## Constructs expressing fluorescent fusion proteins

The *hABH2* and *hABH3* cDNAs were cloned, respectively, into the *Hin*dIII/*Acc*I and *Eco*RI/*Sma*I sites of pEYFP-N1 (Clontech) to produce hABH2–EYFP and hABH3–EYFP. The pGFP–PCNAL2 construct[18] was used to prepare the expression construct pECFP–PCNA by standard cloning techniques[21].

## Confocal microscopy

Transfection of HeLa cells by the $CaPO_4^-$ method was done as described (Profection, Promega). Cells were examined in a Zeiss LSM 510 laser scanning microscope using a 458-nm laser line for excitation of enhanced cyan fluorescent protein (ECFP; detected at 480 nm $< \lambda_{ECFP} <$ 520 nm) and a 514-nm laser line for enhanced yellow fluorescent protein (EYFP; detected at $\lambda_{EYFP} >$ 560 nm). Cells were sectioned at a thickness of 1 μm for the images.

1. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (ASM, Washington DC, 1995).
2. Falnes, P. Ø., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli. Nature* **419**, 178–182 (2002).
3. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
4. Singer, B. & Grunberger, D. *Molecular Biology of Mutagens and Carcinogens* (Plenum, New York, 1983).
5. Boiteux, S. & Laval, J. Mutagenesis by alkylating agents: coding properties for DNA polymerase of poly(dC) template containing 3-methylcytosine. *Biochimie* **64**, 637–641 (1982).
6. Saffhill, R. Differences in the promutagenic nature of 3-methylcytosine as revealed by DNA and RNA polymerising enzymes. *Carcinogenesis* **5**, 691–693 (1984).
7. Rozenski, J., Crain, P. F. & McCloskey, J. A. The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **27**, 196–197 (1999).
8. Agris, P. F. The importance of being modified: roles of modified nucleosides and $Mg^{2+}$ in RNA structure and function. *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 79–129 (1996).
9. Yoshizawa, S., Fourmy, D. & Puglisi, J. D. Recognition of the codon–anticodon helix by ribosomal RNA. *Science* **285**, 1722–1725 (1999).
10. Matsugi, J. & Murao, K. Study on construction of a cDNA library corresponding to an amino acid-specific tRNA and influence of the modified nucleotide upon nucleotide misincorporations in reverse transcription. *Biochim. Biophys. Acta* **1521**, 81–88 (2001).
11. Haug, T. *et al.* Human uracil-DNA glycosylase gene: sequence organization, methylation pattern, and mapping to chromosome 12q22–q24.1. *Genomics* **36**, 408–416 (1996).
12. Wei, Y. F., Carter, K. C., Wang, R. P. & Shell, B. K. Molecular cloning and functional analysis of a human cDNA encoding an *Escherichia coli* AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Res.* **24**, 931–937 (1996).
13. Aravind, L. & Koonin, E. V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, Research0007 [online] 〈http://genomebiology.com/2001/2/3/research/0007〉 (2001).
14. Myllyharju, J. & Kivirikko, K. I. Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.* **17**, 1173–1180 (1997).
15. Valegård, K. *et al.* Structure of a cephalosporin synthase. *Nature* **394**, 805–809 (1998).
16. Nash, T. The colorimetric estimation of formaldehyde by means of the Hantzsch reaction. *Biochem. J.* **55**, 416–421 (1953).
17. Kaasen, I., Evensen, G. & Seeberg, E. Amplified expression of the *tag*+ and *alkA*+ genes in *Escherichia coli*: identification of gene products and effects on alkylation resistance. *J. Bacteriol.* **168**, 642–647 (1986).
18. Leonhardt, H. *et al.* Dynamics of DNA replication factories in living cells. *J. Cell Biol.* **149**, 271–280 (2000).
19. Warner, J. R. Nascent ribosomes. *Cell* **107**, 133–136 (2001).
20. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning. A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
22. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
23. Bjelland, S., Bjørås, M. & Seeberg, E. Excision of 3-methylguanine from alkylated DNA by 3-methyladenine DNA glycosylase I of *Escherichia coli. Nucleic Acids Res.* **21**, 2045–2049 (1993).
24. Huang, X., Powell, J., Mooney, L. A., Li, C. & Frenkel, K. Importance of complete DNA digestion in minimizing variability of 8-oxo-dG analyses. *Free Radicals Biol. Med.* **31**, 1341–1351 (2001).
25. Crain, P. F. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.* **193**, 782–790 (1990).
26. Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Haugan, K. & Valla, S. Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the RK2 replicon. *Appl. Environ. Microbiol.* **63**, 370–379 (1997).
27. Dinglay, S., Trewick, S. C., Lindahl, T. & Sedgwick, B. Defective processing of methylated single-stranded DNA by *E. coli* AlkB mutants. *Genes Dev.* **14**, 2097–2105 (2000).

# The complete folding pathway of a protein from nanoseconds to microseconds

Ugo Mayor*, Nicholas R. Guydosh*, Christopher M. Johnson*, J. Günter Grossmann†, Satoshi Sato*, Gouri S. Jas‡§, Stefan M. V. Freund*, Darwin O. V. Alonso‖, Valerie Daggett‖ & Alan R. Fersht*

\* *MRC Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK*
† *CLRC Daresbury Laboratory, Daresbury, Warrington WA4 4AD, UK*
‡ *Laboratory of Chemical Physics, NIDDK, NIH, Bethesda, Maryland 20892, USA*
‖ *Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195-7610, USA*

Combining experimental and simulation data to describe all of the structures and the pathways involved in folding a protein is problematic. Transition states can be mapped experimentally by φ values[1,2], but the denatured state[3] is very difficult to analyse under conditions that favour folding. Also computer simulation at atomic resolution is currently limited to about a microsecond or less. Ultrafast-folding proteins fold and unfold on timescales accessible by both approaches[4,5], so here we study the folding pathway of the three-helix bundle protein Engrailed homeodomain[6]. Experimentally, the protein collapses in a microsecond to give an intermediate with much native α-helical secondary structure, which is the major component of the denatured state under conditions that favour folding. A mutant protein shows this state to be compact and contain dynamic, native-like helices with unstructured side chains. In the transition state between this and the native state, the structure of the helices is nearly fully formed and their docking is in progress, approximating to a classical diffusion–collision model. Molecular dynamics simulations give rate constants and structural details highly consistent with experiment, thereby completing the description of folding at atomic resolution.

Engrailed homeodomain (En-HD) from *Drosophila melanogaster* is a 61-residue α-helical protein[6]. Measurement of folding kinetics by temperature jump through the thermal denaturation transition region shows En-HD to be the fastest unfolding and one of the fastest folding proteins so far recorded directly[4], and its unfolding and folding are compatible with the present time regime of molecular dynamics simulation. Our previous simulations of the unfolding of En-HD predict the correct order of magnitude of the unfolding rate constant and also significant residual native α-helical structuring in the denatured state ensemble[4]. We have now been able to generate a denatured state, by means of mutation, that is stable under physiological conditions, to analyse it in depth and to show by NMR that the wild-type protein populates an equivalent denatured state. We have measured rate constants for folding and unfolding from some hundreds of nanoseconds to microseconds to show that the protein folds by an intermediate that is, in fact, the denatured state under physiological conditions, and we have simulated the unfolding of the protein by molecular dynamics to reconstruct the structures of the transition and denatured states. We now describe each state along the folding pathway, first by experiment and then by simulation.

We first generated a denatured state that was stable under physiological conditions. The native state for most proteins is favoured by only 5–15 kcal mol$^{-1}$. This balance has been exploited to analyse native and denatured states in equilibrium under

folding conditions by NMR[7]. In principle, however, the equilibrium position for any protein can be shifted by mutation so that the denatured state largely predominates under physiological conditions[8]. We lowered the stability of En-HD by mutating a leucine residue to alanine at position 16 (L16A), which deleted stabilizing tertiary interactions without introducing a side chain that was likely to make new specific interactions. Mutation of leucine to alanine tends to lower the stability of proteins by up to 4–5 kcal mol$^{-1}$ (ref. 9). Wild-type En-HD had a free energy of folding of only 2.5 kcal mol$^{-1}$. A change in free energy of 4–5 kcal mol$^{-1}$ therefore should, and did, radically tip the balance so that the protein was predominantly denatured under physiological conditions.

There were no detectable native tertiary interactions in L16A under physiological conditions; see for example, the aromatic region of the $^{1}$H-NMR spectra (Fig. 1). The backbone was very dynamic, but the protein had most of its native $\alpha$-helical structure, as monitored by circular dichroism (CD) and H$\alpha$ chemical shifts (Fig. 1). The denatured state revealed by L16A had the characteristics of a folding intermediate with much of the correct native

secondary structure. Its radius of gyration increased by 33% compared with the native state (17.0 and 12.8 Å, respectively), as monitored by solution X-ray scattering experiments. The CD and NMR spectra of L16A changed non-cooperatively with increasingly denaturing conditions towards those of a random coil (data not shown).

We then probed the denatured state of wild-type En-HD directly under physiological conditions using native state $^{1}$H/$^{2}$H-exchange experiments[10]. The free energy of what is effectively the complete unfolding of regions of the protein is measured by the rate of exchange of buried backbone amide groups with solvent. The free energy of denaturation of En-HD in $^{2}$H$_2$O measured by conventional procedures for thermal melting was 2.8 $\pm$ 0.2 kcal mol$^{-1}$ at 5 °C. However, the free energy of opening of most of the amide groups in the helices was significantly higher; on average the helices unfolded with a free energy of 3.4 $\pm$ 0.1 kcal mol$^{-1}$, with helix I being the most protected (Fig. 2). There was little protection against exchange for residues 50–55, residues that were highly flexible in L16A (Fig. 1). Thus, the completely open denatured state of En-HD was ~1 kcal mol$^{-1}$ less stable than the dominant denatured state, which had considerable native $\alpha$-helical structure and closely resembled the denatured state of the L16A mutant, namely a folding intermediate. Folding intermediates are often the predominant denatured state of a protein under physiological conditions[11].

Folding intermediates generated in thermal denaturation simulations, and then quenched to 25 °C, of both the wild-type and L16A proteins had extensive secondary structure and few tertiary contacts (Fig. 3). The overall helical content was 57% for the wild-type intermediate ensemble and 64% for the crystal structure. Breakdown of the helix content along the sequence was consistent with the hydrogen protection results: helix I was highly populated with a helix content of 76%, helix II was less stable at 38%, and helix III had a helix content of 51%. Interestingly, however, the intermediate contained two helices much of the time, as the loop between helices I and II adopted helical structure and served to unite the two segments (Fig. 3). Non-native helical structure in the loop between these helices is supported by the NMR chemical shift deviations for L16A (Fig. 1). The radius of gyration of this intermediate state was 30% greater than the native state (14.4 Å for the wild type, 14.2 Å for L16A and 11 Å for the native state), in agreement with experiments on L16A. Further unfolding at high temperature produced a very open unfolded state with little helical structure (21% on average) (Fig. 3). This highly unfolded form had a radius of gyration of 20.0 and 18.8 Å by simulation and experiment (as measured on the acid-denatured state), respectively. Similarly, experiments on En-HD at high temperature or in high concentrations of urea failed to detect regular secondary structure (Fig. 1).

We also detected a folding intermediate from ultra-fast kinetic
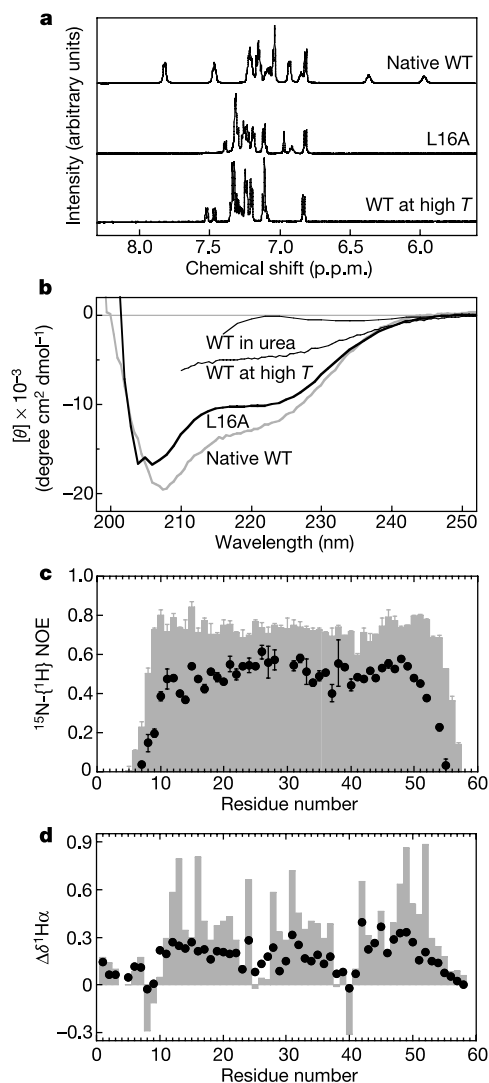
**Figure 1** Characterization of wild-type En-HD (WT) and the L16A mutant. **a**, Aromatic region of the $^{1}$H-NMR spectra of WT and L16A. Dispersion of the aromatic chemical shifts was lost in L16A (25 °C). **b**, Far-ultraviolet CD spectra of WT (25 and 95 °C) and L16A (25 °C). **c**, $^{15}$N-{$^{1}$H} heteronuclear NOE data ($I/I_0$) of WT (grey bars) and L16A (black circles) at 25 °C. The low NOE values for L16A indicate its higher flexibility. **d**, Deviations in chemical shift from reported random coil values ($\Delta\delta\,^{1}$H$\alpha$) in p.p.m. of WT (grey bars) and L16A (black circles) at 25 °C.
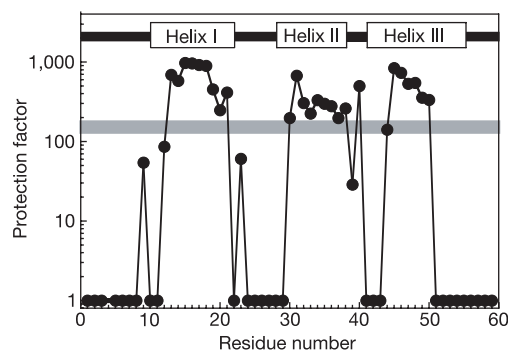
**Figure 2** Protection factors for the amide groups of wild-type En-HD, calculated from the $^{1}$H/$^{2}$H-exchange experiments at 5 °C. A value of 1 is given to those residues for which protection was not observed. The expected protection factor from a two-state fit of the denaturation curves is the thick grey line.
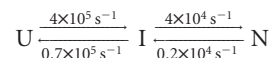
measurements. We showed previously that we could measure the rate constants for folding and unfolding by changing the temperature of a solution of En-HD rapidly (temperature jump) and monitoring the fluorescence of the single tryptophan residue[4]. We
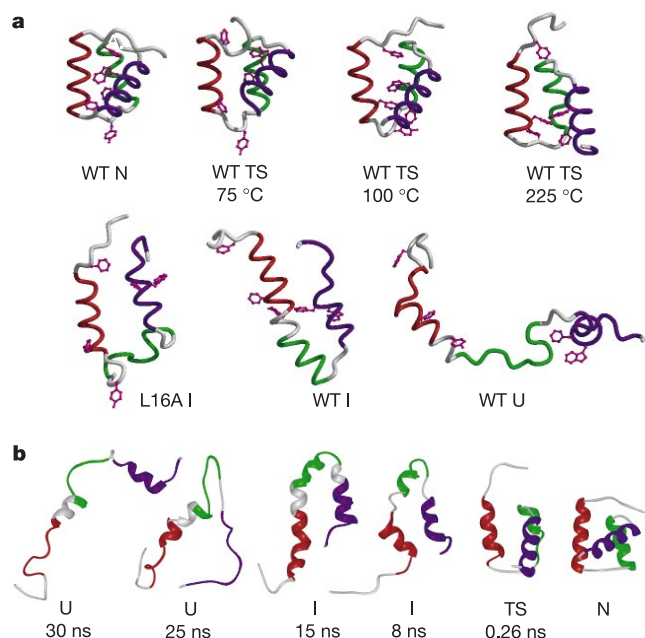


**Figure 3** Representative structures from the molecular dynamics simulations. **a**, Snapshots for the transition-state (TS) ensembles identified from wild-type (WT) simulations at different temperatures (61 ns at 75 °C, 1.985 ns at 100 °C and 0.26 ns at 225 °C), the intermediate (I) state (10-ns snapshots from the wild-type and L16A 225 °C trajectories quenched to 25 °C) and the unfolded (U) state (represented by the 40-ns WT 225 °C structure). The native helical segments are coloured as follows: red, residues 10–22; green, residues 28–38; blue, residues 42–55. Aromatic side chains are shown in magenta. **b**, Structures from the wild-type 225 °C denaturation simulation shown in reverse, to illustrate a probable folding pathway of the protein to reach the native (N) state.

have now increased the resolution of our experiments with the use of laser $T$-jump[12] and improvements to the classical electrical discharge apparatus[4]. In so doing, we detected two-step folding kinetics (Fig. 4). At 25 °C, there was a fast phase of $t_{1/2} \approx 1.5\,\mu$s, in addition to the previously reported phase of 15 μs (ref. 4). We were able to assign these phases by two means. First, $T$-jump kinetics on L16A unfolding and refolding exhibited just the fast relaxation time (Fig. 4), indicating that it corresponded to the transition between its equilibrium denatured state (folding intermediate) and a more unfolded denatured state. Indeed, with increasing temperature the tryptophan fluorescence in both the mutant and the native proteins approached that of an exposed tryptophan residue both in equilibrium and $T$-jump experiments. Second, we measured the rate of formation of the wild-type native structure by the change in line shape[13] of the leucine-16 side-chain resonances between 43 and 56 °C. The rate constants corresponded to those of the slower phase in the $T$-jump kinetics (Fig. 4).

If the fast phase corresponds to the full unfolding of the protein detected by the solvent exchange experiments, then combining the values of the measured equilibrium constants for the overall unfolding and the exchange experiments with the standard formulae for relaxation times gives a minimal kinetic scheme at 25 °C of

$$ U \underset{0.7\times10^5\,s^{-1}}{\overset{4\times10^5\,s^{-1}}{\rightleftarrows}} I \underset{0.2\times10^4\,s^{-1}}{\overset{4\times10^4\,s^{-1}}{\rightleftarrows}} N $$

where U is the unfolded denatured state, I is the folding intermediate with considerable α-helical secondary structure (or D, the denatured state under physiological conditions) and N is the native state. The results of the various kinetic experiments are compiled in Fig. 4. The unfolding experiments were conducted at up to 65 °C and extrapolated to 100 °C. The times to reach the transition state in the unfolding simulations at 75 and 100 °C fit well with experiment: at 75 °C, the simulated time was approximately 60 ns and the extrapolated $t_{1/2}$ was 330 ns; at 100 °C the times were 2 and 5 ns, respectively (Fig. 4).

We analysed the structure of the transition state between N and I by means of $\phi$ values (a measure of the degree of native structure in the transition state; 0 implies denatured, 1 implies fully native)
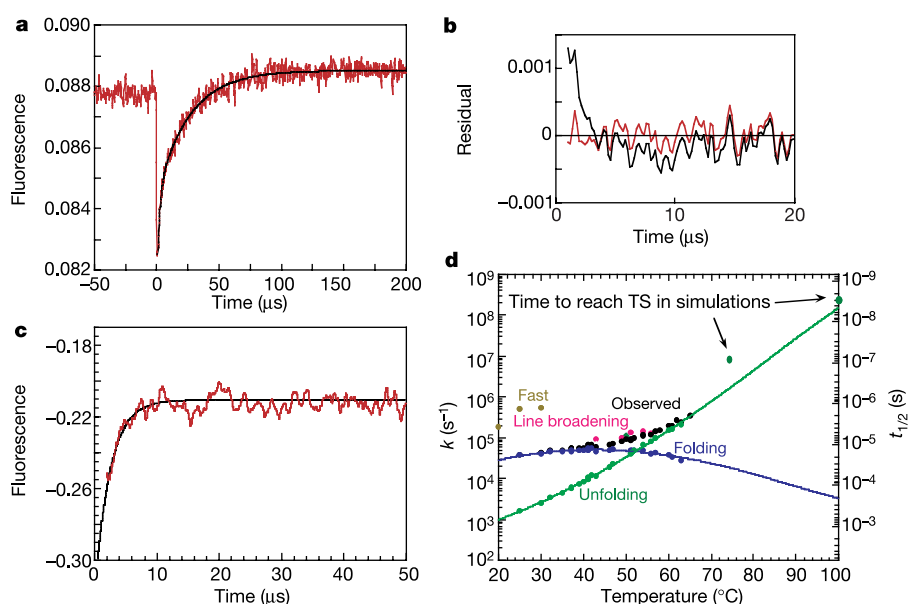


**Figure 4** Kinetics of folding and unfolding. **a**, Laser-heating $T$-jump of En-HD to 25 °C; data after 800 ns fitted to a double exponential. Details of output and errors are available from the authors. **b**, Residuals from **a** (red line), randomly distributed around zero, unlike those of a single-exponential fit (black line). **c**, The L16A mutant of En-HD, $T$-jumped to 25 °C (Joule heating), showed only the fast phase. En-HD that had been $T$-jumped with

the Joule-heating apparatus gave very similar rate constants to those from the laser apparatus. Details of output and errors are available from the authors. **d**, Observed relaxation rate constants ($k$) versus $T$. Solid curves were derived from combining $k$ with equilibrium data from calorimetry. The filled red circles are results from NMR line-broadening analysis. TS, transition state.

**865**
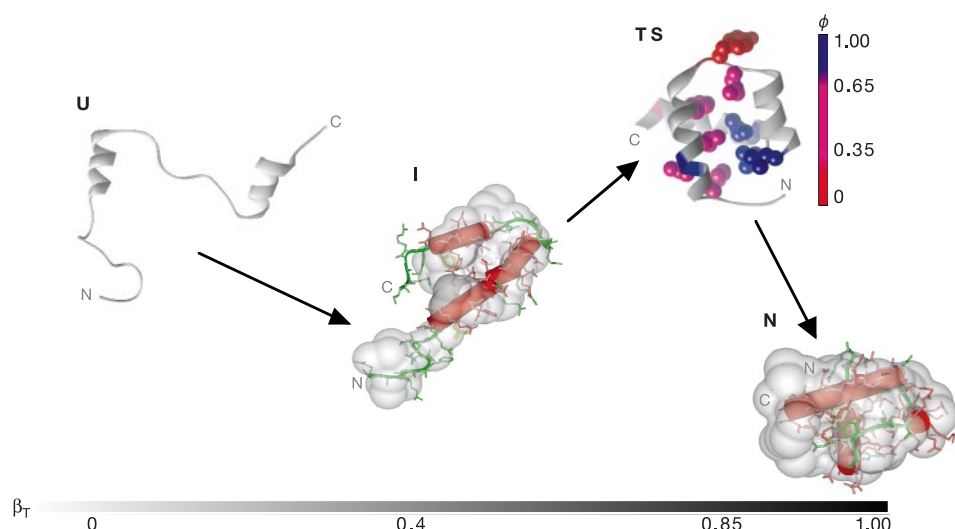
**Figure 5** The complete folding pathway of En-HD by experiment and simulation. U is the quenched 40-ns snapshot from the simulation at 225 °C; I is the denatured state under 'physiological conditions'. The 10-ns structure from the high-temperature molecular dynamics simulation is enclosed in the grey molecular envelope obtained by X-ray scattering. TS is the transition state between I and N, the native state. $\phi$ values are colour-coded, with residues used as probes of secondary structure coloured directly on the backbone. The simulated structure is from the transition state ensemble for the wild-type at 100 °C (1.985 ns). Structures are positioned according to their relative solvent accessibilities ($\beta_T$), as obtained from the dependence of the rate constants on urea concentration. Molecular dynamics simulation gave consistent results.

calculated from the unfolding kinetics[14] for 16 mutations. The transition-state ensembles were also identified from the unfolding simulations at different temperatures, ranging from 75 to 225 °C, and they were very similar (Fig. 3). The experimentally and theoretically derived $\phi$ values are in good agreement: for example, the correlation coefficient between the values calculated for the 100 °C transition-state ensembles and experiment was 0.86. The combined picture is that the main transition state for folding and unfolding was very native-like, as reflected by both the $\phi$ values and the molecular-dynamics-generated structures (Fig. 5), as expected for a transition from a structured intermediate. Helix I, helix III and the turn between helices II and III were largely formed (Fig. 3) (for example, $\phi$ values: A14G in H1 = 0.79; G39A in the turn, 0.92; A43G in H3, 1.05)[14]. The loop between helices I and II was less structured (for example, $\phi$ for A25G was 0.17). The hydrophobic core had fractional $\phi$ values indicative of partial structure (typically more than 0.4). This structural pattern is just what is expected for a framework[15] or diffusion–collision mechanism[16], in which native-like secondary structure forms and then docks in the rate-determining step (Fig. 3). The use of coarse-grained models has shown that En-HD can fold by diffusion–collision[17]. The results here show that a diffusion–collision mechanism is approximated at atomic resolution, both by experiment and simulation by comparing the intermediate and transition states.

The helical contents of isolated synthetic peptides corresponding to helices I, II and III were roughly 30%, 15% and 25%, respectively, as determined both by trifluoroethanol titration experiments[18] and directly from the CD ellipticity at 222 and 208 nm (ref. 19). These abnormally high values, which are probably increased by tertiary contacts in the intact protein, as indicated by our simulations, precipitate a tendency towards the diffusion–collision model. We have argued elsewhere[20] that there is a continuum of mechanisms between the classical stepwise diffusion–collision mechanism and nucleation–condensation, in which there is concurrent consolidation of secondary and tertiary structure. It is probably correct to describe En-HD as being close to the diffusion–collision limit, because the pure mechanism of totally independent formation of secondary and tertiary interactions is most unlikely.

En-HD is an informative protein for studying unfolding and folding. Fast events can be observed directly at the upper limits predicted for folding[21,22] that enter the time range currently accessible by molecular dynamics simulation. The agreement between molecular dynamics simulation and experiment is very encouraging and on the same timescale. Combining molecular dynamics and experiment has now allowed us to characterize all of the necessary structures along the folding pathway. One of the longest-running controversies in mechanistic studies is whether intermediates are on the pathway or are unwanted side products. Here, simulation shows that the intermediate is on the pathway and that the protein does not fold by an intermediate that must unfold to be productive. Simulation could well be the answer in general for solving such problems in mechanism. □

## Methods

### Materials
[15]N-labelled and [13]C–[15]N-labelled wild-type En-HD were prepared[4] to >99% homogeneity as determined by both mass spectrometry and SDS chromatography, as was the L16A mutant. CD spectra of 30-μM samples in 50 mM sodium acetate, pH 5.7, 100 mM NaCl were recorded in an AVIV 202SF instrument. NMR experiments on 0.5–2-mM samples were performed with Bruker DRX 500 and DRX 600 spectrometers equipped with inverse triple-resonance probes and single-axis gradients.

Peptides of sequence AcNH-SSEQLARLKREFNENR-NH$_2$, AcNH-YLTERRRQQ LSSELGL-NH$_2$ and AcNH-NEAQIKIWFQNKRAKI-NH$_2$, corresponding to helices I, II and III of En-HD, were synthesized chemically by G. Bloomberg (University of Bristol). Peptide purity and concentration were determined by mass spectroscopy and amino acid analysis, respectively.

### NMR Methods
NMR resonances of both wild-type and L16A mutant En-HD were assigned by using standard triple-resonance techniques[23]. ^{1}H-NMR spectra in [2]H$_2$O, 50 mM sodium acetate-d$_3$ and 100 mM NaCl were recorded from 1 to 90 °C, with weak presaturation of the residual solvent peak. The spectra were internally referenced to 3-(trimethylsilyl)propionate at 0 p.p.m. Series of ^{1}H,[15]N heteronuclear single-quantum-coherence spectra for solvent exchange experiments into [2]H$_2$O were acquired with 1024 × 90 complex points, four scans per increment and a dead time of 4 min. Solvent exchange data were analysed with standard equations for EX2 conditions[24]. The apparent equilibrium constant for the formation of the open state was calculated as $K_{op} = k_{ex}/k_{int}$, where $k_{ex}$ is the observed exchange rate and $k_{int}$ is the intrinsic exchange rate calculated from model peptide studies[25], with the SPHERE online software[26]. Apparent free energies of opening were calculated from this equilibrium constant. Enhancements of the [15]N-{[1]H} nuclear Overhauser effect (NOE) were determined from spectra obtained with and without presaturation of amide proton resonances. Presaturation was achieved by applying 120° decoupling pulses in 5-ms intervals with a field strength of 25 kHz for 3 s.

Reference spectra were obtained by replacing the presaturation sequence with a simple delay of the same length. For both experiments, the overall recycle delay was set to 5 s. Two sets of interleaved NOE and control experiments were added to obtain the intensities used in the calculation.

The $^1$H-NMR lineshapes of the L16 δ protons from 43 to 56 °C were fitted to a two-site exchange model[13,27]. Relative populations of the native and the denatured protein were determined independently from differential scanning calorimetry measurements; other parameters were obtained from the spectra. The independent variables were a normalization factor, the exchange rate and the frequency for L16 in the denatured state.

## Solution X-ray scattering

Solution X-ray scattering data for wild-type En-HD and the L16A mutant were collected at station 2.1 of the Daresbury Synchrotron Radiation Source, UK, and processed with procedures described previously[28,29]. Several shape restorations for both protein states have been undertaken to evaluate the similarity and reliability of the low-resolution models. Indeed, the calculations resulted in similar shapes that were consistent with those presented here as semitransparent surfaces.

## Kinetics methods

A DIA-RT T-Jump (DiaLog) electrical-discharge temperature-jump instrument[12] and a laser $T$-jump instrument[12] were used. Protein concentrations were in the range 15–300 μM, buffered with 50 mM HEPES buffer, pH 8.0, 100 mM NaCl. The rate constants were independent of concentration.

## Computer simulations

Previous molecular dynamics simulations of the wild-type protein were at 100 and 225 °C for 76 and 42 ns, respectively[4]. New simulations presented here were of the wild type (15 ns at 25 °C, 100 ns at 75 °C, 24 ns at 100 °C, and 21 ns at 225 °C) and the L16A mutant (42 ns at 25 °C, 10 ns at 100 °C, and two independent simulations of 10 and 22 ns at 225 °C). There were further simulations in which structures from these high-temperature simulations were quenched to 25 °C and the water density was adjusted correspondingly (from 0.829 g ml$^{-1}$ at 225 °C to 0.997 g ml$^{-1}$). We performed the following quenches: the 10-ns snapshot from the wild-type simulation at 225 °C was simulated at 25 °C for 40 ns to represent the folding intermediate; the 40-ns snapshot from the wild-type simulation at 225 °C was run for 10 ns to represent the unfolded state; and the 10-ns snapshot from the L16A simulation at 225 °C was simulated for 10 ns to model the equilibrium intermediate state. Many of these simulations were performed as controls and to ensure the reproducibility of the results. The various En-HD simulations amounted to 0.422 μs.

1. Fersht, A. R., Leatherbarrow, R. J. & Wells, T. N. Structure–activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. *Biochemistry* **26**, 6030–6038 (1987).
2. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224**, 771–782 (1992).
3. Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27–34 (1996).
4. Mayor, U., Johnson, C. M., Daggett, V. & Fersht, A. R. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl Acad. Sci. USA* **97**, 13518–13522 (2000).
5. Ferguson, N. et al. Using flexible loop mimetics to extend phi-value analysis to secondary structure interactions. *Proc. Natl Acad. Sci. USA* **98**, 13008–13013 (2001).
6. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L. & Pabo, C. O. Structural studies of the engrailed homeodomain. *Prot. Sci.* **3**, 1779–1787 (1994).
7. Blanco, F. J., Serrano, L. & Forman-Kay, J. D. High populations of non-native structures in the denatured state are compatible with the formation of the native folded state. *J. Mol. Biol.* **284**, 1153–1164 (1998).
8. Spector, S., Rosconi, M. & Raleigh, D. P. Conformational analysis of peptide fragments derived from the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: evidence for nonrandom structure in the unfolded state. *Biopolymers* **49**, 29–40 (1999).
9. Serrano, L. & Fersht, A. R. Principles of protein stability derived from protein engineering methods. *Curr. Opin. Struct. Biol.* **3**, 75–83 (1993).
10. Englander, S. W. & Krishna, M. M. G. Hydrogen exchange. *Nature Struct. Biol.* **8**, 741–742 (2001).
11. Oliveberg, M. & Fersht, A. R. Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry* **35**, 2738–2749 (1996).
12. Thompson, P. A., Eaton, W. A. & Hofrichter, J. Laser temperature jump study of the helix ⇔ coil kinetics of an alanine peptide interpreted with a 'kinetic zipper' model. *Biochemistry* **36**, 9200–9210 (1997).
13. Huang, G. S. & Oas, T. G. Submillisecond folding of monomeric lambda repressor. *Proc. Natl Acad. Sci. USA* **92**, 6878–6882 (1995).
14. Matouschek, A., Kellis, J. T. Jr, Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126 (1989).
15. Baldwin, R. L. & Rose, G. D. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83 (1999).
16. Karplus, M. & Weaver, D. L. Protein-folding dynamics—the diffusion–collision model and experimental-data. *Prot. Sci.* **3**, 650–668 (1994).
17. Islam, S. A., Karplus, M. & Weaver, D. L. Application of the diffusion–collision model to the folding of three-helix bundle proteins. *J. Mol. Biol.* **318**, 199–215 (2002).
18. Kippen, A. D., Arcus, V. L. & Fersht, A. R. Structural studies on peptides corresponding to mutants of the major alpha-helix of barnase. *Biochemistry* **33**, 10013–10021 (1994).
19. Greenfield, N. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).
20. Daggett, V. & Fersht, A. R. Is there a unifying mechanism of protein folding? *Trends Biochem. Sci.* **28**, 19–26 (2003).
21. Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA* **93**, 11615–11617 (1996).
22. Bieri, O. et al. The speed limit for protein folding measured by triplet–triplet energy transfer. *Proc. Natl Acad. Sci. USA* **96**, 9597–9601 (1999).
23. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego, 1996).
24. Itzhaki, L. S., Neira, J. L. & Fersht, A. R. Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J. Mol. Biol.* **270**, 89–98 (1997).
25. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
26. Zhang, Y.-Z. *Structural Biology and Molecular Biophysics*. PhD thesis, Univ. Pennsylvania (1995).
27. Sandstrom, J. *Dynamic NMR Spectroscopy* (Academic, New York, 1982).
28. Grossmann, J. G., Hall, J. F., Kanbi, L. D. & Hasnain, S. S. The N-terminal extension of rusticyanin is not responsible for its acid stability. *Biochemistry* **41**, 3613–3619 (2002).
29. Witty, M. et al. Structure of the periplasmic domain of *Pseudomonas aeruginosa* TolA: evidence for an evolutionary relationship with the TonB transporter protein. *EMBO J.* **21**, 4207–4218 (2002).