M. Saraste. The SH3 protein was expressed in *Escherichia coli* BL21 (DE3), using M9 minimal medium. A total of 0.5 g $^{15}NH_4Cl$, 2.0 g $NaH^{12}CO_3$ and 2.0 g [1,3-^{13}C]glycerol per litre of medium was added in the case of ([1,3-^{13}C]glycerol, ^{15}N) SH3. For the preparation of the ([2-^{13}C]glycerol, ^{15}N) SH3 sample, 0.5 g $^{15}NH_4Cl$, 2.0 g $NaH^{13}CO_3$ and 2.0 g [2-^{13}C]glycerol were used[21]. The proteins were purified as previously described[20]. The final yield was approximately 20 mg of protein per litre of culture. A 200 mM $(NH_4)_2SO_4$ solution (pH 3.5, 0.04% NaN_3) was added to a 1.15 mM SH3 solution (pH 3.5) at a volume ratio of 1:1. The protein samples for MAS measurement were precipitated as previously described[20].

NMR spectroscopy

All solid-state carbon–carbon NMR experiments were performed at a field of 17.6 T on a DMX-750 narrow-bore spectrometer, equipped with a 4-mm double-resonance MAS probe (Bruker, Karlsruhe). The protein was confined to the centre of the rotor with the use of spacers. Approximately 6–7 mg of protein was used in all experiments, except for the fivefold diluted samples, where 9–10 mg of protein was used. All two-dimensional correlation spectra were acquired at MAS frequencies of 8.0 kHz or 13.0 kHz using time-proportional phase incrementation for phase-sensitive detection. Ramped cross-polarization from ^{1}H to ^{13}C created the initial transverse carbon magnetization; spin-lock fields were 36 kHz for ^{1}H and 18–36 kHz for the ^{13}C ramp. After the first ^{13}C evolution period, carbon magnetization was exchanged by using a PDSD mixing scheme[22]. Spin diffusion periods of 50–500 ms were applied. Typical carbon 90° pulse lengths were 5.3 μs. A proton RF field of ~60 kHz was applied for the two-pulse phase modulation decoupling[28] during ^{13}C acquisition and evolution. The two-dimensional ^{13}C–^{13}C spectra were recorded with 32–64 scans, and with ~6 ms evolution in the indirect dimension, leading to experimental times of 16–32 h. The spectra of the 80% unlabelled and 20% labelled protein were recorded with 128 scans in 2.5 days. The ^{15}N–^{15}N PDSD spectrum was recorded under similar experimental conditions, but with a mixing time of 4 s and on a DMX-750 wide-bore spectrometer, equipped with a 4-mm triple-resonance MAS probe (Bruker, Karlsruhe).

The data were processed with the XWINNMR software version 2.6 (Bruker, Karlsruhe) and subsequently analysed using the program Sparky version 3.100 (T. D. Goddard & D. G. Kneller, University of California).

Structure calculations

Structures were calculated with the program CNS version 1.0 (ref. 25). Calculations were performed using the simulating annealing protocol with torsion-angle dynamics, starting with 200 randomized conformers. The 286 ^{13}C–^{13}C restraints were categorized into strong (2.5–4.5 Å), medium (2.5–5.5 Å), weak (2.5–6.5 Å) or very weak (2.5–7.5 Å) classes. The six ^{15}N–^{15}N correlations were constrained to 3–6 Å. Twelve of the fifteen lowest-energy structures with no violations larger than 0.3 Å were used to represent the three-dimensional fold of the α-spectrin SH3 domain. The atomic coordinates and NMR restraints were deposited in the RCSB Protein DataBank under entry number 1M8M.

Received 30 January; accepted 6 August 2002; doi:10.1038/nature01070.

1. Tycko, R. Biomolecular solid state NMR: Advances in structural methodology and applications to peptide and protein fibrils. *Annu. Rev. Phys. Chem.* **52,** 575–606 (2001).
2. Opella, S. J. NMR and membrane proteins. *Nature Struct. Biol.* **4** suppl., 845–848 (1997).
3. Griffin, R. G. Dipolar recoupling in MAS spectra of biological solids. *Nature Struct. Biol.* **5** NMR suppl., 508–512 (1998).
4. De Groot, H. J. M. Solid-state NMR spectroscopy applied to membrane proteins. *Curr. Opin. Struct Biol.* **10,** 593–600 (2000).
5. Pauli, J., Baldus, M., van Rossum, B., de Groot, H. & Oschkinat, H. Backbone and side-chain ^{13}C and ^{15}N signal assignments of the α-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. *ChemBioChem* **2,** 272–281 (2001).
6. Straus, S. K., Bremi, T. & Ernst, R. R. Experiments and strategies for the assignment of fully ^{13}C/^{15}N-labelled polypeptides by solid-state NMR. *J. Biomol. NMR* **12,** 39–50 (1998).
7. Hong, M. Resonance assignment of ^{13}C/^{15}N labelled solid proteins by two-and three-dimensional magic-angle-spinning NMR. *J. Biomol. NMR.* **15,** 1–14 (1999).
8. Hong, M. Determination of multiple φ-torsion angles in proteins by selective and extensive ^{13}C labeling and two-dimensional solid-state NMR. *J. Magn. Reson.* **139,** 389–401 (1999).
9. Jaroniec, C. P., Tounge, B. A., Herzfeld, J. & Griffin, R. G. Frequency selective heteronuclear dipolar recoupling in rotating solids: accurate ^{13}C-^{15}N distance measurements in uniformly ^{13}C,^{15}N-labeled peptides. *J. Am. Chem. Soc.* **123,** 3507–3519 (2001).
10. Brown, S. P. & Spiess, H. W. Advanced solid-state NMR methods for the elucidation of structure and dynamics of molecular, macromolecular, and supramolecular systems. *Chem. Rev.* **101,** 4125–4156 (2001).
11. Ketchem, R. R., Lee, K.-C., Huo, S. & Cross, T. A. Macromolecular structural elucidation with solid-state NMR-derived orientational constraints. *J. Biomol. NMR* **8,** 1–14 (1996).
12. Shochat, S. *et al.* ^{13}C MAS NMR evidence for a homogeneously ordered environment of tyrosine M210 in reaction centres of *Rhodobacter sphaeroides. Spectrochim. Acta* **51A,** 135–144 (1995).
13. Verhoeven, M. A. *et al.* Ultra-high-field MAS NMR assay of a multispin labeled ligand bound to its G-protein receptor target in the natural membrane environment: electronic structure of the retilylidene chromophore in rhodopsin. *Biochemistry* **40,** 3282–3288 (2001).
14. Egorova-Zachernyuk, T. A. *et al.* Characterization of pheophytin ground states in *Rhodobacter sphaeroides* R26 photosynthetic reaction centers from multispin pheophytin enrichment and 2-D ^{13}C MAS NMR dipolar correlation spectroscopy. *Biochemistry* **36,** 7513–7519 (1997).
15. Griffiths, J. M. *et al.* Dipolar correlation NMR spectroscopy of a membrane protein. *J. Am. Chem. Soc.* **116,** 10178–10181 (1994).
16. Hodgkinson, P. & Emsley, L. The accuracy of distance measurements in solid-state NMR. *J. Magn. Reson.* **139,** 46–59 (1999).
17. Kiihne, S. *et al.* Distance measurements by dipolar recoupling two-dimensional solid-state NMR. *J. Phys. Chem. A* **102,** 2274–2282 (1998).
18. Nielsen, N. C., Bildsoe, H., Jakobsen, H. J. & Levitt, M. H. Double-quantum homonuclear rotary resonance: efficient dipolar recovery in magic-angle spinning nuclear magnetic resonance. *J. Chem. Phys.* **101,** 1805–1812 (1994).
19. Costa, P. R., Sun, B. Q. & Griffin, R. G. Rotational resonance tickling: accurate internuclear distance measurement in solids. *J. Am. Chem. Soc.* **119,** 10821–10830 (1997).
20. Pauli, J., van Rossum, B., Förster, H., De Groot, H. J. M. & Oschkinat, H. Sample optimization and identification of signal patterns of amino acid side chains in 2D RFDR spectra of the α-spectrin SH3 domain. *J. Magn. Reson.* **143,** 411–416 (2000).
21. LeMaster, D. M. & Kushlan, D. M. Dynamical mapping of *E. coli* thioredoxin via ^{13}C NMR relaxation analysis. *J. Am. Chem. Soc.* **118,** 9255–9264 (1996).
22. Szeverenyi, N. M., Sullivan, M. J. & Maciel, G. E. ^{13}C Spin exchange by 2D FT ^{13}C CP/MAS. *J. Magn. Reson.* **47,** 462–475 (1982).
23. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
24. van Rossum, B. J., Castellani, F., Rehbein, K., Pauli, J. & Oschkinat, H. Assignment of the non exchanging protons of the α-spectrin SH3 domain by two- and three-dimensional ^{1}H-^{13}C solid-state magic-angle-spinning NMR and comparison of solution and solid-state proton chemical shifts. *ChemBioChem* **2,** 906–914 (2001).
25. Brünger, A. T. *et al.* Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
26. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359,** 851–855 (1992).
27. Reif, B., Jaroniec, C. P., Rienstra, C. M., Hohwy, M. & Griffin, R. G. ^{1}H-^{1}H MAS correlation spectroscopy and distance measurements in a deuterated peptide. *J. Magn. Reson.* **151,** 320–327 (2001).
28. Bennett, A. E., Rienstra, C. M., Auger, M., Lakshmi, K. V. & Griffin, R. G. Heteronuclear decoupling in rotating solids. *J. Chem. Phys.* **103,** 6951–6958 (1995).

Acknowledgements H. de Groot is acknowledged for access to the high-field NMR facility in Leiden. The authors thank R. Kühne, P. Schmieder, G. Krause and C. Glaubitz for discussions, and L. Ball, K. Heuer and K. Zierler for carefully reading the manuscript.

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

Correspondence and requests for materials should be addressed to H.O. (e-mail: oschkinat@fmp-berlin.de) or B.v.R. (brossum@fmp-berlin.de).

Absolute comparison of simulated and experimental protein-folding dynamics

Christopher D. Snow*†, Houbi Nguyen†‡, Vijay S. Pande* & Martin Gruebele‡

** Biophysics Program and Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA*
‡ Departments of Chemistry and Physics, and Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, USA
† These authors contributed equally to this work

Protein folding is difficult to simulate with classical molecular dynamics. Secondary structure motifs such as α-helices and β-hairpins can form in 0.1–10 μs (ref. 1), whereas small proteins have been shown to fold completely in tens of microseconds[2]. The longest folding simulation to date is a single 1-μs simulation of the villin headpiece[3]; however, such single runs may miss many features of the folding process as it is a heterogeneous reaction involving an ensemble of transition states[4,5]. Here, we have used a distributed computing implementation to produce tens of thousands of 5–20-ns trajectories (700 μs) to simulate mutants of the designed mini-protein BBA5. The fast relaxation dynamics these predict were compared with the results of laser temperature-jump experiments. Our computational predictions are in excellent agreement with the experimentally determined mean folding times and equilibrium constants. The rapid folding of BBA5 is due to the swift formation of secondary structure. The convergence of experimentally and computationally accessible timescales will allow the comparison of absolute quantities characterizing *in vitro* and *in silico* (computed) protein folding[6].

To establish a statistically favoured mechanism, map the free energy surface, and to compare absolute rate and equilibrium constants with experimental data requires multiple simulations of any folding reaction. Such multiple sampling has only been achieved for peptide systems[7–10]. As an alternative, simulations in the past have been compared to relative quantities such as Phi-values[11], or high temperature unfolding rates[2]. The prevailing view is that current molecular dynamics cannot find the native state at the free energy minimum owing to limitations of timescale and force field accuracy. For example, performing a molecular dynamics simulation for 10 μs with a 2-fs time step on a simple system in implicit solvent, such as our model protein BBA5, would require decades for a typical modern CPU. Our approach is to compare directly experimental rate and equilibrium constants with distributed computing simulations of the same quantities from thousands of short (5–20 ns) folding trajectories totalling hundreds of microseconds. Absolute quantities such as rate constants provide strong constraints on barriers and stringent folding model tests, but are more difficult to calculate precisely than relative quantities. For instance, agreement between simulation and experiment within a factor of two would be noteworthy for even small-molecule rate prediction. To allow experiment and theory to meet on the microsecond timescale, we have chosen to study mutants of the fast-folding, designed mini-protein BBA5. The choice of a 23-residue protein with strong secondary structure propensities and a small hydrophobic core reduces the impact of force field inaccuracies, and allows kinetics to be resolved both experimentally and computationally.

Although folding during a 20-ns simulation is improbable, collecting a very large number of trajectories has allowed us to observe such improbable events (Fig. 1). Given a small, two-state protein that obeys exponential kinetics with a mean folding time of 10 μs, roughly 10 out of 10,000 individual molecules should have folded conformations after 10 ns (ref. 10). Individual trajectories may not be arbitrarily short. Each simulation must first relax from the initial condition to the unfolded state under native conditions. For BBA5 this relaxation is extremely rapid—the initial extended ensemble collapses to a compact unfolded ensemble within 10 ns. The unfolded ensemble under native conditions has a significant degree of α-helical structure (in both simulation and experiment). Our simulations also satisfy the constraint that each simulation should not be substantially shorter than the minimal passage time through the transition state (about 10 ns, from peptide diffusion data[12,13]). The Folding@Home distributed computing project[14], a cluster of over 30,000 volunteer computers distributed around the world, has been used for several months to accumulate approximately a million CPU days of simulation time (http://folding.stanford.edu/). We observed folding in over 100 independent trajectories. Calculations and laser temperature-jump experiments indicate that our BBA5 mutants are moderately cooperative folders, have a very low melting temperature, and fold on a timescale of several microseconds. Furthermore, the simulations characterize the timescale for the formation of secondary structure, and demonstrate folding by means of a heterogeneous transition state (Supplementary Information).

Two mutants of BBA5 (Fig. 1a), a 23-residue β-hairpin/turn/α-helix motif inspired by the zinc finger fold, were constructed on the basis of the sequence developed by the Imperiali group, Ac-YRVpSYDFSRSDELAKLLRQHAG-NH_2 (ref. 15). BBA5 is a 'mini-protein' containing all three elementary units of secondary structure, and folds without obligate metal-ion binding[16]. A D-proline residue favours the formation of a type II′ β-turn, stabilizing a hairpin centred about residues 4–5. The amino-terminal β-hairpin and carboxy-terminal α-helix interact to form a small hydrophobic core. Native state simulations at 278 K reveal flexibility in the relative orientation of the hairpin and the helix. Our single mutant replaces a phenylalanine at position 8 with a tryptophan, which acts as a fluorescent probe and as a larger aromatic residue for interaction with side chains in the helix. The double mutant additionally replaces valine 3 with a tyrosine. Both mutants show the same basic fold in simulations, and have proton NMR spectra with dispersion similar to BBA5. BBA5 mutants were chemically synthesized, purified and buffered in 10 mM phosphate at pH 7 for all experiments. To study the folded mutants, homology models were constructed from the structure of BBA5 (see Methods).

Circular dichroism and fluorescence spectra of BBA5 mutants were measured to quantify the folding equilibrium (Fig. 2). The folding transition was weakly cooperative, as expected for a mini-protein (midpoint below 0 °C). To assess secondary structure changes during folding, far-ultraviolet circular dichroism spectra were collected from −15 °C to 90 °C. Both mutants fold with increased helical content at low temperature (local minimum at 222 nm). Their circular dichroism spectra closely resemble BBA5 (ref. 15). The N-terminal 1–10 hairpin fragment by itself has a random coil spectrum without cooperative transition (Fig. 2a). The C-terminal 11–23 helix fragment has higher secondary structure propensity. To assess solvent exposure of the core, tryptophan fluorescence as a function of wavelength was measured at temperatures from −2.5 °C to 60 °C. On unfolding at higher temperature, the fluorescence intensity of the double mutant displayed a redshift of 11 nm, indicating increased tryptophan solvent exposure. The fluorescence and circular dichroism spectra have identical melting transitions, as well as isosbestic/isodichroic points where all spectra intersect. This indicates that the BBA5 mutants fold thermodynamically by means of a two-state mechanism. The calculated equili-

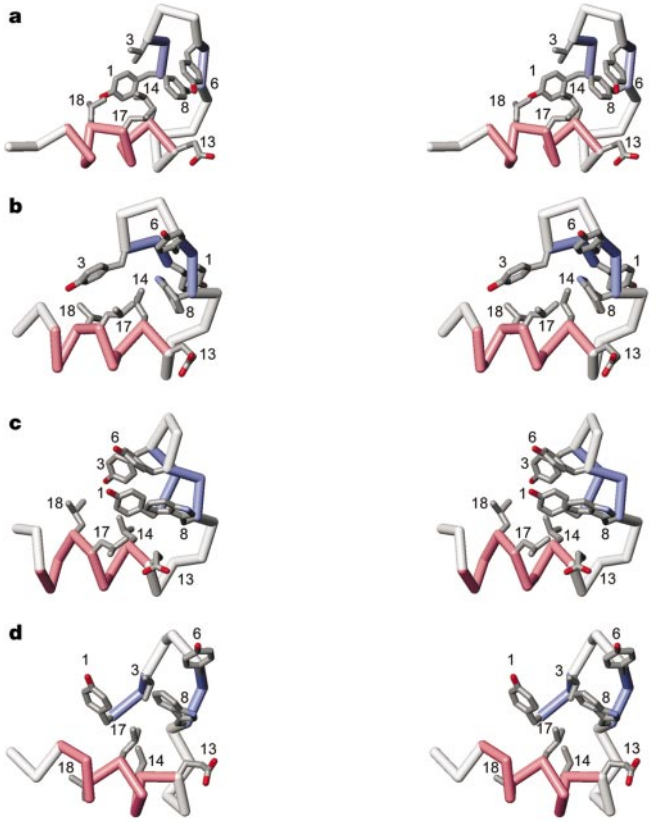

Figure 1 Folding results. Stereo diagrams showing the Cα backbone (blue, 1–3 and 6–8; red, 11–21) and selected side chains (Tyr 1, Val/Tyr 3, Tyr 6, Phe/Trp 8, Glu 13, Leu 14, Leu 17 and Leu 18) for folding trajectories terminating with low r.m.s.d.$_{Cα}$ to BBA5. **a**, NMR structure of BBA5. **b**, Double mutant at 2.2 Å. **c**, Double mutant at 2.4 Å. **d**, Single mutant at 2.5 Å. See Supplementary Information for examples of folding trajectories.

brium constant K_{eq} for the double mutant is shown as a function of temperature in Fig. 2c (see also Methods).

To investigate experimental folding kinetics, laser temperature jumps of 11 ± 1 °C were performed on both BBA5 mutants with final temperatures between 293 and 303 K. Relaxation to the new equilibrium concentration was monitored at 100-ns intervals by following the temporal shift of the fluorescence spectrum with a sub-microsecond, real-time multichannel fluorescence wavelength detector[17]. The double mutant response could be resolved above the 400-ns instrument response function (Fig. 2d; see also Methods).

The same 1.5 ± 0.7 μs relaxation rate was detected whether analysing intensity changes or wavelength shifts of the time-resolved fluorescence spectrum. Standard definitions for the folding rate and the equilibrium constant of a two-state folding protein resulted in a folding rate of 7.5 ± 3.5 μs at 298 K (see Methods). For the single mutant we can establish a limit $\tau < 10$ μs. The kinetic data are entirely compatible with single exponential folding kinetics, although we cannot exclude processes of small amplitude, or speed greater than the 400-ns resolution of our apparatus. Such processes could become relevant for fast folding over a low barrier (for example, a subpopulation of protein molecules might fold directly without need for further thermal excitation to the barrier).

Our choice of a tractable protein, a united atom model, an implicit solvent treatment, and the use of distributed computing techniques allowed aggregate simulation of over 700 μs. Folded state simulations were started from homology models (see Methods). We acquired 2,500 single-mutant folded simulations (10 ns) at both 278 K and 298 K. We also acquired 2,500 double-mutant folded simulations (10 ns) at 278 K, 378 K and 478 K. Twenty-five alternative double-mutant models generated 8,750 trajectories (5 ns) at 278 K. Folding trajectories were started from extended initial conformations. A total of 15,000 single-mutant folding simulations (20 ns) were acquired at both 278 K and 298 K; 9,000 and 8,500 double-mutant folding simulations were acquired, respectively, at 278 K (20 ns) and 338 K (10 ns). Considering all 32,500 folding trajectories, we observed the expected β-hairpin in over 1,100 independent trajectories and the α-helix in over 21,000 independent trajectories. We selected successful folding trajectories that possessed both native secondary structure and low alpha carbon position root mean square deviation (r.m.s.d.$_{C\alpha}$) with BBA5 (see Methods). Of 9,000 double-mutant folding trajectories at 278 K, 16 were folded after 20 ns (Fig. 3). Trajectories that fold quickly are not unphysical[10]; the fact that the average unfolded molecule explores conformational space for microseconds before finding the native conformation does not preclude individual molecules from folding quickly. Quite the opposite, the basic principles of two-state protein kinetics indicate that we must expect to observe a small fraction of molecules to fold on short timescales[18].

Our simulations were consistent with the experimentally determined folding rate and stability. Our two-state assumption was justified by the thermodynamic data. The interpolated simulated folding time at 298 K was about 6 μs for the double mutant, and we computed unfolding rates slightly higher than the folding rates. This is to be compared with the experimental folding time of 7.5 ± 3.5 μs and K_{eq} of 0.25 ± 0.1 (that is, the unfolding rate k_u, is about four times faster than the folding rate, k_f). The primary uncertainty in the computed rate was not the fit, but sensitivity to the folding criteria; varying the r.m.s.d.$_{C\alpha}$ cutoff by ±0.5 Å resulted in expected folding times at 298 K of 3–13 μs. Simulations of the single mutant at 298 K were more difficult to resolve and generated an expected

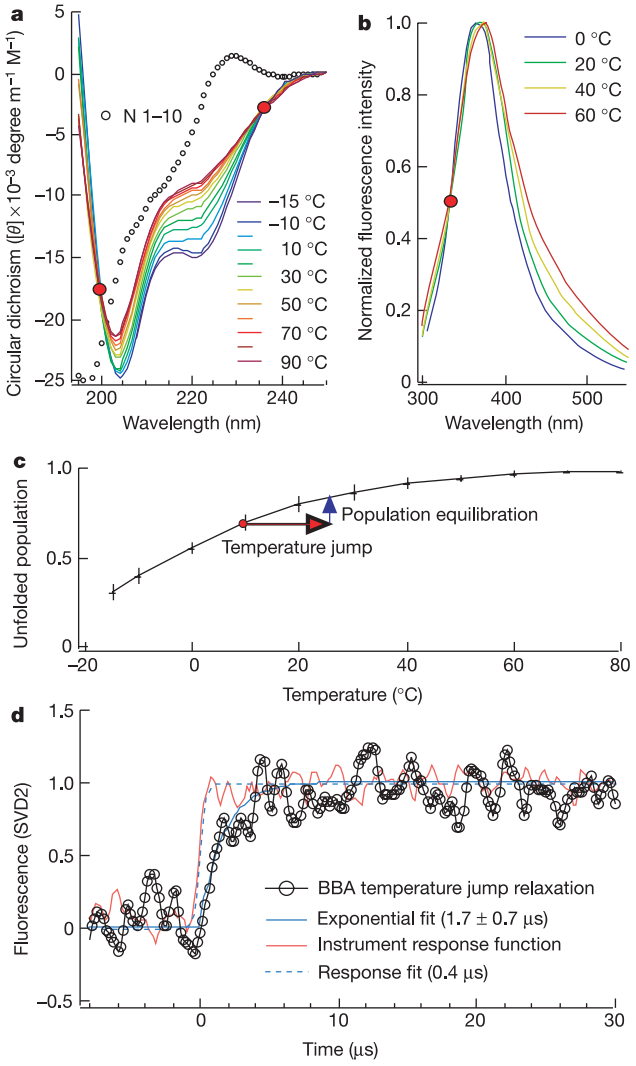

Figure 2 BBA5 double mutant thermodynamics and kinetics. **a**, Circular dichroism spectra from −15 °C to 90 °C. Isodichroic points (red circles) and the largest singular values (S2/S1 = 0.10, S3/S1 = 0.02, S4/S1 = 0.02, S5/S1 = 0.01; see Methods) indicate thermodynamic two-state behaviour. The N-terminal fragment is a random coil. Circular dichroism spectra of the single mutant (not shown) are very similar. **b**, Intensity normalized fluorescence spectra. The emission peak shifts from 366 nm to 377 nm and fluorescence intensity decreases by 2.5 as temperature increases from 0 °C to 60 °C. The 330-nm isosbestic point is further evidence of a two-state mechanism. **c**, Unfolded protein population with error bars. **d**, Temperature-jump relaxation kinetics and tryptophan instrument response at an average temperature of 298 K. The kinetics are normalized second singular value components of the time-evolving fluorescence spectrum ($\Delta\nu \approx 2$ nm). The single exponential fit with a relaxation time constant of 1.7 μs yields a deconvoluted relaxation time constant of 1.5 μs.

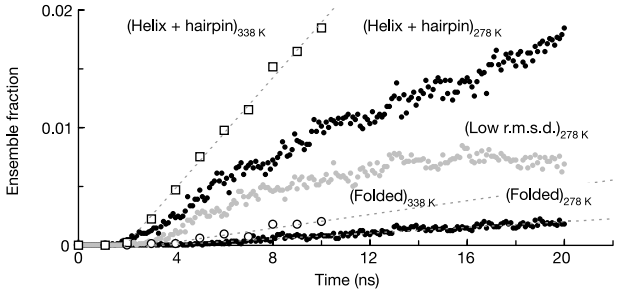

Figure 3 Folded population growth. The growing population of folded double mutant conformations, at 278 K (filled circles, dotted line) and at 338 K (open circles, dotted line), the population with low r.m.s.d.$_{C\alpha}$ (grey circles), and the population with both helix and hairpin at 278 K (filled circles) and 338 K (open squares, dotted line), are shown. The scale of the *y* axis illustrates the need to simulate thousands of trajectories. Note the increased rate of folding at high temperature.

folding time of 16 μs (7–43 μs), which was roughly consistent with experimental observation ($\tau < 10$ μs). A total of 10,000 trajectories at a timescale of tens of nanoseconds seem sufficient to provide a reasonable statistical estimate of the folding rate for this rapidly folding, two-state mini-protein.

It is important to address directly limitations in terms of methods and modelling. First, to discriminate folded structures, we rely on qualitatively reasonable but arbitrary criteria. Also, our simulations do not represent structure prediction, as our folded criterion relies on the Cα positions of BBA5. Besides methodological limitations, it is imperative to address whether established methods, such as the optimized potentials for liquid simulations (OPLS) parameter set, the generalized Born surface area solvent model, and stochastic dynamics within Tinker are able to sample native-like states without further optimization. The observed native state flexibility may be an indication that our force field was insufficiently accurate to describe completely the native structure free energy minimum. Finally, given a 400-ns instrument response function and a 1.5-μs relaxation rate, the uncertainty in the experimental folding rate is still large.

There is a growing body of knowledge about rapid protein dynamics. The fastest helices in apomyoglobin form in approximately 50 ns (ref. 19), and a collapsed hydrophobic core in the same molecule forms in about 10 μs (ref. 20). It has been suggested that helices and turns commonly form on the timescale of about 100 ns and 1 μs, respectively[1]. For the BBA5 single mutant, we can make the rough estimate that α-helix formation above the level in the unfolded state population has a relaxation time of about 500–800 ns (from 278 to 298 K). We further estimate that the β-hairpin has a relaxation time of about 1.7–1.5 μs (from 278 to 298 K). The β-hairpin forms more quickly in the double mutant with a relaxation time of only about 0.5 μs at 278 K, which demonstrates the sensitivity of simulated folding kinetics to a single mutation and explains why the estimate of the double-mutant folding rate exceeds the single mutant estimate. These folding rates approach the theoretical maximal folding rate estimated from protein chain diffusion[12]. The ability to simulate an approximately 10-μs process by sampling many shorter trajectories puts any number of rapidly folding proteins within our reach. Simulated rates could complement and explain quantities correlated with rate, such as contact order[21].

The two principal challenges of simulated protein folding—to build accurate models and to adequately sample conformational space—are not easily decomposed. Here, we have addressed an improved means of sampling; using many thousands of independent trajectories, we have explored the conformational space of two BBA5 mutants. We have observed their folding many times, a process that takes microseconds. We have studied the behaviour of ensembles of thousands of trajectories. Helical structure in the unfolded state, fragment secondary structure propensity, rate of helix formation, rate of hairpin formation, and the rate for folding were all consistent with the available experimental evidence. The ability to investigate directly folding dynamics and heterogeneity with molecular dynamics will bring molecular dynamics to the same prominence in protein folding that it has already achieved in structural modelling. □

Methods

Protein preparation

BBA5 mutants and fragments were synthesized with C-terminal amidation and N-terminal acetylation by Bioworld or the UIUC Biotech Center to 70% purity. High-performance liquid chromatography (HPLC) purification using a Vydac C18 reversed-phase analytical column yielded over 90% purity, verified by mass spectrometry. The samples were lyophilized after purification. To calculate concentrations, extinction coefficients at 280 nm were estimated for the mutants and fragments using the extinction coefficients of tyrosine and tryptophan residues. C-terminal fragment concentrations were determined by amino-acid composition analysis, using the Moore–Stein method[22].

Model preparation

The BBA5 double mutant model 1 was generated by homology to BBA1 (Protein Data Bank code 1HCW)[16] with 74% sequence identity. All applicable torsion values from the most representative member of the BBA1 NMR ensemble were taken for an initial conformation. Mutant residues used the Tinker[23] default torsions of 180° except for Trp 8 where a side-chain torsion angle χ1 of 0° relieved a steric clash. No memory Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton energy optimization[24] to r.m.s. gradient convergence, followed by 100 ps of molecular dynamics, provided a numerically stable initial conformation. The other double mutant models (2–26) were generated by homology to BBA5 (91% identical to the double mutant)[15]. We took each of the 25 NMR structures, removed Val 3 and Phe 8, placed Tyr 3 and Trp 8 simultaneously into their most favoured rotamers (holding the rest of the protein fixed) using a self-consistent mean field method[25], and minimized each structure including a generalized Born surface area energy term[26] at each step. To obtain an initial unfolded structure, a fully extended conformer was generated using Tinker (backbone torsions $\varphi,\psi = -135,135$). A total of 100 steps of minimization relieved a clash between the side chain and carbonyl of D-Pro 4. This extended structure was equilibrated for 100 ps using molecular dynamics, providing a numerically stable denatured structure. The single mutant and fragment simulations were started from similarly prepared extended conformations.

Circular dichroism and fluorescence measurements

Fluorescence spectra were excited at 285 nm in a 1-cm quartz cuvette at 100 μM peptide concentration. A thermocouple and circulating water bath measured temperature to ±0.2 °C. Sixty micromolar circular dichroism samples (0.3–2 mm cuvettes) were temperature equilibrated for 1 min for each temperature point. In addition to isosbestic points, singular value decomposition (SVD) was used to verify that the thermodynamic data could be modelled as a function of temperature with only two significant species; the third and higher SVD basis functions were negligible (see legend to Fig. 2): two species account for most of the observed circular dichroism and fluorescence spectral changes with temperature. The second SVD component was fitted to $K_{eq}(T)$, with and without baselines, to estimate the population error from inaccurate baselines (Fig. 2c). $K_{eq} = \exp[-\Delta G_f(T)/RT]$, where $\Delta G_f(T) = mT(T - T_m)$, ΔG_f is folding energy, T is temperature, $R = 8.31\ \mathrm{J\,K^{-1}\,mol^{-1}}$ and T_m is the melting temperature of BBAW (−2 ± 3 °C). The resulting unfolded fraction $(1 - K_{eq})/(1 + K_{eq})$ (Fig. 2c) changes from under 25% to over 75% over a 45 °C range.

Temperature jumps

Temperature jumps of 11 °C were induced by a 1.5-μm wavelength infrared pulse lasting 10 ns, generated by a Raman-shifted YAG laser. Free tryptophan in phosphate buffer was used to calibrate the temperature jump before each measurement series. Time-resolved, dispersed fluorescence data was taken by channelling the fluorescence (excited by 288-nm pumping with a frequency-tripled, mode-locked Ti:sapphire laser) by means of a liquid ultraviolet light guide through a monochromator, which imaged 60 nm of the spectrum (340–400 nm) into 15 channels of an array photomultiplier tube[17]. The net unfolding process was monitored by an approximately 11-nm fluorescence redshift. SVD as a function of time also revealed only two significant basis functions in the time-evolving fluorescence spectra. The measured relaxation rate was $k_o = k_f + k_u$, where $K_{eq} = k_f/k_u$. The reported rate of 1.5 μs was obtained by deconvoluting the 400-ns gaussian full-width-at-half-maximum instrument response from a 1.7-μs fit.

Molecular dynamics

Our simulations used software adapted from the Tinker molecular modelling package[23] and the OPLS united atom parameter set[27]. Constant temperature stochastic dynamics modelled the viscous drag of water ($\eta = 91\ \mathrm{ps^{-1}}$). We modelled solvation with the generalized Born/surface area implicit solvent model[26]. The electrostatic calculations used 16 Å cutoffs with 12 Å tapers. The bond lengths were constrained with the RATTLE algorithm allowing time steps of 2 fs[28]. His 21 was given neutral charge. Trajectory conformations were recorded at intervals of 0.1–1 ns.

For analysis, a conformation contained α-helix if it had at least four α-helical residues according to the program Dictionary of Protein Secondary Structure (DSSP)[29] with the default H-bond cutoff parameter of 0.5 kcal mol^{-1}. A conformation contained the native hairpin if it had a hairpin with β-bridges between residues 2–7 and 3–6 according to DSSP. To quantify the presence of expected tertiary structure, we aligned each conformation to the Cα positions of the reported low energy BBA5 NMR structure and calculated r.m.s.d.$_{Cα}$ using ProFit. A conformation had low r.m.s.d.$_{Cα}$ if it had r.m.s.d.$_{Cα}$ with BBA5 < 3.622 Å (one standard deviation above the equilibrated (10 ns) folded double-mutant ensemble mean). A conformation had high r.m.s.d.$_{Cα}$ if it had r.m.s.d.$_{Cα}$ with BBA5 > 5.0 Å (one standard deviation below the final unfolded ensemble mean). A conformation was folded if it had the helix, the hairpin, and low r.m.s.d.$_{Cα}$. A conformation was unfolded if it lacked either the helix or the hairpin, and had high r.m.s.d.$_{Cα}$.

Received 19 February; accepted 23 September 2002; doi:10.1038/nature01160.
Published online 20 October 2002.

1. Eaton, W. A., Muñoz, V., Thompson, P. A., Henry, E. R. & Hofrichter, J. Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. *Acc. Chem. Res.* **31,** 745–753 (1998).
2. 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).
3. Duan, Y. & Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282,** 740–744 (1998).
4. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267,** 1619–1620 (1995).
5. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4,** 10–19 (1997).

6. Shea, J. & 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).
7. Ferrara, P., Apostolakis, J. & Caflisch, A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B* **104,** 5000–5010 (2000).
8. Daura, X., Jaun, B., Seebach, D., Gunsteren, W. F. v. & Mark, A. E. Reversible peptide folding in solution by molecular dynamics simulation. *J. Mol. Biol.* **280,** 925–932 (1998).
9. Ferrara, P. & Caflisch, A. Folding simulations of a three-stranded antiparallel β-sheet peptide. *Proc. Natl Acad. Sci. USA* **97,** 10780–10785 (2000).
10. Zagrovic, B., Sorin, E. J. & Pande, V. S. β-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313,** 151–169 (2001).
11. 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).
12. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl Acad. Sci. USA* **97,** 7220–7225 (2000).
13. Bieri, O. *et al.* The speed limit of protein folding measure by triplet-triplet energy transfer. *Proc. Natl Acad. Sci. USA* **96,** 9597–9601 (1999).
14. Shirts, M. & Pande, V. S. Screen savers of the world unite. *Science* **290,** 1903–1904 (2000).
15. Struthers, M., Ottesen, J. J. & Imperiali, B. Design and NMR analyses of compact, independently folded BBA motifs. *Folding Des.* **3,** 95–103 (1998).
16. Struthers, M. D., Cheng, R. C. & Imperiali, B. Design of a monomeric 23-residue polypeptide with defined tertiary structure. *Science* **271,** 342–345 (1996).
17. Ervin, J., Sabelko, J. & Gruebele, M. Submicrosecond real-time fluorescence detection: application to protein folding. *J. Photochem. Photobiol. Biol.* **54,** 1–15 (2000).
18. Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68,** 2959–2970 (1978).
19. Gilmanshin, R., Williams, S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. Fast events in protein folding: relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl Acad. Sci. USA* **94,** 3709–3713 (1997).
20. Ballew, R. M., Sabelko, J. & Gruebele, M. Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA* **93,** 5759–5764 (1996).
21. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277,** 985–994 (1998).
22. Moore, S. & Stein, W. Amino acid determination, methods and techniques. *J. Biol. Chem.* **192,** 663–670 (1951).
23. Ponder, J. W. & Richards, F. M. An efficient Newton-like method for molecular mechanics energy minimization of large molecules. *J. Comput. Chem.* **8,** 1016–1024 (1987).
24. Fletcher, R. & Powell, M. J. D. A rapidly convergent descent method for minimization. *Comput. J.* **6,** 163–168 (1963).
25. Koehl, P. & Delarue, M. On the use of a self-consistent mean field theory to predict protein side chain conformations and estimate their entropies. *J. Mol. Biol.* **239,** 249–275 (1994).
26. Qiu, D., Shenkin, P. S., Hollinger, F. P. & Still, W. C. The GB/SA Continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* **101,** 3005–3014 (1997).
27. Jorgensen, W. L. & Tirado-Rives, J. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110,** 1657–1666 (1988).
28. Andersen, H. C. Rattle: a 'velocity' version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **52,** 24–34 (1983).
29. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

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

Acknowledgements We thank the Folding@Home volunteers whose processor power made this work possible; the members of the Pande, Levitt and Gruebele laboratories for discussion; J. Ottesen for the BBA5 coordinates; B. Imperiali for NMR data; the UIUC Laboratory for Fluorescence Dynamics; and the Suslick group for equipment use. C.S. was supported by a predoctoral Howard Hughes Medical Institute fellowship. V.P. and the Folding@Home project were supported by the National Institutes of Health (NIH), American Chemical Society-Petroleum Research Fund, National Science Foundation Materials Research Science & Engineering Centers, Center on Polymer Interfaces and Macromolecular Assemblies seed funds, and a gift from Intel. H.N. was supported by an NIH biophysics training grant. H.N. and M.G. were also funded by the NIH.

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

Correspondence and requests for materials should be addressed to V.S.P. (e-mail: pande@stanford.edu) or M.G. (e-mail: gruebele@scs.uiuc.edu).

erratum

Simulation of the atmospheric thermal circulation of a martian volcano using a mesoscale numerical model

Scot C. R. Rafkin, Magdalena R. V. Sta. Maria & Timothy I. Michaels

Nature **419**, 697–699 (2002).

In this Letter, "(see Supplementary Information)" should have appeared at the end of the third sentence of the third paragraph. At the end of the Letter, the line "**Supplementary Information** accompanies the paper on *Nature*'s website (http://www.nature.com/nature)." should have been included. □

corrigendum

Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid

Rachel J. Waugh O'Neill, Michael J. O'Neill & Jennifer A. Marshall Graves

Nature **393**, 68–72 (1998).

In this Letter, the maternal species listed for hybrid BE-1 is attributed to *Macropus eugenii*. However, our ongoing studies show that the maternal complement of chromosomes in BE-1 was inherited from a *Macropus rufogriseus* female. The centromeres of *M. rufogriseus* chromosomes are extended in comparison to all other macropod species. The extent, therefore, to which the centromere extensions shown in Fig. 4 can be attributed to hybrid-specific amplification of the retrolement KERV-1 cannot be precisely determined. Nevertheless, Southern analysis confirms that this retroelement is present at a 20% higher copy number in the hybrid's genome compared with that of its parents, and FISH analysis shows KERV-1 localization only to centromeres in the hybrid. Our conclusions regarding hybrid-specific undermethylation in this hybrid individual are not affected because *M. rufogriseus* shows methylation levels typical of species within the macropod group. Hybrid-specific undermethylation and genome rearrangement also remain true for the *Petrogale* hybrids we presented. □

