Determining Denaturation Midpoints in Multiprobe Equilibrium Protein Folding Experiments†

Athi N. Naganathan and Victor Muñoz*

Department of Chemistry and Biochemistry and Center for Biomolecular Structure and Organization, University of Maryland, College Park, Maryland 20742, and Departamento de Ciencia de Proteínas, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maéztu 9, Madrid 28040, Spain

Received February 28, 2008; Revised Manuscript Received April 18, 2008

ABSTRACT: Multiprobe equilibrium unfolding experiments in the downhill regime (i.e., maximal barrier $< 3RT$) can resolve the folding process with atomic resolution [Muñoz (2002) *Int. J. Quantum Chem. 90*, 1522−1528]. Such information is extracted from hundreds of heterogeneous atomic equilibrium unfolding curves, which are characterized according to their denaturation midpoint (e.g., T_m for thermal denaturation). Using statistical methods, we analyze T_m accuracy when determined from the extremum of the derivative of the unfolding curve and from two-state fits under different sets of simulated experimental conditions. We develop simple procedures to discriminate between real unfolding heterogeneity at the atomic level and experimental uncertainty in the single T_m of conventional two-state folding. We apply these procedures to the recently published multiprobe NMR experiments of BBL [Sadqi et al. (2006) *Nature 442*, 317−321] and conclude that for the 122 single transition atomic unfolding curves reported for this protein the mean T_m accuracy is better than 1.8 K for both methods, compared to the 60 K spread in T_m determined experimentally. Importantly, we also find that when the pre- or posttransition baseline is incomplete, the two-state fits systematically drift the estimated T_m value toward the center of the experimental range. Therefore, the reported 60 K T_m spread in BBL is in fact a lower limit. The derivative method is significantly less sensitive to this problem and thus is a better choice for multiprobe experiments with a broad T_m distribution. The results we obtain in this work lay the foundations for the quantitative analysis of future multiprobe unfolding experiments in fast-folding proteins.

Two-state folding results from a free energy surface exhibiting only two major basins of attraction separated by a high free energy barrier. Protein molecules interconvert between the two basins crossing the barrier with activated kinetics, a process that is much slower than the reconfiguration dynamics on either basin (*3*). According to theory, folding barriers arise from the nonsynchronous compensation between conformational entropy and stabilization energy along the folding reaction (*4*). This idea, which has recently found strong empirical support (*5*), implies a maximal folding barrier at the denaturation midpoint. It also leads to the possibility of downhill folding (*6*). In downhill folding conformational entropy and stabilization energy go hand in hand producing a barrierless free energy surface. Downhill folding (or unfolding) is more likely when strongly native or denaturing conditions tilt the free energy surface toward the folded or unfolded basin of attraction (*6*). Further tilting of the free energy surface can be achieved through specifically engineered mutations (*7, 8*) and/or the addition of stabilizing cosolvents (*9*).

† This work has been supported by NIH Grant RO1-GM066800, NSF Grant MCB-0317294, and Marie Curie Excellence Award MEXT-CT-2006-042334.

* Correspondence should be addressed to this author. Phone: 34-91-837-3112, ext 4389. Fax: 34-91-536-0432. E-mail: vmunoz@cib.csic.es or vmunoz@umd.edu.

Table 1: Thermodynamic Parameters for the Simulated Experimental Scenarios[a]

exptl scenario	T_m (K)	ΔH_m (kJ/mol)	temp range (K)	temp interval (K)
broad-short	300	90	268–338	5
broad-long	320	90	268–368	5
sharp-long	320	120	268–368	5

[a] $\Delta C_p = 0$ for all of these calculations.

Proteins that (un)fold downhill under strong native or denaturing bias are also expected to have marginal barriers ($<3RT$) at the denaturation midpoint (*10*). Such marginal barriers should result in significant population of the barrier top conformations and, thus, in signature equilibrium behavior (*10*). When the maximal barrier (i.e., at the midpoint) is below thermal energy (RT), the protein is said to fold globally downhill, or in a one-state fashion (*11*). In this case the free energy surface has a single minimum at all conditions, resulting in a unimodal probability distribution that shifts from high order to disorder with increasing denaturational stress (*1, 11, 12*). The important implication of folding over marginal or zero barriers at the midpoint is that, in principle, the entire unfolding process can be resolved in multiprobe equilibrium experiments (*1, 12*). This is the case when the unfolding process is structurally nonhomogeneous, thus resulting in probe-dependent behavior. The situation is the inverse to the classical premise in the probe-independence tests for two-state folding developed in the

10.1021/bi800336x CCC: $40.75
Published on Web 06/10/2008

1960s (*13, 14*). As it was thoroughly discussed then (*14*), the test hinges on using probes that monitor as many and diverse structural properties of the native state as possible. The argument has been recently exploited to expose the global downhill behavior of the small protein BBL by probing the local unfolding environment of 158 of its protons using NMR[1] (*2*).

The interpretation and analysis of multiprobe protein unfolding experiments require statistical mechanics concepts, as it was well appreciated in the seminal work in this area (*13, 14*). When (and only when) the unfolding process approaches a true "two-state" scenario, the interpretation of the unfolding curves can be streamlined, resulting in the conventional two-state analysis widely used by protein folding experimentalists (*15*). In the conventional two-state analysis of thermal unfolding experiments the pre- and posttransition data define the structural properties of native and unfolded states (baselines), which are extrapolated to the transition region to obtain the single point at which the native:unfolded ratio is 50:50 (two-state T_m) whereas the width of the transition region provides the global change in enthalpy upon unfolding. For marginal barriers at T_m ($<3RT$) such analysis is inappropriate because the native and unfolded ensembles are not constant, but move structurally closer together the lower the T_m barrier (and merging when the T_m barrier is below RT) (*16*). Therefore, pre- and posttransition data also reflect structural changes, and there is neither a single T_m value nor a single width for the unfolding transition. The T_m values of multiple atomic probes simply define the points at which the structural environment around each particular atom is halfway unfolded. The average T_m for an extensive collection of atomic probes corresponds to a global denaturation midpoint (in principle comparable to the single two-state T_m), and the width of the T_m distribution reflects the degree of thermodynamic coupling in the protein (folding cooperativity) (*2*).

From this discussion it becomes obvious that T_m values provide the simplest means to quantitatively analyze multiprobe equilibrium thermal unfolding data in a noncommittal way. The important point is how to accurately determine T_m values of multiple unfolding curves likely to be broad and highly heterogeneous. It is also convenient to develop strategies to estimate T_m accuracy and to calculate the level of incompliance of a given multiprobe unfolding data set with a two-state scenario. There are two basic procedures to obtain the T_m of a thermal unfolding curve: (i) least-squares fitting to a two-state model (adding additional states if more than one apparent transition is observed); (ii) calculation of the extremum(a) in the derivative of the unfolding curve. The latter is equivalent to the two-equal area test originally developed by Hill (*13*). In the recent atom-by-atom analysis of BBL both methods were employed with consistent results (*2, 17*).

Here we investigate the accuracy of these two methods for determining T_m and develop simple procedures to assess the level of compliance of multiprobe data sets with a two-state scenario given specific levels of experimental uncertainty. The extension of these methods to multiprobe equilibrium experiments that use other denaturing agents (e.g., chemical denaturation) is straightforward. The outline of the paper is as follows. In the first section we use extensive numerical simulations of equilibrium unfolding data to evaluate the performance of T_m determination methods. From this analysis we derive simple procedures to estimate the overall T_m accuracy of a multiprobe unfolding data set. In the second section we compare the results expected for multiprobe unfolding experiments in the two extreme folding scenarios (two-state and global downhill) under identical "experimental" conditions. In the third and final section we utilize the concepts and results of sections 1 and 2 to revisit the original experimental results in BBL (*2*).

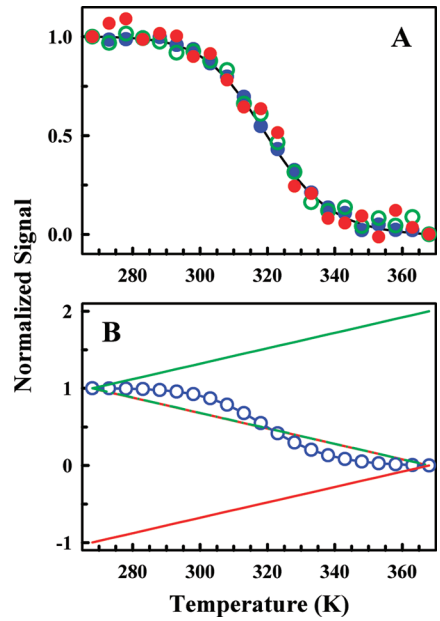

FIGURE 1: Examples of unfolding curves, different levels of noise, and baseline sloping for the broad-long scenario. (A) Different levels of noise: 0.015 (filled blue), 0.03 (open green), and 0.045 (filled red). (B) Range of pre- and posttransition baselines.

METHODS

Numerical Simulation of Thermal Unfolding Curves. Three experimental scenarios with different ratios between dynamic (i.e., temperature) range and transition broadness were investigated. The scenarios correspond to the parameters shown in Table 1. The effect of noise was simulated by generating random numbers according to a Gaussian distribution with standard deviation varying from 0 to 0.1 of the total amplitude of the unfolding curve in steps of 0.005. Slopes for folded and unfolded baselines were generated using randomly generated numbers (flat distribution) with values between 0 and 1% of the total temperature range in steps of 0.1% (see Figure 1). Noise and slope values were added onto the curves generated with the aforementioned parameters. For every noise and sloping level, 100 noise vectors and slope values were generated to produce 100 different unfolding curves. The T_m accuracy for a given level of noise and sloping was determined by calculating the standard deviation of the difference between the true value and the values obtained for the 100 simulated curves.

[1] Abbreviations: T_m, thermal denaturation midpoint; ΔH_m, two-state equilibrium unfolding enthalpy at the thermal denaturation midpoint; SLS, sum of least squares; NMR, nuclear magnetic resonance; σ_{T_m}, standard error in T_m.

T_m Determination with Phenomenological Fits. T_m values were obtained from least-squares minimization to a two-state model. Fully unconstrained fits include six adjustable parameters: T_m, ΔH_m, and linear ($ax + b$) baselines for folded and unfolded states. Bound fits include the same floating parameters but with the slope of the baselines bound to a maximum of 1% of the temperature range. Constrained fits set the slope for the two baselines to zero, thus including only four adjustable parameters.

T_m Determination with Derivative Method. The T_m is defined as the temperature at which the derivative of the unfolding curve reaches its extremum or the point dividing the derivative curve in two segments of equal area. Practically, the T_m value is obtained numerically from the noisy, sloped, and typically incomplete unfolding curves with the following procedure: (1) The first derivative of the synthetic equilibrium unfolding curves is obtained numerically with standard methods. (2) The derivative vector is then linearly interpolated to a high-density mesh with 0.25 K spacing and smoothed with the robust lowest algorithm. (3) Another vector corresponding to the ratio between the local areas on the left (low temperature, A_L) and right (high temperature, A_H) sides of each point in the interpolated derivative curve is calculated. A_L and A_H are obtained for each point integrating over a window of specific width in kelvin. The optimal window width depends on the broadness of the unfolding transition and the position of the denaturation midpoint relative to the temperature range in the experiment. (4) The T_m then corresponds to the point in the A_L/A_H vector that is closest to 1 (see Figure 2).

The procedure is mathematically straightforward and robust. For the calculations with varying degrees of noise and baseline sloping a generic scanning width of 19 K was used for the three experimental regimes. A narrower width of 10 K was used for the analysis of two-state and global downhill folding scenarios. In this case, the narrower width permitted the determination of T_m values that were very close to the edges of the temperature range for the global downhill folding scenario.

Analysis of Two-State and Global Downhill Folding Scenarios. A two-state scenario corresponding to the experimental conditions in Naf-BBL was simulated, generating 300 replicas of a thermal unfolding curve with $T_m = 300$ K, $\Delta H_m = 90$ kJ/mol, and with 15 points between 268 and 338 K every 5 K. The global downhill folding scenario was simulated, generating 300 synthetic unfolding curves with the same dynamic range, a random (flat) distribution of temperature midpoints in the range 273−333 K, and a random (flat) distribution of ΔH_m values in the range 30−180 kJ/mol. Folded and unfolded baselines with a random distribution of slopes varying from 0 to 0.3% of the temperature range were added to both sets of 300 curves. Gaussian noise with standard deviation ranging from 0.001 to 0.025 relative to the total amplitude of the curve was added to 285 curves in each set (95% of the total). Gaussian noise with standard deviations between 0.025 and 0.05 relative to the total amplitude of the curve was added to the remaining 15 curves in each scenario. The range and specific distribution of noise and sloping were chosen to closely mimic the distribution of sums of least squares and Δslope in the BBL experimental data (as shown in Figure 7A).

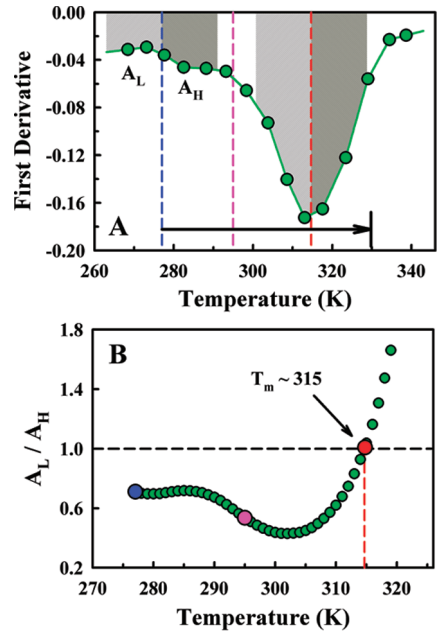

FIGURE 2: (A) The smoothed derivative (green circles) of the equilibrium thermal unfolding curve of Hδ2 from leucine 10 in Naf-BBL (*2*) (green circles in Figure 7B) along with the high-density mesh derivative curve obtained by linear interpolation within experimental data points and linear extrapolation to 5 K outside the highest and lowest experimental temperatures (green curve). In the first step of the extremum calculation a temperature window width (w) is chosen (14 K here). The window width determines the local areas at lower (A_L, light gray shade) and higher temperatures (A_H, dark gray shade) for each point in the high-density derivative curve and, thus, the temperature range accessible to T_m determination (black horizontal arrow). Typically, w should be the narrowest value that is still broader than noise fluctuations in the data. The second step consists in integrating the local areas A_L and A_H for each point in the high-density derivative curve within the accessible range. Dashed lines show three examples: lowest temperature accessible with $w = 14$ K (blue); intermediate temperature with A_L and A_H not shown to avoid graphical overlap (pink); temperature corresponding to the extremum (red). (B) The third step consists in calculating the A_L/A_H ratio for each point of the high-density derivative curve within accessible range (green circles in panel B). The extremum is defined as the point with the A_L/A_H value closest to 1 (314.7 K for this curve, red circle). The three nongreen circles show the A_L/A_H values for the three examples of panel A.

RESULTS AND DISCUSSION

(1) Accuracy in T_m Determination. In this section we analyze the accuracy of different methods for T_m determination under different levels of noise, baseline sloping, and dynamic range. We are interested in multiprobe unfolding experiments for proteins likely to fold by crossing marginal free energy barriers. As an example of very broad unfolding transition with limited dynamic range we chose the experimental conditions of the NMR unfolding experiments in Naf-BBL at pH 5.3 with 70 K dynamic range and overall transition broadness of ∼60 K (*2*). This simulated experiment is termed the *broad-short* scenario (from broad transition and short dynamic range). We also simulate a similarly broad unfolding process but with larger dynamic range (the *broad-long* scenario). As a third case, we simulate an experiment with larger dynamic range and increased sharpness (∼45 K span; the *sharp-long* scenario). The latter is similar to the experimental conditions encountered in various BBL variants at pH 7.0 and moderate ionic strength (*11, 12, 18, 19*). The

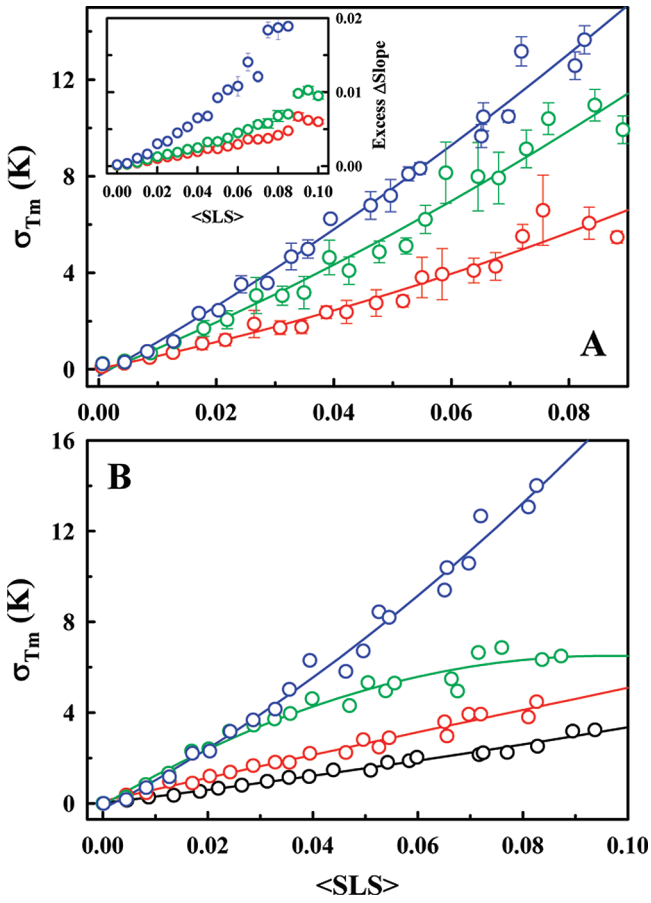

FIGURE 3: (A) Plot of T_m accuracy versus the average SLS per point from two-state fits (⟨SLS⟩) for the three scenarios: broad-short (blue), broad-long (green), and sharp-long (red). The error bars correspond to the variation in σ_{T_m} for the different baseline slopes ranging from 0 to 1% per K. The lines are fits to a quadratic equation. Note that 0.1 noise corresponds to a ⟨SLS⟩ of 0.083 for the broad-short scenario. (Inset) Plot of excess Δslope versus ⟨SLS⟩. (B) Performance of different methods for T_m determination as a function of the ⟨SLS⟩ from two-state fits for the broad-short scenario: free-floating two-state fit (blue); two-state fit with baseline slopes bound to within 1% of the curve amplitude per K (green); derivative method (red); two-state fit with fixed baseline slopes (black). For the derivative method the ⟨SLS⟩ from free-floating two-state fits was used in the plot to facilitate direct comparison.

broad-long scenario together with the range of baseline sloping and examples of noise levels is shown in Figure 1 for illustration. The thermodynamic parameters employed to simulate the three scenarios can be found in Table 1.

We determine T_m accuracy calculating the standard deviation of the difference between estimated and true T_m for 100 curves with random sets of the same noise level (see Methods). This procedure provides a direct estimate of T_m accuracy rather than using the covariance matrix from the fit (i.e., fitting error). Furthermore, this procedure can be applied to the derivative method, which does not involve a fitting routine. To compare results, we plot T_m accuracy against the average of the sum of least squares per data point obtained from the 100 curves (⟨SLS⟩). Although the noise level provides a more direct comparison, using ⟨SLS⟩ is convenient because it allows extending these results to real experimental data for which a direct determination of the noise level is not straightforward. This choice does not alter the results significantly because there is a close relationship between the noise level and the ⟨SLS⟩ of the fits.

As a first step we investigate the performance of conventional two-state fits for the three experimental scenarios. Figure 3A shows that T_m uncertainty (σ_{T_m}) increases quadratically with the noise level (a relative noise level of 0.1 corresponds to ⟨SLS⟩ ∼ 0.083). The accuracy of two-state fits does not depend on the degree of baseline sloping in any of the scenarios. This is expected because the fits include sloped folded and unfolded baselines as fitting parameters. Accordingly, in Figure 3A we plot the average and standard deviation of the T_m accuracy for all tested degrees of baseline sloping. The scatter is due to statistical noise from the limited number of samples (100). An interesting observation is that T_m accuracy is always better than the estimate from the covariance matrix. When the ⟨SLS⟩ is low, the overestimation of the error from the covariance (fitting errors) is not too large (∼20%), but it increases sharply as ⟨SLS⟩ increases. For ⟨SLS⟩ of 0.05 the average error overestimation for the broad-short scenario is 2-fold. Fitting errors overestimate the uncertainty because they are obtained by assuming that the six fitted parameters (T_m, ΔH_m, and two linear baselines) have independent errors (ignoring cross-terms). Such assumption is clearly inadequate for these fits in which variations in T_m and ΔH_m are effectively compensated by changes in the baselines (*17*).

As expected, Figure 3A also shows that T_m values from two-state fits are more accurate the more pre- and posttransition data is included. However, this effect is monotonic. For example, a ⟨SLS⟩ of 0.04 results in σ_{T_m} of 6, 4, and 2 K for the broad-short, broad-long, and sharp-long scenarios, respectively. In fact, Figure 3A indicates that decreasing the noise level in the unfolding curves is a more practical strategy to improve T_m accuracy than expanding the dynamic range.

Another important result is related to the effect of free-floating baselines. To investigate this issue, we introduce a parameter that we term Δslope: the absolute value of the difference between the slope of the native state baseline and that of the unfolded state baseline ($\Delta\text{slope} = |N_{slope} - U_{slope}|$). It is a measure of the degree of orthogonality between the two baselines and provides an estimate of baseline trimming effects on T_m determination (*17*). The baseline slopes do not affect T_m as long as the baselines remain approximately parallel, but when the two baselines cross, the two-state estimation of T_m can differ significantly from the point at which the signal is halfway (see below). For multiprobe experiments such effects can be detected in a plot of the excess in Δslope versus ⟨SLS⟩ (inset in Figure 3A). This plot shows that T_m uncertainty is directly correlated with the excess Δslope (i.e., Δslope of the fit minus original Δslope) introduced by the two-state fits of the noisy data. We will come back to this issue in section 2, where we address the interpretation of global fits to a two-state model.

In Figure 3B we compare different procedures for T_m determination. For simplicity, we eliminate baseline sloping effects and focus on the broad-short experiment, which can be considered a worst case scenario. The data shown in blue correspond to the two-state fits with free-floating baselines. The green curve and symbols show the accuracy of two-state fits in which baseline slopes have been bound not to exceed 1% of the total amplitude of the curve per kelvin. Bound and free baseline fits produce the same results when ⟨SLS⟩ is low, because the excess Δslope is always below the limits imposed by the fit. At ⟨SLS⟩ > 0.025 the bound fits hit the limits imposed on the baseline slopes, and the

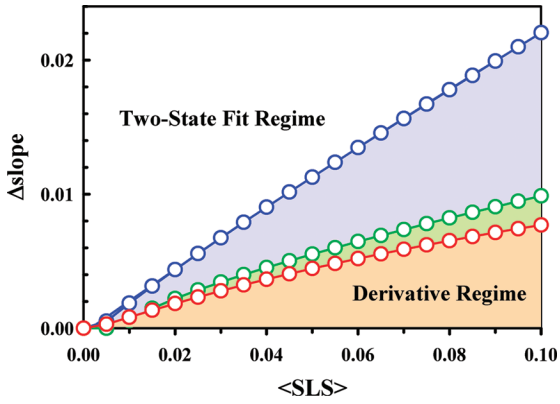

FIGURE 4: Plot of Δslope versus ⟨SLS⟩ showing the regions in which the two-state fit and the derivative method provide more accurate T_m values for the three scenarios: broad-short scenario (blue); broad-long (green); sharp-long (red). The circled lines correspond to the intercept line between T_m accuracy for both methods (i.e., the line of equal accuracy) for each particular scenario. The colored area corresponds to the regime where the derivative method is more accurate with the red area corresponding to the sharp-long scenario, red + green corresponding to the broad-long scenario, and red + green + blue corresponding to the broad-short scenario.

curve flattens out. This result demonstrates the direct link between baseline crossing and T_m accuracy in two-state fits. The noisier the data, the more pronounced the effect. However, these results also indicate that it is possible to obtain reasonably accurate T_m values of noisy and broad unfolding curves with prior knowledge of the expected baseline slopes, as was done for BBL (*2*). By the same token, the T_m accuracy of noisy broad curves increases greatly if the baseline slopes can be completely restricted, resulting in reasonably accurate estimates even at 10% noise levels (black curve in Figure 3B).

Figure 3B also shows the performance of the derivative method. Although the derivative is straightforwardly obtained numerically, it is not trivial to determine its extremum when the unfolding curves are noisy, broad, or even incomplete. The procedure that we use here and in ref *2*, and which is illustrated in Figure 2 with a real experimental example, is halfway between determining the extremum directly and dividing the entire derivative curve in two segments of equal area. This procedure is expected to exhibit low sensitivity to noise and can be applied to incomplete unfolding transitions as long as the midpoint is within the experimental range. The results shown in Figure 3B (red circles) confirm the very low sensitivity to the noise level of this method, in stark contrast with recent claims (*20, 21*). The derivative method is much less sensitive to noise than the two-state fits because it does not require baseline extrapolation.

Ignoring pre- and posttransition slopes, however, may be problematic for experiments in which the spectroscopic signal changes with the denaturing agent. These changes do not reflect structural rearrangements in the protein but contribute to the position of the extremum in the derivative. For two-state folding they would need to be corrected by baseline extrapolation, whereas for one-state folding these changes are intermingled with the continuous unfolding of the protein. To investigate combined noise and baseline sloping effects on the T_m determination of *true two-state* unfolding curves with the derivative method, we looked again at every combination of Δslope and noise level in the three scenarios. These calculations show that even for true two-state unfolding curves there is a range of conditions for which T_m values from the derivative method are still more accurate than those from free-baseline floating two-state fits. This regime corresponds to conditions in which ⟨SLS⟩ is relatively large and Δslope is relatively small and varies depending on the experimental scenario. The summary of these results is given in Figure 4, in which the colored area signifies the range of conditions in which the derivative method is more precise than the fit for true two-state processes. The plot in Figure 4 can be used to determine which method to favor for any given combination of noise and baseline sloping and, of course, assuming that the protein folds two state. For points that lie above the line two-state fits are preferable, whereas for points below the line the derivative method is more accurate. Figure 4 clearly shows that extending the dynamic range preferentially favors the two-state fit method.

(2) Comparison between Two-State and Global Downhill Folding Scenarios. In this section we simulate multiprobe unfolding experiments for two-state and global downhill folding scenarios mimicking the experimental conditions of the BBL NMR experiments at pH 5.3 (broad-short scenario). The two-state multiprobe experiment is simulated with 300 replicas of the same unfolding curve (i.e., T_m = 300 K, ΔH_m = 90 kJ/mol) with random Gaussian noise and native and unfolded baselines added to agree with those of the original BBL data set (*2*) (see Figure 7A). Therefore, this two-state scenario simulates the results expected for the NMR experiments on BBL if the folding of this protein were two state. Figure 5A shows a representative sample of the 300 "two-state" curves normalized to facilitate visual comparison. Although all of these curves correspond to the same unique unfolding behavior, the presence of noise and pre- and posttransition baselines produces some spread. The T_m distribution obtained with conventional two-state fits has the exact mean (300 K) and very low standard deviation (1.35 K) (Figure 5B), but it is clearly non-Gaussian, exhibiting a narrow peak in the center flanked by shallow wings. Its origin is the nonuniform noise distribution of the original BBL data (88% of the data with low noise and the remainder with higher noise; see Figure 7A), which also results in slightly better T_m accuracy (1.35 K) than estimated assuming a uniform noise distribution (i.e., 1.8 K when the 0.014 ⟨SLS⟩ of the experimental BBL data set is plotted on the blue curve of Figure 3A). The derivative method produces a slightly broader, almost Gaussian T_m distribution that corresponds to an accuracy of 1.5 K. The agreement between T_m values from fits and derivative is very good, with a median discrepancy of 0.8 K (Figure 5D). There is a tail in the histogram corresponding to curves with either very high noise or intrinsic Δslope. These results indicate that a two-state scenario comparable to the multiprobe NMR experiments in BBL should result in a very narrow T_m distribution for both two-state fits and the derivative method.

The fitting of this two-state scenario to a global two-state model (single T_m, single ΔH_m, and 300 sets of native and unfolded baselines with linear temperature dependence) is qualitatively very good and only slightly worse than the individual fits (i.e., 11% higher SLS). To determine quantitatively whether the global fit supports a true two-state process, we carried out a standard statistical *F*-test. The *F*-test compares the ratio between fitting residuals with the ratio between degrees of freedom (number of data points minus

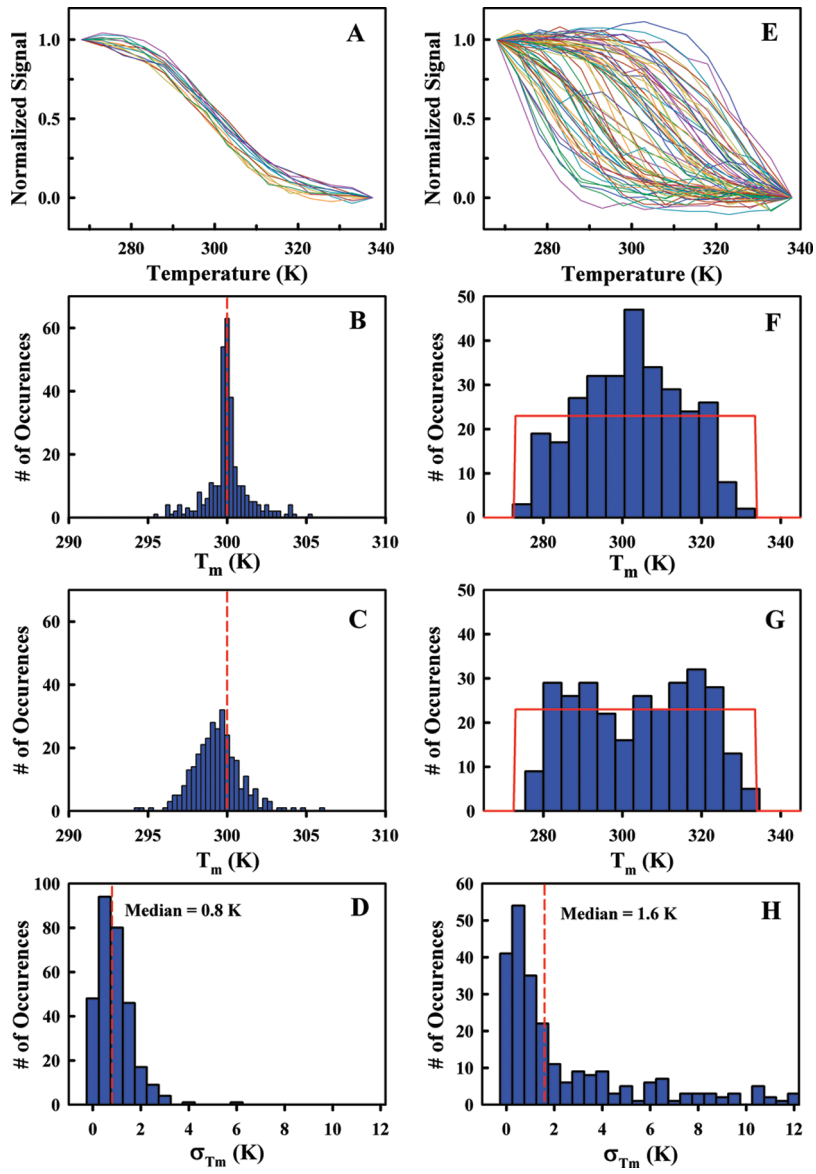

FIGURE 5: Two-state versus global downhill folding scenarios. (A) Representative examples of the 300 synthetic unfolding curves for the two-state scenario. Noise and baseline sloping were chosen to mimic those in the Naf-BBL experimental data (Figure 7A). (B) T_m distribution from two-state fits to the two-state scenario. (C) T_m distribution from the derivative method applied to the two-state scenario. In both cases the red dashed line signals the true value. (D) Histogram of the standard deviation in T_m values between two-state fits and the derivative method for the two-state scenario. The red dashed line signals the median. (E−H) As in (A)−(D) but for the global downhill scenario. The red lines in panels F and G signal the parent T_m distribution used to generate the 300 curves of the global downhill scenario.

number of fitting parameters) of global and individual fits. The output of the *F*-test gives the probability that data originating from the statistically simpler model (in this case the global two-state model) fit the complex model that much better. For the two-state scenario of Figure 5A the *F*-test gives a probability of 99.91% that the data originate from a single two-state transition. Thus, the *F*-test effectively discriminates that the spread in Figure 5A is caused by noise.

The physical relevance of the global fit can be analyzed using the recently introduced ΔΔslope (Δslope of global fit minus Δslope individual fit) versus ΔT_m (midpoint from individual fit minus global midpoint) plot (*17*). The plot is based on the direct connection between excess Δslope and the deviations in T_m introduced by individual two-state fits of noisy data (see inset of Figure 3A). The global two-state fit to a true two-state transition should correct such baseline deviations by enforcing unique T_m and ΔH_m for all the curves, since the data are likely to have uncorrelated noise. Thus, all data points should appear on the diagonal line of the plot with ΔT_m spread equivalent to the T_m accuracy estimated from the average noise (or the ⟨SLS⟩ from individual fits). Plotting the results from the two-state scenario, we obtain such pattern exactly (Figure 6A). The dotted red lines signal the 95% confidence interval for the T_m estimated from the ⟨SLS⟩ of individual fits, which agrees quite closely with the data. The results we show here for the two-state scenario validate the ΔΔslope versus ΔT_m plot as a powerful test for global two-state fits.

To simulate the global downhill folding scenario, we calculate 300 unfolding curves with the same dynamic range, levels of noise, and baseline sloping but with random T_m (from 273 to 333 K) and ΔH_m (from 30 to 180 kJ/mol). A selection of the 300 curves is shown normalized in Figure 5E. This distribution of unfolding behaviors is very similar to the experimental one reported for BBL (*2*) and drastically different to the simulated two-state scenario (Figure 5A). The

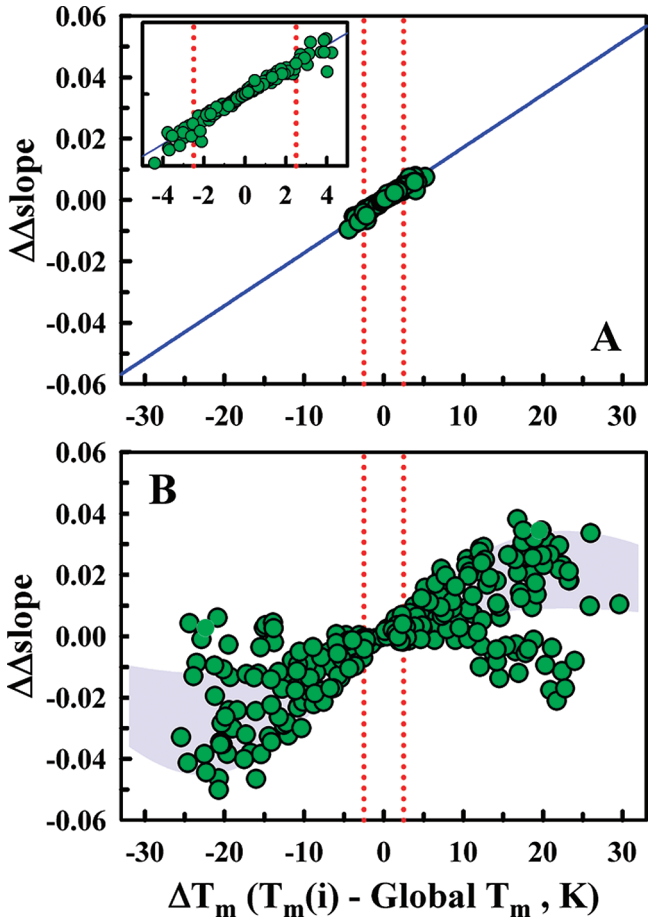

FIGURE 6: (A) ΔΔslope versus ΔT_m plot for the two-state scenario. The dotted red lines delimit the 95% confidence interval in the determination of the midpoint temperature based on the average ⟨SLS⟩ from the fits. The blue line corresponds to the linear regression of the data. (Inset) A blowout of the region within ±5 K. (B) As in (A) but for the global downhill scenario. The gray area shows the bowtie distribution for the global downhill scenario in the absence of experimental noise.

parent T_m distribution for the global downhill scenario is flat (random values) with mean = 304.2 K and full spread (Figure 5F). However, although the distribution obtained from individual fits has similar mean (302.4 K), it is significantly narrower and approximately Gaussian (Figure 5F). What happens here is that two-state fits are much less accurate when the T_m is near the edges of the experimental dynamic range. Furthermore, the two-state fits do not distribute the error uniformly on both sides of the true T_m. Instead, the least-squares minimization procedure shifts the estimated T_m toward the center of the experiment. This effect is apparent in Figure 5F, which shows a migration from the edges to the center of the distribution. In other words, for a global downhill folding scenario the T_m distribution from conventional two-state fits provides a lower bound to the true spread of the parent T_m distribution. The importance of this result cannot be overemphasized, since it is common to assume that the T_m error due to a poorly resolved baseline is distributed uniformly on both sides of the true T_m (*20, 21*).

For the downhill folding scenario the derivative method produces a T_m distribution closer to the parent one (Figure 5G), although it is also slightly depopulated at the edges. The edges correspond to curves in which the T_m is just outside the sensitivity range of the method (i.e., there is no extremum in the derivative). The comparison between the T_m values from two-state fits and the derivative method shows a median discrepancy of ~1.6 K with a long tail (Figure 5H). The median increase relative to the two-state scenario is mostly due to the larger tail, which shows discrepancies of up to 12 K. The tail corresponds to some curves with T_m near the edges and/or large noise levels. Interestingly, the discrepancy between the two methods is mostly caused by the two-state fits, whereas the derivative method is in closer agreement with the true T_m values. Therefore, for multiprobe unfolding experiments with the expectation of large T_m spreads the derivative method is more accurate but still provides a lower bound of the real spread.

The results from fitting of the synthetic global downhill scenario to the global two-state model are summarized in the ΔΔslope versus ΔT_m plot shown in Figure 6B. A striking result of this exercise is that in spite of the large intrinsic spread of this data set (both in T_m and ΔH_m), the fit looks quite reasonable. The curve fits follow the major trends in the data, and only a few curves (typically with very extreme T_m and large noise) exhibit qualitatively bad fits. This result is very similar to the one previously published for 300 synthetic curves with unique ΔH_m and large T_m spread (*17*). The ability of the global two-state model to apparently fit the global downhill scenario resides on the trimming power of free-floating baselines, which can trace baselines through the unfolding transition, thus making one of the true baselines the apparent transition (see Figure 1 in ref *17* for a graphical demonstration). However, the validity of the fit can be effectively discarded using a quantitative statistical test. For instance, the ⟨SLS⟩ of the global fit is 3.8-fold larger than that of the individual fits, which the *F*-test interprets as a probability below 1 in 10000 that the data originate from a noisy two-state transition.

Furthermore, the plot of ΔΔslope versus ΔT_m provides clear indication that the global two-state fit is unphysical. In contrast with the two-state scenario (Figure 6A), the plot for the global downhill scenario displays a large spread with a characteristic bowtie distribution (Figure 6B) (*17*). The range in T_m discrepancy (ΔT_m) extends for ~60 K and is over an order of magnitude larger than the one expected from ⟨SLS⟩ of individual fits to the two-state scenario (vertical dotted red lines in Figure 6B). The bowtie shape is due to the combination of uncorrelated changes in T_m and ΔH_m together with poor definition in one of the baselines for curves with extreme T_m. The gray shade in the figure reflects the spread expected for the same data set without any noise (*17*). Points outside the gray area are then the result of a large noise level. Of particular interest are the outsiders that lie near the horizontal line in the plot (i.e., a ΔΔslope = 0). These outsiders constitute about 10% of the total data set and correspond to noisy curves with T_m near the edges in which the baselines drifted the fitted T_m toward the center of the experiment in both the individual and global fits. In summary, the *F*-test and ΔΔslope vs ΔT_m plot are effective tools to differentiate between global downhill and two-state scenarios even in relatively noisy data sets with broad unfolding and narrow dynamic range.

(3) Analysis of Multiprobe Atomic Resolution Experiments on the Protein BBL. The results obtained in the previous two sections give us the opportunity to revisit the multiprobe NMR data on BBL (*2*). For simplicity, we focus on the 122 (out of

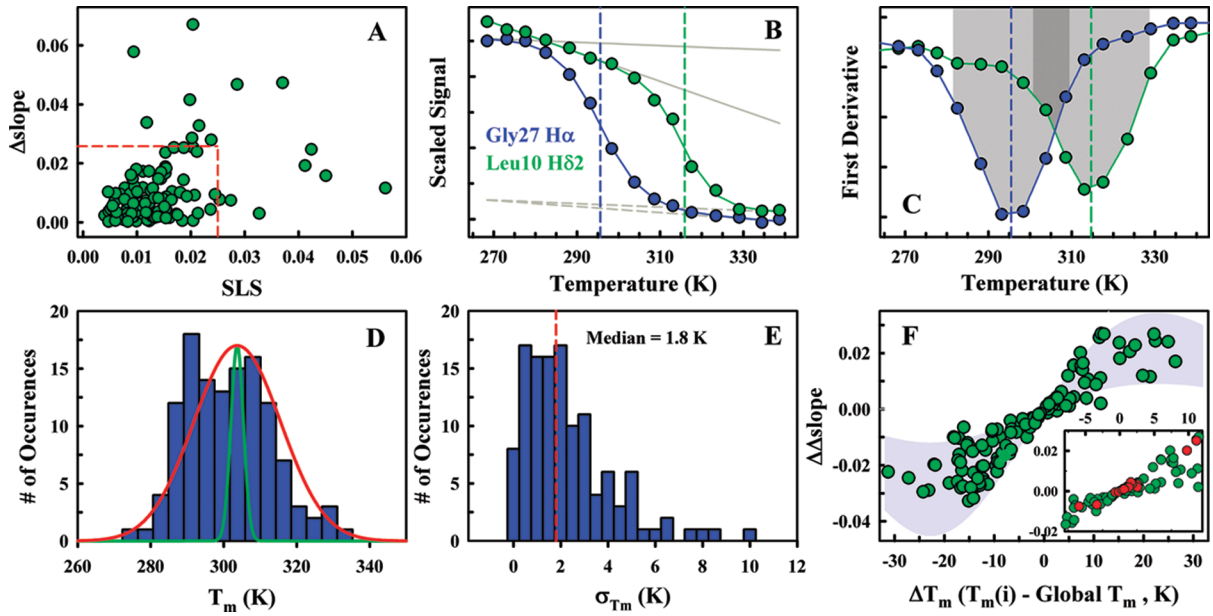

FIGURE 7: Analysis of the 122 single-transition atomic unfolding curves for Naf-BBL (*2*). (A) Distribution of Δslope and SLS from two-state fits for the Naf-BBL single-transition data. (B) Two examples of Naf-BBL curves with well-resolved baselines and very different T_m (T_m values shown as vertical dashed lines). The curves correspond to Hα from glycine 27 and Hδ2 from leucine 10 in Naf-BBL. Fits to a two-state model are shown as continuous lines, and the fitted baselines are in gray. The T_m fitting error for these two curves is less than 1 K. (C) Derivative of the unfolding curves in (B). Vertical lines signal the T_m obtained with the derivative method, and the light gray area shows the sum of the two local areas ($A_L + A_H$) at the midpoint as used in the analysis. (D) T_m distribution from two-state fits of the 122 single-transition curves (blue bars), along with a Gaussian fit to the original distribution shown in Figure 2B of ref *2* (red curve). The green curve shows the broadening of a single T_m value that would be expected from the T_m accuracy estimated from the average SLS of the Naf-BBL data. (D) Histogram of the standard deviation in T_m between the derivative method and the two-state fits for the 122 single-transition curves of Naf-BBL data. (E) Plot of ΔΔslope versus ΔT_m between individual two-state fits and the global two-state fit of the 122 Naf-BBL single-transition curves (filled green circles). The shaded area in the background shows the bowtie distribution expected for a theoretical global downhill scenario (from Figure 6B). (Inset) Blowout of the central region of the plot also displaying the data from Fersht and co-workers (*19*) as red circles.

158) proton unfolding curves that exhibit a single unfolding transition (i.e., the subset of two-state-like curves) (*17*). The other 36 more complex unfolding curves cannot be properly fit to a two-state model and thus are not amenable to the analyses described above. In principle, the analysis of this "pruned" multiprobe data set is a more stringent test of the global downhill character of BBL because we are omitting unfolding curves that are obviously incompatible with a two-state scenario. To compare noise levels and degree of baseline sloping for different unfolding curves, we normalize the signals. This is an essential step because the magnitude of the chemical shift and of its change upon unfolding could vary over an order of magnitude depending on the chemical nature of the proton. The noise level can also vary significantly for different protons, depending on the region of the NMR spectrum, the signal overlap, the presence of line broadening from chemical exchange, and the number of cross-peaks per signal that can be used for averaging the proton chemical shift. As we showed in the first section, T_m accuracy does not depend on the absolute levels of noise and baseline sloping but on their magnitude relative to the total change in signal upon unfolding. Here we used the simplest normalization procedure, which is to subtract the lowest signal value from the whole unfolding curve and then divide subtracted curve by its highest signal value. This procedure is mathematically straightforward and does not involve any *a priori* interpretation of the data.

Figure 7A shows the SLS versus Δslope plot for the individual two-state fits of the 122 normalized unfolding curves of Naf-BBL. The figure shows that the majority of curves have SLS below 0.025 and Δslope below 2%. In fact, the rectangle defined by the dashed red lines includes approximately 88% of the 122 curves. Therefore, the bulk of the single-transition curves of Naf-BBL is of quite high quality in terms of noise and has small differences between the slopes of the two baselines. There are six curves with SLS higher than 0.03 and nine curves with high Δslope (Figure 7A). The distribution of SLS and Δslope shown in Figure 7A was the template we used to generate the synthetic two-state and global downhill folding scenarios of section 2.

The ⟨SLS⟩ of the 122 unfolding curves is 0.014, which according to Figure 3A corresponds to an average T_m accuracy for two-state fits of ~1.8 K (blue curve in Figure 3A). This estimate assumes a uniform distribution of noise, but given the distribution shown in Figure 7A, we can conclude that the T_m accuracy is better for the majority of curves with a few in which it is significantly worse. An error of 1.8 K (68% confidence interval) can hardly explain the T_m spread obtained from the 122 two-state fits. As an illustration, Figure 7B shows 2 of the 122 curves, which have well-resolved pre- and posttransition baselines, low noise, and T_m values 21 K apart. From Figure 7B it is also clear that curves with one unresolved baseline are likely to have T_m values below 285 K or above 320 K. Figure 7C shows the derivative of the two curves with the T_m values estimated with this method and shaded regions signaling the two lobes of equal area used in determining the extrema. The complete T_m distribution spanning 60 K is shown in Figure 7D, which also shows in green a Gaussian curve signifying the width

corresponding to the estimated 1.8 K average accuracy. From this result we can conclude that the large T_m spread reported for BBL is definitely not caused by inaccuracies in T_m determination. It is also interesting to note that the T_m distribution for the 122 single-transition curves is actually broader than the one reported originally (*2*), which included the 36 × 2 T_m from the double-transition curves (red curve in Figure 7D). This is because the more poorly defined three-state fits (involve 10 parameters to reproduce a typically incomplete double unfolding transition) result in larger drifts of the fitted T_m values toward the center of the dynamic range. The comparison between T_m values from two-state fits and derivative results in a median discrepancy of 1.8 K and a histogram with a tail extending up to 10 K (Figure 7E) that is very similar to the one we obtained for the synthetic global downhill scenario.

Finally, we fitted the 122 BBL curves to a global two-state model. As we previously observed for the synthetic global downhill scenario, for most unfolding curves the fit seems reasonable by eye. However, the SLS is over 3 times larger than for the individual two-state fits. Indeed, the *F*-test indicates that the probability that the 122 Naf-BBL curves arise from a two-state process is statistically insignificant (<1 in 10000). Here is important to mention that this *F*-test result corresponds to the normalized 122 two-state-like curves of Naf-BBL. A similar result was obtained before for a fit to the nonnormalized 158 curves (*17*).

The ΔΔslope versus ΔT_m plot (Figure 7F) offers more evidence of the globally downhill folding of Naf-BBL. The spread in ΔT_m for the 122 unfolding curves spans 60 K, and the bowtie distribution is clearly evident. Most data points fall within the gray area that delimits the theoretical expectation for a noise-free global downhill scenario. The few outliers correspond to examples in which the individual fit also skewed the baselines because of high noise combined with T_m near one edge of the experiment. The inset in Figure 7F shows a blowout of the central region of the ΔΔslope versus ΔT_m plot of Naf-BBL, also including the 15 ^{13}C unfolding curves reported by Fersht and co-workers for another BBL variant in different experimental conditions (*19*). This is an interesting figure that highlights the sensitivity of the ΔΔslope versus ΔT_m plot. Out of all the possible ^{13}C signals in BBL (over 200), Fersht and co-workers reported these 15 to support their case for a single T_m in BBL (*18, 19*). Thus, it would be reasonable to assume that these 15 curves are among those that best agree with a two-state scenario. In agreement with this assumption, most of the points from the ^{13}C probes lie within the average T_m accuracy. Nevertheless, even in this reduced multiprobe data set there are some significant deviations from a single T_m (red circles at ∼10 and −6 K ΔT_m in Figure 6B). Finally, it is important to mention that the unfolding curves of protons adjacent to the 15 carbons reported by Fersht and co-workers are also very similar to one another and to the average folding behavior in Naf-BBL at pH 5.3 (see Figure 2d in ref *2*). The agreement between the two experiments strongly suggests that the larger data set obtained for Naf-BBL represents the general folding behavior of BBL, irrespective of experimental conditions and small chemical differences in the unstructured tails.

CONCLUSIONS

The combination of high-resolution NMR experiments with the recent discovery of proteins that fold by crossing marginal folding barriers (i.e., from near downhill to globally downhill) allows resolving their folding−unfolding process at atomic resolution using simple equilibrium unfolding experiments (*2*). This can be achieved by tracking the changes in NMR signals that are sensitive to the local structural environment of individual atoms (e.g., chemical shifts) in response to increasing levels of denaturational stress. A critical point of the analysis is to obtain precise denaturation midpoints of broad and heterogeneous unfolding curves (*10*).

In this work we have investigated the accuracy of various methods for determining T_m values under different levels of noise, baseline sloping, and experimental dynamic range. We have carried out these studies using the NMR multiprobe unfolding experiments in BBL (*2*) as a frame of reference. Our results indicate that the width of the T_m distribution previously reported for BBL is at least an order of magnitude larger than the average T_m uncertainty. We find that T_m estimates from two-state fits are rather accurate even for broad unfolding transitions with limited dynamic range, as long as the noise level is low relative to the total amplitude of the curve. This is especially true when the T_m is near the center of the unfolding curve, as expected for the multiprobe unfolding curves of two-state folding proteins. Not surprisingly, T_m estimates from two-state fits are significantly less accurate when the true T_m is near one of the edges of the curve due to poor tracing of the pre- or posttransition baseline. However, this situation, which already implies a true T_m far away from the two-state expectation, results in T_m estimates systematically drifted toward the center of the unfolding curve rather than in random error. Consequently, T_m estimates from two-state fits underestimate the width of a broad parent distribution such as that previously reported for Naf-BBL.

Another important observation is that the lower accuracy of the T_m estimates from two-state fits of noisy and/or incomplete unfolding curves can be greatly improved, constraining the baseline slopes during the fit. This requires prior knowledge of how the spectroscopic signal is expected to change in response to the denaturing agent. A straightforward example is to use short peptides as models of the unfolded (unstructured) baseline. An alternative is to determine the T_m from the extrema in the derivative of the unfolding curve. In our analysis we find that the derivative method is much less sensitive to noise and, most importantly, does not require baseline tracing. The derivative method can successfully determine the T_m for all curves in which the extremum lies within the experimental range. In practical terms this means that the derivative method can be used for T_m values extending up to 90% of the experimental dynamic range. On the other hand, two-state fits are more accurate, determining the T_m of sharper unfolding curves with little noise and/or with large degree of baseline sloping. In general, good agreement should be found between the two methods when the unfolding curves have reasonably defined baselines, low level of noise, and baseline sloping (for example, the area within the dashed rectangle in Figure 7A).

As a last piece in the puzzle, we have investigated the performance of global two-state fits in discriminating between two-state and global downhill folding scenarios. This is an important exercise because, although global two-state fits should in principle be stringent tests of the two-state compliance of multiprobe experiments, it has been recently shown that such fits can fit deceivingly well large data sets of synthetic unfolding curves with vastly different T_m values (*17*). We find here that a global two-state model can fit global downhill folding scenarios because the free-floating baselines trim the large inconsistencies with a single T_m. The discriminative value of the global two-state fit only emerges after quantitative analysis of the results, for example, by combining a simple statistical test (*F*-test) and a plot of the differences in the degree of baseline crossing versus the T_m differences between global and individual fits (e.g., $\Delta\Delta$slope versus ΔT_m plot). These two tests confirm Naf-BBL as a global downhill folding protein and emerge as a powerful tool for the characterization of folding regimes intermediate between global downhill and two-state folding (*10*).

REFERENCES

1. Muñoz, V. (2002) Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *Int. J. Quantum Chem. 90*, 1522–1528.
2. Sadqi, M., Fushman, D., and Muñoz, V. (2006) Atom-by-atom analysis of global downhill protein folding. *Nature 442*, 317–321.
3. Zwanzig, R. (1997) Two-state models of protein folding kinetics. *Proc. Natl. Acad. Sci. U.S.A. 94*, 148–150.
4. Onuchic, J. N., LutheySchulten, Z., and Wolynes, P. G. (1997) Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem. 48*, 545–600.
5. Akmal, A., and Muñoz, V. (2004) The nature of the free energy barriers to two-state folding. *Proteins: Struct., Funct., Bioinf. 57*, 142–152.
6. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein-folding—a synthesis. *Proteins: Struct., Funct., Genet. 21*, 167–195.
7. Yang, W. Y., and Gruebele, M. (2003) Folding at the speed limit. *Nature 423*, 193–197.
8. Yang, W. Y., and Gruebele, M. (2004) Rate-temperature relationships in lambda-repressor fragment lambda(6−85) folding. *Biochemistry 43*, 13018–13025.
9. Yang, W. Y., and Gruebele, M. (2004) Folding λ-repressor at its speed limit. *Biophys. J. 87*, 596–608.
10. Muñoz, V. (2007) Conformational dynamics and ensembles in protein folding. *Annu. Rev. Biophys. Biomol. Struct. 36*, 395–412.
11. Naganathan, A. N., Perez-Jimenez, R., Sanchez-Ruiz, J. M., and Muñoz, V. (2005) Robustness of downhill folding: Guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry 44*, 7435–7449.
12. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M., and Muñoz, V. (2002) Experimental identification of downhill protein folding. *Science 298*, 2191–2195.
13. Hill, T. (1963) *Thermodynamics of small systems. Part 1*, Chapter *5*, W. A. Benjamin, New York.
14. Lumry, R., Biltonen, R., and Brandts, J. F. (1966) Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers 4*, 917–944.
15. Aune, K. C., and Tanford, C. (1969) Thermodynamics of denaturation of lysozyme by guanidine hydrochloride. 2. Dependence on denaturant concentration at 25 degrees. *Biochemistry 8*, 4586–4590.
16. Naganathan, A. N., Doshi, U., and Muñoz, V. (2007) Protein folding kinetics: Barrier effects in chemical and thermal denaturation experiments. *J. Am. Chem. Soc. 129*, 5673–5682.
17. Sadqi, M., Fushman, D., and Munoz, V. (2007) Structural biology—Analysis of protein-folding cooperativity—Reply. *Nature 445*, E17–E18.
18. Ferguson, N., Schartau, P. J., Sharpe, T. D., Sato, S., and Fersht, A. R. (2004) One-state downhill versus conventional protein folding. *J. Mol. Biol. 344*, 295–301.
19. Ferguson, N., Sharpe, T. D., Schartau, P. J., Sato, S., Allen, M. D., Johnson, C. M., Rutherford, T. J., and Fersht, A. R. (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *J. Mol. Biol. 353*, 427–446.
20. Ferguson, N., Sharpe, T. D., Johnson, C. M., Schartau, P. J., and Fersht, A. R. (2007) Structural biology—Analysis of "downhill" protein folding. *Nature 445*, E14–E15.
21. Zhou, Z., and Bai, Y. W. (2007) Structural biology—Analysis of protein-folding cooperativity. *Nature 445*, E16–E17.

BI800336X