Cooperative folding near the downhill limit determined with amino acid resolution by hydrogen exchange

Wooyoung Yu^a, Michael C. Baxa^{a,b}, Isabelle Gagnon^a, Karl F. Freed^{c,d,e,1}, and Tobin R. Sosnick^{a,b,e,1}

^aDepartment of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637; ^bInstitute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637; ^cDepartment of Chemistry, University of Chicago, Chicago, IL 60637; ^dJames Franck Institute, University of Chicago, Chicago, IL 60637; and ^eComputation Institute, University of Chicago, Chicago, IL 60637

Edited by Robert L. Baldwin, Stanford University, Stanford, CA, and approved March 10, 2016 (received for review November 16, 2015)

The relationship between folding cooperativity and downhill, or barrier-free, folding of proteins under highly stabilizing conditions remains an unresolved topic, especially for proteins such as λ -repressor that fold on the microsecond timescale. Under aqueous conditions where downhill folding is most likely to occur, we measure the stability of multiple H bonds, using hydrogen exchange (HX) in a λ_{YA} variant that is suggested to be an incipient downhill folder having an extrapolated folding rate constant of $2 \times 10^5 \text{ s}^{-1}$ and a stability of $7.4 \text{ kcal}\cdot\text{mol}^{-1}$ at 298 K. At least one H bond on each of the three largest helices ($\alpha 1$, $\alpha 3$, and $\alpha 4$) breaks during a common unfolding event that reflects global denaturation. The use of HX enables us to both examine folding under highly stabilizing, native-like conditions and probe the pretransition state region for stable species without the need to initiate the folding reaction. The equivalence of the stability determined at zero and high denaturant indicates that any residual denatured state structure minimally affects the stability even under native conditions. Using our ψ analysis method along with mutational ϕ analysis, we find that the three aforementioned helices are all present in the folding transition state. Hence, the free energy surface has a sufficiently high barrier separating the denatured and native states that folding appears cooperative even under extremely stable and fast folding conditions.

protein folding | hydrogen exchange | cooperative folding | λ repressor | folding pathway

The nature of the energy landscape when folding occurs on the microsecond timescale continues to be investigated using both experimental and computational approaches. The 80-residue subdomain of the λ -repressor transcription factor, λ_{6-85} (1–11), is an appealing system as it contains five α -helices arranged in a non-symmetric pattern, folds in microseconds, and is nearly a downhill folder (12), a condition where a continuous series of folding events may be characterized. These characteristics along with a folding time that nears the upper limit for current all-atom simulations (13–19) make this protein appropriate for detailed comparisons between experiment and simulations.

The energy landscape of λ_{6-85} is significantly influenced by its sequence. Oas and coworkers found that the wild type and a faster-folding mutant with two helix-promoting substitutions (λ_{AA} with G46A/G48A, Fig. S14) fold cooperatively (1–4). The transition state ensemble (TSE) characterized using ϕ analysis minimally contains $\alpha 1$ and $\alpha 4$ (3). Kinetic H/D isotope effect measurements in the Sosnick laboratory found that both λ_{6-85} and λ_{AA} fold cooperatively with $\sim 70\%$ of the H bonds being formed in the TSE, matching the degree of surface buried in the TSE [beta Tanford value (β_{Tanford})] (5).

Using nanosecond T-jump and other methods, Gruebele and coworkers proposed that faster-folding and stabilized variants, such as λ_{YA} (λ_{AA} + Q33Y) and λ_{D14A} (λ_{YA} + D14A), fold in an incipient or fully downhill manner, especially under highly stable conditions at lower temperatures (6–9). Key signatures of downhill folding are the presence of kinetics that are on the same timescale as the chain dynamics, suggestive of a free energy barrier with near-zero activation energy, and a fast “molecular phase” reflecting the

rapid redistribution of species near the top of the barrier after the T-jump that is more pronounced for the faster-folding variants (8). Additional signatures include probe-dependent kinetics (20) and aggregation tendencies (6). Lapidus and coworkers proposed that the λ_{D14A} variant folds in a downhill manner based on a combination of microfluidic mixing measurements and probe-dependent melting temperatures (10). Various computational methods, ranging from Gō models to all-atom simulations, have been applied to study the folding of λ_{6-85} and its variants, yielding mixed agreement with experiment and each other (13–19).

Here, a high-resolution characterization of the landscape and of the TSE of the λ_{YA} variant is obtained from a combination of equilibrium native state hydrogen exchange (NSHX) data, along with kinetic studies using both ψ analysis with bi-Histidine (biHis) metal ion binding sites and ϕ analysis at 298 K. The hydrogen exchange (HX) rates and their denaturant dependence provide information on individual H-bond stabilities and the structural fluctuations that lead to their breakage. Another advantage of HX is that measurements can be carried out without denaturant at room temperature, thereby enabling the characterization of folding behavior under highly stabilizing conditions where downhill folding is most likely to occur (7). Furthermore, HX has the ability to characterize stable pretransition state species, if they exist, as the surface is studied under equilibrium conditions.

We find that λ_{YA} folds in a manner typical of many proteins with a landscape that does not qualitatively change in destabilizing (urea) or stabilizing (sarcosine) conditions. The protein folds with considerable cooperativity, implying there is a sizable barrier between the native state ensemble (NSE) and the denatured state ensemble (DSE). Three helices ($\alpha 1$, $\alpha 3$, and $\alpha 4$) unfold in the same

Significance

Fast-folding proteins provide a testing ground for theories and simulations of folding at the extreme limit, in particular when it occurs on the timescale of chain diffusion and potentially in the elusive barrier-free limit. Whereas most fast-folding studies probe the reaction near the transition point with limited resolution, we apply hydrogen exchange methods to a microsecond folder, characterizing its energy landscape with amino acid precision under highly stabilizing conditions where barrier-free folding is most probable. Despite folding with $\tau = 5 \mu\text{s}$, we find that the molecule folds and unfolds in a cooperative process matching the properties observed at elevated denaturant concentration, implying that much faster folding rate constants are required to reach the downhill limit.

Author contributions: W.Y., K.F.F., and T.R.S. designed research; W.Y. and I.G. performed research; W.Y., M.C.B., I.G., and T.R.S. analyzed data; and W.Y., M.C.B., K.F.F., and T.R.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence may be addressed. Email: trsosnic@uchicago.edu or freed@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1522500113/-DCSupplemental.

step, representing the major unfolding event. According to kinetic studies, these three helices are present in the TSE, whereas $\alpha 2$ and $\alpha 5$ fold later, consistent with the HX data. These results are compared with those of previous experimental and computational studies.

Results

Cooperative Folding Behavior. Denaturation with urea at 298 K, monitored with far-UV circular dichroism (CD), fits well to a two-state model having a linear denaturant dependence, $\Delta G_{\text{eq}} = \Delta G_{\text{eq}}^0 - m^0[\text{urea}]$ (Table S1). The measured values, $\Delta G_{\text{eq}} = 6.96 \pm 0.02$ kcal·mol⁻¹ and $m^0 = 1.00 \pm 0.02$ kcal·mol⁻¹·M⁻¹, closely match their counterparts derived from kinetic measurements using chevron analysis (21–23), $\Delta G_{\text{eq}} = RT \ln(k_f/k_u) = 6.96 \pm 0.15$ kcal·mol⁻¹ and $m^0 = 1.01 \pm 0.02$ kcal·mol⁻¹·M⁻¹ (extrapolated $k_f^{\text{water}} \sim 2 \times 10^5$ s⁻¹ and $\beta_T = 0.76 \pm 0.03$, Fig. 1), where R is the gas constant, and T is the absolute temperature. The agreement between equilibrium and kinetic data indicates that all of the free energy and surface burial are accounted for in the observed kinetic phase. This finding is a generally accepted demonstration that folding is well approximated by a double-well energy surface with a single barrier and that no species with appreciable population exists before the millisecond observation window (21–23).

Thermal denaturation is more complicated, in part due to aggregation (SI Text and Fig. S2). Rapid CD and fluorescence measurements before aggregation suggest thermal denaturation is adequately described as proceeding through a single transition, consistent with Ma and Gruebele's findings (20). Whereas two-state, barrier-limited folding behavior near the thermal or urea-induced melting transition is possible even for downhill folders (24), the major issue under investigation is whether the cooperativity persists even under highly stabilizing conditions, which we will determine using HX.

Whereas the major folding phase satisfies a two-state folding model, we do note that a minor (5–20%) slow refolding phase of ~ 2 – 10 s⁻¹ is observed for λ_{YA} , which persists beyond 6 M urea with little to no denaturant dependence (Fig. S3). This phase may be due to a subpopulation folding along a parallel pathway (25) that accumulates due to a misfolding event (26, 27), e.g., potentially the rearrangement of $\alpha 2$ or $\alpha 5$. Prigozhin and Gruebele (11) proposed the presence of a trapped intermediate containing nonnative β structure based on similar results and as observed in some simulations (18). Interestingly, this minor phase is not observed in the biHis and A/G variants used to characterize the TSE. Because we are interested in the major folding pathway (and the minor phase is absent in the various mutants), we focus on properties of the major phase.

HX Studies. A detailed characterization of the folding landscape is obtained using NSHX/NMR measurements in 0–4 M urea and in 0.5 M and 1 M sarcosine, a stabilizing osmolyte ($\Delta\Delta G_{\text{CD}}^{\text{IM sarc}} =$

1.5 kcal·mol⁻¹, Tables S1 and S2). The analysis of HX data assumes that amide protons (NH) exist in states that are either exchange competent (open) or incompetent (closed) (28). The observed HX rate (k_{ex}) depends on the rates of opening (k_{op}) and closing (k_{cl}) and the intrinsic exchange from the open state (k_{chem}) according to $[\text{Closed}] \xrightleftharpoons[k_{\text{cl}}]{k_{\text{op}}} [\text{Open}] \xrightarrow{k_{\text{chem}}} [\text{Exchanged}]$. Values for k_{chem} are calculated from peptide studies (29). Often $k_{\text{cl}} \gg k_{\text{chem}}$ (including our experiments where $k_{\text{fold}} = k_{\text{cl}} > 5k_{\text{chem}}$), and k_{ex} reduces to the “EX2” limit; i.e., $k_{\text{ex}} = k_{\text{chem}}k_{\text{op}}/(k_{\text{op}} + k_{\text{cl}}) = k_{\text{chem}}K_{\text{op}}/(K_{\text{op}} + 1)$, where $K_{\text{op}} = k_{\text{op}}/k_{\text{cl}}$ is the equilibrium constant, and the free energy of exchange, $\Delta G^{\text{HX}} = RT \ln K_{\text{op}}$.

The use of denaturants enables the quantification of the amount of accessible surface area (ASA) present in the various conformational states that are exchange competent, i.e., an open state. Denaturants stabilize states with more exposed area (30) and thus provide a quantitative measure of exposure (28). Residue-specific m^{HX} values reflect the size of the opening event for each NH undergoing HX and depend upon whether exchange occurs in a near native state, the DSE, or an intermediate state, i.e., local, global, and subglobal openings, respectively (28, 31). In a local event, m^{HX} is near zero, implying that HX occurs from the NSE via a small-scale perturbation that exposes little additional surface area. Sites that require complete unfolding exchange with global properties, $\Delta G^{\text{HX}} = \Delta G^{\text{global}}$ and $m^{\text{HX}} = m^0$. The third class, subglobal exchange, is more often found in larger proteins and occurs when residues in one or more pieces of secondary structure, or “foldons,” collectively exchange with $\Delta G^{\text{HX}} < \Delta G^{\text{global}}$ and $0 < m^{\text{HX}} < m^{\text{global}}$. These openings represent intermediate states that can provide information on the folding pathway. In the case of apparent two-state folding reactions, these species lie on the native side of the major barrier.

Whereas some residues may predominantly exchange via only one of the three classes of openings, other residues may exchange from two or all three classes to varying degrees that depend on the relative values of ΔG^{global} , ΔG^{local} , and $\Delta G^{\text{subglobal}}$. For example, increasing the denaturant concentration preferentially promotes the global unfolding event such that exchange may first occur through a local event but merge with the global “isotherm” at higher denaturant concentrations, $[\text{denaturant}]_{\text{switch}} \sim (\Delta G^{\text{global}} - \Delta G^{\text{local}})/m^0$ (32). More complex behavior can occur if all three classes contribute, which results in the merging of three isotherms. Alternatively, a continuum of states suggestive of a downhill folding landscape would result in a continuous spread in ΔG_i^{HX} and m^{HX} values over the entire protein (Fig. 2).

HX rates were obtained using ¹H-¹⁵N heteronuclear single-quantum correlation (HSQC) spectra for fully-¹H λ_{YA} in 80% D₂O at 298 K, pD_{read} 5.8 ~ 7.5, a condition where the protein is 0.4 kcal·mol⁻¹ more stable than in 100% H₂O (Table S1). The urea and sarcosine data are combined using a conversion factor $m_{\text{urea}}/m_{\text{sarc}} = -1.5$ (33), providing points for apparent [urea] of -0.75 M and -1.5 M.

HX data for nine residues on $\alpha 1$, $\alpha 3$, and $\alpha 4$ can be fitted allowing for only a global unfolding process from 0 M to 4 M urea. The HX data for three residues on $\alpha 1$ and $\alpha 4$ (I21, L65, and V71) track with the global unfolding even in -1.5 M urea (1 M sarcosine). To fit the combined urea/sarcosine data for the other six residues, a local unfolding process is added with $\Delta G^{\text{local}} \sim 7$ kcal·mol⁻¹ and $\Delta G^{\text{global}} \geq 7.3$ kcal·mol⁻¹ with $m^{\text{HX}} \leq 1.0$ kcal·mol⁻¹·M⁻¹ (Fig. 3). These values match the equilibrium values of $\Delta G_{\text{eq}}^0 = 7.3$ – 7.4 kcal·mol⁻¹ and $m^0 = 1.0$ kcal·mol⁻¹·M⁻¹ obtained in the same solvent condition. The results indicate that the most stable sites on $\alpha 1$, $\alpha 3$, and $\alpha 4$ exchange via global unfolding and preclude the presence of a stable intermediate having one or more of these helices unfolded.

One site, N27, has a slightly higher stability than the global exchanging sites, $\Delta\Delta G^{\text{HX}} \sim 0.5$ – 1 kcal·mol⁻¹, yet exhibits the same global urea dependence. The N27 amide proton is located in a distorted region at the C terminus of $\alpha 1$ and lacks an obvious H-bond partner (Fig. 3D). This superprotection could be due to an inaccurate estimate of its k_{chem} . Alternatively, this proton may be

Fig. 1. Folding behavior of λ_{YA} . Chevron plot of folding and unfolding rate constants using tryptophan fluorescence (Left) and chemical denaturation monitored using fluorescence and CD (Right), obtained in 100% H₂O (blue) and 80% D₂O (black, as used in HX measurements). Units for ΔG and m are kcal·mol⁻¹ and kcal·mol⁻¹·M⁻¹, respectively.

Fig. 2. Free energy surface (Upper Left) and predicted HX patterns for four barrier heights: $<0 RT$, $0 RT$, $2 RT$, and $5 RT$. The indicated m^{HX} values are the fitted slopes from 0 to 4 M urea for the isotherms associated with the various states. The set of states located between the DSE-TSE and the set between the TSE-NSE are assumed to possess free energy differences and m values that are uniformly spaced within each set. Insets depict the energy surface for each condition.

partially buried or participating in an H bond that is formed 50–80% of the time in an otherwise unfolded chain. The persistence of the superprotection out to 4 M urea necessitates that any possible burial or H-bond stability in the DSE be denaturant independent. As this requirement seems unlikely, we believe instead that k_{chem} is inaccurately estimated for this residue.

The remaining observable HX data for the five α -helices and the 3_{10} helix can be fitted with a model containing both global and local unfolding processes with $\Delta G_{\text{local}} < 7 \text{ kcal}\cdot\text{mol}^{-1}$. Of these residues, two on the 3_{10} helix (E75, F76), have the highest stability with $|\Delta G_{\text{global}} - \Delta G_{\text{local}}| \sim 1 \text{ kcal}\cdot\text{mol}^{-1}$ below the global value.

The fit of the data for M40 in $\alpha 2$ can be slightly improved by allowing for a subglobal unfolding process with $\Delta G_{\text{sub}} = 5.8 \text{ kcal}\cdot\text{mol}^{-1}$ and $m_{\text{sub}} = 0.15 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$. The kinetic value $m_u = 0.24 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$, obtained from the chevron analysis (Fig. 1), reflects the surface exposed going from the NSE to the TSE. Because $m_{\text{sub}} < m_u$, the possible intermediate reflected by the subglobal opening must

exist between the TSE and the NSE. This finding is consistent with the kinetic data described below and the other HX data. The three observable NHs on $\alpha 2$ exchange with $\Delta G^{\text{HX}} = 4.3\text{--}5.9 \text{ kcal}\cdot\text{mol}^{-1}$ whereas the two NHs on $\alpha 5$ exchange with $\Delta G^{\text{HX}} = 3.6\text{--}4.4 \text{ kcal}\cdot\text{mol}^{-1}$. This difference, in conjunction with equilibrium mutation studies (34), suggests that $\alpha 5$ folds after $\alpha 2$, although $\alpha 5$ just may be more dynamic in the NSE. Regardless, the significant HX protection found in $\alpha 5$ indicates that it is folded at least 99.9% of the time.

The TSE and Folding Pathway. We applied a combination of ψ and ϕ analyses to characterize the TSE. To conduct ψ analysis, five biHis sites were individually introduced into the five helices of λ_{YA} (*biHis1*–*biHis5*, Fig. 4 and Table S3). The addition of zinc or nickel ions, which coordinate the pair of histidines, alters the protein's stability and activation free energy for folding ($\Delta\Delta G_{\text{eq}}$ and $\Delta\Delta G_{\text{f}}$, respectively) due to differences in binding constants K^{DSE} , K^{N} , and K^{TSE} (35–37). The ion-induced changes in $\Delta\Delta G_{\text{eq}}$ and $\Delta\Delta G_{\text{f}}$ are used to define the ψ value, a parameter analogous to the mutational ϕ value, with ψ being the instantaneous change in $\Delta\Delta G_{\text{f}}$ relative to $\Delta\Delta G_{\text{eq}}$.

$$\phi_{\text{mutation}} = \frac{\Delta\Delta G_{\text{f}}}{\Delta\Delta G_{\text{eq}}}\bigg|_{\text{mutation}}; \quad \psi = \frac{\partial\Delta\Delta G_{\text{f}}}{\partial\Delta\Delta G_{\text{eq}}}\bigg|_{\text{divalent ion}} \quad [1]$$

Any potential artifacts related to metal ion binding are alleviated by evaluating ψ in the limit of zero perturbation, ψ_0 . The value of ψ_0 reflects the intrinsic degree of contact formation in the TSE in the absence of metal ions. ψ_0 values of zero and unity indicate that the binding affinity of the biHis site in the TSE is the same as found in the DSE and NSE, respectively. These two behaviors are interpreted as the biHis site being absent or native-like in the TSE, respectively. A fractional ψ_0 indicates that the biHis site either is native-like in a subpopulation of the TSE or contains nonnative binding affinity (e.g., a distorted site with less favorable binding geometry or a flexible site that must be restricted before ion binding) or some combination thereof (35, 38–42) (*SI Materials and Methods*). Mutational ϕ^{biHis} values also are calculated for each biHis site, using double-mutant data (in the absence of metal ions) with the wild-type protein data serving as the reference point.

The CD spectra of the biHis variants are very similar to those for λ_{YA} except for *biHis5*, the variant with the biHis site on $\alpha 5$ (Fig. S1B). Its spectrum suggests that $\alpha 5$ may be unfolded before the addition of metal ions. The refolding fluorescence signal for *biHis2* was too small to measure, whereas for *biHis4*, the addition of cations did not produce a readily interpretable shift in the chevron plot (Fig. S1C). To compensate, we measured $\phi^{\text{Ala}\rightarrow\text{Gly}}$ to

Fig. 3. Denaturant dependence of HX. (A and B) Data are fitted using a model allowing for global and local unfolding processes except for M40 and E75 where additional subglobal processes are allowed (dashed-dotted lines). The denaturant dependence of the equilibrium global stability from Fig. 1 is shown as a black dashed line. (C) Fitted values for ΔG^{HX} in zero denaturant are shown along with the m values for sites in A that exchange via the global unfolding process. Sites not requiring a local unfolding process to fit the data down to 0 M urea are highlighted with an "x." Units for ΔG and m are $\text{kcal}\cdot\text{mol}^{-1}$ and $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$, respectively. (D) The locations of associated H bonds are highlighted with a cyan solid line. N27 amide does not have an H bond in PDB structure and best candidate H-bond partners for N27 are highlighted with blue dashed lines.

Fig. 4. ψ , ϕ analysis. (A) The ψ_0 , ϕ values are mapped onto the protein structure by coloring the biHis side chains and ribbon according to the associated ϕ^{biHis} and ψ_0 values, respectively. The D38 and A63 C α atoms are denoted with colored spheres according to their $\phi^{\text{A} \rightarrow \text{G}}$ values. (B) Measured ψ , ϕ values. (C) Proposed folding pathway. The 3_{10} helix is not shown for simplicity and lack of information; HX measurements are consistent with it forming before $\alpha 2$. (D) Chevron plots. Shown are biHis mutant data without (black) and with 1 mM Zn^{2+} (red) and glycine (green) and alanine (blue) variants; orange vertical lines designate the [urea] at which $\Delta\Delta G_f$ and $\Delta\Delta G_u$ are used for the ψ calculation, obtained by simultaneously folding and unfolding data (SI Materials and Methods).

probe for the presence of $\alpha 2$ and $\alpha 4$ helices in the TSE. This application of ϕ analysis using helix promoting/disrupting substitutions at solvent-exposed positions avoids many of the factors that typically confound the interpretation of ϕ values (38) and has proved consistent with ψ analysis (43, 44). We also calculated ϕ^{biHis} to further probe the presence of $\alpha 2$ and $\alpha 4$ in the TSE.

The combination of ψ and ϕ data indicates that the TSE contains $\alpha 1$, $\alpha 3$, and $\alpha 4$ but likely lacks the two smaller helices (Fig. 4). Specifically, we obtain $\psi_0^{\text{biHis}} = 0.98 \pm 0.28$ and $\psi_0^{\text{biHis}} = 0.76 \pm 0.06$ whereas $\psi_0^{\text{biHis}} = 0.03 \pm 0.02$. The corresponding ϕ^{biHis} values for $\alpha 1$, $\alpha 3$, and $\alpha 5$ are 0.63 ± 0.02 , 0.55 ± 0.02 , and 0.30 ± 0.03 , respectively. The site on $\alpha 4$ yields $\phi^{\text{biHis}} = 0.66 \pm 0.03$, whereas the A/G pairs in $\alpha 2$ and $\alpha 4$ produce $\phi^{\text{A38G}} = 0.40 \pm 0.05$ and $\phi^{\text{A63G}} = 0.98 \pm 0.07$. The intermediate ϕ^{A38G} value is challenging to interpret; it could reflect fractional helix formation or the loss of some backbone entropy due to the chain becoming constrained upon the folding of the flanking helices $\alpha 1$ and $\alpha 3$.

Regardless, our identification of the presence of $\alpha 1$, $\alpha 3$, and $\alpha 4$ as the major constituents in the TSE is consistent with the HX data as they are the same three helices that unfold in the global unfolding process. The folding cooperativity of these three helices identified by HX precludes the presence of a stable intermediate having one or more of these helices unfolded. Rigorously, local unfolding events do not provide pathway information. However, the possible subglobal opening for $\alpha 2$ along with lower stability of sites in $\alpha 5$ suggests that $\alpha 2$ folds before $\alpha 5$ (Fig. 4).

HX Pattern for Cooperative and Downhill Folding. To assist in the interpretation of the data, we model the HX pattern for a landscape with barrier heights of $\Delta G_f = 0 RT$, $2 RT$, and $5 RT$ (Fig. 2). The model has five sequential states (other than the DSE), with each folding event representing the formation of an additional helix. The folding order of the first three helices is unknown so that their order is left ambiguous (labeled a–c in Fig. 2) whereas the HX data suggest $\alpha 2$ folds before $\alpha 5$. The state containing helices a–c is the TSE with $\beta_T = 0.75$, in accord with the chevron data.

For a barrier height of $5 RT$, the HX isotherms for helices a–c display global or near global behavior, whereas the helix d isotherm is notably lower and flatter. With decreasing barrier height,

the three highest isotherms diverge. The observed HX data are noticeably closer to those of the models with higher barriers. The addition of alternative pathways or more states would only accentuate the difference with the observed data by increasing the number of states from which HX could occur and further separate the highest three isotherms from the global isotherm. Although the experimental data cannot identify the barrier height, this model illustrates that the folding reaction must overcome a significant barrier to be consistent with the observed HX pattern.

Discussion

The HX, kinetic, and equilibrium data indicate that the λ_{YA} variant folds cooperatively, which is well approximated by a double-well energy landscape that is typical of many globular proteins. Folding is cooperative even under highly stabilizing conditions where the (extrapolated) folding rate constant is extremely fast, $(5 \mu\text{s})^{-1}$. Our HX data indicate that the last major unfolding barrier reflects the concerted unfolding of $\alpha 1$, $\alpha 3$, and $\alpha 4$. Kinetic data using ψ and ϕ analysis indicate that folding events occur in the reverse order, with the three helices forming early, being present in the TSE. A possible subglobal opening was measured for $\alpha 2$, suggesting that this helix folds after the TSE, consistent with the apparent two-state folding behavior and the mutational data.

λ_{YA} has been suggested to be an incipient downhill folder (6). Ma and Gruebele (20) observed probe-dependent kinetics for its major folding phase. λ_{YA} shares some properties of λ_{D14A} , a protein suggested to be a downhill folder (6, 10), including an initial 2- μs signal change. This phase is proposed to be a molecular phase resulting from the fast redistribution of molecules residing at the top of the barrier immediately following the T-jump, in part justified by the phase being more pronounced in the faster-folding variants. Alternatively, this phase, which is observed upon a shift to unfolding conditions, could be reporting on a redistribution of states in the native energy well rather than states near the top of the major barrier. These events could include helical fraying or even the unfolding of $\alpha 2$ or $\alpha 5$. Despite the better agreement between the HX data and models of cooperative folding (Fig. 2), it is difficult to rule out a small fraction of molecules existing near the TSE on an incipient downhill landscape for λ_{YA} .

We previously investigated the possibility of downhill folding for the cross-linked GCN4-p2 coiled coil (45) and similarly inferred that folding remains barrier limited using HX even though $k_f = (5 \mu\text{s})^{-1}$ and $\Delta G_{\text{eq}} = 6 \text{ kcal}\cdot\text{mol}^{-1}$ at 313 K. These two HX studies on fast-folding proteins indicate that significant barriers remain even in the absence of denaturants and that the prefactor (k_{attempt}) used to estimate barrier heights (46), $k_f = k_{\text{attempt}} e^{-\Delta G^\ddagger/RT}$, must be faster than $(5 \mu\text{s})^{-1}$.

TSE Structure. In our previous studies of kinetic H/D isotope effects on λ_{AA} , we obtained $\phi^{\text{H/D}} = 73 \pm 4\%$, which is interpreted as three-quarters of the helical H bonds are formed in its TSE (5). This magnitude is the same as the β_T value and the percentage of λ_{YA} 's H bonds contained within the three helices present in the TSE. Our results also are largely consistent with those of Burton et al. (3) for λ_{6-85} , where the seven $\phi^{\text{A/G}}$ values measured on $\alpha 1_{2\text{sites}}$, $\alpha 2$, $\alpha 3$, $\alpha 4_{2\text{sites}}$, and $\alpha 5$ have $\phi^{\text{A/G}} = 0.5/1.0$, 0.2 , 0.3 , $0.8/1.2$, and 0.6 , respectively. Only one of two A/G sites on $\alpha 1$ and on $\alpha 4$ has a highly exposed C β , a desirable feature for probing helix formation independent of tertiary contacts. These positions produce large ϕ values in agreement with our identification of these two helices as members of the TSE ($\phi^{\text{A20G}} = 1.0$ and $\phi^{\text{A63G}} = 0.8$).

The high ψ and ϕ values in $\alpha 1$, $\alpha 3$, and $\alpha 4$ point to a TSE with limited conformational diversity. However, we are hesitant to entirely rule out the presence of $\alpha 2$ (or the 3_{10} helix) in the TSE, in part, as we observed a nonzero ϕ value in $\alpha 2$ ($\phi^{\text{A38G}} = 0.4 \pm 0.1$). Furthermore, excluding $\alpha 2$ from the TSE requires that all of the H bonds in $\alpha 1$, $\alpha 3$, and $\alpha 4$ be present in the TSE to satisfy the H/D isotope effect data (5). This possibility seems unlikely especially as $\alpha 1$ is 20 residues long with a poorly packed amino terminus that is

possibly frayed in the TSE. Hence, we suspect other portions of the protein could be partially structured in the TSE in a native or nonnative manner and contribute to the 74% H-bond formation inferred from our prior isotope effect measurements (5).

In recent folding studies of λ_{YA} , Prigozhin et al. (47) compared the fluorescence quenching of W22 on $\alpha 1$ by Y32 on $\alpha 2$ with two new W-Y pairs on $\alpha 1$ – $\alpha 3$ and $\alpha 3$ – $\alpha 2$. The k_f for the $\alpha 1$ – $\alpha 3$ construct is slower by up to 20% than the other pairs, suggesting that the contacts involving $\alpha 2$ form earlier. We find, however, that the TSE largely lacks $\alpha 2$. It is possible that the quenching signals do not require that $\alpha 2$ be folded. Alternatively, the changes in rates could be viewed from the perspective of ϕ analysis where rate differences reflect energetic perturbations to the TSE. This view suggests that $\alpha 1$ – $\alpha 3$ contacts are the most significant, consistent with our finding of a TSE that contains $\alpha 1$, $\alpha 3$, and $\alpha 4$.

The present study produces a picture of an extensive TSE, as observed for six other proteins studied using ψ (36, 37, 41, 48–50). Their TSEs adopt a high fraction of the native topology, defined using the relative contact order (RCO) parameter, $RCO^{TSE} \sim 0.7 \cdot RCO^N$. A TSE model composed only of the contacts between $\alpha 1$, $\alpha 3$, and $\alpha 4$ follows this trend, yielding an $RCO^{TSE} \sim 0.85 \cdot RCO^N$, further supporting this relationship and the well-known correlation between k_f and RCO (51).

Residual Structure. The importance of residual structure elicits diverse views. Both the stability and the m -value determined by HX for λ_{YA} even in 1-M sarcosine match those determined by equilibrium denaturation at high [urea]. This equivalence implies that residual structure has no (additional) thermodynamic effect in the absence of denaturant and the level of surface exposure in the DSE is the same in zero- and high-urea concentrations.

For a denatured analog of λ_{AA} , Chugha et al. (52) measured a $CD_{222\text{ nm}} = -4,500 \text{ deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1}$. This level is equivalent to ~13–15% helical content. By comparison, a helix-coil theory based on primary sequence (53) predicts an average helicity of 19% for λ_{YA} , most of which is concentrated in $\alpha 1$ and $\alpha 4$. These two helices have a stretch of 12 or 11 residues with an average helical probability of 66% or 30%, respectively. The remaining helices have minimal helicity (<5%), including $\alpha 3$, which is present in the TSE. There is no obvious correlation between intrinsic helicity and a higher ΔG^{HX} , indicating that residual helicity, if present, affords no measurable HX protection.

Evidence of a thermodynamically equivalent DSE under stabilizing and denaturing conditions can be found for many other proteins. We found that the HX-determined stabilities and m values match those derived from global denaturation measurements for ubiquitin (54), a coiled coil (45), and protein G (32). In addition, Huyghues-Despointes et al. (55) observed that the stability in water obtained from HX data agrees with values determined by traditional chemical and thermal denaturation studies for 19 of 20 proteins. This equivalence indicates that the two methods probe the same folding transition to a DSE that is devoid of stable H bonds. Moreover, data from small-angle X-ray scattering imply that the dimensions in the DSE of many [but not all (56, 57)] small proteins remain unchanged upon shifting to a lower level of denaturant (32, 58–60). Furthermore, most [but not all (61)] small proteins satisfy the two-state chevron criterion, sometimes even down to 0 M denaturant (21, 22), indicating that the DSE remains highly solvated under native conditions for many proteins.

Comparison with Other Simulations. Previous simulations on λ repressor and variants used a variety of methodologies from all-atom

molecular dynamics (MD) simulations to coarse-grain models (13–19, 62) (Fig. S4 and Table S4). Some all-atom simulations produce a DSE that is overly collapsed with a considerable amount of helix formation, in contrast with experimental data, as discussed in *Residual Structure*. Hence, it is challenging to identify the folding order; in these cases, folding is ranked by when the entire helix becomes folded.

Most, but not all, simulations (17) predict that $\alpha 1$ and $\alpha 4$ fold first, in good agreement with our data. These two helices are the longest, have the highest intrinsic helicity, and provide a suitable scaffold for the folding of $\alpha 2$ and $\alpha 3$. A Gō model, however, describes the folding of $\alpha 2$ and $\alpha 3$ as effectively commensurate with $\alpha 1$ and $\alpha 4$ (15). The C-terminal helix, $\alpha 5$, tends to fold last and to fluctuate in the native state.

Some of the variability in the folding order of $\alpha 2$ and $\alpha 3$ may be due to minor sequence differences. Our *TerFix* algorithm that is based on the principle of sequential stabilization (13, 14) predicts a folding pathway for λ_{D14A} that is in agreement with the experimental data for λ_{YA} . However, using slightly different protocols for passing information to the next round of trajectories switches the order of helix formation between $\alpha 2$ and $\alpha 3$. This switch suggests that, at least computationally, the relative folding order of $\alpha 2$ and $\alpha 3$ is sensitive to small energetic differences and may be sequence dependent.

Conclusion

The cooperative folding behavior observed for λ_{YA} is consistent with an energy surface containing a single major energy barrier. Prior simulations generally support the experimental findings regarding the order of helix formation with nearly all studies identifying $\alpha 1$ and $\alpha 4$ as the first to fold. However, simulations typically describe the DSE as highly compact and highly structured in contrast to experimental data where any possible residual structure is thermodynamically silent and affords little, if any, HX protection. We have also observed cooperative folding for λ_{YA} , which also folds with an extrapolated time constant, $\tau = 5 \mu\text{s}$, implying that much faster folding-rate constants are required to achieve the downhill limit.

Materials and Methods

The sequence of λ_{YA} was taken from Protein Data Bank (PDB) 3KZ3 (6) and cloned into a pet21 expression vector with a His6-tag and an expression tag with a TEV cleavage site for tag removal. See *SI Materials and Methods* for additional details.

HX was initiated by buffer exchange from H_2O to a final condition of 80% $\text{D}_2\text{O}/20\%$ H_2O , 50 mM sodium phosphate, using sephedex G-15 spin columns (0.1–0.2 mM protein, 1 mM sodium azide). Proton levels were measured with a Bruker AVANCE III console at 500 MHz, using SOFAST-HMQC (heteronuclear multiple-quantum correlation spectroscopy) (63) and HSQC pulse sequences.

Equilibrium denaturation was monitored with a Jasco J-715 CD spectrometer in a 1-cm path length cuvette at a protein concentration of 2 μM and 50 mM sodium phosphate buffer at pH 7.0 by CD at $227 \pm 5 \text{ nm}$ and fluorescence at $\lambda_{\text{excite}} = 280$; fluorescence was measured downstream of a 310-nm cutoff filter. Kinetic measurements were obtained using a Bio-Logic SFM-400, -4000 stopped-flow instrument integrated with a PTL light source to monitor tryptophan fluorescence at $\lambda_{\text{excite}} = 285 \text{ nm}$. ψ_0 values were determined from a simultaneous fit to the vanishing and high Zn^{2+} chevrons, with ψ_0 included as one of the fitting parameters and using nonlinear least-squares algorithms implemented in the Origin software package (OriginLab).

ACKNOWLEDGMENTS. We thank M. Gruebele, G. Bowman, and members of our groups for useful comments. Trajectories for λ repressor were kindly provided by D. E. Shaw Research. This work was supported by National Institutes of Health Research Grant R01 GM055694 and by the NSF (CHE-1363012). W.Y. was supported in part by National Creative Research Initiatives (Center for Proteome Biophysics) of the National Research Foundation, Korea (Grant 2011-0000041).

- Huang GS, Oas TG (1995) Structure and stability of monomeric lambda repressor: NMR evidence for two-state folding. *Biochemistry* 34(12):3884–3892.
- Ghaemmhami S, Word JM, Burton RE, Richardson JS, Oas TG (1998) Folding kinetics of a fluorescent variant of monomeric lambda repressor. *Biochemistry* 37(25):9179–9185.
- Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG (1997) The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat Struct Biol* 4(4):305–310.

- Burton RE, Huang GS, Daugherty MA, Fullbright PW, Oas TG (1996) Microsecond protein folding through a compact transition state. *J Mol Biol* 263(2): 311–322.
- Krantz BA, et al. (2002) Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat Struct Biol* 9(6):458–463.
- Liu F, Gao YG, Gruebele M (2010) A survey of lambda repressor fragments from two-state to downhill folding. *J Mol Biol* 397(3):789–798.

7. Yang WY, Gruebele M (2004) Folding lambda-repressor at its speed limit. *Biophys J* 87(1):596–608.
8. Liu F, Gruebele M (2007) Tuning lambda6-85 towards downhill folding at its melting temperature. *J Mol Biol* 370(3):574–584.
9. Yang WY, Gruebele M (2003) Folding at the speed limit. *Nature* 423(6936):193–197.
10. DeCamp SJ, Naganathan AN, Waldauer SA, Bakajin O, Lapidus LJ (2009) Direct observation of downhill folding of lambda-repressor in a microfluidic mixer. *Biophys J* 97(6):1772–1777.
11. Prigozhin MB, Gruebele M (2011) The fast and the slow: Folding and trapping of lambda6-85. *J Am Chem Soc* 133(48):19338–19341.
12. Gelman H, Gruebele M (2014) Fast protein folding kinetics. *Q Rev Biophys* 47(2):95–142.
13. Adhikari AN, Freed KF, Sosnick TR (2013) Simplified protein models: Predicting folding pathways and structure using amino acid sequences. *Phys Rev Lett* 111(2):028103.
14. Adhikari AN, Freed KF, Sosnick TR (2012) De novo prediction of protein folding pathways and structure using the principle of sequential stabilization. *Proc Natl Acad Sci USA* 109(43):17442–17447.
15. Shoemaker BA, Wang J, Wolynes PG (1999) Exploring structures in protein folding funnels with free energy functionals: The transition state ensemble. *J Mol Biol* 287(3):675–694.
16. Allen LR, Krivov SV, Paci E (2009) Analysis of the free-energy surface of proteins from reversible folding simulations. *PLoS Comput Biol* 5(7):e1000428.
17. Liu Y, Strümpfer J, Freddolino PL, Gruebele M, Schulten K (2012) Structural characterization of lambda-repressor folding from all-atom molecular dynamics simulations. *J Phys Chem Lett* 3(9):1117–1123.
18. Bowman GR, Voelz VA, Pande VS (2011) Atomistic folding simulations of the five-helix bundle protein lambda6-85. *J Am Chem Soc* 133(4):664–667.
19. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517–520.
20. Ma H, Gruebele M (2005) Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proc Natl Acad Sci USA* 102(7):2283–2287.
21. Jackson SE, Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30(43):10428–10435.
22. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* 324(2):359–371.
23. Jackson SE (1998) How do small single-domain proteins fold? *Fold Des* 3(4):R81–R91.
24. Kim SJ, Matsumura Y, Dumont C, Kihara H, Gruebele M (2009) Slowing down downhill folding: A three-probe study. *Biophys J* 97(1):295–302.
25. Kiefhaber T (1995) Kinetic traps in lysozyme folding. *Proc Natl Acad Sci USA* 92(20):9029–9033.
26. Sosnick TR, Mayne L, Hiller R, Englander SW (1994) The barriers in protein folding. *Nat Struct Biol* 1(3):149–156.
27. Krishna MM, Englander SW (2007) A unified mechanism for protein folding: Pre-determined pathways with optional errors. *Protein Sci* 16(3):449–464.
28. Englander SW, Mayne L, Bai Y, Sosnick TR (1997) Hydrogen exchange: The modern legacy of Linderström-Lang. *Protein Sci* 6(5):1101–1109.
29. Bai Y, Milne JS, Mayne L, Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17(1):75–86.
30. Auton M, Holthauzen LM, Bolen DW (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA* 104(39):15317–15322.
31. Englander SW, Sosnick TR, Englander JJ, Mayne L (1996) Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol* 6(1):18–23.
32. Skinner JJ, et al. (2014) Benchmarking all-atom simulations using hydrogen exchange. *Proc Natl Acad Sci USA* 111(45):15975–15980.
33. Auton M, Rösger J, Sinev M, Holthauzen LM, Bolen DW (2011) Osmolyte effects on protein stability and solubility: A balancing act between backbone and side-chains. *Biophys Chem* 159(1):90–99.
34. Larios E, Pitera JW, Swope WC, Gruebele M (2006) Correlation of early orientational ordering of engineered lambda6-85 structure with kinetics and thermodynamics. *Chem Phys* 323(1):45–53.
35. Sosnick TR, Krantz BA, Dothager RS, Baxa M (2006) Characterizing the protein folding transition state using psi analysis. *Chem Rev* 106(5):1862–1876.
36. Baxa MC, et al. (2015) Even with nonnative interactions, the updated folding transition states of the homologs Proteins G & L are extensive and similar. *Proc Natl Acad Sci USA* 112(27):8302–8307.
37. Pandit AD, Jha A, Freed KF, Sosnick TR (2006) Small proteins fold through transition states with native-like topologies. *J Mol Biol* 361(4):755–770.
38. Sosnick TR, Dothager RS, Krantz BA (2004) Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc Natl Acad Sci USA* 101(50):17377–17382.
39. Fersht AR (2004) ϕ value versus ψ analysis. *Proc Natl Acad Sci USA* 101(50):17327–17328.
40. Bodenreider C, Kiefhaber T (2005) Interpretation of protein folding psi values. *J Mol Biol* 351(2):393–401.
41. Krantz BA, Dothager RS, Sosnick TR (2004) Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* 337(2):463–475.
42. Krantz BA, Dothager RS, Sosnick TR (2004) Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* 337(2):463–475, and erratum (2005) 347(5):1103.
43. Moran LB, Schneider JP, Kentsis A, Reddy GA, Sosnick TR (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc Natl Acad Sci USA* 96(19):10699–10704.
44. Krantz BA, Sosnick TR (2001) Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat Struct Biol* 8(12):1042–1047.
45. Meisner WK, Sosnick TR (2004) Barrier-limited, microsecond folding of a stable protein measured with hydrogen exchange: Implications for downhill folding. *Proc Natl Acad Sci USA* 101(44):15639–15644.
46. Kramers HA (1940) Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7:284–304.
47. Prigozhin MB, Chao SH, Sukenik S, Pogorelov TV, Gruebele M (2015) Mapping fast protein folding with multiple-site fluorescent probes. *Proc Natl Acad Sci USA* 112(26):7966–7971.
48. Yoo TY, et al. (2012) The folding transition state of protein L is extensive with non-native interactions (and not small and polarized). *J Mol Biol* 420(3):220–234.
49. Sosnick TR, Barrick D (2011) The folding of single domain proteins—Have we reached a consensus? *Curr Opin Struct Biol* 21(1):12–24.
50. Baxa MC, Freed KF, Sosnick TR (2009) Psi-constrained simulations of protein folding transition states: Implications for calculating. *J Mol Biol* 386(4):920–928.
51. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* 277(4):985–994.
52. Chugha P, Sage HJ, Oas TG (2006) Methionine oxidation of monomeric lambda repressor: The denatured state ensemble under nondenaturing conditions. *Protein Sci* 15(3):533–542.
53. Lacroix E, Viguera AR, Serrano L (1998) Elucidating the folding problem of alpha-helices: Local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J Mol Biol* 284(1):173–191.
54. Zheng Z, Sosnick TR (2010) Protein vivisection reveals elusive intermediates in folding. *J Mol Biol* 397(3):777–788.
55. Huyghues-Despointes BMP, Scholtz JM, Pace CN (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* 6(10):910–912.
56. Kimura T, et al. (2005) Specific collapse followed by slow hydrogen-bond formation of beta-sheet in the folding of single-chain monellin. *Proc Natl Acad Sci USA* 102(8):2748–2753.
57. Luan B, Lyle N, Pappu RV, Raleigh DP (2014) Denatured state ensembles with the same radii of gyration can form significantly different long-range contacts. *Biochemistry* 53(1):39–47.
58. Yoo TY, et al. (2012) Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. *J Mol Biol* 418(3–4):226–236.
59. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR (2007) Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 367(3):609–615.
60. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR (2004) Early collapse is not an obligate step in protein folding. *J Mol Biol* 338(2):369–382.
61. Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR (2005) Solution structure of a protein denatured state and folding intermediate. *Nature* 437(7061):1053–1056.
62. Pogorelov TV, Luthey-Schulten Z (2004) Variations in the fast folding rates of the lambda-repressor: A hybrid molecular dynamics study. *Biophys J* 87(1):207–214.
63. Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* 33(4):199–211.
64. Prigozhin MB, et al. (2013) Misplaced helix slows down ultrafast pressure-jump protein folding. *Proc Natl Acad Sci USA* 110(20):8087–8092.
65. Loftus D, Gbenle GO, Kim PS, Baldwin RL (1986) Effects of denaturants on amide proton exchange rates: A test for structure in protein fragments and folding intermediates. *Biochemistry* 25(6):1428–1436.
66. Lim WK, Rösger J, Englander SW (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc Natl Acad Sci USA* 106(8):2595–2600.