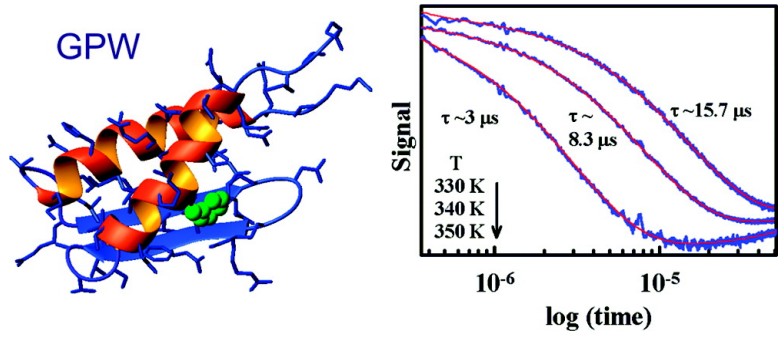
Article

Expanding the Realm of Ultrafast Protein Folding: gpW, a Midsize Natural Single-Domain with #+# Topology that Folds Downhill

Adam Fung, Peng Li, Raquel Godoy-Ruiz, Jose M. Sanchez-Ruiz, and Victor Mun#oz

J. Am. Chem. Soc., **2008**, 130 (23), 7489-7495 • DOI: 10.1021/ja801401a • Publication Date (Web): 14 May 2008



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES

Published on Web 05/14/2008

Expanding the Realm of Ultrafast Protein Folding: gpW, a Midsize Natural Single-Domain with α+β Topology that Folds Downhill

Adam Fung,[†] Peng Li,[†] Raquel Godoy-Ruiz,[†,‡] Jose M. Sanchez-Ruiz,[‡] and Victor Muñoz*[,†,§]

Contribution from the Department of Chemistry and Biochemistry, and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, Departamento de Química-Física, Facultad de Ciencias, Universidad de Granada, Granada 18071, Spain, and Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas (CSIC), Ramiro de Maeztu 9, Madrid 28040, Spain

Received February 27, 2008; E-mail: vmunoz@umd.edu; vmunoz@cib.csic.es

Abstract: All ultrafast folding proteins known to date are either very small in size (less than 45 residues), have an α-helix bundle topology, or have been artificially engineered. In fact, many of them share two or even all three features. Here we show that gpW, a natural 62-residue α+β protein expected to fold slowly in a two-state fashion, folds in microseconds (i.e., from $\tau = 33$ μs at 310 K to $\tau = 1.7$ μs at 355 K). Thermodynamic analyses of gpW reveal probe dependent thermal denaturation, complex coupling between two denaturing agents, and differential scanning calorimetry (DSC) thermogram characteristic of folding over a negligible thermodynamic folding barrier. The free energy surface analysis of gpW folding kinetics also produces a marginal folding barrier of about thermal energy (*RT*) at the denaturation midpoint. From these results we conclude that gpW folds in the downhill regime and is close to the global downhill limit. This protein seems to be poised toward downhill folding by a loosely packed hydrophobic core with low aromatic content, large stabilizing contributions from local interactions, and abundance of positive charges on the native surface. These special features, together with a complex functional role in bacteriophage λ assembly, suggest that gpW has been engineered to fold downhill by natural selection.

Introduction

Following the development of ultrafast folding techniques,[1–4] several proteins have been found to fold in microseconds.[5] These timescales are in contrast with the tens of milliseconds to seconds considered typical for single domain two-state proteins.[6] The discovery of ultrafast folding has shifted the emphasis to directly asking what is the folding speed limit and what physical factors determine folding rates. Current empirical estimates set the folding speed limit at ∼1 μs for a medium-sized domain at typical laser T-jump temperatures (e.g., 330−340 K).[7,8] Such folding speed limit suggests that microsecond folding proteins cross marginal barriers or even no barrier at all.[9] Proteins with maximal folding barrier ≤3*RT* (i.e., at the denaturation midpoint) can be practically considered within the downhill regime.[10] When the maximal barrier is near zero (or the surface is concave) the protein is said to fold globally downhill,[11] whereas a marginal barrier at the midpoint should disappear in sufficiently strong native conditions.[12] Interestingly, the downhill folding regime results in characteristic thermodynamic signatures[13] so that thermodynamic analyses can be used both to diagnose downhill folding[14,15] and to achieve an atom-by-atom characterization of the unfolding process.[16] Downhill folding proteins are also attractive targets for single molecule experiments.[17]

However, inspection of the existing ultrafast folding database suggests that this club has very stringent admission rules. The available microsecond folding proteins have one or several of these features: very small size (<45 residues), α-helix bundle

[†] University of Maryland.
[‡] Universidad de Granada.
[§] Consejo Superior de Investigaciones Científicas (CSIC).

(1) Eaton, W. A.; Muñoz, V.; Thompson, P. A.; Chan, C. K.; Hofrichter, J. *Curr. Opin. Struct. Biol.* **1997**, *7*, 10–14.
(2) Dyer, R. B.; Gai, F.; Woodruff, W. H. *Acc. Chem. Res.* **1998**, *31*, 709–716.
(3) Gruebele, M. *Annu. Rev. Phys. Chem.* **1999**, *50*, 485–516.
(4) Roder, H.; Maki, K.; Cheng, H. *Chem. Rev* **2006**, *106*, 1836–1861.
(5) Muñoz, V. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 395–412.
(6) Jackson, S. E. *Folding Des.* **1998**, *3*, R81–R91.
(7) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76–88.
(8) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193–197.
(9) Naganathan, A. N.; Doshi, U.; Muñoz, V. *J. Am. Chem. Soc.* **2007**, *129*, 5673–5682.
(10) Gruebele, M. In *Protein Folding, Misfolding and Aggregation. Classical Themes and Novel Approaches*; Muñoz, V., Ed.; RSC: Cambridge, 2008.
(11) Naganathan, A. N.; Perez-Jimenez, R.; Sanchez-Ruiz, J. M.; Muñoz, V. *Biochemistry* **2005**, *44*, 7435–7449.
(12) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167–195.
(13) Muñoz, V. *Int. J. Quantum Chem.* **2002**, *90*, 1522–1528.
(14) Garcia-Mira, M. M.; Sadqi, M.; Fischer, N.; Sanchez-Ruiz, J. M.; Muñoz, V. *Science* **2002**, *298*, 2191–2195.
(15) Naganathan, A. N.; Doshi, U.; Fung, A.; Sadqi, M.; Muñoz, V. *Biochemistry* **2006**, *45*, 8466–8475.
(16) Sadqi, M.; Fushman, D.; Muñoz, V. *Nature* **2006**, *442*, 317–321.
(17) Schuler, B.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

topology, and artificial origin (i.e., either designed de novo or mutagenized toward increased folding speed). The small size is easily rationalized by the strong size-scaling of protein folding barriers.[18,19] Helix bundles exemplify a simple topology with largest contributions from local interactions and, thus, low contact order.[20] De novo designed proteins fold faster than their natural counterparts,[21] whereas Gly→Ala substitutions in α-helices speed folding.[22] In fact, the only known microsecond folding proteins with β-structure are WW domains,[23–25] which in turn have the smallest size (~32 residues) and are relatively slow. The important question emerging from these considerations is whether the realm of ultrafast folding is indeed so restrictive as to exclude the majority of known structural scaffolds.

Here we address this question investigating the folding of gpW, a protein that does not exhibit any of such trademarks. GpW is a single gene product from bacteriophage λ.[26] With 62 residues, gpW is a medium-sized domain similar in length to several slow two-state folding proteins.[6] Its structure and folding topology with two α-helices and two β-strands is organized in an α+β fold[26] (Figure 1A). Detailed topological features of gpW can be best observed in the contact map of Figure 1B, which illustrates the many long-range contacts involved in bringing together the two α-helices on top of the central hairpin. Accordingly, the relative contact order[20] for gpW suggests folding in the millisecond range. However, during exploratory stopped-flow refolding experiments we observed that the far-uv circular dichroism (f-CD) signal of native gpW was fully recovered within the 2 ms instrumental dead-time. These preliminary experiments pointed to gpW as a candidate of new natural fold with ultrafast folding and as a model case to test recent work in downhill folding.[15]

Materials and Methods

Protein Samples. The gpW gene was subcloned from the original construct provided by Alan Davidson[26] into the pBAT vector using 5′ NcoI and 3′ *Hin*dIII restriction sites and eliminating the N-terminal hexahistidine tag and six unstructured C-terminal residues that induced aggregation.[26] All the experiments were performed in 20 mM sodium phosphate buffer pH 6.0. Protein concentration was determined by absorbance using $\epsilon_{280} = 1888$ M^{-1} cm^{-1}.

Circular Dichroism. F-CD experiments were performed at 25 μM protein concentration with 1 mm path length. Near-UV circular dichroism (n-CD) experiments were performed at 80 μM with 10 mm path length. Spectra were measured in a thermostatted Jasco J-810 spectropolarimeter in continuous mode at 10nm/min with 2nm bandwidth at temperatures from 268 to 362 K every 3 K.

Fluorescence. Fluorescence emission spectra were collected at 6 μM protein concentration in a thermostatted Flurolog-3 spectrofluorimeter (Jobin Yovin, Inc.) exciting at 280 nm, 2 nm slit widths, and 0.25 s integration time at temperatures from 273 to 361 every 2 K.

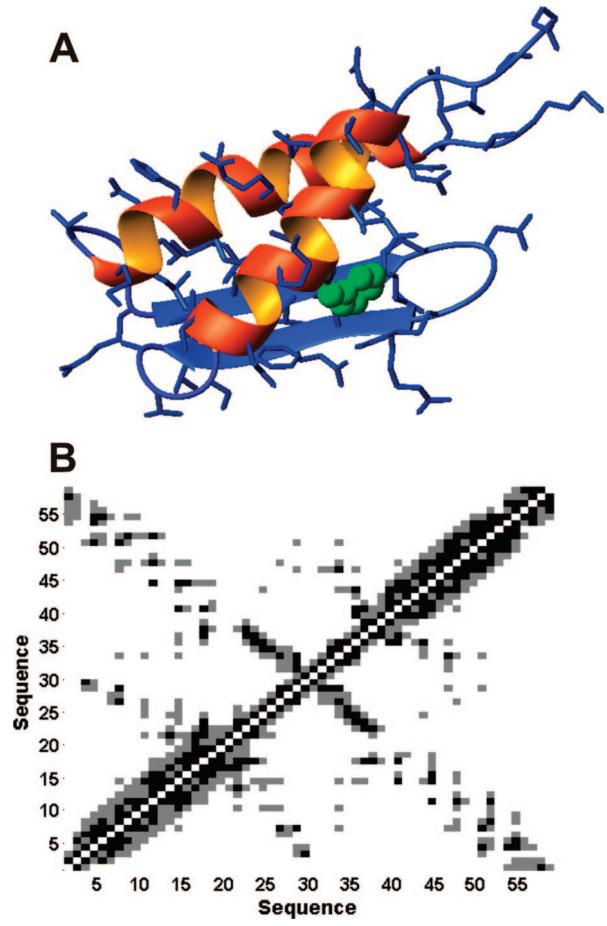


Figure 1. (A) Schematic representation of gpW three-dimensional structure with the side-chain of Y46 shown as green spheres. (B) Contact map of gpW.

Fourier Transformed Infrared (FTIR). Spectra were measured on a thermostatted FTS-300 IR Spectrometer (BioRad) at 2 cm^{-1} resolution using $CaCl_2$ windows and a 50 μm Teflon spacer between 277 and 365 K. The sample was prepared at 4 mg/mL protein concentration in 99.9% D_2O and same buffer.

Differential Scanning Calorimetry (DSC). Experiments were performed with a VP-DSC calorimeter from MicroCal (Northampton, MA) at a scan rate of 1.5 K/min. Protein samples were prepared by exhaustive dialysis against the buffer. The experiments, controls, and analysis were performed exactly as explained in ref 27.

IR Laser T-Jump Kinetics. Time-resolved IR experiments were performed on a custom built version of the instrument developed by Feng Gai and co-workers.[28] Briefly, the fundamental of a Continuum Surelite I-10 Nd:YAG laser run at 2 Hz is shifted to ~1.9 μm with a 1 m path length Raman cell (Lightage) filled with a mixture of Ar and H_2 at 1000 psi to heat water by vibrational excitation. Heating pulses of ~20 mJ were used to generate ~11 K jumps. A CW lead salt diode laser (Laser Components) and a MCT (Mercury−Cadmium-Telluride) detector with 50 MHz bandwidth are used to monitor time-dependent changes in IR absorption at 1632 cm^{-1}. D_2O buffer was used for background subtraction and as an internal thermometer. Infrared cell and sample preparation were identical to the equilibrium FTIR experiments.

(18) Naganathan, A. N.; Muñoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 480–481.
(19) Kouza, M.; Li, M. S.; O'Brien, E. P.; Hu, C. K.; Thirumalai, D. *J. Phys. Chem. A* **2006**, *110*, 671–676.
(20) Plaxco, K. W.; Simons, K. T.; Baker, D. *J. Mol. Biol.* **1998**, *277*, 985–994.
(21) Scalley-Kim, M.; Baker, D. *J. Mol. Biol.* **2004**, *338*, 573–583.
(22) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305–310.
(23) Crane, J. C.; Koepf, E. K.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2000**, *298*, 283–292.
(24) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13002–13007.
(25) Nguyen, H.; Jager, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3948–3953.
(26) Maxwell, K. L.; Yee, A. A.; Booth, V.; Arrowsmith, C. H.; Gold, M.; Davidson, A. R. *J. Mol. Biol.* **2001**, *308*, 9–14.
(27) Guzman-Casado, M.; Parody-Morreale, A.; Robic, S.; Marqusee, S.; Sanchez-Ruiz, J. M. *J. Mol. Biol.* **2003**, *329*, 731–743.
(28) Huang, C. Y.; Getahun, Z.; Zhu, Y. J.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2788–2793.

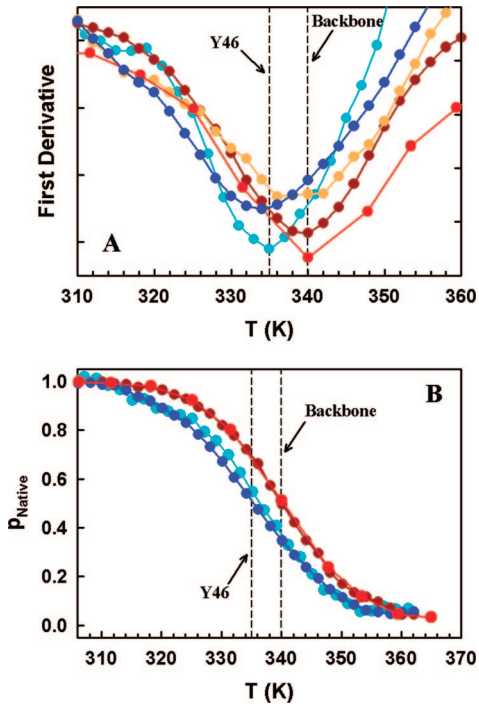


Figure 2. (A) Derivative of equilibrium unfolding curves of gpW: (dark red) f-CD; (red) FTIR; (cyan) fluorescence; (blue) amplitude of 1st SVD component of n-CD; (orange) amplitude of 2nd SVD component of n-CD. (B) Native probability from phenomenological two-state fits to the equilibrium unfolding curves (color code as in A).

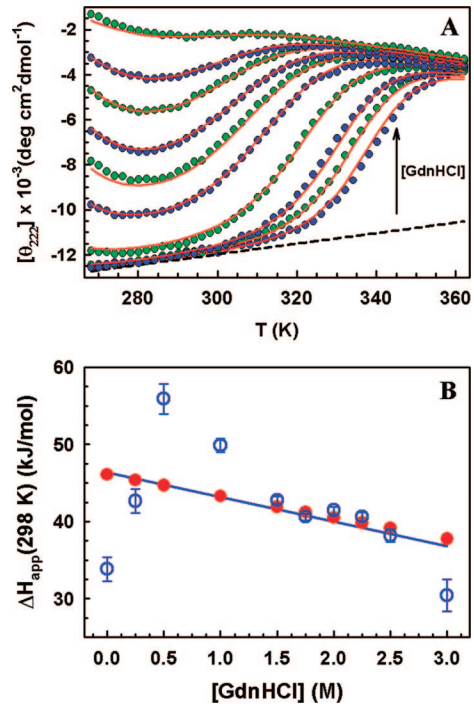


Figure 3. (A) Temperature-GnHCl double perturbation experiment of gpW. The GnHCl concentrations are 0, 0.25, 0.5, 1, 1.5, 1.75, 2, 2.25, 2.5, and 3 M. The red lines are the best fit to a global two-state model with native baseline shown as a dashed black line. (B) Plot of the apparent unfolding enthalpy at 298 K versus GnHCl concentration: (open blue circles) individual two-state fits; (blue line) linear regression of the values from individual two-state fits; (closed red circles) global two-state fit.

Fitting Kinetic Data to Free Energy Surface Model. The analysis was done by fitting the relaxation rates and amplitudes from the IR data to the same model implementation used and described in detail before.[9] The fixed parameters were $\Delta S_{res} = 16.5$ J/(mol·K·res),[9] $\kappa\Delta C_p = 4.3$,[9] $\Delta C_{p,res} = 40$ J/(mol·K·res) (obtained from the double perturbation experiment). The fitted parameters were $\Delta H^{340K} = 4.75$ kJ/(mol·res), $\kappa_{\Delta H} = 1.22$, $D_0 = 7108\ n^2 \cdot s^{-1}$ (at 333 K), and $E_a = 59.2$ kJ/mol.

Thermodynamic Analysis of Marginal Folding Barriers

The first thermodynamic test that was developed to identify downhill folding is based on the observation and quantitative analysis of probe dependence on equilibrium unfolding experiments.[13,14] GpW has no tryptophans but does have a single tyrosine (Y46) that sits on the beginning of α-helix 2 and interacts with the central β-hairpin (shown in green in Figure 1A). Therefore, Y46 is a useful probe of tertiary environment that can be investigated both by fluorescence and near-uv circular dichroism (n-CD). The average changes in secondary structure of gpW upon unfolding can be also investigated with two independent techniques: f-CD and FTIR.

The equilibrium thermal unfolding experiments of gpW monitored by the four techniques (f-CD, n-CD, fluorescence, and FTIR) do show signs of probe dependence (Figure 2). The derivative of the f-CD and FTIR unfolding curves exhibit their extremum (indicative of the midpoint temperature, T_m) at ~340 K, whereas the derivative of the fluorescence experiment and that of the first singular value decomposition (SVD) component of the n-CD experiment have their extremum at ~335 K (Figure 2A). Therefore, the tertiary environment around Y46 and the backbone secondary structure (mostly the two α-helices) do not unfold concertedly in gpW, with a higher thermal stability for the backbone. Similar decoupling between melting of tertiary and backbone structure, but at a larger scale, has been reported for the global downhill folder BBL.[16] The determination of the T_m from the derivative of the unfolding curve has shown to be a powerful method, which does not require fitting the data to a given model nor the tracing of pre- and post-transition baselines.[29] Nevertheless, the same 5 K T_m mismatch between structural probes is found when the data is fitted to a phenomenological two-state model (Figure 2B). These experiments also have an internal control: we find very good agreement between T_m values obtained with different instruments (f-CD with FTIR and n-CD with fluorescence) for the same probe, but a 5 K difference for the two probes measured with the same CD instrument. The control can be taken even one step beyond. Because Y46 is located in one of the helices, its side-chain environment is also affected by helix melting resulting in a minor second SVD component for the n-CD signal that exhibits the "backbone" T_m (see orange curve in Figure 2A). The structural decoupling observed in gpW unfolding is a first indication of marginal folding barrier.

In a second test, the coupling between two protein denaturation agents is investigated to determine whether it follows a simple Maxwell relationship, as expected for two-state folding, or is more complex due to conformational ensemble shifts during unfolding.[30] Figure 3A shows the effect of increasing the guanidinium chloride (GdnHCl) concentration on the thermal denaturation of gpW monitored by f-CD. The experiments have been fitted to a global two-state model, which renders an apparently reasonable fit (red lines in Figure 3A). However, closer inspection reveals systematic deviations, especially below 2 M GdnHCl. For example, the fit underestimates the stability

(29) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2007**, *445*, E17−E18.
(30) Oliva, F. Y.; Muñoz, V. *J. Am. Chem. Soc.* **2004**, *126*, 8596–8597.

at 0 M and overestimates it at 1 M. Indeed, the global fit SLS is 5.7 times higher than for the sum of individual two-state fits (keeping ΔC_p constant). Accordingly, the statistical F-test[29] indicates a probability $<10^{-10}$ that such deviations from two-state behavior are not significant. Figure 3B shows the comparison between the apparent unfolding enthalpy at 298 K (ΔH_{app}) obtained from individual and global fits. As documented before for BBL,[11] the ΔH_{app} versus denaturant concentration plot for gpW is characteristically curved (blue circles in Figure 3B). Interestingly, the linear regression of this data (blue line in Figure 3B) is almost identical to the output from the global fit (red circles). Thus, the global two-state fit reproduces the first order behavior but fails to account for the second order effects arising from complex coupling between denaturing agents.

In another recently developed method, the unfolding enthalpy fluctuations measured by DSC are used to derive a one-dimensional folding free energy surface with enthalpy as order parameter.[31] This quantitative analysis provides a direct estimate of the thermodynamic folding barrier (understood as the relative population at the top of the free energy surface).[31] Moreover, thermodynamic barriers obtained with this method correlate strongly with folding rates for 15 single-domain proteins.[32] The DSC thermogram of gpW is shown in Figure 4A in absolute heat capacity units, as required for this analysis.[31] Although the thermogram is nicely peaked, the DSC transition is broad for a protein of this size and T_m. This can be best seen by direct comparison with the DSC thermogram of spectrin's SH3 domain,[33] a very slow two-state folder of similar size (61 residues) and T_m (green curve in Figure 4A). Moreover, the heat capacity values at low temperature (what would be considered "native" baseline in a two-state analysis) are higher and more sloped than expected from Freire's correlation[34] (black open circles in Figure 4A). The low temperature slope is also higher than for the two-state spectrin SH3. In fact, fitting gpW's thermogram to a two-state model results in "native" and "unfolded" baselines that cross within the transition (black lines in Figure 4A). These observations clue in the presence of a marginal ($\leq 3RT$) thermodynamic folding barrier in gpW.

Quantitative analysis with the variable-barrier model[31] confirms this interpretation. The model fits the experimental data remarkably well with only four fitting parameters (red curve in Figure 4A). The folding free energy surface at the characteristic temperature (T_0, temperature at which the free energy difference between barrier top and the bottom of the two wells is equal) from the best fit has an extremely shallow barrier of 0.6 kJ/mol (inset of Figure 4B). To better estimate the confidence range, we also fitted the data with native baselines shifted one standard deviation up or down from Freire's baseline (error bars in the black symbols of Figure 4A).[34] Down- and up-shifted baselines resulted in much poorer fits (data not shown) and rendered barriers of 0 and 2.7 kJ/mol, respectively. Therefore, gpW has a thermodynamic folding barrier that is positive, but below thermal energy (i.e., the full free energy scale in the inset of Figure 4B). As a consequence the probability density is never

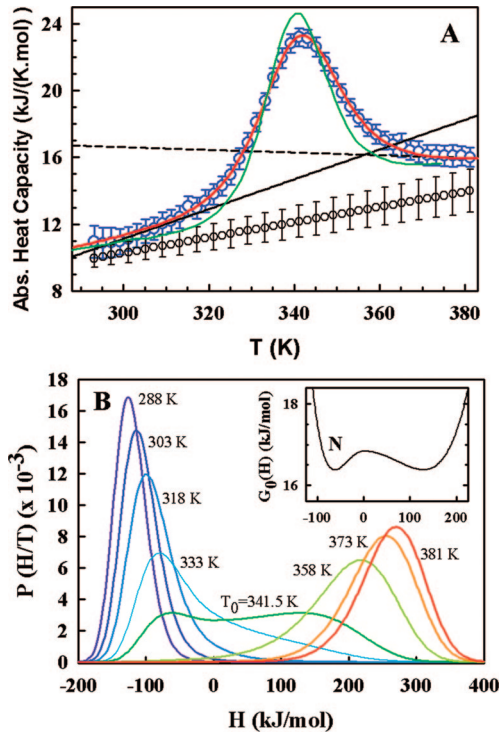


Figure 4. (A) gpW DSC thermogram (blue open circles) with error bars showing the standard deviation from six experiments at different protein concentrations; gpW native baseline (black open circles) predicted by Freire's empirical correlation;[34] best fit to the variable-barrier mode (red curve); DSC thermogram for spectrin SH3 (green curve) obtained from ref 33; native (black lines; continuous) and unfolded (dashed) baselines from a two-state fit. (B) Probability density as a function of the order parameter enthalpy at different temperatures from the fit. (Inset) One-dimensional free energy surface for gpW folding at the characteristic temperature (T_0). The letter N signals the native state.

bimodal (Figure 4B). The distribution moves from the folded basin as temperature raises, develops a higher enthalpy tail that grows into a second hump at T_0 and then again becomes a single moving peak corresponding to an increasingly unstructured unfolded basin. In summary, from a thermodynamic standpoint gpW exhibits the exact features expected for a protein that approaches the global downhill limit (maximal barrier $<RT$).

Ultrafast Folding Kinetics

The issue that remains is whether the thermodynamic signatures of near downhill folding observed in gpW correspond to ultrafast folding kinetics. To address this question we performed nanosecond laser-induced T-jump experiments employing infrared absorption at 1632 cm^{-1} as a probe of backbone secondary structure and following the instrumental setup developed by the Dyer and Gai groups.[2,28] In particular, we performed T-jumps of ~11 K from initial temperatures in the 300−350 K range. The infrared relaxation after the T-jump for these experiments is very well fit to single exponential decay with relaxation times (τ) between 33 and 1.7 μs (Figure 5A). The temperature dependence of the relaxation rate ($1/\tau$) is shown in Figure 5B (blue circles). The rate versus $1/T$ plot reveals that gpW folds indeed in the microsecond regime characteristic of ultrafast folding.[9] The smooth trends in the plot and the small fitting errors (error bars are smaller than the size of the circles in Figure 5B) highlight the high data quality, which permits to distinctly observe a slight increase in slope above the T_m (below 2.94 in Figure 5B). A precise determination of the curvature in

(31) Muñoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646–17651.
(32) Naganathan, A. N.; Sanchez-Ruiz, J. M.; Muñoz, V. *J. Am. Chem. Soc.* **2005**, *127*, 17970–17971.
(33) Viguera, A. R.; Martinez, J. C.; Filimonov, V. V.; Mateo, P. L.; Serrano, L. *Biochemistry* **1994**, *33*, 2142–2150.
(34) Freire, E. *Protein stability and folding*; Humana Press: Totowa, NJ, 1995.

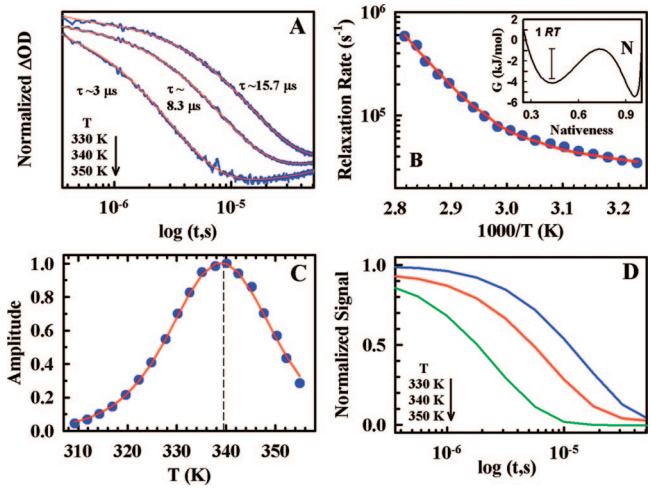


Figure 5. (A) IR relaxation decays after T-jumps of ~11 K. The red curves correspond to fits to a single exponential decay. The slight drift at the longest times corresponds to the beginning of recooling after the T-jump. (B) Relaxation rates obtained from single exponential fits to the relaxation decays of gpW versus $1/T$ (blue circles); fit to phenomenological one-dimensional free energy surface model (red curve). (Inset) 1D folding free energy surface at the midpoint temperature (i.e., 340 K) from fit to the kinetic data of gpW. The letter N signals the native state. (C) Kinetic amplitude of the T-jump relaxation as a function of the final temperature (blue circles); fit to free energy surface model (red curve); T_m of the equilibrium FTIR data (dashed line). (D) Simulation of the relaxation decays shown in panel A with the fitted free energy surface model.

the rate versus $1/T$ plot is important because it is directly related to the height of the kinetic folding barrier.[9] In fact, the slight curvature in the gpW data is already suggestive of a marginal folding barrier.

Another important piece of information is the amplitude of the relaxation decay as a function of the final temperature (Figure 5C). The kinetic amplitude is directly related to the changes in population after the T-jump and, thus, it is a kinetic determination of the derivative of the equilibrium unfolding curve. The kinetic amplitude permits to independently assess the compliance between kinetic and equilibrium data, rather than enforcing their agreement as is typically done in standard two-state analyses. For gpW the agreement between equilibrium derivative (red curve in Figure 2A) and kinetic amplitude (blue circles in Figure 5C) is excellent. The $T_m \sim 340$ K for the FTIR experiment (shown in Figure 5C as a dashed line) corresponds exactly to the maximum in the amplitude data and so agrees the overall shape of both curves. Equilibrium and kinetics should be consistent with each other regardless of whether folding is two-state or downhill, and thus, this is an important experimental control. Moreover, the combination of rate and amplitude makes possible the complete independent analysis of gpW folding kinetics with an empirical free energy surface approach.

Using the simple phenomenological model previously applied to other ultrafast folding proteins,[9] we can fit all of the kinetic observations of gpW and obtain a kinetic estimate of its 1-D folding free energy surface. As implemented originally,[9] the model has several fixed empirical parameters and only two fitting parameters ($\kappa_{\Delta H}$, ΔH^{385}) to determine shape and temperature dependence of the free energy surface. ΔH^{385} is the change in enthalpy upon unfolding at the reference temperature of 385 K. $\kappa_{\Delta H}$ determines the shape of free energy surface and the folding barrier height. T-jump kinetics are then calculated as diffusion on the free energy surface at the final temperature with a diffusion coefficient of the form $D = D_0 \exp(-E_a(1/T - {}^1/_{333})/R)$, where E_a and D_0 are also fitting parameters. The amplitude of the relaxation is determined by the population redistribution upon the T-jump and a switching function to represent the spectroscopic signal.[9]

The four parameters fit to the model reproduces the gpW experimental data very well: (1) the exponential relaxation decays (Figure 5D); (2) the changes in rate with temperature (red curve in Figure 5B); (3) and the kinetic amplitudes (red curve in Figure 5C). The activation energy of the diffusion coefficient (E_a) produced by the fitting is 59 kJ/mol, in close agreement with the empirical estimate of ~1 kJ/mol per residue obtained from other ultrafast folding proteins.[9] Other fitted parameters are given in Materials and Methods. From the fit to the rate and amplitude data we obtain a gpW folding free energy surface with a very small free energy barrier of ~2.8 kJ/mol at the T_m (inset to Figure 5B). Therefore, the kinetically determined folding barrier for gpW is also marginal (~RT) at its maximal value (i.e., at the midpoint). The agreement between this kinetic barrier and the DSC thermodynamic barrier is noteworthy, cross-validating the two approaches.

Discussion

A series of thermodynamic tests for the identification of downhill folding[15] indicate that the naturally single domain $\alpha+\beta$ protein gpW folds within the downhill folding regime. The variable-barrier analysis[31] of gpW DSC thermogram produces a negligible thermodynamic folding barrier (i.e., $<RT$). In parallel, T-jump kinetic experiments show folding kinetics in the tens to one microsecond range, thus placing gpW within the ultrafast folding group.[5,7] Furthermore, the free energy surface analysis of gpW folding kinetics also produces a marginal barrier (i.e., ~RT). From the combination of all these results we classify gpW as a downhill folder approaching the global downhill limit.

Thus, gpW is an excellent model case to compare existing tests for the identification of downhill folding and to address some misconceptions that have emerged in the recent literature. Protein folding barriers range from the many RT, resulting in two-state-like behavior to zero.[35] Therefore, the downhill folding issue is a quantitative one that must be addressed through thorough quantitative analysis. When the maximal (i.e., midpoint) folding barrier is below $3RT$, the equilibrium behavior exhibits signature features simply because the population at the top of the barrier becomes experimentally significant.[5] Such features unambiguously indicate that the folding barrier is marginal,[5] and thus likely to result in downhill folding in favorable conditions.[3,9,36] However, whether the midpoint barrier is $2RT$ or the protein folds globally downhill can only be established quantitatively, as it was done originally,[14] and subsequently corroborated with other approaches,[15,16,31] for the protein BBL. This simple point has been seriously misconstrued in recent work,[37–39] in which the quantitative aspect is ignored altogether by setting an artifactual dichotomy between global

(35) Akmal, A.; Muñoz, V. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 142–152.
(36) Cho, S. S.; Weinkman, P.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 118–123.
(37) Ferguson, N.; Schartau, P. J.; Sharpe, T. D.; Sato, S.; Fersht, A. R. *J. Mol. Biol.* **2004**, *344*, 295–301.
(38) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L. M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 123–127.
(39) Yu, W.; Chung, K.; Cheon, M.; Heo, M.; Han, K.-H.; Ham, S.; Chang, I. J. *Proc. Nat. Acad. Sci. U.S.A.* **2008**, *105*, 2397–2402.

downhill and what is loosely characterized as "cooperative" folding (even though global downhill folding also exhibits cooperativity[16]).

We show here that the equilibrium thermal unfolding of gpW is probe-dependent, with 5 K T_m difference between probes monitoring tertiary interactions and secondary structure. Although small compared to the T_m spread found for BBL,[16] this result is very robust because each structural probe produces the same T_m with different instruments and with different analytical methods (phenomenological two-state fit or derivative), whereas the same instrument and analytical method reproduce the 5 K mismatch between probes. In turn, 5 K is likely a lower bound to the overall T_m spread in gpW because it comes from only two probes (using four experimental techniques) that report on global structural features. From this result alone we can conclude that the midpoint barrier of gpW is within the 3*RT* limit that defines marginal barrier folding (also termed downhill generically[10]). To go beyond this point requires a detailed quantitative analysis. The limited structural information provided by the spectroscopic probes available in gpW does not warrant a full-blown analysis of probe dependence, as it was done for BBL.[14,16] It is from the quantitative analyses of DSC thermogram and T-jump kinetics that we can establish that the midpoint folding barrier of gpW is positive but of the order of thermal energy (near global downhill). The double perturbation test[30] also successfully detects that gpW folds crossing a marginal barrier (Figure 2B), but seems not to be sufficiently quantitative to distinguish between the 1*RT* bump of gpW and global downhill.

The inherently quantitative nature of the downhill folding issue highlights the importance of employing high quality experimental data in the analysis. It is also convenient to establish reference points for both extremes (global downhill and two-state-like). For the DSC analysis, it is critical to produce thermograms in absolute heat capacity units that lead to direct estimates of the native baseline.[31] The uncertainty in the DSC native baseline of gpW is very small, as indicated by the large degradation in fit quality after minimal down- or up-shifts of Freire's estimate. This same analysis recently applied to the 35-residue ultrafast folding villin headpiece subdomain produced a higher thermodynamic barrier (2*RT*) and suggested larger uncertainty in the native baseline for this protein.[40] Such small proteins, which are most likely to fold ultrafast via marginal barriers,[18] have intrinsically broad unfolding transitions and lack reference examples of truly two-state slow folding.[9] Being gpW a midsize domain, its DSC thermogram can be directly compared with those of slow two-state folding proteins of similar size. Such comparison nicely illustrates the differences between downhill and two-state thermograms (Figure 4A), and the clear connection between thermogram shape and overall folding rate for natural single domain proteins.[32]

The estimate of the folding barrier from the analysis of kinetic data mostly depends on the relative temperature dependence of the relaxation rate before and after the T_m.[9] As we show here for gpW, by obtaining the T_m directly from the kinetic amplitude, we eliminate the need to enforce agreement with the equilibrium data. This fact together with the low noise in the rate versus temperature data result in a precise estimate of the kinetic folding barrier. In fact, an agreement within *RT* between the thermodynamic and kinetic estimates of the folding barrier is rather remarkable, indicating that gpW has a smooth folding free energy landscape.[41] A relaxation rate of 1/(33 μs) at 310 K for gpW is also in reasonable agreement with the previously reported correlation between folding rates and DSC thermodynamic barriers.[32] The kinetic analysis of gpW also addresses an important issue regarding the common misconception of exponential decays as signature of barrier crossing ("activated") kinetics. For gpW, we find perfect exponential decays throughout (Figure 5A), but this protein crosses a minimal ($\sim$*RT*) kinetic barrier at the midpoint and (un)folds completely downhill below 325 K or above 355 K. As shown in Figure 5D, the kinetic simulations on the free energy surfaces obtained from the analysis of gpW data reproduce exactly the exponential decays observed experimentally. Therefore, exponential decays cannot be used as evidence of kinetic barriers, a point that has also been made using simulations in various kinetic models.[13,42] A great deal of confusion could be avoided in the folding literature simply by taking this issue into consideration.

Our results in gpW show that ultrafast folding and, more specifically, downhill folding are more ubiquitous than previously thought. Both by size and structural criteria, gpW is expected to fold in a two-state fashion,[15] yet gpW folds nearly downhill in microseconds. What is the origin of this behavior? There are three structural features that gpW shares with the global downhill folder BBL and, thus, may be involved in inducing downhill folding. The first one is a large fraction of positively charged amino acids. The resulting long-range electrostatic repulsions could broaden the native basin of attraction thereby lowering the folding barrier. It has recently been put on the spotlight by theoretical analyses of BBL[36] and gpW[43] that these proteins also share a loosely packed hydrophobic core. In such loose core, contributions from many-body interactions to overall stability are presumably small. This effect should lower the folding barrier, as it has been described theoretically,[44] observed in simulation,[45] and proposed empirically.[35] Finally, gpW appears to have large contributions from local interactions, showcased by the very high intrinsic helical propensity predicted by AGADIR[46] for its two α-helices. Obviously, the more local interactions, the higher the level of residual structure in the unfolded state and the lower the folding barrier.[9]

The rather special properties of gpW suggest a high degree of engineering in its amino acid sequence. Given that this protein is a natural single domain gene product, we can argue that the downhill folding of gpW is a result of selective pressure during evolution and, as such, is important for its biological function. It has been argued that the broad easily tunable conformational ensembles of downhill folding proteins could be used by nature as molecular control systems (e.g., rheostats) for complex biological processes that require regulation at the level of individual particles.[14,15] According to this hypothesis, one would expect biologically relevant downhill folding domains to carry

(40) Godoy-Ruiz, R.; Henry, E. R.; Kubelka, J.; Hofrichter, J.; Muñoz, V.; Sanchez-Ruiz, J. M.; EatonW. A. *J. Phys. Chem. B* **2008**, DOI: 10.1021/jp0757715.

(41) Onuchic, J. N.; LutheySchulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.

(42) Hagen, S. J. *Proteins: Struct., Funct., Bioinf.* **2007**, *68*, 205–217.

(43) De Sancho, D.; Rey, A. *J. Comput. Chem.* **2008**, DOI 10.1002/jcc.20924.

(44) Ejtehadi, M. R.; Avall, S. P.; Plotkin, S. S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15088–15093.

(45) Shimizu, S.; Chan, H. S. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 15–30.

(46) Muñoz, V.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 399–409.

out complex functions involving synchronization, coordination, or regulation of large molecular assemblies. This is the case for the global downhill folder BBL and seems to be the case for gpW as well. Although less well-known, genetic evidence indicate that this little protein from the λ-phage might be in charge of the connector assembly, channeling the genomic DNA through it during capsid packing and perhaps also in tailspike anchoring.[26] Hence, gpW emerges as another molecular rheostat candidate.

Acknowledgment. The authors thank Athi N. Naganathan for assistance in fitting the kinetic data and Alan Davidson for the original plasmid containing the gpW gene. This work was supported by NIH Grant RO1-GM066800 and Marie Curie Excellence Award MEXT-CT-2006-042334 to V.M. and Grants BIO2006-07332 from the Spanish Ministry of Science and Education, CVI-771 from the Junta de Andalucia, and Feder funds to J.M.S.R. R.G.R. has been partly supported by a HFSP postdoctoral fellowship.

JA801401A