

# Differentiating between the sequence of structural events on alternative pathways of folding of a heterodimeric protein

Rupam Bhattacharjee<sup>1,2</sup>  | Jayant B. Udgaonkar<sup>1,2</sup> 

<sup>1</sup>National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, Karnataka, India

<sup>2</sup>Indian Institute of Science Education and Research, Pune, Maharashtra, India

## Correspondence

Jayant B. Udgaonkar, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, Karnataka, India.

Email: [jayant@iiserpune.ac.in](mailto:jayant@iiserpune.ac.in)

## Funding information

Department of Science and Technology, Government of India., Grant/Award Number: 30119476; Indian Institute of Science Education and Research Pune; National Centre for Biological Sciences

**Review Editor:** John Kuriyan

## Abstract

Distinguishing between competing pathways of folding of a protein, on the basis of how they differ in their progress of structure acquisition, remains an important challenge in protein folding studies. A previous study had shown that the heterodimeric protein, double chain monellin (dcMN) switches between alternative folding pathways upon a change in guanidine hydrochloride (GdnHCl) concentration. In the current study, the folding of dcMN has been characterized by the pulsed hydrogen exchange (HX) labeling methodology used in conjunction with mass spectrometry. Quantification of the extent to which folding intermediates accumulate and then disappear with time of folding at both low and high GdnHCl concentrations, where the folding pathways are known to be different, shows that the folding mechanism is describable by a triangular three-state mechanism. Structural characterization of the productive folding intermediates populated on the alternative pathways has enabled the pathways to be differentiated on the basis of the progress of structure acquisition that occurs on them. The intermediates on the two pathways differ in the extent to which the  $\alpha$ -helix and the rest of the  $\beta$ -sheet have acquired structure that is protective against HX. The major difference is, however, that  $\beta$ 2 has not acquired any protective structure in the intermediate formed on one pathway, but it has acquired significant protective structure in the intermediate formed on the alternative pathway. Hence, the sequence of structural events is different on the two alternative pathways.

## KEYWORDS

alternative pathways, hydrogen exchange, intermediate, kinetics, monellin

## 1 | INTRODUCTION

A major challenge in the field of protein folding is to detect and characterize structurally the intermediates that populate during folding. The nature of the

intermediates and the role they play during folding are not yet fully understood. Several small proteins appear to fold by a two-state mechanism, without populating any intermediates to detectable extents, calling into question the significance of intermediates during folding.<sup>1–9</sup> Very

**Abbreviations:** D, deuterium; DcMN, double-chain monellin; ETD, electron transfer dissociation; GdnDCl, guanidine deuteriochloride; GdnHCl, guanidine hydrochloride; HX-MS, hydrogen exchange coupled to mass spectrometry; MNEI, single-chain monellin.

often, it is the transient nature of these intermediates that prevents them from being detected by the commonly used experimental methods; hence, the folding process appears to be two-state. For example, protein L was believed to be a two-state folder, but it was shown recently to fold through an intermediate.<sup>10</sup> On the other hand, many proteins have been shown to fold through one or multiple intermediates.<sup>11–17</sup> In most of the cases, the intermediate was proposed to be an on-pathway productive intermediate.<sup>18–25</sup> However, a few studies have also identified intermediates that possess non-native interactions, and which must unfold first, at least partially, in order to reach the native state.<sup>26–28</sup> Hence, they were suggested to be off the folding pathway. Clear experimental distinction between on and off-pathway intermediates is difficult to obtain, and usually, only indirect arguments can be made against an intermediate being off-pathway, based on a comparison between the structures of intermediates and transition states.<sup>29</sup>

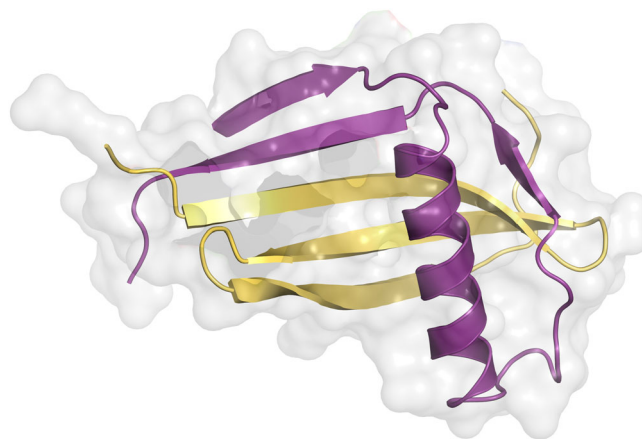
Structural characterization of intermediates is challenging with the ensemble averaging probes that are typically used to study protein folding reactions. Ensemble averaging probes measure the average change in the measured property and cannot distinguish between multiple sub-populations of conformations that may be present together. High-resolution probes such as hydrogen exchange (HX) in conjunction with NMR are of great utility in this regard. HX-NMR studies can provide residue-level information on the structure of folding intermediates.<sup>30,31</sup> Such studies have not only provided important evidences for the existence of folding<sup>32–36</sup> and unfolding<sup>37</sup> intermediates but also, indirectly for unproductive dead-end intermediates,<sup>27</sup> molten globule intermediates,<sup>38,39</sup> and for multiple pathways of folding<sup>11,40</sup> and unfolding.<sup>37,41</sup>

While multiple pathways for folding and unfolding have been observed now for many proteins, both in experimental<sup>11,37,42,43</sup> and molecular dynamics simulation studies,<sup>44–47</sup> it has been difficult to distinguish experimentally between the detected pathways on the basis of how they differ in the progress of structure acquisition that occurs on them. Competing folding and unfolding pathways have been distinguished under different folding conditions on the basis of the differences in the solvent exposure of the transition states (TSs) on the pathways but could not be distinguished on the basis of how they differ in the acquisition or loss of secondary structure. The problem of determining how folding pathways are distinct from each other is compounded by the structures of TSs<sup>48</sup> and folding intermediates<sup>49,50</sup> being malleable and thereby different under different folding conditions. It is even more difficult to distinguish between pathways that compete under the same folding conditions. Only

recently has it been possible to show experimentally that for pathways whose origin lies in the heterogeneity of the unfolded state and early folding intermediates, the sequence of structural events is different on folding pathways that operate simultaneously.<sup>51</sup>

When HX is coupled to mass spectrometry (MS), it can distinguish between multiple conformations and their relative populations. HX-MS measurements have also provided direct evidence for the existence of multiple folding pathways<sup>52</sup> and have enabled the characterization of folding<sup>53–55</sup> and unfolding<sup>56–59</sup> intermediates. Native state HX-MS has revealed the lack of cooperativity in protein unfolding reactions and has led to the detection of high-energy intermediates populated during unfolding under native and native-like (marginally denaturing) conditions.<sup>57–60</sup> It has also brought out the effects of mutations and co-solutes, such as denaturant and osmolytes, on the cooperativity of protein (un)folding reactions.<sup>61,62</sup> Pulse labeling HX-MS studies allowed the detection of the population of transiently formed intermediates, as well as the extent of structure formation on the folding pathways of lysozyme<sup>53</sup> and the PI3K SH3 domain.<sup>63</sup>

The small protein monellin has been used productively as a model protein for unfolding studies.<sup>42,57,59,64–66</sup> Monellin has been studied in its naturally occurring heterodimeric form (dcMN)<sup>48,59,67</sup> as well as in its artificially created single chain form.<sup>68–71</sup> It is an  $\alpha$ - $\beta$  heterodimeric protein in which the sole  $\alpha$ -helix is packed against five  $\beta$  strands (Figure 1). The folding mechanism of single-chain monellin is complex, partly because of heterogeneity in the unfolded state, with multiple intermediates and pathways.<sup>42,43,51</sup> Native-state HX-MS and time-resolved



**FIGURE 1** Structure of double chain monellin (dcMN). Chains B and A are shown in purple and yellow, respectively. The  $\beta$ 1- $\alpha$ 1- $\beta$ 2 segment of Chain B associates with the  $\beta$ 3- $\beta$ 4- $\beta$ 5 segment of Chain A by noncovalent interchain interactions. The structure was drawn using the Pymol software and PDB ID 3MON.

fluorescence resonance energy transfer studies have revealed much about how folding cooperativity is modulated by protein stability<sup>57,61,70</sup> and how structural steps during protein folding can occur both in a cooperative manner and in a completely noncooperative manner.<sup>60,71</sup>

The folding mechanism of dcMN also appears to be similarly complex, with folding occurring in multiple steps.<sup>48</sup> As the individual polypeptide chains of dcMN are unstructured in isolation, its folding is an example of binding-induced folding.<sup>67</sup> The unstructured encounter complex formed when the chains bind to each other, folds to the native state via transient intermediates. Folding was shown to switch from a pathway characterized by a less compact TS to one with a more compact TS, upon an increase in the denaturant concentration in which folding was carried out.<sup>48</sup> Since folding intermediates were found to be populated at both low and high denaturant concentrations, it seemed possible that the folding pathways could be distinguished structurally on the basis of the secondary structural content of the intermediates in the two folding conditions.

In this study, the refolding of dcMN has been studied using pulse labeling HX coupled with mass spectrometry. It is shown that dcMN folds via a three-state triangular mechanism, at both low and high GdnHCl concentration. The structures of the intermediates formed in 0.1 M ( $I_1$ ) and 0.4 M ( $I_2$ ) GdnDCl have been characterized.  $I_1$  and  $I_2$  have been differentiated on the basis of the protection of their amide hydrogens belonging to different secondary structural elements. The assembly of the  $\beta$ -sheet has also been resolved temporally, and the sequence of structure acquisition has been determined.

## 2 | RESULTS

### 2.1 | Labeling of U and N

It is important for any kinetic reaction to determine the start and end point of the reaction. Hence, the mass distributions of the U and N states of dcMN, after the application of different HX labeling pulses, were first determined.

The protein was unfolded in deuterated buffer containing 4 M GdnDCl for 12 hr. When the unfolded protein was subjected to a 5 s HX labeling pulse in 4 M GdnHCl at pH 9, the mass profile showed a single peak with the centroid  $m/z$  at  $853.65 \pm 0.1$  and  $770.15 \pm 0.1$  for Chains B and A, respectively (+7 charge state for both the chains; Figure S1). The centroid  $m/z$  of the mass distribution of the fully protonated protein was at  $853.2 \pm 0.1$  and  $769.86 \pm 0.1$  for Chains B and A, respectively (+7 charge state for both the chains). Thus, unfolded

Chain B retained three deuteriums and unfolded Chain A retained two deuteriums when labeled by a 5 s HX pulse at pH 9. This was not unexpected as the final HX reaction contained 90% H and 10% D.

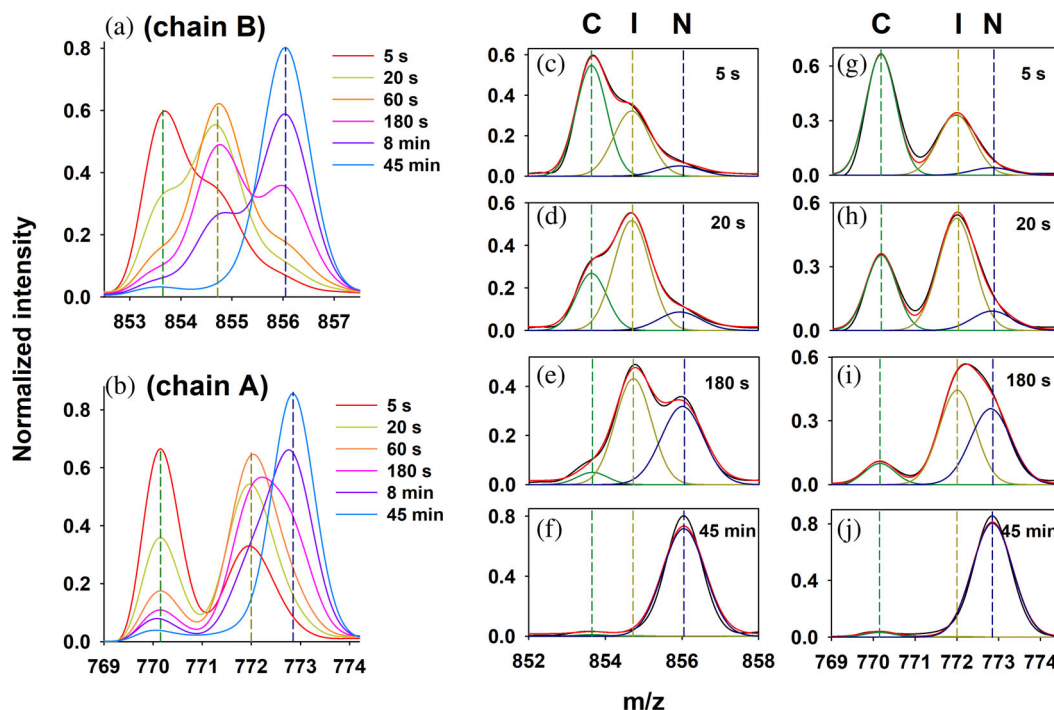
When native deuterated protein was subjected to a 5 s HX labeling pulse in both 0.1 and 0.4 M GdnDCl at pH 9, the mass distribution showed a single peak centered at  $m/z$   $855.95 \pm 0.1$  for Chain B, and at  $m/z$   $772.75 \pm 0.1$  for Chain A (+7 charge state for both the chains; Figure S1). This indicated that a total of  $34 \pm 1$  deuteriums were protected in N against HX by the labeling pulse at pH 9, out of which 16 deuteriums were protected in Chain B and 18 deuteriums were protected in Chain A.

HX labeling of U and N was also carried out by a 5 s pulse at pH 7 to determine the strength of the interactions in the native and intermediate structures. The number of deuteriums retained by U after HX labeling by the pH 7 pulse was the same as by the pH 9 pulse (Figure S1). But the N state retained more deuteriums when labeled with a HX pulse at pH 7 than at pH 9. The mass profiles of Chain B and Chain A were centered at  $m/z$   $857.15 \pm 0.1$  and  $773.25 \pm 0.1$ , respectively (+7 charge state for both the chains; Figure S1). This indicated that 24 deuteriums were protected against HX in Chain B, and 22 deuteriums were protected in Chain A, for the labeling pulse at pH 7. In the case of the mass distributions of both U and N, the centroid  $m/z$  and peak width were observed to vary only by  $\pm 0.1$ .

### 2.2 | Refolding kinetics of dcMN in 0.1 and 0.4 M GdnDCl monitored using a 5 s HX labeling pulse at pH 9

It was shown in a previous study that dcMN folds via alternative pathways at low (<0.2 M) and high (>0.2 M) GdnHCl concentrations, and that the transition states and the intermediates populated on these pathways were different in their compactness.<sup>48</sup> Thus, any difference in their secondary structures was expected to be detected by pulsed HX labeling. In this study, 5 s pulses of HX labeling were applied at different times of refolding in 0.1 and 0.4 M GdnHCl, starting with free Chains B and A present in the unfolded state (U) in 4 M GdnHCl. It is known that immediately upon transfer to refolding conditions, Chains B and A associate to form the encounter complex C, which is U-like in its Trp fluorescence.<sup>48</sup>

Figure 2a,b shows the mass distributions of Chains B and A, respectively, obtained using a HX labeling pulse at pH 9 for different times of refolding in 0.1 M GdnDCl, and Figure 3a,b do likewise for different times of refolding in 0.4 M GdnDCl. In both refolding conditions, the mass distributions showed the presence of at least three



**FIGURE 2** Kinetics of refolding of double chain monellin (dcMN) in 0.1 M GdnDCl at pH 7, 25°C monitored using a 5 s hydrogen exchange (HX) labeling pulse at pH 9. Representative mass spectra are shown for Chain B (a) and Chain A (b) at different times of refolding. The number of deuteriums protected in the C, I, and N states were 3, 7, and 16, respectively, for Chain B and 2, 12, and 18, respectively, for Chain A. Three-state fits to the mass spectra obtained at different times of refolding are shown in (c–f) for Chain B and in (g–j) for Chain A. In each panel, the black line represents the experimentally determined mass profile data, and the red line is the fit of the data to the sum of three Gaussian mass distributions. In all the panels, the vertical dashed green, yellow and blue lines represent the centers of the mass distributions corresponding to the C, I, and N states, respectively.

different conformations at any time of refolding. C molecules were the least protected and showed peaks centered at  $m/z$  853.65 and 770.15 for chains B and A, respectively. The  $m/z$  values for the labeled C molecules were the same as those for labeled U molecules (Figure S1) indicating that the amide hydrogen sites were as unprotected in C as they were in U. The molecules having the native conformation showed peaks at  $m/z$  856 and 772.8 for Chains B and A, respectively, and retained a larger number of deuteriums (16 in Chain B and 18 in Chain A) indicating that they were the most protected against HX (Figures 2 and 3).

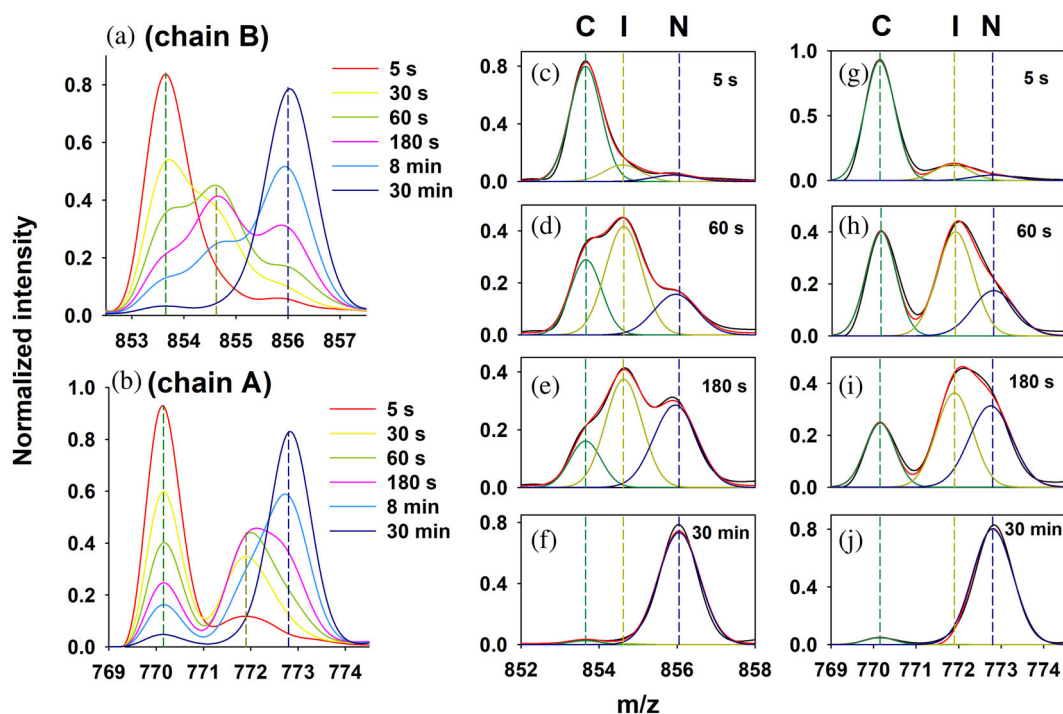
The peak at intermediate  $m/z$  corresponded to the molecules having intermediate protection. The centroid  $m/z$  of the intermediate observed in 0.1 M GdnHCl was not significantly different from that of the intermediate observed in 0.4 M GdnHCl (Table 1). In the intermediate, 7 and 12 deuteriums were protected in Chains B and A, respectively (Figures 2 and 3). The number of deuteriums protected in these two intermediates was the same (Table 1). It was possible that the deuteriums were protected in different regions in the intermediates seen in 0.1 and 0.4 M GdnDCl. Hence, it was necessary to

structurally delineate the differences in secondary structure between the intermediates.

The mass profiles were fit to the sum of three Gaussian mass distributions to determine the fractional areas of the mass distributions corresponding to fractional populations of C, I, and N states. For the fit, the centroid  $m/z$  and width of the mass distributions were not allowed to vary across the different times of refolding at which labeling pulse was applied; only the area under each mass distribution was allowed to vary. Figures 2c–j and 3c–j show the deconvoluted mass distributions of the C, I, and N states of both the chains at different refolding times, during refolding in 0.1 and 0.4 M GdnDCl, respectively.

The fractional populations of C, I, and N were plotted against time of refolding in 0.1 M GdnDCl (Figure 4a,b) and in 0.4 M GdnDCl (Figure 4c,d). The fractional population of C decreased with time of refolding, while that of I first increased and then decreased with time. The fractional population of N increased with time, with no lag being observed in the formation of N for both the chains. This indicated that a direct pathway from C to N must be present in addition to a  $C \leftrightarrow I \leftrightarrow N$  pathway, otherwise a





**FIGURE 3** Refolding kinetics of double chain monellin (dcMN) in 0.4 M GdnDCl at pH 7, 25°C monitored using a 5 s hydrogen exchange (HX) labeling pulse at pH 9. Representative mass spectra are shown for Chain B (a) and Chain A (b) for when the labeling pulse was applied at different times of refolding. The number of deuteriums protected in the C, I, and N states were 3, 7, and 16, respectively, for Chain B and 2, 12, and 18, respectively, for Chain A. Three-state fits to the mass spectra obtained at different times of refolding are shown in (c–f) for Chain B, and in (g–j) for Chain A. In each panel, the black line, which represents the experimentally determined mass profile data, was fit to the sum of three Gaussian mass distributions (red line). In all the panels, the vertical dashed green, yellow and blue lines represent the centers of the mass distributions corresponding to the populations of the C, I, and N states, respectively.

lag would have been observed in the formation of N. Hence, refolding under both refolding conditions could be described by a triangular three-state mechanism (Scheme 1).

The mass profiles of both the chains at different times of refolding in both 0.1 and 0.4 M GdnDCl were globally fit to the triangular folding scheme to obtain the kinetic parameters of refolding which are listed in Table 1. The changes in the fractional populations of C, I, and N with time of refolding obtained from global fitting of the data (Figure 4) to a three-state triangular mechanism agreed well with that obtained from discrete fitting of the mass profiles to sum of three Gaussian mass distributions (Figures 2 and 3). The mass distributions simulated using these kinetic parameters and the triangular folding scheme also agreed well with the experimentally observed mass distributions for both the chains in both 0.1 (Figure S2) and 0.4 M GdnDCl (Figure S3). This suggested that a triangular three-state folding mechanism is the simplest and minimal kinetic model that can account for the data.

The pulse-labeling HX data (Figure 4) showed that I was populated to an extent of ~75% for refolding in 0.1 M

GdnDCl and to an extent of ~60% for refolding in 0.4 M GdnDCl. The rate constant for the formation of I was 0.1 and 0.03 s<sup>-1</sup> in 0.1 and 0.4 M GdnDCl, respectively. But the rate constant of the slow kinetic phase did not change in these two different solvent conditions, which suggested that it could be associated with proline isomerization.

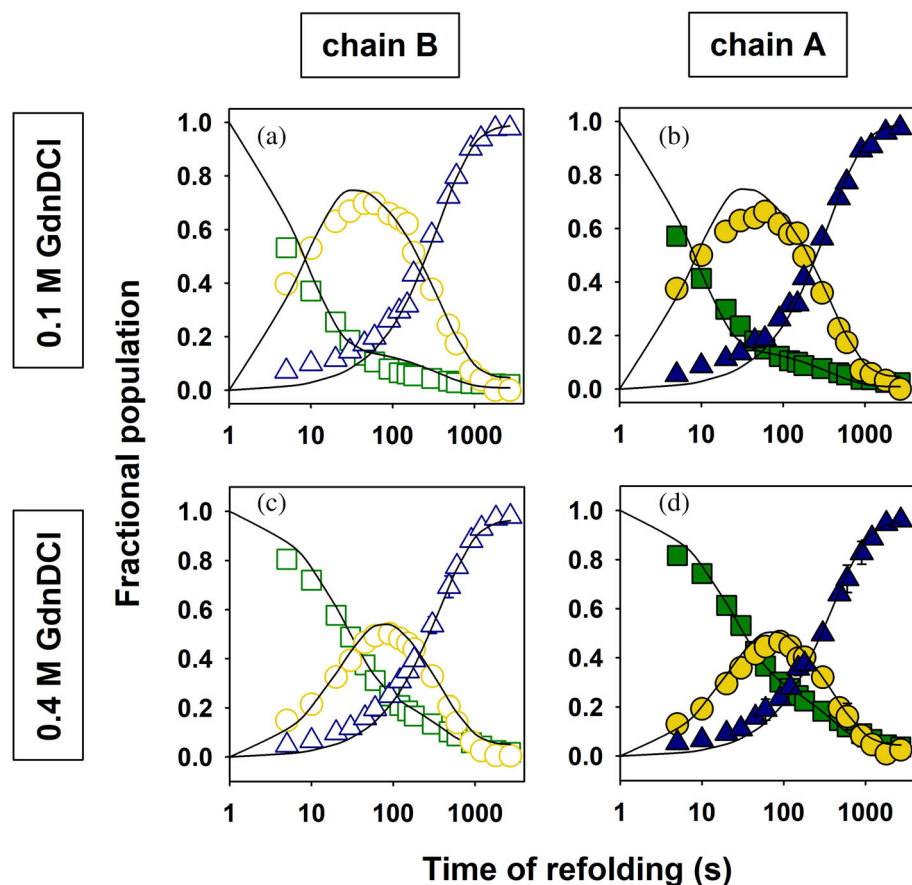
When monitored by measurement of the intrinsic fluorescence of Trp, refolding was found to occur in a fast and a slow phase. The relative amplitude and rate constant of the fast phase was 87% and 0.07 s<sup>-1</sup> in 0.1 M GdnHCl, and 70% and 0.02 s<sup>-1</sup> in 0.4 M GdnHCl (Figure S4). The rate constants indicated that the fast phase of fluorescence change corresponded to the formation of I (see above). The relative amplitude of the fast phase of fluorescence change was accounted for by the populations of I and N present at the time when I was maximally populated, in both 0.1 and 0.4 M GdnHCl. The rate constants of the slow phase of refolding monitored by pulsed-HX labeling, was also similar to the slow phase of refolding monitored by fluorescence (Tables 1 and S1).

Rate constants (s <sup>-1</sup> ) <sup>b</sup>		Parameters describing the mass distributions <sup>c</sup>		
			Chain B	Chain A
k <sub>CI</sub>	0.1 ± 0.014	W <sub>C</sub>	0.81	0.71
	0.026 ± 0.001		0.81	0.71
k <sub>IN</sub>	(2.65 ± 0.07) × 10 <sup>-3</sup>	W <sub>I</sub>	0.94	0.90
	(2.45 ± 0.07) × 10 <sup>-3</sup>		0.98	0.90
k <sub>CN</sub>	(3.05 ± 0.07) × 10 <sup>-3</sup>	W <sub>N</sub>	1.00	0.92
	(2.65 ± 0.07) × 10 <sup>-3</sup>		1.00	1.00
k <sub>IC</sub>	(1.75 ± 0.07) × 10 <sup>-2</sup>	C <sub>C</sub>	853.65	770.15
	(1.07 ± 0.014) × 10 <sup>-2</sup>		853.65	770.15
k <sub>NI</sub>	(1.25 ± 0.07) × 10 <sup>-4</sup>	C <sub>I</sub>	854.75	771.95
	(1.35 ± 0.07) × 10 <sup>-4</sup>		854.65	771.95
k <sub>NC</sub>	(2.53 ± 0.24) × 10 <sup>-5</sup>	C <sub>N</sub>	856	772.8
	(6.0 ± 0.2) × 10 <sup>-5</sup>		855.95	772.7

<sup>a</sup>The white and gray colored rows denote the values of the parameters obtained for refolding in 0.1 and 0.4 M GdnDCI, respectively.

<sup>b</sup>The same set of rate constants was used to globally fit the refolding data obtained for intact Chain B and Chain A under each refolding condition, as it has been shown that the two chains bind together before the start of any structure formation.<sup>48,67</sup>

<sup>c</sup>W represents the width of the Gaussian distribution, and C represents the centroid m/z value of the Gaussian distribution for C, I, and N populated during refolding. The width and centroid m/z values of the different mass distributions were allowed to vary only by ±0.05 and ±0.1, respectively, when the mass profiles were fit either to the sum of three Gaussian mass distributions (discrete analysis) or globally to the three-state triangular mechanism.



**TABLE 1** Kinetic parameters obtained from global fitting of the refolding data of the intact chains in 0.1 and 0.4 M GdnDCI at pH 7, 25°C, monitored using a pH 9 HX labeling pulse<sup>a</sup>

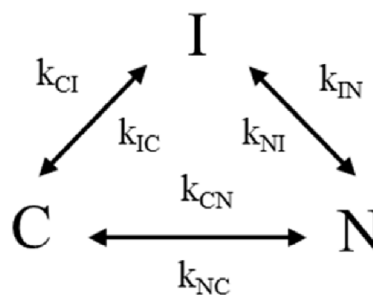
**FIGURE 4** Kinetic data obtained for the refolding of double chain monellin (dcMN) in 0.1 M (a, b) and 0.4 M (c, d) GdnDCI at pH 7, 25°C monitored using a hydrogen exchange (HX) labeling pulse at pH 9. Panels (a) and (c) represent the data for Chain B, whereas (b) and (d) represent the data for Chain A. Green squares, yellow circles and blue triangles represent the fractional populations of C, I, and N, respectively, at different times of refolding. These fractional populations were obtained from the deconvoluted mass distributions shown in Figures 2 and 3. The lines passing through the data points represent the kinetics of the change in the populations of C, I, and N obtained from a global fit of the mass profiles to a three-state triangular folding mechanism. The values obtained for the kinetic parameters from the global fitting are listed in Table 1. The error bars show the standard deviations in the data obtained from two separate experiments.

### 2.3 | Intermediate does not lose its structure during the duration of the labeling pulse

It was shown in a previous study that HX into dcMN occurs in the EX1 regime at pH 9.<sup>59</sup> The intrinsic rate constant of HX is  $4,324 \text{ s}^{-1}$  at pH 9. Thus, a 5 s HX labeling pulse at pH 9 will label all the amide sites, which are unprotected (unstructured) in any of the conformations (C, I, or N) at the time of application of the pulse. The duration of the HX pulse (5 s) used in this study was eightfold shorter than the time constant of the fast phase of refolding (40 s). Hence, no significant folding would have occurred during the duration of the pulse. Hence, the protein molecules could not have cycled between the C and I states during the duration of the pulse, which would have led to an incorrect determination of the fractional populations of C and I. Under folding conditions, I was not expected to unfold during the duration of the labeling pulse. If it did, the population of the least protected state, C would be overestimated and that of the more protected state, I would be underestimated. Indeed, the time constant of unfolding of I ( $1/k_{IC}$ ) was found to be almost 20-fold slower than the pulse duration (Table 1) which suggested that no significant unfolding of I occurred during the pulse duration. This was also confirmed by applying a longer labeling pulse of 10 s duration at pH 9. It was observed that the fractional populations of C and I were not significantly different from when the pulse duration was 5 s (data not shown), confirming that I did not undergo significant unfolding during the duration of the labeling pulse.

### 2.4 | Refolding kinetics of dcMN in 0.4 M GdnDCI monitored using a 5 s HX labeling pulse at pH 7

Refolding was also monitored by the application of a HX labeling pulse at pH 7 to probe the stability of the intermediate. Figure S5a,b shows the mass distributions obtained upon applying the HX labeling pulses at different times of refolding, for Chains B and A, respectively. The mass profiles could be fit to the sum of three Gaussian mass distributions (Figure S5c–j). The deconvoluted mass distributions of C, I, and N at different times of refolding are shown in Figure S5c–f for Chain B and Figure S5g–j for Chain A. Again, the least protected conformation was C having the lowest centroid  $m/z$ , and the most protected conformation was N having the highest centroid  $m/z$ . The mass distribution having the intermediate  $m/z$  arose from the pulse labeling of I. The fractional populations corresponding to C, I, and N, for Chain B (Figure S5k) and Chain A (Figure S5l) were



SCHEME 1 Mechanism of refolding.

obtained and plotted against the time of refolding. The mass distribution was globally fit to a triangular three-state mechanism (Scheme 1) and the global fit was found to be in good agreement with the data (Figures S5k,l). The kinetic parameters obtained from the global fit are listed in Table S2. The simulated mass distributions according to a triangular folding mechanism using these kinetic parameters also agreed well with the experimental data (Figure S6).

The mechanism of HX into intact dcMN is known to be in the EX1 regime in the pH range of 7–9.<sup>59</sup> It was therefore very unlikely that the HX into the protected amide sites of I occurred by the EX2 mechanism in this pH range. The closing rate constant should be much faster than the intrinsic rate constant of HX in case of the exchange occurring in EX2 regime (see SI text). The intrinsic rate constant of HX is 100-fold slower at pH 7 than at pH 9, yet it is around 1700 times faster than the refolding (closing) rate constant of the fast phase. Hence, HX into I must occur in the EX1 regime. The number of deuteriums protected from exchange in I was 32 (16 in Chain B and 16 in Chain A) for a pH 7 pulse, and only 19 (7 in Chain B and 12 in Chain A) when the labeling pulse was at pH 9. If HX into I was indeed occurring in the EX1 regime, then the observation that a larger number of deuteriums were protected in I at pH 7 than at pH 9, would mean that the structure opening rate constant is significantly slower at pH 7 than at pH 9. This would mean that part of the protective structure in I is kinetically more unstable at pH 9 than at pH 7. The N state also retained a greater number of deuteriums at pH 7 (46 deuteriums protected) than at pH 9 (34 deuteriums protected). A similar observation had been made in a previous study, in which the N state was seen to retain fewer deuteriums at pH 9 than at pH 7.<sup>59</sup> Indeed, it is known that dcMN is less stable at higher pH.<sup>67</sup>

### 2.5 | Peptide map of dcMN by ETD

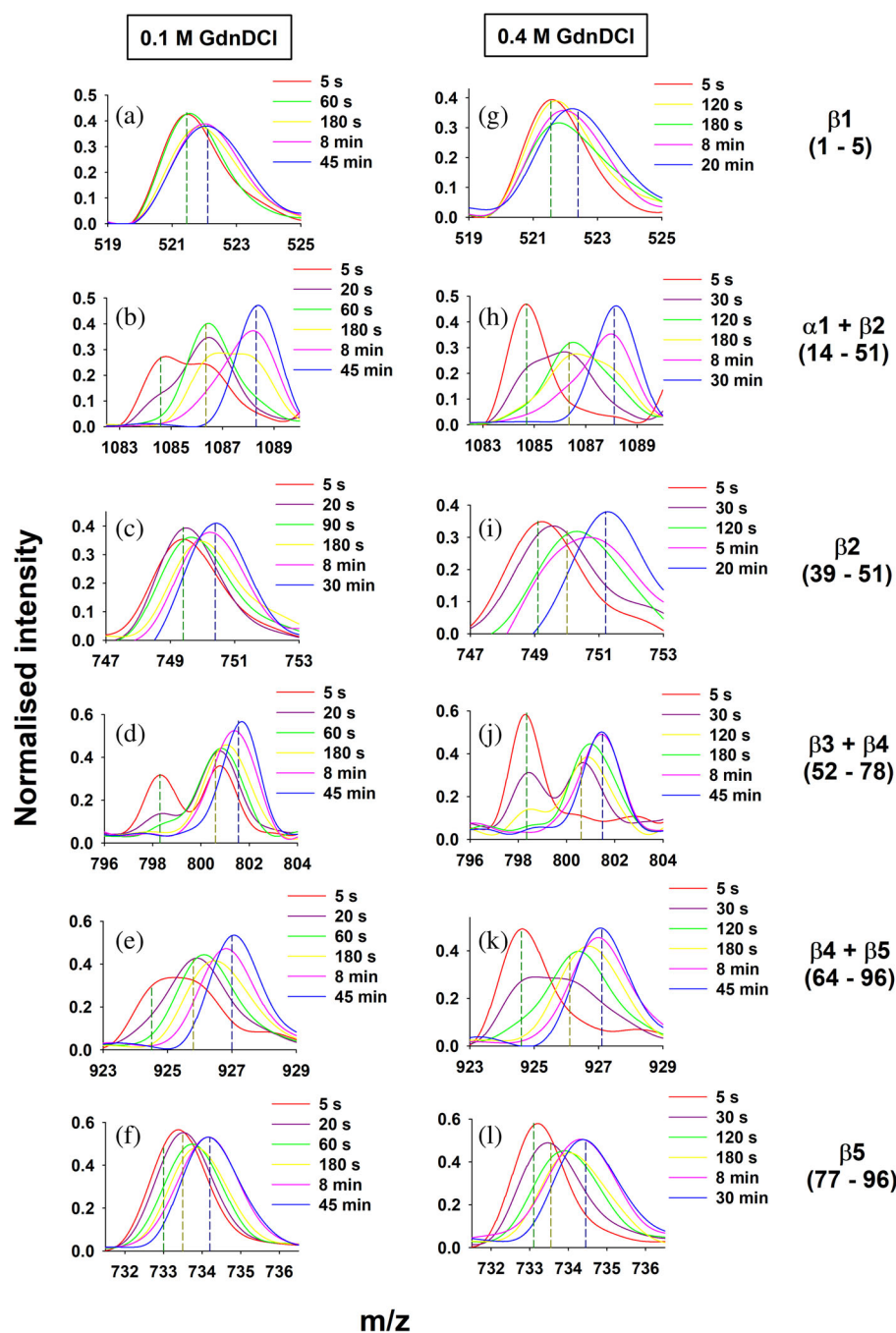
It was important to characterize the structure of the intermediate formed during refolding in 0.1 and 0.4 M

GdnDCI by determining the sequence segments that were structured (protected) and unstructured (unprotected). Hence, electron transfer dissociation (ETD) of intact dcMN was carried out subsequent to HX, to generate c ions (N-terminal fragments) and z ions (C-terminal fragments). These ions are generated by the breakage of the N-C $\alpha$  bond.<sup>72</sup> Reliably identified c and z ions were mapped on to the sequence of the protein in order to generate a peptide map (Figure S7). These different c and z ions were used to delineate the structural events of different sequence segments during refolding of the native protein, and the number of deuteriums protected in each of

the sequence segment was calculated according to Table S3.

## 2.6 | Refolding kinetics of different sequence segments in 0.1 and 0.4 M GdnDCI

Since significantly more deuteriums were protected in the I and N conformations of both Chains B and A (Tables 1 and S2), when labeled with a pH 7 pulse than with a pH 9 pulse, it would have been desirable to carry



**FIGURE 5** Kinetics of structure formation in different sequence segments of double chain monellin (dcMN) during refolding in 0.1 and 0.4 M GdnDCI at pH 7, 25°C, monitored using a 5 s hydrogen exchange (HX) labeling pulse at pH 9. Peptide fragments were generated by electron transfer dissociation (ETD) after the HX had occurred. The mass profiles of different peptide fragments at different time of refolding are shown. The sequence segments corresponding to the mass profiles are indicated beside the panels. In all the panels, the dashed vertical green, yellow, and blue lines represent the centroid  $m/z$  values of the C, I, and N-like conformations of the corresponding sequence segments.



out pulsed HX-labeling at pH 7 than at pH 9, before fragmenting the protein to determine the sequence segments getting labeled. Unfortunately, however, C and I differed in mass ( $m/z$ ) by more than 1  $m/z$  when pulse-labeled at pH 7 than at pH 9 (Figures 2, 3 and S5), and the quadrupole of the mass spectrometer allows only a  $\pm 1$   $m/z$  range at a given  $m/z$ . It was therefore not possible to select a  $m/z$  value that allowed both C and I through for ETD fragments arising from C to be quantified with adequate sensitivity. Hence, the kinetics of the change in the populations of fragments arising from C could not be obtained. On the other hand, the centroid  $m/z$  values of the mass spectra corresponding to C, I, and N were near to each other when pulsed HX-labeling was carried out at pH 9 (Figures 2 and 3), which allowed the entire kinetics of change in population to be monitored reliably for all the sequence segments of the protein.

Figure 5 compares the mass profiles of different sequence segments obtained at different times of refolding in 0.1 and 0.4 M GdnDCl. Sequence segments 1–5 ( $\beta_1$ ), 14–51 ( $\alpha_1$  and  $\beta_2$ ), 39–51 ( $\beta_2$ ), 52–78 ( $\beta_3$  and  $\beta_4$ ), 64–96 ( $\beta_4$  and  $\beta_5$ ) and 77–96 ( $\beta_5$ ) were represented by c4 (+1 charge state), z37 (+4 charge state), z12 (+2 charge state), c26 (+4 charge state), z32 (+4 charge state), and z19 (+3 charge state) ions, respectively (Figure S7). The mass profiles were found to fit well to the sum of either two or three Gaussian mass distributions (Figures S8 and S9).

In both 0.1 and 0.4 M GdnHCl, the mass profiles for sequence segment 1–5 ( $\beta_1$ ), shown in Figure 5a,g, fit well to the sum of two Gaussian mass distributions, where the least protected state corresponded to C and the most protected state corresponded to N (Figures S8a–c and S9a–c). The observation of only a 1 Da change in mass between the two conformations demanded that a two-state fit to be used. The sequence segment was unstructured during the C  $\leftrightarrow$  I transition, and acquired protection only during the I  $\leftrightarrow$  N step. This indicated that  $\beta_1$  was completely bereft of protective structure in the intermediate populated in both 0.1 and 0.4 M GdnDCl.

The mass profiles of sequence segments, 14–51 ( $\alpha_1$  and  $\beta_2$ ), 52–78 ( $\beta_3$  and  $\beta_4$ ), 64–96 ( $\beta_4$  and  $\beta_5$ ), and 77–96 ( $\beta_5$ ) after pulse labeling at different times of refolding in 0.1 and 0.4 M GdnDCl are also shown in Figure 5. The mass profiles of these sequence segments fit well only to the sum of three Gaussian mass distributions (Figures S8 and S9), as in the case of the intact protein. This suggested that these sequence segments acquired protection during both the C  $\leftrightarrow$  I and I  $\leftrightarrow$  N steps and were partially protected in the intermediate. The centroid  $m/z$  and peak width of the deconvoluted mass distributions were allowed to vary by only  $\pm 0.3$  and  $\pm 0.2$ , respectively. No peptide covering the sequence segment 6–13 (loop

between  $\beta_1$  and  $\alpha_1$  and N-terminal  $\alpha$ -helix) could be obtained.

The mass profiles of sequence segment 39–51 ( $\beta_2$ ), after pulse labeling at different times of refolding in 0.1 and 0.4 M GdnDCl, are shown in Figure 5c,i, respectively. The mass profiles obtained in 0.1 M GdnDCl fit well to the sum of two Gaussian mass distributions, where the least protected state corresponded to C and the most protected state corresponded to N (Figures S8g–i). The sequence segment was unstructured during the C  $\leftrightarrow$  I transition and acquired protection only during the I  $\leftrightarrow$  N step. This indicated that  $\beta_2$  was completely bereft of protective structure in the intermediate populated in 0.1 M GdnDCl. On the other hand, the mass profiles of this sequence segment obtained in 0.4 M GdnDCl could be fit well only to the sum of three Gaussian mass distributions (Figures S9g–i). This indicated that the sequence segment acquired protection during both the C  $\leftrightarrow$  I and I  $\leftrightarrow$  N steps and was partially protected in the intermediate formed in 0.4 M GdnDCl. The number of deuteriums that underwent protection in each phase of refolding, was also similar for both intact Chains B and A and their constituent sequence segments (Table 2).

The sequence segment 39–51 ( $\beta_2$ ) was completely devoid of protective structure in the intermediate populated in 0.1 M GdnDCl, while it was partially structured in 0.4 M GdnDCl (see above). Surprisingly, the sequence segment 14–51 had the same number of deuteriums protected during the fast C  $\leftrightarrow$  I step during refolding in both 0.1 and 0.4 M GdnDCl (Table 2). Since sequence segment 39–51 ( $\beta_2$ ) was unfolded in I in 0.1 M GdnDCl, the bigger sequence segment 14–51 should also have retained fewer deuteriums in I than it did in 0.4 M GdnDCl. But it was observed that the sequence segment 14–38 retained more deuteriums in 0.1 M GdnDCl than in 0.4 M GdnDCl. Thus, the total number of deuteriums protected in sequence segment 14–51 did not change, but what changed were the numbers of deuteriums undergoing protection in the two regions (14–38 and 39–51) of this large sequence segment 14–51. It was possible that the sequence segments were differentially stabilized in these two different solvent conditions.

The fractional populations of C, I, and N-like conformations for all the sequence segments were obtained in the same way as was done for the intact protein, and were analyzed according to the same three-state triangular folding mechanism (Figures 6 and 7) to obtain the kinetic parameters listed in Table 3. The rate constants used to fit the intact protein data and the fragment data were found to be in good agreement with each other, suggesting that the different sequence segments also underwent structural transitions during the fast and the slow phases of refolding.

TABLE 2 Number of deuteriums that get protected during each phase of refolding in 0.1 and 0.4 M GdnDCl at pH 7, 25°C<sup>a</sup>

Sequence segments	Deuteriums protected					
	0.1 M GdnDCl			0.4 M GdnDCl		
	Native protein control	Fast phase of refolding (C ↔ I <sub>1</sub> )	Slow phase of refolding (I <sub>1</sub> ↔ N)	Native protein control	Fast phase of refolding (C ↔ I <sub>2</sub> )	Slow phase of refolding (I <sub>2</sub> ↔ N)
Intact chain B (1–51)	16 ± 0.02	7 ± 0.2	9 ± 0.01	16 ± 0.7	7 ± 0.1	9 ± 0.7
1–5 (β1)	0.6 ± 0.06	0	0.63 ± 0.03	0.8 ± 0.1	0	1 ± 0.05
6–13 (loop + N-terminal of α-helix)	1.1 ± 0.2	0	1.3 ± 0.13	1 ± 0.4	0.25 ± 0.12	1.3 ± 0.2
14–38 (α-helix + loop)	12 ± 0.3	7.2 ± 0.35	5 ± 0.01	10.3 ± 0.7	5.2 ± 0.01	5 ± 0.3
39–51 (β2)	2.6 ± 0.01	0	2.1 ± 0.2	3.9 ± 0.5	1.5 ± 0.12	2 ± 0.2
Intact chain A (52–96)	18.5 ± 0.1	12 ± 0.2	6 ± 0.01	18 ± 0.1	12 ± 0.3	6 ± 0.3
52–63 (β3)	7.1 ± 0.3	5.6 ± 0.4	1.5 ± 0.5	7 ± 0.2	5 ± 0.2	2.3 ± 0.12
64–76 (β4)	6.3 ± 0.4	3.8 ± 0.12	2.2 ± 0.2	6.2 ± 0.2	4.2 ± 0.13	1.5 ± 0.12
77–96 (β5)	3.6 ± 0.2	2 ± 0.3	2 ± 0.01	4.2 ± 0.2	1.4 ± 0.2	2.8 ± 0.12

<sup>a</sup>The numbers in the white and gray colored rows were obtained for HX into Chain B and Chain A of dcMN, respectively.

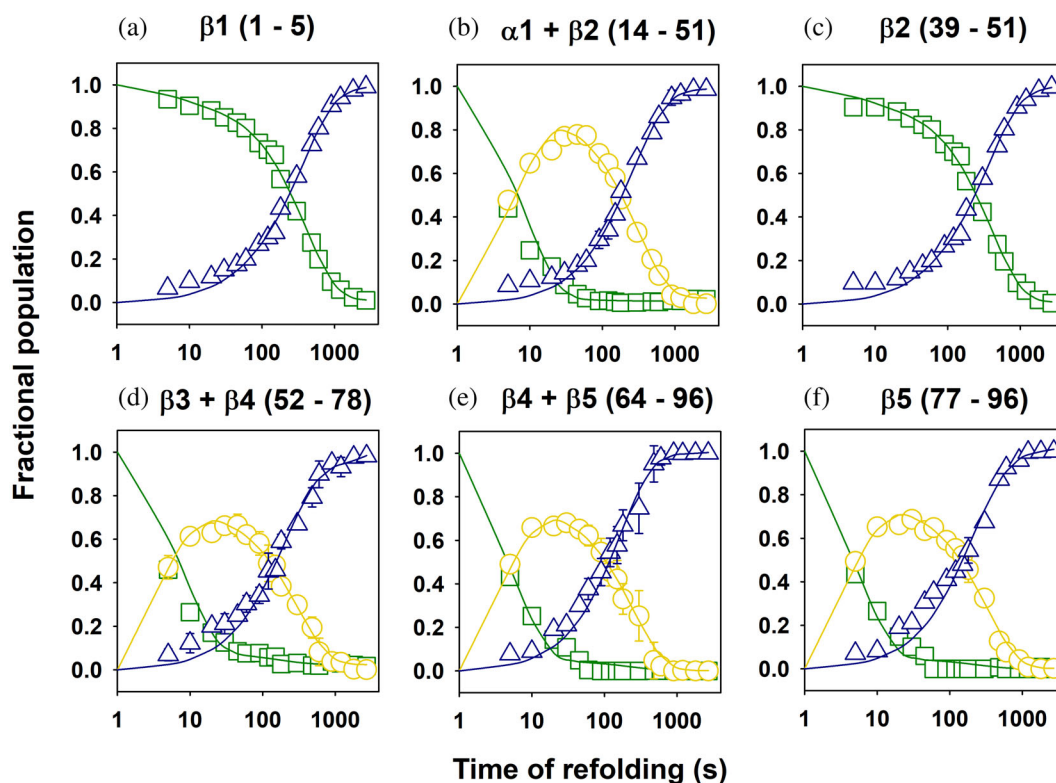
### 3 | DISCUSSION

In a previous study of the folding of dcMN using intrinsic Trp fluorescence as the probe, it was shown that refolding from U commenced by the fast association of the Chains B and A to form an initial encounter complex C.<sup>48</sup> Further folding was shown to occur via kinetic partitioning along two pathways, on one of which an intermediate, I, is populated, while the other pathway is a direct pathway from C to N. The C ↔ I ↔ N pathway was shown to have a more compact transition state at low GdnHCl than at high GdnHCl concentrations.<sup>48</sup> This indicated that I differed in its compactness at low and high GdnHCl concentrations. It appeared therefore that the C ↔ I ↔ N pathway at low GdnHCl concentration is distinct from that at high GdnHCl concentration, but the previous study could not delineate the difference in the structure of I and consequently the difference in the sequence of structural events defining the pathways operating at low and high GdnHCl concentrations. The current study has utilized pulsed HX labeling and mass spectrometry to characterize the progress in structure acquisition on the C ↔ I ↔ N pathway at low (0.1 M) as well as at high (0.4 M) GdnHCl concentration, where the pathways are known to be different.<sup>48</sup> The two intermediates, which accumulate on the two parallel pathways have been distinguished on the basis of how their structures differ in the extents to which they afford protection against HX.

#### 3.1 | Mechanism of folding of dcMN

In this study, the changes in the populations of three conformations, C, I, and N, were directly monitored and quantified during refolding. The observation that the kinetics of the change in the population of I is the same whether monitored for Chain B or for Chain A indicate that I is formed from the encounter complex C and not from the free chains present in U (Figure 4). Furthermore, the observation that the rate constants of formation of I and N are not dependent on protein concentration (data not shown), supports I being the product of the folding of C and not U, as in the latter case a dependence on protein concentration would have been observed for both the rate constants.

The minimal model for folding, in both 0.1 and 0.4 M GdnHCl, which adequately describes the kinetic data, is a triangular three-state mechanism with an indirect C ↔ I ↔ N pathway on which I is found to accumulate, and a direct C ↔ N pathway (see Section 2). Calculations of the flux of folding on each pathway indicate that in 0.1 M GdnDCl, 97% of the molecules fold via the C ↔ I ↔ N pathway, and 3% via the C ↔ N pathway. In 0.4 M GdnHCl, 91% of the molecules fold via the C ↔ I ↔ N pathway, and 9% via the C ↔ N pathway. Hence, the C ↔ I ↔ N pathway is the major folding pathway. Triangular three-state mechanisms have also been used to describe the folding of other proteins,<sup>13,24</sup> but this is the first instance where detailed structural



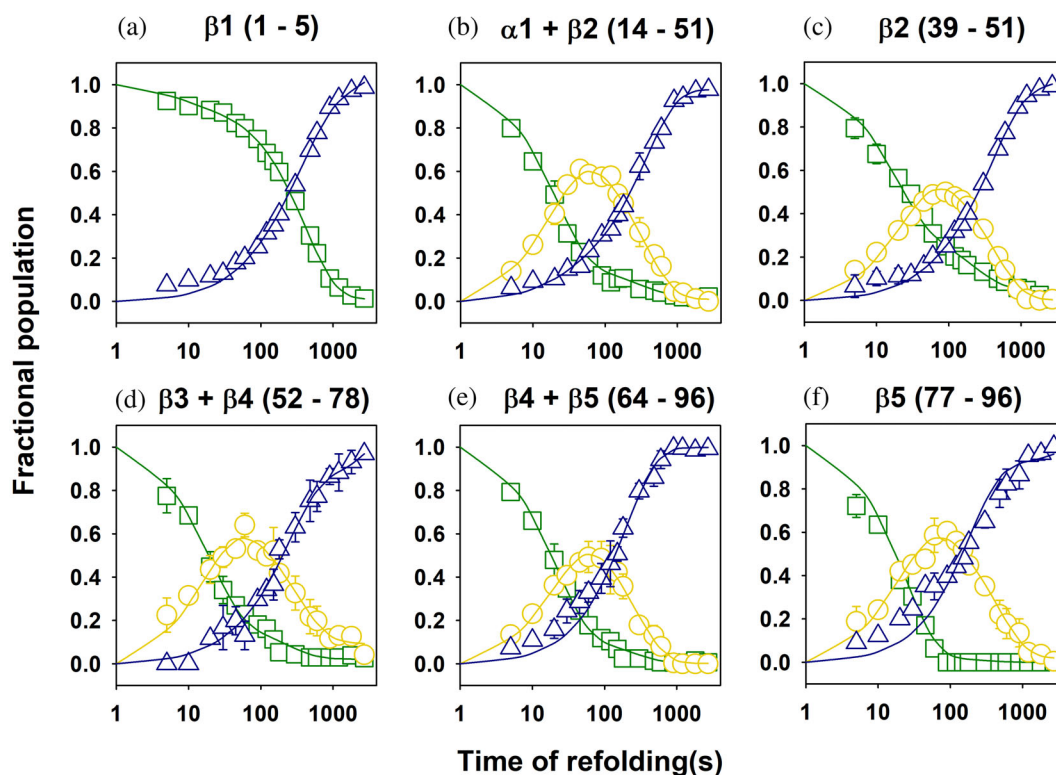
**FIGURE 6** Kinetics of structure formation in different sequence segments of double chain monellin (dcMN) during refolding in 0.1 M GdnDCI at pH 7, 25°C. In all the panels, the green squares, yellow circles, and blue triangles represent the fractional changes in the populations of the C, I, and N-like conformations of different sequence segments at various refolding times. The fractional populations were determined by dividing the area of the deconvoluted mass distribution corresponding to the sequence segment in C, I, or N (Figure S8) by the total area under the mass profile at a given time point. The solid lines (same color codes as for the symbols) were obtained from the global fit of the data to a three-state triangular folding mechanism. The rate constants obtained from the fits are listed in Table 3. The error bars represent the standard deviations in the data obtained from three independent experiments.

characterization of the intermediate has been carried out. Importantly, it is shown that only the population of I changes during the course of folding, while its protective structure remains the same during the folding process (Figures 2 and 3).

In most protein folding reactions, intermediates are seen to form very rapidly, usually forming too fast to measure.<sup>12,19,24,29,39,69,73–78</sup> Consequently, U and I are effectively at equilibrium before further folding occurs, and it becomes very difficult to establish whether I is on the direct pathway from U to N or whether it is a nonproductive side-product that has to unfold back to U for folding to proceed to N.<sup>26–28</sup> In the current study, I is seen to form slowly enough for the rate constants of its formation and disappearance to be precisely determined. The individual rate constants for the formation of N, from I on one pathway and from C on the other pathway (Scheme 1), are about the same (Figures 2 and 3; Table 1) and hence, equal to the net rate constant for the formation of N. The observation that the rate constant of the disappearance of I is the same as the rate constant of the

formation of N, is very strongly supportive of I being a productive on-pathway intermediate (Table 1). As expected for such an intermediate, the structure of I (Figures 2 and 3, Table 2) is intermediate between that of C and N and is native-like. It should be noted that it is because the pulsed HX methodology enables not only the structural identification of I but also the quantification of the extent to which it is populated, it is possible to firmly establish that I is a productive on-pathway intermediate.

The triangular three-state mechanism is based on C being a single kinetically homogenous population of molecules, in which the transitions between the molecules are very rapid relative to the folding reactions of C. Folding molecules kinetically partition from C along the direct and indirect pathways, in accordance with the rate constants defining the two pathways. It should be noted a mechanism with two sub-populations in the C ensemble in slow equilibrium with each other would also account for the data, but it is disregarded because it would be more complex. Nevertheless, it is necessary to consider heterogeneity in the C ensemble, because of the



**FIGURE 7** Kinetics of structure formation in different sequence segments of dcMN during refolding in 0.4 M GdnDCl at pH 7, 25°C. In all the panels, green squares, yellow circles, and blue triangles represent the fractional changes in the populations of the C, I, and N-like conformations of different sequence segments at various refolding times. The fractional populations were determined by dividing the area of the deconvoluted mass distribution corresponding to the sequence segment in C, I, or N (Figure S9) by the total area under the mass profile at a given time point. The solid lines (same color codes as for the symbols) were obtained from the global fit of the data to a three-state triangular folding mechanism. The rate constants obtained from the fits are listed in Table 3. The error bars represent the standard deviations in the data obtained from three separate experiments.

**TABLE 3** Rate constants obtained from fitting of the fragments data to a three-state triangular folding mechanism

Rate constants ( $s^{-1}$ )	0.1 M GdnDCl	0.4 M GdnDCl
$k_{CI}$	$0.13 \pm 0.017$	$0.032 \pm 0.006$
$k_{IN}$	$(3.4 \pm 0.7) \times 10^{-3}$	$(3.3 \pm 0.9) \times 10^{-3}$
$k_{CN}$	$(3.0 \pm 1.4) \times 10^{-3}$	$(2.9 \pm 1.7) \times 10^{-3}$
$k_{IC}$	$(2.1 \pm 1.9) \times 10^{-2}$	$(1.0 \pm 0.3) \times 10^{-2}$
$k_{NI}$	$(3.1 \pm 2.6) \times 10^{-4}$	$(3.0 \pm 3.0) \times 10^{-4}$
$k_{NC}$	$(3.3 \pm 3.1) \times 10^{-5}$	$(8.3 \pm 1.5) \times 10^{-5}$

possibility that *cis* to *trans* proline isomerization may lead to it having sub-populations of molecules.

Pro41 and Pro92 are known to be *cis* in dcMN, and it is likely that U as well as C have at least two-subpopulations that differ in possessing *cis* versus *trans* X-Pro bonds. Indeed, the rate constants of the folding of I to N on the indirect pathway and of C to N on the direct

folding pathway are similar to the rate constant expected for a *cis* to *trans* proline isomerization reaction,<sup>20,42,79–82</sup> and are independent of GdnHCl concentration in the range of 0.1–0.4 M, as expected for a proline isomerization reaction.<sup>83–87</sup> It should be noted that for other proteins, protein molecules with both the correct *cis* Pro isomer and with the incorrect *trans* Pro isomer can form structure at the same rate.<sup>88,89</sup>

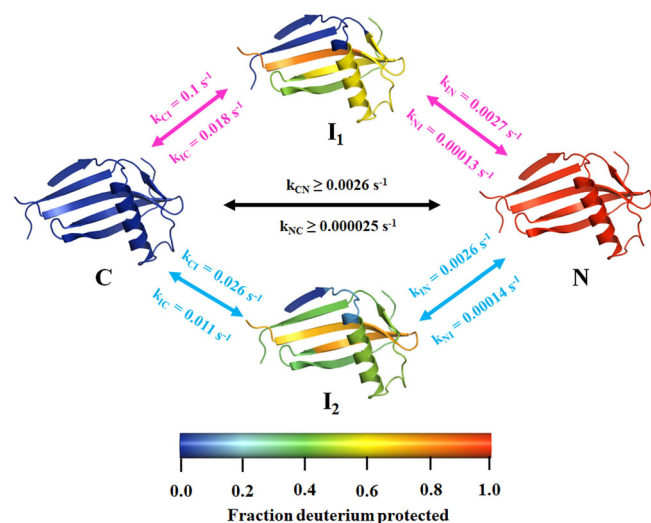
### 3.2 | Step-wise and slow assembly of secondary structural elements

A defining feature of the assembly of secondary structure on the  $C \leftrightarrow I \leftrightarrow N$  pathway is that the  $\alpha$ -helix and four of the five  $\beta$ -strands do not gain all their protective structure in one folding step. Only  $\beta 1$  gains its protective structure in one step, during the folding of I to N (Figure 5a,g). Protective structure forms to different extents in the other secondary structures in each of the two steps.  $\beta 3$  and  $\beta 4$  form most of their protective structure during the C to I



step, while the  $\alpha$ -helix as well as  $\beta$ 2 and  $\beta$ 5 form about equal amounts of secondary structure in both steps (Figure 8 and Table 2). It is likely that differentially stabilizing tertiary interactions are responsible for the nonuniform protection against HX of amide sites on the same  $\beta$ -strand<sup>30</sup> as well on different  $\beta$ -strands.<sup>90–92</sup> The observation that protection against HX develops in steps implies that tertiary interactions form in a step-wise manner. Native-state HX-MS<sup>60</sup> and SX-MS<sup>43,66</sup> studies of the single chain variant of monellin had also indicated that the secondary structural units assemble in stages. These studies suggested that the helix in monellin consolidates its structure in stages because a specific tertiary packing interaction has to form to stabilize each helical segment.

In a previous pulsed HX labeling MS study of the folding of the PI3K SH3 domain, the  $\beta$ -sheet was also observed to assemble in a step-wise manner.<sup>63</sup> The assembly of a  $\beta$ -sheet has been shown to occur in steps for other proteins too, including hFGF-1,<sup>92</sup> CBTX,<sup>93</sup> and RNase H,<sup>94</sup> with different  $\beta$ -strands appearing to each form fully in different steps. In the case of RNase A<sup>95</sup> and ubiquitin,<sup>12</sup> the  $\beta$ -sheet appears to form cooperatively, and it is possible that this cooperativity may be a



**FIGURE 8** Mechanism of refolding of double chain monellin (dcMN). The fraction of deuterium protected, which represents the fractional structural change that has occurred, has been mapped on to the structure of native dcMN for the C state, the N state and the intermediate states. Refolding occurs via a triangular mechanism populating the intermediate  $I_1$  in 0.1 M GdnDCl (pathway shown with pink arrows), and shifts to an alternate pathway in 0.4 M GdnDCl where the intermediate  $I_2$  is populated (pathway shown with cyan arrows). A direct  $C \leftrightarrow N$  pathway, shown with the black arrow, operates in both the solvent conditions. The color bar at the bottom indicates the fraction of deuterium protected in the different secondary structural elements. The structures of dcMN were drawn using the Pymol software and the PDB ID 3MON.

consequence of folding beginning from a collapsed unfolded state. In the case of RNase A,<sup>95</sup> the  $\beta$ -sheet is formed early but is unstable when it first forms and becomes stable only later as tertiary interactions develop during folding. The stabilizing effect of tertiary interactions on  $\beta$ -sheet structure is especially notable in the case of the  $\beta$ -domain of staphylococcal nuclease.<sup>96</sup>

There does not appear to be any correlation between the  $\beta$ -sheet propensity of the residues in a  $\beta$ -strand, and when the  $\beta$ -strand starts to form. The residues in  $\beta$ 3,  $\beta$ 4, and  $\beta$ 5 all have high  $\beta$ -sheet propensities (Figure S10), and these  $\beta$ -strands start acquiring protection against HX during the fast phase of refolding. In contrast, the residues of  $\beta$ 1 have very high  $\beta$ -sheet propensities (Figure S10), but  $\beta$ 1 starts acquiring protection against HX only during the slow phase of folding (Figures 5–7). It cannot, however be ruled out that  $\beta$ 1 also forms during the fast phase, but it acquires protection only during the slow phase of refolding.

The formation of the  $\alpha$ -helix and the  $\beta$ -sheet appears to be remarkably slow during the folding of dcMN (Figures 6 and 7). For many proteins, secondary structure has been observed to form very rapidly.<sup>11,97–101</sup> Even in the case of the single-chain variant of monellin, secondary structure has been shown to form in the millisecond time domain,<sup>68–69,71</sup> in contrast to the tens of second time domain observed here in the case of dcMN. It is, however, possible, that labile secondary structures form more rapidly during the folding of dcMN but become protective against HX only later, when tight packing occurs as a consequence of significant tertiary structure formation. It should be noted that the opening of secondary structure to HX during the 5 s labeling pulse would slow down only when the structure is stabilized by packing and other tertiary interactions.

The surmise that the secondary structural units form early but are kinetically unstable when they first form is supported by the observation that the intrinsic fluorescence of Trp4 is the same in both I and N, which is possible only if the chemical environment of Trp4 is very similar in I and N. Trp4 is present in  $\beta$ 1, which has not acquired any protective structure in I. The side chain of Trp4 is about 60% buried in N, interacting with Met43 and Lys45 in  $\beta$ 2, and with Gln61 in  $\beta$ 3. Protective structure in  $\beta$ 2 is absent in I in 0.1 M GdnHCl. It is difficult to envisage that the native interactions of the side-chains of Trp4 with the side chains of the other residues would have formed in I if the segments of the main chain with the interacting residues are not in their native  $\beta$ -strand conformation. It is therefore likely that the main chain has at least partially adopted  $\beta$ -strand structure, albeit unstable, pertaining to  $\beta$ 1,  $\beta$ 2 and  $\beta$ 3 in I, which stabilizes and becomes protective against HX later during folding.

In this context, it should be noted that in the case of several proteins, including interleukin-1 $\beta$ <sup>90</sup> and dihydrofolate reductase (DHFR),<sup>91</sup>  $\beta$ -sheet structure appears to form early as detected by circular dichroism but affords no discernible protection against HX. The  $\beta$ -sheet structure becomes stable only later during folding.

### 3.3 | The C $\leftrightarrow$ I $\leftrightarrow$ N pathways are different and structurally distinct in 0.1 and 0.4 M GdnHCl

Remarkably, the structure of the intermediate I is different in 0.1 and 0.4 M GdnHCl, as reflected in the distribution of protected amide hydrogens that afford protections against HX. It is assumed that the degree of protective structure corresponds to the extent of secondary structure formation. The principal differences in structure appear to be in the  $\alpha$ -helix and the loop connecting the  $\alpha$ -helix to  $\beta$ 2 and  $\beta$ 2 (Table 2 and Figure 8).  $\beta$ 3 and  $\beta$ 4 have also formed to different extents in I in 0.1 M GdnHCl compared to in 0.4 M GdnHCl (Table 2 and Figure 8). The  $\alpha$ -helix appears to be more structured in I formed in 0.1 M than in 0.4 M GdnHCl, suggesting that the interactions stabilizing it are sensitive to denaturant concentration. Of course, for many proteins, it has been observed that  $\alpha$ -helix formation starting from a collapsed globule occurs faster at low denaturant concentration as expected.<sup>39,53,98</sup>

It is intriguing that  $\beta$ 2 is not protected against HX in I formed in 0.1 M GdnHCl, but it is protected in I formed in 0.4 M GdnHCl (Figure 8). Since residues in  $\beta$ 2 interact with the side chain of Trp4 in N (see above), the observation that the fluorescence of Trp4 is slightly higher in 0.4 M than in 0.1 M GdnHCl (Figure S4), is consistent with  $\beta$ 2 being more structured in 0.4 M than in 0.1 M GdnHCl (Table 2). It should be noted that the residues of  $\beta$ 2 have a very low  $\beta$ -sheet propensity (Figure S10).  $\beta$ 2 (Asn36–Tyr48) is split into two segments by the Arg40–Pro41 stretch (Figure 1), and the current study is also unable to determine in which segment the protective structure is present.

In an earlier native-state HX-MS study of dcMN, unfolding in the absence of GdnHCl, leading to the loss of protective structure, was seen to occur in four steps involving three high-energy intermediates that were very sparsely populated.<sup>59</sup> In this study, refolding was carried out under strongly stabilizing conditions, in the presence of 0.1 M GdnHCl. It was expected that the same sequence of structural events would be observed as in the native-state HX experiments, but in the reverse order. The high-energy intermediates that had been observed in the

native-state HX study are populated too sparsely to be detected in the current pulse labeling HX study. Instead, a stable very significantly populated intermediate has been observed. At present, it is not clear why this intermediate was not observed in the earlier native-state HX study. It would appear that the manner in which Chains B and A have collapsed together to form the encounter complex C from U in high GdnHCl is responsible for the formation of the stable I observed in the current study. Nonetheless, two important features of the folding reaction of dcMN are seen both in the earlier native-state HX study and in the current pulsed HX labeling study. In both studies, secondary structural units are seen to form in steps, and  $\beta$ 1 forms late during folding.

## 4 | CONCLUSIONS

The present HX-MS pulse labeling study structurally characterizes the refolding of dcMN in different GdnHCl concentrations. The mechanism was shown to be the same (three-state triangular mechanism) in both lower (0.1 M) and higher (0.4 M) GdnHCl concentrations, but the structure of the intermediate was different at the two GdnHCl concentrations. The intermediates populated in both the solvent conditions were shown to be productive. The temporal order of progress of structure formation in the different secondary structural elements was also resolved in this study. It is shown that the  $\beta$ -strands acquire structure faster than the sole  $\alpha$ -helix although they possess different propensity to be in a  $\beta$ -strand. There is one striking difference in the sequence of structural events on the two alternative pathways operating in the different solvent conditions.  $\beta$ 2 has no protective structure in the intermediate on one pathway and attains all its protective structure only when the intermediate folds to N, while it attains significant protective structure early during the formation of the intermediate on the alternative pathway.

## 5 | MATERIALS AND METHODS

### 5.1 | Protein expression and purification

Double-chain monellin was expressed and purified as reported previously.<sup>102</sup> The mass and purity of the protein were verified by electrospray ionization mass spectrometry (ESI-MS). The protein was found to be >95% pure. Protein concentration was determined by measuring the absorbance at 280 nm, using an extinction coefficient value of 14,600 M<sup>-1</sup> cm<sup>-1</sup>.

## 5.2 | Buffers and reagents

All the chemicals used in the current study were of high purity grade, and were purchased from Sigma. Ultra-pure GdnHCl was obtained from United States Biochemicals and was of the highest purity grade. All the experiments were carried out at 25°C. Sodium phosphate buffer (50 mM) and Glycine-NaOH buffer (50 mM), were used as the labeling buffer at pH 7 and pH 9, respectively. Ice-cold 100 mM glycine-HCl buffer containing 8 M GdnHCl was used as the quench buffer to stop the labeling reaction. GdnHCl was added to the quench buffer to unfold the protein in order to achieve higher ion count upon ETD fragmentation. The concentration of GdnHCl in each buffer was determined by measuring the refractive index using an Abbe refractometer. GdnHCl was deuterated by carrying out three cycles of HX in D<sub>2</sub>O followed by lyophilization. For all D<sub>2</sub>O buffers, the reported pH values were not corrected for any isotope effect.

## 5.3 | Fluorescence monitored equilibrium and kinetic studies

GdnHCl-induced equilibrium unfolding and kinetic refolding transitions were monitored using a stopped-flow model (SFM4) attached to the MOS-450 optical system from Biologic, as described previously.<sup>67</sup> The excitation wavelength was 280 nm, and the fluorescence emission was collected at 340 nm using a 10 nm band-pass filter (Asahi Spectra). The final protein concentration was 40 μM.

## 5.4 | Refolding kinetics monitored by pulse-labeling HX

The protein was deuterated by unfolding it in the deuterated unfolding buffer containing 50 mM sodium phosphate and 4 M GdnDCl for at least 12 h. dcMN contains a Cys residue at position 42 (in Chain B). To prevent interchain dimerization by disulfide linkage formation, DTT (prepared in D<sub>2</sub>O) was added to the unfolded protein to a final concentration of 2 mM. Refolding was initiated by diluting the deuterated unfolded protein in refolding buffer (50 mM sodium phosphate buffer prepared in D<sub>2</sub>O, pH 7) to a final GdnDCl concentration of 0.1 M or 0.4 M. In all the cases, the final protein concentration was 40 μM. At different refolding times (from 5 s to 45 min), a 5 s HX labeling pulse was given by 10-fold dilution into 50 mM sodium phosphate buffer (prepared in H<sub>2</sub>O, pH 7) or 50 mM Glycine-NaOH buffer (prepared in H<sub>2</sub>O, pH 9). The HX reaction was quenched by

addition of ice-cold quench buffer (twofold dilution), and the final pH of the solution became 2.8 on ice. The quenched reaction was incubated for 1 min on ice.

## 5.5 | Sample processing for mass spectrometry

After incubation for 1 min on ice, the quenched reaction was desalted using a Sephadex G-25 Hi-trap desalting column from GE, equilibrated with ice-cold distilled water at pH 2.6 (0.1% formic acid was added), using a Postnova AF4 system at a flow rate of 5 ml/min. The desalted sample was then injected within 5–10 s into the HDX module coupled to a nanoAcquity UPLC (from Waters Corporation) and analyzed using a Synapt G2 HD mass spectrometer (Waters Corporation). The protein was first loaded on to a C18 reverse phase trap column using water containing 0.05% formic acid at a flow rate of 100 μl/min for 1 min. The two chains of the protein were eluted from the column using a gradient of 35%–95% acetonitrile (0.1% formic acid) at a flow rate of 40 μl/min in 3 min. The temperature of the entire chromatography assembly was maintained at 4°C inside the HDX module (Waters Corporation) to minimize back exchange during the processing of the sample.<sup>103,104</sup>

For ETD reactions, Chains B and A were first trapped in a trap column with water containing 0.05% formic acid at a flow rate of 100 μl/min for 1.5 min and separated using an analytical column (C18 reverse phase), with a gradient of 35%–95% acetonitrile containing 0.1% formic acid, in 9 min. Separation of the two chains was necessary as only one precursor mass could be allowed to undergo ETD fragmentation at a time inside the mass spectrometer.

## 5.6 | Data acquisition by ESI-MS

The source parameters were set to the following values for the ionization of intact protein: capillary voltage, 3 kV; sample cone voltage, 40 V; extraction cone voltage, 4 V; source temperature, 80°C and desolvation temperature, 200°C. For ETD reactions, 1,4-dicyanobenzene was used as the ETD reagent, and the radical anions were generated from it using a glow discharge current of 50 μA, and a makeup gas (nitrogen) with a flow rate of 35 ml/min was used to obtain reagent ion counts of >10<sup>6</sup> per scan. The +7 charge states for both Chains B and A were used to achieve optimum fragmentation. The other instrument parameters were set to the following values: trap wave velocity, 300 m/s; trap wave height, 0.2 V. Source parameters values were set to those mentioned

above for the intact protein. The transfer collision energy was ramped from 4 eV to 10 eV for Chain A, and from 2 eV to 6 eV for Chain B in order to achieve higher ion counts of the fragment ions. One precursor mass (+7 charge state for both the chains) from each chain was allowed at a time to undergo ETD fragmentation inside the trap cell for ~60 s, and the fragmentation data for the two chains were collected in two channels.

## 5.7 | Data analysis

### 5.7.1 | Analysis of intact protein data

The protein mass spectrum at each timepoint was generated by combining ~40 scans, each 1 s long, from the elution peak of the total ion count (TIC) chromatogram. Each mass spectrum was then processed further by background subtraction and smoothing using the MassLynx version 4.1 software. The +7 charge state peak, from smoothed spectra, which was the highest intensity peak for both Chain A and Chain B, was taken for further analysis. The signal of the +7 charge state peak was normalized by its total area at each timepoint using the Origin software. Then the mass distributions were analyzed by discrete fitting as well as global fitting using MATLAB (see SI methods). The mass distributions were fitted to the sum of three Gaussian equations, where each Gaussian distribution corresponded to a particular conformation C, I, or N, and the width(s), height(s), and centroid(s) of the mass distributions were determined. The centroid  $m/z$  positions and widths of the mass distributions arising from a particular conformation were allowed to vary by only  $\pm 0.1$  and  $\pm 0.05$ , respectively. The fractional area under each Gaussian distribution at a particular timepoint corresponded to the fractional population of the conformation giving rise to that distribution at that timepoint.

The number of deuteriums protected in C, I, and N, for both the chains was calculated by subtracting the protonated chain mass from the mass obtained from the centroid of the mass distribution corresponding to each conformation (C, I, or N).

### 5.7.2 | Analysis of the fragments generated by ETD

The individual c and z ions obtained from both Chains B and A were identified using the BioLynx software, and a peptide map of the protein was made. The mass profiles of the peptides were fitted to the sum of either two or three Gaussian mass distributions. The centroid  $m/z$  and width of the mass distributions of a particular

conformation were allowed to vary by only  $\pm 0.3$  and  $\pm 0.2$ , respectively, across different times of refolding. The fractional area under each Gaussian distribution corresponded to the fractional population of the respective conformation. The fractional population kinetic curves were analyzed using a triangular three-state folding mechanism to obtain the kinetic parameters of refolding.

The number of deuteriums protected in each sequence segments for the U, I, and N states was determined in the same way as for the intact protein. The number of deuteriums protected in different sequence segments were calculated by doing multiple subtractions with consecutive c and z ions. The operations are summarized in Table S3.

## AUTHOR CONTRIBUTIONS

**Rupam Bhattacharjee:** Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); software (lead); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Jayant B. Udgaonkar:** Conceptualization (supporting); funding acquisition (lead); project administration (supporting); resources (lead); supervision (lead); validation (supporting); visualization (supporting); writing – review and editing (equal).

## ACKNOWLEDGMENTS

The authors thank members of our laboratory for discussions, and Nilesh Aghera for help with MATLAB simulation. Jayant B. Udgaonkar is a recipient of a JC Bose National Research Fellowship from the Government of India. This work was funded by the Indian Institute of Science Education and Research, the Tata Institute of Fundamental Research and by the Department of Science and Technology, Government of India.

## CONFLICT OF INTEREST

The authors declare no competing financial interest.

## DATA AVAILABILITY STATEMENT

Data available on request from the authors.

## ORCID

Rupam Bhattacharjee  <https://orcid.org/0000-0001-5481-0381>

Jayant B. Udgaonkar  <https://orcid.org/0000-0002-7005-224X>

## REFERENCES

- Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*. 1991; 30(43):10428–10435.



2. Otzen DE, Itzhaki LS, ElMasry NF, Jackson SE, Fersht AR. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci.* 1994; 91(22):10422–10425.
3. Milla ME, Sauer RT. P22 arc repressor: Folding kinetics of a single-domain, dimeric protein. *Biochemistry.* 1994;33(5): 1125–1133.
4. Huang GS, Oas TG. Structure and stability of monomeric. Lambda. Repressor: NMR evidence for two-state folding. *Biochemistry.* 1995;34(12):3884–3892.
5. Huang GS, Oas TG. Submillisecond folding of monomeric lambda repressor. *Proc Natl Acad Sci.* 1995;92(15):6878–6882.
6. Schindler T, Herrler M, Marahiel MA, Schmid FX. Extremely rapid protein folding in the absence of intermediates. *Nat Struct Biol.* 1995;2(8):663–673.
7. Robinson CR, Sauer RT. Equilibrium stability and sub-millisecond refolding of a designed single-chain arc repressor. *Biochemistry.* 1996;35(44):13878–13884.
8. Kuhlman B, Boice JA, Fairman R, Raleigh DP. Structure and stability of the N-terminal domain of the ribosomal protein L9: Evidence for rapid two-state folding. *Biochemistry.* 1998; 37(4):1025–1032.
9. Perl D, Welker C, Schindler T, et al. Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol.* 1998;5(3): 229–235.
10. Aviram HY, Pirchi M, Barak Y, Riven I, Haran G. Two states or not two states: Single-molecule folding studies of protein L. *J Chem Phys.* 2018;148(12):123303.
11. Radford SE, Dobson CM, Evans PA. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature.* 1992;358(6384):302–307.
12. Briggs MS, Roder H. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci.* 1992;89(6): 2017–2021.
13. Wildegger G, Kiefhaber T. Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol.* 1997;270(2):294–304.
14. Rami BR, Udgaonkar JB. Mechanism of formation of a productive molten globule form of barstar. *Biochemistry.* 2002; 41(6):1710–1716.
15. Sridevi K, Lakshmikanth G, Krishnamoorthy G, Udgaonkar JB. Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble. *J Mol Biol.* 2004;337(3):699–711.
16. Went HM, Benitez-Cardoza CG, Jackson SE. Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett.* 2004;567(2–3):333–338.
17. Vu N-D, Feng H, Bai Y. The folding pathway of barnase: The rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry.* 2004;43(12):3346–3356.
18. Laurents DV, Bruix M, Jamin M, Baldwin RL. A pulse-chase-competition experiment to determine if a folding intermediate is on or off-pathway: Application to ribonuclease a. *J Mol Biol.* 1998;283(3):669–678.
19. Park S-H, Shastry M, Roder H. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat Struct Biol.* 1999;6(10):943–947.
20. Capaldi AP, Ferguson SJ, Radford SE. The Greek key protein apo-pseudoazurin folds through an obligate on-pathway intermediate. *J Mol Biol.* 1999;286(5):1621–1632.
21. Bai Y. Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c. *Proc Natl Acad Sci.* 1999;96(2): 477–480.
22. Bai Y. Kinetic evidence of an on-pathway intermediate in the folding of lysozyme. *Protein Sci.* 2000;9(1):194–196.
23. Heidary DK, O'Neill JC, Roy M, Jennings PA. An essential intermediate in the folding of dihydrofolate reductase. *Proc Natl Acad Sci.* 2000;97(11):5866–5870.
24. Capaldi AP, Shastry M, Kleanthous C, Roder H, Radford SE. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol.* 2001;8(1):68–72.
25. Rami BR, Krishnamoorthy G, Udgaonkar JB. Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar. *Biochemistry.* 2003;42(26): 7986–8000.
26. Bollen YJ, Sánchez IE, van Mierlo CP. Formation of on-and off-pathway intermediates in the folding kinetics of *Azotobacter vinelandii* apoflavodoxin. *Biochemistry.* 2004;43(32): 10475–10489.
27. Bollen YJ, Kamphuis MB, van Mierlo CP. The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates. *Proc Natl Acad Sci.* 2006;103(11):4095–4100.
28. Houwman JA, Westphal AH, Visser AJ, Borst JW, van Mierlo CP. Concurrent presence of on-and off-pathway folding intermediates of apoflavodoxin at physiological ionic strength. *Phys Chem Chem Phys.* 2018;20(10):7059–7072.
29. Khorasanizadeh S, Peters ID, Roder H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol.* 1996; 3(2):193–205.
30. Englander SW, Mayne L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu Rev Biophys Biomol Struct.* 1992;21(1):243–265.
31. Chamberlain AK, Handel TM, Marqusee S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol.* 1996;3(9):782–787.
32. Udgaonkar JB, Baldwin RL. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease a. *Nature.* 1988;335(6192):694–699.
33. Roder H, Elöve GA, Englander SW. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature.* 1988;335(6192):700–704.
34. Hughson FM, Wright PE, Baldwin RL. Structural characterization of a partly folded apomyoglobin intermediate. *Science.* 1990;249(4976):1544–1548.
35. Bycroft M, Matouschek A, Kellis JT, Serrano L, Fersht AR. Detection and characterization of a folding intermediate in barnase by NMR. *Nature.* 1990;346(6283):488–490.
36. Hosszu L, Craven CJ, Parker MJ, et al. Structure of a kinetic protein folding intermediate by equilibrium amide exchange. *Nat Struct Biol.* 1997;4(10):801–804.
37. Juneja J, Udgaonkar JB. Characterization of the unfolding of ribonuclease a by a pulsed hydrogen exchange study: Evidence for competing pathways for unfolding. *Biochemistry.* 2002;41(8):2641–2654.

38. Baum J, Dobson CM, Evans PA, Hanley C. Characterization of a partly folded protein by NMR methods: Studies on the molten globule state of Guinea pig. Alpha-lactalbumin. *Biochemistry*. 1989;28(1):7–13.
39. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*. 1993;262(5135):892–896.
40. Morozova-Roche LA, Jones JA, Noppe W, Dobson CM. Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. *J Mol Biol*. 1999;289(4):1055–1073.
41. Juneja J, Udgaonkar JB. NMR studies of protein folding. *Curr Sci*. 2003;84:157–172.
42. Patra AK, Udgaonkar JB. Characterization of the folding and unfolding reactions of single-chain monellin: Evidence for multiple intermediates and competing pathways. *Biochemistry*. 2007;46(42):11727–11743.
43. Jha SK, Dasgupta A, Malhotra P, Udgaonkar JB. Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry. *Biochemistry*. 2011;50(15):3062–3074.
44. Davis R, Dobson CM, Vendruscolo M. Determination of the structures of distinct transition state ensembles for a  $\beta$ -sheet peptide with parallel folding pathways. *J Chem Phys*. 2002;117(20):9510–9517.
45. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci*. 2009;106(45):19011–19016.
46. Zanetti-Polzi L, Davis CM, Gruebele M, Dyer RB, Amadei A, Daidone I. Parallel folding pathways of Fip35 WW domain explained by infrared spectra and their computer simulation. *FEBS Lett*. 2017;591(20):3265–3275.
47. Hsu DJ, Leshchev D, Kosheleva I, Kohlstedt KL, Chen LX. Unfolding bovine  $\alpha$ -lactalbumin with T-jump: Characterizing disordered intermediates via time-resolved x-ray solution scattering and molecular dynamics simulations. *J Chem Phys*. 2021;154(10):105101.
48. Aghera N, Udgaonkar JB. Kinetic studies of the folding of heterodimeric monellin: Evidence for switching between alternative parallel pathways. *J Mol Biol*. 2012;420(3):235–250.
49. Pradeep L, Udgaonkar JB. Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar. *J Mol Biol*. 2002;324(2):331–347.
50. Pradeep L, Udgaonkar JB. Osmolytes induce structure in an early intermediate on the folding pathway of barstar. *J Biol Chem*. 2004;279(39):40303–40313.
51. Bhatia S, Krishnamoorthy G, Udgaonkar JB. Mapping distinct sequences of structure formation differentiating multiple folding pathways of a small protein. *J Am Chem Soc*. 2021;143(3):1447–1457.
52. Arrington CB, Teesch LM, Robertson AD. Defining protein ensembles with native-state NH exchange: Kinetics of interconversion and cooperative units from combined NMR and MS analysis. *J Mol Biol*. 1999;285(3):1265–1275.
53. Miranker A, Robinson CV, Radford SE, Aplin RT, Dobson CM. Detection of transient protein folding populations by mass spectrometry. *Science*. 1993;262(5135):896–900.
54. Heidary DK, Gross LA, Roy M, Jennings PA. Evidence for an obligatory intermediate in the folding of Interleukin-1 $\beta$ . *Nat Struct Biol*. 1997;4(9):725–731.
55. Wintrode PL, Rojsajjakul T, Vadrevu R, Matthews CR, Smith DL. An obligatory intermediate controls the folding of the  $\alpha$ -subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol*. 2005;347(5):911–919.
56. Wani AH, Udgaonkar JB. Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration. *Proc Natl Acad Sci*. 2009;106(49):20711–20716.
57. Malhotra P, Udgaonkar JB. Tuning cooperativity on the free energy landscape of protein folding. *Biochemistry*. 2015;54(22):3431–3441.
58. Jethva PN, Udgaonkar JB. Modulation of the extent of cooperative structural change during protein folding by chemical denaturant. *J Phys Chem B*. 2017;121(35):8263–8275.
59. Bhattacharjee R, Udgaonkar JB. Structural characterization of the cooperativity of unfolding of a heterodimeric protein using hydrogen exchange-mass spectrometry. *J Mol Biol*. 2021;433(23):167268.
60. Malhotra P, Udgaonkar JB. Secondary structural change can occur diffusely and not modularly during protein folding and unfolding reactions. *J Am Chem Soc*. 2016;138(18):5866–5878.
61. Malhotra P, Jethva PN, Udgaonkar JB. Chemical denaturants smoothen ruggedness on the free energy landscape of protein folding. *Biochemistry*. 2017;56(31):4053–4063.
62. Jethva PN, Udgaonkar JB. The osmolyte TMAO modulates protein folding cooperativity by altering global protein stability. *Biochemistry*. 2018;57(40):5851–5863.
63. Aghera N, Udgaonkar JB. Stepwise assembly of  $\beta$ -sheet structure during the folding of an SH3 domain revealed by a pulsed hydrogen exchange mass spectrometry study. *Biochemistry*. 2017;56(29):3754–3769.
64. Jha SK, Dhar D, Krishnamoorthy G, Udgaonkar JB. Continuous dissolution of structure during the unfolding of a small protein. *Proc Natl Acad Sci*. 2009;106(27):11113–11118.
65. Jha SK, Udgaonkar JB. Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci*. 2009;106(30):12289–12294.
66. Malhotra P, Udgaonkar JB. High-energy intermediates in protein unfolding characterized by thiol labeling under nativelylike conditions. *Biochemistry*. 2014;53(22):3608–3620.
67. Aghera N, Earanna N, Udgaonkar JB. Equilibrium unfolding studies of monellin: The double-chain variant appears to be more stable than the single-chain variant. *Biochemistry*. 2011;50(13):2434–2444.
68. Goluguri RR, Udgaonkar JB. Rise of the helix from a collapsed globule during the folding of monellin. *Biochemistry*. 2015;54(34):5356–5365.
69. Goluguri RR, Udgaonkar JB. Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein. *J Mol Biol*. 2016;428(15):3102–3117.
70. Bhatia S, Krishnamoorthy G, Udgaonkar JB. Site-specific time-resolved FRET reveals local variations in the unfolding mechanism in an apparently two-state protein unfolding transition. *Phys Chem Chem Phys*. 2018;20(5):3216–3232.
71. Bhatia S, Krishnamoorthy G, Dhar D, Udgaonkar JB. Observation of continuous contraction and a metastable misfolded state during the collapse and folding of a small protein. *J Mol Biol*. 2019;431(19):3814–3826.
72. Rand KD, Pringle SD, Morris M, Engen JR, Brown JM. ETD in a traveling wave ion guide at tuned Z-spray ion source conditions allows for site-specific hydrogen/deuterium exchange

- measurements. *J Am Soc Mass Spectrom.* 2011;22(10):1784–1793.
73. Agashe VR, Shastry M, Udgaonkar JB. Initial hydrophobic collapse in the folding of barstar. *Nature.* 1995;377(6551):754–757.
  74. Houry WA, Rothwarf DM, Scheraga HA. The nature of the initial step in the conformational folding of disulphide-intact ribonuclease a. *Nat Struct Biol.* 1995;2(6):495–503.
  75. Parker MJ, Spencer J, Clarke AR. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J Mol Biol.* 1995;253(5):771–786.
  76. Nölting B, Golbik R, Fersht AR. Submillisecond events in protein folding. *Proc Natl Acad Sci.* 1995;92(23):10668–10672.
  77. Hamada D, Segawa S-I, Goto Y. Non-native  $\alpha$ -helical intermediate in the refolding of  $\beta$ -lactoglobulin, a predominantly  $\beta$ -sheet protein. *Nat Struct Biol.* 1996;3(10):868–873.
  78. Sauder JM, MacKenzie NE, Roder H. Kinetic mechanism of folding and unfolding of *Rhodobacter capsulatus* cytochrome c 2. *Biochemistry.* 1996;35(51):16852–16862.
  79. Ridge JA, Baldwin RL, Labhardt AM. Nature of the fast and slow refolding reactions of iron (III) cytochrome c. *Biochemistry.* 1981;20(6):1622–1630.
  80. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry.* 1991;30(43):10436–10443.
  81. Dodge RW, Scheraga HA. Folding and unfolding kinetics of the proline-to-alanine mutants of bovine pancreatic ribonuclease a. *Biochemistry.* 1996;35(5):1548–1559.
  82. Houry WA, Scheraga HA. Nature of the unfolded state of ribonuclease a: Effect of cis–trans X–pro peptide bond isomerization. *Biochemistry.* 1996;35(36):11719–11733.
  83. Garel J-R, Nall BT, Baldwin RL. Guanidine-unfolded state of ribonuclease a contains both fast-and slow-refolding species. *Proc Natl Acad Sci.* 1976;73(6):1853–1857.
  84. Nall BT, Garel J-R, Baldwin RL. Test of the extended two-state model for the kinetic intermediates observed in the folding transition of ribonuclease a. *J Mol Biol.* 1978;118(3):317–330.
  85. Schmid FX, Baldwin RL. The rate of interconversion between the two unfolded forms of ribonuclease a does not depend on guanidinium chloride concentration. *J Mol Biol.* 1979;133(2):285–287.
  86. Garel J-R. Evidence for involvement of proline cis-trans isomerization in the slow unfolding reaction of RNase a. *Proc Natl Acad Sci.* 1980;77(2):795–798.
  87. Sridevi K, Udgaonkar JB. Surface expansion is independent of and occurs faster than core solvation during the unfolding of barstar. *Biochemistry.* 2003;42(6):1551–1563.
  88. Schreiber G, Fersht AR. The refolding of cis-and trans-peptidylprolyl isomers of barstar. *Biochemistry.* 1993;32(41):11195–11203.
  89. Shastry M, Agashe VR, Udgaonkar JB. Quantitative analysis of the kinetics of denaturation and renaturation of barstar in the folding transition zone. *Protein Sci.* 1994;3(9):1409–1417.
  90. Varley P, Gronenborn AM, Christensen H, Wingfield PT, Pain RH, Clore GM. Kinetics of folding of the all- $\beta$  sheet protein interleukin-1 $\beta$ . *Science.* 1993;260(5111):1110–1113.
  91. Jones BE, Robert Matthews C. Early intermediates in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange and NMR. *Protein Sci.* 1995;4(2):167–177.
  92. Samuel D, Kumar TKS, Balamurugan K, Lin W-Y, Chin D-H, Yu C. Structural events during the refolding of an all  $\beta$ -sheet protein. *J Biol Chem.* 2001;276(6):4134–4141.
  93. Sivaraman T, Kumar T, Tu Y, et al. Secondary structure formation is the earliest structural event in the refolding of an all  $\beta$ -sheet protein. *Biochem Biophys Res Commun.* 1999;260(1):284–288.
  94. Hu W, Walters BT, Kan Z-Y, et al. Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci.* 2013;110(19):7684–7689.
  95. Udgaonkar JB, Baldwin RL. Early folding intermediate of ribonuclease a. *Proc Natl Acad Sci.* 1990;87(21):8197–8201.
  96. Jacobs MD, Fox RO. Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc Natl Acad Sci.* 1994;91(2):449–453.
  97. Ptitsyn O, Pain RH, Semisotnov G, Zerovnik E, Razgulyaev O. Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.* 1990;262(1):20–24.
  98. Elove GA, Chaffotte AF, Roder H, Goldberg ME. Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry.* 1992;31(30):6876–6883.
  99. Mann CJ, Matthews CR. Structure and stability of an early folding intermediate of *Escherichia coli* trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. *Biochemistry.* 1993;32(20):5282–5290.
  100. Yamasaki K, Ogasahara K, Yutani K, Oobatake M, Kanaya S. Folding pathway of *Escherichia coli* ribonuclease HI: A circular dichroism, fluorescence, and NMR study. *Biochemistry.* 1995;34(51):16552–16562.
  101. Parker MJ, Dempsey CE, Lorch M, Clarke AR. Acquisition of native  $\beta$ -strand topology during the rapid collapse phase of protein folding. *Biochemistry.* 1997;36(43):13396–13405.
  102. Aghera N, Udgaonkar JB. Heterologous expression, purification and characterization of heterodimeric monellin. *Protein Expr Purif.* 2011;76(2):248–253.
  103. Wales TE, Fadgen KE, Gerhardt GC, Engen JR. High-speed and high-resolution UPLC separation at zero degrees Celsius. *Anal Chem.* 2008;80(17):6815–6820.
  104. Walters BT, Ricciuti A, Mayne L, Englander SW. Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment. *J Am Soc Mass Spectrom.* 2012;23(12):2132–2139.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Bhattacharjee R, Udgaonkar JB. Differentiating between the sequence of structural events on alternative pathways of folding of a heterodimeric protein. *Protein Science.* 2022;31(12):e4513. <https://doi.org/10.1002/pro.4513>