An experimental survey of the transition between two-state and downhill protein folding scenarios

Feng Liu[‡], Deguo Du[§], Amelia A. Fuller[§], Jennifer E. Davoren[¶], Peter Wipf[¶], Jeffery W. Kelly[§], and Martin Gruebele[‡‖‡‡]

[‡]Center for Biophysics and Computational Biology and [‖]Departments of Chemistry and Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [§]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road BCC265, La Jolla, CA 92037; and [¶]Center for Chemical Methodologies and Library Development, University of Pittsburgh, Pittsburgh, PA 15260

Communicated by Peter G. Wolynes, University of California at San Diego, La Jolla, CA, December 18, 2007 (received for review November 25, 2007)

A kinetic and thermodynamic survey of 35 WW domain sequences is used in combination with a model to discern the energetic requirements for the transition from two-state folding to downhill folding. The sequences used exhibit a 600-fold range of folding rates at the temperature of maximum folding rate. Very stable proteins can achieve complete downhill folding when the temperature is lowered sufficiently below the melting temperature, and then at even lower temperatures they become two-state folders again because of cold denaturation. Less stable proteins never achieve a sufficient bias to fold downhill because of the onset of cold denaturation. The model, considering both heat and cold denaturation, reveals that to achieve incipient downhill folding (barrier <3 *RT*) or downhill folding (no barrier), the WW domain average melting temperatures have to be ≥50°C for incipient downhill folding and ≥90°C for downhill folding.

activated rate | alkene peptide isosteres | molecular rate | speed limit | stability

Energy landscape theory predicts that the entropic and enthalpic contributions to the free energy of a protein may be able to compensate for one another to the point where no significant (>3 *RT*) barrier appears along the folding reaction coordinate (1). Such folding is now referred to as "type 0" or "downhill" folding (1, 2). The possibility of downhill folding has been supported by a number of experiments (3–7). Kinetic measurements have focused on the transition from simple single exponential (two-state) to nonexponential (low barrier) back toward simpler (pure downhill) kinetics as the thermodynamic bias toward the native state is increased (3, 6, 8, 9). Thermodynamic measurements with probe-dependent baselines and transition temperatures suggest that downhill folding may be possible even at the melting temperature of a protein (10–12). Recently, two engineered proteins with identical melting temperatures were compared by both kinetic and thermodynamic criteria, showing that one can be classified as a two-state folder, whereas the other can be classified as a downhill folder (13). Such results have been debated extensively in the literature (14–17).

It has been suggested that the kinetics and thermodynamics of downhill folders must be very sensitive to sequence and environment because of the low barriers involved (18). Indeed, the fast folder lambda repressor has been shown to fold by either two-state, framework intermediate, or downhill mechanisms, depending on solvent conditions and sequence (6, 9, 13, 19, 20). Models for protein BBL, another fast folding protein, also indicate that it folds either in a two-state or downhill manner, depending on the exact sequence and solvent conditions (21, 22).

The diversity of observations suggests that criteria for downhill folding can be developed only by examining a large number of fast-folding proteins. Here, we take a survey approach to the experimental study of downhill folding. We examine a series of 35 engineered WW domains with variation in loops 1 and 2, where sequence changes have the largest effect on folding kinetics without disrupting the fold topology (23–25). These proteins exhibit a wide range of melting temperatures, T_m (20–85°C) and span a 600-fold range of rate coefficients at the temperature of the maximum folding rate. Four of the fastest folders we studied switch from single exponential kinetics to having an additional fast molecular rate coefficient, k_m, below T_m, indicating incipient downhill folding (6).

We deduce a rather sharp transition from two-state folding to incipient downhill folding as a function of melting temperature and group proteins into three temperature zones proposed in our earlier work: apparent two-state folders, incipient downhill folders (<3-RT_0 barrier), and downhill folders (9, 13). These zones for T_m can be understood by using a quantitative model that accounts for both heat denaturation (at T_m) and cold denaturation (at T_{cd}). As illustrated in Fig. 1, a very stable protein (Fig. 1 *Left*) might be a two-state folder near its T_{cd} and T_m but have a sufficiently strong native bias at temperatures between T_{cd} and T_m to become a downhill folder. A less stable protein (Fig. 1 *Right*) might barely achieve a sufficient native bias to become an incipient downhill folder or it might even remain a two-state folder over the entire temperature range between T_{cd} and T_m. Two-state/downhill/two-state transitions for stable proteins owe their existence to the temperature dependence of solvent-averaged interaction free energies, as was first discussed for phosphoglycerate kinase by Sabelko *et al.* (3). Our multiprotein survey enables us to predict at what characteristic melting temperatures to expect the transitions from two-state folding to incipient downhill folding, and finally to complete downhill folding. We hope that this work will stimulate tests of our prediction for two-state, incipient downhill, and downhill folding zones to see whether they apply quantitatively beyond the WW domain fold shown in Fig. 2.

Results

Proteins. To obtain meaningful statistics, we used engineered WW domains exhibiting a wide range of melting temperatures. Because loop 1 formation is rate-limiting for WW domain folding, we used WW domain sequences harboring both natural and unnatural sequences in lieu of the WT loop 1 sequence. In addition, we include a number of previously studied WW domains in our analysis (25–27). Table 1 shows the 35 WW Pin1-derived domain sequences we use in our survey. Further details are discussed in *Materials and Methods*.

Unfolding Thermodynamics Are Two-State at T_m. Thermal titration curves (see *Materials and Methods*) were fitted to a two-state folding model as described in detail in ref. 24. The free energy difference between the native state and the denatured state was

Author contributions: J.W.K. and M.G. designed research; F.L., D.D., A.A.F., and J.E.D. performed research; J.E.D. and P.W. contributed new reagents/analytic tools; F.L. and M.G. analyzed data; and F.L., J.W.K., and M.G. wrote the paper.

The authors declare no conflict of interest.

[‡‡]To whom correspondence should be addressed. E-mail: gruebele@scs.uiuc.edu.

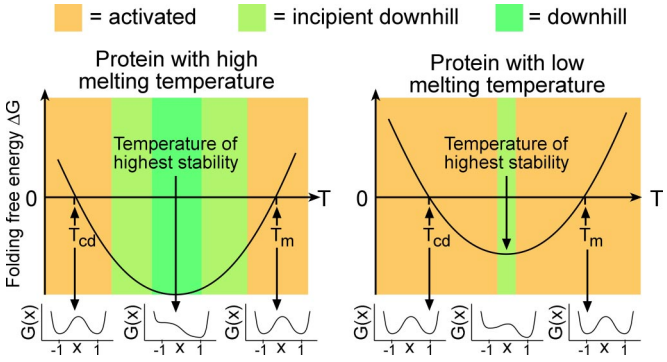

Fig. 1. Free energy difference between the native and the denatured state as a function of temperature. The temperature of maximal stability lies between the cold and heat denaturation temperatures. A very stable protein may reach downhill folding there, whereas a less stable protein may still fold over an activation barrier. Near the two melting points T_{cd} and T_m, both proteins fold over an activation barrier, as illustrated by the free energy surfaces $G(x)$ as a function of reaction coordinate x shown at the bottom.

expanded about the melting temperature T_m as a quadratic function

$$\Delta G(T) \approx -\Delta S(T_m)(T - T_m) - \Delta C_p(T_m)(T - T_m)^2/2T_m. \quad [1]$$

The equilibrium constant is $K = \exp[-\Delta G(T)/RT]$. K is equal to 1 at T_m but also at the cold denaturation temperature

$$T_{cd} = T_m\left(1 - \frac{2\Delta S(T_m)}{\Delta C_p(T_m)}\right) \quad [2]$$

in this quadratic model. T_{cd} typically lies well below 0°C (Fig. 1 shows both T_m and T_{cd}). The CD and fluorescence thermal denaturation curves show that there are no probe-dependent differences near T_m. Even the fastest folders in this study undergo the Bryngelson *et al.* (1) type 0 to type 1 (downhill to two-state) transition when the native state is destabilized by raising the temperature toward the heat denaturation transition.

T-Jump Kinetics. Protein folding kinetics were measured with our home-built laser T-jump instrument in 10 mM phosphate buffer. The observed relaxation was fitted by least squares to single exponential ($A_m = 0$) or stretched plus exponential models:

$$S(t) = (1 - A_m)e^{-k_a t} + A_m e^{-(k_m t)^\beta}. \quad [3]$$

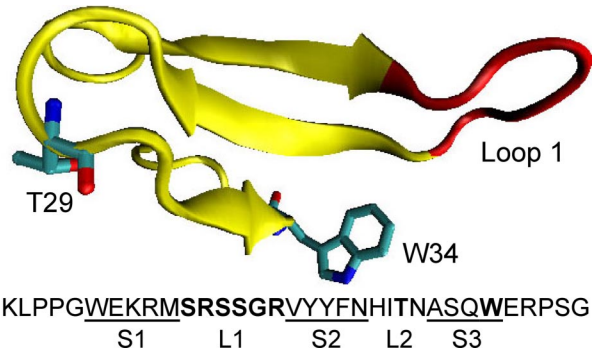

Fig. 2. Protein sequence and structure of wild-type hPin WW domain protein (Protein Data Bank entry 1PIN; variant 1 in Table 1). The mutation sites—loop 1 (L1), residue T29 at loop 2 (L2), and the residue W34—are highlighted as bolded letters in the sequence.

Table 1. Thirty-five proteins derived from hPin1 WW domain by substitutions in loop1 (residues 16–21), residue T29 in loop 2, and residue W34 in strand 3

No.	Sequence of residues 16–21 in loop 1	29	34	T_m, °C	T_{max}, °C	τ_f, μs
1‡	SRSSGR	T	W	59	45	69
2§	SRSSGR	T	F	57	38	56
3¶	SRSSGR	A	F	45	38	93
4‖	SRSSGR	D	W	43	32	96
5‖	SRSSGR	A	W	44	31	124
6‖	GRSSGR	T	W	48	39	236
7‖	ARSSGR	T	W	54	43	185
8‖	TRSSGR	T	W	53	46	333
9‖	SGSSGR	T	W	57	38	135
10‖	SRSEGR	T	W	55	41	200
11‖	SRSSGGR	T	W	48	40	325
12‖	SRSSGGGR	T	W	51	40	193
13¶	CSRSSGRC	A	F	53	54	13
14‖	S-SSGR	T	W	61	50	65
15‡	SRSS-R	T	W	51	36	758‡‡
16‡	S-SSGR	T	W	69	58	25
17‖	S-SSGR	A	W	56	36	19
18‡	S-ARGR	T	W	62	56	55‡‡
19‡	S-ADGR	T	W	78	64	12
20¶	S-ADGR	A	F	64	65	10
21§	S-ADGR	T	F	85	78	14‡‡
22‡	S-RDGR	T	W	74	67	13
23§	S-RDGR	T	F	78	72	13.9
24‡	S–SGR	T	W	62	41	45‡‡
25‡	S–RGR	T	W	63	58	51
26‡	S–NGR	T	W	68	55	20
27§	S–NGR	T	F	65	52	14.5
28¶	S–NGR	A	F	52	51	35.9
29¶	S–ᴅPGR	A	F	65	68	11
30§	S-ᴅA=AGR§§	T	F	56	50	105
31§	S-ᴅA=A-R§§	T	F	74	74	14
32§	–ᴅA=A–§§	T	F	21	13	221‡‡
33§	S-A=ᴅAGR§§	T	F	55	40	113‡‡
34§	S-A=ᴅA-R§§	T	F	51	36	160‡‡
35§	–A=ᴅA–§§	T	F	38	28	212

T_m is the melting temperature from a CD two-state fit. T_{max} is the temperature of the maximum folding rate. τ_f is the activated folding time at T_{max}. Folding times were obtained based on a two-state model as described in the text. The uncertainty of T_m and T_{max} is 0.5°C and 1°C, respectively. The relative uncertainty of τ_f is estimated to be <10%.

‡Kinetic/thermodynamic data from ref. 25.

§Kinetic/thermodynamic data from this work.

¶Kinetic/thermodynamic data from ref. 35.

‖Kinetic/thermodynamic data from ref. 39.

‡‡The highest measured folding rate is tabulated when the folding rate does not have the maximum value within the measured temperature range.

§§"ᴅA=A" and "A=ᴅA" represent trisubstituted alkene dipeptide isosteres of ᴅA-A and A-ᴅA, respectively. The isosteres replace amide bonds with Ψ[(*E*)-C(CH$_3$)=CH].

k_a and k_m are the observed activated and molecular rate coefficients, and β is the stretching factor for the molecular phase (1 for normal diffusion, <1 for anomalous diffusion). The slower fitted activated rate coefficient was analyzed by a two-state model $k_a = k_f + k_u$, where $k_f = k_a K/(1 + K)$ and k_u denote the activated folding and activated unfolding rate coefficients. The Arrhenius plots of the activated folding rate could all be fitted to quadratic polynomials, yielding a maximum folding rate coefficient $k_f(T_{max}) = \tau_f^{-1}$ at the temperature T_{max}, which lies between T_{cd} and T_m for almost all proteins (Table 1). As discussed previously, downhill folding dynamics characterized by

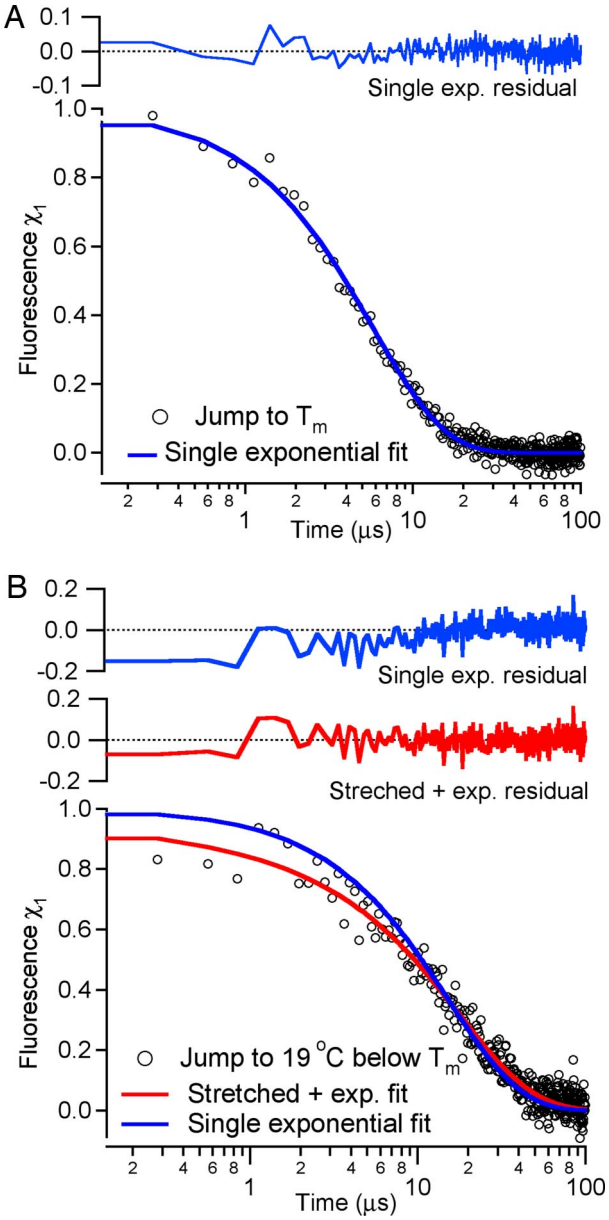

Fig. 3. Protein relaxation kinetics of variant 20 recorded with a time resolution of 280 ns. (*A*) Relaxation is fitted very well to a single exponential decay when temperature jumped to T_m. (*B*) The stretched plus exponential fit becomes better than a single exponential fit when the temperature jumps to 19°C below T_m.

k_m can be stretched or simple exponential (9, 17, 26, 27). A fit within the experimental signal-to-noise ratio required $\tau_m = 1.4$ μs and the stretching factor $\beta = 0.4$ in Fig. 3. Reasonable fits could be obtained by setting $\beta = 1$, thus characterizing the molecular phase by an average time constant. The resulting $\tau_m = 1/k_m$ are listed in Table 2 for the four molecules with $A_m > 10\%$. Note that under incipient downhill conditions, where the fast molecular phase k_m takes over from the activated phase k_a, $K \gg 1$ and the downhill folding rate equals k_m according to ref. 6.

Correlation Between Maximal Folding Rate and T_m. Mutants with higher stability have higher maximum folding rates on average (Fig. 4). The correlation between $\ln(\tau_f(T_{max}))$ and T_m is sigmoidal, as indicated by the dotted envelopes. A transition to faster folding occurs at $T_m^* \approx 50$°C. Near T_m^*, as indicated by the vertical/red dotted line in Fig. 4, the folding time $\tau_f(T_{max})$ has a very broad distribution, ranging from 10 μs (variant 20) to 758 μs (variant 15). The distribution of $\tau_f(T_{max})$ is much narrower above this temperature. Folding times above T_m^* are on average more than an order of magnitude faster than for proteins below that transition temperature. The minimum activated folding time of all of the mutants is 10 μs (variant 20 with a T_m of 64°C). The folding time of variant 21 with the highest T_m (85°C) is 14 μs at the lowest temperature where we could measure it.

Table 2. Incipient downhill folders: Temperature of fitted kinetics in Fig. 3, molecular time scale, and activated folding time

No.	T, °C	τ_m, μs	τ_f, μs
13	44	5.0 (0.8)	25.0 (1.7)
20	45	1.2 (0.5)	19.7 (0.5)
20	53	4.7 (1.1)	17.5 (2.0)
23	64	1.5 (0.3)	15.8 (1.6)
31	67	2.1 (1.0)	14.5 (1.5)

Parentheses are estimated fitting uncertainties.

Four Among the Fastest Folders Make a Transition from Exponential to Nonexponential Kinetics at High Native Bias. At temperatures below T_m, but well above the cold denaturation temperature T_{cd}, we identified a transition from exponential to nonexponential kinetics for four mutants. All of them have high melting temperatures. At their T_m, these proteins fold with single exponential relaxation kinetics. Nonexponential relaxation begins when the temperature is lowered at least 10–20°C below T_m. At even lower temperature, the cold denaturation process illustrated in Fig. 1 presumably would lead back to two-state folding, but we cannot measure kinetics at the required subfreezing temperatures.

Fig. 3 shows the onset of a single exponential-to-nonexponential transition: Protein variant 20 has a minimum folding time of 10 μs near its T_m. Its relaxation kinetics can be fitted to a single exponential at T_m. When the temperature decreases to 19°C

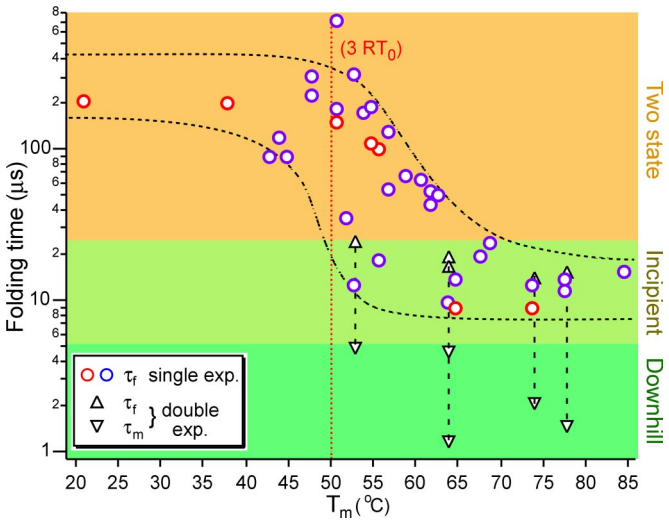

Fig. 4. The correlation between the minimum activated folding time (τ_f at the temperature of maximum folding rate) versus the melting temperature T_m. All of the data are shown in Table 1. Blue circles are variants 1–28, and red circles are variants 29–35 measured here, which include non-natural amino acids loops. The triangles correspond to double exponential fits to Eq. **3** for the four proteins of Table 2. The upper triangle of a pair is the activated folding time, and the lower triangle of a pair is the molecular (downhill) time scale.

below T_m, the double exponential begins to fit the data significantly better than the single exponential fit. For pure downhill folding, the prediction (9) is that the activated and molecular phases merge. We observe only incipient downhill folding (barrier <3 RT), where both phases are still observed, at the temperatures we were able to reach. The fitted molecular rate ranges from $(1–5\ \mu s)^{-1}$, as summarized in Table 2.

The Molecular Phase for Diffusive Relaxation Becomes Faster at Higher Temperature. The fitted stretching coefficient ranges from 0.4 to 1. If we fix it to 1, the molecular rate increases from $k_m = (4–5\ \mu s)^{-1}$ to $(1–2\ \mu s)^{-1}$ at higher temperatures (Table 2). The increase is slightly larger than the viscosity scaling expected in the normal Kramers regime ($k_m \sim \eta^{-1}$). Solvent viscosity scaling only accounts for a decrease of about 40% between 40°C to 70°C. However, the k_m values extracted are not reliable enough currently to allow a quantitative comparison with models of viscosity scaling for the prefactor (28).

Discussion

Downhill folding has been investigated experimentally by looking at thermodynamic or kinetic anomalies of a few individual proteins. A statistical analysis covering a wide range of stabilities and folding rates reported here using 35 WW domains yields more concrete insight into the requirements for downhill folding. The salient observations that need to be explained are as follows. Why is there a change from only two-state folders being observed at $T_m < 50$°C to incipient downhill folding at T_m between 50 and 85°C? Why is this change so sharp? And why has no pure downhill folder been observed below $T_m = 85$°C? We begin the discussion with a model that allows us to define protein stability zones for "two-state," "incipient downhill," and "downhill" folding. Finally, we discuss some of the features of the molecular rate coefficients k_m and their association with downhill folding in more detail.

Fig. 4 shows the two-state, incipient downhill, and downhill zones according to the rate criterion of Yang and Gruebele (6). Proteins with a single slow phase k_a are apparent two-state folders. Proteins showing two phases (triangles in Fig. 4) are incipient downhill folders: The slow phase is the remnant of activated folding, and the fast molecular phase k_m results from a sizeable protein population diffusing over a low barrier (<3 RT) (6, 9). Proteins with a single fast phase k_m would be pure downhill folders.

It appears from Fig. 4 that such zones can also be defined by a single thermodynamic criterion; namely, T_m. The transition at ≈50°C suggests that incipient downhill folding cannot be observed below that temperature. Moreover, the transition to complete downhill folding appears to lie at or above 85°C, because even the highest T_m folder shows traces of both activated and molecular phases.

A fitting model that includes both heat denaturation and cold denaturation explains why less stable proteins cannot be made to fold downhill. As illustrated in Fig. 1 and Eq. **1**, the native bias is a quadratic function of temperature: Proteins not only heat-denature at T_m, but also cold-denature at T_{cd} (Eq. **2**). If a protein is very stable, the region between the two denaturation temperatures is wide enough to allow downhill folding near its center. If a protein has a low melting point, then the region between T_m and T_{cd} is small, so a strong native bias cannot be achieved, and the protein might even remain a two-state folder everywhere between T_m and T_{cd}. This hypothesis was introduced earlier, qualitatively, by Sabelko and coworkers (3); however, the present survey provides sufficient data to test this notion quantitatively. Specifically, we show that the characteristic temperatures $T_m^* \approx 50$°C (two-state to incipient downhill) and $T_m^* \geq 85$°C (incipient downhill to pure downhill) can be predicted from our data.

We approximate the free energy profile as a function of a collective reaction coordinate x (not specified structurally) and temperature T by a sum of two terms:

$$G(x, T) = G_{T_m}(x) + \Delta G(T)x/2. \quad [4]$$

The first term represents a symmetric double-well at T_m (Fig. 1, bottom), when $\Delta G = 0$:

$$G_{T_m}(x) = \Delta G^{\dagger}x^4 - 2\Delta G^{\dagger}x^2. \quad [5]$$

$\Delta G^{\dagger}$ is the barrier between the native and denatured states when $\Delta G = 0$. The second term in Eq. **4** is the free energy difference between the folded and denatured ensembles, assuming the reaction coordinate x is normalized so $x = -1$ in the denatured states and $x = +1$ in the native state. Because heat and cold denaturation are both possible, $\Delta G(T)$ is a quadratic function of temperature (Eq. **1**).

Finally, we need an expression for the activation energy $\Delta G^{\dagger}$ in Eq. **5**. A plot of the activated folding $k_f^{-1}(T_m)$ in Fig. 5*A* shows that $\Delta G^{\dagger}$ decreases as the melting temperature of the protein increases. We approximate its dependence on the T_m by

$$\Delta G^{\dagger} = \Delta G_0^{\dagger} + \lambda^{\dagger}(T_m - T_0) + \Delta G_{random}^{\dagger}. \quad [6]$$

The first term is the average barrier of mutants at $T_0 = 298$ K (for reference), $\lambda^{\dagger}$ describes the linear dependence on the melting temperature T_m, and $\Delta G_{random}^{\dagger}$ describes random sequence-dependent fluctuations of the barrier.

Numerical values for the constants in Eqs. **1**, **2**, and **4**-**6** can be estimated quite accurately from the experimental data. Neglecting random fluctuations, $\Delta G^{\dagger} \approx 5.5 - 0.046\ (T_m - T_0)$ from Fig. 5*A* (units of RT_0). WW domains have similar cooperativity and reach the maximum stability at ≈5°C; thus, $T_{cd} \approx 10$°C $- T_m$ according to Fig. 1. We obtain $\Delta S(T_m) \approx -0.11 - 6.8 \cdot 10^{-4}\ T_m$ (in units of RT_0/Kelvin) from the thermodynamic fit of Eq. **1** to the circular dichroism data. The sign of the slope is as expected if hydrophobic interactions are weakened at lower temperature. ΔC_p can be evaluated from Eq. **2**.

To compute the downhill folding temperature, $T_{downhill}$, for a given T_m, one evaluates Eq. **4** and finds the temperature for which the folding barrier drops below 3 RT_0 (incipient downhill folding) or for which the barrier is exactly equal to 0 (completely downhill folding). Fig. 5*B* shows the resulting plot of $T_{downhill}$ vs. T_m for incipient and completely downhill folding, with cold denaturation included (solid curves) or neglected (dashed lines). When the possibility of cold denaturation is neglected, downhill folding always occurs at sufficiently low temperature. When cold denaturation is included, our model yields $T_m^* \approx 50$°C for incipient downhill folding (red curve in Fig. 5*B*) and $T_m^* \approx 90$°C for complete downhill folding (blue curve in Fig. 5*B*).

These results agree well with the experimental survey: No incipient downhill folders are observed that have a T_m below 52°C, and no complete downhill folders are observed up to a T_m of 85°C. In essence, the onset cold denaturation prevents downhill folding from occurring in insufficiently stable proteins. The two free energy surfaces in Fig. 5*C* computed from our model illustrate the extremes of behavior. Fig. 5*C Upper* corresponds to a melting temperature below the T_m^* for incipient downhill folding, and a barrier >3 RT_0 occurs at all temperatures between T_{cd} and T_m. The surface in Fig. 5*C Lower* corresponds to $T_m \approx 90$°C $> T_m^*$ for completely downhill folding, and no barrier occurs near $(T_{cd} + T_m)/2$, the temperature of maximal stability. For $T > 1.1\ T_m$ or $T < 0.8\ T_{cd}$, downhill unfolding is obtained.

Shen *et al.* (29) recently studied downhill folding and unfolding temperatures for a large number of two-state folders with a native structure-based model that excludes cold denaturation. Fig. 5*B* (blue dotted line) shows our fitting results with cold

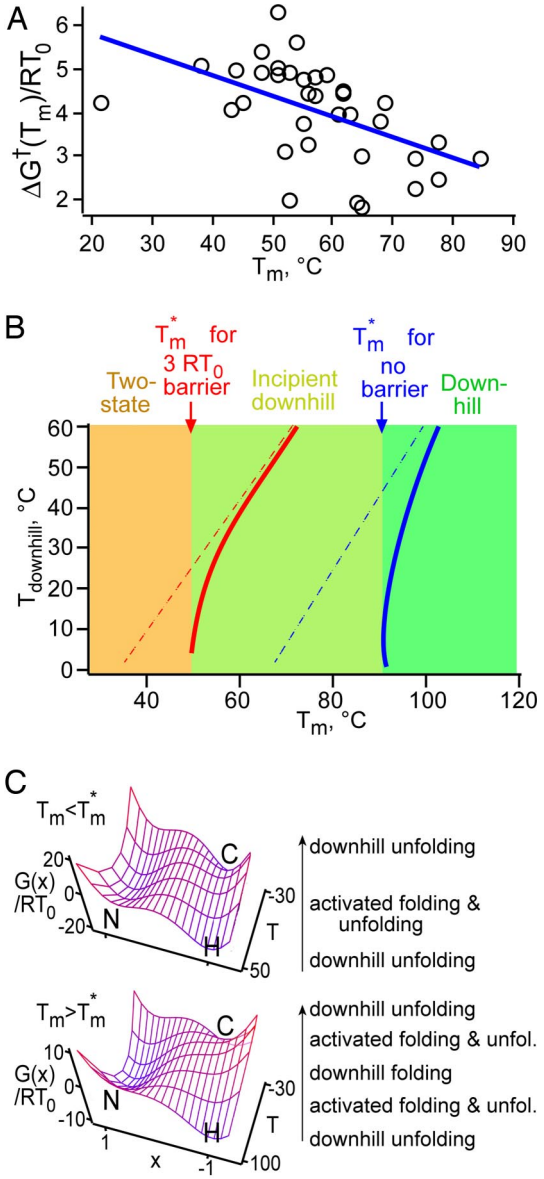

Fig. 5. Two-state to downhill folding transition model. (*A*) The activation barrier at T_m of the 35 variants of WW protein, $\Delta G^{\ddagger}$, is approximated by a linear function of T_m. (*B*) Calculated $T_{downhill}$ for incipient downhill folding (3-RT_0 barrier, red) and complete downhill folding (0-RT_0 barrier, blue) as a function of melting temperature T_m based on the model. (*C*) Free energy surfaces of the proteins with low melting temperature and high melting temperature. Only proteins of sufficient stability (naturally biased proteins) undergo complete downhill folding. N, native; C, cold denatured; H, heat denatured state.

denaturation excluded (no quadratic term in ΔG). We find that for a WW domain of typical stability ($T_m \approx 50$°C) to fold downhill, $T_{downhill}/T_m \approx 229$ K/323 K ≈ 0.71, compared with their estimate of <0.58 for the proteins they studied. Our estimate for the onset of downhill heat denaturation is 1.29 T_m, compared with their prediction of 1.25 T_m. The upper temperature ratio is in very good agreement between experiment and theory, while downhill folding turns out to be significantly easier experimentally than the calculation would indicate. It will be very interesting to see what results are obtained from statistical mechanical folding models when the temperature dependence for the interaction parameters is included.

We conclude our discussion with the properties of the molecular rate coefficient k_m. The best fits (e.g., Fig. 3) were obtained by allowing anomalous diffusion with $\beta \leq 1$, but to characterize downhill folding by a single average rate coefficient, we report fits with β fixed to 1 in Table 2. According to Yang and Gruebele (6), the average rate coefficient can be taken as the prefactor for activated kinetics. Values greater than 0.1 μs indicate reduced diffusion caused by longitudinal (along the reaction coordinate) or transverse (along other coordinates) roughness of the free energy surface (30). In a recent article, Qi and Portman (31) enhanced the ability of native topology-based folding models to predict the full range of folding rates by addition of excluded volume effects. From this model, they estimate a more realistic range of prefactors for the collective reaction coordinate, 1–5 μs. Our range of fitted molecular phase rate coefficients for the four proteins in Table 2 is in good agreement with this estimate, as well as earlier models and experimental estimates of the "speed limit" (32–34).

The fitted k_m values of this small β-sheet protein (34 residues) are comparable to those of the larger helical downhill folder λ^*_{6-85} (82 residues). They are slower at lower temperature, in qualitative agreement with viscosity-dependent scaling of the diffusion coefficient. The scatter in Table 2 is too large to fit any specific viscosity model. For the folding speed limit, both linear scaling with chain length (N/100 μs) (2) and exponential scaling with contact order (18) have been proposed. The $k_m \approx$ 1–2 μs observed for the helical λ^*_{6-85} is very close to the limit of 0.8 μs estimated by linear scaling. For the Pin WW β-sheet domain, the measured molecular phase ranges from 1 to 5 μs, 3–10 times slower than the linear chain length scaling model (0.36 μs).

In summary, we have derived a simple thermodynamic model from which melting temperature zones for two-state folding and downhill folding can be determined. Application of the model to the folding kinetics of hPin WW domain yields quantitative agreement with the experimental minimum temperature required for incipient downhill folding (50°C) and complete downhill folding (≥85°C). It remains to be tested how generally this model applies to the boundary between two-state and downhill protein folding.

Materials and Methods

Protein Engineering. Our main focus in the survey is on loop 1 of the Pin WW domain (entry 1, Table 1). This loop adopts a rare (4:6) loop conformation (Fig. 2), harboring a type-II four-residue turn sequence (–Ser-16–Arg-17–Ser-18–Ser-19–) incorporated into the six-residue loop sequence (–Ser-16–Arg-17–Ser-18–Ser-19–Gly-20–Arg-21–), apparently for functional reasons (25). Sequence alignments of over 100 WW domain family members reveal that a five-membered type 1 G1 bulge turn is the most common loop 1 structure. Given the importance of loop 1 in the folding transition state of numerous WW domain sequences, we have varied this loop in several studies to four-, five-, and six-residue loop sequences to understand structure–function–folding relationships in WW domains (25, 35). Some Pin-derived WW domain sequences exhibit an additional slower concentration-dependent phase, apparently resulting from transient reversible oligomer formation; thus, we have used Thr29Ala and/or W34F mutations to eliminate this phase. Table 1 depicts several six-, five-, and four-residue sequences prepared to test numerous hypotheses about the importance of this region in folding and function. The last six entries are part of a more extensive survey of the efficacy of β-turn mimetics prepared by the organic chemistry community over the last two decades (A.A.F., D.D., F.L., J.E.D., Gerard Kroon, Evan T. Powers, P.W., M.G., and J.W.K., unpublished results). In particular, the last six WW sequences (entries 30–35) incorporate trisubstituted *E*-olefin dipeptide isosteres in place of two α-amino acid residues. Protein expression and purification methods are as described in ref. 25. Protein identity was confirmed by low-resolution ESI and MALDI mass spectrometry, and purity was established chromatographically. For variants 30–35, trisubstituted alkene dipeptide isosteres in which the amide bond was replaced with Ψ[(*E*)-C(CH_3)=CH] were incorporated into the protein by using methods described in refs. 36 and 37.

Thermal Titrations and Melting Temperatures. Circular dichroism (CD)-detected thermal titration curves were obtained by monitoring at 227 nm. Lyophilized protein samples were dissolved in 10 mM phosphate buffer (pH 7) with concentrations near 10 μM. The temperature scans ranged from 2°C to 108°C (under an oil film for mutants with $T_m >$ 75°C).

Far-UV CD spectra were similar for all of the mutants under native conditions (data not shown). Furthermore, the crystal structures of several mutants (variant 19 and 26) have been solved and are superimposable on the wild-type Pin WW domain structure, except for loop 1 (25), the substructure of which was purposefully altered. Substantial biophysical data not depicted here are also consistent with the hypothesis that all mutants conserve the native structure of the wild-type Pin1 WW domain, except for local perturbations in loop 1 upon modification of its sequence.

Laser T-Jump Kinetics. Protein folding kinetics were measured in 20% D_2O/80% H_2O, 10 mM phosphate buffer. The concentration of the protein samples ranged from 80 to 150 μM to avoid any transient aggregation (6). A Raman-shifted YAG laser pulse generated a temperature jump of 5–12°C within several nanoseconds. The relaxation of the protein to the new equilibrium was probed by tryptophan fluorescence following excitation of the protein by 280-nm UV laser pulses. The fluorescence decays were digitized with 500-ps time resolution. Several decays were averaged, and data were binned into 140- to 280-ns bins to improve the signal-to-noise ratio. The evolution of the fluorescence lifetime profile was extracted by χ analysis as described in ref. 38, to reveal deviations from two-state folding via nonexponential decays (Fig. 3). The temperature range of the experiments was limited by the low signal-to-noise ratio at temperatures far below T_m and by photoacoustic cavitation at temperatures well above T_m.

ACKNOWLEDGMENTS. We thank Drs. Houbi Nguyen and Marcus Jäger for measuring the kinetics and thermodynamics of WW domains from ref. 39. This work was supported by National Science Foundation Grant MCB 0313643 (to F.L. and M.G.), National Institutes of Health Grant GM 051105, the Skaggs Institute of Chemical Biology (D.D., A.A.F., and J.W.K.), the Lita Annenberg Hazen Foundation (D.D., A.A.F., and J.W.K.), and National Institutes of Health P50 Grant GM067082 (to J.E.D. and P.W.). A.A.F. is supported by a Ruth L. Kirschstein National Research Service Award fellowship.

1. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) funnels, pathways, and the energy landscape of protein folding: A Synthesis. *Proteins Struct Funct Genet* 21:167–195.
2. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding ''speed limit.'' *Curr Opin Struct Biol* 14:76–88.
3. Sabelko J, Ervin J, Gruebele M (1999) Observation of strange kinetics in protein folding. *Proc Natl Acad Sci USA* 96:6031–6036.
4. Leeson DT, Gai F, Rodriguez HM, Gregoret LM, Dyer RB (2000) Protein folding on a complex energy landscape. *Proc Natl Acad Sci USA* 97:2527–2532.
5. Garcia-Mira M, Sadqi M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) Experimental identification of downhill protein folding. *Science* 298:2191–2195.
6. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423:193–197.
7. Zhu Y, Alonso DOV, Maki K, Huang, C.-Y., Lahr SJ, Daggett V, Roder H, DeGrado WF, Gai F (2003) Ultrafast folding of alpha-3-D, a *de novo* designed three-helix bundle protein. *Proc Natl Acad Sci USA* 100:15486–15491.
8. Ma H, Gruebele M (2005) Kinetics are probe-dependent during downhill folding of an engineered lambda-6–85 protein. *Proc Natl Acad Sci USA* 102:2283–2287.
9. Yang W, Gruebele M (2004) Folding lambda repressor at its speed limit. *Biophys J* 87:596–608.
10. Sadqi M, Fushman D, Muñoz V (2006) Atom-by-atom analysis of global downhill protein folding. *Nature* 442:317–321.
11. Naganathan AN, Perez-Jimenez R, Sanchez-Ruiz JM, Munoz V (2005) Robustness of downhill folding: Guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* 44:7435–7449.
12. Muñoz V (2002) Thermodynamics and kinetics of downhill protein folding investigated with a simple statistical mechanical model. *Int J Quant Chem* 90:1522–1528.
13. Liu F, Gruebele M (2007) Tuning λ6–85 towards downhill folding at its melting temperature. *J Mol Biol* 370:574–584.
14. Sadqi M, Fushman D, Muñoz V (2007) Structural biology—Analysis of protein-folding cooperativity—Reply. *Nature* 445:E16–E18.
15. Ferguson N, Sharpe TD, Johnson CM, Schartau PJ, Fersht AR (2007) Structural biology—Analysis of 'downhill' protein folding. *Nature* 445:E14–E16.
16. Hagen S (2007) Probe-dependent and nonexponential relaxation kinetics: Unreliable signatures of downhill protein folding. *Proteins Struct Funct Bioinformatics* 68:205–217.
17. Gruebele M (2007) Comment on probe-dependent and nonexponential relaxation kinetics: Unreliable signatures of downhill protein folding. *Proteins Struct Funct Bioinformatics* 70:1099–1102.
18. Gruebele M (2005) Downhill protein folding: Evolution meets physics. *Comp Rend Biol* 328:701–712.
19. Yang WY, Gruebele M (2004) Rate-temperature relationships in lambda repressor fragment 6–85 folding. *Biochemistry* 43:13018–13025.
20. Dumont C, *et al.* (2006) Solvent-tuning collapse and helix formation time scales of lambda6–85. *Protein Sci* 15:2596–2604.
21. Bruscolini P, Pelizzola A, Zamparo M (2007) Downhill versus two-state protein folding in a statistical mechanical model. *J Chem Phys* 126:215103.
22. Cho SS, Weinkam P, Wolynes PG (2008) Origin or barriers and barrierless folding in BBL. *Proc Natl Acad Sci USA* 105:118–123.
23. Nguyen H, Jager M, Moretto A, Gruebele M, Kelly JW (2003) Tuning the energy landscape of a WW domain by mutation, temperature tuning, and truncation. *Proc Natl Acad Sci USA* 100:3948–3953.
24. Jäger M, Nguyen H, Crane J, Kelly J, Gruebele M (2001) The folding mechanism of a beta-sheet: The WW domain. *J Mol Biol* 311:373–393.
25. Jäger M, *et al.* (2006) The structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci USA* 108:10648–10653.
26. Hagen SJ (2003) Exponential decay kinetics in ''downhill'' protein folding. *Proteins Struct Funct Genet* 50:1–4.
27. Ma H, Gruebele M (2005) Low barrier kinetics: Dependence on observables and free energy surface. *J Comput Chem* 27:125–134.
28. Qiu LL, Hagen SJ (2005) Internal friction in the ultrafast folding of the tryptophan cage. *Chem Phys* 312:325.
29. Shen T, Zong C, Portman JJ, Wolynes PG (2008) Variationally determined free energy profiles for structural models of proteins: Characteristic temperatures for folding and trapping. *J Phys Chem B*, in press.
30. Portman JJ, Takada S, Wolynes PG (2001) Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach. *J Chem Phys* 114:5069–5081.
31. Qi X, Portman JJ (2007) Excluded volume, local structural cooperativity, and the polymer physics of protein folding rates. *Proc Natl Acad Sci USA* 104:10841–10846.
32. Hagen SJ, Hofrichter J, Szabo A, Eaton WA (1996) Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc Natl Acad Sci USA* 93:11615–11617.
33. Chang IJ, Lee JC, Winkler JR, Gray HB (2003) The protein-folding speed limit: Intrachain diffusion times set by electron-transfer rates in denatured Ru(NH3)(5)(His-33)-Zn-cytochrome c. *Proc Natl Acad Sci USA* 100:3838–3840.
34. Camacho CJ, Thirumalai D (1995) From minimal models to real proteins: Time scales for folding kinetics. *J Physique I* 5:1457–1467.
35. Nguyen H, Jaeger M, Kelly JW, Gruebele M (2005) Engineering a beta-sheet protein towards the folding speed limit. *J Phys Chem B* 109:15182–15186.
36. Wipf P, Pierce JG (2006) Expedient synthesis of the alpha-C-glycoside analogue of the immunostimulant galactosylceramide. *Org Lett* 8:3375–3378.
37. Fu Y, Gao J, Bieschke J, Dendle MA, Kelly JW (2006) Amide-to-E-olefin versus amide-to-ester backbone H-bond perturbations: Evaluating the O-O repulsion for extracting H-bond energies. *J Am Chem Soc* 128:15948–15949.
38. Ballew RM, Sabelko J, Reiner C, Gruebele M (1996) A single-sweep, nanosecond time resolution laser temperature-jump apparatus. *Rev Sci Instrum* 67:3694–3699.
39. Nguyen H (2004) PhD Thesis (Univ of Illinois at Urbana–Champaign, Urbana), p 320.