyproline provides a rigid spacer between donor and acceptor[6–8], which means that the interdye distance is independent of denaturant concentration, whereas all other parameters are expected to vary in the same way that they do in the protein.

Figure 2a and b shows parts of two typical data sets for the polyproline control. For $(Pro)_6$, the bursts of counts (that is, the detected photons) above background resulting from diffusion of single molecules into the illuminated volume come mostly from the red fluorescing acceptor dye, which shows that FRET is occurring (Fig. 2a). In $(Pro)_{20}$ there is a larger separation between the dyes (Fig. 1) and, on average, comparable numbers of green and red counts are observed in each burst (Fig. 2b). The apparent FRET efficiency (E_{app}) for each burst is calculated as the ratio of acceptor counts to the sum of acceptor plus donor counts. Figure 2c and d shows the E_{app} distributions for $(Pro)_6$ and $(Pro)_{20}$ as a function of the concentration of guanidinium chloride (GdmCl). For $(Pro)_{20}$, the distribution of E_{app} peaks near 50%. This E_{app} is higher than might be expected from the 6.2-nm polyproline helix, considering that R_0 (the distance at which our dye pair is expected to exhibit 50% transfer) is 5.4 nm. It results from the long flexible linkers of the dyes (Fig. 1), which allow the dyes to approach each other during the fluorescence lifetime of the donor. The additional maximum at $E_{app} \approx 0$ arises from $(Pro)_{20}$ molecules in which the acceptor dye has been chemically altered by photodestruction, or from $(Pro)_{20}$ molecules labelled only with the donor dye that were not completely removed during preparation.

For Csp*Tm*, three subpopulations are clearly resolved in the histograms of E_{app} (Fig. 2e), corresponding to folded molecules with large E_{app}, unfolded molecules with intermediate E_{app}, and molecules with $E_{app} \approx 0$ owing to a missing or inactive acceptor, as found for $(Pro)_{20}$. The finding of only folded and unfolded populations, whose relative proportions change with increasing GdmCl concentration, is exactly what is expected for this protein, which, like unlabelled Csp*Tm*[5,9], shows two-state behaviour in ensemble equilibrium and kinetic measurements (Supplementary Information). Similarly, two populations have been resolved in single-molecule FRET measurements[4] of chymotrypsin inhibitor 2 (CI2), which is also known to show two-state behaviour from ensemble measurements[10].

Each E_{app} distribution for Csp*Tm* was fit with a sum of one gaussian and two lognormal functions, for the unfolded, folded and donor-only protein peaks, respectively. Figure 3 shows how the means and widths of the E_{app} distributions for $(Pro)_{20}$ compare with those of the protein. As pointed out previously[4], resolving the FRET efficiencies for the folded and unfolded subpopulations can expose changes in $\langle E_{app} \rangle$ for the unfolded protein that cannot be extracted from equilibrium ensemble measurements. For unfolded Csp*Tm*, $\langle E_{app} \rangle$ clearly increases between 3 and 0 M GdmCl (Fig. 3b). Solvent effects on the dyes can be excluded as a source of the increased FRET efficiency for the unfolded protein, because $\langle E_{app} \rangle$ for $(Pro)_{20}$ shows almost no change over this range of GdmCl

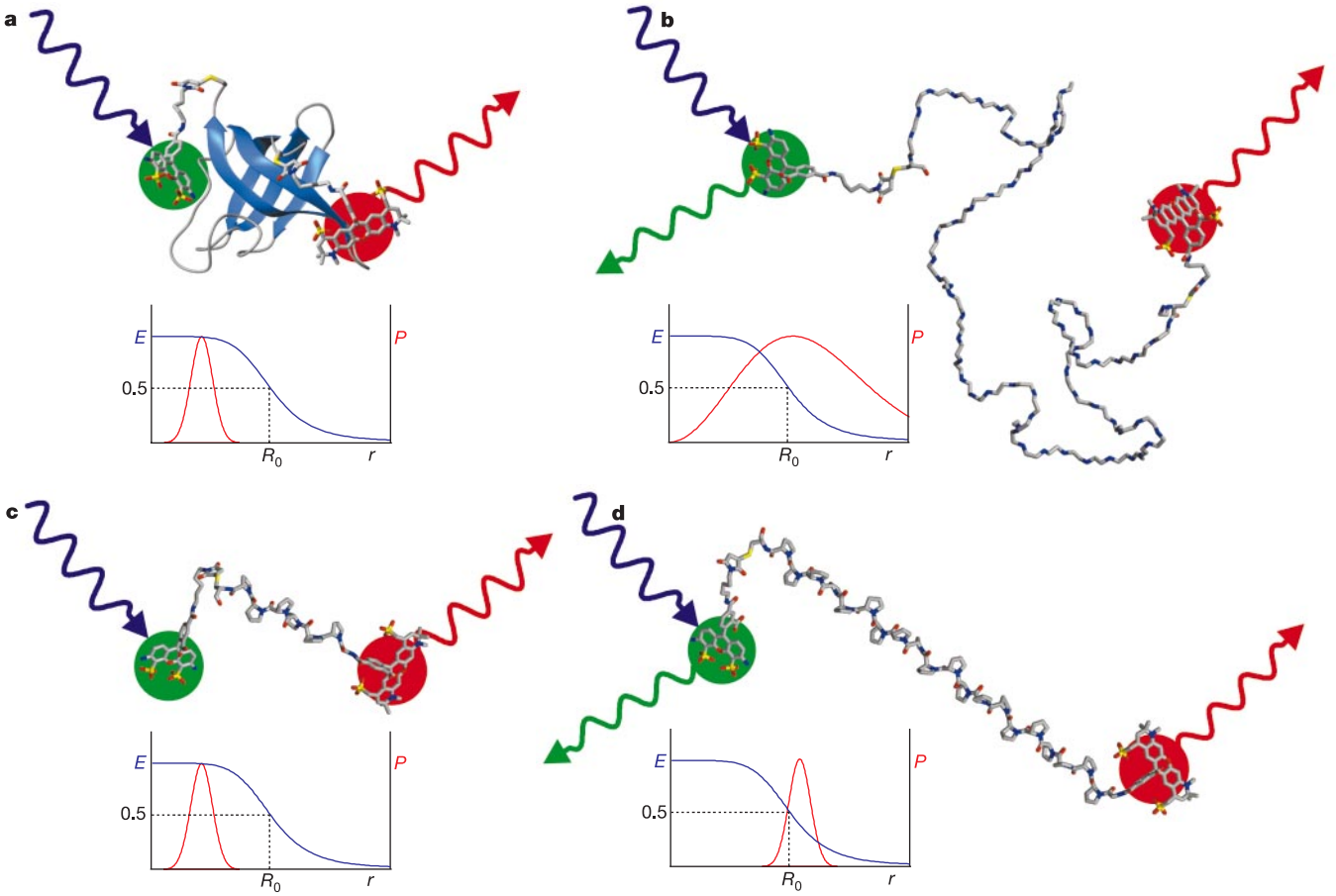

Figure 1 Schematic structures of protein and polyproline helices labelled with donor (Alexa 488) and acceptor (Alexa 594) dyes (using the program MOL/MOL). **a**, Folded Csp*Tm*, a 66-residue, five-stranded β-barrel protein (Protein Data Bank accession code 1G6P)[31]; **b**, unfolded Csp*Tm*; **c**, $(Pro)_6$; and **d**, $(Pro)_{20}$. A blue laser excites the green-emitting donor dye, which can transfer excitation energy to the red-emitting acceptor dye at a rate that depends on the inverse sixth power of the interdye distance[1,8]. In each case, the functional form of the FRET efficiency *E* versus distance (blue curve) is shown, as well as a representation of the probability distribution of distances between donor and acceptor dyes, *P* (red curve).

concentrations (Fig. 3a). Therefore, we must be observing a decrease in the average end-to-end distance, which indicates collapse to more compact denatured structures at low denaturant concentration. The change in the size of the unfolded protein is qualitatively consistent with predictions of an approximate mean field theory for a weakly hydrophobic sequence[11]. But the predicted continuous expansion of the chain between 3 and 6 M GdmCl is not observed for Csp*Tm* (Fig. 3b) and is also not observed in either ensemble FRET[12] or small-angle X-ray scattering experiments on unfolded cytochrome *c* (ref. 13). The most obvious explanation is that above 3.0 M GdmCl the unfolded polypeptide binding sites become saturated with denaturant molecules.

Important dynamical information is contained in the widths of the E_{app} distributions, that is, the range of FRET efficiencies observed for individual bursts of photons. Notably, the width of the peak corresponding to the unfolded protein is the same to within experimental error as that observed for $(Pro)_{20}$ (Fig. 3c), in spite of the large difference in flexibility between the two polypeptides. Because the end-to-end distance of polyproline is fixed, the FRET efficiency is nearly the same for all molecules and so a sharp distribution would be expected (Fig. 1). The width of the distribution observed for $(Pro)_{20}$ arises primarily from fluctuations in E_{app} owing to the small number of photons detected in each burst (see below). By contrast, an unfolded protein has a broad distribution of end-to-end distances (Fig. 1b), which makes the resulting transfer efficiency distributions sensitive to the polypeptide dynamics (Fig. 4). If the chain motion were infinitely slow relative to the observation time *t*, then every molecule diffusing through the focus would show a different FRET efficiency, resulting in an extremely broad distribution (Fig. 4a). If, in the opposite limit, the chain motion were sufficiently rapid for the unfolded protein molecule to explore most of its accessible conformational space during the observation time, then the same FRET efficiency would result for every molecule (in the absence of noise and other effects not related to distance, Fig. 4b).

The lack of any width in the E_{app} distributions for the unfolded protein in excess of that observed for $(Pro)_{20}$ (Fig. 3c) indicates that the end-to-end distance distribution for the unfolded protein does not contribute to the distribution of transfer efficiencies. The unfolded protein must therefore be reconfiguring fast relative to *t*. The relaxation time of the FRET efficiency autocorrelation function, τ_E, is given by $t(\sigma^2_{app} - \sigma^2_0)/2\sigma^2_E$ for $\tau_E < t$ (ref. 14), where σ^2_{app} is the variance in E_{app} for the unfolded protein, σ^2_0 is the variance owing to noise and other effects not dependent on the interdye distance, and σ^2_E is the variance of the FRET efficiency resulting from the underlying equilibrium distance distribution (Fig. 4a). For a gaussian chain with a mean squared end-to-end distance $\langle r^2 \rangle = R^2_0$, this relationship yields a polypeptide reconfiguration time of $\tau_0 = 9.8t(\sigma^2_{app} - \sigma^2_0)$. Equating σ^2_0 with the variance for $(Pro)_{20}$ and recognizing that, to within experimental error, σ_{app} cannot be more than 25% larger than σ_0 (Fig. 3c), the maximum value for τ_0 consistent with our observations is $t/40$. Variation in the time intervals used in the data collection indicates that the average *t* is about 1 ms, so that the upper limit for τ_0 is about 25 μs.

Our interpretation of the E_{app} distributions is different from those given previously[3,4], in which the additional width beyond that predicted from shot noise (the variation in count rates about fixed

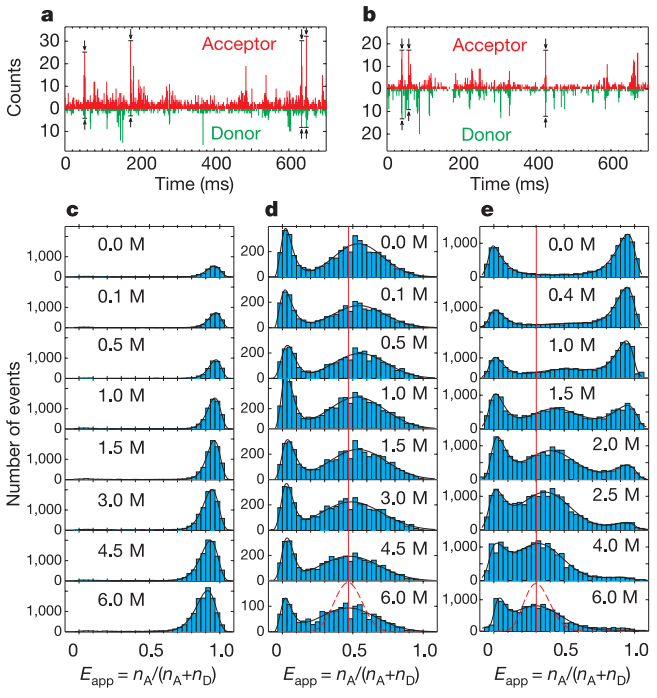

Figure 2 FRET trajectories and histograms. **a**, **b**, Donor (green) and acceptor (red) channel time traces using 1-ms bins for labelled $(Pro)_6$ (**a**) and $(Pro)_{20}$ (**b**). Arrows indicate photon bursts for which the sum of the counts in the two channels is greater than 25. **c**–**e**, Histograms of measured FRET efficiencies (E_{app}) at various GdmCl concentrations for labelled $(Pro)_6$ (**c**), $(Pro)_{20}$ (**d**) and Csp*Tm* (**e**). The black curves are the best fits to the data using lognormal and/or gaussian functions. The red dashed curves were calculated from the β-distribution[15] $P(E_{app}) = E_{app}^{\langle n_A \rangle}(1 - E_{app})^{\langle n_D \rangle}$, where $\langle n_A \rangle$ and $\langle n_D \rangle$ are the average number of detected acceptor and donor photons in the significant bursts.

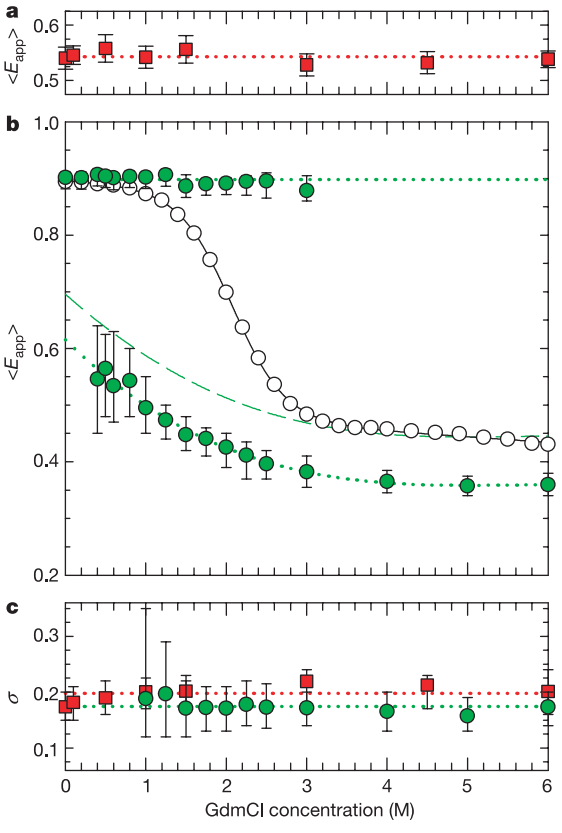

Figure 3 Dependence of the means and widths of the measured FRET efficiency (E_{app}) on the concentration of GdmCl. **a**, $\langle E_{app} \rangle$ for $(Pro)_{20}$. **b**, Single molecule mean values (filled circles), ensemble FRET efficiencies (open circles), and associated two-state fit (unbroken curve) for Csp*Tm*. The dotted curve is a third-order polynomial fit to the unfolded protein data that was matched (dashed curve) to the ensemble data between 4 and 6 M GdmCl. **c**, Standard deviations (σ) taken from the gaussian fits to the $(Pro)_{20}$ data (squares) and unfolded Csp*Tm* data (circles) in Fig. 2. Error bars indicate uncertainty in the fits.

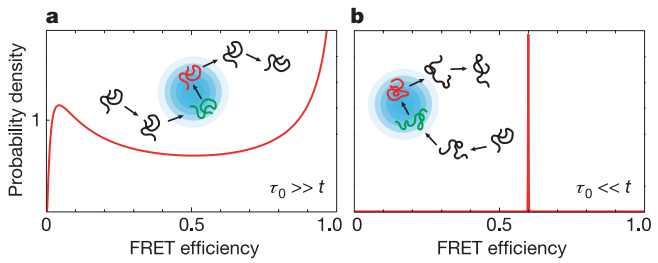

Figure 4 Two limiting cases for polypeptide dynamics in experiments on freely diffusing molecules. **a**, If the end-to-end distance of the protein does not change during the observation period in the illuminated volume (blue), then the distribution of transfer efficiencies reflects the equilibrium end-to-end distance distribution of the molecules and consequently results in a very broad probability distribution of FRET efficiencies, shown here for a gaussian chain (red curve, see Supplementary Information). **b**, If the molecule reconfigures fast relative to the time it takes to diffuse through the illuminated volume, then the FRET efficiency averages completely and (in the absence of shot noise and other broadening effects) is the same for every molecule.

means due to the discrete nature of the signals; ref. 15 and Fig. 2d, e) was attributed to fluctuations of interdye distances on a timescale comparable to or slower than the observation time. Although the possible pitfalls were recognized[3,4,15], Förster's equation, $E = 1/(1 + r^6/R_0^6)$ (ref. 1), was used in those studies to transform the E_{app} histograms for the folded and unfolded populations into histograms of interdye distances, r. But E_{app} widths very similar to those of unfolded CI2 (ref. 4) were observed previously for double-stranded DNA labelled with the same donor and acceptor dyes[16]. Consequently, it is unlikely that there is a significant contribution to the E_{app} widths for CI2 from a distribution of distances, in agreement with our interpretation of the widths measured for unfolded Csp*Tm*. Therefore, the widths in excess of shot noise of the E_{app} distributions for Csp*Tm* and CI2 must be caused by other sources, such as nonrandom photon emission intervals resulting from triplet state formation or intensity variation across the focal volume. Quantifying these contributions to the width of the E_{app} distributions will require detailed experimental and theoretical investigation, but this issue does not affect any of the conclusions reached below.

The finding of a reconfiguration time less than 25 μs can be related directly to important theoretical considerations in protein folding. Extensive studies suggest that essential features of the dynamics of folding can be captured by describing the process as diffusion on a low-dimensional free-energy surface. The free energy as a function of global coordinates, such as the fraction of native inter-residue interactions[17–19], contains the depths and average properties of the wells that correspond to thermodynamic states, and the positions and heights of the free-energy barriers separating these states which dictate the rates. Conventional equilibrium and kinetic studies yield the free-energy differences between the thermodynamic states, and the rates connecting them, but not the heights of the free-energy barriers. But Kramers theory, which accurately reproduces the kinetics for simplified representations of proteins[20–22], shows how the free-energy barrier height can be obtained if the polypeptide reconfiguration time is known. Kramers equation[23] for the folding time τ_f is

$$\tau_f = \frac{2\pi\omega_{min}\tau_0}{\omega_{max}}\exp\left(\frac{\Delta}{k_BT}\right) \approx 2\pi\tau_0\exp\left(\frac{\Delta}{k_BT}\right) \quad (1)$$

where ω_{min} and ω_{max} are, respectively, the frequencies that characterize the curvature of the (one-dimensional) free-energy surface in the harmonic well of the unfolded state and at the barrier top, Δ is the height of the folding free-energy barrier, k_B is Boltzmann's constant, T is the absolute temperature, $\tau_0 (= k_BT/m\omega_{min}^2 D)$ is the reconfiguration time in the unfolded well, and D is the diffusion constant for motion along the coordinate. Assuming the simplest case in which $\omega_{min} \approx \omega_{max}$, as found in lattice[20] and off-lattice models[21], we can determine Δ from τ_f and τ_0. A lower bound of $4k_BT$ is obtained from the upper limit on τ_0 of 25 μs (calculated above) and the corresponding ensemble folding time of $\tau_f = 12$ ms at 20 °C (at 0 M GdmCl) for the dye-labelled protein from stopped-flow kinetic measurements (Supplementary Information). An upper bound can be estimated from the reconfiguration time for a gaussian chain, which is simply $\langle r^2\rangle/3D$, where $\langle r^2\rangle \approx 22\,nm^2$ is the mean-squared end-to-end distance for the unfolded protein that can be calculated from $\langle E_{app}\rangle \approx 0.7$ in the absence of denaturant (Fig. 3b and Supplementary Information), and D is the relative end-to-end diffusion constant. Using $D < 1.7 \times 10^{-6}\,cm^2\,s^{-1}$ from a study of end-to-end contact rates in disordered peptides[24], $\tau_0 > 40$ ns and $\Delta < 11k_BT$. We regard this value of D as an upper limit because it has been determined for peptides containing about 30% glycines; it is expected to be smaller for Csp*Tm* because of side chain interactions and the smaller fraction of glycines (15%) in the protein sequence.

Using different and less rigorous arguments, estimates of the pre-exponential factor in equation (1)—that is, $2\pi k_BT/m\omega_{min}\omega_{max}D$—have ranged from 0.05 μs (ref. 25) to 1 μs (ref. 26) for a generic small protein, as compared with 0.3 μs to 0.2 ms for the specific protein studied here. The final result is that we have placed bounds on the free-energy barrier to folding of $11k_BT > \Delta > 4k_BT$. Together with the (viscosity-corrected) activation enthalpy for folding of $5k_BT$ (Supplementary Information), these bounds on the free-energy barrier correspond to an activation entropy between $-6k_B$ and $+1k_B$. The small activation entropy presumably results from cancellation of two large contributions upon forming the transition state ensemble—the entropy loss from ordering the polypeptide and the entropy gain from forming hydrophobic contacts.

The determination of bounds on a free-energy barrier height provides a previously unavailable benchmark for theoretical free-energy surfaces[27]. Although there are no detailed theoretical calculations for Csp*Tm* as yet, two other cold-shock proteins, CspA and CspB, that have the same structure and folding rates as Csp*Tm* have been studied. Using a simple statistical mechanical model that considers only inter-residue interactions present in the folded structure, free-energy barriers of $15k_BT$, $11k_BT$ and $7k_BT$ have been calculated for CspB as the allowed number of native stretches of polypeptide in the molecule was increased from one to two to three, respectively[28]. Our experimental results are, not surprisingly[27], inconsistent with the single-sequence approximation but are consistent with double- and triple-sequence approximations. In contrast, calculations of a free-energy surface derived from molecular dynamics simulations that also only consider native interactions in an α-carbon representation of CspA show no free-energy barriers under folding conditions[29]. In addition, the radius of gyration of the simulated unfolded CspA protein at the folding temperature is only 1.4 nm (where a barrier exists but is only $1.5k_BT$), compared with 2.3 nm calculated from $\langle E_{app}\rangle$ for unfolded Csp*Tm* at the midpoint concentration of denaturant. The free-energy surface for CspA calculated from all-atom molecular dynamics simulations at room temperature also predicts the absence of a barrier[29], whereas the observation of exponential folding kinetics[30] for CspA suggests a free-energy barrier crossing, as found for Csp*Tm*.

Note added in proof: C. L. Brooks (ref. 29) has communicated to us that the calculated free energy surface for an α-carbon representation of CspB is more consistent with our experimental results, showing a barrier under folding conditions of about $1k_BT$ and a radius of gyration of 2.0 nm for the unfolded state at the folding temperature. □

Methods

Synthesis and labelling of Csp*Tm* and polyproline

Cysteine residues were introduced at the N terminus after Met 1 and at the C terminus by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes. To exclude potential complications owing to proline *cis–trans* isomerization in Csp*Tm*, Pro 57 was replaced with glycine. We carried out expression and purification as described[31]. Dye labelling was carried out by procedures described by the manufacturer (Molecular Probes). Alexa Fluor 488 maleimide was reacted with the protein, and singly labelled protein was separated from unlabelled and doubly labelled protein by ion exchange chromatography (MonoQ HR 5/5, Amersham Pharmacia). The fractions containing singly labelled Csp*Tm*, as confirmed by electrospray ionization mass spectroscopy, were labelled with Alexa Fluor 594 maleimide. Doubly labelled protein was again separated from singly labelled protein by ion exchange chromatography.

We synthesized polyproline peptides containing 6 and 20 prolines using standard FastMoc chemistry, and incorporated a C-terminal cysteine and an N-terminal glycine as reactive groups for dye labelling. After cleavage and HPLC purification, fractions containing the pure peptide, as confirmed by mass spectroscopy, were labelled with Alexa Fluor 488 maleimide. Singly labelled peptide was purified by size-exclusion chromatography (Amersham Pharmacia) and labelled with Alexa Fluor 594 succinimidyl ester. Doubly labelled peptide was again purified by size-exclusion chromatography.

For single-molecule experiments, samples of labelled protein or peptide were diluted to a concentration of 75 pM or 38 pM, respectively, in 50 mM sodium phosphate (pH 7) plus 0.01% Tween 20 to prevent surface adhesion of the polypeptides. We carried out ensemble experiments on a spectrofluorometer (Spex Fluorolog 2) under identical buffer conditions at protein concentrations of 10 nM. Sequanal Grade GdmCl (Pierce) was used for denaturation experiments. Steady-state polarization measurements of the attached dyes resulted in anisotropies between 0.06 and 0.09 for all samples, indicating sufficient rotational averaging during the fluorescence lifetime of the dyes to justify using $\kappa^2 = 2/3$ (ref. 4). We calculated[1] a value of $R_0 = 5.4$ nm from the donor emission and acceptor absorption spectra of singly labelled Csp*Tm*, and a donor fluorescence quantum yield of 0.5, which was measured relative to the manufacturer's quantum yield for the free dye. The raw data shown in Fig. 2 were corrected for the refractive index change with increasing GdmCl concentration, according to the n^{-4} dependence of Förster theory[1], to determine the values in Fig. 3.

Confocal fluorescence microscope

Observations of single-molecule fluorescence were made using a custom confocal optical system with a 1.4 NA ×100 microscope objective (Nikon CFN Plan Apo 85025) and the 488-nm line of an argon ion laser (Lexel 95-5) as an excitation source. Sample fluorescence passed through a 100-μm diameter pinhole in the image plane of the objective. The fluorescence was separated into donor and acceptor components using a dichroic mirror (Omega 560DCLP), and two final filters (Omega 525AF45 for the donor and Omega 600ALP for the acceptor). Each component was focused onto a photon-counting avalanche photodiode (PerkinElmer Optoelectronics SPCM-AQR-15), and output pulses were collected in intervals of 1 ms. To minimize saturation effects owing to excitation of the donor dye before the acceptor dye had returned to the ground state, we reduced the laser intensity to an acceptable level on the basis of power dependence experiments.

Data reduction

Background (usually between 0.5 and 2 counts), obtained in each experiment from independent measurements of solutions without labelled samples, was subtracted from the counts in each 1-ms interval. The signal was considered significant if the sum of counts in the donor and acceptor channels was greater than 25, thereby ensuring that no more than one burst per thousand was due to background. We combined significant signals from adjacent intervals, because the probability that they arose from different molecules was negligible. Finally, the sums of donor counts (n_D) and acceptor counts (n_A) for each burst were used to compute an apparent FRET efficiency using the relation $E_{app} = n_A/(n_A + \gamma n_D)$, where $\gamma = (\phi_A \eta_A)/(\phi_D \eta_D)$ is a factor that corrects for the difference in the fluorescence quantum yields of the acceptor (ϕ_A) and the donor (ϕ_D) (in the absence of acceptor), as well as differences in the detection efficiencies of the donor and acceptor channels of the instrument, η_D and η_A, respectively. Comparison of the single-molecule instrument with a detector-calibrated fluorometer was made using solutions of the protein singly labelled with donor and acceptor dyes at concentrations that resulted in equal optical densities at 488 nm (100 nM and 2.2 μM, respectively). These experiments showed that at the laser intensities used in the single-molecule measurements γ is close to 1, the value we therefore used in our analysis. Slight deviations of γ from 1 were observed with increasing radiant flux.

Received 4 April; accepted 23 July 2002; doi:10.1038/nature01060.

1. Van Der Meer, B. W., Coker, G. III & Chen, S. S.-Y. *Resonance Energy Transfer. Theory and Data* (VHC, New York, 1994).
2. Jia, Y. W. *et al.* Folding dynamics of single GCN4 peptides by fluorescence energy transfer confocal microscopy. *Chem. Phys.* **247,** 69–83 (1999).
3. Talaga, D. S. *et al.* Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proc. Natl Acad. Sci. USA* **97,** 13021–13026 (2000).
4. Deniz, A. A. *et al.* Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA* **97,** 5179–5184 (2000).
5. Perl, D. *et al.* Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5,** 229–235 (1998).
6. Mandelkern, L. in *Poly-α-amino Acids. Protein Models for Conformational Studies* (ed. Fasman, G. D.) 675–724 (Marcel Dekker, New York, 1967).
7. Jacob, J., Baker, B., Bryant, R. G. & Cafiso, D. S. Distance estimates from paramagnetic enhancements of nuclear relaxation in linear and flexible model peptides. *Biophys. J.* **77,** 1086–1092 (1999).
8. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA* **58,** 719–726 (1967).
9. Wassenberg, D., Welker, C. & Jaenicke, R. Thermodynamics of the unfolding of the cold-shock protein from *Thermotoga maritima*. *J. Mol. Biol.* **289,** 187–193 (1999).
10. Fersht, A. *Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman, New York, 1999).
11. Alonso, D. O. V. & Dill, K. A. Solvent denaturation and stabilization of globular proteins. *Biochemistry* **30,** 5974–5985 (1991).
12. Chan, C. K. *et al.* Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl Acad. Sci. USA* **94,** 1779–1784 (1997).
13. Millett, I. S., Doniach, S. & Plaxco, K. W. Towards a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv. Prot. Chem.* (in the press).
14. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* 192 (Clarendon, Oxford, 1987).
15. Dahan, M. *et al.* Ratiometric measurement and identification of single diffusing molecules. *Chem. Phys.* **247,** 85–106 (1999).
16. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations. *Proc. Natl Acad. Sci. USA* **96,** 3670–3675 (1999).
17. Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl Acad. Sci. USA* **84,** 7524–7528 (1987).
18. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Genet.* **21,** 167–195 (1995).
19. Dobson, C. M., Sali, A. & Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Edit.* **37,** 868–893 (1998).
20. Socci, N. D., Onuchic, J. N. & Wolynes, P. G. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J. Chem. Phys.* **104,** 5860–5868 (1996).
21. Klimov, D. K. & Thirumalai, D. Viscosity dependence of the folding rates of proteins. *Phys. Rev. Lett.* **79,** 317–320 (1997).
22. Nymeyer, H., Socci, N. D. & Onuchic, J. N. Landscape approaches for determining the ensemble of folding transition states: Success and failure hinge on the degree of frustration. *Proc. Natl Acad. Sci. USA* **97,** 634–639 (2000).
23. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7,** 284–304 (1940).
24. Lapidus, L. J., Steinbach, P. J., Eaton, W. A., Szabo, A. & Hofrichter, J. Effects of chain stiffness on the dynamics of loop formation in polypeptides. *J. Phys. Chem. B* (in the press).
25. Jäger, M., Nguyen, H., Crane, J. C., Kelley, J. W. & Gruebele, M. The folding mechanism of a β-sheet: the WW domain. *J. Mol. Biol.* **311,** 373–393 (2001).
26. Hagen, S. J., Hofrichter, J. & Eaton, W. A. The rate of intrachain diffusion of unfolded cytochrome *c*. *J. Phys. Chem. B* **101,** 2352–2365 (1997).
27. Portman, J. J., Takada, S. & Wolynes, P. G. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J. Chem. Phys.* **114,** 5082–5096 (2001).
28. Muñoz, V. & Eaton, W. A. A simple model for calculating the kinetics of protein folding from three-dimensional structure. *Proc. Natl Acad. Sci. USA* **96,** 11311–11316 (1999).
29. Shea, J. E. & Brooks, C. L. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52,** 499–535 (2001).
30. Reid, K. L., Rodriguez, H. M., Hillier, B. J. & Gregoret, L. M. Stability and folding properties of a model β-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7,** 470–479 (1998).
31. Kremer, W. *et al.* Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur. J. Biochem.* **268,** 2527–2539 (2001).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com/nature).

Acknowledgements We thank A. Szabo and I. Gopich for suggestions and guidance on theoretical issues; S. Doniach, J. Hofrichter and G. Hummer for discussion and comments on the manuscript; L. Davis and R. Zare for advice regarding the single-molecule instrument; and L. Pannell for mass spectroscopy measurements. B.S. was supported by an Emmy Noether fellowship from the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.A.E. (e-mail: eaton@helix.nih.gov).