Better biomolecule thermodynamics from kinetics

Kiran Girdhar,[1] Gregory Scott,[2] Yann R. Chemla,[1,3] and Martin Gruebele[1,2,3,a)]
[1]*Center for Biophysics and Computational Biology, University of Illinois, Urbana, Illinois 61801, USA*
[2]*Department of Chemistry, University of Illinois, Urbana, Illinois 61801, USA*
[3]*Department Physics, University of Illinois, Urbana, Illinois 61801, USA*

(Received 27 March 2011; accepted 15 June 2011; published online 7 July 2011)

Protein stability is measured by denaturation: When solvent conditions are changed (e.g., temperature, denaturant concentration, or *p*H) the protein population switches between thermodynamic states. The resulting denaturation curves have baselines. If the baselines are steep, nonlinear, or incomplete, it becomes difficult to characterize protein denaturation. Baselines arise because the chromophore probing denaturation is sensitive to solvent conditions, or because the thermodynamic states evolve structurally when solvent conditions are changed, or because the barriers are very low (downhill folding). Kinetics can largely eliminate such baselines: Relaxation of chromophores, or within thermodynamic states, is much faster than the transition over activation barriers separating states. This separation of time scales disentangles population switching between states (desired signal) from chromophore or population relaxation within states (baselines). We derive simple formulas to extract unfolding thermodynamics from kinetics. The formulas are tested with model data and with a difficult experimental test case: the apparent two-state folder PI3K SH3 domain. Its melting temperature T_m can be extracted reliably by our "thermodynamics from kinetics approach," even when conventional fitting is unreliable. © *2011 American Institute of Physics*. [doi:10.1063/1.3607605]

I. INTRODUCTION

Protein or nucleic acid denaturation is often approximated by a series of transitions between thermodynamic states.[1–4] In addition to population moving among states, the states and the chromophores probing the transitions are sensitive to solvent conditions.[5] Thus baselines arise even when there is no population exchange between states.

These baselines are fitted traditionally by linear functions, adequate if the signal used to probe folding thermodynamics varies slowly along the reaction coordinate.[6,7] If not, the baseline can become difficult to separate from the actual folding transition. This is particularly true for small proteins, where the unfolding transition may be broad due to finite-size effects. Schuler and Eaton measured the unfolded state baseline under solvent conditions favoring the native state, showing that the baseline is far from linear.[8]

Even linear baselines may not always be available experimentally: a protein may have a high melting point, preventing access to the unfolded state baseline,[9,10] or protein folding is measured in a living cell, where high temperatures or denaturant concentrations cannot be reached without killing the cell,[11,12] or a protein mutant may be destabilized, preventing access to the full native state baseline.[13,14]

In such cases, kinetics can come to the rescue as long as all the activated folding/unfolding events can be resolved. Kinetics helps because of a separation of time scales. For concreteness we discuss the two-state scenario in Fig. 1, where a temperature jump induces folding. Immediately following a jump from T_2 to T_1, relaxation of the chromophore and within the well is rapid, and creates the unwanted baseline. Relaxation between wells (dotted arrow "*U*" to "*F*" in Fig. 1) is slow, and corresponds to the desired population transfer among thermodynamic states. If the activation barriers are large enough, kinetics can separate the fast relaxation within thermodynamic states from the slow transitions among thermodynamic states. A small kinetic jump automatically separates baseline from folding amplitude. In the same spirit, Mücke and Schmid previously proposed distinguishing native populations from faster-unfolding intermediates.[15]

We derive the necessary equations to allow accurate melting temperatures T_m to be measured when baselines are incomplete, nonlinear, or noisy. We compare this "thermodynamics from kinetics" approach with conventional thermodynamic fitting, showing that our approach is more robust when baselines are incomplete. To do so, we first analyze simulated thermodynamic and kinetic data obtained from the same model by solving the Smoluchowski equation. Next we illustrate the approach by analyzing experimental thermal denaturation curves and T-jump kinetics of the small two-state folder PI3K SH3 (Ref. 16) labeled by a Förster resonant energy transfer (FRET) pair. This is a difficult case for conventional thermodynamic fitting: the distance between the FRET pair does not increase much when the protein unfolds, but the quantum efficiencies of donor and acceptor depend on temperature, yielding a small signal relative to the baseline. The kinetic approach yields a reliable T_m.

II. EXPERIMENTAL AND COMPUTATIONAL METHODS

A. Protein engineering

The recombinant protein used in this study has 83 residues of the sarcoma homology 3 domain (SH3) from

a)Author to whom correspondence should be addressed. Electronic mail: gruebele@scs.uiuc.edu. FAX: (001) 217 244 3186.

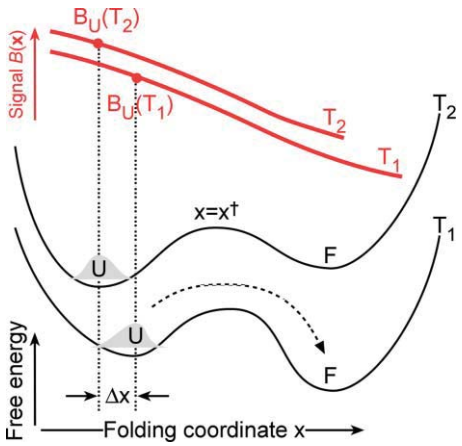

FIG. 1. Kinetic separation of time scales. Black free energy curves and red (light) signal baselines are shown at two temperatures (T_1 and T_2) as a function of the reaction coordinate x. The unfolded portion of the thermal population is highlighted (shaded). Upon T-jump, the population rapidly relaxes within a well (amount Δx); this fast relaxation does not populate a new thermodynamic state (e.g., F), it merely shifts the thermodynamic state (U). Likewise the signal baseline rapidly switches due to chromophore relaxation. After the T-jump, the population relaxes much more slowly over the barrier (dotted arrow), shifting population between thermodynamic states.

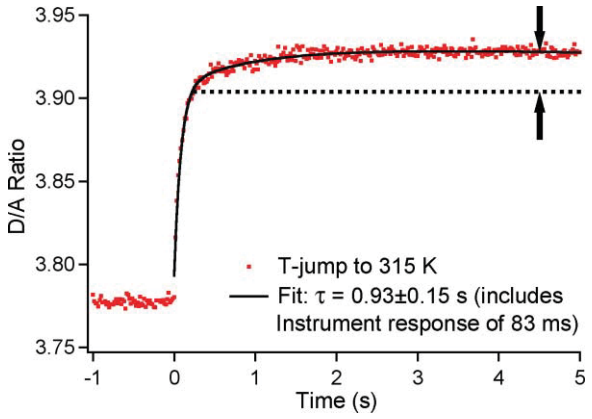

FIG. 2. Kinetic transient at 315 K (red/light dots) and fit (black curve). Protein concentration was 5 μM. The dotted line and arrows indicate the amplitude of the resolved kinetic phase as in Fig. 3(a). The rapid rise is the 83 ms instrument response.

phosphoinositide 3-kinase (PI3K) plus a two residue N-terminal extension (GS).[16] It is an apparent two-state folder with a fast barrier-free initial chain collapse.[17] To remove slow proline isomerization kinetics, the two prolines were mutated to G and W residues (underlined), yielding the sequence GSMSAEGYQY RALYDYKKER EEDIDLHLGD ILTVNKGSLV ALGFSDGQEA K**G**EEIGWLNG YNETTGERGD F**W**GTYVEYIG RKKIS.

The FRET plasmid for PI3K-SH3 was created by ligating the genes for the AcGFP1 FRET donor (Clontech) and mCherry FRET acceptor (Clontech) to the 5′ and 3′ ends, respectively. The fusion gene was cloned into the pDream 2.1 vector (Genscript). A 6-histidine tag was added to the N-terminus of AcGFP1 to enable purification on a Ni-NTA column for *in vitro* studies. The plasmid was transformed in BL21-CodonPlus (DE3)-RIPL cells (Stratagene) in 2-YT (yeast extract-tryptone) media. The transformed bacterial culture was grown at 37 °C until the optical density reached 1.0. Protein expression was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (Inalco). After induction, the culture was grown at 25 °C for 24 to 48 h, followed by centrifugation at 10 000 rpm for 10 min. The centrifuged cells were lysed by ultrasonication. Cell lysate was run through a Ni-NTA column and purified according to the manufacturer's protocol (Qiagen). Purity of the protein was determined by electrospray ionization mass spectrometry and matrix-assisted laser desorption/ionization. For both the *in vitro* thermodynamics and kinetics experiments, we used a 5 μM solution of the SH3 fusion protein.

B. Thermodynamic and kinetic experiment

Thermal denaturation and temperature jump kinetics of PI3K-SH3 were collected over the 295 to 333 K range on our fast relaxation imaging (FreI) instrument, described in detail elsewhere.[12] Briefly, protein solution was placed in an imaging chamber whose temperature was controlled by heating resistors and monitored with thermocouples. Donor (AcGFP) and acceptor (mCherry) fluorescence was imaged independently onto a CCD camera. As the protein unfolds, the donor fluorescence intensity (D) increases relative to the acceptor fluorescence intensity (A). Kinetics was initiated by jumping the temperature with a programmable 2200 nm diode laser. For an apparent two-state folder like PI3K SH3,[16] upward and downward jumps provide equivalent rate information ($k_{observed}(T) = k_{forward}(T) + k_{backward}(T)$). Upward jumps have better time resolution and were analyzed here. Figure 2 shows an upward jump from $T = 313$ to 315 K. A correction for slow differential bleaching of the AcGFP donor and mCherry acceptor was applied by subtracting the small linear downward trend in the D/A signal (see Fig. S4 in supplementary material[26]). The instrument response of 83 ms was determined from jumps at room temperature and held constant while fitting kinetics at higher temperature. It was about 10 times faster than the measured relaxation times (Fig. 2).

The D and A signals of a two-state folder are each linear combinations of the folded fraction f_F and unfolded fraction f_U, with coefficients dependent on the FRET efficiency between the green and red chromophores and on chromophore quantum yields.[12] The ratio D/A eliminates the dependence on the donor quantum yield, while D$a(T)$/A also eliminates the dependence on the acceptor quantum yield $a(T)$. D/A is strictly speaking a ratio of linear combinations of f_F and f_U. For small folding amplitude changes such as in Fig. 2, D/A is proportional to the folding population, so we plot it here for fitting.

C. Computed test data

To compare conventional fitting of T_m with the "thermodynamics from kinetics" approach, we generated simulated thermodynamic melts and kinetic amplitude data as a function of temperature from the same free energy model. A one-dimensional two-state free energy surface $G(x,T)$ was modeled as the sum of two Gaussians with the depth of one

well allowed to vary linearly with temperature, similar to the schematic in Fig. 1. The signal function, $B(x,T)$ was modeled as a Taylor series in x and T. For the case shown in Fig. 4, which resembles our PI3K SH3 experimental data, the temperature dependence of the signal function was linear. The supplementary material[26] describes $G(x,T)$ and $B(x,T)$ in detail, and shows additional simulation results with nonlinear signal functions and near-downhill free energies.

Thermodynamic curves and temperature jumps were computed at approximately 2° intervals. The equilibrium populations $\rho(x,T)$ on the free energy surface $G(x,T)$ were used to determine the thermodynamic titration signal $B(T)$ (see Eqs. (1) and (2) below). For the kinetic data, 4° temperature jumps from T_1 to T_2 were simulated by switching the free energy suddenly from $G(x,T_1)$ to $G(x,T_2)$ and allowing the initial population $\rho(x,T_1)$ to relax to the final population $\rho(x,T_2)$. The population relaxation was simulated by Smoluchowski dynamics using a singular-value basis method.[18] An exponential decay function was fitted to determine the resolved kinetic amplitude (indicated by arrow pairs in Figs. 3(a)–3(c)). The final temperature after each jump corresponded to a point in the thermodynamic titration curve for comparison.

To assess the effect of noisy data, random Gaussian-distributed noise was added to the signal function from which the thermodynamic melt and kinetic amplitude data were computed. The maximum noise level was chosen so that the signal-to-noise-ratio of the kinetic traces near T_m would be ≤ 2 (see supplementary material[26] in Fig. S3(a)); it is unlikely that an experimentalist would accept noisier data. 10 thermodynamic and kinetic data sets with different pseudo-random noise were generated. In the plots of kinetic amplitude vs. temperature, all amplitudes with a fitting uncertainty greater than the amplitude were set to zero, the others were weighted by the uncertainty. Each of the 10 thermodynamic melt or kinetic amplitude data sets was fitted by least squares to determine T_m, and the fit with the best chi-squared (out of 20 initial parameter guesses) was used. At each noise level, the 10 "thermodynamic" T_m and 10 "kinetic" T_m were then used to compute average T_m and standard deviation of T_m. The difference ΔT_m between the average T_m and the known T_m of the model is a systematic error. The standard deviation is a random error. The magnitudes of both were summed to represent the error by a single value with units of temperature.

III. THE MODEL

The fitting procedure discussed here can be used to improve accuracy for any denaturation measurement, such as by temperature, *p*H, or denaturant concentration. For concreteness, we will consider temperature as our example. Kinetic separation of in-well and barrier-crossing dynamics improves thermodynamic fitting for two-state or multistate proteins. Again for concreteness, we describe the two-state case.

Consider the model represented in Fig. 3. The free energy along the reaction coordinate x leads to a normalized equilibrium population,

$$\rho(x,T) = \frac{e^{-G(x,T)/RT}}{\int_{-\infty}^{\infty} dx e^{-G(x,T)/RT}}. \quad (1)$$

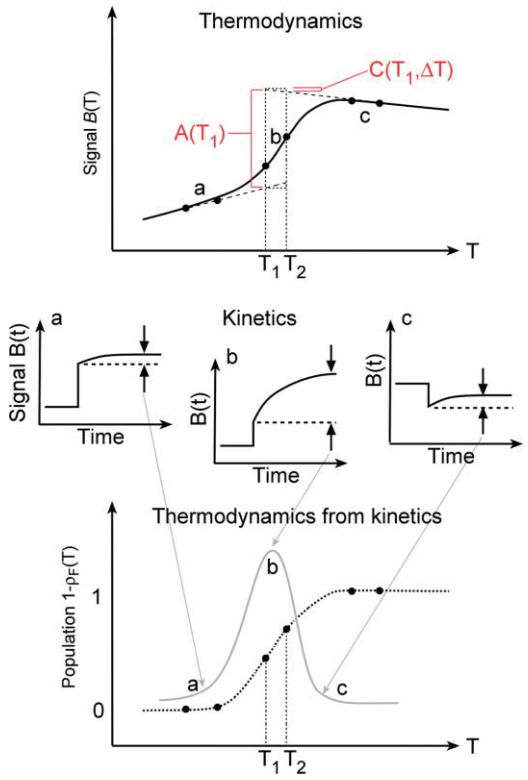

FIG. 3. Comparison of traditional heat denaturation and "thermodynamics from kinetics." (Top) Thermodynamic folded baseline, transition and unfolded baseline. Red labels are defined by the equations in the text. a, b, and c label kinetic T-jumps shown in (Middle): at "a," little population switches states, and the resolved phase amplitude (arrows) is small. The initial amplitude jump is proportional to the baseline slope. At "b," a large population switch between states is resolved (arrows). At "c," the resolved phase is small again. The initial amplitude has switched sign because the baseline slope has switched sign. (Bottom) By plotting only the resolved amplitude (solid curve), the baselines are reduced and only population changes (maximized at b) are selected. Integrating the amplitude curve yields the (unnormalized) population as a function of temperature. The melting point of the protein is easily identified near the maximum at b, even if the baseline at a or c cannot be fully measured.

The free energy and population depend on temperature (or some other perturbing variable). The signal function $B(x,T)$ along the reaction coordinate yields an observed signal (e.g., tryptophan fluorescence, IR absorption, FRET, etc.) as a function of temperature,

$$B(T) = \int_{-\infty}^{\infty} dx B(x,T)\rho(x,T). \quad (2)$$

If the two-state (or n-state) approximation is valid, i.e., if the barriers are high enough, we can divide the reaction coordinate into states, such as a "folded" and "unfolded" side in the two-state case. The optimal dividing surface will lie near the barrier at $x^\dagger$ in Fig. 1.

$$\begin{aligned} B(T) &= \int_{-\infty}^{x^\dagger} dx B(x,T)\rho(x,T) + \int_{x^\dagger}^{\infty} dx B(x,T)\rho(x,T) \\ &= B_U(T)f_U(T) + B_F(T)f_F(T). \end{aligned} \quad (3)$$

Here f_U and f_F are the unfolded and folded fractions, and B_U and B_F are their signal baselines. The second

line of Eq. (3) is obtained by defining $f_U(T) = \int_{-\infty}^{x^\dagger} dx \rho(x,T)$ (and $f_N(T)$ accordingly), and setting $B_U(T) = [\int_{-\infty}^{x^\dagger} dx\, B(x,T)\rho(x,T)]/f_U(T)$ (and $B_N(T)$ accordingly). The observed signal depends on an unfolded baseline $B_U(T)$ weighted by the unfolded population, and a folded baseline $B_N(T)$ weighted by a folded population.

The population in state "*i*" (U or F above) is related to the free energy by an equation analogous to Eq. (1),

$$f_i(T) = \frac{e^{-G_i(T)/RT}}{\sum e^{-G_i(T)/RT}}. \tag{4}$$

For the two-state case, a simple and frequently used model for the free energy, equilibrium constant, and population is the Taylor series expansion[19]

$$\Delta G = G_U - G_F = g^{(1)}(T - T_m) + g^{(2)}(T - T_m)^2 \cdots,$$
$$K_{eq}(T) = e^{-\Delta G/RT}, \quad f_F(T) = \frac{K_{eq}}{1 + K_{eq}}, \quad f_U = 1 - f_F. \tag{5}$$

The goal is to determine the fitting parameters T_m and $g^{(1)}$ by extracting the population $f_F(T)$ from the measured signal $B(T)$. In particular, linear baselines with adjustable slopes m have been used in a vast literature to approximate $B_U(T)$ and $B_F(T)$,

$$B(T) \approx (B_{U0} + m_U(T - T_m))f_U + (B_{F0} + m_F(T - T_m))f_F. \tag{6}$$

The problem is that if data cannot be collected to map out full baselines, the baseline fitting parameters B_{i0} and m_i cannot be determined accurately. Even if the full baseline is observed, it may be steep or nonlinear, and a linear extrapolation, for example, of the unfolded baseline into the folded region (low temperature) is inaccurate or inadequate.[8] The result is that T_m and other fitting parameters cannot be determined accurately.

Kinetics can help out. If kinetics can be measured that capture all well-to-well relaxation events, easily possible nowadays with nanosecond temperature jumps,[20] the baseline to a significant extent can be separated from the population change. This is illustrated by Fig. 3: at low or high T where only in-well relaxation occurs, kinetics mostly shows a sudden jump, due to the baseline. If kinetics is measured near T_m, the well-to-well dynamics produce a resolvable amplitude. At T_m, the resolved amplitude is largest. Plotting the resolved amplitude thus maps out the thermodynamic transition with a peak at T_m (Fig. 3, bottom). The amplitude curve can be integrated to yield the population curve.

Mathematically, the kinetic signals just before the jump, just after the jump, and at equilibrium long after the jump, are

$$\begin{aligned} B(t = 0^-) &= B_F(T_1)f_F(T_1) + B_U(T_1)f_U(T_1), \\ B(t = 0^+) &= B_F(T_2)f_F(T_1) + B_U(T_2)f_U(T_1), \\ B(t = \infty) &= B_F(T_2)f_F(T_2) + B_U(T_2)f_U(T_2). \end{aligned} \tag{7}$$

Measuring kinetics independently determines the resolved and unresolved amplitudes

$$\begin{aligned} &B(t = \infty) - B(t = 0^+) \\ &\quad = [B_F(T_2) - B_U(T_2)][f_F(T_2) - f_F(T_1)], \\ &B(t = 0^+) - B(t = 0^-) = [B_F(T_2) - B_U(T_2) + B_U(T_1) \\ &\quad - B_F(T_1)]f_F(T_1) + [B_U(T_2) - B_U(T_1)] \end{aligned} \tag{8}$$

(using $f_U = 1 - f_F$), or

$$\begin{aligned} &B_{resolved}(T) = A(T)[f_F(T) - f_F(T - \Delta T)], \\ &B_{unresolved}(T - \Delta T \to T) \\ &\quad = [A(T) - A(T - \Delta T)]f_F + C(T, \Delta T), \end{aligned} \tag{9}$$

where $\Delta T = T_2$-T_1 is the size of the T-jump. Eq. (9), together with a model for the population f_F such as Eq. (5), constitutes the kinetic fitting model. There are still two baseline functions (here $A(T) = B_F(T) - B_U(T)$ and $C(T) = B_U(T) - B_U(T - \Delta T)$), just as in the thermodynamic fit in Eq. (3). Unlike Eq. (3), two data points at each temperature constrain the baseline functions and populations, instead of a single data point.

Also unlike Eq. (6), the baseline functions A and C are differences between baselines. C is much more likely than the original baselines to be approximated by linear functions if the temperature jump ΔT is small. The situation for A depends on how different the folded and unfolded baselines are. If the folded and unfolded baselines are far apart, or if they are nearly parallel, then $A(T)$ is approximately constant. Otherwise $A(T)$ still has to be approximated by a linear baseline or a polynomial, but at least there is only one function $A(T)$ on which the resolved amplitude depends, instead of two functions $B_F(T)$ and $B_U(T)$.

The following are good approximations when a small jump ΔT is applied

$$\begin{aligned} &B_{resolved}(T) \approx \Delta T\, A(T) \left.\frac{\partial f_F}{\partial T}\right|_{T - \Delta T/2} \\ &B_{unresolved}(T - \Delta T \to T) \\ &\quad \approx \Delta T \left(\left.\frac{\partial A}{\partial T}\right|_{T - \Delta T/2} f_F(T - \Delta T) + \left.\frac{\partial B_U}{\partial T}\right|_{T - \Delta T/2} \right). \end{aligned} \tag{10}$$

$B_{resolved}$ contains a lot of useful information on its own. It is proportional to the derivative of the thermodynamic population curve. Inserting the two-state population model of Eq. (5) into Eq. (10) yields

$$\begin{aligned} B_{resolved}(T) = &\frac{-g^{(1)}\Delta T\, T_m}{R(T - \Delta T/2)^2} [A_0 + m_A(T - T_m)] \\ &\times \frac{e^{-g^{(1)}(T - \Delta T/2 - T_m)/R(T - \Delta T/2)}}{(1 + e^{-g^{(1)}(T - \Delta T/2 - T_m)/R(T - \Delta T/2)})^2}. \end{aligned} \tag{11}$$

There are four fitting parameters: $g^{(1)}, T_m, A_0, m_A$, including the last two for the baseline of A. This model requires 2

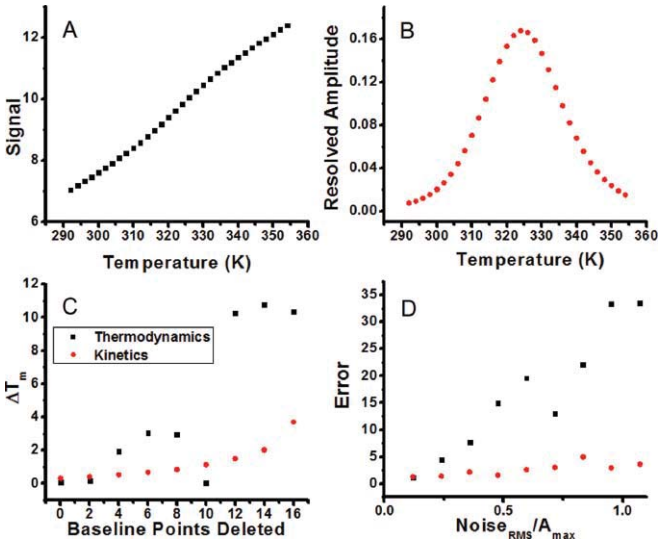

FIG. 4. Model data derived by Smoluchowski dynamics from a two-state free energy and signal function. The data were chosen to correspond to a "difficult" case, such as FRET with a small protein. (a) Conventional thermodynamic melt. (b) Resolved kinetic amplitude. (c) Deviation ΔT_m of the fitted from the model melting temperature as function of deleted points in the high temperature baseline. (d) Combined random and systematic error in the melting temperature as a function of noise added to calculate the data in (a) and (b).

fewer fitting parameters than Eqs. (4) and (5), the standard linear baseline two-state fit, and is expected to yield T_m and $g^{(1)}$ more reliably. In Eqs. (10) and (11), the "$-\Delta T/2$" arise from best approximating a finite difference by the derivative at its center.

IV. RESULTS

A. Fits to simulated data

To compare the standard thermodynamic fitting (Eq. (6) combined with the two-state model in Eq. (5)) with the "thermodynamics from kinetics" fitting (Eq. (11)), a set of test data similar to the PI3K SH3 experimental data was simulated. The model T_m was 323 K in both cases. The results are summarized in Fig. 4. Panels (a) and (b) show the thermodynamic titration and the resolved kinetic amplitude $B_{resolved}$. Fig. 4 corresponds to a "difficult" case very similar to the FRET experiments discussed below. In cases like these, a small amount of noise can make it difficult to identify the cooperative region in a thermal titration experiment.

The "thermodynamics from kinetics" method works well as long as the region of maximum amplitude near T_m is available for fitting, even if there is very little baseline. Panel (c) of Fig. 4 shows the deviation of the fitted melting temperature from the model value as data points were removed from the high-temperature unfolded baseline. This is a common difficulty with *in vivo* or living cell experiments, where high temperature measurements cannot be made.[12] The thermodynamic model became unstable quickly as the baseline was removed. The kinetics model was able to fit T_m reliably until the peak of the resolved amplitude was truncated (removal of half of the data points). Analogous results for both models were obtained by removing the folded baseline, or by symmetrically removing both baselines progressively.

Because experiments do not produce an infinite signal to noise ratio, we tested how the thermodynamic and "thermodynamics from kinetics" fits performed with varying amounts of noise added to the signal function used to generate both thermodynamic and kinetic data. "Thermodynamics from kinetics" generally makes smaller errors than a conventional thermodynamic fit at the same noise level. Panel (d) of Fig. 4 shows the error in T_m as a function of noise level. Two sources of error (random and systematic) are included (see Methods). The noise level is given relative to the kinetic amplitude at T_m in panel (b), so an rms noise of 1 corresponds to a signal-to-noise ratio of 1. Figure S3 in the supplementary material[26] shows the raw simulated kinetic traces for the lowest nonzero noise level and the highest noise level in Fig. 4(c). Even the highest noise level yielded a fairly reliable T_m, and it is unlikely that experimentalists would accept noisier kinetic data. The thermodynamic data is completely unable to provide a reliable T_m when the rms noise approaches 1.

We carried out additional simulations analogous to Fig. 4, discussed in more detail in the supplementary material.[26] Figure S1 in the supplementary material corresponds to an easy case for conventional thermodynamics, where the folded and unfolded baselines have different slope, so the transition is clearly discernible. Regardless, the kinetic model again performed better. Figure S2 illustrates a nearly downhill folding case. The barrier crossing rate becomes comparable to the intra-well relaxation rate or the barrier transit rate (k_m or $k_{transit}$, both estimated experimentally by various methods[21–24]). The "thermodynamics from kinetics" approach becomes prone to noise but still outperforms the standard thermodynamic analysis.

B. Fits to experimental data

To illustrate "thermodynamics from kinetics" with real data, we measured the thermal denaturation of the 83 residue two-state folder protein PI3K SH3, labeled by a AcGFP/mCherry donor/acceptor pair at the termini. This is a relatively difficult case for thermodynamic fitting: The termini of the small SH3 protein expand from 8 Å (Protein Data Bank distance from residue 2 to 82) to only 33 Å (assuming a random coil in the unfolded state), compared to $R_0 = 52$ Å for the donor/acceptor pair. In addition, the baseline caused by the temperature-dependent quantum yields of AcGFP and mCherry is large. Thus the presence of noise makes it difficult to discern the melting transition (Fig. 5(a)). We also measured temperature jump kinetics of PI3K SH3 (Fig. 2) and plotted the resolved kinetic amplitude as a function of temperature (Fig. 5(b)), yielding the characteristic plot with a maximum near T_m.

The kinetic fit yields a melting temperature of 312 K, whereas the conventional thermodynamic fit yields a higher 317 K. Since the "thermodynamics from kinetics" plot shows a very clear peak, whereas the thermodynamic data appears almost as a straight line with noise, we believe that the kinetics fit provides a better approximation to the melting

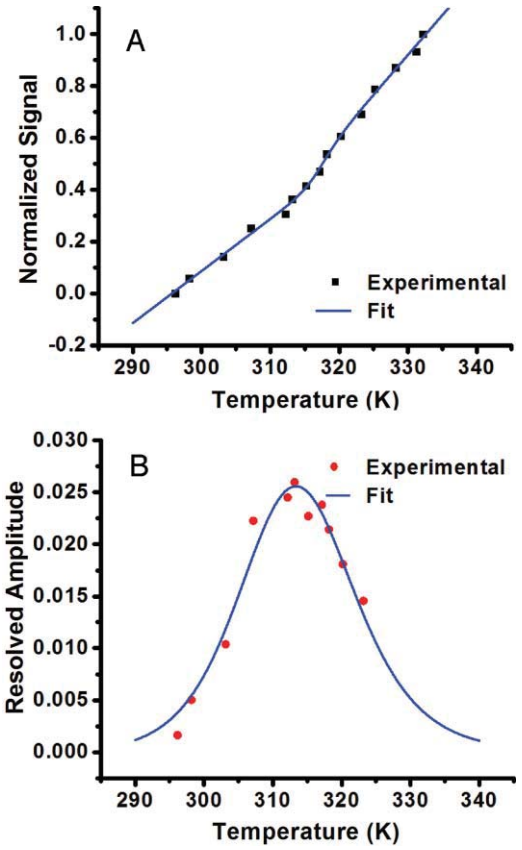

FIG. 5. Experimental data for FRET-labeled PI3K SH3 protein and model fits. (a) Conventional thermal titration experiment and fit with $T_m = 316.9 \pm 0.2$ K. (b) Resolved kinetic amplitude and fit with $T_m = 311.8 \pm 2.4$ K.

temperature of PI3K SH3. This should hold especially because the folded and unfolded baselines in Fig. 5(a) have about the same slope, leading to a cancellation of the slope in Eq. (11) so the kinetics can be fitted by only three parameters (T_m, $g^{(1)}$ and A_0).

The better quality of the "thermodynamics from kinetics" is further supported by Table I. There we show the effect of truncating the high temperature baseline. The kinetic fit consistently returns T_m in the 311-312 K range and $g^{(1)}$ of 0.14, with two standard deviation errors to match. The thermodynamic fit varies all over the place, well outside the estimated uncertainties. The results in Table I are the lowest chi-squared fits out of many initial parameter guesses, to give the thermodynamic fit the best chance of finding the smalles chi-squared. Unlike the thermodynamic nonlinear least-squares fit, which gets trapped in various local minima, the kinetic fit finds the same minimum of chi-squared for almost all initial guesses.

TABLE I. Fitted melting temperatures and free energy parameters by standard thermodynamic fitting and "thermodynamics from kinetics" as a function of upper baseline truncation. The root-mean-square variation of T_m or $g^{(1)}$, indicates how consistently parameters can be determined with different amounts of baseline. The ± errors are two standard deviations of the mean.

	Thermodynamics		Kinetics	
	T_m (K)	$g^{(1)}$ (kJ mol^{-1} K^{-1})	T_m (K)	$g^{(1)}$ (kJ mol^{-1} K^{-1})
Data up to 332 K	316.9 ± 0.2	1.79 ± 0.92	…	…
Data up to 323 K	317.7 ± 0.2	2.17 ± 0.99	311.8 ± 2.4	0.14 ± 0.02
Data up to 320 K	310.3 ± 0.2	> 10[a]	311.5 ± 2.5	0.14 ± 0.02
Data up to 318 K	332.6 ± 0.2	0.55 ± 0.02	311.3 ± 2.7	0.14 ± 0.02
rms variation	9.4	0.7	0.25	0.02

[a]Any large value fits the data; this value was not included in the rms variation, which would be even larger otherwise.

V. DISCUSSION

The measurement of protein melting temperatures can be a challenge when the high temperature baseline is missing, or when the baselines are steep. These problems can occur in many situations: Some proteins aggregate at high temperature, and need to be measured at low temperature only. Some proteins have melting points above the boiling point of the solvent.[9] And with an increasing interest in folding in living cells, or even *in vivo*, the high temperature baseline may be inaccessible without killing the cell or organism under study.[25] In such cases, it would be nice to have a method for measuring protein stability without the need for scanning the whole baseline.

For all of the cases we simulated, we found that the method of fitting "thermodynamics from the kinetics" was more reliable than conventional thermal melts when baseline information is limited. Kinetics eliminates much of the baseline by shifting it into the unresolved phase, whereas the population switch among wells falls into the resolved phase. The latter yields a nearly baseline-free signal with a peak near T_m, although large baselines of opposite sign could shift the peak. In the latter case, conventional thermodynamic fitting should be at its most effective.

The "thermodynamics from kinetics" model yields a more consistent increase of the error with baseline truncation. In Fig. 4(c) (and Fig. S1 in the supplementary material[26]), the conventional fit produces a nonmonotonic error. As further data is removed, the unfolded baseline is lost and the thermodynamic model is unable to fit the melting temperature. In contrast, the kinetic fit has a slowly and predictably increasing error when its high temperature points are truncated, and generally performs well until the truncation reaches the denaturation midpoint. The fits to experimental data in Table I also show that kinetics fitting produces more consistent thermodynamic parameters and more realistic error estimates.

We added the equivalent noise mathematically to the thermodynamics and the kinetics test data. We found that the "thermodynamics from kinetics" method is less susceptible to experimental noise. This is due to the fact that experimental noise is added to a derivative of the thermodynamic signal (T-jump) rather than adding the noise to a smooth function and then taking the derivative (thermal melt).

Ironically, the experimentally relevant noise is often worse for the thermodynamic signal than for the time-resolved kinetic signal, even though the thermodynamic signal is collected over a longer time period with more signal averaging. In fact, slow data collection can amplify

systematic errors such as sample bleaching, laser intensity drift, or sample temperature fluctuations during data collection.

The kinetic model can be constrained more easily than the thermodynamic model. For a two-state system, the resolved kinetic amplitudes approach zero away from the melting temperature; for the thermodynamics, the baselines have no similar consistent constraint.

The major drawback of "thermodynamics from kinetics" is that all barrier-crossing phases must be resolved. Otherwise, genuine population transfer is lumped together with dynamics within a single state in the sudden response right after the T-jump (or whichever relaxation method is used). Thus methods with good time resolution must be employed for the measurement. However, if excellent signal-to-noise can be achieved, the method is useful right down to the folding speed limit (downhill folding).

In summary, "thermodynamics from kinetics" provides a simple formula (Eq. (11)) for fitting the resolved kinetics amplitude to obtain thermodynamic parameters such as T_m. The result is more robust to truncation of the baseline and noise than are conventional thermodynamic fits. The only drawback is that the kinetics must resolve all barrier crossing events to distinguish them from intra-well relaxation and chromophore relaxation.

ACKNOWLEDGMENTS

K.G, Y.R.C., and M.G. were funded by the National Science Foundation (NSF) Center for Physics of Living Cells. G.S. was supported by a grant from the National Science Foundation, Grant No. MCB 1019958.

[1] T. Y. Tsong, R. L. Baldwin, and E. L. Elson, Proc. Natl. Acad. Sci. U.S.A. **68**, 2712 (1971).
[2] J. SantaLucia, Jr. and D. Hicks, Annu. Rev. Biophys. Biomol. Struct. **33**, 415 (2004).
[3] D. T. Leeson, F. Gai, H. M. Rodriguez, L. M. Gregoret, and R. B. Dyer, Proc. Natl. Acad. Sci. U.S.A. **97**(6), 2527 (2000).
[4] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995).
[5] J. R. Alcala, E. Gratton, and F. G. Prendergast, Biophys. J. **51**, 925 (1987).
[6] C. N. Pace, Methods Enzymol. **131**, 266 (1986).
[7] M. M. Santoro and D. W. Bolen, Biochemistry **31**, 4901 (1992).
[8] B. Schuler, E. A. Lipman, and W. A. Eaton, Nature (London) **419**, 743 (2002).
[9] R. Hiller, Z. H. Zhou, M. W. W. Adams, and S. W. Englander, Proc. Natl. Acad. Sci. U.S.A. **94**, 11329 (1997).
[10] M. Elwell and J. Schellman, Biochim. Biophys. Acta **386**(1), 309 (1975).
[11] Z. Ignatova and L. M. Gierasch, Proc. Natl. Acad. Sci. U.S.A. **101**(2), 523 (2004).
[12] S. Ebbinghaus, A. Dhar, J. D. McDonald, and M. Gruebele, Nat. Methods **7**, 319 (2010).
[13] J. L. Marmorino and G. J. Piel, Biochemistry **34**, 3140 (1995).
[14] A. E. Eriksson, W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthews, Science **255**(5041), 178 (1992).
[15] M. Mücke and F. X. Schmid, Biochemistry **33**, 12930 (1994).
[16] J. I. Guijarro, C. J. Morton, K. W. Plaxco, I. D. Campbell, and C. M. Dobson, J. Mol. Biol. **276**, 657 (1998).
[17] A. Dasgupta and J. B. Udgaonkar, J. Mol. Biol. **403**(3), 430 (2010).
[18] G. Scott and M. Gruebele, J. Comput. Chem. **31**(13), 2428 (2010).
[19] J. Ervin and M. Gruebele, J. Biol. Phys. **28**, 115 (2002).
[20] H. Ma, J. Ervin, and M. Gruebele, Rev. Sci. Instrum. **64**(2), 486 (2004).
[21] S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton, Proc. Natl. Acad. Sci. U.S.A. **93**, 11615 (1996).
[22] W. Y. Yang and M. Gruebele, Nature (London) **423**(6936), 193 (2003).
[23] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgartel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, and B. Schuler, Proc. Nat. Acad. Sci. U.S.A. **104**, 105 (2007).
[24] F. Liu and M. Gruebele, J. Chem. Phys. **131**, 195101 (2009).
[25] S. Ebbinghaus and M. Gruebele, J. Phys. Chem. Lett. **2**, 314 (2011).
[26] See supplementary material at http://dx.doi.org/10.1063/1.3607605 for additional figures and text describing the kinetics measurements and simulations.