Protein folding kinetics and thermodynamics from atomistic simulation

Stefano Piana1,2, Kresten Lindorff-Larsen1,2, and David E. Shaw3,2

1D. E. Shaw Research, New York, NY 10036; and 2Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10032

Edited by Peter G. Wolynes, Rice University, Houston, TX, and approved June 13, 2012 (received for review March 2, 2012)

Advances in simulation techniques and computing hardware have created a substantial overlap between the timescales accessible to atom-level simulations and those on which the fastest-folding proteins fold. Here we demonstrate, using simulations of four variants of the human villin headpiece, how simulations of spontaneous folding and unfolding can provide direct access to thermodynamic and kinetic quantities such as folding rates, heat capacities, pre-exponential factors, and temperature-jump relaxation profiles. The quantitative comparison of simulation results with various forms of experimental data probing different aspects of the folding process can facilitate robust assessment of the accuracy of the calculations while providing a detailed structural interpretation for the experimental observations. In the example studied here, the analysis of folding rates, pre-exponential factors, and unfolding pathways provides support for the notion that a norleucine double mutant of villin folds five times faster than the wild-type sequence, but following a slightly different pathway. This work showcases how computer simulation has now developed into a mature tool for the quantitative computational study of protein folding and dynamics that can provide a valuable complement to experimental techniques.

Amber ff99SB*-ILDN | enthalpy | heat capacity | pre-exponential factor | transition path time

Proteins are synthesized in the cell or in vitro as unstructured polypeptide chains that, in most cases, self-assemble into their functionally active three-dimensional shapes. This process, called protein folding, occurs on a broad range of timescales ranging from microseconds to seconds and higher. From a purely physical-chemical perspective, it should be possible in principle to characterize the folding mechanism of a given protein at atomistic resolution and to reconstruct its free-energy landscape, given only its primary sequence, through molecular dynamics (MD) simulations based on elementary physical principles. This direct approach has been rarely pursued because even the simplest systems representing a protein immersed in water consist of several thousand atoms, and simulating their behavior on the timescales typical of protein folding is computationally extremely demanding. The discovery and design of fast-folding proteins (1) significantly narrowed the timescale gap between simulations and experiments, making such simulations feasible, at least for the fastest-folding proteins.

The C-terminal fragment of the villin headpiece [referred to in the remainder of this paper simply as “villin” (2)], one of the fastest-folding protein domains known (3), has proven to be an excellent target for folding simulations with physics-based force fields and an atomistically detailed representation of both the solute and the surrounding solvent (4–7). Until recently, the length of such simulations was limited to a few microseconds—a timescale sufficient to capture, at best, a single folding event (8, 9). With this limitation, it has been difficult to directly connect the data produced by short, non-equilibrium simulations to experimental observations, unless sufficient statistics were generated to allow the construction of coarse-grained kinetic models that approximate the underlying folding dynamics (10, 11). While these models have proven useful for obtaining certain insights into the folding mechanism, it has been difficult to produce consistent predictions of the thermodynamics of villin folding (10–13), and where comparison has been possible between kinetic models built from multiple short simulations and independent, long-equilibrium MD simulations (14), substantial differences have been observed in the folding free energies.

Recent advances in computer hardware have, however, extended the timescale accessible to simulation up to the millisecond (15), thus creating a broad overlap with the microsecond timescale characteristic of fast-folding proteins such as villin, and allowing for the direct calculation of equilibrium thermodynamic and kinetic properties from the simulation data (16, 17). Substantial improvements have also been made in the molecular mechanics force fields used in MD simulations (14, 18–20).

Taking advantage of improvements in both simulation speeds and force fields, we have employed equilibrium MD simulations to study the folding kinetics and thermodynamics of villin. Some researchers have suggested that temperature-jump experiments may underestimate the folding time of villin (4, 5) and may be more sensitive to processes other than protein folding. It should be noted, however, that some of these theoretical estimates were based on models that combine the results of non-equilibrium, short, individual trajectories, thereby introducing additional sources of uncertainty in the quantitative comparisons between the simulation results and the experimental observations. The equilibrium simulations presented here overcome this difficulty as in each trajectory at least 30 folding and unfolding events are observed, making it possible to directly compute thermodynamic and kinetic quantities without the need to build approximate models to describe the system. Further, we compared simulations of different variants allowing the direct calculation of ΔΦ-values in a manner analogous to experiments (17, 21).

Most of the results presented here are in good agreement with previous experimental findings, with the notable exception of the heat capacity for folding, which appears to be smaller than the value extracted from calorimetric data. We find that the double norleucine (Nle/Nle) mutant (3) folds approximately five times faster than the wild-type protein, thus supporting the original interpretation of the experimental data (3). In agreement with our previous observations (14), the results reported here also indicate that both the number of helical residues and the Trp side-chain environment are sensitive to the folding/unfolding process, supporting the notion that experiments that probe these quantities, like infrared (IR) and fluorescence-detected temperature-jump, may be used to determine folding and unfolding rates (22).

Author contributions: S.P., K.L.-L., and D.E.S. designed research; S.P. and K.L.-L. analyzed data; and S.P., K.L.-L., and D.E.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

S.P. and K.L.-L. contributed equally to this work.

To whom correspondence may be addressed. E-mail: Stefano.Piana-Agostinetti@DEShawResearch.com or Kresten.Lindorff-Larsen@DEShawResearch.com or David.Shaw@DEShawResearch.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201811109/-/DCSupplemental.
Results and Discussion

Equilibrium Reversible Folding Simulation of the Villin Headpiece C-terminal Fragment. We validated that the Amber ff99SB-ILDN (18, 19, 23) force field appears to be reasonably transferable across different protein classes (SI Text) and used it to investigate computationally the kinetics and thermodynamics of villin folding. In particular, we performed equilibrium MD simulations, at multiple temperatures, of the human villin C-terminal fragment (24), the Nle/Nle double mutant (3), and variants where we introduced an F10L mutant into either the wild-type protein or the Nle/Nle variant. Each simulation was run for at least 300 μs and contained between 30 and 150 folding and unfolding events (Table 1).

The simulation trajectories were analyzed using the stable state approximation (25) and partitioned into folded and unfolded segments using a transition-based assignment (14, 26). The folding free energy (ΔGf, kcal mol⁻¹) is calculated from the ratio of the folded and unfolded fractions. Folding enthalpies (ΔHf, kcal mol⁻¹) were calculated either from the folding free energy at different temperature using the van’t Hoff equation (27) or as ΔH = ΔU + VΔP where ΔU is the difference in average force-field energy in the folded and unfolded states (MD). The heat capacities (ΔCv, kcal mol⁻¹ K⁻¹) were calculated either from the difference in fluctuations of the force-field energy between the folded and unfolded states (MD) or from the temperature dependency of the folding enthalpy (ΔΔH). The folding time (τf) is calculated as the average waiting time in the unfolded state. The pre-exponential factor for folding (k0) is estimated from the folding time and the mean transition path time (τTP) using Kramers’ theory (40).

The trajectories have been partitioned into folded and unfolded segments using a transition-based assignment (14, 26). The folding free energy can be compared to the folding enthalpies (ΔHf, kcal mol⁻¹) measured value of 14, 26. The folding free energy at different temperature using the van’t Hoff analysis for the Nle/Nle double mutant (3). The temperature of each MD simulation is reported together with the total length (L) and the number of observed folding and unfolding events (n). The trajectories have been partitioned into folded and unfolded segments using a transition-based assignment (14, 26). The folding free energy (ΔGf, kcal mol⁻¹) is calculated from the ratio of the folded and unfolded fractions. Folding enthalpies (ΔHf, kcal mol⁻¹) were calculated either from the folding free energy at different temperature using the van’t Hoff equation (27) or as ΔH = ΔU + VΔP where ΔU is the difference in average force-field energy in the folded and unfolded states (MD). The heat capacities (ΔCv, kcal mol⁻¹ K⁻¹) were calculated either from the difference in fluctuations of the force-field energy between the folded and unfolded states (MD) or from the temperature dependency of the folding enthalpy (ΔΔH). The folding time (τf) is calculated as the average waiting time in the unfolded state. The pre-exponential factor for folding (k0) is estimated from the folding time and the mean transition path time (τTP) using Kramers’ theory (40).

Table 1. Folding kinetics and thermodynamics from equilibrium MD simulations of wild-type (WT) and two variants of villin headpiece C-terminal fragment

<table>
<thead>
<tr>
<th>Variant</th>
<th>ΔGf (K)</th>
<th>ΔHf (K)</th>
<th>ΔCv (K)</th>
<th>τf (μs)</th>
<th>k0 (μs⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT (HP-35)</td>
<td>345</td>
<td>398</td>
<td>30</td>
<td>0.8(2)</td>
<td>15(4)</td>
</tr>
<tr>
<td>WT (HP-35)</td>
<td>360</td>
<td>319</td>
<td>31</td>
<td>1.6(2)</td>
<td>18(3)</td>
</tr>
<tr>
<td>WT-F10L</td>
<td>345</td>
<td>371</td>
<td>30</td>
<td>1.6(3)</td>
<td>10(2)</td>
</tr>
<tr>
<td>Nle/Nle</td>
<td>360</td>
<td>305</td>
<td>61</td>
<td>0.6(2)</td>
<td>16(3)</td>
</tr>
<tr>
<td>Nle/Nle-F10L</td>
<td>370</td>
<td>395</td>
<td>150</td>
<td>0.0(1)</td>
<td>18(2)</td>
</tr>
<tr>
<td>Nle/Nle-F10L</td>
<td>380</td>
<td>301</td>
<td>140</td>
<td>0.7(1)</td>
<td>21(2)</td>
</tr>
<tr>
<td>Nle/Nle-F10L</td>
<td>360</td>
<td>301</td>
<td>110</td>
<td>0.4(1)</td>
<td>14(7)</td>
</tr>
<tr>
<td>Nle/Nle-F10L</td>
<td>370</td>
<td>300</td>
<td>130</td>
<td>0.8(1)</td>
<td>16(1)</td>
</tr>
</tbody>
</table>

The temperature of each MD simulation is reported together with the total length (L) and the number of observed folding and unfolding events (n). The trajectories have been partitioned into folded and unfolded segments using a transition-based assignment (14, 26). The folding free energy (ΔGf, kcal mol⁻¹) is calculated from the ratio of the folded and unfolded fractions. Folding enthalpies (ΔHf, kcal mol⁻¹) were calculated either from the folding free energy at different temperature using the van’t Hoff equation (27) or as ΔH = ΔU + VΔP where ΔU is the difference in average force-field energy in the folded and unfolded states (MD). The heat capacities (ΔCv, kcal mol⁻¹ K⁻¹) were calculated either from the difference in fluctuations of the force-field energy between the folded and unfolded states (MD) or from the temperature dependency of the folding enthalpy (ΔΔH). The folding time (τf) is calculated as the average waiting time in the unfolded state. The pre-exponential factor for folding (k0) is estimated from the folding time and the mean transition path time (τTP) using Kramers’ theory (40).
is not necessarily an intrinsic property of all proteins simulated with this force field. In all cases, it should be noted that, while errors in the enthalpy and heat capacity affect the temperature-dependence of calculated properties, most of this study was carried out in a relatively narrow range of temperatures (between 345 and 380 K), and we do not observe dramatic deviations from experiment when comparing simulations performed in this limited range.

**Kinetics.** As described above, we used a transition-based assignment (14, 26) to define folded and unfolded segments of the trajectories, and calculated folding and unfolding rates from the waiting times in these states (Table 1). Experimental measurements of folding and unfolding rates are, however, generally based on observing the time-resolved relaxation of a spectroscopic signal after a small external perturbation; for fast-folding proteins this is typically a sudden increase in temperature of 5–10 K. For a two-state folder, the relaxation rate obtained from such an experiment is the sum of the folding and unfolding rates, and this relaxation rate can—together with a measured equilibrium constant—thus be used to determine the folding and unfolding rates. In the case of villin, such temperature-jump experiments have been performed using either tryptophan fluorescence (3, 27) or the amide I band in IR spectroscopy (32) as spectroscopic signals.

We recently described an approach by which the kinetics observed in a long MD simulation can be compared more directly with the results of temperature-jump experiments (14). In particular, in the limit of a small temperature perturbation, the relaxation kinetics observed in a temperature-jump experiment can be compared more directly to an MD simulation by calculating the autocorrelation function (ACF) of a quantity that mimics the experimentally employed spectroscopic probe (14). This analysis requires (i) simulations that are much longer than the timescale of the relaxation of interest (so that the ACFs can be calculated with a reasonable statistical accuracy) and (ii) an efficient method that can be used to calculate the spectroscopic signal from the structures observed in the simulation. Most of the simulations presented here satisfy the first requirement, as they are one to two orders of magnitude longer than the folding and unfolding times (Table 1). To the best of our knowledge, however, no computationally affordable method has been reported that allows for a highly accurate calculation of IR spectra or excited-state energies for a large number of conformations in a complex protein system. We have thus limited ourselves to calculating quantities that are expected to be correlated with the spectroscopic signal of interest, although the exact nature of the correlation is unknown. This approach is expected to be sufficient for the estimation of relaxation times, but may prevent quantitative comparisons with the amplitude of the signal observed in experiment, in particular when relaxations on multiple timescales are present.

The spectroscopic signal from the amide I band in IR spectroscopy is primarily sensitive to the number and geometry of backbone hydrogen bonds, and is therefore a global measure of the secondary structure content in the protein. In our calculations we approximate this signal by the total number of amino acid residues that are in a helical geometry (33). The fluorescence properties of tryptophan residues are strongly influenced by fluctuations of the electric field surrounding the indole ring, which in turn is determined by the type of environment that surrounds the Trp side chain (34). In the absence of an accurate and efficient method to predict fluorescence properties, we approximate this signal by the solvent-accessible surface area (SASA) of the indole ring (14), although we note that a very recent study suggests that a more detailed description of the local geometry may be required to fully capture the fluorescence properties of villin (35). We therefore calculated the ACFs of the number of helical residues and the indole SASA from our equilibrium simulations of villin folding. Fitting of the ACFs to a single exponential decay resulted in considerable discrepancies for short lag times. They could, however, be fitted well to a double exponential decay (Fig. 1A) with a fast phase (time constants between 40 ns and 200 ns) and a slow phase (time constants between 0.4 μs and 5 μs). A similar two-exponential fit was deemed necessary in the analysis of temperature-jump experiments (3).

The faster relaxation was originally attributed to intra-basin dynamics (3, 27) and has been the subject of extensive triplet–triplet energy transfer (TTET) studies (36) and simulations (10, 28). We have simulated TTET experiments using the definition of TTET-active states proposed in (10) and the trajectory data of wild-type villin at 345 K, as these are the simulation data most readily comparable to the TTET experiment; analysis of the Nle/Nle double mutant trajectory at 360 K gave very similar results. The calculated triplet decay profiles are all fully consistent with the experimental results obtained in 2 M guanidinium hydrochloride (36). For probes placed on the N-terminus and residue 23 (0/23) and residues 7/23 (Fig. 1D) the triplet decay profiles are very similar (Fig. 1C) and can be modelled by two exponential decays with timescales of 60–100 ns and approximately 4.5 μs. A decomposition of the decay into the folded and unfolded state components shows that the slow phase corresponds to the unfolding transition and the fast phase to contact formation in the unfolded state. The calculated triplet decay profiles generated by probes placed on residues 23/35 and on the N-terminus and residue 35 (0/35) (Fig. 1D) can also be modelled by double-exponential decays (Fig. 1C), but this time with timescales of a few tens of nanoseconds and 190–230 ns. In these cases, the fast phase can be attributed to contact formation in the unfolded state, while the slower phase is generated by partial melting of helix 3 in the folded state, consistent with previous interpretations (10).

The timescale of the slow relaxation is very similar for the two spectroscopic probes and suggests that both the Trp side-chain environment and number of helical residues are sensitive to the global folding and unfolding reaction. To quantify the extent to which this is the case, we plotted the long-timescale relaxation time constants obtained from fitting the ACFs against the relaxation time calculated directly from the folding and unfolding rates (Fig. 1B). The results show a strong correlation between the relaxation time obtained from the ACFs and the values obtained from the mean waiting times in the folded and unfolded states, consistent with recent experimental studies indicating that Trp fluorescence properties can be used to monitor the folding/unfolding transition in villin (22). Over the temperature range studied here, we find—in agreement with experiments—that the folding rates are only weakly dependent on temperature (3, 27), whereas the unfolding rates increase substantially as the temperature is increased (Table 1). We also find that wild-type villin folds substantially slower than the Nle/Nle double mutant. At 360 K—where we have data for both proteins—the wild-type protein folds in 16 μs and the double mutant in 3 μs; the relative folding rates of the two proteins are in excellent agreement with the experimental measurements (3, 27), while the absolute folding rates appear to be a factor of three slower than the rates experimentally determined at 300 K. The ability to calculate folding rates and equilibrium constants directly from simulations of reversible folding also allows us to calculate protein-engineering Φ-values in a manner analogous to experiments (17, 21, 37). We performed simulations of a variant of villin in which we introduced the F10L mutation into both the wild type and Nle/Nle backgrounds (Fig. 1D). This mutation disrupts several key hydrophobic interactions between helix 1 and 2 and is expected to have a sizable effect on the stability of the protein. As described above, this mutation gives rise to a 0.8 to 1 kcal mol⁻¹ destabilization when introduced both in the Nle/Nle variant or in wild-type villin (Table 1). We find that at 345 K the F10L mutation appears to decrease the folding rate of wild-type
villin, and the resulting Φ-value of 0.2 ± 0.2 is consistent with the fractional Φ-values measured experimentally at 310 K (Φ = 0.3) and 340 K (Φ = 0.6) (29). In contrast, this mutation has no observable effect on the folding rate of the Nle/Nle variant, and the calculated Φ-value for folding is therefore small (0.0 ± 0.2).

The small difference in the Φ-value calculated for wild-type villin and for the Nle/Nle variant might suggest that the introduction of two stabilizing Nle-residues in helix 3 in the Nle/Nle double mutant subtly shifts the folding pathway relative to that in the wild-type protein (38). In our previous analysis of the Nle/Nle variant in the Amber ff99SB*-ILDN force field we found that helix 3 in general forms early during the folding pathway, and that helix 1 almost invariably forms last (14); a result that is in good agreement with the very low Φ-value that we here obtained computationally for F10L in the Nle/Nle background. A comparable analysis of the order of helix formation in wild-type villin shows that helix 2 forms first in 80% of the folding events (as compared to only 30% in the Nle/Nle double mutant). Thus, the larger Φ-value for F10L in the WT background reflects genuine, but subtle, differences in the folding mechanisms of the two proteins in our simulations. Future experiments probing the folding kinetics and thermodynamics of the F10L mutant in the Nle/Nle variant could be used to validate or disprove the pathway shift proposed on the basis of MD simulations.

The transition path time is the time it takes the molecule to transition between the folded and unfolded basin, and can be substantially shorter than the mean waiting times between such folding and unfolding events (39, 40). The mean value of the transition path time is of considerable interest as it contains useful information regarding the properties of the free-energy barrier. It also determines the time resolution needed in experiments in order to resolve individual folding events in single-molecule studies (39, 40). For each folding and unfolding event we determined the transition path time as the time required to transition fully between the RMSD cutoffs used to define these two states. The calculated values range between 120 and 460 ns, and show an approximately exponential dependence on the temperature (Fig. 2A). The calculated mean transition path times can be used to estimate the pre-exponential factor for folding (k₀) using the relation (40):

$$\langle \tau_{TP} \rangle = \frac{1}{2\pi k_0} \ln \left(\frac{k_0}{k_f}\right)^3$$

The values of k₀ estimated with this approach range between (0.5 μs)⁻¹ and (1.5 μs)⁻¹, in remarkable agreement with previous estimates (1), and also appear to be temperature-dependent (Fig. 2B). For diffusion in a rough potential, it has been proposed that there should be some kind of exponential dependence of k₀ with respect to the inverse or the square of the inverse of the temperature (41). The exponent is expected to be related to the roughness of the free-energy surface (41). We note the exact functional form describing the temperature-dependence of k₀ is not known and a substantial amount of additional data spanning a larger temperature interval would be required to infer it from simulation. A crude estimate of the roughness of the energy landscape can be obtained from an Arrhenius plot of k₀ (Fig. 2B), giving “effective activation energies” for diffusion of

---

**Fig. 1.** Simulated T-jump and TTET experiments. (A) Simulated IR T-jump experiment from the wild-type villin simulation performed at 345 K. The autocorrelation function for the number of helical residues as reported by STRIDE (33) (black) is used as a proxy for the decay of the amide band IR absorption following a T-jump and can be fitted to the sum of two exponential decays (red) with timescales of 0.12 and 5.2 μs. The long timescale describes the folding/unfolding transition, as demonstrated by the excellent agreement (B) between relaxation times calculated from the simulated IR and fluorescence T-jump experiments and the sum of the folding and unfolding rates (k⁺ and k⁻) obtained from a two-state analysis of the simulation trajectories. (C) Simulated TTET experiments from the wild-type villin simulation at 345 K. Four experiments with probes located on different pairs of residues were simulated. The decay profile of probe absorbance, monitored in TTET experiments, was calculated from the kinetics of contact formation using the contact definitions of (10). The dashed and dotted black lines show the decomposition of the solid black curve (probes located on residues 0 and 23) in the contributions coming from the folded (dashed) and unfolded (dotted) states. (D) Native state structure of the villin headpiece showing in red the side chains corresponding to the residues where donor and acceptor probes were attached in TTET experiments (38). The side chain of residue Phe10 is shown in cyan.
produce a reasonable description of the kinetics of the system.

for the simulations performed here, a two-state analysis will pro-

are substantially longer than the transition path time, as is the case

simulations. As long as the folding and unfolding relaxation times

short-end timescale of this distribution sets a natural lower limit

mine which of these theories fits the simulation results best. The

ever, require a substantially larger number of transitions to deter-

that predicted by a number of theories (41)

150 folding and unfolding events observed for the Nle/Nle double

characterized by a $kcal mol^{-1}$ for wild-type villin, approximately

5 kcal mol$^{-1}$ for the Nle/Nle double mutant, and approximately

1 kcal mol$^{-1}$ for the F10L mutant, when assuming a simple expo-

ential dependency on the inverse temperature. [For a model in

which the exponent scales with the inverse square of the tempera-

ture (41) we find the corresponding numbers to be approximately

2 kcal mol$^{-1}$ for wild-type villin, approximately 2 kcal mol$^{-1}$ for

the Nle/Nle double mutant, and approximately 1 kcal mol$^{-1}$ for

the F10L mutant.]

Finally, we find that the distribution of transition-path times for

individual folding and unfolding events is rather broad and

characterized by a “lag” at short times where no events are found and

a roughly exponential decay at longer timescales. In Fig. 3 we report the free energy profiles

for the wild-type villin simulations (1.7 kcal mol$^{-1}$ and 2.3 kcal mol$^{-1}$ at 345 K and 360 K, respectively) can be

Fig. 3. Distribution of transition path times. The plot shows the distribution of the observed transition path times for the 150 folding/unfolding events observed in the Nle/Nle double mutant simulation at 370 K. On the longest timescales the distribution is roughly exponential. On the shortest timescales, however, there is a clear “lag time” so that almost no events are observed at the shortest timescales. The distribution can be fitted to a simple model with two intermediate states on the transition path (red) (40), although some discrepancy is observed for short timescales, suggesting that the actual transition is a more complex process.

Estimation of the Folding Free-Energy Barrier. We used a previously described variational approach (45, 46) to estimate the free-energy barrier for the folding of villin. Briefly, in this approach, a reaction coordinate is optimized to separate maximally the transition paths from the stable folded and unfolded states. The free energy profile calculated along the optimized reaction coordinate is therefore expected to provide a reasonable estimate of the actual folding free-energy barrier. The free-energy barriers for folding calculated with this approach range between 1.1 kcal mol$^{-1}$ and 2.3 kcal mol$^{-1}$ (Fig. 4). These values are consistent with the pre-exponential factors of $(0.5 \mu s)^{-1}$ to $(1.5 \mu s)^{-1}$ estimated from the mean transition path times and the calculated folding times of 2.3 to 20 $\mu$s. In Fig. 4 we report the free energy profiles for the Nle/Nle double mutant, where simulations are available at three temperatures. While the three profiles are remarkably similar, the shape of the calculated barrier changes with temperature. In the 360-K simulation of the Nle/Nle mutant, a sparsely populated folding intermediate can be observed; analysis of the trajectory suggests that this intermediate corresponds to formation of helix 3 and the turn between helices 2 and 3. As the temperature is increased, this intermediate disappears and at 380 K the top of the barrier moves towards the native state, suggestive of a Hammond behavior. The free-energy barriers calculated for the wild-type villin simulations (1.7 kcal mol$^{-1}$ and 2.3 kcal mol$^{-1}$ at 345 K and 360 K, respectively) can be

Fig. 4. Free energy profiles for the fast-folding Nle/Nle double mutant of villin. Free energy profiles for the folding of the Nle/Nle double mutant of villin at three temperatures, projected along an optimized one-dimensional coordinate. The reaction coordinate was optimized separately for each simulation. To facilitate comparison, the profiles have been rescaled on the x-axis and translated so that the folded state has a coordinate value of approximately 0 and the unfolded state has a coordinate value of approximately 1. Folding free-energy barriers calculated from the profiles are 1.10 kcal mol$^{-1}$ at 360 K, 1.14 kcal mol$^{-1}$ at 370 K, and 1.70 kcal mol$^{-1}$ at 380 K.
Initial structures of the proteins were built in an extended conformation. Chignolin (1), Trp-cage (2), and FiP35 (3, 4) were solvated in a cubic boxes containing approximately 1900, 1700, and 4300 water molecules, respectively, and the charge of each protein was neutralized by sodium or chloride ions. Villin (5) was solvated in 40 mM NaCl in a cubic box of side length approximately 54 Å containing approximately 4400 water molecules. The Amber f99SB*–ILDN (6–8) force field was used to represent all systems. The systems were equilibrated in the NPT ensemble and production runs were performed in the NVT ensemble using the Anton specialized hardware (9) as described in (10). A single equilibrium simulation was run for each system, with the exception of the F10L mutant of wild-type villin, where two independent 180 μs simulations were performed and their results aggregated. Folded and unfolded states and transition pathways were defined using a transition-based assignment (10–13). IR amide I band and fluorescence intensity relaxation profiles were calculated from the autocorrelation function of the number of helical residues [as estimated using STRIDE (14)] and the solvent accessible surface area of the indole ring in the sole tryptophan residue (10). Optimal reaction coordinates to describe the folding/unfolding transition were obtained as described in (11).

Supporting Information

Table S1. Folding kinetics and thermodynamics from equilibrium MD simulations of three small proteins of α and β secondary structure content

<table>
<thead>
<tr>
<th>Protein</th>
<th>$T_{on}$ (K)</th>
<th>L (μs)</th>
<th>$T_{off}$ (K)</th>
<th>$\Delta G_f$</th>
<th>$\Delta H_f$</th>
<th>$\tau_f$ (μs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chignolin</td>
<td>370</td>
<td>60</td>
<td>343 *</td>
<td>423</td>
<td>$-0.7$</td>
<td>$-11.4^*$</td>
</tr>
<tr>
<td>Trp-cage</td>
<td>320</td>
<td>16</td>
<td>315 †</td>
<td>358</td>
<td>$-1.0$</td>
<td>$-13.4^*$</td>
</tr>
<tr>
<td>FiP35</td>
<td>385</td>
<td>1110</td>
<td>1360 ‡</td>
<td>386</td>
<td>$-0.1$</td>
<td>ND</td>
</tr>
</tbody>
</table>

The folding time ($\tau_f$) is calculated as the average waiting time in the unfolded state. Energies are reported in kcal mol$^{-1}$.

8. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB-ILDN force fields that have previously been used to study WW and villin folding (11, 17), as the melting temperatures of the FiP35 WW domain and—in particular—villin are closer to the experimental values.