



Protein Vivisection Reveals Elusive Intermediates in Folding

Zhongzhou Zheng¹ and Tobin R. Sosnick^{1,2*}

¹*Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA*

²*Institute for Biophysical Dynamics, Computation Institute, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA*

Received 17 November 2009;
received in revised form
19 January 2010;
accepted 25 January 2010
Available online
6 February 2010

Although most folding intermediates escape detection, their characterization is crucial to the elucidation of folding mechanisms. Here, we outline a powerful strategy to populate partially unfolded intermediates: A buried aliphatic residue is substituted with a charged residue (e.g., Leu → Glu[−]) to destabilize and unfold a specific region of the protein. We applied this strategy to ubiquitin, reversibly trapping a folding intermediate in which the β 5-strand is unfolded. The intermediate refolds to a native-like structure upon charge neutralization under mildly acidic conditions. Characterization of the trapped intermediate using NMR and hydrogen exchange methods identifies a second folding intermediate and reveals the order and free energies of the two major folding events on the native side of the rate-limiting step. This general strategy may be combined with other methods and have broad applications in the study of protein folding and other reactions that require trapping of high-energy states.

© 2010 Elsevier Ltd. All rights reserved.

Edited by K. Kuwajima

Keywords: Psi-analysis; ubiquitin; protein engineering; native-state hydrogen exchange; NMR

Introduction

Multiple points on the free energy surface must be structurally and thermodynamically characterized to fully delineate protein folding mechanisms. However, this task is extremely challenging, in particular for proteins that fold in an apparent two-state manner. Folding intermediates tend to be transient and weakly populated in both kinetic and equilibrium studies. Two major strategies are used to address this challenge. The most straightforward and information-rich method is direct trapping followed by structural characterization. However, trapping often involves the use of

unnatural conditions (e.g., low pH, cosolvents)^{1,2} or relies on protein-specific mutagenesis.^{3–8}

The other major strategy involves the detection of sparsely populated intermediates—for example, using NMR relaxation dispersion (RD),^{7,9} native-state hydrogen exchange (NSHX),^{10,11} chemical modification,^{12,13} or proteolysis.^{14,15} RD methods can characterize excited states that are populated as little as 0.5% and provide interconversion rates, stability, and the chemical shifts of the excited state.⁹ RD methods have characterized intermediates that reside either before or after the rate-limiting transition state (TS) for a number of proteins.^{16–18}

NSHX can detect partially unfolded intermediates by examining the denaturant dependence of the HX rates. The regions of the protein that unfold cooperatively in subglobal openings are termed foldons. Foldons are typically associated with elements of secondary structure. Trapped intermediates have been created by selectively destabilizing or deleting a foldon.^{6,19–22} The resulting intermediates generally are on-pathway, located before or after the rate-limiting step in two-state or multi-state folding, respectively.

RD and NSHX methods are powerful but have certain limitations. RD requires appropriate interconversion rates and chemical shift differences, and the intermediate must comprise at least ~0.5% of the total population. NSHX requires that the hydrogen

*Corresponding author. Department of Biochemistry and Molecular Biology, University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA. E-mail address: trsosnic@uchicago.edu.

Abbreviations used: HSQC, heteronuclear single quantum coherence; HX, hydrogen exchange; NSHX, native-state hydrogen exchange; pWTUb, pseudo-wild-type ubiquitin; R_g , radius of gyration; RD, relaxation dispersion; SAXS, small-angle X-ray scattering; TSE, transition-state ensemble; Ub, mammalian ubiquitin; UbL50E, Leu50Glu version of pseudo-wild-type Ub.

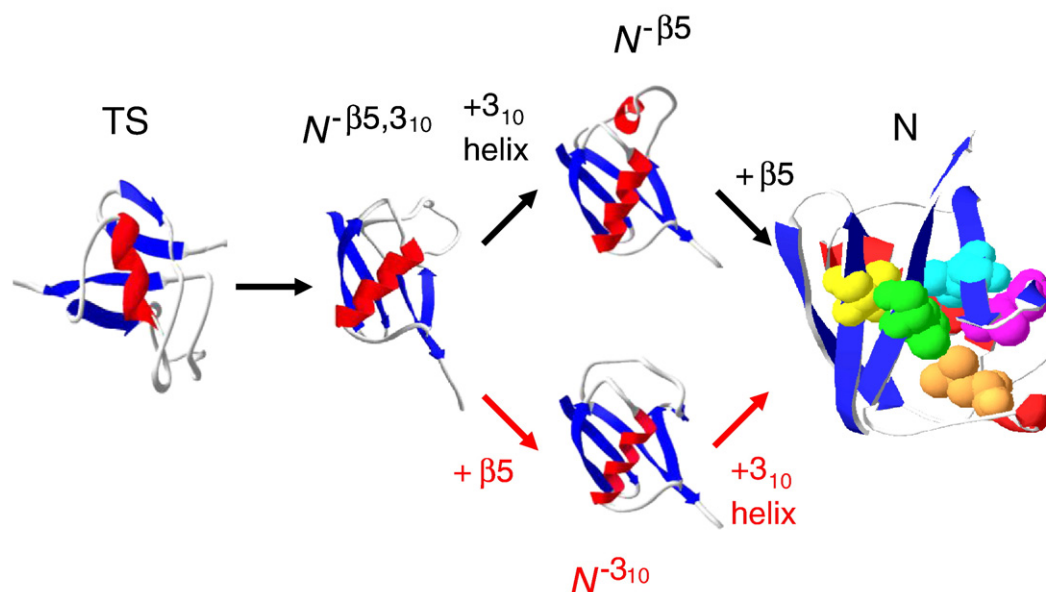


Fig. 1. Post-TS event in Ub folding. This study indicates that the folding of the 3_{10} -helix and that of the $\beta 5$ -strand preferentially occur along the upper pathway. Mutations that were employed in the study strategy are highlighted (V5, yellow; L43, blue; L50, magenta; L56, orange; L67, green), except I30E on the interior face of the helix. The three-dimensional models of the folding intermediates were created by changing the ϕ, ψ angles of the disrupted regions while leaving the remainder of the protein unchanged.

bonds broken in the subglobal opening are not hidden by faster exchange processes, including smaller-scale (local) openings (Supplementary Fig. 1). In addition, the free energy and the size of the subglobal opening must be such that its HX “isotherm” is distinct from those of global and local unfolding processes. These requirements may explain why NSHX studies on mammalian ubiquitin (Ub) did not reveal any subglobal openings.²³

These issues call for new strategies for trapping folding intermediates. In this study, we explored a strategy that involves reversibly destabilizing a foldon. The destabilization is accomplished by replacing a buried aliphatic residue with a nearly isosteric but charged acidic residue (e.g., Leu \rightarrow Glu[−]).^{24,25} At neutral pH, the energetic penalty associated with the burial of the acidic group destabilizes the foldon, thereby creating an equilibrium intermediate amenable for detailed characterization. We applied this strategy to Ub to reversibly populate an intermediate that had escaped detection by NSHX. Kinetic folding experiments indicate that it is a late folding intermediate on the native side of the rate-limiting barrier. The application of NSHX to the trapped folding intermediate identifies a second unfolding intermediate. These results reveal a hierarchical unfolding pathway beginning with the loss of the $\beta 5$ -strand followed by the disruption of a small 3_{10} -helix.

Results

Rationally populating intermediates

We have previously proposed a detailed folding pathway for Ub using a combination of results

including HX¹ and ψ -analysis, our bihistidine metal binding method that identifies inter-residue contacts in the TS ensemble (TSE).^{26–28} Each major pathway step both on the way to and from the TSE is the addition or consolidation of regions of secondary structure. These events occur with a commensurate level of hydrogen-bond formation and surface burial, as indicated by our kinetic isotope effect studies.^{29,30} The only secondary structures that form after the TS are the small 3_{10} -helix and the $\beta 5$ -strand (Fig. 1). This identification suggests the presence of one or two post-TS, kinetically silent intermediates lacking one or both of these structural elements. Such intermediates are likely to be unfolding intermediates whose thermodynamics may be characterized using HX. An earlier HX study¹ indicated that hydrogen bonds between $\beta 4$ and $\beta 5$ generally are ~ 1 kcal mol^{−1} less stable than bonds in the 3_{10} -helix. This difference suggests that the $\beta 5$ -strand unfolds prior to the helix on the unfolding pathway (N \rightarrow N^{− $\beta 5$} \rightarrow N^{− $\beta 5,3_{10}$} \rightarrow TS; Fig. 1).

In order to investigate the sequential order of unfolding events, we set out to trap one or both of these intermediates. The buried leucine on $\beta 5$ is replaced with a glutamic acid (UbL50E). The energetic penalty for burying the charged Glu[−], or, more precisely, the energy required to shift acid's pK_a so that it remains neutral at a pH above its intrinsic pK_a,²⁵ is anticipated to unfold the β -strand, thereby populating the N^{− $\beta 5$} conformation. This unfolding event would be analogous to the disruption of the LOV2 protein's signaling helix upon an I \rightarrow E substitution²⁴ and the partial unfolding of a helix in SNase,³¹ but it is possibly different from other results with this protein as nearly all substitutions are well tolerated.²⁵ For UbL50E at pH below

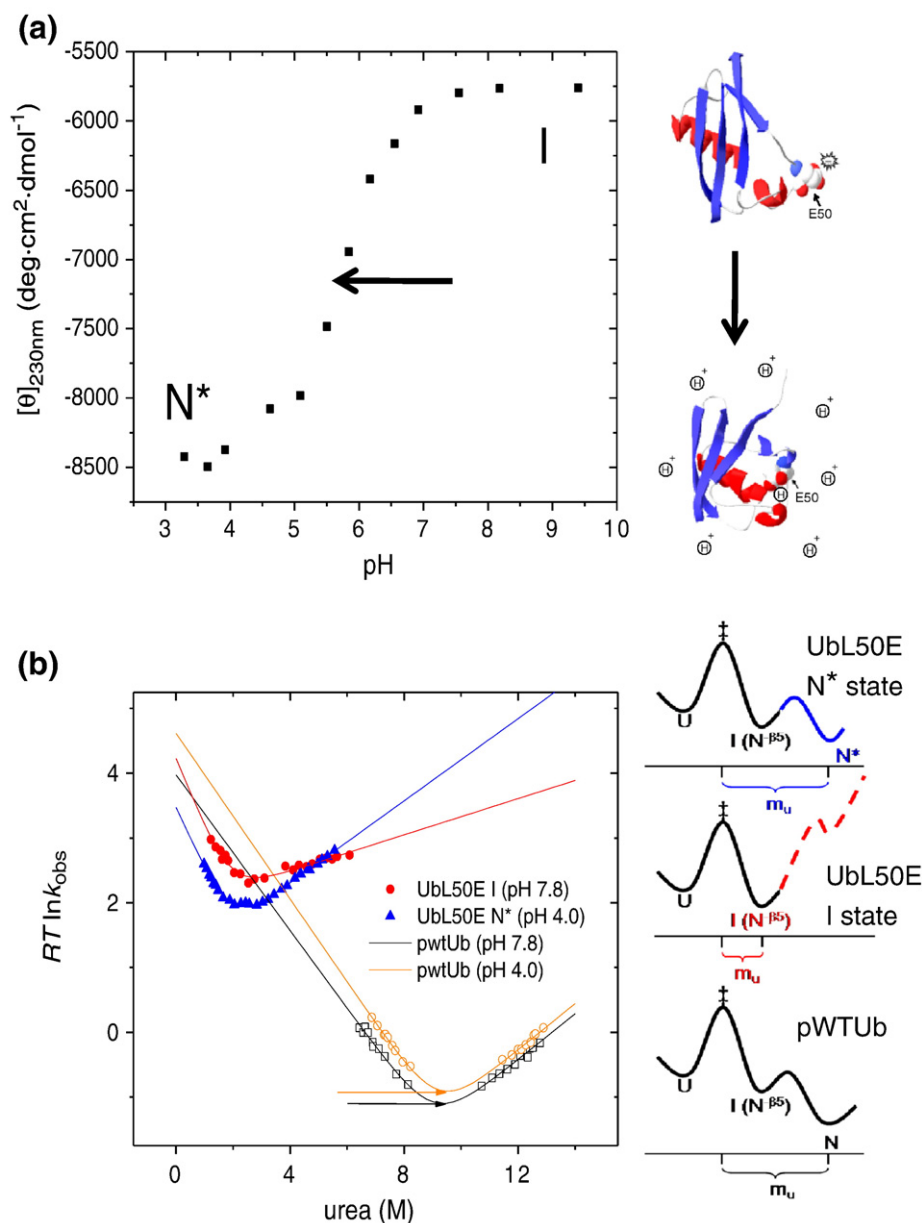


Fig. 2. Equilibrium and kinetic folding behaviors of UbL50E. (a) Upon acidification, the protein shifts from an intermediate to a native-like state. (b) At both high and low pH, the chevron plots of UbL50E and pWTUb have nearly overlapping folding arms but distinct unfolding arms. At high pH, the UbL50E unfolding arm is shallower, suggesting disruption of $\beta 5$ -strand in the intermediate state. Illustrative folding energy landscapes are shown at the right. Chevron values for the pWTUb have been shifted by the amount indicated by the arrows to account for the addition of 2 or 2.3 M guanidine hydrochloride required to unfold the pWTUb.

glutamic acid's intrinsic pK_a , the nearly isosteric Glu⁰ may adequately mimic the Leu, allowing the β -strand to refold and generate the native-like N* conformation. If successful, this strategy provides a system whereby pH can be used to reversibly convert Ub from a native-like state to a late folding intermediate, which can be characterized using equilibrium methods.

Thermodynamic and kinetic properties

Structural, kinetic, and thermodynamic measurements are performed on UbL50E to investigate the

ability of the charge-burial strategy to trap a partially unfolded state. A pH titration of UbL50E at 4 °C, monitoring near-UV circular dichroism (CD), indicates the presence of a minor unfolding transition with a midpoint near pH 6.0 (Fig. 2a; a similar transition is observed by NMR following the chemical shift of the E50 amide proton, see below). At pH 4.0, the protein has approximately the wild-type $[\theta]_{230\text{ nm}}$ value, while the high-pH species has lost structure ($\Delta[\theta]_{230\text{ nm}} = 2700\text{ deg cm}^2\text{ dmol}^{-1}$). The observed pH transition midpoint is 1.7 units above the intrinsic pK_a^{Glu} because $\beta 5$'s folding energy helps stabilize the folding transition by

Table 1. Kinetic parameters obtained from chevron analysis

Protein	Solvent condition	m_f (kcal mol ⁻¹ M ⁻¹)	m_u (kcal mol ⁻¹ M ⁻¹)	$\Delta G_f^{\ddagger}(0)$ (kcal mol ⁻¹)	ΔG° (kcal mol ⁻¹)
pWTUb (Fig. 2)	N state, pH 7.8	0.60±0.04	0.35±0.02	3.97±0.24	-8.64±0.35
pWTUb (Fig. 2)	N state, pH 4.0	0.64±0.03	0.36±0.02	4.62±0.21	-9.18±0.28
L50E (Fig. 2)	I state, pH 7.8	1.08±0.15	0.14±0.02	4.22±0.21	-2.29±0.17
L50E (Fig. 2)	N* state, pH 4.0	0.97±0.05	0.32±0.01	3.47±0.06	-2.43±0.04
L50E (Supplementary Fig. 3, HX condition)	I state, pD _{read} 7.5	0.71±0.02	0.18±0.01	4.18±0.05	-2.98±0.07
L67E (Fig. 8)	I state, pH 7.8	0.63±0.02	0.21±0.01	4.31±0.04	-4.00±0.07
L67E (Fig. 8)	N* state, pH 4.0	0.68±0.02	0.28±0.01	4.10±0.08	-4.39±0.07

2 kcal mol⁻¹. This amount is 4 kcal mol⁻¹ less than the free energy for opening $\beta 5$ in the F45W pseudo-wild-type Ub (pWTUb; see HX results below), presumably because the Glu⁰ does not fully mimic the Leu³¹ in the low-pH N* species.

The folding rates of UbL50E are measured as a function of denaturant concentration using chevron analysis and stopped-flow methods following F45W fluorescence (Fig. 2) to further characterize the low- and high-pH conformations. At pH 7.8, the folding rates of UbL50E and pWTUb are statistically identical ($\Delta\Delta G^{\ddagger} = -0.25 \pm 0.32$ kcal mol⁻¹) in spite of the considerable destabilization imparted by the substitution of the acidic residue ($\Delta\Delta G_{eq}^{L50E}$ of $\sim 6.4 \pm 0.4$ kcal mol⁻¹). At pH 4.0, the folding rate of UbL50E is mildly slower than that of pWTUb. In energetic terms, however, the difference ($\Delta\Delta G^{\ddagger} = 1.1 \pm 0.2$ kcal mol⁻¹) remains much smaller than the destabilization ($\Delta\Delta G_{eq}^{L50E} = 6.8 \pm 0.3$ kcal mol⁻¹). The resulting mutational ϕ^{L50E} values are 0.16 ± 0.03 and 0.04 ± 0.05 at pH 4.0 and pH 7.8, respectively. These low values are consistent with our identification of $\beta 5$ as a structure formed after the TS along the folding pathways of pWTUb and UbL50E.

At both low and high pH, the folding rates of the two proteins have the same urea dependence (m_f -value; Table 1). This correspondence indicates that a comparable amount of urea-sensitive surface area is buried in their TSs (relative to their denatured states). This result, along with the similarity of their folding rates and the lack of $\beta 5$ in their TS, suggests that the two proteins have very similar TS structures at both pH conditions.

Although the folding rate of UbL50E is largely unchanged, its unfolding rate at both pH 4.0 and pH 7.8 is much faster than the unfolding rate of pWTUb. This acceleration is due to the stabilizing effect of burying the acidic group in UbL50E. In addition at higher pH, the slope of the chevron unfolding arm (m_u -value) for UbL50E is shallower by $\sim 15\%$ of the equilibrium m_o -value as compared with pWTUb. This decrease in slope for UbL50E indicates that the starting state of its unfolding reaction has lost some structure, presumably the destabilized $\beta 5$ -strand. This result is suggestive that high-pH conformation of UbL50E is a folding intermediate on the native side of the major folding barrier.

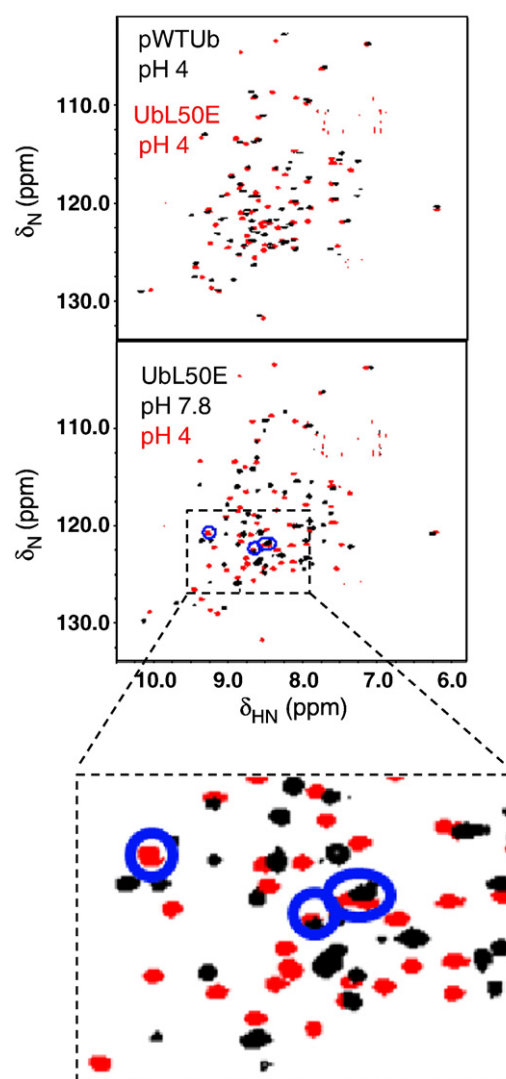


Fig. 3. NMR analysis. Overlay of ¹H-¹⁵N HSQC spectra of pWTUb and UbL50E at pH 4.0 (upper panel). The extensive peak overlap indicates that the two proteins have similar conformations. At pH 7.8, UbL50E has sharp and well-dispersed peaks, although with fewer resonances than at pH 4.0 (lower panel). The four peaks (blue circles) observable at pH 4.0, but not at pH 5.0, are attributed to the four unassigned residues on $\beta 5$.

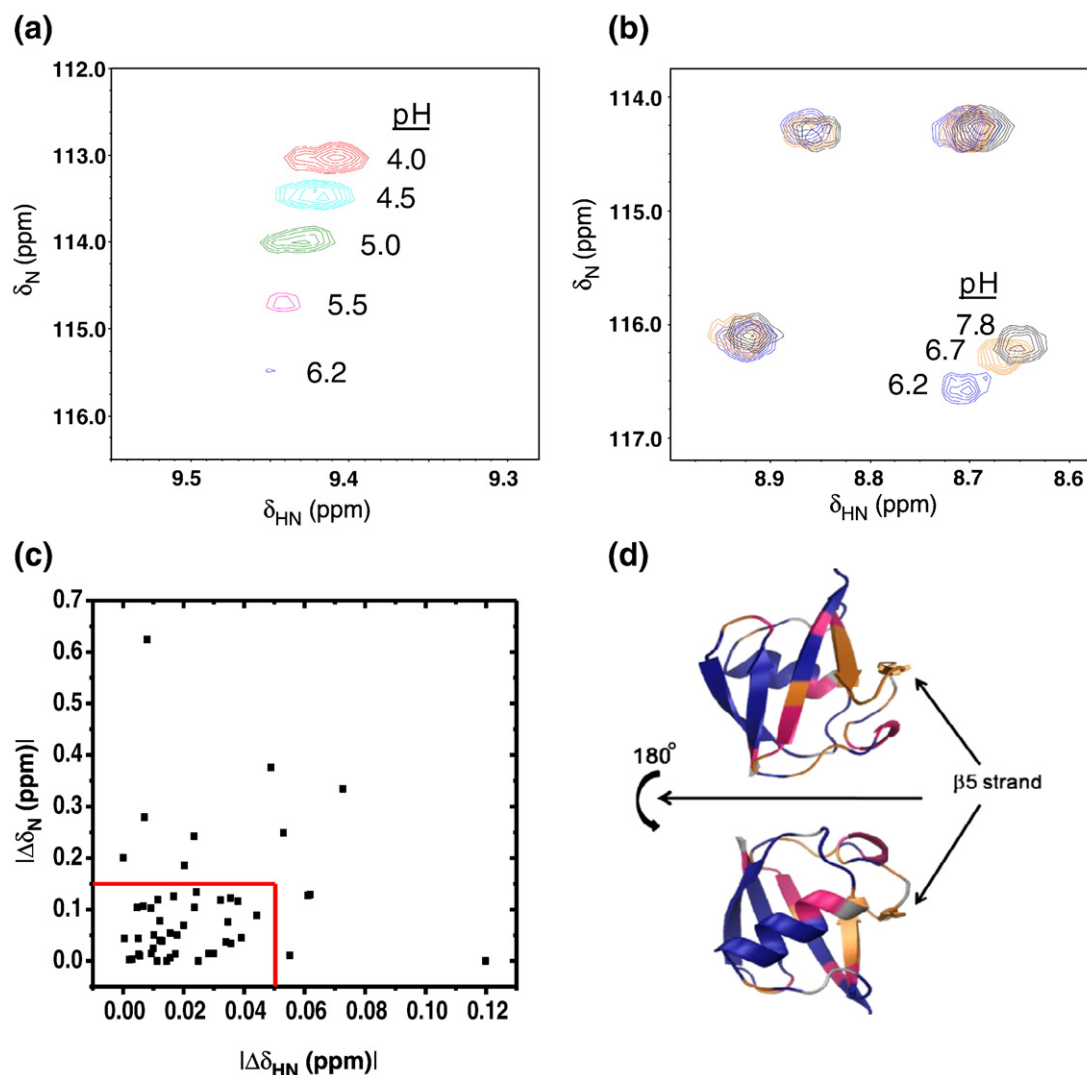


Fig. 4. pH titration of UbL50E from pH 4.0 to pH 7.8. (a) Representative resonance that shifts and then disappears due to exchange broadening. (b) In the pH range of 6.2–7.8, some resonances move, while others do not. (c) Resonances were classified as non-moving if their chemical shift remained unchanged from pH 6.2 to pH 7.8 (red box). (d) The three classes of behavior of UbL50E peaks are mapped onto the structure: unchanged (blue), changed (magenta), and chemical shift changed with peak disappearing at neutral pH due to exchange broadening (orange).

NMR analysis

A series of ^1H – ^{15}N heteronuclear single quantum coherence (HSQC) spectra were acquired from pH 4 to pH 7.8 (Figs. 3 and 4) to characterize the low- and high-pH forms of UbL50E in detail. The spectra of the high- and low-pH states contain well-dispersed sharp peaks, indicating that both states contain well-structured regions. At pH 4.0 (Fig. 3) and slightly higher pH (Fig. 4), most of the amide resonances overlapped or remained close to the pWTUb's resonances, suggesting that the UbL50E with Glu⁰ retains a very native-like structure.

Resonances were assigned using triple resonance HNCACB and CBCACONH measurements in conjunction with the pWTUb assignments (F. Massi, unpublished data). To obtain a concentrated protein sample for the three-dimensional measurements, we used a solubility-enhancing buffer (22) (15 mM

sodium phosphate, 225 mM sodium chloride, 50 mM glutamic acid, and 50 mM arginine). The enhancement is best near neutral pH, but the NMR spectra gradually deteriorate from pH 5.0 to pH 7.8. Accordingly, assignments were conducted at pH 5.3, where all NH resonances are assignable except for residues 48–51 on the $\beta 5$ -strand. These four resonances were not observed, presumably due to exchange broadening (pH 5.3 is in the middle of the pH-induced folding transition). At pH 4.0, four additional peaks appeared in the region appropriate for amide NH resonances. Because the other amide resonances are accounted for, we associated the new resonances with the four missing $\beta 5$ resonances (Fig. 3). The assignment of peaks at other solvent conditions was accomplished by tracking peak movement during the pH titration (Fig. 4a and b).

In the pH titration, all the NH peaks shift, and some start to disappear near pH 6.0. The shifts

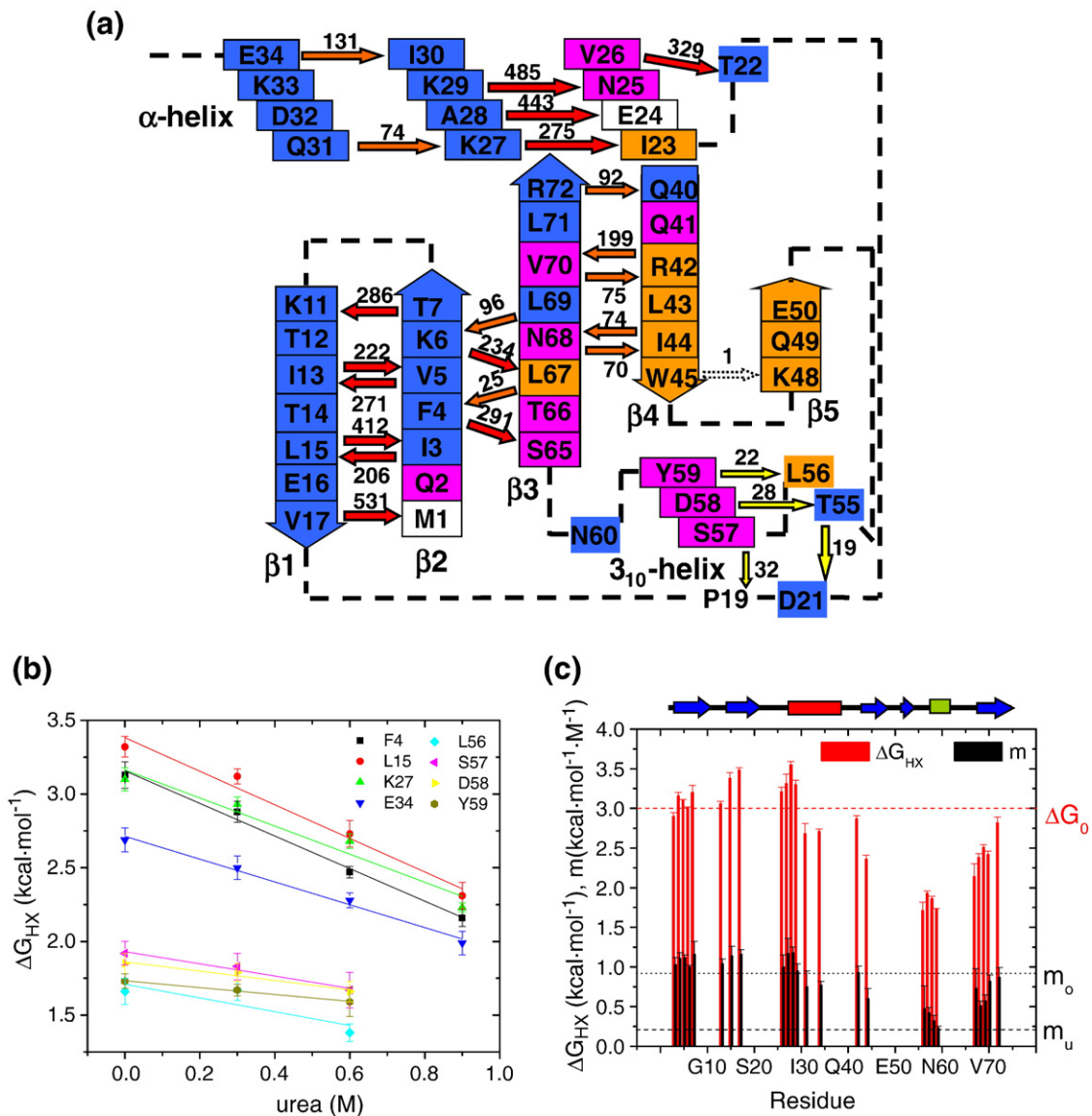


Fig. 5. NSHX on UbL50E I state at pD_{read} 7.5. (a) Hydrogen-bond network with direction of bond (NH→O=C) and K_{eq} noted. Arrows are color-coded according to their stability (red indicates global; yellow, 3₁₀ opening; orange, other; white, absent), while boxes are color-coded according to the type of chemical shift movement during the pH titration (adapted from Fig. 4). (b) Representative NSHX denaturant-dependent isotherms. Isotherms for protons on the 3₁₀-helix have lower ΔG_{HX} and m -values compared with the rest of the isotherms. (c) Histogram of ΔG_{HX} and m -values. The positions of the strands (blue), α -helix (red), and 3₁₀-helix (green) are noted at the top.

reflect both the folding transition and the less interesting ionization of the surface-exposed Glu and Asp groups. Fortunately, their ionization equilibria are ~2 pH units below the midpoint of the CD-monitored folding transition. Consequently, the surface Glu and Asp side chains are nearly fully ionized by pH 6.0, and hence the remaining chemical shift movement at higher pH levels reflects the partial unfolding transition.

Across the pH 6.2–pH 7.8 range, 48%, 28%, and 24% of the observable NH resonances remained unchanged ($|\Delta\delta_{HN}|$, <0.05 ppm; $|\Delta\delta_N|$, <0.15), shifted, and shifted and then disappeared, respectively (Fig. 4). These three classes are interpreted as minimal or significant backbone conformational

change, which may also be accompanied by extreme exchange broadening due to millisecond timescale dynamics. The residues composing each class map onto the native structure of Ubs according to their distance from the β 5-strand (Fig. 4d) and ψ -values.²⁶ The resonances for the β 1– β 2 hairpin and carboxy-terminus of the α -helix, the most distal regions, are the most invariant, while the resonances for β 4, 3₁₀-helix, located adjacent to β 5, are the most perturbed. The exchanged broadened peaks do not reappear as sharp peaks with random-coil chemical shifts. Hence, the β 5-strand is not behaving as a completely unfolded loop. Presumably, the loop interacts dynamically with the rest of the protein.

Table 2. NSHX on UbL50E intermediate, $pD_{\text{read}} 7.5$

Secondary structure of H-bond donor	H-bond donor (-NH)	H-bond acceptor (-C=O)	ΔG_{HX} (kcal mol ⁻¹)	m (kcal mol ⁻¹ M ⁻¹)
$\beta 2$ -strand	I3	L15	2.90±0.05	1.03±0.09
	F4	S65	3.16±0.04	1.11±0.07
	V5	I13	3.10±0.01	1.12±0.03
	K6	L67	3.01±0.01	1.00±0.02
	T7	K11	3.20±0.09	1.16±0.16
$\beta 1$ -strand	I13	V5	3.06±0.03	1.04±0.06
	L15	I3	3.38±0.07	1.16±0.06
	V17	M1	3.48±0.03	1.16±0.06
α -helix	V26	T22	3.21±0.06	1.00±0.15
	K27	I23	3.31±0.12	1.17±0.19
	A28	E24	3.55±0.04	1.18±0.07
	K29	N25	3.30±0.06	0.95±0.09
	Q31	K27	2.68±0.13	0.75±0.20
$\beta 4$ -strand	E34	I30	2.71±0.03	0.77±0.05
	R42	V70	2.87±0.04	0.93±0.08
	I44	N68	2.36±0.05	0.60±0.13
	W45	K48	~0 ^a	ND ^b
3_{10} -helix	L56	D21	1.71±0.11	0.47±0.29
	S57	P19	1.93±0.03	0.42±0.07
	D58	T55	1.86±0.03	0.32±0.07
	Y59	L56	1.73±0.01	0.23±0.02
$\beta 3$ -strand	L67	F4	2.14±0.16	0.73±0.25
	N68	I44	2.38±0.05	0.51±0.06
	L69	K6	2.51±0.03	0.57±0.08
	V70	R42	2.42±0.04	0.82±0.08
	R72	Q40	2.82±0.07	0.87±0.12

^a $k_{\text{HX}}/k_{\text{int}}$ of ~1.^b Not determined.

Hydrogen exchange

HX measurements are conducted on the intermediate form of UbL50E to characterize its structure and thermodynamics, and potentially identify other partially unfolded states. HX probes thermally activated fluctuations that break backbone hydrogen bonds. Once the bond is broken, HX between amide and solvent protons occurs according to:^{11,32}



where k_{open} and k_{close} are the opening and closing rates of the fluctuation, respectively, and k_{int} is the intrinsic HX rate, which depends upon the amino acid sequence and solvent conditions.^{33,34} For a stable protein in the EX2 limit where $k_{\text{close}} \gg k_{\text{int}}$, the protein establishes an equilibrium between the ground and excited states. The observed HX rate reduces to the fraction of time the protein is in the excited state multiplied by the intrinsic rate of exchange, $k_{\text{HX}} = k_{\text{int}} / (1 + k_{\text{close}}/k_{\text{open}}) = k_{\text{int}} / (1 + 1/K_{\text{eq}})$. The free energy of the excited state is calculated according to $\Delta G_{\text{HX}} = -RT \ln(K_{\text{eq}}) = -RT \ln(k_{\text{HX}} / (k_{\text{int}} - k_{\text{HX}}))$, where RT is the product of the gas constant and temperature.

Under our experimental conditions ($pD_{\text{read}} 7.5$, 4 °C), intrinsic exchange rates are from seconds to minutes, which are faster than standard HSQC acquisition times. Accordingly, a pulse-labeling strategy is employed to measure HX rates. HX is initiated with a 1:4 dilution with D₂O to a pD_{read} of

7.5. After a specified time, H-to-D exchange is quenched by reducing the pD_{read} to 2.9 with DCl. In addition to quenching HX, acidification also generates the N* state where the exchange-broadened resonances in the intermediate state become sharp and assignable. Hence, the HX rates for otherwise NMR invisible residues in the intermediate can be determined using this protocol. HX likely occurs via the EX2 mechanism as k_{int} is much slower than global refolding rates. Hence, the stability of the hydrogen bonds can be determined for ~80% of the intramolecular hydrogen bonds observable in the crystal structure.

In the intermediate, significant HX protection is observed for all of the native secondary structures except for the single measured hydrogen bond between strands $\beta 4$ and $\beta 5$ ($\text{NH}^{\text{W45}} \rightarrow \text{O}=\text{C}^{\text{K48}}$). For this bond, $k_{\text{HX}}/k_{\text{int}} = 0.3 \pm 0.3 \text{ s}^{-1} / 0.3 \text{ s}^{-1}$. Hence, the protection factor or essentially is unity. The four unassigned peaks associated with $\beta 5$ also have a $k_{\text{HX}}/k_{\text{int}}$ of ~1 (observed HX is complete within 3 s; k_{int} of ~0.5–3 s⁻¹). Hence, the hydrogen bonds between $\beta 4$ and $\beta 5$ are broken most of the time. For the remaining hydrogen bonds, ΔG_{HX} is between 1.7 and 3.5 kcal mol⁻¹ (Fig. 5; Table 2; Supplementary Fig. 2).

Small-angle X-ray scattering

We compared the global dimensions of pWTub with the UbL50E intermediate using small-angle X-ray scattering (SAXS) (Fig. 6). At pH 7.8, the measured radii of gyration (R_g 's) are 13.2±0.2 and 13.8±0.1 Å for the two proteins, respectively. This difference matches the 0.5-Å difference between R_g values of the native state and the model for the N- $\beta 5$ intermediate but is less than the 1.6-Å difference for the N- $\beta 5,3_{10}$ intermediate (Fig. 1).

In summary, a combination of results—the loss of secondary structure (CD, pH titration) that is formed

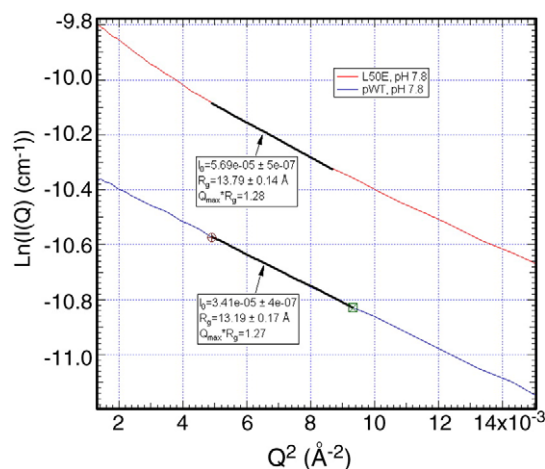


Fig. 6. Dimensions of the native pWTub and the L50E intermediate measured using SAXS. R_g values were obtained from a Guinier analysis performed over the data range highlighted. Protein concentration is ~2 mg/mL. Solvent condition was the same as that used for NMR assignments except at pH 7.8, 0% D₂O.

L67E; Figs. 1 and 8). At neutral pH, the L56E substitution on the 3_{10} -helix is anticipated by folding hierarchy to unfold both this helix and the $\beta 5$ -strand, thereby populating the $N^{-\beta 5,3_{10}}$ state. This species is unstable and prone to precipitate with a virtually blank HSQC spectrum presumably due to extreme line broadening. Nevertheless, this species refolds at pH 4.0 with an NMR spectrum having similar dispersion as the native protein. The potential loss of the 3_{10} -helix could be tested with HX labeling at neutral pH followed by NMR analysis at low pH, as done for UbL50E. Irrespective of the exact structural content, the NMR spectrum of the L56E species lacks sharp and dispersive peaks at neutral pH, suggesting that the protein becomes more dynamic on the unfolding pathway back toward the TS.

The L67E and L43E substitutions are located on $\beta 3$ and $\beta 4$, respectively. These two strands are present in the TS according to ψ -analysis. Both mutants have well-dispersed HSQC spectra at both pH 4.0 and pH 8.0. The L67E species has a native number of peaks, while L43E has slightly less at pH 8.0. The kinetic folding properties of the L67E mutant are similar to those of L50E, in particular possessing a shallower chevron unfolding arm slope at pH 8.0 as compared with pWTUb (Fig. 8b; Table 1). However, the $\sim 10\%$ reduction of the m_u -value for L67E is insufficient to account for disruption of the entire $\beta 3$ -strand, which makes multiple interactions with $\beta 2$ and $\beta 4$ in the TS. We infer that the L67E mutation induces some local structural fraying of $\beta 3$ at high pH that exposes some buried surface area, rather than the loss of the entire strand.

The V5D and I30E mutations are located on the amino-terminal hairpin and the α -helix, respectively. These two secondary structures form prior to the TS in the folding pathway, and the two substituted positions are located in the core of the protein. The HSQC spectrum of V5D exhibits extensive peak broadening at pH 8.5, whereas I30E's spectrum remains dispersive with extra sets of peaks. At pH 4.0, the V5D spectrum has less dispersion than I30E, which is similar to pWTUb. The V5D substitution is disruptive, indicating that charge burial can identify regions of the protein that cannot be unfolded without disrupting the rest of the protein. Such regions are likely to form early in the folding pathway.

Discussion

Rather than relying on serendipitously found mutations, our charge-burial strategy is a rational, active, and generally applicable approach for reversibly trapping partially unfolded states and delineating folding pathways. For a protein where little is known about its folding behavior, the charge-burial strategy can be extremely revealing. Leu/Ile \rightarrow Glu or Val \rightarrow Asp substitutions can be introduced into each potential foldon. The acquisition of NMR spectra under acidic and neutral conditions can identify which parts of the protein can be selectively disrupted without unfolding the

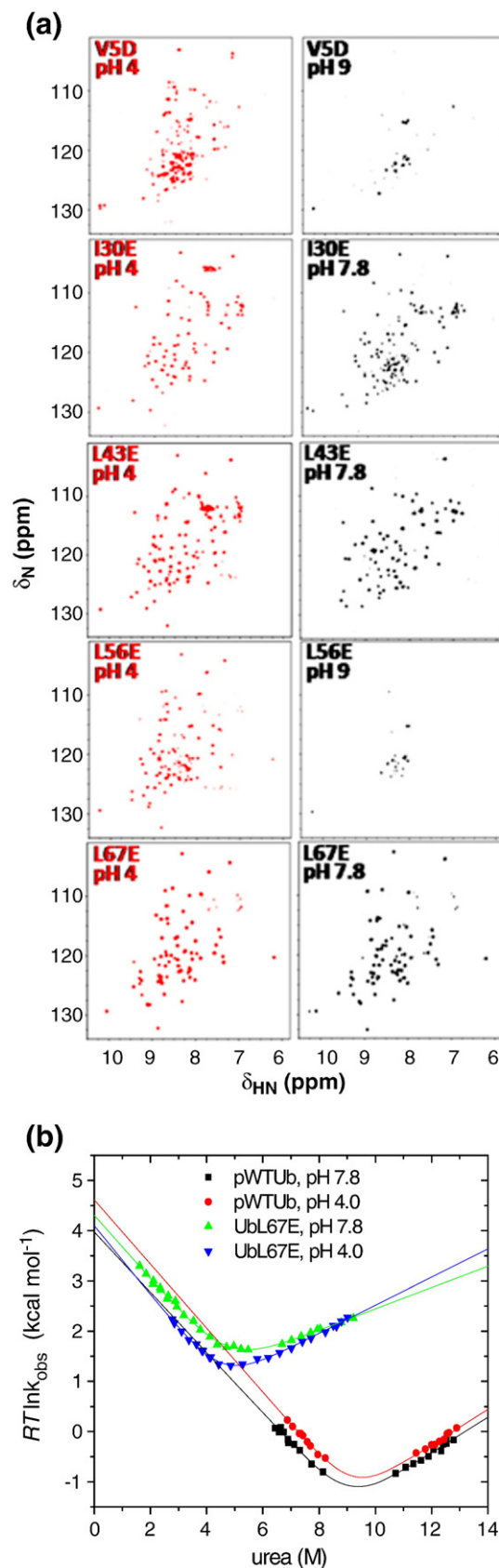


Fig. 8. Charge-burial partial unfolding strategy applied at other positions. (a) NMR ^1H - ^{15}N HSQC spectra. (b) Corresponding chevron plots of UbL67E at low and high pH compared with pWTUb.

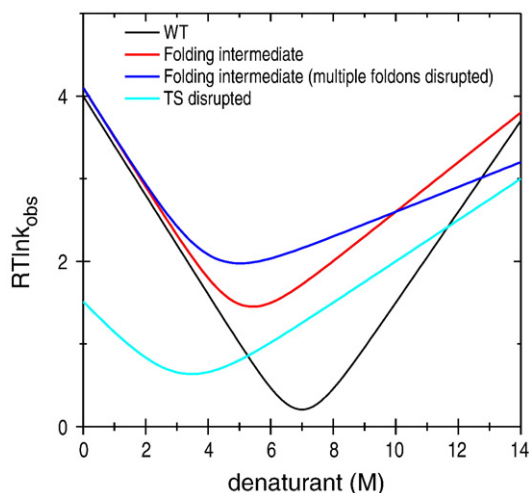


Fig. 9. Interpretation of possible chevrons generated by the charge-burial strategy. For an aliphatic-to-charged substitution, the chevron can have a similar folding arm but an altered unfolding arm at neutral pH. The change in slope can be used to estimate the amount of structure lost due to the substitution. If the decrease of slope is consistent with the estimate (red), it is suggestive of the creation of a folding intermediate lacking structure near the site of the substitution that otherwise is formed after the TS in the folding of the wild-type protein. If the decrease of slope is smaller than the estimate (not shown), it may be indicative of a locally frayed conformation rather than a genuine folding intermediate. Alternatively, the decrease may be consistent with the loss of an additional foldon (blue) if folding is hierarchical, being dependent on the perturbed foldon. This possibility can be tested by further characterization (e.g., with NSHX). Changes in the folding arm indicate that the TS is destabilized by the substitution. When the slope of the folding is decreased, some structure has been lost in both the TS and the ground state (cyan). More complicated scenarios are also possible.

entire protein. Once a potential folding intermediate is found through such substitutions, it can be further interrogated with HX, chevron analysis (Fig. 9), and, possibly, structure determination. The intermediate can also be used as the new ground state to perform additional NSHX and RD measurements to “walk” up the energy surface, identify other folding intermediates, and test tentative conclusions from the preliminary scan using charged residues.

By introducing the L50E substitution into Ub, we trapped the $N^{-\beta 5}$ species, a late folding intermediate, and characterized its structure, thermodynamics, and dynamics using CD, NMR, SAXS, and chevron analysis. This trapped species has very native-like regions on the opposite side of the protein (no shift in the sharp ^{15}N - ^1H resonances). The disrupted $\beta 5$ -strand and neighboring regions, however, undergo extreme exchange broadening due to millisecond timescale dynamics, although the backbone hydrogen bonds remain in the 3_{10} -helix and between strands $\beta 3$ and $\beta 4$.

We performed NSHX measurements on the intermediate, taking advantage of the well-behaved NMR spectrum of the low-pH native-like species.

These measurements identified an additional unfolding intermediate lacking the 3_{10} -helix. As far as we are aware, this is the first report of a folding intermediate identified through NSHX applied to another intermediate.

Even for an equilibrium study that identifies an extensive ladder of unfolding intermediates, the kinetic connectivity can only be inferred. In the case of Ub, we have additional kinetic information. The only secondary structures that form between the TS native state are the $\beta 5$ -strand and the 3_{10} -helix.²⁶ Hence, the unfolding of either of these elements is highly likely to be the first step along the unfolding pathway. Our work demonstrates that a species lacking only the $\beta 5$ -strand lies at a lower free energy than the state lacking the 3_{10} -helix and that the $\beta 5$ -strand can be individually disrupted. Furthermore, the stability of the 3_{10} -helix responds to the L50E substitution to the same degree as does the $\beta 5$ -strand, thereby demonstrating that a structural hierarchy exists between these two foldons. We also found that the unfolding rate starting from the $N^{-\beta 5}$ species is faster than that from N yet the pathways go through the same TS, and the trapped species is in between the native state and TS when considering urea-sensitive surface burial (m -value). All these results point to the early stages of the kinetic unfolding pathway going through two intermediates, $N \rightarrow N^{-\beta 5} \rightarrow N^{-\beta 5, 3_{10}} \rightarrow \text{TS}$.

When NSHX was performed with wild-type Ub, neither of these two intermediates was readily detectable.²³ Similarly, RD methods on wild-type protein would not have detected these intermediates because their populations are too low ($<0.002\%$). However, RD measurements could be applied to the trapped L50E intermediate. Such measurements should detect the $N^{-\beta 5, 3_{10}}$ species as it now comprises 3% of the total population. This potential application illustrates the possibilities of combining the charge-burial strategy with other methods to enable structural and thermodynamic characterization of folding intermediates.

Other trapping strategies and dynamics in intermediate states

Other protein engineering strategies have been used to trap intermediates. However, these methods often lack the degree of control possible with the charge-burial strategy, and they typically require advanced knowledge of the structural content of the intermediate. For example, Bai and coworkers used HX in conjunction with truncation and glycine substitutions to create analogs of intermediates of *Thermus thermophilus* RNase H³⁵ and T4 lysozyme.²⁰ They conducted NSHX on the T4 lysozyme analog but found no additional intermediate.²⁰ Based on HX results on *Escherichia coli* RNase H, Marqusee and coworkers generated a stable mini-core²² and a fractionally populated intermediate using a single I25A substitution.⁸ Other late intermediates have been created using glycine substitutions in cyt-b562²¹ and IM9.⁶

The NMR characterization of the IM9 intermediate indicated that it has a fluid core with fluctuations on the NMR timescale due to the repacking of the remaining three helices. This repacking is similar to that of the remaining helices in the cyt-b562 intermediate.²¹ However, this intermediate and the *T. thermophilus* RNase H intermediate³⁵ remain very native-like according to NMR line widths, as do the regions distal to the perturbation in the UbL50E intermediate. In contrast, the *E. coli* I25A intermediate has more dynamic behavior.⁸ We suggest that an I25D charged substitution would fully populate the *E. coli* intermediate and help address the apparent contradiction between the RNase H studies of the groups of Bai and Marqusee. This example further highlights the utility of the charge-burial strategy.

Conclusions

The buried aliphatic-to-charged substitution strategy is a general method for trapping and characterizing folding intermediates under equilibrium conditions. In the case of Ub, the intermediate's stability relative to a native-like species is controllable, with acidification generating a well-dispersed NMR spectrum under conditions where HX is conveniently quenched. As a result, NSHX measurements can be performed on intermediates with poor NMR spectra. The strategy is particularly powerful in combination with other methods such as NSHX and RD. Although we applied the strategy with foreknowledge of Ub's folding behavior, we envision that even without this foreknowledge, a series of charged substitutions throughout a protein can help identify foldons and their relationship to folding pathways in other proteins.

Materials and Methods

UbL50E and other mutants, based on a pWTUb (with F45W and H68N), were prepared according to Krantz *et al.*²⁶ Proteins in inclusion bodies were solubilized in 8 M urea buffer before HPLC purification.

Equilibrium and kinetic measurements

CD experiments were performed with a Jasco 715 spectropolarimeter with a path length of 1 cm in 15 mM sodium phosphate and 225 mM sodium chloride. Rapid-mixing fluorescence experiments used a Biologic SFM-400 stopped-flow apparatus connected via a fiber optic cable to a PTI A101 arc lamp. Fluorescence spectroscopy used excitation and emission wavelengths of 280–290 nm and 300–400 nm, respectively.

NMR spectroscopy and HX chemistry

All the NMR experiments were run on a 600-MHz Varian Unity Inova spectrometer equipped with a cryoprobe. HNCACB and CBCACONH experiments were run on ¹⁵N,¹³C-labeled UbL50E at 0.5 mM concentration in 15 mM sodium phosphate, 225 mM sodium chloride, 50 mM Glu, 50 mM Arg, pH 5.3, and 5% D₂O to obtain

resonance assignments. The two-dimensional ¹⁵N HSQC spectra were taken with ¹⁵N-labeled UbL50E at 0.1 mM protein concentration in the same buffer with added 10 mM Glu and Arg. The protein at 0.5 mM, pH 6.6, was diluted 1:4 to a final condition of 15 mM sodium phosphate, 225 mM sodium chloride, 10 mM Glu, 10 mM Arg, 80% D₂O, and pD_{read} 7.5 to initiate HX. Exchange was quenched by the addition of 3.5 μL of 6 M DCl to 500 μL of protein solution, which dropped pD_{read} to 2.9.

The HX rate at the quenching condition (*k_q*) was also measured for a fully protonated sample and fit with the following equation in order to correct for back-exchange during the NMR measurement:

$$V_T = 0.8 \cdot V_0 \cdot [(1 - \exp(-k_q \cdot L)) / (k_q \cdot L)] \cdot \exp(-k_q \cdot T_o) + 0.2 \cdot V_0,$$

where *T_o* is the HSQC starting time, *L* is the length of the HSQC measurement, *V_T* is the NMR peak volume at *T_o*, and *k_q* and *V₀* are fitting parameters. The HX results were fit with the equation:

$$V_t = 0.8 \cdot V_0 \cdot [(1 - \exp(-k_q \cdot L)) / (k_q \cdot L)] \cdot \exp(-k_{HX} \cdot t) + 0.2 \cdot V_0,$$

where *t* is the length of the labeling period at pD_{read} 7.5, *V_t* is the peak volume at time *t*, and *k_{HX}* and *V₀* are fitting parameters.

Small-angle X-ray scattering

Data were collected at the BioCAT beamline at the Advanced Photon Source, Argonne National Laboratory. Samples were flowed through a 1.5-mm capillary at a rate of 2 μL/s, and 15 exposures of ~1-s duration were collected at room temperature. The exposures were averaged, buffer blanks were subtracted, and Guinier analysis was performed using IGOR Pro.

Acknowledgements

This work was supported by grants from the National Institutes of Health. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-ENG-38. BioCAT is a National Institutes of Health-supported research center (RR-08630). The content of this article is solely the responsibility of the authors and does not necessarily reflect the official views of the National Center for Research Resources or the National Institutes of Health. We thank Josh Kurutz for assistance in NMR measurements; L. Guo for help with the SAXS experiments; and B. Garcia-Moreno, N. Kallenbach, S. W. Englander, Y. Bai, A. Robertson, F. Massi, and members of the Sosnick laboratory for helpful discussions and comments on the manuscript.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2010.01.056](https://doi.org/10.1016/j.jmb.2010.01.056)

References

- Pan, Y. & Briggs, M. S. (1992). Hydrogen exchange in native and alcohol forms of ubiquitin. *Biochemistry*, **31**, 11405–11412.
- Jourdan, M. & Searle, M. S. (2001). Insights into the stability of native and partially folded states of ubiquitin: effects of cosolvents and denaturants on the thermodynamics of protein folding. *Biochemistry*, **40**, 10317–10325.
- Vallee-Belisle, A. & Michnick, S. W. (2007). Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol.* **374**, 791–805.
- Rea, A. M., Simpson, E. R., Crespo, M. D. & Searle, M. S. (2008). Helix mutations stabilize a late productive intermediate on the folding pathway of ubiquitin. *Biochemistry*, **47**, 8225–8236.
- Eliezer, D., Yao, J., Dyson, H. J. & Wright, P. E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **5**, 148–155.
- Whittaker, S. B., Spence, G. R., Gunter Grossmann, J., Radford, S. E. & Moore, G. R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol.* **366**, 1001–1015.
- Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, **430**, 586–590.
- Connell, K. B., Horner, G. A. & Marqusee, S. (2009). A single mutation at residue 25 populates the folding intermediate of *E. coli* RNase H and reveals a highly dynamic partially folded ensemble. *J. Mol. Biol.* **391**, 461–470.
- Neudecker, P., Lundstrom, P. & Kay, L. E. (2009). Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.* **96**, 2045–2054.
- Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science*, **269**, 192–197.
- Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. (1997). Hydrogen exchange: the modern legacy of Linderström-Lang. *Protein Sci.* **6**, 1101–1109.
- Silverman, J. A. & Harbury, P. B. (2002). Rapid mapping of protein structure, interactions, and ligand binding by misincorporation proton-alkyl exchange. *J. Biol. Chem.* **277**, 30968–30975.
- Sridevi, K. & Udgaonkar, J. B. (2002). Unfolding rates of barstar determined in native and low denaturant conditions indicate the presence of intermediates. *Biochemistry*, **41**, 1568–1578.
- Park, C. & Marqusee, S. (2004). Probing the high energy states in proteins by proteolysis. *J. Mol. Biol.* **343**, 1467–1476.
- Wang, L. & Kallenbach, N. R. (1998). Proteolysis as a measure of the free energy difference between cytochrome *c* and its derivatives. *Protein Sci.* **7**, 2460–2464.
- Korzhnev, D. M. & Kay, L. E. (2008). Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41**, 442–451.
- Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., 3rd & Raleigh, D. P. (2006). Multistate folding of the villin headpiece domain. *J. Mol. Biol.* **355**, 1066–1077.
- Grey, M. J., Tang, Y., Alexov, E., McKnight, C. J., Raleigh, D. P. & Palmer, A. G., 3rd (2006). Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: contribution of His41 to the pH-dependent stability of the N-terminal subdomain. *J. Mol. Biol.* **355**, 1078–1094.
- Bai, Y., Feng, H. & Zhou, Z. (2007). Population and structure determination of hidden folding intermediates by native-state hydrogen exchange-directed protein engineering and nuclear magnetic resonance. *Methods Mol. Biol.* **350**, 69–81.
- Kato, H., Feng, H. & Bai, Y. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J. Mol. Biol.* **365**, 870–880.
- Feng, H., Vu, N. D. & Bai, Y. (2004). Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome *b562*: native fold with non-native hydrophobic interactions. *J. Mol. Biol.* **343**, 1477–1485.
- Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, A. S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci.* **8**, 2251–2257.
- Sidhu, N. S. & Robertson, A. D. (2004). Exploring the conformational manifold of ubiquitin by native state hydrogen exchange. Master degree thesis, University of Iowa.
- Harper, S. M., Christie, J. M. & Gardner, K. H. (2004). Disruption of the LOV- α helix interaction activates phototropin kinase activity. *Biochemistry*, **43**, 16184–16192.
- Isom, D. G., Cannon, B. R., Castaneda, C. A., Robinson, A. & Garcia-Moreno, B. (2008). High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc. Natl Acad. Sci. USA*, **105**, 17784–17788.
- Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **337**, 463–475.
- Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using Psi analysis. *Chem. Rev.* **106**, 1862–1876.
- Baxa, M. C., Freed, K. F. & Sosnick, T. R. (2009). Psi-constrained simulations of protein folding transition states: implications for calculating. *J. Mol. Biol.* **386**, 920–928.
- Krantz, B. A., Moran, L. B., Kentsis, A. & Sosnick, T. R. (2000). D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nat. Struct. Biol.* **7**, 62–71.
- Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat. Struct. Biol.* **9**, 458–463.
- Karp, D. A., Gittis, A. G., Stahley, M. R., Fitch, C. A., Stites, W. E. & Garcia-Moreno, E. B. (2007). High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* **92**, 2041–2053.
- Englander, S. W. & Kallenbach, N. R. (1984). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **16**, 521–655.
- Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins*, **17**, 75–86.
- Connelly, G. P., Bai, Y., Jeng, M.-F., Mayne, L. & Englander, S. W. (1993). Isotope effects in peptide group hydrogen exchange. *Proteins*, **17**, 87–92.
- Zhou, Z., Feng, H., Ghirlando, R. & Bai, Y. (2008). The high-resolution NMR structure of the early folding intermediate of the *Thermus thermophilus* ribonuclease H. *J. Mol. Biol.* **384**, 531–539.