

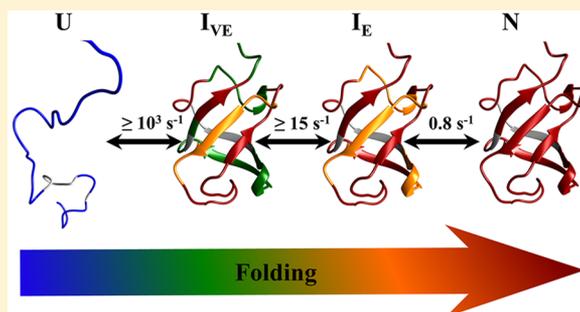
Stepwise Assembly of β -Sheet Structure during the Folding of an SH3 Domain Revealed by a Pulsed Hydrogen Exchange Mass Spectrometry Study

Nilesh Aghera[†] and Jayant B. Udgaonkar^{*†}

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India

Supporting Information

ABSTRACT: Dissecting temporally the sequence of secondary structural changes, and determining how these specific changes modulate conformational heterogeneity, remain major goals of protein folding studies. In this study, the folding of the SH3 domain of PI3 kinase has been characterized using pulsed hydrogen exchange mass spectrometry (HX-MS). The folding could be described as a four-state process, $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$, where I_{VE} and I_E are structurally heterogeneous intermediate ensembles. Compared to U , early intermediate I_{VE} has a marginally increased level of protection against HX of amides along the entire length of the polypeptide. Sequential assembly into β -sheet structure has been resolved temporally. Three of the five β -strands acquire nativelike structure before the rate-limiting step. β -Strands 2 and 5 acquire nativelike structure in I_{VE} , while β -strand 4 does so in I_E . β -Strand 1 acquires nativelike structure only during the last step of the folding process. Hence, the HX-MS study has resolved the order of assembly of the β -strands for the formation of the two β -sheets, which previous studies utilizing Φ -value analysis of several different SH3 domains had been unable to accomplish. Moreover, it is shown that structural heterogeneity decreases in a stepwise manner during the three stages of folding.



The kinetics and thermodynamics of protein folding reactions have been studied extensively, but the temporal sequence of the specific structural transitions that lead to the formation of the native state is not fully understood. An understanding of the mechanism of protein folding requires not only a delineation of the chronological order of the structural arrangements that occur but also a determination of how structural heterogeneity is reduced at each structural transition. Some methodologies used to study protein folding can report on structural heterogeneity,^{1–5} but they do not provide detailed structural information. In contrast, NMR-assisted studies^{6–9} can describe structural transitions at the resolution of individual amino acid residues, but they reveal little about how structural heterogeneity is decreased during folding. A quantitative determination of the heterogeneity associated with the structures populated at progressive stages of the folding process is needed to understand the role of heterogeneity in modulating the free energy landscape of protein folding. How the decrease in heterogeneity (entropy change) is linked to intramolecular contact formation (enthalpy change) remains to be ascertained. To obtain useful insight into this unresolved question, it is important that a probe with the requisite structural and temporal resolution, which can also measure structural heterogeneity, be used to study the folding of a suitable model protein.

The hydrogen exchange (HX) methodology used in conjunction with NMR to characterize protein folding

pathways not only legitimized the existence of folding^{8,9} and unfolding¹⁰ intermediates but also directly provided residue-level information about their structures. Application of the pulsed HX-NMR methodology not only revealed the stepwise assembly and stabilization of secondary structures during folding^{8,9,11–13} but also enabled determination of local stabilities of intermediates.^{11,14,15} Such studies have confirmed that molten globules are populated on direct folding pathways,¹⁶ have unveiled the utilization of competing folding^{8,17,18} and unfolding^{10,19} pathways and of unproductive dead-end pathways,^{15,20} and have also revealed heterogeneity in the early folding intermediates.³ Native-state HX-NMR studies have also been of great utility. They have revealed the fluctuations that occur in native functional protein structure^{14,21–23} and have dissected the structures of high-energy intermediates populated after the rate-limiting step of folding.^{21,22,24,25} When used in conjunction with mass spectrometry (MS), the HX methodology can distinguish between multiple conformations present together and can monitor the temporal variation in their relative populations.^{25–30} Application of the HX-MS methodology has provided some of the best proof that folding intermediates are populated on direct folding pathways,^{26,31} has revealed how the cooperativity of protein folding and unfolding

Received: April 24, 2017

Revised: June 14, 2017

Published: June 30, 2017



reactions can be modulated by stabilization of the folded, unfolded, and transition states,²⁹ and has revealed that unfolding can occur via multiple folding routes.³² The pulsed HX-MS methodology has great potential to characterize the structures and heterogeneity of intermediates preceding the TS on protein folding pathways.

SH3 domains have been widely exploited as model proteins for the study of protein folding. Although they originally appeared to be classic “two-state” folders,^{6,33} more careful and detailed studies have revealed the complexity of their folding mechanisms.^{30,34–36} Hydrogen exchange and NMR studies have identified partially structured intermediates,^{7,30,37–41} including some with non-native structures^{42–44} on the unfolding pathways of SH3 domains. Φ -Value analysis of the transition state (TS) on the folding pathway has revealed that the hydrophobic core composition^{45,46} as well as non-native hydrophobic contacts^{47,48} can modulate folding rates, that native-state topology plays an important role in folding,⁴⁹ and that the TS is largely hydrated⁵⁰ and appears to be highly polarized. A computational study guided by Φ -value analysis has suggested that β -strands 2–4 are structured in the TS, whereas β strands 1 and 5 are unstructured.⁵¹ The temporal order in which the β -strands form and assemble into β -sheets is, however, not yet known, because structural intermediates preceding the rate-limiting step have not been characterized structurally. Other spectroscopic measurements have also revealed that the transient formation of non-native structure might modulate subsequent folding,⁵² but how this might occur is unknown. Clearly, SH3 domains still have great utility as model proteins for answering important questions concerning fundamental mechanistic issues in protein folding.

The folding of the PI3K SH3 domain has been shown to commence by a gradual, nonspecific collapse of the polypeptide chain that is complete within 150 μ s.³⁴ The initially collapsed ensemble (U_C) is devoid of specific structure, has unfolded-like fluorescence properties, and is long-lived: specific structure formation accompanied by further chain compaction occurs on the 100–1000 ms time scale. It appears that when folding conditions are made more stabilizing by addition of salt, U_C converts to an early intermediate ensemble with specific structure within the first millisecond of folding.³⁵ A second late intermediate, first identified in both equilibrium and kinetic unfolding studies,³⁶ forms ephemerally after the rate-limiting step of folding. This late intermediate appears to possess natively like secondary structure, has unfolded-like fluorescence properties and substantial solvent-exposed surface area, and is partially stabilized by non-native interactions.⁵³ It appears that hydrophobic core consolidation accompanied by water expulsion occurs only when the late intermediate folds to N, as also suggested by computational studies.^{51,54} The observation that both U_C and the late intermediate have unfolded-like fluorescence suggests the possibility that other folding intermediates with unfolded-like fluorescence properties might also be populated on the folding pathway. There has, however, been little direct characterization of the structural events that occur progressively on the folding pathways of the PI3K SH3 domain as well as of other SH3 domains.

In this study, the folding of the PI3K SH3 domain has been characterized by pulsed HX labeling followed by peptic digestion coupled to mass spectrometric analysis. Four discrete conformations, the unfolded (U) state, the native (N) state, and two partially structured conformations, I_{VE} and I_E , are populated and have been individually tracked during the course

of the folding reaction. The mechanism of folding can be quantitatively described by a four-state, $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$ model, and conformational heterogeneity is shown to decrease in such a stepwise manner with an increase in the level of natively like structure. Structure (protective against HX) formation in different sequence segments of the protein has been delineated for each of the three folding steps, including those preceding the rate-limiting step of folding. The temporal order of β -strand formation into β -sheets in the SH3 domain has been resolved, thereby providing general insights into the nature of β -sheet assembly.

MATERIALS AND METHODS

Guanidine hydrochloride (GdnHCl) was purchased from USB and was of the highest purity grade. D_2O with 99.9% D content was purchased from BRIT (Board of Radiation and Isotope Technology, Navi Mumbai, India). GdnHCl was deuterated by three cycles of HX in D_2O followed by freeze-drying. All other reagents used were from Sigma and were of the highest purity grade. All the experiments were performed in 50 mM phosphate buffer at pH 7.2 or pD 7.2 (corrected for the isotope effect).

Refolding Kinetics and Pulse Labeling. The pulse labeling experiments were performed using the SFM-4/Q quench-flow machine from Biologic at 25 °C. The protein was unfolded for at least 12 h in unfolding buffer (prepared in D_2O) containing 50 mM phosphate buffer and 2.4 M GdnDCl. Refolding of the protein was initiated at a final concentration of 10–20 μ M in the presence of 0.3 M GdnHCl by rapidly mixing unfolded protein in a 1:7 ratio with native buffer containing 50 mM phosphate buffer (prepared in D_2O). It is important to note that earlier fluorescence monitored folding studies had shown that the folding kinetics were independent of protein concentration in the range of 10–200 μ M.^{35,36} It should also be noted that in 0.3 M GdnHCl, 99.5% of the protein is folded at equilibrium.³⁶ After different time periods of refolding (from 200 ms to 20 s), the protein was subjected to a HX labeling pulse by 10-fold dilution into 50 mM glycine buffer prepared in H_2O , at pH 9 or 10. The HX reaction was quenched following the 12 ms pulse of exchange by reducing the pH to 2.6 using a glycine-HCl buffer (prepared in H_2O). All the pulse labeling experiments were performed using an interrupted-flow technique. The quenched sample was diluted 2-fold into ice-cold buffer at pH 2.6 and immediately analyzed using an UPLC-MS system with the HX module from Waters. The mass spectra were acquired in ESI-MS mode using a Synapt G2 HD mass spectrometer from Waters. The populations of the +11 charge state and +7 charge state were used for all the analysis performed on the intact protein.

In a quench-flow experiment with pulsed HX labeling, a small fraction (7–8%) of labeled protein remains in the flow system of the mixing module and is carried over into the next pulse labeling experiment. To determine the extent of carryover, immediately after the pulse labeling experiment with protein, a dummy pulse labeling experiment in which the unfolded protein solution was replaced with buffer was performed. The mass profile of the dummy run revealed the extent of carryover of labeled protein.

In a competition refolding experiment aimed to delineate the early events of structure formation, refolding was initiated by diluting fully deuterated unfolded protein into native buffer prepared in H_2O . The dilution triggered the refolding and exchange processes simultaneously. The refolding and exchange

were performed at pH 8 (50 mM Tris buffer), pH 9 and 10 (50 mM glycine buffer) for 12 ms, and HX was subsequently quenched by reducing the pH to 2.6.

Data Analysis. Raw MS data were processed and smoothed using the MassLynx software (Waters). All mass profiles were determined after applying the 12 ms HX labeling pulses at different times of folding and corrected by subtracting out the carryover spectrum (see above). The corrected and smoothed mass profiles were normalized with respect to the area under the curve. MS data were analyzed by discrete fitting as well as by global fitting (Supporting Information), assuming that the mass distributions of all the species populated during refolding could be described as Gaussian distributions. Analysis of pulsed HX-MS data by fitting mass profiles to the sum of Gaussian distributions, each arising from HX into a distinct protein conformation, is an effective way to delineate the mechanism of exchange and structural transitions.⁵⁵ The mass profiles obtained from the HX studies for the intact protein and for the different segments of the protein were analyzed as the sum of two, three, or four Gaussian distributions using eq 1:

$$y = \sum_{i=1}^n \left(\frac{a_i}{c_i \sqrt{\pi/2}} \right) \times e^{-2(x-b_i/c_i)^2} \quad (1)$$

where a represents the area under each Gaussian distribution and b and c represent the values of the peak m/z position and the width, respectively, of each of n Gaussian distributions. $n = 2$ for a two-state analysis; $n = 3$ for a three-state analysis, and $n = 4$ for a four-state analysis of the mass profile data. For the four-state analysis, the four Gaussian distributions represented U (unfolded state), I_{VE} (first intermediate or unfolded-like intermediate), I_E (second intermediate or nativelylike intermediate), and N (native state). In the analysis using constrained fits, the peak m/z positions and widths of the Gaussian distributions were constrained such that across time, the variations of m/z and width were ± 0.1 and $\pm 10\%$, respectively. The fractional area under each Gaussian distribution corresponded to the fractional population of the species giving rise to the distribution. Fractional population curves were analyzed using a single-exponential equation to obtain the kinetic parameters.

Analysis of the Kinetics of Protein Segments. The mass profiles of the peptides were analyzed using two-, three-, or four-state models, i.e., as the sum of either two, three, or four Gaussian distributions. During the analysis of each sequence segment, the m/z and width values for different Gaussian distributions were held constant (m/z , ± 0.1 ; width, $\pm 10\%$) during the different durations of folding. The area under each Gaussian distribution corresponded to the fractional population of the respective species. The exchange kinetics observed for all the analyzed protein segments were consistent with the exchange kinetics observed for the intact protein, irrespective of the number of Gaussian distributions used for analysis. Sequence segments whose labeling kinetics during refolding of the intact protein could be described as a four-state process were found to have kinetics identical to that of the intact protein. Many sequence segments were found to follow three-state labeling kinetics during the refolding of the intact protein, where one of the three fractional populations corresponded to the sum of two fractional populations seen for the intact protein (i.e., either $I_{VE} + I_E$ or $I_E + N$). Some protein segments were found to follow labeling kinetics that could be described by a two-state process, where the fractional population of the

species with a higher level of protection corresponded to the sum of the fractional populations of the two intermediates and the native state (i.e., $I_{VE} + I_E + N$).

RESULTS

HX studies are performed either by subjecting native protein to continuous HX labeling (native-state HX) or by subjecting folding protein to pulses of HX labeling (pulse labeling) at different times during the folding process. Of these two alternative methodologies, native-state HX studies are easier to perform but more difficult to interpret. The interpretation of native-state HX labeling requires that it be known whether HX occurs in the EX1 regime or in the EX2 regime,⁵⁶ and it is often not easy to distinguish between these two mechanisms. In contrast, in pulse labeling studies, each discrete conformation becomes labeled to the same fractional extent at every time of folding, independent of the mechanism of labeling during the short labeling pulse. Because the duration and pH of the labeling pulse are not varied during the course of the folding reaction, each discrete conformation can be easily identified in a pulse HX-MS study, because it will have the same unique HX-labeled mass distribution at every time of folding, as reflected by its m/z and width. Only the extent to which each conformation is populated will vary with the time of application of the labeling pulse, which will be reflected in the area under its mass distribution.

In this study, the refolding kinetics of the PI3K SH3 domain (Figure 1) was investigated using the HX-MS pulse labeling methodology to obtain insights into the structural transitions occurring during the folding of the protein. Refolding of the deuterated protein was initiated in a buffer containing D_2O at pD 7.2, and the protein was labeled at different times by applying a 12 ms HX labeling pulse of H_2O at pH 9 or 10. Labeling was performed at pH 9 and 10 to determine the strength of interactions in the structures of the folding intermediates. Different conformations of the protein were labeled to different extents by the HX labeling pulse, depending on the extent of protective structure present at the time of the application of the pulse.

In pulse labeling HX-MS studies of protein folding, the mass distributions corresponding to the different conformations populated during folding are sometimes well separated, but they can overlap, leading to complex mass profiles of overlapping mass distributions. To simplify the deconvolution of the complex mass profile into individual mass distributions each corresponding to a distinct conformation, it is necessary to first determine the mass distributions of the N and U states populated at the beginning and end of the folding reactions. Once this is done, it becomes easier to identify the number of intermediate conformations populated transiently at any time during the course of folding.

Labeling of N and U. When the native deuterated PI3K SH3 domain was subjected to a 12 ms HX labeling pulse in 0.3 M GdnDCI at pH 9, the mass profile showed a single peak with a maximum at m/z 848.9 ± 0.1 for the +11 charge state (Figure 2a and Figure S1a). Deuterated native protein that had not been subjected to the 12 ms HX labeling pulse, but otherwise had been subjected to the same workup conditions, showed a single peak centered at m/z 849.4. This indicated that the HX labeling pulse at pH 9 labeled approximately five amide sites in the native protein. After a 12 ms HX labeling pulse at pH 10 to native PI3K SH3, the single peak in the mass profile was centered at m/z 847.9 (Figure S1b). Because the m/z for the

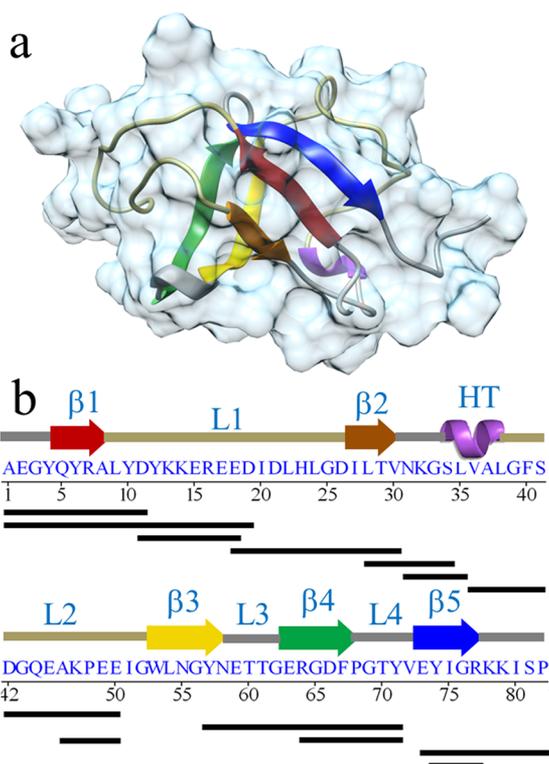


Figure 1. Structure of the PI3K SH3 domain and its peptide map. (a) The tertiary structure was drawn using the UCSF Chimera package and Protein Data Bank entry 1PNJ. The domain is an 82-residue protein consisting of two β -sheets, two helical turns, and long loops.⁵⁷ (b) Primary structure of the PI3K SH3 domain aligned with the secondary structure map. The horizontal black bars below the primary structure represent the peptides obtained upon pepsin digestion at pH 2.6 and 4 °C (Supporting Information), which cover 91% of the PI3K SH3 sequence. These peptide fragments were used for analysis of the structure formed in the corresponding sequence segments of the intact protein. β , L, and HT denote the β -strand, loop, and helical turn, respectively.

+11 charge state of fully protonated protein was 844.3, these results indicated that of 56 protected amide sites, 51 amide sites in N remained protected against HX by the labeling pulse at pH

9, while 40 amide sites remained protected against HX by the labeling pulse at pH 10.

When the unfolded deuterated PI3KSH3 domain in 2.4 M GdnHCl was subjected to a 12 ms HX labeling pulse at pH 9, the mass profile showed a single peak centered at m/z 844.8 \pm 0.1 (Figure 2b). HX-labeled U therefore had a mass marginally (\sim 6 Da) higher than that of the fully protonated PI3K SH3 domain, which was expected because the HX reaction was performed in buffer containing 90% H and 10% D.

For the mass distributions of both N and U, it was found by multiple repetition of the experiments shown in Figure 2 that the variations in m/z and width were \pm 0.1 and \pm 10%, respectively.

To make sure that the pulse labeling methodology correctly determines the fractions of protein present as U and N when both forms are present together, deuterated protein incubated in different concentrations of GdnDCl was subjected to the 12 ms labeling pulse at pH 9. It is known that the time constants of interconversion between N and U at GdnHCl concentrations straddling the unfolding transition zone are more than 100-fold longer than the 12 ms duration of the labeling pulse.³⁶ Hence, only protein molecules that were present as U at the time of application of the 12 ms labeling pulse were expected to become labeled. Figure 2c shows that the fraction of protein that was unfolded at each of nine different denaturant concentrations agreed well with the value obtained from a fluorescence-monitored equilibrium unfolding curve determined for deuterated protein in deuterated buffer. At each GdnDCl concentration, the mass spectrum showed only two well-resolved peaks (data not shown) corresponding to N (at m/z 848.9) and U (at m/z 844.8). It should be noted that although the protein samples at different GdnHCl concentrations were pulse-labeled consecutively in the quench-flow machine, the \sim 7% carryover (Figure S2) from one mixing event to the next did not affect the estimation of U (Figure 2c): it was found that it was not necessary to subtract the carryover mass spectrum (Figure S2) from the mass spectrum of each labeled sample. Nevertheless, for all mass profiles obtained after the application of 12 ms HX labeling pulses at different times of folding, the \sim 7% carryover mass spectrum was subtracted out, to reduce the possibility that conformations labeled to only small extents might be wrongly estimated.

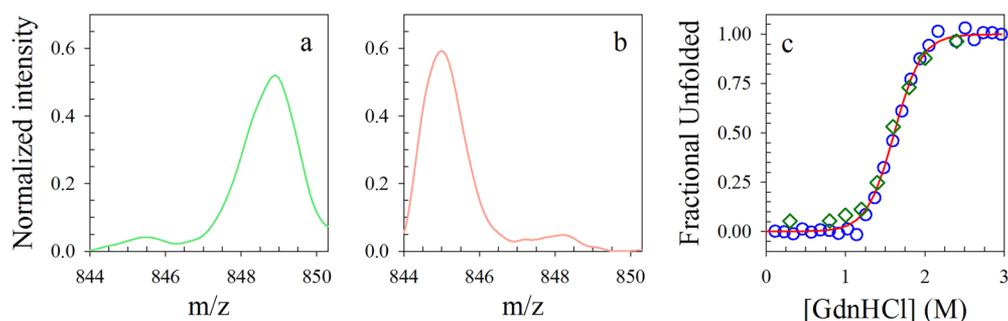


Figure 2. HX pulse labeling of N and U. Panel a shows the mass profile of native protein in 0.3 M GdnHCl that had been subjected to a 12 ms HX labeling pulse at pH 9. Approximately 7% of the protein in the mass profile is seen to be completely labeled, with its mass distribution centered at m/z 844.8. This 7% labeled protein represents carryover protein from the previous labeling reaction performed in the quench-flow machine, as shown in control experiments (see Materials and Methods). Panel b shows the mass profile of unfolded protein in 2.4 M GdnDCl that had been subjected to a 12 ms HX labeling pulse at pH 9. Panel c shows the values of the fraction of protein that is unfolded, which was determined by application of the 12 ms labeling pulse to protein that had been equilibrated at nine different GdnHCl concentrations (green diamonds), overlaid on the fractional change in the magnitude of the fluorescence signal at 310 nm upon excitation at 280 nm (blue circles). The solid red line is a fit through the data using a two-state model.

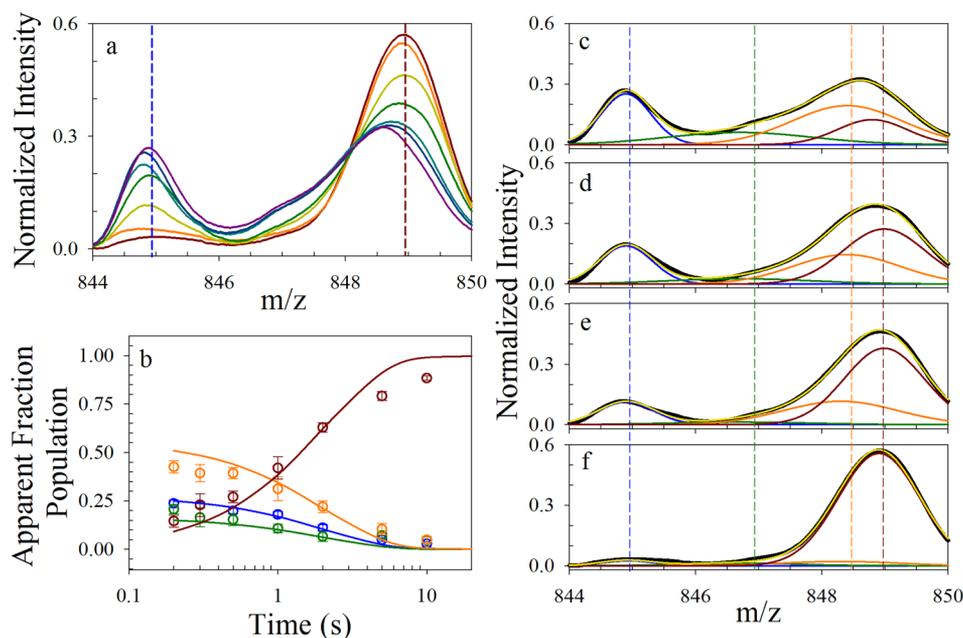


Figure 3. Refolding kinetics of the PI3K SH3 domain in 0.3 M GdnCl (pD 7.2) monitored using a HX labeling pulse at pH 9. (a) Mass spectra (+11 charge state) obtained upon subjecting deuterated refolding protein to a 12 ms HX labeling pulse at refolding times of 0.2, 0.3, 0.5, 1, 2, 5, and 10 s (from bottom to top, respectively, in the right-most peak). The vertical dashed blue and dark red lines represent the centers of the mass distributions of the U and N states, respectively (see Figure 2). (b) Global fitting of the kinetic data obtained using HX labeling pulses at pH 9 to a four-state folding model ($U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$). The lines (same color code as for the symbols) represent the kinetics of the changes in the different populations of molecules, obtained from the global fit (Supporting Information). The values obtained for the kinetic parameters from global fitting are listed in Table 1. Error bars in the figure represent the standard errors from two or more experiments. Representative four-state fits to the mass spectra obtained at refolding times of 0.2, 1, 2, and 10 s are shown in panels c–f, respectively. The black line represents the experimentally determined mass profile data, which were fit to the sum of four Gaussian mass distributions (solid yellow line) according to a four-state folding model. The blue and dark red distributions represent the populations of the unfolded (least protected) and native (most protected) states, respectively. The vertical dashed lines mark the position of the m/z maxima of U, I_{VE} , I_E , and N. The dark green and orange distributions represent the populations of I_{VE} and I_E , respectively. All mass spectra (experimental data) were normalized using the areas under their mass profiles. For analysis, each labeled population was assumed to have a Gaussian mass distribution centered at a fixed m/z value and at a full width at half-maximum (fwhm), but whose area changed with time of refolding.

Refolding of the PI3K SH3 Domain Monitored by HX-MS. Figure 3a shows the mass profiles of the +11 charge state upon application of the 12 ms HX labeling pulse at pH 9 at different times of refolding in 0.3 M GdnHCl. At each time of refolding, the mass profile obtained was complex, with some molecules labeled at m/z 848.9 as was native protein (Figure 2a), some molecules labeled at m/z 844.8 as was unfolded protein (Figure 2b), and other protein molecules labeled to an intermediate extent. The intermediate extent of labeling suggested the presence of one or two intermediates, minimally with a structure intermediate between U and N. Just as in the case of U and N, which have unique conformations and hence unique HX-labeled mass distributions, any intermediate with a unique conformation whose structure similarly does not change with time of folding will show up in a mass distribution whose m/z and width will not change with time of folding.

To determine the number of intermediate conformations, the mass distributions of both the +11 and +7 charge state were fit to the sum of the N and U mass distributions, plus either zero, one, or two additional Gaussian mass distributions (Figures S3 and S4). For the mass distributions of the +7 charge state (Figure S4), the separation in m/z units between the N and U mass distributions was 1.5-fold larger than for the +11 charge state (Figure S3); hence, the mass distributions obtained for the +7 charge state are better suited for discriminating among two, three, and four Gaussian fits. In the constrained fits shown in Figure 3 and Figures S3 and S4, the position and width of each

Gaussian in the mass profile were not allowed to change with the time of refolding at which the labeling pulse was applied. The positions and widths of the mass distributions arising from the labeling of N and U were kept identical to those shown in Figure 2. Only the area under each individual Gaussian distribution, which represented the population of the conformation from which it originated, was allowed to change.

It is seen in Figures S3 and S4 that the data at all times of application of the labeling pulse could not be adequately described by the sum of only the U and N distributions, ruling out a two-state $U \leftrightarrow N$ model for refolding. When the data were fit to the sum of the N and U distributions plus one additional intermediate mass distribution whose center and width were not allowed to vary (three-Gaussian constrained fit), it was seen that the fits were not good, particularly for the mass distributions of the +7 charge state (Figure S4). The root-mean-square deviations (RMSDs) of the three-Gaussian constrained fits from the data were significantly worse than the RMSDs seen for the four-Gaussian constrained fits across all times of application of the labeling pulse. When, however, the mass profiles of the +7 charge state were fit to the sum of the N and U mass distributions plus one additional intermediate mass distribution whose center and width were allowed to vary (three-Gaussian free fit), the fits were better than the three-Gaussian constrained fits (Figure S4). The RMSD values for the three-Gaussian free fits were, however, not as good as the RMSD values obtained from four-Gaussian

constrained fits, in which the mass profiles of the +7 charge state were fit to the sum of the N and U distributions plus two additional intermediate mass distributions whose center and width were not allowed to vary (Figure S4).

Although the four-Gaussian constrained fit appeared to be better than the three-Gaussian free fit in terms of RMSD values (Figure S4), it was only marginally better, and it was not straightforward to rule out a three-state $U \leftrightarrow I \leftrightarrow N$ model in favor of a four-state $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$ model purely on the basis of the RMSD values of the three-Gaussian and four-Gaussian fits, respectively, demanded by these models. The three-state model could, however, be excluded by a simple experimental observation. Both three-Gaussian fits indicated that 30% of the protein molecules would have to form N in a burst phase (Figure S5). If this were true, then a 30% burst phase change in fluorescence arising from the burst phase formation of N would have been observed when the refolding of the protein was monitored by fluorescence under identical refolding conditions. No burst phase change in fluorescence was observed,³⁶ ruling out any three-state model for folding. It should also be noted that the three-Gaussian free fit is to a $U \leftrightarrow I \leftrightarrow N$ model in which I is not a discrete conformation but increases continuously in level of structure and hence level of protection with time of refolding. In other words, I would represent a continuum of intermediate forms. While such a scenario cannot be ruled out, the three-state $U \leftrightarrow I \leftrightarrow N$ model in which I undergoes a change in structure continuously with time of folding is certainly less simple than a four-state $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$ model in which I_{VE} and I_E are discrete conformations. In the study presented here, the simplest and minimal kinetic model that could account for the data (the four-Gaussian constrained fit) was the four-state model. It should be noted that in earlier HX pulse labeling studies of the folding of ribonuclease H, an apparently moving peak could be more simply described as the sum of two mass distributions.^{58,59}

It therefore appeared that the simplest description of the data was that four different conformational ensembles differing in the protection they afford against HX were populated during refolding. The fits to the mass profiles at the different times of folding indicated that two partially folded conformations were populated, which afforded protection that was intermediate between that afforded by U and N. These intermediates, I_{VE} and I_E , afforded protection to 25 and 41 more amide sites, respectively, than did the U form, but their Gaussian mass distributions were very broad, indicating that both intermediates were heterogeneous ensembles of molecules with varying levels of protection and, hence, structure. In comparison, N in 0.3 M GdnHCl had 47 more protected amide sites than U did. Figure 3b shows the observed fractional populations of U, I_{VE} , I_E , and N at different times of folding, determined from four Gaussian fits to the mass profiles. Figure S6 shows that if the mass profiles shown in Figure 3a had not been corrected for carryover (see above), as they were, it would have made little difference to the estimations of the four conformations.

The mass profiles obtained at different times of folding (Figure 3) were globally fit to a $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$ model for refolding. When the mass profiles were fit to the four-state model, the effect of the duration of the labeling pulse (12 ms) on the extent of labeling was taken into account, as described in the Supporting Information. Briefly, the analysis took into account the fact that if the pulse duration was long compared to the time constants of interconversion between any two states,

protein molecules would cycle between the two states during the duration of the pulse. If this cycling is not corrected for, then the fractional population of the less protected state would be overestimated at the expense of the more protected state. The values of the kinetic parameters defining the model, which were determined using the global fitting procedure taking into account the molecular cycling that occurs during the labeling pulse, are listed in Table 1. From Table 1, it is clear that the rate

Table 1. Parameters Obtained from Global Fitting of the Kinetic HX-MS Data for Folding in 0.3 M GdnHCl^a

kinetic parameters		properties of the species	
K_{N-U}	1450	W U	0.82
K_1	10.2	W I_{VE}	2.2
K_2	1.5	W I_E	1.8
K_3	94	W N	1.38
k_1	$\geq 1000 \text{ s}^{-1}$		
k_2	$\geq 15 \text{ s}^{-1}$	m/z U	844.9
k_3	0.82 s^{-1}	m/z I_{VE}	846.8
k_{-1}	$\geq 67 \text{ s}^{-1}$	m/z I_E	848.45
k_{-2}	$\geq 15 \text{ s}^{-1}$	m/z N	848.95
k_{-3}	0.009 s^{-1}		

^a K_{N-U} is the equilibrium constant between N and U, whose value was allowed to vary in a range of the experimentally obtained values.³⁶ All other kinetic parameters listed were obtained by global fitting of the data (Supporting Information) to the folding mechanism shown in Figure 7. The global fitting routine took into account labeling that occurred during the 12 ms labeling pulse because of molecular cycling (see Results). Prefix W represents the width of the Gaussian distribution, and prefix m/z represents the m/z value at the maximum of the Gaussian distribution. During the discrete analysis performed using these values, the m/z maximum and fwhm were allowed to change by only 0.1 and 10%, respectively.

constants for the $U \leftrightarrow I_{VE}$ transition are fast compared to the 12 ms duration of the labeling pulse, which results in an overestimation of the fractional population of U and an underestimation of the fractional population of I_{VE} (Figure 3b). The actual fractional populations of U, I_{VE} , I_E , and N at the different times of application of the labeling pulse are shown in Figure S7. A comparison of Figure 3b and Figure S7 indicates that molecular cycling affects the estimation of only U and I_{VE} , because of the values of the rate constants defining the transition.

Probing the Stabilities of I_{VE} and I_E by Varying the pH of the Labeling Pulse. The 12 ms HX labeling pulse will label all amide sites, whether in U, I_{VE} , I_E , or N, which are unstructured (unprotected) at the time of application of the pulse. Labeling would occur at the intrinsic HX rate, which is calculated for the PI3KSH3 domain to be 1500 s^{-1} at pH 9 and 15000 s^{-1} at pH 10. If an amide site is protected in I_{VE} , I_E , or N, then the fraction of protein molecules in which this site will become labeled will depend on whether exchange at that amide site follows an EX1 or EX2 mechanism.⁵⁶

To probe the stabilities of amide sites in I_{VE} and I_E , refolding was also monitored by the application of HX labeling pulses at pH 10 (Figure 4). Mass profiles obtained upon applying the HX labeling pulses at different times of refolding (200 ms to 10 s) could again be fitted to the sum of four-Gaussian mass distributions (Figure 4). From their centroid values, the lowest and highest Gaussian mass distributions could be shown to arise from the labeling of U and N molecules, respectively. The two intermediate mass distributions, whose widths were less

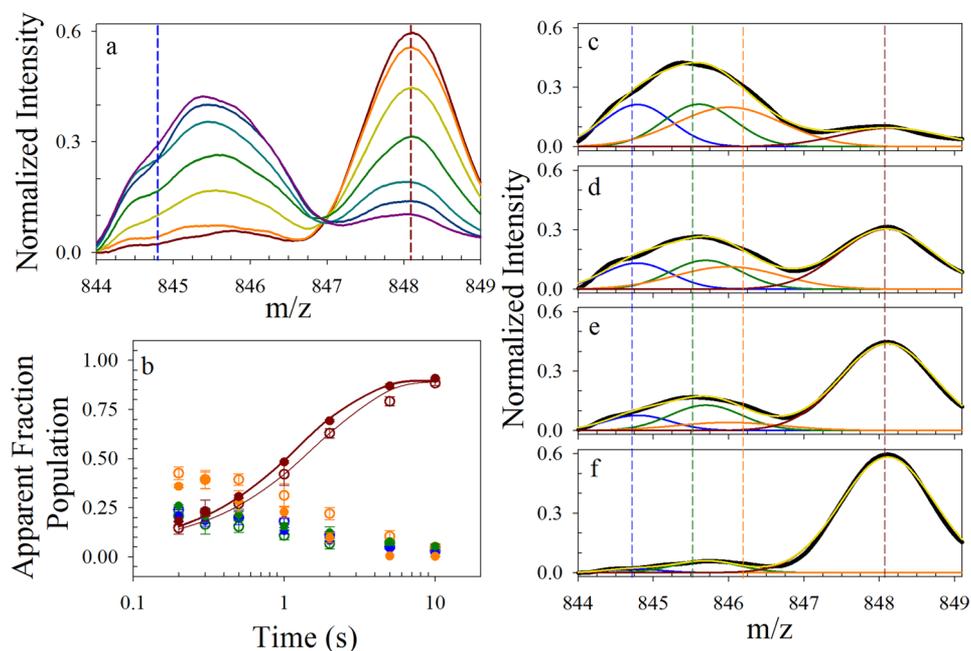


Figure 4. Kinetics of folding of the PI3K SH3 domain in 0.3 M GdnCl (pD 7.2) monitored using pulsed HX labeling at pH 10. (a) Mass spectra (+11 charge state) obtained upon subjecting deuterated refolding protein to a 12 ms HX labeling pulse at refolding times of 0.2, 0.3, 0.5, 1, 2, 5, and 10 s (from bottom to top, respectively, in the right-most peak). (b) Comparison of the time courses of the changes in populations of U, I_{VE} , I_E , and N, determined using pulsed HX labeling at pH 9 (empty circles) and pH 10 (filled circles) at the indicated times of refolding. The mass profiles obtained by pulsed HX labeling at pH 9 (as shown in Figure 3) and pH 10 were fitted to a sum of four-Gaussian mass distributions centered at fixed m/z values. The fractional populations of U (blue), I_{VE} (green), I_E (orange), and N (dark red) at different times of folding were determined by the fractional areas under the mass distributions. The thin and thick lines (same color code as for the symbols) through the data points represent single-exponential fits to data obtained using pH 9 and 10 HX labeling pulses, respectively. Error bars in the figure represent standard errors from two or more experiments. Representative four-state fits to the mass spectra obtained at refolding times of 0.2, 1, 2, and 10 s are shown in panels c–f, respectively. The color coding of the lines, as well as the data analysis, is as described in the legend of Figure 3. All mass spectra (experimental data) were normalized using the areas under their mass profiles.

than the intermediate mass distributions seen when the labeling pulse was at pH 9, could be shown to arise from I_{VE} and I_E . This was possible because these mass distributions disappeared, as did the mass distribution arising from the labeling of U, at the same rate as that of the intermediate mass distributions observed when labeling was performed using the pH 9 pulse (Figures 3b and 4b).

While HX into amide sites in native PI3K SH3 domain has been shown to occur by the EX1 mechanism even at pH 7 and 8,³⁰ it is conceivable, although unlikely, that HX into protected amide sites in I_{VE} and I_E could occur by the EX2 mechanism even at pH 9 and 10, if these sites opened and closed much faster in the intermediates than in N under these conditions. If the mechanism of HX into a protected amide site in I_{VE} or I_E were EX2, then a 12 ms pulse at pH 9 would completely label all amide sites that had a protection factor of <5 and would insignificantly label amides with a protection factor of >30 . A 12 ms pulse at pH 10 would completely label all amide sites that have a protection factor of <50 and insignificantly label amide sites with a protection factor of >300 . Hence, nine amide sites in I_{VE} would have protection factors of >300 , while 16 ($25 - 9 = 16$) amide sites would have protection factors between 30 and 300. In I_E , 17 amide sites would have protection factors of >300 and 24 ($41 - 17 = 24$) amide sites would have protection factors between 30 and 300.

The mechanism of HX into protected amide sites in the native PI3K SH3 domain has been shown to be EX1 even at pH 7.2³⁰ and is therefore certainly expected to be EX1 at pH 9 or 10, the pH value of the labeling pulse. If the mechanism of

HX into protected amide sites in I_{VE} and I_E is also EX1, then the fraction of I_{VE} or I_E molecules that become labeled at that site will depend on the rate constant of structure opening at that site in each of these species. While 25 amide sites in I_{VE} and 41 amide sites in I_E were sufficiently structured to not exchange out during the pH 9 pulse, only nine amide sites in I_{VE} and 17 in I_E were sufficiently structured to not exchange out during the pH 10 pulse. If the EX1 mechanism is indeed applicable, which is likely, then the observation that a larger number of amide sites are labeled by the pH 10 pulse than by the pH 9 pulse must mean that structure opening at these sites is significantly faster at pH 10 than at pH 9 in I_{VE} , I_E , and N. In other words, at least part of the protective structure in I_{VE} , I_E , and N would be more unstable kinetically at pH 10 than at pH 9. Indeed, it is known that native PI3KSH3 domain is less stable at pH 10 than at pH 9 (unpublished results).

I_{VE} Appears To Form within the Initial 10 ms Burst Phase. The data in Figure 3 indicated that I_{VE} and I_E had fully formed at 200 ms. To determine whether the intermediates had formed within the 10 ms dead time of mixing itself, the 12 ms HX labeling pulse was given directly to unfolded deuterated protein under refolding conditions (i.e., at refolding time zero). For a 12 ms pulse at pH 9, the mass profile showed two peaks, indicating that two populations of molecules had been labeled (Figure S1c). From the centroid of its m/z distribution, the lower-mass peak appeared to have originated from the labeling of U. The higher-mass distribution was broad and originated from the labeling of primarily one partially protected intermediate, which appeared to be I_{VE} . The fractional area

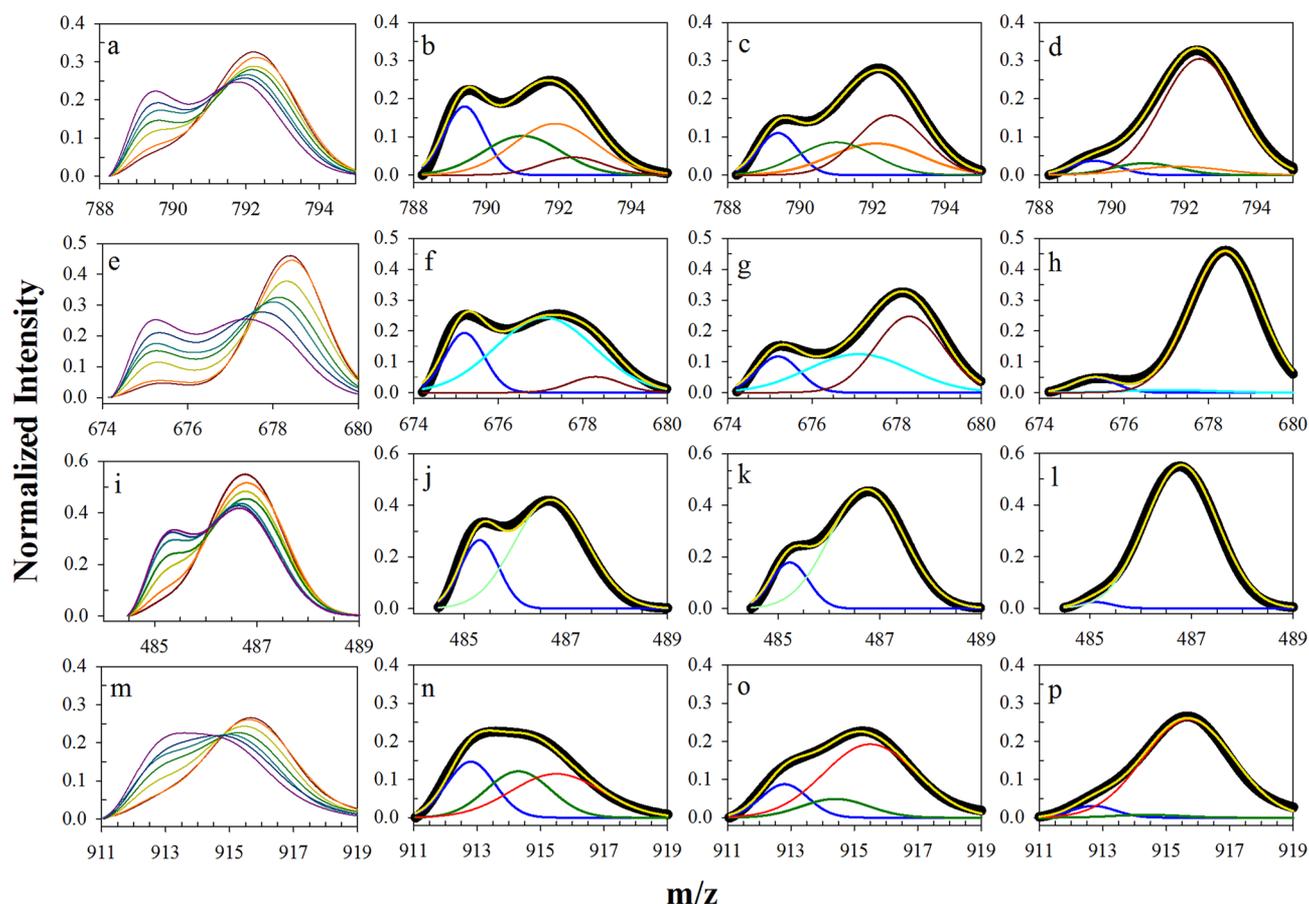


Figure 5. Kinetics of structure formation in different sequence segments of the PI3K SH3 domain, monitored using HX labeling. The 12 ms HX labeling pulses at pH 9 were applied at different times of refolding (0.2, 0.3, 0.5, 1, 2, 5, and 10 s, from top to bottom, respectively, in the left-most peak) in 0.3 M GdnCl (pD 7.2), followed by pepsin digestion of the labeled protein (see the Supporting Information) before mass spectrometry. Panels a, e, i, and m represent the mass spectra of the peptide fragments corresponding to sequence segments 36–50, 1–11, 18–30, and 64–71, respectively, of the intact protein. Panels b–d show the mass profiles of peptide fragments (sequence segments) 36–50 (+2 charge state) obtained after pulsed HX labeling at refolding times of 0.2, 1, and 10 s, respectively. The solid black line and the solid yellow line represent the experimentally obtained mass spectra and the fit of the data to the sum of four-Gaussian mass distributions, respectively. The dark red and blue mass distributions represent the populations having the sequence segment in the most (N-like) and least protected (U-like) conformations, respectively, whereas the dark green and orange mass distributions represent the populations corresponding to I_{VE} and I_E , respectively. Panels f–h show the mass profiles of peptide fragments (sequence segments) 1–11 (+2 charge state) obtained after pulsed HX labeling at refolding times of 0.2, 1, and 10 s, respectively. The solid black line and the solid yellow line represent the experimentally obtained mass spectra and the fit of the data to the sum of three-Gaussian mass distributions, respectively. The dark red and blue mass distributions represent the populations having the sequence segment in the most (N-like) and least protected (U-like) conformations, respectively, whereas the cyan mass distribution represents the populations having the sequence segment possessing an intermediate level of protection. Similarly, panels j–l and n–p represent sequence segments 18–30 and 64–71, respectively. The mass spectra (+3 charge state) for sequence segment 18–30 (panels e–h) fit well to the sum of two-Gaussian mass distributions, where the blue and light green distributions represent the populations having the sequence segment in its least (U-like) and most protected conformations, respectively. The mass spectra (+1 charge state) for peptide 64–71 (panels i–l) fit well to the sum of three-Gaussian mass distributions, where the blue, dark green, and red mass distributions represent populations having the sequence segment with the least, intermediate, and most protection, respectively.

under the U peak at refolding time zero was the same as at 200 ms, indicating that I_{VE} had formed within the 10 ms mixing dead time. In contrast, labeling by a pH 10 HX pulse led to a single peak in the mass profile (Figure S1d). The peak was centered at an m/z value higher than that of the U peak seen when the labeling pulse was applied at pH 9, and it also had a larger width. The peak could be fit to a sum of two Gaussian distributions, one arising from the labeling of U and the other from a species that has a level of protection higher than that of U but lower than that of I_{VE} (Figure S1d). The more protected species could be the collapsed state U_C , which is known to form on the submillisecond time scale.³⁴ It was, however, more likely to have been I_{VE} , showing an apparently smaller number of

protected amides due to competition between the refolding and exchange processes. It was also possible that I_{VE} and I_E both had less protective structure when they first formed (within the initial 10 ms) than after refolding for 200 ms (or longer). The possibility that this protected species represents an additional intermediate preceding I_{VE} and I_E cannot, however, be ruled out at present.

Refolding Profiles of Different Sequence Segments. It was important to determine which sequence segments were structured and, hence, protected from HX in I_{VE} as well as in I_E . To delineate the structural events during refolding, a HX labeling pulse at pH 9 was applied at varying times of refolding, and the protein (after HX was quenched) was fragmented by

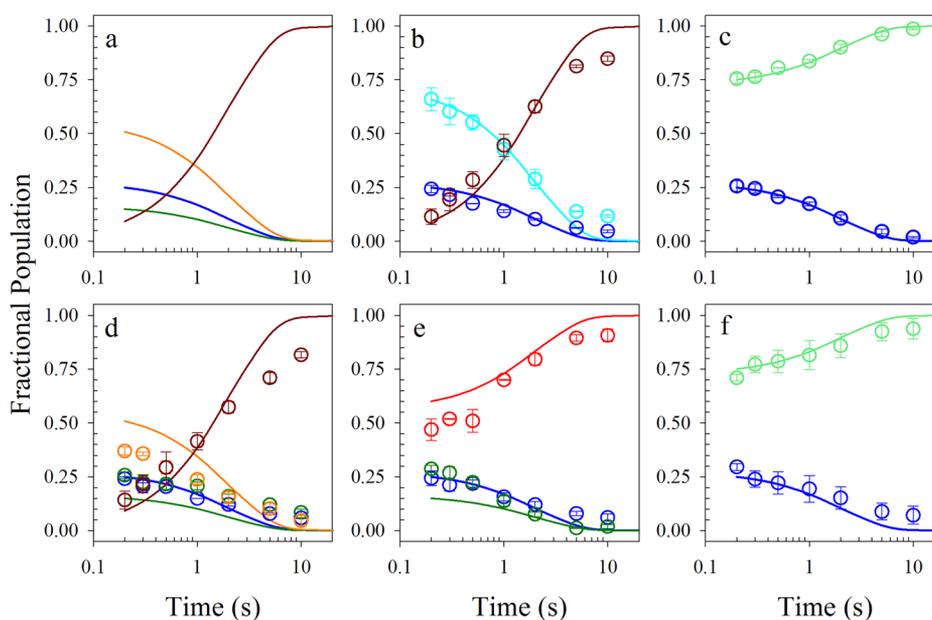


Figure 6. Kinetics of structure formation in different sequence segments of the PI3K SH3 domain during refolding in 0.3 M GdnDCl (pD 7.2). Panel a shows the kinetics of the changes in the populations of U (blue), I_{VE} (dark green), I_E (orange), and N (dark red) as determined by global analysis of the refolding data for the intact protein (Figure 3b) according to a four-state $U \leftrightarrow I_{VE} \leftrightarrow I_E \leftrightarrow N$ mechanism. Panel b shows the changes in the fractional areas under the three mass distributions seen for peptide (sequence segment) 1–11 with time of refolding. The fractional areas under the mass distributions represent the populations having sequence segment 1–11 in its least protected and most labeled (blue circles), partially protected and labeled (cyan circles), and fully protected and least labeled (dark red circles) conformations. The kinetics of disappearance of U (blue line) and I_{VE} and I_E (cyan line), respectively (as shown in panel a). The kinetics of appearance of the population having sequence segment 1–11 in its most structured conformation are described well by the kinetics of appearance of the N state (dark red line) (as shown in panel a). This sequence segment does not show any change in labeling (protection) during the transition from I_{VE} to I_E . Similarly, panel c shows that the kinetics of the changes in the populations of protein molecules having sequence segment 18–30 in its least protected (blue circles) and protected (light green circles) conformations (Figure 5) correspond to the kinetics of the changes in the populations of U (blue line) and I_{VE} , I_E , and N (light green line), respectively (as shown in panel a). This sequence segment attains N-like protection during the formation of I_{VE} . Similarly, panel d shows that the kinetics of the changes in the populations of protein molecules having sequence segment 36–50 in its least protected (blue circles), partially protected (dark green circles and orange circles), and protected (dark red circles) conformations (Figure S10) correspond to the kinetics of the changes in the populations of U (blue line), I_{VE} (dark green line), I_E (orange line), and N (dark red line), respectively (as shown in panel a). Similarly, panel e shows that the kinetics of the changes in the populations of protein molecules having sequence segment 64–71 in its least protected (blue circles), partially protected (dark green circles), and protected (red circles) conformations (Figure 5) correspond to the kinetics of the changes in the populations of U (blue line), I_{VE} (dark green line), and I_E and N (red line), respectively (as shown in panel a). This sequence segment attains N-state-like protection during the formation of I_E . Panel f shows the kinetics of the changes in the populations of protein molecules for sequence segment 74–77 (Figure S11), as shown for sequence segment 18–30 in panel c. All error bars represent standard errors from two independent experiments.

pepsin digestion at pH 2.6 (see Materials and Methods). The mass profiles of the fragments were obtained at different times of refolding, ranging from 200 ms to 10 s. The extent of labeling of each fragment reported on the structure of the corresponding sequence segment of the intact protein.

The mass profiles of the 12 analyzed sequence segments could be fitted to the sum of either two-, three-, or four-Gaussian mass distributions (Figures 5 and 6 and Figures S8–S11). For the peptide fragments derived from these sequence segments, typically only one charge state was observed. Only for two peptide fragments (1–19 and 18–30) were two charge states observed. For the 12 analyzed sequence fragments, RMSD values of the fits from the data were determined both as values averaged over all times of application of the labeling pulse and as values averaged over only the 0.2, 0.3, and 0.5 s time points at which U, I_{VE} , I_E , and N are all significantly populated (Table S13). The RMSD values were used to determine whether the labeling data for a sequence segment were best described by a two-state, three-state, or four-state model. Fortunately, it was known that HX into the intact

protein could be described only by the four-state model (see above), and it was straightforward to show that the fractional populations of the four states derived from the peptide data fell on either the fractional populations of the corresponding states derived from the intact protein data or the sum of the fractional populations of two or more states derived from the intact protein data.

The peak widths of the mass distributions of the peptide fragments were not considered, as the creation and infusion of peptic fragments into the mass spectrometer took more time than the infusion of intact protein did. Consequently, the back exchange contribution was significantly larger. This was further supported by the observation that the peak width, in the case of the many peptides, was correlated to the number of protected amides.

The RMSD values indicated that a two-state model was best for sequence segment 28–34. For sequence segments 18–30 and 46–50, the RMSD values suggested that a two-state fit was much better than a four-state fit but better (albeit marginally) than a three-state fit, and the simpler two-state fit was chosen

(Table S13). For sequence segment 74–77, the observation of an only 1 Da change in mass demanded that only a two-state model be used. For sequence segment 11–18, the observation of an only 2 Da change in mass clearly precluded a four-state model, and a two-state model, which was marginally better than a three-state model, was used. For each of these five sequence segments (11–18, 18–30, 28–34, 46–50, and 74–77), at each time of folding, the apparent fractional population of the conformation giving rise to the lower-mass distribution was found to match the apparent fractional population of U determined from the labeling of intact protein (Figure 3b). The higher-mass distribution in the mass profile of each sequence segment arose from the combined labeling present in I_{VE} , I_E , and N (Figures 5i–l and 6c,f and Figure S8) at each time of refolding. Hence, for these five sequence segments, protective structure appeared to have formed during the transition from U to I_{VE} itself.

The RMSD values indicated that a three-state fit was appropriate for sequence segments 1–19, 31–35, 64–71, and 73–82 (Table S13). Although the RMSD values could not distinguish between a three-state and four-state model for sequence segment 1–11, the three-state model was chosen as this sequence segment is part of sequence segment 1–19 that could be described as three-state. For these five sequence segments (1–11, 1–19, 31–35, 64–71, and 73–82), the mass profile at each time of refolding appeared to fit best to the sum of three-Gaussian mass distributions. (Figure 5e–h,m–p and Figures S9–S12), but they manifested different kinetic behaviors. The fractional population of the conformation that was partially protected in sequence segments 1–11 and 1–19 corresponded to the sum of the apparent fractional populations of I_{VE} and I_E observed for the intact protein at the same time point (Figure 6b and Figure S8). Hence for these two sequence segments, partially protective structure formed in I_{VE} , was retained in I_E and was augmented during the $I_E \rightarrow N$ transition. For the other three sequence segments exhibiting mass profiles that fit to the sum of three-Gaussian mass distributions, the apparent fractional population of the segment conformation that was most protected corresponded to the sum of the fractional populations of I_E and N observed for the intact protein at the same time point (Figure 6e). For these three sequence segments, some protective structure was formed in I_{VE} , and was augmented during the $I_{VE} \rightarrow I_E$ transition.

The RMSD values indicated that a four-state fit was best for the two sequence segments 36–50 and 57–71 (Table S13). For these two sequence segments, the apparent fractional populations of the conformations giving rise to the distributions corresponded to the apparent fractional populations of U, I_{VE} , I_E and N observed for the intact protein at all the times of refolding (Figure 6d and Figure S8). For these two sequence segments, incremental protective structure formation occurred during each of the three folding transitions, U to I_{VE} , I_{VE} to I_E , and I_E to N. In the case of sequence segment 57–71, the RMSD values indicated that the four-state fit was only marginally better than a three-state fit when the values were those averaged over all times of application of the labeling pulse. When averaged over only the 0.2, 0.3, and 0.5 s time points at which U, I_{VE} , I_E , and N are known to be significantly populated (from data obtained for the whole protein), the RMSD values provided better support for the four-state fit. The four-state fit appeared to be more appropriate than the three-state model, because in addition to showing marginally lower RMSD values, it yielded fractional populations that agreed

better with the fractional populations of U, I_{VE} , I_E , and N determined from the global four-state fit to the data for the whole protein (Figure S13).

DISCUSSION

In previous studies of the refolding of the PI3K SH3 domain, optical measurements had shown that a collapsed intermediate is formed within a few milliseconds of folding³⁴ and that a largely swollen and hydrated intermediate is formed during the first few milliseconds of unfolding.^{36,53,60} The folding and unfolding data could together be analyzed on the basis of a four-state mechanism, with an early intermediate forming before, and a late intermediate after, the rate-limiting step of refolding.³⁵ The goal of this study was to structurally characterize the sequence of structure formation along the folding pathway using pulsed HX labeling and mass spectrometry. The methodology has allowed the detection of two intermediates that are populated before the rate-limiting step of refolding. The early intermediate was seen to form within the first few milliseconds of refolding, while the second was formed in a subsequent optically silent kinetic phase that was at least 10-fold faster than the kinetic phase corresponding to the rate-limiting step of folding. The intermediates could be distinguished from each other and from the N state on the basis of the extent of protective (to HX) structure they possessed. Hence, unlike in most folding studies in which the presence of intermediates can be inferred only from complex changes in ensemble averaging probes such as fluorescence, in the current study utilizing HX-MS, the population of each intermediate could be directly and quantitatively measured at each time of folding.

Formation of Collapsed I_{VE} . Previous multisite fluorescence resonance energy transfer measurements have shown that the folding of the PI3K SH3 domain begins with the formation of the nonspecifically collapsed, unfolded form under refolding conditions, U_C .^{34,35} The U to U_C transition is known to be completed within 150 μ s.³⁴ It is not known at present whether all molecules of U transform into U_C or whether U and U_C coexist at this time because they are separated by a small free energy barrier. It is possible that all molecules of U have collapsed to U_C ⁶¹ and that U_C possesses some nonspecific structure that is absent in U and affords some protection to HX. This protection is likely to be marginal because U_C appears to be solvated.³⁴ In any case, even if U and U_C coexist initially during refolding, because the transitions between them are very rapid ($\tau < 50 \mu$ s) and because the duration of the HX pulse was 12 ms, the cycling of molecules between U and U_C during the HX pulse would make it impossible to differentiate between U and U_C on the basis of any differences in their HX properties. Hence, in the study presented here, all molecules that become fully protonated are termed U.

The folding and unfolding transitions between U and I_{VE} are very fast (Table 1). Because of this, molecules cycle between these two states several times during the 12 ms labeling pulse. Consequently, a certain fraction of the molecules that had folded to I_{VE} at the time of application of the HX labeling pulse become labeled. The molecular cycling during the labeling pulse leads to an overestimation of the population of U, and a corresponding underestimation of the population of I_{VE} . It is for this reason that although from the rate constants listed in Table 1, U is expected to be populated to <5% at 200 ms of folding, its population is seen to be significantly larger. Furthermore, although the rate constants and equilibrium constant listed in

Table 1 indicate that the populations of I_{VE} and I_E should be the same at all times of folding, the population of I_E appears to be larger than that of I_{VE} in the pulse labeling data only because the population of I_{VE} is underestimated due to the cycling during the labeling pulse.

In this study, when the HX labeling pulse at pH 9 is given directly to unfolded protein under refolding conditions, two mass distributions, one arising from U and another from a partially structured intermediate, were observed (Figure S1c). Because the intrinsic HX time constant at pH 9 is 0.7 ms (corresponding to an intrinsic HX rate of 1500 s^{-1}) and the folding of U to the partially structured intermediate, I_{VE} , would compete with HX into U during the HX labeling pulse, this observation means that the formation of I_{VE} from U must occur in the submillisecond time domain; otherwise, only the mass distribution arising from the labeling of U would have been observed.

Formation of the More Structured Intermediate, I_E .

The observation that the mass profiles of protein that had been subjected to pulsed HX labeling at different times of refolding could be fitted to a minimum of four mass distributions (Figures 3 and 4), three of which could be attributed to the labeling of U, I_{VE} , and N, suggests that a second partially structured intermediate, I_E , is also populated before the rate-limiting step of refolding. I_{VE} and I_E differ in the amount of protective structure they possess. Importantly, the observation that the number of amide sites that are labeled in I_{VE} and I_E does not change with time of folding (Figure 3), and that only the extent to which I_{VE} and I_E are populated changes, suggests that both intermediates are discrete conformational ensembles whose average structures are invariant with the time of folding. I_E has a larger number of protected amide sites than I_{VE} does, and many amide sites are significantly more protected in I_E than they are in I_{VE} (see Results, section on stabilities of I_{VE} and I_E). It is therefore likely that I_E forms from I_{VE} on the $U \rightarrow I_{VE} \rightarrow I_E \rightarrow N$ pathway, and not separately on a parallel $U \rightarrow I_E \rightarrow N$ pathway. If this is true, then the observation that both I_{VE} and I_E disappear at the same rate as N forms indicates that on the sequential pathway, the $I_{VE} \rightarrow I_E$ transition must be substantially faster than the $I_E \rightarrow N$ transition.

For technical reasons, the earliest folding time at which the labeling pulse could be applied in this study was 200 ms. From this time onward, the populations of both I_{VE} and I_E were observed to decrease with time of folding. Thus, I_E must have become fully populated at or before 200 ms of folding, and the time constant for its formation should therefore be $\lesssim 60$ ms. I_E could conceivably form in <1 ms, but there is little evidence of any I_E having formed to any significant extent at a few milliseconds of folding (Figure S1c). When folding is monitored by measurement of fluorescence, no kinetic phase occurring with such a time constant is observed; only the kinetic phase corresponding to the formation of N is observed. It appears then that the formation of I_E is silent to fluorescence change, even though I_E appears to have substantial protective (to HX) structure. Although this might appear to be surprising, it has been shown that the late intermediate populated after the rate-limiting step of folding has U-like fluorescence;^{35,36} hence, any intermediate such as I_{VE} and I_E preceding the late intermediate on the folding pathway would also be expected to have U-like fluorescence.

Formation of Native State N of the PI3K SH3 Domain.

When the refolding of the PI3K SH3 domain under identical folding conditions was monitored by intrinsic fluorescence

measurement, two kinetic phases were observed.³⁶ One kinetic phase accounting for 70% of the fluorescence change was complete within 10 s, while the other kinetic phase accounting for the remaining fluorescence change was complete only at 300 s. It is known that the slow phase corresponds to the folding of unfolded protein with a non-native X–Pro bond. The observation in this study that all protein molecules become as protected as native protein at 10 s of folding (Figures 3 and 4 and Figure S1a) suggests that an intermediate, I_S , which differs from N in its X–Pro bond isomerization and is as well protected as is the N state, becomes fully populated at 10 s of refolding of the 30% unfolded protein with the non-native X–Pro bond. The slow phase of intrinsic fluorescence change³⁶ must occur during the $I_S \rightarrow N$ transition. The HX results indicate that the proline isomerization reaction that occurs during the $I_S \rightarrow N$ transition cannot be captured in this HX-MS study and also does not affect the native structure already present in I_S .

Mechanism of Folding. The mechanism of folding of the PI3K SH3 domain must accommodate the following observations. (1) Two partially structured intermediates, I_{VE} and I_E , form before the rate-limiting step of folding. (2) I_{VE} forms within a few milliseconds and is in pre-equilibrium with U. (3) I_E must form at least 10-fold slower than I_{VE} does, but within ~ 200 ms. Therefore, the lower limit for the rate of formation of I_E is 15 s^{-1} . (4) The N state is populated in a single-exponential manner without any apparent lag phase.

These observations can be explained by different possible folding mechanisms. The folding of the unfolded protein to the native state could occur through multiple pathways, which proceed with and without populating intermediates.^{62,63} It is possible that I_{VE} and I_E are populated on competing pathways, and in fact, previous studies have suggested that SH3 domains may fold using multiple pathways.^{30,64} However, it is more likely that I_E forms from I_{VE} on a single pathway, given that it is substantially more structured (see above), and structure is expected to develop progressively.

Alternatively and more likely, folding could occur via a sequential folding model in which U equilibrates with intermediates much faster than the formation of N, and hence, a lag phase in the formation of the native state is not seen. Indeed, a simple sequential four-state model can explain the absence of a lag phase in the formation of the native state. In this model, both I_{VE} and I_E are on-pathway folding intermediates. The mass profiles shown in Figure 3 at different times of folding are described adequately by a $U \rightarrow I_{VE} \rightarrow I_E \rightarrow N$ mechanism (Figure 3b). It is the minimal kinetic model that describes all the data. The lower limits obtained for the rate constants of formation of I_{VE} and I_E suggest that they form much faster than the N state does, suggesting that a pre-equilibrium is established among U, I_{VE} , and I_E before N starts to slowly form. It should be noted that the kinetic data can also be fitted equally well to a kinetic model with I_E being an off-pathway species or to a kinetic model with I_{VE} and I_E being on parallel pathways. The three different models cannot be discriminated on the basis of the quality of fits. Nevertheless, the sequential model describing progressive structure formation appears to be conceptually the simplest model that adequately describes the data.

It is apt to compare the sequential $U \rightarrow I_{VE} \rightarrow I_E \rightarrow N$ mechanism determined in this study to the $U \rightarrow L \rightarrow M \rightarrow N$ mechanism determined in an earlier study.³⁵ In the earlier study, in which folding in the presence of added salt was

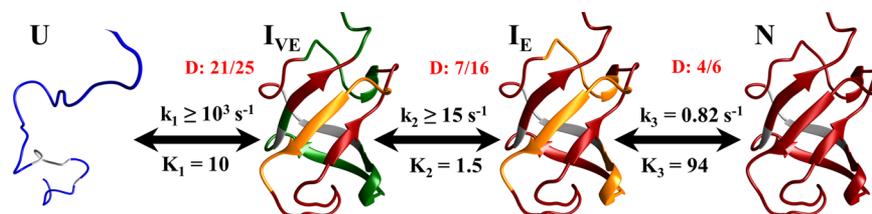


Figure 7. Kinetic mechanism of structure formation during the refolding of the PI3K SH3 domain. The folding of the PI3K SH3 domain begins with the formation of a collapsed state I_{VE} from U, on the submillisecond time scale. I_{VE} is heterogeneous in structure. I_{VE} further folds to form I_E, which undergoes structural rearrangements to form the native state in a slow phase that corresponds to the observed refolding phase in fluorescence studies. Fully unstructured regions are colored blue, while regions with native-like structure are colored dark red. I_{VE}- and I_E-like structures are colored dark green and orange, respectively. The protein regions colored gray represent the regions not covered during peptide mapping. Twenty-five amide sites in the protein become protected during the U to I_{VE} transition, and of these, 21 can be observed in the individual peptic fragments. Sixteen amide sites in the intact protein become protected during the I_{VE} to I_E transition, and of these, seven can be observed at the fragment level. Six amide sites in the intact protein become protected in the I_E to N transition, and of these, four can be observed at the fragment level. The values of the kinetic parameters shown have been obtained from global fitting of the intact protein data.

monitored by the ensemble averaging probe, fluorescence, L was shown to be part of the initially collapsed ensemble formed after folding for a few milliseconds, which was preceded by the nonspecifically collapsed U_C.³⁶ Under the folding conditions utilized in this study (absence of added salt), L, like U_C that precedes it, had U-like fluorescence, as did M that follows it. It is therefore likely that I_{VE} and I_E, which appear to form on the same time scale as L does, are components of the L ensemble. M, also termed I_U because it was first detected as an unfolding intermediate,^{30,36} was shown to form transiently after the rate-limiting step of folding^{35,36} and would therefore be populated to too small an extent to be detected in this study.

Conformational Heterogeneity Decreases with Progressive Folding. Conformational heterogeneity decreases drastically during folding, but the manner in which this happens is not well understood,⁵ because the ensemble averaging probes that are typically used to monitor refolding report only on the average structure present at any time of folding. HX-MS reports, however, on both the average structure (peak *m/z* position) and conformational heterogeneity (width of the *m/z* distribution) present, because different members of a conformational ensemble of similar energy will differ in the type and extent of protective (to HX) structure they possess.

In a pulsed HX labeling experiment, the peak width is determined by many factors, such as the pulse strength, protection factor, ratio of H to D in the labeling buffer, and back exchange.⁵⁶ Back exchange and the presence of 10% D₂O in the pulse buffer cause a small increase in the peak width for all four species. The U state manifests the smallest peak width, which is marginally larger than that seen for the fully protonated state. Although the U state is the most heterogeneous, its mass distribution upon labeling has the smallest width because amides in the U state have a homogeneous environment and are fully accessible to exchange. The observation that the N-state population manifests a peak width slightly larger than that manifested by the U state upon pulsed HX labeling at pH 9 can be attributed to either back exchange or the presence of amides with lower protection factors. Nevertheless, N-state heterogeneity cannot be ruled out.

The peak widths exhibited by I_{VE} and I_E are significantly larger than those exhibited by U and N. Furthermore, the peak widths are much larger than expected if they were determined solely by exchange at 24 amide sites (I_{VE}) and 41 amide sites (I_E).⁵⁶ I_{VE}, which has the least protection of amides (among I_{VE}, I_E, and N), shows the largest peak width (Figure 3 and Table

1). The observation that the width of the peak decreases from I_{VE} to I_E to N, as the number of protected amide sites increases, suggests that the increase in peak width from U to I_{VE} cannot be due to back exchange or the presence of residual D₂O in the pulse buffer: a larger number of protected residues will allow more random back exchange and hence an increased width. Thus, for structured I_{VE}, I_E, and N, the peak widths reflect conformational heterogeneity, and this decreases with progressive folding from I_{VE} to I_E to N. Hence, the presence of the intermediates causes conformational entropy to decrease in discrete steps during folding. This is an important result. It suggests an important productive role for intermediates. With the large decrease in the conformational entropy barrier to folding being broken down into smaller sequential entropy barriers, folding is likely to be accelerated significantly.

It is important to note that if HX into the PI3K SH3 domain during the 12 ms labeling pulse occurred by both the EX1 and EX2 mechanisms, i.e., the EXX mechanism, that too might result in an increased peak width. This is, however, unlikely to be the case, because HX into the PI3K SH3 domain has been shown to occur by the EX1 mechanism at pH 7.2 itself³⁰ and will therefore certainly occur by the EX1 mechanism at pH 9, the pH of the 12 ms labeling pulse.

Structural Delineation of the Folding Pathway of the PI3K SH3 Domain. The HX studies aided by the peptide analysis (Supporting Information) have allowed delineation of the order of structure formation. Although the data could not provide information at the level of single amino acid residues, they provide considerable detail at the secondary structure level. Different sequence segments of the protein attain native-like protective structure in either one, two, or three steps on the three-step folding pathway, U → I_{VE} → I_E → N (see Results). Five sequence segments acquire full protection during the formation of I_{VE} itself, which does not change during subsequent folding (Figure 7), indicating that these segments attain native-like interactions in I_{VE}. Five sequence segments acquire progressive protection in two steps: during the transition from U to I_{VE} and during the transition from either I_{VE} to I_E or I_E to N. Two sequence segments acquire progressively more structure with each successive folding step.

By assuming that the extent of protection directly corresponds to the extent of the structure, we can conclude that β-strand 2, β-strand 5, loop 1, and a partial region of loop 2 (Figure 1) acquire complete structure during the U to I_{VE} transition (Figure 7). Subsequent folding of PI3K SH3 from I_{VE} to I_E is associated with the acquisition of native structure in β-

strand 4, loop 4, and the C-terminal tail. The final step of folding from I_E to N involves the formation of the native structure in β -strand 1, part of loop 2, and loop 3 (Figure 7). Information about the protein segment corresponding to β -strand 3 could not be obtained because the corresponding peptide fragment could not be identified in this study. The peptide analysis could show $\sim 80\%$ of the protected amides observed in intact protein analysis, for the U to I_{VE} and I_E to N transitions. However, for the I_{VE} to I_E transition, only $\sim 45\%$ of the protected amides seen for the intact protein could be traced (Figure 7). This large difference indicates, very likely, that β -strand 3 acquires most of its protection during the I_{VE} to I_E transition. This interpretation is consistent with previous HX-MS studies, where β -strands 3 and 4 were reported to unfold simultaneously in the second step of unfolding.³⁰ Furthermore, it is also in agreement with the Φ -value-dependent computational analysis performed with the Src, Fyn, and spectrin SH3 domains, where β -strands 3 and 4 were reported to fold simultaneously.⁵¹ It should, however, be noted that the Φ -value, and Φ -value-guided, computational studies could not resolve the temporal sequence of assembly of β -strand into β -sheets, which has been achieved in the study presented here.

While the different β -strands form sequentially, the structure formation process in the individual β -strands appears to be cooperative. The β -strands attain complete or significant protection either in a single step or in two steps. β -Strands 2 and 5 clearly appear to fold in a single step during the U to I_{VE} transition. Thus, the β -sheet that constitutes β -strands 1, 2, and 5 appears to attain natively like structure early during the folding of the PI3K SH3 domain. Structure formation in β -strand 1 appears to occur in two steps in which more residues attain protection during the U to I_{VE} transition (Figure 5) than in the I_E to N transition. The folding of β -strand 4 also appears to be a two-step process, in which more residues attain protection during the transition from U to I_{VE} than during the I_{VE} to I_E transition. The peptide segments that report on structure formation in the case of β -strands 1 and 4 consist of sequences that cover coil/loop regions in addition to β -strands. It is therefore possible that the apparently two-step folding behavior of these β -strands could be due to sequential folding of the β -strand region and the coil region.

It is difficult to determine the role of the short helix (residues 33–38) in the assembly of the β -strands. Sequence segments 28–34 and 31–35 report on the labeling of the helix, and hence, it appears that the helix forms in a multistep, noncooperative manner. It would appear that the helix forms in steps because it needs to be stabilized by tertiary interactions that develop only as the β -sheet forms in multiple steps.

It is remarkable that all sequence segments acquire partial or full protection during the formation of the first intermediate, I_{VE} . The level of protection for most amide sites in I_{VE} is, however, quite low (see Results), typical of a molten globule that is collapsed but hydrated in the core.^{65,66} Only a few amide sites appear to be substantially protected in that they open slowly to HX. This result is not surprising given that I_{VE} appears to form from a nonspecific chain-collapsed U_C . It is possible that the protection against HX in I_{VE} is afforded by the formation of specific interactions within the hydrophobic clusters that are known to have formed in U_C .³⁴

Mechanism of β -Sheet Assembly. Little is understood about the mechanism of formation of β -sheet structure in a protein. The time scale of β -sheet formation ranges from milliseconds to several seconds in different proteins.^{15,67} Both

fast folding^{15,68} and slow folding¹⁸ β -sheets appear to fold in a cooperative manner. Nevertheless, studies utilizing Φ -value analysis as well as molecular dynamics suggest that the transition state of folding of many all- β proteins is polarized.^{51,69–71} Furthermore, HX-NMR studies have indicated that β -sheets assemble in multiple steps in the case of several proteins, including hFGF-1, CBTX, and RNaseH: a few β -strands attained natively like protection very early during the folding, whereas other β -strands attained natively like protection in late folding steps.^{58,72,73} In this study, too, it is seen that the β -sheets in the PI3K SH3 domain assemble in multiple steps. The roles of native-state topology, the stabilities of β -hairpins, and the packing interactions in determining the time scale and mechanism of β -sheet formation remain to be determined.

In the study presented here, the folding rate constants for the β -strands were found not to be correlated with their depth⁷⁴ or solvent accessibility (data not shown). β -Strands 2 and 5, which form the earliest during folding, appear to have the highest β -sheet (Figure S14a) propensity^{75–77} and hydrophobicity (Figure S14b),^{78–80} but β -strand 1 is the slowest folding β -strand even though it also has a very high β -sheet propensity. β -Strands 3 and 4 fold before β -strand 1, even though they have the lowest β -sheet propensities. The folding rates of β -strands 3 and 4 appear to be determined by the formation of interactions that are predominantly local (Figure S15). The rate of folding of β -strand 1 appears to be limited by the necessity to form interactions that are predominantly long-range. However, the rate of β -strand formation does not have a good correlation with contact order.^{81,82} Furthermore, the loop regions between β -strands 1 and 2, as well as between β -strands 2 and 3, acquire structure very early on during the folding process, although these regions have the highest helical (data now shown) propensity.^{75–77} It appears that the rate constants of folding of the β -strands in the PI3K SH3 domain are governed by a combination of factors: β -sheet propensities and the extent of short-range versus long-range interactions.

CONCLUSIONS

This study has yielded a detailed depiction of the structural transitions involved in the folding of the PI3K SH3 domain. The HX-MS studies in conjunction with peptic fragmentation have allowed the delineation of the structural transitions at the secondary structure level. The observed structural transitions are consistent with previous studies of the folding of the Src, Fyn, and spectrin SH3 domains, which revealed the extent of structure formation in the TS but not the temporal order in which this structure was formed. This study has resolved temporally the formation of individual β -strands, as well as the assembly of the β -strands into β -sheets. The revelation that structural heterogeneity, the manifestation of conformational entropy, decreases in a stepwise manner suggests a productive role for folding intermediates: the breaking down of the large conformational entropy barrier to folding into several smaller barriers would result in acceleration of the folding reaction.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00374.

Experimental procedure for protein purification and details of data analysis, various supporting mass profiles

(Figures S1 and S2), comparison of different Gaussian fits for +11 and +7 charge states (Figures S3 and S4), kinetics of population changes for the three-state fittings (Figure S5), effects of integrating corrections into analysis (Figures S6 and S7), fits to the mass profiles of sequence segments and kinetics of different populations obtained from those fits (Figures S8–S13), propensities of the primary sequence of the PI3K SH3 domain and its contact map (Figures S14 and S15), properties of different populations obtained from analysis of different sequence segments (Tables S1–S12), and RMSD values obtained from the fits using different models to various sequence segments (Table S13) (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: 91-80-23666150. Fax: 91-80-23636662. E-mail: jayant@ncbs.res.in.

ORCID

Jayant B. Udgaonkar: 0000-0002-7005-224X

Present Address

[†]N.A.: Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, India.

Funding

This work was funded by the Tata Institute of Fundamental Research and by the Department of Science and Technology, Government of India. J.B.U. is a recipient of a J. C. Bose National Fellowship from the Government of India.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

GdnHCl, guanidine hydrochloride; GdnDCI, guanidine deuterium chloride; HX, hydrogen exchange; MS, mass spectrometry; NMR, nuclear magnetic resonance.

REFERENCES

- (1) Jha, S. K., Dhar, D., Krishnamoorthy, G., and Udgaonkar, J. B. (2009) Continuous dissolution of structure during the unfolding of a small protein. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11113–11118.
- (2) Kuzmenkina, E. V., Heyes, C. D., and Ulrich Nienhaus, G. (2006) Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.* 357, 313–324.
- (3) Nishimura, C., Dyson, H. J., and Wright, P. E. (2002) The apomyoglobin folding pathway revisited: structural heterogeneity in the kinetic burst phase intermediate. *J. Mol. Biol.* 322, 483–489.
- (4) Sridevi, K., Lakshminanth, G. S., Krishnamoorthy, G., and Udgaonkar, J. B. (2004) Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J. Mol. Biol.* 337, 699–711.
- (5) Udgaonkar, J. B. (2008) Multiple routes and structural heterogeneity in protein folding. *Annu. Rev. Biophys.* 37, 489–510.
- (6) Grantcharova, V. P., and Baker, D. (1997) Folding dynamics of the src SH3 domain. *Biochemistry* 36, 15685–15692.
- (7) Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M., and Kay, L. E. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* 430, 586–590.
- (8) Roder, H., Elove, G. A., and Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* 335, 700–704.
- (9) Udgaonkar, J. B., and Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335, 694–699.
- (10) Juneja, J., and Udgaonkar, J. B. (2002) Characterization of the unfolding of ribonuclease a by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding. *Biochemistry* 41, 2641–2654.
- (11) Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) An amino acid code for protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 98, 105–112.
- (12) Uzawa, T., Nishimura, C., Akiyama, S., Ishimori, K., Takahashi, S., Dyson, H. J., and Wright, P. E. (2008) Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast H/D exchange coupled with 2D NMR. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13859–13864.
- (13) Vu, N. D., Feng, H., and Bai, Y. (2004) The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry* 43, 3346–3356.
- (14) Chu, R., Pei, W., Takei, J., and Bai, Y. (2002) Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry* 41, 7998–8003.
- (15) Udgaonkar, J. B., and Baldwin, R. L. (1990) Early folding intermediate of ribonuclease A. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8197–8201.
- (16) Jennings, P. A., and Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262, 892–896.
- (17) Morozova-Roche, L. A., Jones, J. A., Noppe, W., and Dobson, C. M. (1999) Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. *J. Mol. Biol.* 289, 1055–1073.
- (18) Radford, S. E., Dobson, C. M., and Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358, 302–307.
- (19) Juneja, J., and Udgaonkar, J. B. (2003) NMR studies of protein folding. *Curr. Sci.* 84, 157–172.
- (20) Bollen, Y. J., Kamphuis, M. B., and van Mierlo, C. P. (2006) The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4095–4100.
- (21) Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science* 269, 192–197.
- (22) Bhuyan, A. K., and Udgaonkar, J. B. (1998) Two structural subdomains of barstar detected by rapid mixing NMR measurement of amide hydrogen exchange. *Proteins: Struct., Funct., Genet.* 30, 295–308.
- (23) Moulick, R., Das, R., and Udgaonkar, J. B. (2015) Partially Unfolded Forms of the Prion Protein Populated under Misfolding-promoting Conditions: CHARACTERIZATION BY HYDROGEN EXCHANGE MASS SPECTROMETRY AND NMR. *J. Biol. Chem.* 290, 25227–25240.
- (24) Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Mol. Biol.* 3, 782–787.
- (25) Deng, Y., and Smith, D. L. (1999) Rate and equilibrium constants for protein unfolding and refolding determined by hydrogen exchange-mass spectrometry. *Anal. Biochem.* 276, 150–160.
- (26) Heidary, D. K., Gross, L. A., Roy, M., and Jennings, P. A. (1997) Evidence for an obligatory intermediate in the folding of interleukin-1 beta. *Nat. Struct. Mol. Biol.* 4, 725–731.
- (27) Pan, H., Raza, A. S., and Smith, D. L. (2004) Equilibrium and kinetic folding of rabbit muscle triosephosphate isomerase by hydrogen exchange mass spectrometry. *J. Mol. Biol.* 336, 1251–1263.
- (28) Hamid Wani, A., and Udgaonkar, J. B. (2006) HX-ESI-MS and optical studies of the unfolding of thioredoxin indicate stabilization of a partially unfolded, aggregation-competent intermediate at low pH. *Biochemistry* 45, 11226–11238.

- (29) Malhotra, P., and Udgaonkar, J. B. (2015) Tuning Cooperativity on the Free Energy Landscape of Protein Folding. *Biochemistry* 54, 3431–3441.
- (30) Wani, A. H., and Udgaonkar, J. B. (2009) Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20711–20716.
- (31) Wintrode, P. L., Rojsajakul, T., Vadrevu, R., Matthews, C. R., and Smith, D. L. (2005) An obligatory intermediate controls the folding of the alpha-subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* 347, 911–919.
- (32) Arrington, C. B., Teesch, L. M., and Robertson, A. D. (1999) Defining protein ensembles with native-state NH exchange: kinetics of interconversion and cooperative units from combined NMR and MS analysis. *J. Mol. Biol.* 285, 1265–1275.
- (33) Guijarro, J. I., Morton, C. J., Plaxco, K. W., Campbell, I. D., and Dobson, C. M. (1998) Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* 276, 657–667.
- (34) Dasgupta, A., and Udgaonkar, J. B. (2010) Evidence for initial non-specific polypeptide chain collapse during the refolding of the SH3 domain of PI3 kinase. *J. Mol. Biol.* 403, 430–445.
- (35) Dasgupta, A., and Udgaonkar, J. B. (2012) Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry* 51, 4723–4734.
- (36) Wani, A. H., and Udgaonkar, J. B. (2009) Revealing a concealed intermediate that forms after the rate-limiting step of refolding of the SH3 domain of PI3 kinase. *J. Mol. Biol.* 387, 348–362.
- (37) Bezsonova, I., Korzhnev, D. M., Prosser, R. S., Forman-Kay, J. D., and Kay, L. E. (2006) Hydration and packing along the folding pathway of SH3 domains by pressure-dependent NMR. *Biochemistry* 45, 4711–4719.
- (38) Neudecker, P., Zarrine-Afsar, A., Davidson, A. R., and Kay, L. E. (2007) Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 104, 15717–15722.
- (39) Sadqi, M., Casares, S., Abril, M. A., Lopez-Mayorga, O., Conejero-Lara, F., and Freire, E. (1999) The native state conformational ensemble of the SH3 domain from alpha-spectrin. *Biochemistry* 38, 8899–8906.
- (40) Viguera, A. R., and Serrano, L. (2003) Hydrogen-exchange stability analysis of Bergerac-Src homology 3 variants allows the characterization of a folding intermediate in equilibrium. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5730–5735.
- (41) Wales, T. E., and Engen, J. R. (2006) Partial unfolding of diverse SH3 domains on a wide timescale. *J. Mol. Biol.* 357, 1592–1604.
- (42) Li, J., Shinjo, M., Matsumura, Y., Morita, M., Baker, D., Ikeguchi, M., and Kihara, H. (2007) An alpha-helical burst in the src SH3 folding pathway. *Biochemistry* 46, 5072–5082.
- (43) Neudecker, P., Robustelli, P., Cavalli, A., Walsh, P., Lundstrom, P., Zarrine-Afsar, A., Sharpe, S., Vendruscolo, M., and Kay, L. E. (2012) Structure of an intermediate state in protein folding and aggregation. *Science* 336, 362–366.
- (44) Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R., and Kay, L. E. (2006) Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* 363, 958–976.
- (45) Northey, J. G., Maxwell, K. L., and Davidson, A. R. (2002) Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* 320, 389–402.
- (46) Ventura, S., Vega, M. C., Lacroix, E., Angrand, I., Spagnolo, L., and Serrano, L. (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nat. Struct. Biol.* 9, 485–493.
- (47) Viguera, A. R., Vega, C., and Serrano, L. (2002) Unspecific hydrophobic stabilization of folding transition states. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5349–5354.
- (48) Zarrine-Afsar, A., Wallin, S., Neculai, A. M., Neudecker, P., Howell, P. L., Davidson, A. R., and Chan, H. S. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9999–10004.
- (49) Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* 6, 1016–1024.
- (50) Lindorff-Larsen, K., Vendruscolo, M., Paci, E., and Dobson, C. M. (2004) Transition states for protein folding have native topologies despite high structural variability. *Nat. Struct. Mol. Biol.* 11, 443–449.
- (51) Hubner, I. A., Edmonds, K. A., and Shakhnovich, E. I. (2005) Nucleation and the transition state of the SH3 domain. *J. Mol. Biol.* 349, 424–434.
- (52) Matsumura, Y., Shinjo, M., Kim, S. J., Okishio, N., Gruebele, M., and Kihara, H. (2013) Transient helical structure during PI3K and Fyn SH3 domain folding. *J. Phys. Chem. B* 117, 4836–4843.
- (53) Dasgupta, A., and Udgaonkar, J. B. (2012) Transient non-native burial of a Trp residue occurs initially during the unfolding of a SH3 domain. *Biochemistry* 51, 8226–8234.
- (54) Cheung, M. S., Garcia, A. E., and Onuchic, J. N. (2002) Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U. S. A.* 99, 685–690.
- (55) Witten, J., Ruschak, A., Poterba, T., Jaramillo, A., Miranker, A. D., and Jaswal, S. S. (2015) Mapping Protein Conformational Landscapes under Strongly Native Conditions with Hydrogen Exchange Mass Spectrometry. *J. Phys. Chem. B* 119, 10016–10024.
- (56) Weis, D. D., Wales, T. E., Engen, J. R., Hotchkko, M., and Ten Eyck, L. F. (2006) Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *J. Am. Soc. Mass Spectrom.* 17, 1498–1509.
- (57) Booker, G. W., Gout, I., Downing, A. K., Driscoll, P. C., Boyd, J., Waterfield, M. D., and Campbell, I. D. (1993) Solution structure and ligand-binding site of the SH3 domain of the p85 alpha subunit of phosphatidylinositol 3-kinase. *Cell* 73, 813–822.
- (58) Hu, W., Walters, B. T., Kan, Z. Y., Mayne, L., Rosen, L. E., Marqusee, S., and Englander, S. W. (2013) Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 110, 7684–7689.
- (59) Walters, B. T., Mayne, L., Hinshaw, J. R., Sosnick, T. R., and Englander, S. W. (2013) Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18898–18903.
- (60) Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J. Phys. Chem. B* 118, 6380–6392.
- (61) Udgaonkar, J. B. (2013) Polypeptide chain collapse and protein folding. *Arch. Biochem. Biophys.* 531, 24–33.
- (62) Aghera, N., and Udgaonkar, J. B. (2012) Kinetic studies of the folding of heterodimeric monellin: evidence for switching between alternative parallel pathways. *J. Mol. Biol.* 420, 235–250.
- (63) Wildegger, G., and Kiefhaber, T. (1997) Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270, 294–304.
- (64) Lam, A. R., Borreguero, J. M., Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. (2007) Parallel folding pathways in the SH3 domain protein. *J. Mol. Biol.* 373, 1348–1360.
- (65) Mok, Y. K., Kay, C. M., Kay, L. E., and Forman-Kay, J. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* 289, 619–638.
- (66) Nishimura, C., Dyson, H. J., and Wright, P. E. (2008) The kinetic and equilibrium molten globule intermediates of apoleghemoglobin differ in structure. *J. Mol. Biol.* 378, 715–725.
- (67) Carlsson, U., and Jonsson, B. H. (1995) Folding of beta-sheet proteins. *Curr. Opin. Struct. Biol.* 5, 482–487.

(68) Briggs, M. S., and Roder, H. (1992) Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* 89, 2017–2021.

(69) Lammert, H., Noel, J. K., and Onuchic, J. N. (2012) The dominant folding route minimizes backbone distortion in SH3. *PLoS Comput. Biol.* 8, e1002776.

(70) Piana, S., Lindorff-Larsen, K., and Shaw, D. E. (2013) Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5915–5920.

(71) Went, H. M., and Jackson, S. E. (2005) Ubiquitin folds through a highly polarized transition state. *Protein Eng., Des. Sel.* 18, 229–237.

(72) Samuel, D., Kumar, T. K., Balamurugan, K., Lin, W. Y., Chin, D. H., and Yu, C. (2001) Structural events during the refolding of an all beta-sheet protein. *J. Biol. Chem.* 276, 4134–4141.

(73) Sivaraman, T., Kumar, T. K., Tu, Y. T., Wang, W., Lin, W. Y., Chen, H. M., and Yu, C. (1999) Secondary structure formation is the earliest structural event in the refolding of an all beta-sheet protein. *Biochem. Biophys. Res. Commun.* 260, 284–288.

(74) Tan, K. P., Nguyen, T. B., Patel, S., Varadarajan, R., and Madhusudhan, M. S. (2013) Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res.* 41, W314–321.

(75) Chou, P. Y., and Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* 47, 45–148.

(76) Deleage, G., and Roux, B. (1987) An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng., Des. Sel.* 1, 289–294.

(77) Levitt, M. (1978) Conformational preferences of amino acids in globular proteins. *Biochemistry* 17, 4277–4285.

(78) Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179, 125–142.

(79) Sweet, R. M., and Eisenberg, D. (1983) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. Mol. Biol.* 171, 479–488.

(80) Tanford, C. (1962) Contribution of Hydrophobic Interactions to Stability of Globular Conformation of Proteins. *J. Am. Chem. Soc.* 84, 4240–4247.

(81) Gromiha, M. M., and Selvaraj, S. (2001) Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J. Mol. Biol.* 310, 27–32.

(82) Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277, 985–994.