Molecular Determinants of Site-Specific Conformational Heterogeneity in Protein Folding and Unfolding

विद्या वाचस्पति की
उपाधि की अपेक्षाओं की आंशिक पूर्ति में प्रस्तुत शोध प्रबंध

A thesis submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy

द्वारा / By

अनुष्का कौशिक
Anushka Kaushik

पंजीकरण सं. / Registration No.: 20193673

शोध प्रबंध पर्यवेक्षक / Thesis Supervisor:

प्रो. जयंत बी. उदगांवकर/ Prof. Jayant B. Udgaonkar

सह पर्यवेक्षक / Co-supervisor:

प्रो. अर्नब मुखर्जी/ Prof. Arnab Mukherjee

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान पुणे
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE
2025

Dedicated to

My Family

IISER PUNE

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान, पुणे

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE

(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)

Dr. Homi Bhabha Road, Pune 411008.

CERTIFICATE

Certified that the work incorporated in the thesis entitled **"Molecular Determinants of Site-Specific Conformational Heterogeneity in Protein Folding and Unfolding"** submitted by **Anushka Kaushik** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other university or institution.

Date: 11.02.2026

Place: Pune

Prof. Jayant B. Udgaonkar

(Supervisor)

Prof. Arnab Mukherjee

(Co- supervisor)

boomi@iiserpune.ac.in •Ph: +91-20-25908208

IISER PUNE

भारतीय विज्ञान शिक्षा एवं अनुसंधान संस्थान, पुणे

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pune 411008.

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented, fabricated, or falsified any idea/data/fact/ source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 11.02.2026

Place: Pune

Anushka Kaushik

ID: 20193673

Acknowledgements

As I approach the end of this long and meaningful journey, there are several people I would like to thank, without whom this experience would not have been the same.

I would like to begin by expressing my deepest gratitude to Prof. Jayant Udgaonkar for being an exceptional PhD mentor. He took a personal interest in ensuring that I developed a strong understanding of the experimental methodologies and consistently encouraged me to push myself further. More than teaching me how to analyze and interpret data, he taught me how to think: carefully, critically, and independently.

His style of guidance is far from ordinary. In his own subtle way, he pushes his students to become confident, self-reliant individuals, not just researchers working toward a degree. I can clearly see this change in myself over the course of my PhD. I would strongly encourage anyone reading this thesis to read his article "Advice to PhD Students", which I found myself returning to at various stages of my doctoral journey. His uncompromising insistence on high-quality data and scientific rigor has shaped my approach to research in a fundamental way. I am truly grateful to have been mentored by him.

I would like to thank Prof. Arnab Mukherjee for being my co-supervisor. During the initial months of my PhD, while working with him, I learned to think quantitatively, which has helped me immensely in my projects. I also learned coding during this time and was able to write my own codes for analyzing experimental data independently. I am grateful for his insightful feedback and the many discussions that were instrumental during the early stages of my PhD.

I thank the MHRD for providing financial support during my PhD, and Prof. Srabanti Chaudhury for being a highly supportive PMRF coordinator and for her guidance through the review process and other funding-related matters.

I sincerely thank my RAC members, Prof. Saikrishnan Kayarat and Prof. Santosh K. Jha, for their valuable feedback and insightful discussions during RAC meetings. I am especially grateful to Prof. Kayarat for arranging DBT funding support after my tenure ended.

Special thanks to Sandhya for her guidance during the initial phase of my PhD, particularly for training me in the tr-FRET methodology and for teaching me the foundational skills required to frame and address scientific problems. I learned from her how to critically filter information and identify the most relevant aspects. I would also like to thank Rupam for helping me with some of my experiments and for guiding me with data analysis from time to time. I am grateful to both of them for their valuable comments on my manuscripts and for always being just a call away whenever I needed to discuss anything.

I have had an enriching time in the JBU lab and have been fortunate to work with wonderful colleagues. I would like to thank all the current and past lab members, including Abhilasha, Biswajit, Harish, Ishtiyaq, Lokesh, Maithili, Preeti, Rupam, Sagarika, Sandhya, Shivani, Sreemantee, Suman and Younus. I also had the opportunity to mentor and learn from several students during my PhD, including Anubhuti, Abhilasha, Ananda, Priyanka, and Sagarika, and I am grateful for those experiences.

During the later stages of my PhD, lunch and coffee breaks with Biswajit and Ishtiyaq became a regular and much-needed pause in the day. I thank them for the many scientific discussions over coffee, spanning not only proteins but also broader questions in biology and chemistry.

As they say, friends away from home become family. I cannot express how deeply grateful I am for my closest friends here, who never made me feel alone during this journey. I sincerely thank Kannan, Megha, Purva and Thejas for simply making my life easier and brighter throughout my PhD. A special thanks to Kannan for the many discussions on my data and for always encouraging me, especially during the final stages of my PhD when progress felt slow and motivation was often needed the most.

I am equally thankful to my batchmates, as well as my friends outside IISER, including Rashmi, Rhythm, Aarushi, Vishakha, and Rajan, whose constant phone calls, messages, and unwavering support kept me going even from afar.

Last and most importantly, I am grateful to my parents and my younger brother, Anshul. Their steady support, quiet sacrifices, and belief in me formed the foundation on which this work rests. Everything I have achieved here is deeply tied to them.

Contents

Tables and Figures

Chapter 2. Replacement of the native *cis* prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein

Figures

Chapter 3. Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding

Tables and figures

Chapter 4. Chain entropy modulates cooperativity selectively within intermediate sub-populations during protein folding

Tables and figures

Chapter 5. Concurrent continuous and activated chain collapse during folding of a small heterodimeric protein

Figures

List of Abbreviations

1. N: Native
2. U: Unfolded
3. I: Intermediate
4. MG: Molten globule
5. MNEI: Single-chain monellin
6. dcMN: Double-chain monellin
7. PDB: Protein data bank
8. TNB: Thionitrobenzoate
9. FRET: Förster resonance energy transfer
10. D: Donor
11. A: Acceptor
12. Tr: Time-resolved
13. Ss: Steady-state
14. FCS: Fluorescence correlation spectroscopy
15. HX: Hydrogen exchange
16. MS: Mass spectrometry
17. H: Helix
18. B: β-sheet
19. C: Core
20. E: End-to-end distance
21. MEM: Maximum entropy method
22. SVD: Singular value decomposition
23. PET: Photo-induced electron transfer
24. SAXS: Small angle X-ray scattering
25. NMR: Nuclear magnetic resonance
26. CD: Circular dichroism
27. MRE: Molar residue ellipticity
28. R_g: Radius of gyration
29. R_{DA}: Intra-molecular distance
30. ANS: 1-anilinonapthalene-8-sulfonate
31. 2D-FLCS: Two-dimensional fluorescence lifetime correlation spectroscopy
32. Sm: Single-molecule
33. NATA: N-Acetyl L-tryptophanamide
34. SFM: Stopped flow mixer
35. WT: Wild type
36. EDTA: Ethylenediaminetetraacetic acid
37. DTT: Dithiothreitol
38. GdnHCl: Guanidine hydrochloride
39. TCSPC: Time-correlated single photon counting
40. IRF: Instrument response function
41. FWHM: Full width at half maximum
42. IDP: Intrinsically disordered protein
43. CIDER: Classification of intrinsically disordered ensemble regions
44. NCPR: Net charge per residue

Synopsis

Conformational heterogeneity is an intrinsic feature of protein folding energy landscapes; however, its molecular origins and the factors that control the relative populations of coexisting conformations have remained poorly understood. This thesis, entitled *"**Molecular Determinants of Site-Specific Conformational Heterogeneity in Protein Folding and Unfolding**,"* investigated the molecular origins of conformational heterogeneity in protein folding and unfolding reactions and examined how modulation of coexisting sub-populations shaped folding and unfolding pathways and cooperativity. Using site-specific time-resolved fluorescence resonance energy transfer analyzed by the maximum entropy method in single-chain and heterodimeric variants of monellin, this work directly resolved coexisting conformational sub-populations. Local backbone rigidity and chain connectivity were identified as key molecular determinants of site-specific conformational heterogeneity in protein folding and unfolding.

Background and Research Gap

Conformational heterogeneity in protein folding and unfolding

Protein folding is a stochastic process in which a polypeptide chain explores a rugged free-energy landscape to reach a uniquely structured native state. While classical descriptions often approximate folding and unfolding as transitions between discrete unfolded (U) and native (N) ensembles, it is now evident that multiple conformational sub-populations coexist within the unfolded, intermediate, and native ensembles (Dill et al. 2008; Udgaonkar 2008). Such conformational heterogeneity is an intrinsic feature of protein energy landscapes and plays a central mechanistic role in both productive folding as well as misfolding reactions (Dobson 2004; Malhotra and Udgaonkar 2016; Chiti and Dobson 2017; Brini, Simmerling, and Dill 2020). However, ensemble-averaging experimental probes report only averaged observables and therefore mask coexisting conformations, limiting direct access to heterogeneous sub-populations and their relative contributions to folding mechanisms. Consequently, folding and unfolding reactions often appear two-state in nature, despite underlying non-cooperative structural changes, hidden intermediates, and multiple folding pathways. Resolving heterogeneity therefore requires methodologies that simultaneously provide structural and temporal resolution, together with population-level quantification.

One such methodology is time-resolved fluorescence resonance energy transfer (trFRET), which, when analyzed using the maximum entropy method (MEM), enables direct extraction of intramolecular distance distributions rather than population-averaged mean distances (Lakshmikanth et al. 2001; Jha et al. 2009; Bhatia, Krishnamoorthy, and Udgaonkar 2021). This approach provides population-resolved structural information with picosecond–nanosecond temporal sensitivity and nanometer spatial resolution (Bhatia and Udgaonkar 2022). In this thesis, site-specific trFRET-MEM measurements are used systematically to resolve conformational heterogeneity during folding and unfolding reactions.

Central motivation of the thesis

While conformational heterogeneity during protein folding is now well established, the molecular determinants that encode this heterogeneity and regulate population partitioning between coexisting sub-ensembles remain poorly understood. More fundamentally, whether heterogeneity simply reflects passive ruggedness of energy landscapes or is actively encoded by sequence and topology remains an open mechanistic question. In particular, it is unclear:

- how local backbone rigidity imposed by native *cis*-prolines modulates folding pathways and intermediate sub-populations,
- whether intermediate-state heterogeneity governs the coupling between chain compaction and structure formation,
- how global chain connectivity and topology influence unfolding cooperativity, and
- how chain association in heterodimeric proteins generates and reshapes heterogeneous folding mechanisms.

The central objective of this thesis is therefore to identify the molecular origins of site-specific conformational heterogeneity in protein folding and unfolding reactions, and to determine how this heterogeneity mechanistically controls folding pathways, cooperativity, and structural evolution.

Model system and experimental strategy

To address these questions, the small sweet protein monellin serves as a powerful model system for dissecting the molecular origins of conformational heterogeneity. Its engineered single-chain variant (MNEI) and the naturally occurring heterodimeric form (dcMN) possess essentially the same native structure and sequence, but differ in chain connectivity: in MNEI

the two chains are covalently linked, whereas in dcMN they associate through non-covalent interactions (Patra and Udgaonkar 2007; Aghera, Earanna, and Udgaonkar 2011; Bhattacharjee and Udgaonkar 2021). Comparison of these two forms therefore provides a direct means to isolate the effects of chain connectivity and inter-chain coupling on folding mechanisms and conformational heterogeneity. While MNEI enables systematic site-specific perturbation of local backbone constraints, dcMN allows examination of chain-specific structural responses during folding and unfolding.

Together, these systems, in combination with trFRET-MEM measurements, provide a quantitative framework for resolving coexisting conformational sub-populations and their structural evolution during folding and unfolding.

Key Mechanistic Findings of the Thesis

1. **Substitution of native *cis*-prolines simplifies folding by eliminating not only slow isomerization-limited phases but also the very fast phase of folding.**

   Pro→Ala mutations remove U-state heterogeneity arising from *cis–trans* isomerization and destabilize folding intermediates, revealing that native *cis*-prolines control early and late kinetic complexity during folding.

2. **Folding heterogeneity in MNEI is not eliminated entirely by removal of native *cis*-prolines.**

   Although *cis–trans* proline isomerization contributes to slow kinetic phases and U-state heterogeneity, substitution of the native *cis*-prolines of MNEI to alanine does not abolish pathway heterogeneity. Multiple parallel folding routes persist even in the absence of *cis*-prolines, demonstrating that folding complexity cannot be attributed solely to proline isomerization.

3. **Backbone rigidity encoded by proline residues acts as a tunable molecular determinant of population partitioning during folding.**

   Proline-imposed local backbone constraints selectively stabilize or destabilize specific intermediate sub-populations at different stages of folding, thereby quantitatively modulating their relative populations within folding intermediates.

4. **Intermediate-state heterogeneity governs the coupling between chain compaction and structure formation during later stages of folding.**

Folding intermediates of MNEI comprise coexisting sub-populations with different degrees of compaction. Stabilization of minor compact sub-populations by site-specific Pro→Ala substitutions that reduce local backbone rigidity increases the fraction of molecules undergoing specific contraction without altering the extent of structure formation, demonstrating that compaction can proceed independently of secondary and tertiary structure formation during later stages of folding.

5. **Unfolding of dcMN is intrinsically heterogeneous despite its apparent two-state behaviour.**

   Although ensemble-averaged measurements suggest two-state unfolding, trFRET-MEM analysis reveals coexistence of compact native-like and expanded unfolded-like conformations throughout the unfolding transition, demonstrating that unfolding proceeds in a heterogeneous non-cooperative manner.

6. **Unfolding cooperativity is governed by inter-chain coupling and chain connectivity.**

   When interacting structural elements remain coupled across the two chains of dcMN, unfolding proceeds cooperatively across multiple segments. Upon disruption of inter-chain interactions, the two chains undergo chain-specific, non-cooperative structural responses. In contrast, covalent linkage in the single-chain variant preserves cooperative unfolding even within partially expanded ensembles, demonstrating that chain connectivity controls the extent of unfolding cooperativity.

7. **Refolding of dcMN proceeds through concurrent barrier-limited and continuous structural contraction.**

   Chain association partitions the early folding ensemble into coexisting expanded and compact intermediates. Conversion from expanded to compact sub-populations occurs via activated kinetics, while the compact sub-population itself undergoes gradual contraction, revealing barrier-limited transitions and continuous structural evolution operating simultaneously during folding.

8. **Local backbone constraints and global chain topology constitute the molecular determinants that generate heterogeneous intermediate ensembles during folding.**

   Perturbation of proline-encoded local backbone rigidity and chain connectivity redistributes sub-populations within intermediate states, which in turn shapes folding

pathways, coupling between compaction and structure formation, and the apparent cooperativity of folding and unfolding transitions.

The mechanistic findings listed above are derived from experimental investigations presented in Chapters 2–5, which are summarized below.

Chapter-wise Summary of Results

Replacement of the native *cis* prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein (Chapter 2) (Kaushik and Udgaonkar 2023)

The folding mechanism of MNEI, a single-chain variant of naturally occurring double-chain monellin, is complex, with multiple parallel refolding channels. To determine whether its folding energy landscape could be simplified, the two native *cis*-prolines, Pro41 and Pro93, were mutated, singly and together, to Ala. The stability of P93A was the same as that of the wild-type protein, pWT; however, P41A and P41AP93A were destabilized by ~ 0.9 kcal mol^{-1}. The effects of the mutations on the very fast, fast, slow, and very slow phases of folding were studied. They showed that heterogeneity in the unfolded state arises due to *cis* to *trans* isomerization of the Gly92-Pro93 peptide bond. The Pro41 to Ala mutation abolished the very slow phase of folding, whereas surprisingly, the Pro93 to Ala mutation abolished the very fast phase of folding. Double-jump, interrupted folding experiments indicated that two sequential *trans* to *cis* proline isomerization steps, of the Gly92-Pro93 peptide bond followed by the Arg40-Pro41 peptide bond, lead to the formation of the native state. They also revealed the accumulation of a late native-like intermediate, N*, which differs from the native state in the isomeric status of the Arg40-Pro41 bond, as well as in a few tertiary contacts as monitored by near-UV CD measurements. The Pro to Ala mutations not only eliminated the cis to trans Pro isomerization reaction in the unfolded state, but also the two *trans* to *cis* Pro isomerization reactions during folding. By doing so, and by differentially affecting the relative stabilities of folding intermediates, the mutations resulted in a simplification of the folding mechanism. The two Pro to Ala mutations together accelerate folding to such an extent that the native state forms more than 1000-fold faster than in the case of pWT.

Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding (Chapter 3) (Kaushik and Udgaonkar, 2026a)

Polypeptide chains undergo both compaction and structure formation during folding, but the extent to which these processes are mechanistically coupled remains unclear. Although initial chain collapse can precede structure formation, the two processes invariably appear coupled at later stages of folding. This raises the question of whether the fraction of molecules that undergo initial collapse, as well as the degree of coupling between compaction and structure formation later during folding, are regulated by sequence-encoded structural constraints. To examine this, the folding of the small protein monellin was investigated using time-resolved fluorescence resonance energy transfer (trFRET) analyzed with the maximum entropy method to resolve sub-populations of molecules with native-like and unfolded-like dimensions. Mutation of Pro41 to Ala, or Pro93 to Ala, which relieve local backbone rigidity, selectively stabilized hidden minor conformations within the initial and later intermediate ensembles, respectively. In each case, the minor conformation had a segment that was more compact than in the major one, and its stabilization increased the number of molecules undergoing specific contraction to form the intermediate ensemble, without altering the extent of structure formation. Consequently, sub-populations within these intermediate ensembles could undergo chain contraction independently of structure formation. These findings identify intermediate-state heterogeneity, modifiable by backbone rigidity, as the basis for tunable coupling between chain compaction and structure formation during protein folding.

Chain entropy modulates cooperativity selectively within intermediate subpopulations during protein unfolding (Chapter 4) (Kaushik and Udgaonkar, 2026b)

Protein unfolding invariably appears to be a cooperative transition; yet, the molecular basis by which structural elements could unfold in a coordinated manner remains unresolved. Here, the unfolding mechanism of the naturally occurring heterodimeric protein double-chain monellin (dcMN) was characterized using site-specific time-resolved FRET and fluorescence anisotropy decay measurements made under equilibrium conditions. Although ensemble-averaged measurements suggested an apparently cooperative transition, population-level analysis using the maximum entropy method coupled to time-resolved FRET revealed pronounced conformational heterogeneity, with partially contracted (N-like) coexisting with partially expanded (U-like) sub-populations during unfolding. Time-resolved fluorescence anisotropy decay measurements independently demonstrated that local motional constraints are lost

gradually and asynchronously across different regions of the protein. The N-like sub-populations underwent cooperative expansion across both intra- and inter-chain segments, indicating coordinated responses when inter-chain coupling is maintained. In contrast, the U-like sub-populations displayed pronounced chain-specific, non-cooperative behavior, consistent with independent unfolding of the two chains following loss of coupling. Comparison with a covalently linked single-chain variant demonstrates that chain connectivity suppresses heterogeneity and enforces coordinated unfolding. These results identify restriction of chain entropy arising from inter-chain coupling and covalent connectivity as a molecular determinant that governs whether heterogeneous intermediate sub-populations unfold cooperatively or in a chain-specific manner.

Concurrent continuous and activated chain collapse during folding of a small heterodimeric protein (Chapter 5)

Folding of heterodimeric proteins often couples chain association with structure formation, but the evolution of conformational heterogeneity during refolding remains unclear. The refolding mechanism of the heterodimeric protein double-chain monellin (dcMN) was therefore investigated using site-specific time-resolved fluorescence resonance energy transfer (tr-FRET) analysed by the maximum entropy method (MEM). Ensemble-averaged fluorescence measurements reported the biphasic refolding kinetics previously observed for dcMN, but provided no direct information on the hidden conformational heterogeneity. Population-level analysis instead revealed that following chain association, the early folding ensemble exists as coexisting partially contracted (N-like) and partially expanded (U-like) sub-populations at the first observable time point. During refolding, conversion of U-like to N-like conformations occurred *via* barrier-limited kinetics, while the N-like sub-population itself underwent continuous contraction during the fast phase. The distance distributions throughout refolding could not be described as weighted sums of equilibrium U and N state distributions, demonstrating that the observed sub-populations correspond to structurally distinct intermediates rather than a two-state mixture. These results reveal that the refolding mechanism of dcMN involves barrier-limited conversion of U-like to N-like sub-populations, together with continuous structural contraction within the N-like sub-population.

Conceptual Advances and Significance of the Thesis

This thesis establishes that conformational heterogeneity in protein folding is not merely a passive consequence of rugged energy landscapes, but an actively encoded and tunable

property governed by local backbone constraints and global chain topology. By resolving population-level conformational sub-ensembles with high temporal and spatial resolution, the work demonstrates that folding pathways, cooperativity, and structural evolution arise directly from controlled redistribution of heterogeneous conformations at the sub-population level. A central conceptual advance is therefore the identification of specific molecular determinants that encode and regulate folding heterogeneity.

Together, these findings define a unified mechanistic framework in which protein folding is governed by redistribution of heterogeneous conformational ensembles specified by local structural constraints and global topology. More broadly, the population-resolved experimental strategy employed here provides a general approach for resolving hidden intermediates and heterogeneous ensembles within complex folding and unfolding landscapes. The principles established are likely to extend to misfolding, aggregation, and phase-separation phenomena, where heterogeneous conformational ensembles play central functional and pathological roles.

References

Aghera, Nilesh, Ninganna Earanna, and Jayant B Udgaonkar. 2011. 'Equilibrium unfolding studies of monellin: the double-chain variant appears to be more stable than the single-chain variant', *Biochemistry*, 50: 2434-44.

Bhatia, Sandhya, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. 2021. 'Mapping distinct sequences of structure formation differentiating multiple folding pathways of a small protein', *Journal of the American Chemical Society*, 143: 1447-57.

Bhatia, Sandhya, and Jayant B Udgaonkar. 2022. 'Heterogeneity in protein folding and unfolding reactions', *Chemical Reviews*, 122: 8911-35.

Bhattacharjee, Rupam, and Jayant B Udgaonkar. 2021. 'Structural characterization of the cooperativity of unfolding of a heterodimeric protein using hydrogen exchange-mass spectrometry', *Journal of molecular biology*, 433: 167268.

Brini, Emiliano, Carlos Simmerling, and Ken Dill. 2020. 'Protein storytelling through physics', *Science*, 370: eaaz3041.

Chiti, Fabrizio, and Christopher M Dobson. 2017. 'Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade', *Annual review of biochemistry*, 86: 27-68.

Dill, Ken A, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. 2008. 'The protein folding problem', *Annu. Rev. Biophys.*, 37: 289-316.

Dobson, Christopher M. 2004. "Principles of protein folding, misfolding and aggregation." In *Seminars in cell & developmental biology*, 3-16. Elsevier.

Jha, Santosh Kumar, Deepak Dhar, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. 2009. 'Continuous dissolution of structure during the unfolding of a small protein', *Biophysical Journal*, 96: 81a.

Kaushik, Anushka, and Jayant B Udgaonkar. 2023. 'Replacement of the native cis prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein', *Biophysical Journal*, 122: 3894-908.

Kaushik, Anushka, and Jayant B Udgaonkar. 'Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding'. *Protein Science,* 35(3), p.e70512.

Kaushik, Anushka, and Jayant B Udgaonkar. 'Chain entropy modulates cooperativity selectively within intermediate subpopulations during protein unfolding'. *Biochemistry. DOI: 10.1021/acs.biochem.6c00188.*

Kaushik, Anushka, and Jayant B Udgaonkar. 'Concurrent continuous and activated chain collapse during folding of a small heterodimeric protein'. *Manuscript in preparation*.

Lakshmikanth, GS, K Sridevi, G Krishnamoorthy, and Jayant B Udgaonkar. 2001. 'Structure is lost incrementally during the unfolding of barstar', *nature structural biology*, 8: 799-804.

Malhotra, Pooja, and Jayant B Udgaonkar. 2016. 'How cooperative are protein folding and unfolding transitions?', *Protein Science*, 25: 1924-41.

Patra, Ashish K, and Jayant B Udgaonkar. 2007. 'Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways', *Biochemistry*, 46: 11727-43.

Udgaonkar, Jayant B. 2008. 'Multiple routes and structural heterogeneity in protein folding', *Annu. Rev. Biophys.*, 37: 489-510.

Chapter 1

Protein folding, unfolding, and conformational heterogeneity: concepts and approaches

1.1 Background

1.1.1 The protein folding problem

Proteins perform a wide range of structural, catalytic, and regulatory functions in biological systems, and these functions are linked intimately to their ability to adopt specific three-dimensional structures. They are synthesized in the cell as linear polypeptide chains composed of amino acids linked by peptide bonds (Creighton 1990). For many small, single-domain globular proteins, these initially unstructured chains undergo a spontaneous self-assembly process to form a well-defined three-dimensional native state under physiological conditions (Anfinsen 1973; Onuchic and Wolynes 2004; Dill et al. 2008; Udgaonkar 2008). This conversion from an unfolded state to the native state occurs typically over timescales ranging from microseconds to minutes (Finkelstein 2018).

The native state of a protein is stabilized by a delicate balance of numerous weak, non-covalent interactions, including hydrogen bonding, hydrophobic interactions, Van der Waals forces, and electrostatic contacts (Dill et al. 2008; Dill and MacCallum 2012). Consequently, protein folding differs fundamentally from simple chemical reactions involving the formation or breakage of covalent bonds. Instead, it represents a disorder-to-order transition in which a large loss of conformational entropy of the polypeptide chain is compensated for by the formation of stabilizing intramolecular interactions (Dill et al. 2008; Dill and MacCallum 2012).

The collective balance of these interactions defines the free-energy surface that governs protein folding and unfolding behavior. This surface describes not only the thermodynamic stability of the native state, but also the kinetics of folding and the degree of cooperativity exhibited during the process (Gō 1984; Portman, Takada, and Wolynes 2001; Udgaonkar 2008; Malhotra and Udgaonkar 2016a). Importantly, this free-energy surface is inherently multidimensional, allowing multiple conformational states and folding pathways to coexist even under the same solvent condition (Bryngelson et al. 1995; Wolynes, Onuchic, and Thirumalai 1995; Udgaonkar 2008). As a result, folding often involves multiple pathways rather than a single, well-defined trajectory.

Understanding how proteins fold is of broad biological and biomedical importance. Under certain conditions, such as cellular stress or sequence perturbations arising from mutations or chemical modifications, proteins may fail to fold correctly, leading to the accumulation of misfolded conformations (Dobson 2001; Chiti et al. 2002; Dobson 2003). Such species can

aggregate *in vivo* and are implicated in a wide range of pathological conditions, including neurodegenerative disorders (Eisenberg and Jucker 2012; Knowles, Vendruscolo, and Dobson 2014; Chiti and Dobson 2017). A mechanistic understanding of protein folding is therefore essential not only for elucidating fundamental principles of molecular self-assembly, but also for developing strategies to prevent misfolding and aggregation, and for enabling the rational design of proteins with tailored structural and functional properties (Sosnick 2011; Dill and MacCallum 2012).

Together, these considerations underscore the need for conceptual and experimental frameworks that can capture the complexity of protein folding, including the coexistence of multiple conformational states and pathways, as well as folding processes influenced by chain connectivity and inter-chain interactions (Udgaonkar 2008; Eaton and Wolynes 2017).

1.1.2 Theories of protein folding

The mechanism by which a polypeptide chain folds into its native three-dimensional structure has been described using several models that emphasize different aspects of structure formation and chain dynamics (Ptitsyn 1973; Wetlaufer 1973; Chan and Dill 1990; Bryngelson et al. 1995). These models were developed to rationalize how chain compaction, secondary structure formation, and tertiary interactions are coordinated during folding (Figure 1.1). Rather than representing mutually exclusive descriptions, they capture distinct but complementary features of folding behavior and have collectively shaped contemporary understanding of protein folding mechanisms.

Framework model

The framework model emphasizes hierarchical structure formation, in which folding proceeds through the stepwise assembly of structure acquisition. In this view, local interactions along the polypeptide chain promote the early formation of secondary structural elements such as α-helices, β-sheets, and turns. These preformed secondary structural units subsequently diffuse and associate to generate the native tertiary structure. Folding is thus guided primarily by sequence-encoded propensities for local structure formation, with tertiary interactions emerging as a consequence of the spatial organization and packing of secondary elements (Ptitsyn 1973; Karplus and Weaver 1976, 1994; Udgaonkar and Baldwin 1988; Kim and Baldwin 1990; Dyson and Wright 1993; Baldwin and Rose 1999a, 1999b).

Nucleation-based models

Nucleation-based models propose that folding is initiated by the formation of a critical nucleus stabilized by a combination of secondary and tertiary interactions (Wetlaufer 1973). Formation of this nucleus constitutes a key event in the folding reaction, after which the remainder of the native structure develops cooperatively around it. In the nucleation–condensation model, the folding nucleus is diffuse and only partially structured at early stages, with consolidation of native-like packing occurring concurrently with further structure formation (Jackson and Fersht 1991a; Fersht 1995, 1997; Kiefhaber et al. 1997; Daggett and Fersht 2003). These models emphasize the cooperative nature of folding and the central role of specific native-like interactions in directing the folding process.

Hydrophobic collapse model

The hydrophobic collapse model proposes that folding begins with a rapid collapse of the polypeptide chain driven by the burial of hydrophobic residues. Specific secondary and tertiary structures are subsequently formed within this compact ensemble through slower structural rearrangements (Chan and Dill 1990; Agashe, Shastry, and Udgaonkar 1995; Ptitsyn 1996; Sarkar, Udgaonkar, and Krishnamoorthy 2013). Thus, reduction of conformational entropy through early collapse facilitates subsequent structure formation. However, during later stages of folding, chain compaction is assumed to proceed concomitantly with structure formation.

Energy landscape framework

While the phenomenological models described above differ in the sequence in which chain contraction and secondary and tertiary structure formation occur during folding, protein folding can be best described within the energy landscape framework. In this framework, folding occurs on a multidimensional free-energy surface that describes the stabilities of different conformations as a function of their structural and entropic properties. The unfolded, intermediate, and native states are treated not as single structures, but as ensembles of conformations occupying distinct regions of this landscape. Folding corresponds to diffusive motion on this surface toward the native basin, accompanied by a reduction in free energy and conformational entropy (Dill and Chan 1997).

Within this framework, folding can proceed along multiple routes on the free-energy surface. Landscapes may be relatively smooth, allowing folding to proceed in an apparently two-state manner, or rugged, containing local minima corresponding to partially folded intermediates separated by free-energy barriers (Bryngelson and Wolynes 1987; Bryngelson et al. 1995; Dill and Chan 1997; Pande et al. 1998). The degree of ruggedness influences folding

and unfolding kinetics, the population of intermediates, and the apparent cooperativity of the folding reaction. Factors such as amino acid sequence and solution conditions can modulate the shape of the energy landscape and thereby alter folding behavior (Bryngelson et al. 1995; Onuchic, Luthey-Schulten, and Wolynes 1997; Klimov and Thirumalai 1996; Thirumalai, Klimov, and Woodson 1997).

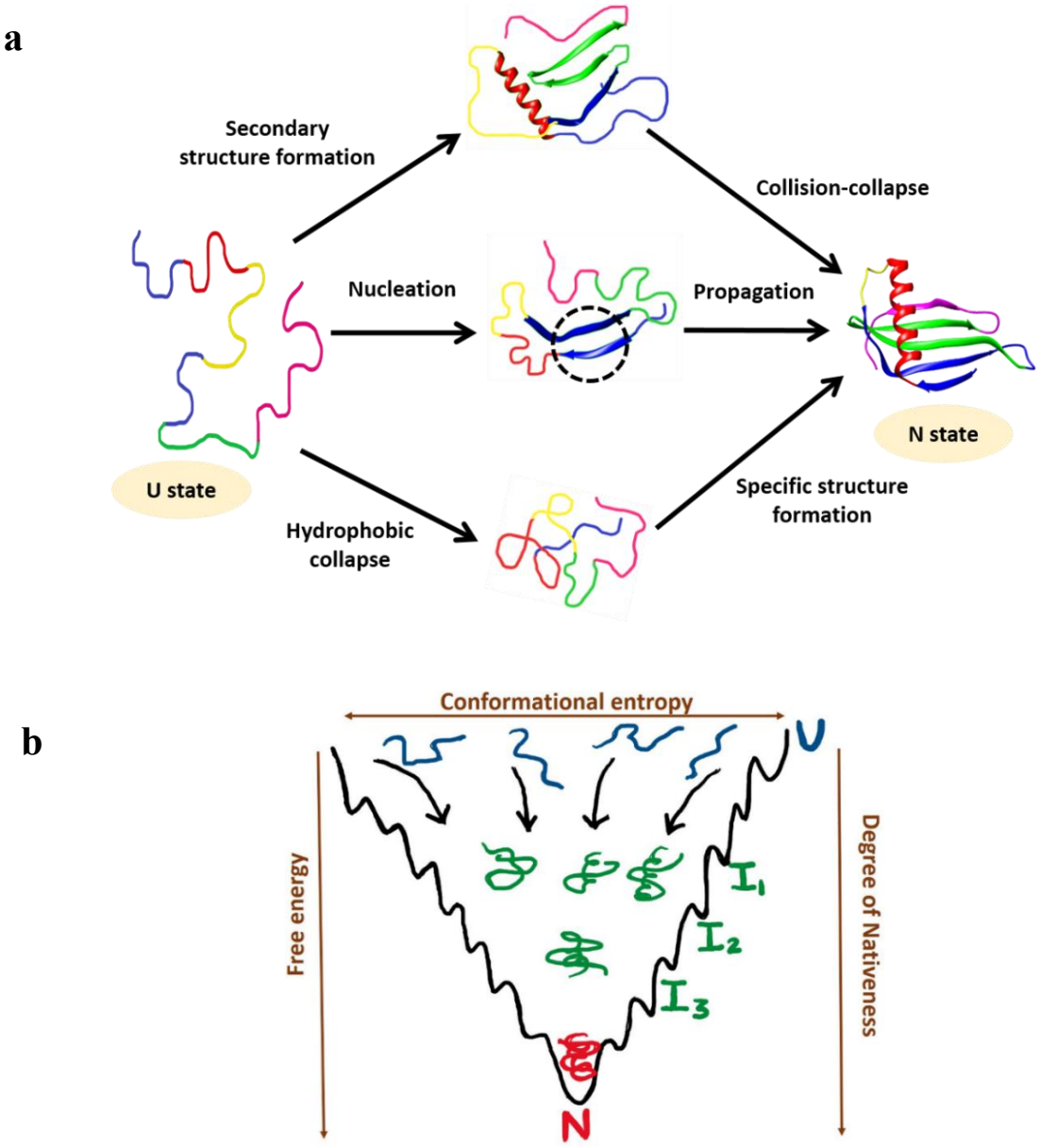


Figure 1.1. Conceptual models and the free energy landscape for protein folding. (a) Schematic comparison of three classical folding models: framework (top-most), nucleation–condensation (middle), and hydrophobic collapse (bottom-most). These models differ in the proposed temporal ordering of secondary structure formation, chain compaction, and tertiary packing en route to the native state. This figure was adapted from (Udgaonkar 2008). **(b)** Folding-funnel representation from energy landscape theory, illustrating an ensemble of high-entropy unfolded conformations converging toward a low-entropy native basin through multiple possible trajectories and local minima corresponding to partially structured intermediates. This figure was adapted from (Dill and MacCallum 2012).

1.1.3 Model systems: monomeric versus oligomeric proteins

To elucidate the fundamental principles governing protein folding and unfolding reactions, experimental studies have relied on carefully chosen model protein systems. In particular, small, single-domain monomeric proteins have played a central role in establishing many of the core concepts of protein folding, including two-state behavior, folding cooperativity, transition-state structure, and the relationship between thermodynamic stability and folding kinetics (Anfinsen 1973; Matouschek et al. 1989; Jackson and Fersht 1991a). Because folding in these systems involves only intrachain interactions, monomeric proteins offer experimentally tractable systems in which folding kinetics and thermodynamics can be directly related to sequence and structure. As a result, much of the current understanding of protein folding has been derived from studies on monomeric proteins.

Most proteins in living cells, however, function as oligomers rather than as isolated monomers. Proteome-wide analyses indicate that only a minority of cellular proteins exist as monomers. In *Escherichia coli*, the average oligomeric state is tetrameric (Goodsell 1991; Goodsell and Olson 2000), while in yeast, most proteins function as dimers or higher-order oligomeric assemblies (Gavin et al. 2002; Krogan et al. 2006). Consequently, extending folding concepts derived from monomeric proteins to oligomeric systems remains a critical challenge in protein folding research.

Oligomeric proteins have several functional and evolutionary advantages. These include cooperative regulation of activity (Schachman 1988), enhanced catalytic efficiency through substrate channelling, as exemplified by multienzyme assemblies (Perham 2000), and the ability to perform mechanical and transport functions, as observed in cytoskeletal filaments and motor protein complexes (Tomishige, Klopfenstein, and Vale 2002).

Oligomeric proteins introduce mechanistic features that are absent in single-chain systems. Folding may be obligatorily coupled to chain association, and native structure formation can involve a competition or coordination between intrachain folding and interchain interactions. As a result, the folding free-energy landscape of an oligomeric protein is shaped jointly by the intrinsic folding energetics of individual chains and by the stability and topology of inter-subunit interfaces. Such coupling can alter folding cooperativity, stabilize partially folded intermediates, and increase the complexity of the folding landscape (Tsai, Xu, and Nussinov 1998; Jaenicke and Lilie 2000; Aguilar 2014; Wodak, Malevanets, and MacKinnon 2015). Nevertheless, relatively few studies have systematically compared the folding thermodynamics and kinetics of monomeric and oligomeric forms of the same protein (Aghera, Earanna, and

Udgaonkar 2011; Aghera and Udgaonkar 2012; Bhattacharjee and Udgaonkar 2021). Such comparisons require carefully designed model systems in which the monomeric and oligomeric forms share the same overall fold and chemistry, differing primarily in chain connectivity and inter-subunit interactions.

Monellin represents a rare and powerful model system in this regard. Naturally occurring monellin exists as a heterodimeric protein, dcMN, whereas single-chain engineered variants, SCM and MNEI, are monomeric forms with very similar native structures (Sung et al. 2001; Peon, Pal, and Zewail 2002; Patra and Udgaonkar 2007; Jha et al. 2009; Aghera and Udgaonkar 2011; Aghera, Earanna, and Udgaonkar 2011; Aghera and Udgaonkar 2012; Bhatia et al. 2019; Bhattacharjee and Udgaonkar 2021). This correspondence allows folding and unfolding reactions to be compared directly between monomeric and heterodimeric forms of the same protein, isolating the effects of chain connectivity and interchain interactions. Previous studies on monellin have shown that such differences influence folding kinetics, the stabilities of intermediates, and folding cooperativity (Aghera, Earanna, and Udgaonkar 2011; Bhattacharjee and Udgaonkar 2021). Comparing the monomeric variant with the heterodimeric variant therefore provides a direct way to probe how interchain coupling and the dimer interface contributes to conformational heterogeneity during folding. It also provides a useful framework to examine how the two chains behave independently during folding. In this context, dcMN and MNEI serve as key model systems in this thesis.

1.2 Heterogeneity in protein folding and unfolding

1.2.1 Overview

Protein folding and unfolding reactions are often described in terms of discrete thermodynamic states such as the unfolded (U), intermediate (I), and native (N) states. Importantly, each of these states represents an ensemble of interconverting conformations rather than a single well-defined structure (Frauenfelder, Parak, and Young 1988; Udgaonkar 2008). Different molecules within the same ensemble may populate distinct conformational sub-populations that interconvert on different timescales, or sample different regions of the folding free-energy landscape (Udgaonkar 2008; Gruebele 2014). At any given time, protein molecules exist in multiple sub-populations of molecules having conformations that differ in free energy, structure, and dynamics, with their relative abundances governed by free-energy differences and their interconversion rates determined by the heights of the intervening barriers (Gibbs 1902; Arrhenius 1889). Folding can thus be described as a redistribution of protein

molecules occurring on a multidimensional free-energy landscape, rather than occupy on a single deterministic pathway connecting U and N (Bryngelson et al. 1995; Wolynes, Onuchic, and Thirumalai 1995). Resolving and quantifying the distinct sub-populations that coexist during folding remains experimentally challenging, but is essential for a detailed mechanistic understanding of protein folding and unfolding reactions. This forms the motivation for the experimental approach used in this thesis.

1.2.2 Experimental approaches to resolve heterogeneity

An overview of the commonly used spectroscopic probes for monitoring different aspects of polypeptide chain folding is given in Table 1.1. Much of our current understanding of protein folding has been derived from ensemble-averaged measurements, including circular dichroism, intrinsic fluorescence, and steady-state fluorescence anisotropy. These approaches report the average signal of a population of molecules and have been instrumental in establishing fundamental principles of folding thermodynamics and kinetics (Creighton 1990; Dill et al. 1995; Greenfield 2006; Woody 2004). Ensemble-averaged probes can resolve heterogeneity only when different conformations contribute measurably distinct signals. However, coexisting conformational sub-states that share similar spectroscopic signatures, or that interconvert faster than the temporal resolution of the probe, remain indistinguishable in the measured signal. Consequently, ensemble-averaged experiments can indicate heterogeneity but cannot directly resolve and quantify the underlying sub-population distributions, or exchange dynamics within conformational ensembles.

These limitations have motivated the use of approaches with higher spatial and temporal resolution that can quantify coexisting conformations and their interconversion dynamics, including hydrogen–deuterium exchange–based methods, time-resolved fluorescence measurements, fluorescence correlation spectroscopy, and single-molecule techniques, as described below:

Hydrogen–deuterium exchange (HX)–based methods, combined with nuclear magnetic resonance (NMR) or mass spectrometry (MS), exploit differences in solvent accessibility and hydrogen bonding to report on structure formation during folding. HX–NMR provides residue-specific information, whereas HX–MS measures deuterium uptake across peptide segments, enabling identification of protected regions, detection of partially folded intermediates, and characterization of structural heterogeneity within folding ensembles (Englander et al. 2016;

Table 1.1. Experimental techniques for probing protein folding, grouped into ensemble-averaging and high-resolution probes.[a]

Technique	Primary observable	Information obtained	Key limitation
ENSEMBLE-AVERAGING PROBES			
Far-UV CD	Mean secondary structure content	Global folding/unfolding transitions; secondary-structure gain/loss	Difficult to determine absolute secondary structure content as can be complicated by contributions due to aromatic residues
Near-UV CD	Mean tertiary packing around aromatic side chains	Tertiary structure formation	Insensitive to dynamic/weak tertiary contacts
Intrinsic tryptophan fluorescence intensity and lifetime	Tryptophan environment	Folding kinetics; compaction/packing near probe	Different conformations can share similar signals
ANS (1-aniline-8-naphthalene sulfonic acid) binding	Exposure of hydrophobic patches	Detection of molten-globule-like intermediates	External fluorophore such as ANS may perturb folding itself
SAXS (small-angle X-ray scattering)	Radius of gyration, global dimensions	U vs compact intermediates vs N; chain collapse	Masks co-existing conformations, yielding only an average size
Steady-state FRET (Fluorescence resonance energy transfer)	Mean FRET efficiency (⟨E⟩) → mean distance	Average chain dimensions / compaction changes	Different mixtures of compact and expanded conformations can yield the same ⟨E⟩; sub-populations are not resolved
HIGH-RESOLUTION PROBES			
Time-resolved FRET coupled with Maximum Entropy Method	Fluorescence lifetime distributions → distance distributions	Coexisting sub-populations with different dimensions; heterogeneity within ensembles	Requires labelling - need to make sure that structure and stability is minimally perturbed upon labelling
Single-molecule FRET	Single-molecule FRET efficiency trajectories	Static vs dynamic heterogeneity; rare states; interconversion	Requires labelling - need to make sure that structure and stability is minimally perturbed upon labelling; photo-physics artefacts
FCS (Fluorescence correlation spectroscopy)/ PET- FCS (Photoinduced electron transfer-FCS)	Fluorescence fluctuations / quenching dynamics	Fast intramolecular motions; exchange rate within ensembles over µs timescales	Sensitive to dynamics, but provides limited direct structural information
2D-FLCS (Fluorescence lifetime correlation spectroscopy)	Lifetime correlation of FRET states	Population exchange at equilibrium; distinct sub-states	Requires specialized instrumentation/analysis
HX–MS (Hydrogen-exchange coupled to mass spectrometry)	Deuterium uptake distributions (peptide-level)	Protection patterns; partially folded regions	Limited spatial resolution relative to NMR; back-exchange
HX–NMR (Hydrogen-exchange coupled to nuclear magnetic resonance)	Residue-specific exchange rates	Site-specific protection; structured regions in intermediates	Requires good NMR spectra; limited time window for very fast exchange; difficult to resolve sub-populations of molecules
Time-resolved anisotropy	Rotational correlation times	Local and global rotations in a molecule; details about size and rigidity	Sensitive to local motional freedom, but does not directly resolve sub-populations

[a] Modified and adapted from (Bartlett and Radford, 2009).

Malhotra and Udgaonkar 2016b; Englander 2000; Krishna et al. 2004; Hamid Wani and Udgaonkar 2006; Wani and Udgaonkar 2009).

Fluorescence resonance energy transfer (FRET)–based approaches report on intramolecular distances and are particularly well suited for monitoring chain compaction and segment-specific structural rearrangements. While steady-state FRET measurements provide ensemble-averaged distances, time-resolved FRET (trFRET) measures fluorescence lifetimes on the 10 ps–10 ns timescale, thereby providing a "snapshot" of the conformational ensemble because these measurements are much faster than the timescale associated with conformational changes (100 ns to 1 μs) (Lakshmikanth et al. 2001; Bhatia and Udgaonkar 2022). Distance distributions can be extracted from trFRET decays using the maximum entropy method (MEM) or singular value decomposition (SVD) analysis (Henry and Hofrichter 1992; Livesey and Brochon 1987; Brochon 1994; Navon et al. 2001; Lakshmikanth et al. 2001; Bhatia and Udgaonkar 2022), enabling quantification of structural heterogeneity within conformational ensembles (Förster 1948; Lakowicz et al. 1991; Haas et al. 1975; Krishnamoorthy 2018) (Figure 1.2). In this thesis, MEM-derived distance-distributions are used to directly resolve and quantify conformational heterogeneity. Such measurements have revealed conformational heterogeneity in several proteins, including barstar (Lakshmikanth et al. 2001; Sridevi et al. 2004), SH3 domains (Kishore, Krishnamoorthy, and Udgaonkar 2013), MNEI (Jha et al. 2009; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a), NTL9 (Peran et al. 2019) and a TIM barrel (Halloran et al. 2019).

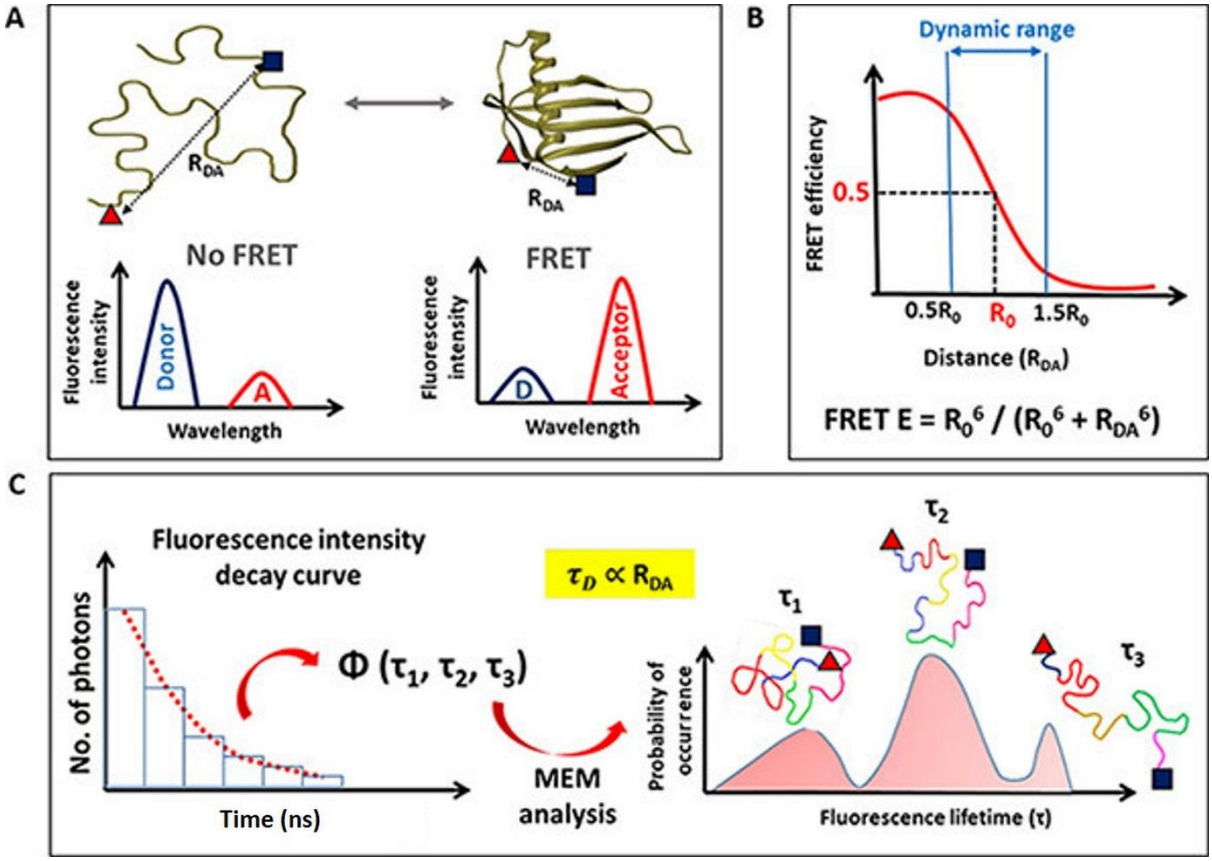


Figure 1.2. Schematic representation of the time-resolved FRET methodology. The donor (D) and acceptor (A) fluorophores used for monitoring FRET are represented as a blue square and a red triangle, respectively. **(a)** In the U state, the D and A fluorophores are separated by large distances, resulting in low FRET and correspondingly higher fluorescence intensity of the D fluorophore. In the N state, reduced D–A separation (R_{DA}) increases FRET, leading to decreased fluorescence intensity of D relative to that of A. **(b)** FRET efficiency has a sigmoidal dependence on the R_{DA} described by the Förster equation (Lakowicz 2006). The Förster radius (R_0) is defined as the distance at which the FRET efficiency is 0.5. **(c)** In trFRET measurements, fluorescence intensity decay curves are recorded and represented as a plot of photon counts *versus* time. The decays are then analyzed using mathematical models for extracting information about different fluorescence lifetimes that would correspond to populations with differential extents of FRET or R_{DA}. This figure was reproduced, with permission, from (Bhatia and Udgaonkar 2022).

Time-resolved fluorescence anisotropy measurements report on local conformational dynamics by measuring rotational correlation times on the nanosecond timescale. Because fluorescence anisotropy decays depend on the rotational mobility of the fluorophore and its coupling to the protein matrix, these measurements are sensitive to local rigidity, segmental flexibility, and changes in packing. Time-resolved fluorescence anisotropy measurements therefore can probe, in a site-specific manner, dynamic heterogeneity within folding ensembles (Lakowicz 2006; Bilsel et al. 1999; Yengo and Berger 2010).

Single-molecule fluorescence techniques, particularly single-molecule FRET (smFRET), probe one molecule at a time and therefore report the underlying distribution of conformations directly. By resolving molecule-to-molecule variability, smFRET can distinguish coexisting sub-populations, differing in intramolecular distances, which would otherwise be obscured in bulk measurements. In addition, analysis of single-molecule trajectories can reveal time-dependent interconversion between conformations, allowing static heterogeneity to be separated from dynamic heterogeneity. Using such approaches, smFRET studies have revealed structural heterogeneity in unfolded, intermediate, and native states of proteins (Deniz et al. 2001; Lipman et al. 2003; Schuler 2007; Schuler et al. 2016).

Fluorescence correlation spectroscopy (FCS) probes conformational dynamics by analysing fluorescence fluctuations. FCS has been used to demonstrate conformational heterogeneity in both native-state ensembles (Elson and Magde 1974; Maiti, Haupts, and Webb 1997; Haupts et al. 1998; Ries et al. 2014; Goluguri, Sen, and Udgaonkar 2019) and unfolded-state ensembles (Chattopadhyay, Elson, and Frieden 2005; Daidone et al. 2010; Voelz et al. 2010; Sherman and Haran 2011; Ries et al. 2014; Borgia et al. 2016). When combined with photo-induced electron transfer (PET-FCS), these methods provide access to fast intramolecular motions and exchange dynamics within ensembles (Doose, Neuweiler, and Sauer 2009; Neuweiler, Johnson, and Fersht 2009; Daidone et al. 2010; Sherman and Haran 2011; Goluguri, Sen, and Udgaonkar 2019). Recently introduced two-dimensional fluorescence lifetime correlation spectroscopic (2D-FLCS) measurements, which couple trFRET and FCS methodologies, also have an ability to identify distinct populations at equilibrium and hence resolve heterogeneity (Otosu et al. 2017).

1.2.3 Complex folding mechanisms as a consequence of heterogeneity

Many proteins show deviations from idealized two-state folding transitions and instead populate discrete intermediates and exhibit multistep folding kinetics (Kim and Baldwin 1990;

Udgaonkar 2008). Folding may proceed through sequential or parallel routes, populating intermediates to different extents. Such complexity in the folding mechanisms arises when conformational sub-states within the unfolded or intermediate ensembles interconvert slowly relative to subsequent folding events, so that different sub-states fold with different rate constants and give rise to kinetically distinguishable folding routes (Ellison and Cavagnero 2006; Udgaonkar 2008).

Experimental evidence for such behavior has been obtained across a wide range of systems. Several monomeric globular proteins, including lysozyme (Radford, Dobson, and Evans 1992), barstar (Shastry and Udgaonkar 1995; Zaidi, Nath, and Udgaonkar 1997), SH3 domains (Dasgupta and Udgaonkar 2012; Aghera and Udgaonkar 2017), RNase T1 (Kiefhaber, Grunert, et al. 1990; Kiefhaber, Quaas, et al. 1990a), and MNEI (Patra and Udgaonkar 2007; Bhatia, Krishnamoorthy, and Udgaonkar 2021a), have been shown to fold *via* parallel pathways and populate multiple structurally distinct intermediates. Folding complexity can be further amplified in oligomeric proteins because intrachain folding is coupled to inter-subunit association. Depending on the intrinsic folding propensities of individual chains, association can occur between unfolded or partially folded subunits, or after substantial structure formation within each chain (Jaenicke 1995; Jaenicke and Lilie 2000; Aghera and Udgaonkar 2012). Consistent with this, the folding kinetics of oligomeric proteins often show changes in the rate-limiting step from subunit association to conformational conversion with a change in concentration (Milla and Sauer 1994; Zeeb et al. 2004). Both monomeric and oligomeric proteins may fold *via* multiple routes (Figure 1.3). Crucially, folding complexity is strongly sensitive to solvent conditions (pH, denaturant concentration, and presence of salts), which can selectively stabilize or destabilize intermediates and redistribute flux among competing pathways (Jamin and Baldwin 1998; Pradeep and Udgaonkar 2002, 2004a, 2004b; Sinha and Udgaonkar 2005; Dasgupta and Udgaonkar 2012; Jha et al. 2011; Aghera and Udgaonkar 2012).

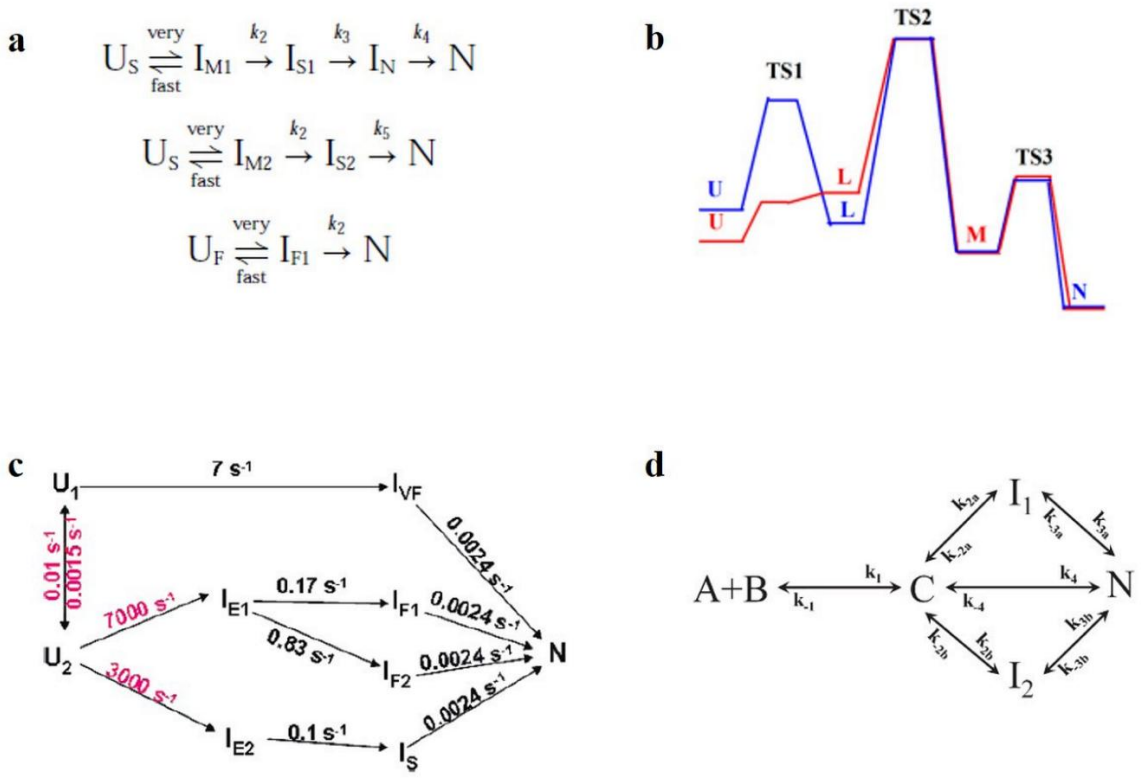


Figure 1.3. Complex folding mechanisms of different protein model systems. (a) Refolding of barstar proceeds through multiple competing pathways that arise from heterogeneity in the U-state ensemble. The slow-refolding unfolded form (U_S) folds *via* two parallel routes in which a late N-like intermediate (I_N) accumulates before conversion to the N state in one of these pathways. In parallel, the fast-refolding unfolded form (U_F) folds via an early intermediate (I_{F1}) to N. This figure was reproduced, with permission, from (Shastry and Udgaonkar 1995). (b) Free-energy profile of the folding of the PI3K SH3 domain in 0 M GdnHCl derived from a multistate (four-state) mechanism (U ↔ L ↔ M ↔ N). The profile reveals intermediate states (L and M), consistent with non-two-state folding behavior. Red: no salt; blue: 0.5 M Na_2SO_4. This figure was reproduced, with permission, from (Dasgupta and Udgaonkar 2012). **(c)** Refolding of single-chain monellin (MNEI) occurs *via* multiple competing pathways arising from heterogeneity in the U-state ensemble. Refolding kinetics resolve three parallel folding phases with kinetic partitioning also originating from the early formed intermediate (I_{E1}). This figure was reproduced, with permission, from (Patra and Udgaonkar 2007). (d) Refolding of heterodimeric monellin (dcMN) proceeds through competing parallel pathways that arise after an initial association step between chains A and B, forming an encounter complex (C), which then partitions into two alternative folding routes involving distinct intermediate ensembles (I_1 and I_2). Refolding switches from one pathway to the other as the GdnHCl concentration is increased. This figure was reproduced, with permission, from (Aghera and Udgaonkar 2012).

1.3 Understanding heterogeneity at different stages of folding

1.3.1 The unfolded (U) state ensemble

The U state represents the starting point of the folding reaction and can influence folding kinetics (Dill and Shortle 1991; Kohn et al. 2004; Sherman and Haran 2006; Udgaonkar 2008, 2013; Schuler et al. 2016). Under strongly denaturing conditions, the U-state ensemble is often approximated as a random-coil–like polymer lacking persistent secondary or tertiary structure (De Gennes 1975; Bower 2002; Rubinstein and Colby 2003). Consistent with polymer theory, ensemble-averaged measurements of unfolded proteins under such conditions generally conform to the scaling laws expected for self-avoiding chains in good solvent (Kohn et al. 2004; Hofmann et al. 2012; Aznauryan et al. 2016; Alessandro et al. 2016). Thus, the U state appears structurally homogeneous when assessed by global parameters such as chain dimensions.

However, this view obscures important features revealed by site-specific and high-resolution experimental probes (Fitzkee and Rose 2004; Meier et al. 2007; Bhavesh et al. 2004; Huang and Grzesiek 2010; Bowler 2012; Peran et al. 2019). Numerous studies have shown that even under denaturing conditions, the U state is best described as an ensemble of conformations that differ locally and globally in structure and dynamics (Udgaonkar 2008, 2013; Schuler et al. 2016; Takahashi, Yoshida, and Oikawa 2018). These conformations can interconvert over a wide range of timescales, from picoseconds to seconds, at rates comparable to or slower than those of early folding events, giving rise to kinetic heterogeneity (Hagen et al. 1996; Lapidus, Eaton, and Hofrichter 2000; Kuzmenkina, Heyes, and Nienhaus 2005; Möglich, Joder, and Kiefhaber 2006; Soranno et al. 2009; Soranno et al. 2017; Waldauer, Bakajin, and Lapidus 2010; Borgia et al. 2012; Schuler et al. 2016).

The timescale of conformational dynamics within the unfolded ensemble is a critical determinant of the folding mechanism (Ellison and Cavagnero 2006; Udgaonkar 2008). When interconversion among unfolded conformations is fast relative to folding, the U-state ensemble behaves effectively as a single population. In contrast, slow conformational exchange gives rise to kinetically distinct unfolded sub-populations, leading to multiple folding phases and pathway heterogeneity (Schuler, Lipman, and Eaton 2002; Kuzmenkina, Heyes, and Nienhaus 2005; Udgaonkar 2008; Jha et al. 2009; Bhatia et al. 2019). U-state dynamics have been observed to span many orders of magnitude across different proteins (Hagen et al. 1996; Lapidus, Eaton, and Hofrichter 2000; Möglich, Joder, and Kiefhaber 2006; Soranno et al. 2009;

Soranno et al. 2017; Waldauer, Bakajin, and Lapidus 2010; Borgia et al. 2012; Schuler et al. 2016; Takahashi, Yoshida, and Oikawa 2018).

A major source of local heterogeneity in the U state ensemble is peptidyl–prolyl *cis–trans* isomerization. The small free-energy difference between *cis* and *trans* isomers allows both forms to be significantly populated (Brandts, Halvorson, and Brennan 1975; Schmid and Baldwin 1979; Wedemeyer, Welker, and Scheraga 2002; Alderson et al. 2018; Waudby et al. 2018). Because interconversion between these isomers is slow relative to subsequent folding steps, distinct unfolded sub-populations can fold independently, giving rise to kinetic heterogeneity (Brandts, Halvorson, and Brennan 1975; Schönbrunner, Koller, and Kiefhaber 1997; Osváth and Gruebele 2003; Udgaonkar 2008). Additional sources of local heterogeneity include alternative rotameric states of aromatic residues (Eis and Lakowicz 1993; Swaminathan et al. 1996; Moors et al. 2006), misligation in prosthetic-group–containing proteins (Pierce and Nall 2000; Sosnick et al. 1994), and scrambling of disulfide bonds in cysteine-rich proteins (Goldenberg and Creighton 1984; Weissman and Kim 1991, 1992; Weissman 1995).

Beyond these local sources, the U state may exhibit global conformational heterogeneity. U-state ensembles can contain sub-populations that differ in their degree of compaction and residual structure (Schuler, Lipman, and Eaton 2002; Merchant et al. 2007; Hofmann et al. 2012; Oikawa et al. 2015; Saito et al. 2016; Schuler et al. 2016). Single-molecule and time-resolved spectroscopic measurements have shown that unfolded proteins can populate both relatively expanded and partially compact conformations (Kuzmenkina, Heyes, and Nienhaus 2005; Hofmann et al. 2012; Aznauryan et al. 2016; Bhatia, Krishnamoorthy, and Udgaonkar 2018). These sub-populations may interconvert on timescales comparable to or slower than early folding events giving rise to a heterogeneous unfolded ensemble (Ellison and Cavagnero 2006; Udgaonkar 2008; Jha et al. 2009; Bhatia et al. 2019). Residual structure in the U state can further contribute to heterogeneity. Both native-like and non-native interactions have been detected in U-state ensembles of several proteins, including ubiquitin, NTL9, lysozyme and barstar (Meier et al. 2007; Huang and Grzesiek 2010; Meng et al. 2013; Klein-Seetharaman et al. 2002; Bowler 2012; Jensen et al. 2014; Halloran et al. 2019; Bhavesh et al. 2004). Native-like residual structure can bias folding toward particular routes by pre-organizing segments of the chain (Neri et al. 1992; Schwalbe et al. 1997; Wirmer et al. 2006; Bowler 2012; Yagi-Utsumi et al. 2020), whereas non-native interactions can trap molecules in conformations that slow folding or redirect it along alternative pathways (Shortle and Ackerman 2001; Jensen et al. 2014; Chung et al. 2015; Takahashi, Yoshida, and Oikawa 2018). The balance between these

interactions is highly sequence-dependent and sensitive to solvent conditions (Schuler et al. 2016; Holehouse and Pappu 2018).

Hence, the U state is neither structurally uniform nor dynamically trivial. Instead, it is a heterogeneous ensemble in which differences in structure and dynamics can strongly influence folding kinetics and pathway selection (Udgaonkar 2008, 2013; Schuler et al. 2016). In particular, U-state heterogeneity can give rise to multiple folding routes.

1.3.2 The Intermediate (I) state ensemble

Protein folding involves progressive contraction and acquisition of structural order as proteins transition from the U-state ensemble to the N-state ensemble (Sherman and Haran 2006; Udgaonkar 2013; Kirmizialtin et al. 2020). In many cases, folding proceeds *via* compact, intermediate ensembles whose chain dimensions and structural order are intermediate between the U and N ensembles (Chan and Dill 1990; Udgaonkar 2008, 2013; Ziv, Thirumalai, and Haran 2009; Haran 2012; Das et al. 2013). Intermediate ensemble comprise heterogeneous sets of compact conformations, can form at different stages of folding and be associated with distinct mechanistic roles.

Upon transfer from denaturing to native solvent conditions, many proteins undergo rapid reduction in chain dimensions to form a collapsed ensemble, Uc (Agashe, Shastry, and Udgaonkar 1995; Sherman and Haran 2006; Sinha and Udgaonkar 2008; Udgaonkar 2013; Goluguri and Udgaonkar 2016). Such early collapse reflects a change in the balance between chain–solvent and chain–chain interactions, driven largely by hydrophobic forces with possible contributions from intramolecular hydrogen bonds (De Gennes 1975; Chan and Dill 1990; Agashe, Shastry, and Udgaonkar 1995; Udgaonkar 2013; Holehouse and Pappu 2018). Uc represents the unfolded ensemble under folding conditions (Chan and Dill 1990; Welker et al. 2004; Udgaonkar 2008, 2013).

A key question is whether early collapse facilitates productive folding or is primarily a solvent-driven response (Sosnick et al. 1994; Udgaonkar 2013; Riback et al. 2017; Takahashi, Yoshida, and Oikawa 2018; Thirumalai et al. 2019; Clark, Plaxco, and Sosnick 2020). Collapse can reduce the conformational search space, but non-specific compaction can also generate misfolded or aggregation-prone states (Bowman et al. 2020; Clark, Plaxco, and Sosnick 2020). A major determinant of which outcome is likely, is the extent to which collapsed intermediates contain native-like structure. In some systems, collapse is accompanied by the formation of native-like interactions, producing intermediates that are on-pathway and promote folding (Ptitsyn and Uversky 1994; Chan and Dill 1990; Shastry and Roder 1998). In other cases,

collapse is largely nonspecific, yielding compact conformations stabilized by hydrophobic clustering with little native structural bias (Qi, Sosnick, and Englander 1998; Agashe, Shastry, and Udgaonkar 1995; Dasgupta and Udgaonkar 2010; Mizukami et al. 2013). Examples of intermediates with signatures of specific collapse include those formed during the folding of apomyoglobin, RNase A, DHFR, and CspTm (Uzawa et al. 2004; Welker et al. 2004; Arai et al. 2007; Hoffmann et al. 2007), whereas nonspecific collapse has been reported for barstar, α-synuclein, PI3K SH3, and staphylococcal nuclease (Agashe, Shastry, and Udgaonkar 1995; Morar et al. 2001; Dasgupta and Udgaonkar 2010; Mizukami et al. 2013). In some systems, both contributions coexist within the same collapsed ensemble, as observed in the case of MNEI (Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Maity and Reddy 2018; Bhatia, Krishnamoorthy, and Udgaonkar 2021b; Bhatia et al. 2019).

The extent of collapse is commonly quantified using SAXS-derived radii of gyration (Rg) or site-specific distance probes such as FRET (R_{DA}) (Plaxco et al. 1999; Sinha and Udgaonkar 2008; Dasgupta and Udgaonkar 2010; Goluguri and Udgaonkar 2015) (Yoo et al. 2012; Riback et al. 2017). Many studies report that Uc is more compact than the unfolded ensemble under denaturing conditions (Sinha and Udgaonkar 2008; Ziv, Thirumalai, and Haran 2009; Yoo et al. 2012; Udgaonkar 2013; Goluguri and Udgaonkar 2016; Zheng et al. 2016; Bhatia et al. 2019). However, SAXS-based studies have suggested only limited initial collapse for some proteins (Plaxco et al. 1999; Yoo et al. 2012; Riback et al. 2017; Bowman et al. 2020; Clark, Plaxco, and Sosnick 2020). Discrepancies between SAXS- and FRET-derived dimensions, observed for example in protein L, indicate that Uc is a heterogeneous ensemble in which R_g and R_{DA} are not uniquely related, such that the same R_{DA} can be consistent with multiple R_g values (and vice versa) (Yoo et al. 2012; Fuertes et al. 2017; Ruff and Holehouse 2017; Song et al. 2017).

Studies on several proteins have shown that Uc is conformationally heterogeneous, with different regions compacting to different extents across molecules (Wu et al. 2008; Arai et al. 2011; Yamada et al. 2013; Halloran et al. 2019; Peran et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021b). Such heterogeneity is often reflected in broad distance distributions and slow interconversion between compact substates (Navon et al. 2001; Flanagan et al. 1992; Lietzow et al. 2002; Bhatia et al. 2019). Multi-site distance measurements have shown non-uniform collapse in proteins including MNEI, adenylate kinase, cytochrome c, ubiquitin and barstar (Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Ratner et al. 2005; Pletneva, Gray, and Winkler 2005; Reddy and Thirumalai 2017; Bhatia, Krishnamoorthy, and Udgaonkar 2021b; Sinha and Udgaonkar 2005; Sinha and Udgaonkar 2009).

Interestingly, further contraction of the polypeptide chain after the initial collapse is widely assumed to occur concomitantly with structure forming steps; however, whether these processes are necessarily coupled and the degree to which they are, remains unresolved. Decoupling of chain contraction from structure formation at later stages of folding is possible only when solvent conditions or mutations selectively modulate compaction without altering the global folding energetics of the protein. Identifying conditions that enable such selective modulation has remained a long-standing challenge in the field. Consequently, contraction and structure formation during later stages of folding continue to be treated as intrinsically coupled, despite representing distinct physical processes (Figure 1.4).

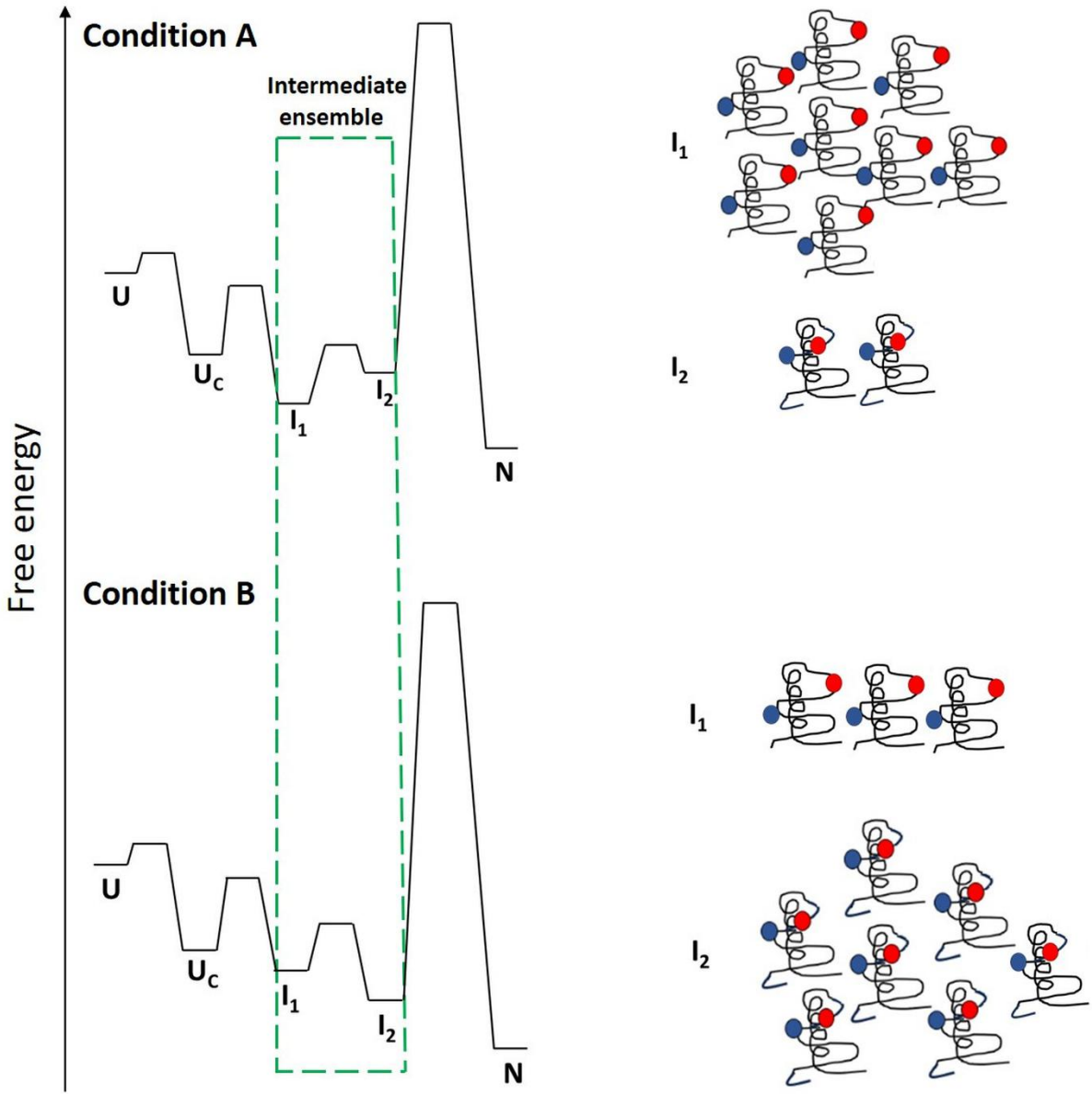


Figure 1.4. Free energy profiles illustrating how the coupling between chain contraction and structure formation can be modulated during protein folding. U and U_C denote the unfolded state and the collapsed non-structured state, respectively. The intermediate ensemble, I, comprises two sub-populations, I_1 and I_2, and accumulates before the rate-determining step in folding. These two subpopulations are similar in the extent to which structure has formed, but differ in the extent to which the chain has contracted in a specific region. The figure shows more chain contraction having occurred in I_2, as apparent by a shorter intra-molecular distance spanned by the blue and red spheres. I_1 and I_2 differ in their stabilities, which also depend on the folding conditions. Under condition A, I_1 is more stable and is therefore predominantly populated in the intermediate ensemble before further folding occurs. The intermediate ensemble will therefore appear to have contracted to only a small extent. Under condition B, I_2 is more stable, and is therefore predominantly populated in the intermediate ensemble. The intermediate ensemble will therefore appear to have contracted to a substantial extent.

1.3.3 The Native (N) state ensemble

Under native conditions, proteins populate a compact, globular folded state stabilized by a defined network of intra-molecular interactions (Chan and Dill 1990; Gō 1984; Udgaonkar 2013; Kirmizialtin et al. 2020). High-resolution structural methods such as X-ray crystallography have been instrumental in defining the N state architecture and have often led to it being described as a unique, well-defined structure. While such structures accurately describe the dominant conformation, they represent static snapshots and do not capture the full range of conformational fluctuations accessible to a native protein under physiological conditions (Frauenfelder, Parak, and Young 1988; Henzler-Wildman and Kern 2007).

Experimental and theoretical studies have established that the N state is dynamic and is best described as an ensemble of interconverting conformations rather than a single static structure (Frauenfelder, Parak, and Young 1988; Lakshmikanth et al. 2001; Englander et al. 2016; Ries et al. 2014; Goluguri, Sen, and Udgaonkar 2019; Malhotra and Udgaonkar 2016b). Even under strongly native conditions, proteins sample higher-energy conformational substates, including locally unfolded or partially destabilized forms (Englander and Kallenbach 1983; Bai et al. 1994; Bai et al. 1995; Englander et al. 2016; Malhotra and Udgaonkar 2016b). These fluctuations arise from the marginal stability of native proteins and reflect the ruggedness in the underlying free-energy landscape (Bowman and Geissler 2014; Ries et al. 2014).

Evidence for N-state heterogeneity has emerged from multiple experimental approaches. Hydrogen–deuterium exchange experiments showed that proteins transiently access partially unfolded conformations in which backbone amide protons become solvent-exposed (Englander and Kallenbach 1983; Bai et al. 1994; Maity et al. 2003; Hernández and LeMaster 2008; Skinner et al. 2012; Englander et al. 2016), demonstrating local structural opening (Bai et al. 1994; Bai et al. 1995; Malhotra and Udgaonkar 2016b). FRET-derived intramolecular distance measurements have further shown that the native ensemble can populate multiple conformations with distinct intramolecular distances. Such heterogeneity has been observed in the case of barstar, SH3 domains, BdpA, and MNEI, where intramolecular distance distributions broaden and shift under mildly destabilizing conditions (Lakshmikanth et al. 2001; Kishore, Krishnamoorthy, and Udgaonkar 2013; Mazouchi et al. 2016; Otosu et al. 2017; Bhatia, Krishnamoorthy, and Udgaonkar 2018).

Dynamic heterogeneity within the N-state ensemble has been revealed by methods that directly probe conformational fluctuations. Fluorescence correlation spectroscopy (FCS) and related approaches have shown that native proteins undergo motions spanning a wide range of

timescales, from fast local fluctuations to slower collective rearrangements (Haupts et al. 1998; Doose, Neuweiler, and Sauer 2009; Ries et al. 2014; Goluguri, Sen, and Udgaonkar 2019). In some systems, the dynamics exhibit slow exchange and glassy behavior, consistent with substantial free-energy barriers separating native substates. In the case of MNEI and the N-terminal domain of spider silk protein, the barriers within the N-state ensemble can be comparable to that for global unfolding (Ries et al. 2014; Bhatia et al. 2019; Goluguri, Sen, and Udgaonkar 2019). Complementary evidence for native-state heterogeneity has also emerged from NMR relaxation-dispersion experiments. Carr–Purcell–Meiboom–Gill (CPMG) and related NMR methods have revealed exchange between the dominant native conformation and low-populated native substates on the microsecond to millisecond timescale (Palmer III, Kroenke, and Loria 2001; Palmer III 2004; Baldwin and Kay 2009; Noguera et al. 2017). These measurements provide direct, residue-specific evidence for dynamic heterogeneity within the native-state ensemble.

N-state heterogeneity can arise from multiple sources. Slow peptidyl–prolyl *cis–trans* isomerization can generate distinct N-like conformers that interconvert on long timescales, as observed for staphylococcal nuclease, calbindin D9k and β2-microglobulin (Evans et al. 1987; Eftink et al. 1989; Koerdel et al. 1990; Svensson, Thulin, and Forsén 1992; Mukaiyama et al. 2013). Local heterogeneity can also arise from alternative side-chain packing (Li and Frieden 2005). At the global level, helix fraying and subtle rearrangements of tertiary contacts can produce native substates with distinct dimensions and stabilities (Serrano, Bilsel, and Gai 2012; Noguera et al. 2017; Otosu et al. 2017). These fluctuations can also be functionally relevant, including in ligand binding, allostery, and aggregation (Henzler-Wildman and Kern 2007; Goluguri, Sen, and Udgaonkar 2019).

1.4 Conceptual basis of cooperativity of folding

As discussed in the preceding sections, proteins vary in folding cooperativity: some appear to undergo effectively all-or-none (two-state) folding transitions, but when these transitions are characterized by suitable probes, high energy intermediates can be detected (Malhotra and Udgaonkar 2014). Some proteins populate stable structured intermediates, and their multi-step folding mechanisms are more apparent (Kim and Baldwin 1990; Jackson and Fersht 1991b; Malhotra and Udgaonkar 2016a). If the U- and N-state ensembles are separated by a single dominant free-energy barrier, folding proceeds in an all-or-none manner and appears highly cooperative, consistent with two-state behavior (Jackson and Fersht 1991a). In contrast, when the folding landscape contains multiple barriers comparable to thermal energy, folding can

proceed through discrete intermediates or via gradual structural change, reflecting limited cooperativity (Kim and Baldwin 1990; Udgaonkar 2008; Malhotra and Udgaonkar 2016a). At the extreme limit of reduced cooperativity lies downhill folding, in which the activation barrier separating U and N becomes negligible and folding occurs *via* continuous population redistribution on a free-energy surface (Bryngelson et al. 1995; Yang and Gruebele 2004; Gruebele 2008) (Figure 1.5).

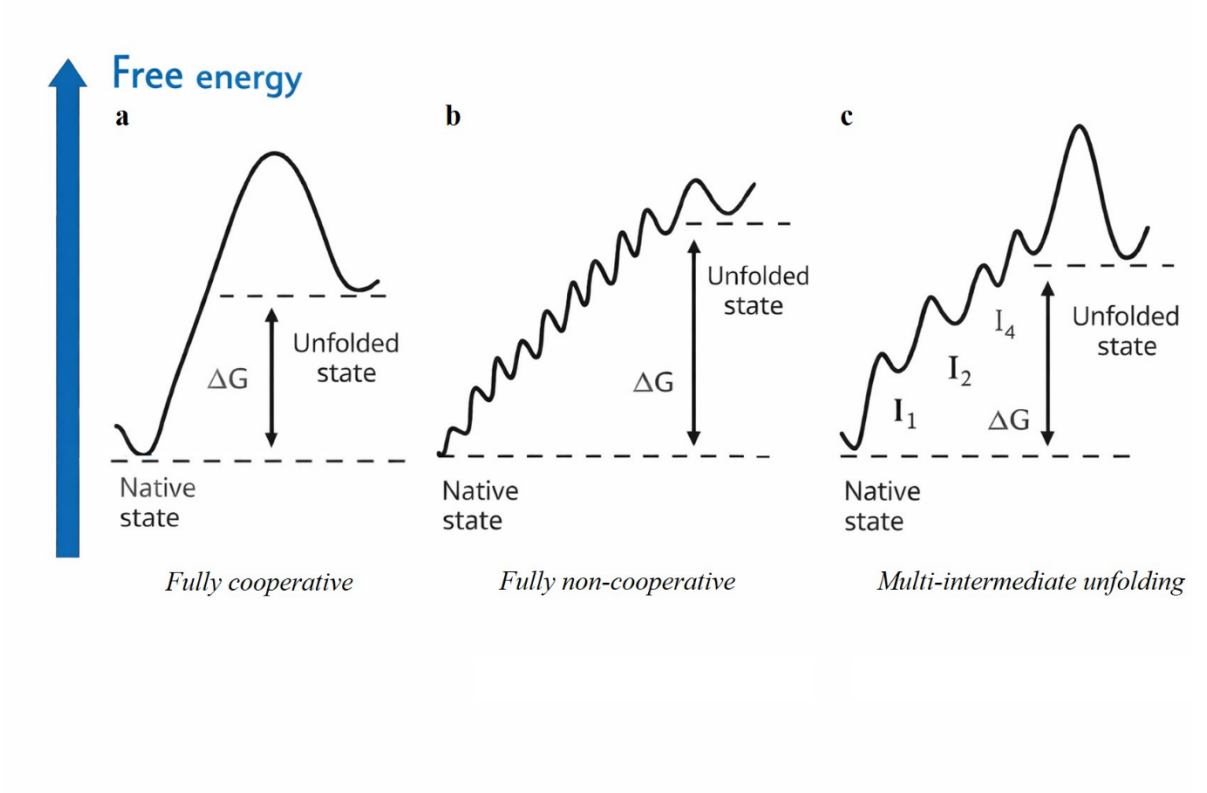


Figure 1.5. Schematic free-energy landscapes illustrating different degrees of cooperativity. (a) Fully cooperative (two-state) folding, in which the unfolded and native ensembles are separated by a single dominant free-energy barrier. **(b)** Fully non-cooperative (uphill) unfolding, in which the native ensemble relaxes continuously toward the unfolded ensemble by crossing only small barriers (typically < ~3 kBT), populating a continuum of conformations that interconvert gradually. **(c)** Reduced cooperativity, in which multiple high-energy intermediates are populated along a rugged landscape. Early steps can be gradual and weakly barrier-limited, whereas the final step may remain strongly cooperative, as shown by a large barrier separating the unfolded ensemble from a late intermediate (I_4). This figure was adapted from (Malhotra and Udgaonkar 2016a).

1.4.1 Cooperativity from equilibrium unfolding measurements

Conventional ensemble-averaging probes often give the impression of two-state folding behavior because the measured signal is dominated by the highly populated states, whereas minor sub-populations within the U- and N-state ensembles remain difficult to resolve (Jackson and Fersht 1991a; Kim and Baldwin 1990; Malhotra and Udgaonkar 2016a). Consequently, many proteins appear to unfold in a two-state manner in urea or GdnHCl when monitored using such probes (Jackson and Fersht 1991a; Pace and Scholtz 1997).

Deviations from two-state behavior become evident when unfolding is monitored using high-resolution methods (Kim and Baldwin 1990; Khurana and Udgaonkar 1994; Sadqi, Fushman, and Munoz 2006; Hamid Wani and Udgaonkar 2006; Lakshmikanth et al. 2001; Kishore, Krishnamoorthy, and Udgaonkar 2013; Bhatia, Krishnamoorthy, and Udgaonkar 2018). Site-specific trFRET measurements have revealed gradual equilibrium structural changes during the unfolding of barstar (Lakshmikanth et al. 2001), and analogous multi-site trFRET measurements showed non-cooperative equilibrium unfolding transitions in PI3K SH3 and MNEI (Kishore, Krishnamoorthy, and Udgaonkar 2013; Bhatia, Krishnamoorthy, and Udgaonkar 2018). Similar conclusions have been drawn from smFRET measurements that show continuous expansion of U-state ensemble with increasing denaturant concentration for several globular proteins and IDPs, including ubiquitin and other model systems (Schuler, Lipman, and Eaton 2002; Hofmann et al. 2012; Aznauryan et al. 2016; Schuler et al. 2016; Saito et al. 2016).

Equilibrium unfolding can also display segment-specific cooperativity within a single protein. Multi-site trFRET studies on MNEI have shown that different structural elements unfold with distinct degrees of cooperativity, indicating a mixture of cooperative and gradual transitions within the protein (Bhatia, Krishnamoorthy, and Udgaonkar 2018). Similar behavior has been observed in equilibrium unfolding experiments monitored by NMR, which have revealed residue-specific differences in stability and non-uniform loss of native structure across the polypeptide chain (Holtzer et al. 1997; Sadqi, Fushman, and Munoz 2006; Sborgi et al. 2015). Thus, thermodynamic cooperativity need not be a single global property; it can vary along the chain, with different regions of a protein differing in local stability (Lakshmikanth et al. 2001; Kishore, Krishnamoorthy, and Udgaonkar 2013; Bhatia, Krishnamoorthy, and Udgaonkar 2018).

1.4.2 Cooperativity from kinetic measurements

Many proteins fold *via* discrete intermediate ensembles, consistent with limited kinetic cooperativity, as observed for lysozyme, barstar, and MNEI (Radford, Dobson, and Evans 1992; Zaidi, Nath, and Udgaonkar 1997; Jha et al. 2009; Bhatia et al. 2019). A prominent example is the formation of molten-globule-like intermediates, which are compact states with substantial secondary structure but only weakly stabilized native tertiary packing (Kuwajima 1989; Ptitsyn and Uversky 1994; Sarkar, Udgaonkar, and Krishnamoorthy 2013; Goluguri and Udgaonkar 2015; Bhatia et al. 2019; Jha et al. 2009). In particular, early folding events such as sub-millisecond chain collapse can exhibit reduced cooperativity (Dasgupta and Udgaonkar 2010; Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Konuma et al. 2015; Bhatia et al. 2019). Probe-dependent burst-phase amplitudes (Georgescu et al. 1998; Pradeep and Udgaonkar 2004a), a non-sigmoidal denaturant dependence of these amplitudes (Sinha and Udgaonkar 2007; Dasgupta and Udgaonkar 2010), and weak denaturant dependence of rate constants have been interpreted as signatures of gradual early events (Sinha and Udgaonkar 2008). Importantly, exponential kinetics alone does not imply a dominant cooperative barrier, because gradual structural relaxation can also yield exponential time courses (Parker and Marqusee 1999; Hagen 2007; Gruebele 2008). Consistent with this, smFRET and trFRET studies on CspTm and MNEI provided evidence that initial collapse can be barrier-limited and continuous, depending on which structural segment is monitored (Schuler, Lipman, and Eaton 2002; Bhatia et al. 2019). These results indicate that kinetic cooperativity is not necessarily uniform across a protein and can vary between early and late folding events (Udgaonkar 2008; Bhatia et al. 2019).

1.4.3 Modulation of cooperativity

The extent of cooperativity is tunable; it can be modulated by perturbations such as mutations (Gruebele 2005; Malhotra, Jethva, and Udgaonkar 2017), temperature changes (Sabelko, Ervin, and Gruebele 1999; Liu and Gruebele 2007), and solvent conditions (Malhotra and Udgaonkar 2015; Jethva and Udgaonkar 2017). Stabilization of an on-pathway intermediate, either by changing solvent conditions, or by mutations can convert an apparently two-state reaction into a three-state mechanism (Nath and Udgaonkar 1995; Pradeep and Udgaonkar 2002; Hamid Wani and Udgaonkar 2006; Spudich, Miller, and Marqusee 2004). Changes in conditions can also redistribute flux between competing unfolding pathways by altering transition-state stabilities (Wright et al. 2003; Aghera and Udgaonkar 2013).

Cooperativity can also be modulated at the level of individual structural elements within a protein. In the case of MNEI and dcMN, HX-MS experiments showed that secondary-structure

formation and dissolution can proceed gradually and diffusively under strongly native conditions, whereas denaturants can promote more cooperative unfolding behavior for specific regions (Malhotra and Udgaonkar 2016b; Bhattacharjee and Udgaonkar 2021). Similar modulation of site-specific cooperativity by the addition of denaturant has also been demonstrated for the PI3K SH3 domain (Jethva and Udgaonkar 2017). Importantly, the transition from cooperative two-state to fully non-cooperative gradual folding can be achieved by perturbing the relative stabilities of the N and U states (Gruebele 2008; Malhotra and Udgaonkar 2015; Jethva and Udgaonkar 2017; Malhotra, Jethva, and Udgaonkar 2017). Overall, apparent cooperativity and folding complexity can be tuned by mutations and solvent conditions through changes in the relative stabilities of the U-, I- and N- state ensembles, as well as the kinetic barriers separating them (Udgaonkar 2008; Malhotra and Udgaonkar 2016a; Gruebele 2008).

1.5 Determinants of conformational heterogeneity

Aim of the current thesis

It is now clear that protein folding and unfolding reactions are often highly heterogeneous, with multiple conformational sub-populations coexisting within the unfolded, intermediate, and native ensembles. However, the determinants that encode this heterogeneity and thereby regulate the population balance between co-existing conformations remain only partially understood. In particular, it is unresolved whether conformational heterogeneity can be modulated by local structural constraints such as backbone rigidity provided by prolines and by global topology/chain connectivity. This thesis investigates the molecular origins of site-specific conformational heterogeneity in protein folding and unfolding reactions, using monellin as a model system. Using MNEI enables systematic site-specific perturbation of local backbone constraints, while comparison with dcMN isolates the effects of chain connectivity. Site-specific time-resolved FRET, analysed using the maximum entropy method, is used throughout to quantify distance distributions and directly resolve co-existing sub-populations during folding and unfolding.

Specific objectives are summarized below:

(i) Native-*cis* prolines as determinants of heterogeneity during folding (Chapter 2).

Proline isomerization is widely regarded as a dominant source of folding heterogeneity in proteins containing native-*cis* prolines, and substitution of these prolines (e.g., Pro→Ala) has

been shown to eliminate slow phases and simplify folding mechanisms. However, it remains unclear whether substitution of native-*cis* prolines is sufficient to eliminate pathway heterogeneity in systems where proline isomerization is expected to be the primary contributor to folding complexity. In addition, native-*cis* prolines enforce specific local backbone constraints in the native state; their substitution can therefore alter the stability and structure of folding intermediates and transition states, potentially affecting not only slow isomerization-limited steps but also early folding events. Effects of Pro substitutions on the fast phases of folding have been much less explored. Chapter 2 addresses these issues by replacing the native-*cis* prolines of MNEI to Ala and examining how folding heterogeneity is affected across all kinetic phases.

(ii) Backbone rigidity as a determinant of the coupling between chain compaction and structure formation in folding intermediates (Chapter 3).

Chain compaction and structure formation are central features of protein folding reactions; yet, the extent to which they are mechanistically coupled remains unclear. Although early chain collapse can precede structure formation, later stages of folding are often assumed to involve concomitant compaction and structure formation. A key unresolved question is whether intermediate ensembles comprise distinct sub-populations with different degrees of compaction, and whether this heterogeneity enables compaction to proceed independently of structure formation. Chapter 3 addresses this question by resolving sub-populations within intermediate ensembles of MNEI using MEM-derived distance distributions, and by testing whether local backbone rigidity, tuned through site-specific Pro→Ala substitutions, redistributes these sub-populations and modulates the coupling of chain compaction and structure formation during folding.

(iii) Chain connectivity/topology as a determinant of unfolding cooperativity (Chapter 4).

Most studies on the cooperativity of the unfolding transition have focused on small monomeric proteins. Much less is known about how cooperativity manifests in heterodimeric proteins, where unfolding can be coupled to inter-chain dissociation and chain-specific responses. It therefore remains unclear whether cooperativity can be tuned by chain connectivity and topology, independent of sequence and native structure. In particular, it is not known whether interacting structural elements unfold more cooperatively when they are covalently linked within a single polypeptide, compared to when they reside on separate chains. Chapter 4 directly tests this by characterizing the equilibrium unfolding of dcMN using site-specific trFRET measurements analysed by the maximum entropy method, together with time-

resolved fluorescence anisotropy decay measurements. The results are then compared with MNEI to examine how chain connectivity modulates apparent unfolding cooperativity.

(iv) Revealing complexity in the folding of dcMN by resolving conformational heterogeneity (Chapter 5).

Conformational heterogeneity during the folding of heterodimeric proteins remains poorly understood, particularly as chain association is intrinsically coupled to structure formation. A key unresolved issue is whether refolding involves barrier-limited interconversion between distinct sub-populations, continuous contraction within a given sub-population, or both processes operating in parallel. Chapter 5 addresses this issue by resolving distance distributions during refolding of dcMN.

Chapter 2

Replacement of the native *cis* prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein

Reproduced with permission from:

Kaushik, A. and Udgaonkar, J.B., 2023. Replacement of the native *cis* prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein. *Biophysical Journal*, *122*(19), pp.3894-3908.

2.1 Significance Statement

Proteins often fold *via* complex mechanisms that feature multiple folding pathways, and it is important to identify the origin of such heterogeneity in protein folding reactions. In proteins containing *cis* Pro residues, folding complexity is invariably attributed to proline isomerization, both during unfolding and during subsequent folding. In this study, it is shown that, even upon mutating both the two *cis* Pro residues of monellin to Ala, two out of three folding pathways still remain operational, suggesting an alternative source of complexity. The mutations do, however, simplify the overall folding mechanism: the very slow phase of folding is eliminated, and so is the slow phase of folding on one of the competing pathways that still operate, as a result of destabilization of a folding intermediate. As a consequence, monellin not having any *cis* Pro residue folds more than a thousand-fold faster than the wt protein. The mutations also reveal how structure formation can be coupled to proline isomerization during folding. Importantly and surprisingly, the mutations also abolish the very fast phase of folding by their effect on the unfolded state. This study therefore demonstrates a new way in which substitution of Pro residues can reduce folding complexity, by eliminating not only the slow phases but also the fast phases of folding.

2.2 Introduction

The mechanism of protein folding from an unstructured, random-coil unfolded (U) state to a uniquely structured native (N) state is still not fully understood. In particular, while various experimental, theoretical and computational studies have highlighted the existence of parallel folding channels (Wallace and Matthews 2002; Udgaonkar 2008; Aksel and Barrick 2014), little is understood about the sequence determinants responsible for multiple folding pathways. Multiple folding pathways can arise when the sub-ensembles within unfolded or intermediate ensembles interconvert on time scales much slower than the subsequent folding steps (Udgaonkar 2008; Eaton and Wolynes 2017). Intermediate ensembles populated at different stages during the folding reaction may be composed of structurally distinct sub-ensembles, resulting in multiple routes to the N state (Chavez et al. 2006; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). Determination of the cause of heterogeneity at the early and intermediate stages of a folding reaction, would lead to a better understanding of the interactions dictating the sequence of structure-forming events during folding. A major cause of the heterogeneity in the U state ensemble for many proteins, which results in parallel channels of folding, is the isomerization of peptidyl-prolyl bonds (Brandts, Halvorson, and

Brennan 1975; Garel, Nall, and Baldwin 1976; Kiefhaber, Quaas, et al. 1990a; Kiefhaber, Kohler, and Schmid 1992; Kiefhaber and Schmid 1992; Shastry, Agashe, and Udgaonkar 1994; Wu and Matthews 2002b). The *trans* isomer is much more stable than the *cis* isomer in the case of a non-prolyl bond, but is only slightly more stable for a prolyl bond. Hence, both the isomeric states for the peptidyl-prolyl bonds may be populated significantly in the U state ensemble (Kim and Baldwin 1982; Eberhardt, Loh, and Raines 1993; Wedemeyer, Welker, and Scheraga 2002). Of course, there can be causes of heterogeneity in the U state ensemble, other than proline isomerization (Bhuyan and Udgaonkar 1998; Pierce and Nall 2000; Welker et al. 2001).

Of the twenty naturally occurring amino acid residues found in proteins, proline is unique in that its side chain is covalently linked to the backbone nitrogen, thereby forming a cyclic ring. Thus, up to 25 % of X-Pro bonds are found in the *cis* conformation in the N states of different proteins (Stewart, Sarkar, and Wampler 1990). They typically occur in tight turns and kinks essential for maintaining the structural integrity of the protein (Stewart, Sarkar, and Wampler 1990). In the U state, however, an X-Pro bond that is *cis* in the N state, can undergo *cis* to *trans* isomerization, resulting in significant population of the *trans* conformation. The important step of *trans* to *cis* prolyl isomerization then becomes inevitable along the folding pathway, as shown for many proteins (Brandts, Halvorson, and Brennan 1975; Wedemeyer, Welker, and Scheraga 2002; Kim and Baldwin 1982; Shastry, Agashe, and Udgaonkar 1994; Schreiber and Fersht 1993; Wu and Matthews 2002a). Such prolyl isomerization steps (either *cis* to *trans* or *trans* to *cis*) are usually much slower than the structure-forming steps, to which they are coupled (Veeraraghavan, Nall, and Fink 1997; Maki et al. 1999; Bhuyan and Udgaonkar 1999). Thus, the occurrence of *cis* prolyl bonds in the N state, though structurally significant, can result in the retardation of the folding reaction. However, the interpretation of studies of the folding of proteins in which the rate-limiting Pro residues were replaced by other residues has not been straight-forward (Herning et al. 1991; Tweedy et al. 1993; Mayr et al. 1994; Odefey, Mayr, and Schmid 1995; Vanhove et al. 1996; Walkenhorst, Green, and Roder 1997; Pappenberger et al. 2001). Nonetheless, it is worthwhile to explore the potential simplification of the folding mechanism upon replacing Pro residues, and to thereby provide additional detail on the influence of the amino acid sequence on the mechanism of protein folding.

Single-chain monellin (MNEI) is a variant of naturally occurring heterodimeric monellin, in which the C-terminus of chain B is linked to the N-terminus of chain A *via* a Gly-Phe dipeptide linker (Patra and Udgaonkar 2007). It is a small globular protein (97 amino acid residues) which consists of an α-helix packed against a five-stranded β-sheet (Figure 2.1a).

Detailed characterization of the folding and unfolding mechanisms of MNEI using optical probes (Kimura, Uzawa, et al. 2005; Patra and Udgaonkar 2007; Jha et al. 2009; Goluguri and Udgaonkar 2016) and mass spectrometry (Malhotra and Udgaonkar 2016b) has been achieved in previous studies. Relevant to the current study, a previous fluorescence-based study (Patra and Udgaonkar 2007) identified five distinct kinetic phases of refolding of MNEI: an ultrafast ($> 1000\ s^{-1}$) phase monitored using ANS fluorescence, very fast ($\sim 10\ s^{-1}$), fast ($\sim 1\ s^{-1}$) and slow ($\sim 0.1\ s^{-1}$) phases monitored using Trp4 fluorescence, and a very slow ($< 0.01\ s^{-1}$) phase monitored by double-jump, interrupted refolding experiments. The U state ensemble was shown to consist of two sub-ensembles, with U_1 folding in the very fast phase, and U_2 folding in the competing fast and slow pathways, and a folding scheme (Patra and Udgaonkar 2007) was proposed. A very recent trFRET study, which monitored the compaction of different structural segments of MNEI (Bhatia, Krishnamoorthy, and Udgaonkar 2021a), also independently arrived at a similar folding scheme as shown in Figure 2.1b.

The intermediate, I_{F1}, populating on the fast phase arising from an early intermediate, I_{E1} was shown to be heterogeneous, consisting of two sub-ensembles that form in parallel (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). It was shown that the N-like intermediate, N* formed directly from the U_1 sub-ensemble, but *via* intermediates along two different parallel pathways from the U_2 sub-ensemble. It was suggested that heterogeneity in the U state ensemble arises due to different isomeric states of Pro residue(s). The molecules folding from the U_1 sub-ensemble differ in the sequence of structure forming steps, from the molecules folding from the U_2 sub-ensemble (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). A possible reason for this could be that different isomeric states of Pro residues in these sub-ensembles, provide different structural biases for subsequent folding events.

It is important to identify the Pro residue(s) responsible for the kinetic heterogeneity in the U state ensemble, which gives rise to parallel pathways of folding. Peptidyl-prolyl bonds that are *cis* in the N state are more likely to cause heterogeneity in the U state than those present in the *trans* conformation(s) (Kim and Baldwin 1982; Wedemeyer, Welker, and Scheraga 2002). Of the six Pro residues present in MNEI, Pro41 and Pro93 are in the *cis* conformation in the N state. Pro41 is present in the middle of the chain and participates in the formation of the β-sheet, whereas, Pro93 is present in the C-terminus region (Figure 2.1a). It was shown in a previous study that the very slow rate of formation of the N state (Figure 2.1b) from the late intermediates is independent of the GdnHCl concentration in which the protein was refolded, as is known to be the case for proline isomerization reactions (Patra and Udgaonkar 2007).

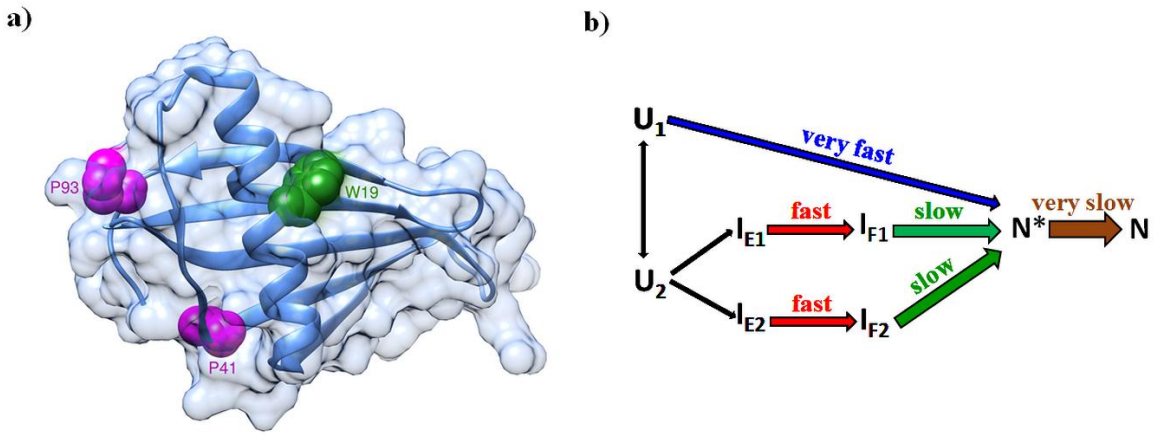


Figure 2.1. Structure and complex folding mechanism of the single-chain monellin (MNEI). (a) Structure of MNEI showing the locations of the two *cis* Pro residues mutated to Ala and the single Trp residue. The side-chains of Pro41 and Pro93 are buried to extents of 97% and 40%, respectively. The structure is drawn using Chimera (PDB ID: 1IV7). (b) The complex folding mechanism of MNEI with multiple parallel refolding channels (Bhatia, Krishnamoorthy, and Udgaonkar 2021a; Patra and Udgaonkar 2007).

Moreover, the rate constant of the formation of the N state was observed to lie in the range of 0.01 to 0.001 s^{-1}, which is that expected for prolyl bond isomerization (Schmid and Baldwin 1979). It is therefore crucial to demonstrate the role of the native *cis* prolines at the later stages of folding of MNEI, and segregate the slow proline isomerization events from the structure-forming folding steps.

In the current study, the potential roles of Pro41 and Pro93 in retarding the overall folding kinetics of pWT, and also in giving rise to multiple pathways has been explored. Pro41 and Pro93 were replaced with Ala using mutagenesis, and the folding kinetics of the mutant variants P41A, P93A and P41AP93A were studied with the aim of determining whether and how elimination of the Pro isomerization steps, could result in a simplification of the folding mechanism of MNEI.

2.3 Materials and Methods

Buffers and Reagents. The native buffer used in all the experiments contained 20 mM Tris, 250 μM EDTA at pH 8.0. The unfolding buffer contained GdnHCl (ultrapure grade, 99.5+ % from Alfa Aesar) in native buffer. 1 mM DTT was added to both the buffers. The concentrations of GdnHCl solutions were determined from the measurement of refractive indices on an Abbe 3L refractometer from Milton Roy. All the buffers and solutions were filtered using 0.22 μm filters, and degassed before use. All the experiments were carried out at room temperature (25°C).

Site-directed mutagenesis. The pseudo wild-type, pWT, (W4YF19W) mutant variant of MNEI was used in this study (Bhatia, Krishnamoorthy, and Udgaonkar 2019). Mutant variants of pWT were generated using the Quickchange site-directed mutagenesis method developed by Stratagene. Three Pro mutant variants were prepared: P41A, P93A and P41AP93A. The mutations were verified by DNA sequencing.

Protein expression and purification. pWT and the Pro mutant variants were expressed and purified as reported previously (Patra and Udgaonkar 2007). The purity of each protein was confirmed by electrospray ionization mass spectrometry (Waters corporation). Each protein had its expected mass. Protein concentrations were determined by the measurement of absorbance at 280 nm, and using an extinction coefficient of 18300 M^{-1} cm^{-1} (Bhatia, Krishnamoorthy, and Udgaonkar 2019).

Fluorescence spectra. Fluorescence spectra were collected on a Fluoromax 4 (Horiba) spectrofluorimeter. Fluorescence was excited at 295 nm, and the emission spectra were collected from 305 to 400 nm, with a response time of 1 s, and excitation and emission bandwidths of 1 nm and 5 nm, respectively. The protein concentration was 10 μM. Each spectrum was recorded as the average of three fluorescence emission wavelength scans.

Far-UV CD and Near-UV CD spectra. Measurements of CD spectra were carried out on a Jasco J-815 spectropolarimeter. Far-UV CD spectra were collected using a 0.2 mm path length cuvette, a bandwidth of 1 nm, a response time of 1 s, and a scan speed of 20 nm min^{-1}. Thirty wavelength scans were averaged for each spectrum.

Near-UV CD spectra were collected using a 10 cm path length cuvette, a bandwidth of 1 nm, a response time of 1 s, and a scan speed of 20 nm min^{-1}. Fifty wavelength scans were averaged for each spectrum. The protein concentration was ~ 10 μM.

Near-UV CD monitored kinetic folding experiments. Folding kinetics was monitored at 270 nm, using a 10 cm path length cuvette, with a response time of 1 s and an integration time of 16 s. The protein was unfolded in 4 M GdnHCl for 6 h. Folding was initiated by dilution with the native buffer such that the final GdnHCl concentration and the protein concentration were 0.4 M and 10 μM, respectively. The data obtained were averaged for every 300 s. The data were fit to a single-exponential.

Fluorescence-monitored equilibrium and kinetic folding experiments. For GdnHCl-induced equilibrium unfolding experiments, 10 μM of protein was incubated for > 6 h in different concentrations of GdnHCl (0 to 4 M). The equilibrated protein samples were excited at 295 nm, and the fluorescence signals were monitored on a MOS-450 optical system (Biologic), using a 360 ± 10 nm band-pass filter (Asahi spectra). A two-state, N↔U model was used to fit the sigmoidal equilibrium unfolding transitions (Agashe and Udgaonkar 1995) to obtain the values for the free energy of unfolding in water, $\Delta G_{U,}$ and the midpoint of transition (Cm).

Millisecond kinetic folding experiments were carried out on a SFM-4 stopped-flow machine (Biologic). A mixing dead time of ~ 10 ms was achieved using a cuvette of path length 0.20 cm, with a flow rate of 5 mL s^{-1}. The final protein concentration was ~ 10 μM. The excitation wavelength was 295 nm, and the fluorescence emission was collected using a 360 ±10 nm band-pass filter (Asahi spectra). The protein was unfolded in 4 M GdnHCl for 6 h. Folding was initiated by rapidly diluting the GdnHCl solution with the native buffer to obtain

different final concentrations of GdnHCl. The kinetic traces obtained were fit to the minimum number of statistically significant exponentials.

Double-jump, interrupted folding experiment. Folding was initiated by manually mixing 30 µL of equilibrium-unfolded protein in 4 M GdnHCl with 270 µL of native buffer such that the final GdnHCl concentration was 0.4 M. The folding mixture was aged for different folding times, 't'. After aging for time t, 300 µL of the aged-folding mixture were manually mixed with 300 µL of 8 M GdnHCl, and unfolding in ~ 4.2 M GdnHCl was monitored by measurement of intrinsic Trp fluorescence. This procedure was followed for pWT and P93A. A similar protocol was used to carry out experiments for P41A, except that the unfolding was carried out at a final GdnHCl concentration of 3.5 M. The final protein concentration was ~ 10 µM. For P41AP93A, a Biologic SFM-4 mixing module was used to mix 30 µL of equilibrium-unfolded protein in 4 M GdnHCl with 270 µL of native buffer, such that the final concentration of GdnHCl was 0.4 M. The folding mixture was aged for different durations in a delay loop of 190 µL volume. After different folding times, 50 µL of the solution in the delay loop were mixed with 305 µL of ~ 4.8 M GdnHCl unfolding buffer, and unfolding was monitored at a final GdnHCl concentration of ~ 4.2 M in a 0.15 cm path length cuvette. The final protein concentration was ~ 6 µM. A dead-time of 35 ms was achieved using flow rates of 1.4 mL s^{-1}.

The kinetic traces for unfolding in 4.2 M GdnHCl (for pWT, P93A and P41AP93A) and 3.5 M (for P41A) GdnHCl, after first folding in 0.4 M GdnHCl for different lengths of time were fit to a single-exponential for an exponential phase corresponding to the unfolding of their N state (the unfolding rate constants of the N state for all the four protein variants were determined separately from manual-mixing kinetic unfolding experiments). The signal change of the unfolding kinetic trace at different times of refolding was normalized with that obtained from unfolding of the refolded state and, the fractional change in signal was plotted as a function of the time at which the refolding was interrupted.

2.4 Results

The Pro to Ala mutations have a minor effect on structure and do not significantly alter protein stability

As can be seen from the fluorescence spectra of the N state of all four protein variants (Figure 2.2a), there was no major change in the fluorescence properties upon mutation of Pro41 and Pro93 to Ala. Hence, the mutations did not significantly affect the environment of Trp19 in the different variants.

The far-UV CD spectrum of each of the four protein variants had a minimum and a maximum at 214 nm and 192 nm, respectively, characteristic of a β-sheet rich protein (Figure 2.2b). This suggested that no significant changes in the polypeptide backbone have occurred upon mutation. The far-UV CD spectrum of P93A was very similar to that of pWT; however, an increase in the amplitude of the minimum and the maximum peak in the case of P41A and P41AP93A was seen. Also, the Pro41 to Ala mutation resulted in a minor change in the shape of the spectrum, which will be discussed later. The extra plateau-like small hump observed at ~ 205 nm in the far-UV CD spectra of pWT and P93A arises because of aromatic amino acid residue contribution (Woody 1978; Chakrabartty et al. 1993). Since this peak was absent in the case of P41A and P41AP93A, it is speculated that some stabilizing interaction between Pro41 and an aromatic amino residue has been perturbed by the mutation.

The near-UV CD spectrum of P41AP93A was similar to that of pWT (Figure 2.2c). However, there were differences in the spectra of P41A and P93A, especially in the 265-285 nm wavelength region, indicating changes in the packing of one or more Tyr residues which occur upon Pro41 and Pro93 mutations.

Figure 2.3 shows the fluorescence-monitored equilibrium unfolding transitions of all four protein variants. The stability of P93A was the same as that of pWT, however, P41A and P41AP93A were slightly destabilized. The stabilities of P41A and P41AP93A were ~ 0.9 and ~ 0.8 kcal mol^{-1} respectively, lower than that of pWT. Because Pro41 is present in the middle of the chain and participates in the β-sheet of the protein, mutation of this residue to Ala might have disrupted some stabilizing interactions, resulting in a slight decrease in conformational stability.

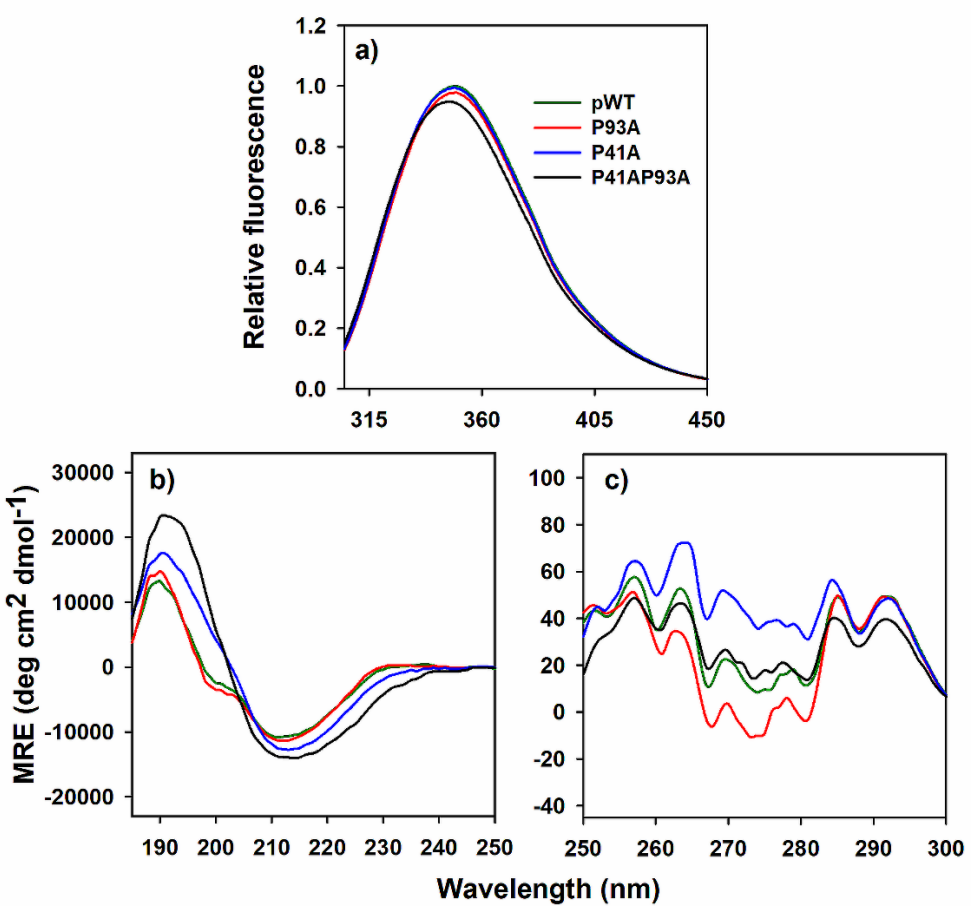


Figure 2.2. Spectroscopic characterization of pWT and the Pro mutant variants. Fluorescence (a), far-UV CD (b), and near-UV CD (c) spectra. The fluorescence spectra of the Pro mutant variants are normalized to that of pWT. The different types of lines represent the different mutant variants as indicated.

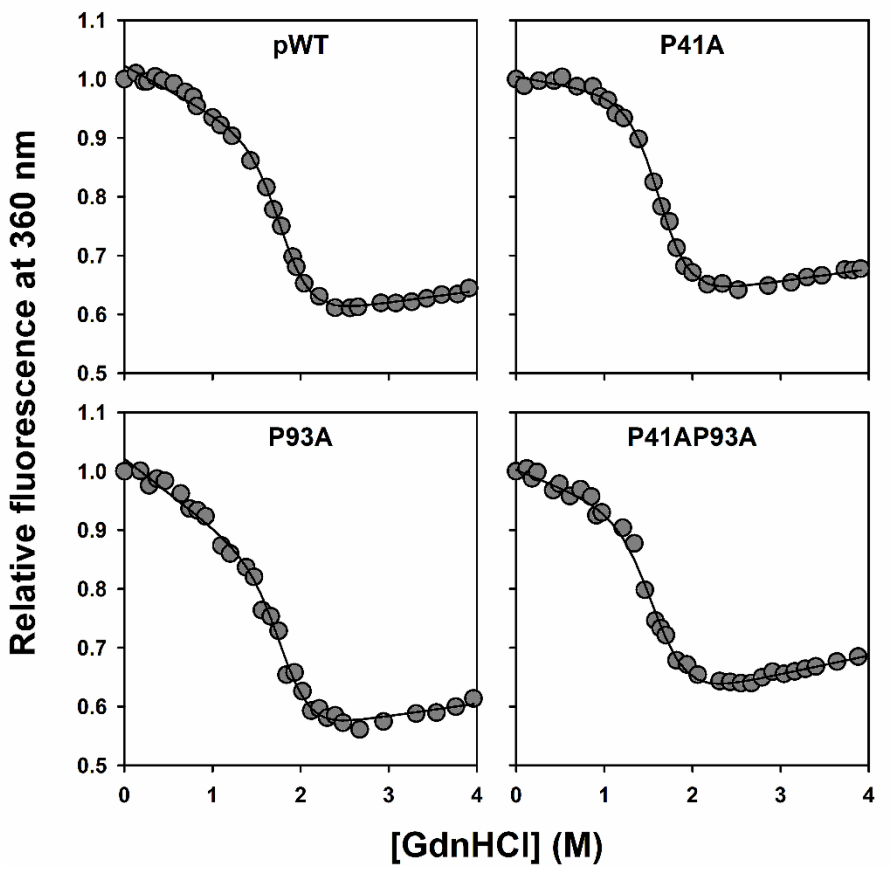


Figure 2.3. Equilibrium unfolding transitions of pWT and the Pro mutant variants. The circles represent the data points, and the solid lines through them are fits to an equation describing a two-state, N↔U unfolding model (Agashe and Udgaonkar 1995). The fits gave values for ΔG_U of 6.3 ± 0.2, 5.4 ± 0.1, 6.2 ± 0.1 and 5.5 ± 0.1 kcal mol^{-1} for pWT, P41A, P93A and P41AP93A, respectively. The mid-points (C_m) of the unfolding transition for pWT, P41A, P93A and P41AP93A are at 1.88 ± 0.04, 1.56 ± 0.02, 1.81 ± 0.02 and 1.61 ± 0.02 M, respectively. The errors in the values of the thermodynamic parameters represent the spread obtained from at least two independent experiments.

The very fast phase of folding is absent for P93A and P41AP93A

The time courses of the increase in Trp19 fluorescence for the four protein variants during folding in two different concentrations of GdnHCl are shown in Figure 2.4. The kinetic folding traces for pWT and P41A were found to fit best to the sum of three exponentials, suggesting that the folding of these proteins in the millisecond to second time domain occur in three distinct kinetic phases: very fast (~ 10 s^{-1}), fast (~ 1 s^{-1}) and slow (~ 0.1 s^{-1}). However, in the case of P93A and P41AP93A, the folding traces fit best to the sum of two exponentials, indicating that folding occurs only in two distinct kinetic phases: fast (~ 1 s^{-1}) and slow (~ 0.1 s^{-1}). Inspection of the residuals for the 'sum of three exponentials' fit *versus* the 'sum of two exponentials' fit for the folding kinetic traces of the four protein variants also highlighted the absence of the very fast phase in the case of P93A and P41AP93A (Figure 2.4). Figure 2.5 shows the dependence of the observed rate constants on GdnHCl, and the variation of the relative amplitudes of the kinetic phases in different GdnHCl concentrations for all four protein variants. The extrapolated rate constants of the three phases of folding in water, and the dependence of the relative amplitudes on GdnHCl concentration of pWT were similar to that of WT, as reported earlier (Patra and Udgaonkar 2007; Bhatia et al. 2019). Similar results were obtained for P41A, P93A and P41AP93A, except that for P93A and P41AP93A, the very fast phase was not observed. The amplitude of the signal change observed in the kinetic folding experiment corresponded well to that observed in the equilibrium unfolding experiment (not shown).

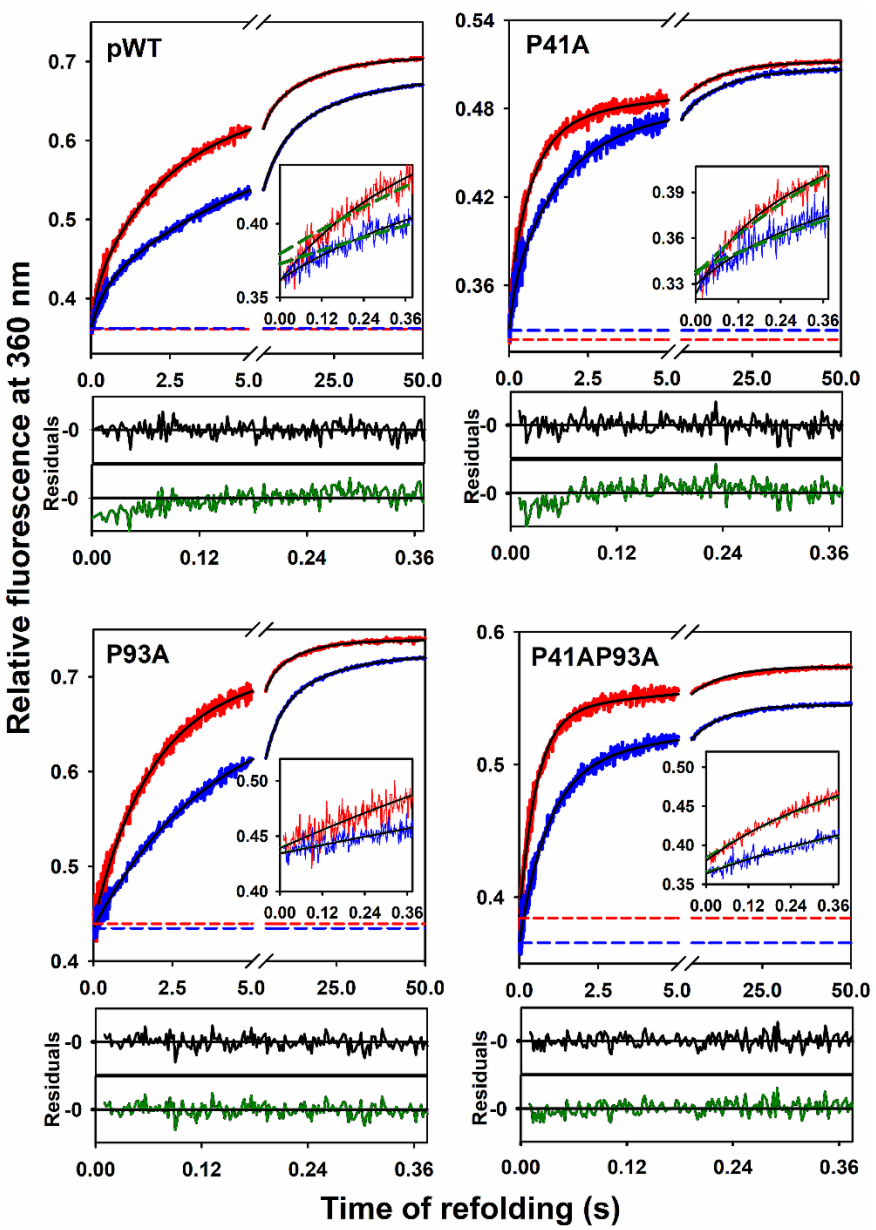


Figure 2.4. Representative folding kinetic traces of pWT and the Pro mutant variants. Relative fluorescence at 360 nm is plotted against the time of refolding. In each panel, the kinetic traces of folding in 0.4 M (red) and 0.6 M (blue) GdnHCl are shown. The extrapolated unfolded protein signals at 0.4 and 0.6 M GdnHCl are shown as red and blue dashed lines, respectively. The fast and slow phases of folding are shown, and the inset in each panel shows the fluorescence change at early times of folding. In the inset and the residuals (for refolding in 0.4 M GdnHCl) below, the dark green and the black colours represent the fit to the 'sum of two exponentials' and the 'sum of three exponentials,' respectively. Inspection of the fits and the residuals shows that the data for pWT and P41A fit well to the sum of three exponentials, with the insets showing the very fast phase of folding. However, the folding traces of P93A and P41AP93A fit well to the sum of two exponentials.

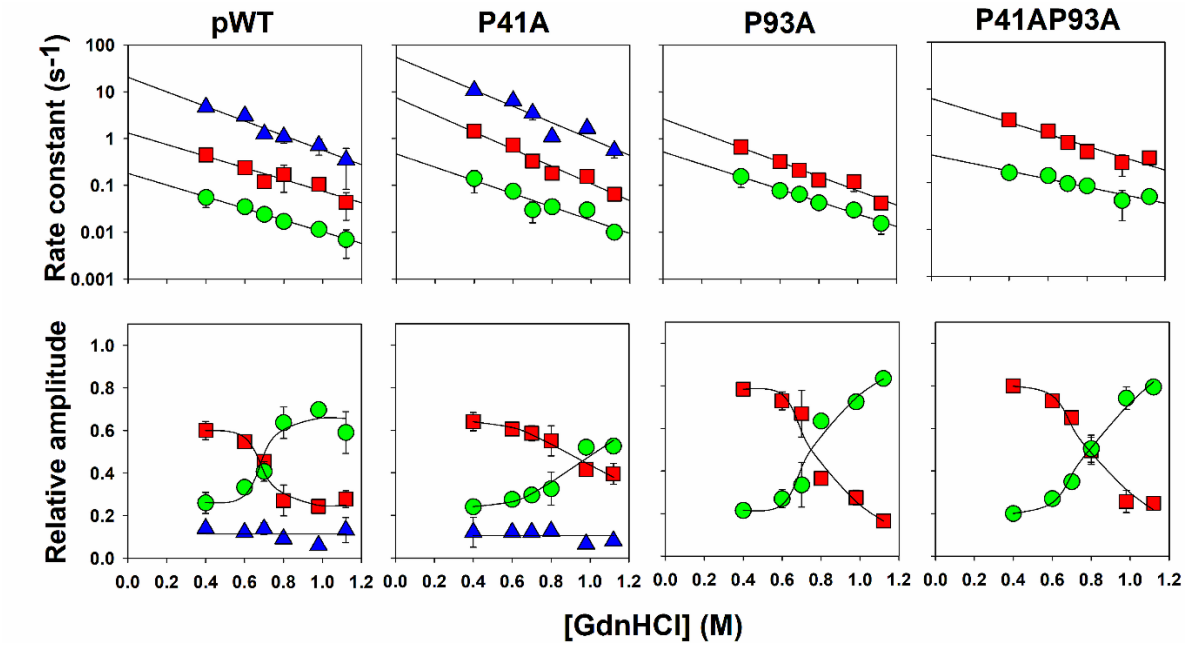


Figure 2.5. Folding kinetics of the MNEI variants. The dependences on GdnHCl concentration of the observed rate constants (top row) and the relative amplitudes (bottom row) for pWT and the Pro mutant variants are shown. The very fast, fast and slow kinetic phases of folding monitored by measurement of the change in Trp fluorescence, are shown by the blue triangles, red squares, and green circles, respectively. The solid lines through the data points were drawn to guide the eye. The error bars represent the spread in the data obtained from at least two independent experiments.

The Pro to Ala mutations lead to overall faster refolding kinetics

Figure 2.6 shows the kinetics of formation of the N state monitored by double-jump, interrupted folding experiments, for all four protein variants. The N state of pWT as well as of P93A formed with the apparent rate constants of 0.0009 ± 0.00001 s^{-1} and 0.0006 ± 0.00007 s^{-1}, respectively. The N states of P41A and P41AP93A formed with rate constants of 0.032 ± 0.002 s^{-1} (~ 35-fold faster than pWT) and 1.1 ± 0.1 s^{-1} (~ 1200-fold faster than pWT), respectively. These results indicate that the slowest and hence, the rate-determining step on the folding pathway of pWT is the last step leading directly to the formation of the N state with a rate constant of 0.0009 ± 0.00001 s^{-1}. A similar rate constant for the formation of the N state was observed in the case of P93A. The rate constant of formation of the N state in the case of P41A and P41AP93A has increased many-fold and appears to lack the rate-determining step seen for pWT and P93A: the mutation of Pro41 to Ala has undoubtedly led to overall faster refolding kinetics. Mutant variants lacking native Pro isomers have been shown to accelerate the (un)folding kinetics for many proteins (Brandts, Brennan, and Lin 1977; Kiefhaber, Grunert, et al. 1990; Wu and Matthews 2003; Roderer et al. 2015). The results shown in Figure 2.6 indicate the involvement of Pro41 at the later stages of the folding of pWT. In the case of other proteins too (Shastry, Agashe, and Udgaonkar 1994; Wedemeyer, Welker, and Scheraga 2002; Kim and Baldwin 1982; Schreiber and Fersht 1993; Brandts, Halvorson, and Brennan 1975; Wu and Matthews 2002a), proline isomerization has been shown to play a role late in folding.

Tertiary contacts are formed in the rate-determining step of refolding

To identify any structural changes accompanying the slowest step on the folding pathway, manual-mixing kinetic studies of folding monitored by far-UV and near-UV CD, were carried out. No change was observed in the far-UV CD monitored kinetics (data not shown), suggesting that the secondary structure of the protein had formed early, and that the slowest step does not involve any structural change at the secondary level. For pWT and P93A, the near-UV CD signal was seen to decrease with rate constants of 0.0008 ± 0.0003 s^{-1} and 0.0005 ± 0.00001 s^{-1}, respectively, suggesting that tight packing of some aromatic amino residues occurs concurrently with the isomerization of Pro41 during the slowest step of refolding. No change in the near-UV CD signal was observed for P41A and P41AP93A (Figure 2.7), on the timescale it was observed in the case of pWT and P93A.

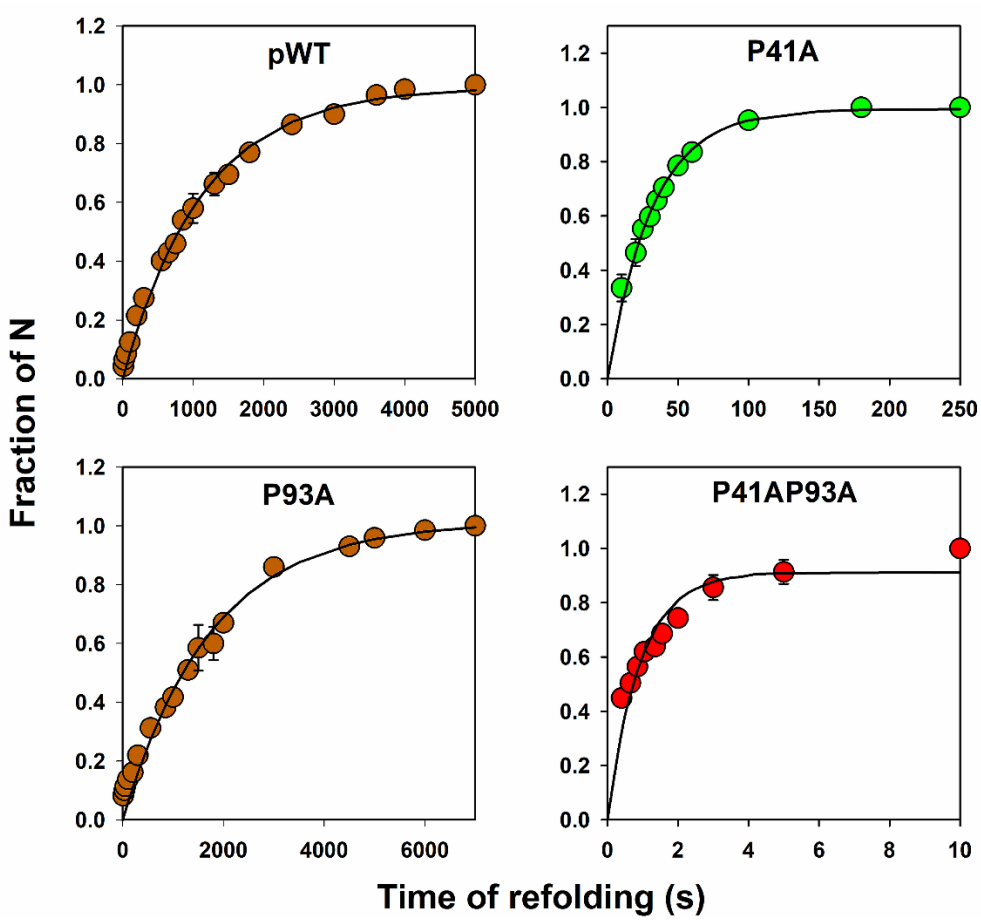


Figure 2.6. Kinetics of formation of the N state. Folding was initiated by effecting a drop in GdnHCl concentration from 4 to 0.4 M. Folding was interrupted at different times, with a jump in GdnHCl concentration from 0.4 to 4.2 M for pWT, P93A and P41AP93A and from 0.4 to 3.5 M for P41A. The solid lines in all the panels describe the formation of N according to a single-exponential process, with apparent rate constants of 0.0009 ± 0.00001 s^{-1}, 0.032 ± 0.002 s^{-1}, 0.0006 ± 0.00007 s^{-1} and 1.1 ± 0.1 s^{-1} for pWT, P41A, P93A and P41AP93A, respectively. The error bars represent the spread in the data obtained from at least two independent experiments.

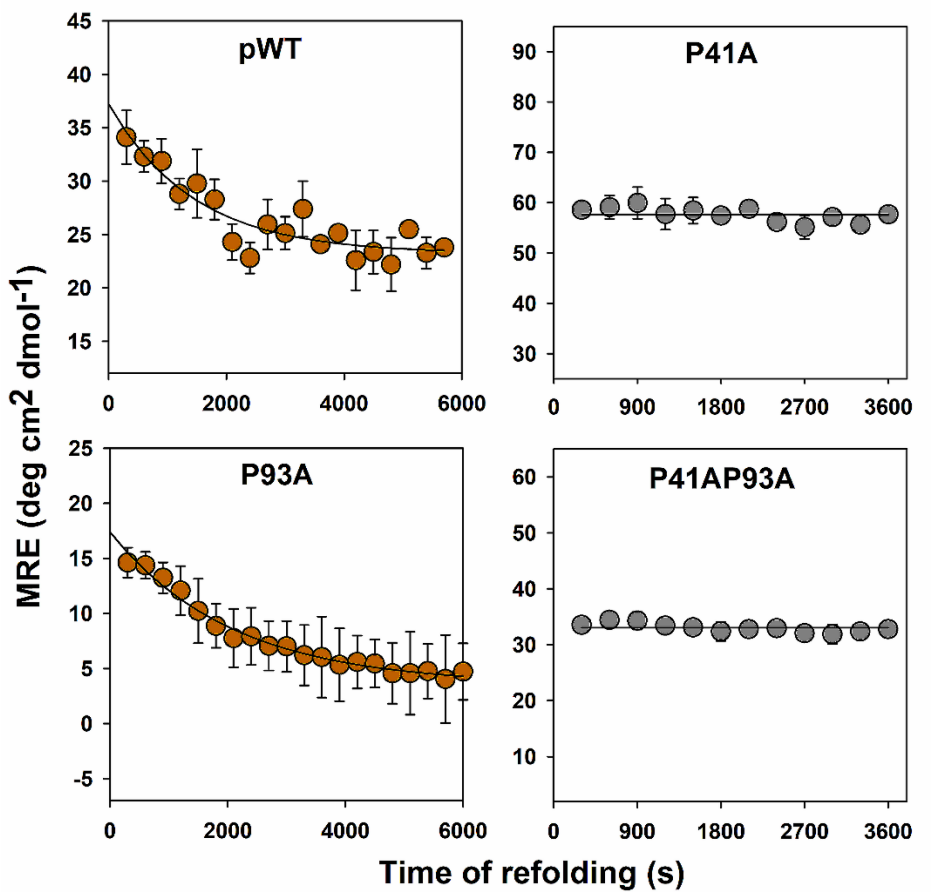


Figure 2.7. Kinetics of formation of N monitored by measurement of near-UV CD at 270 nm for pWT and the Pro mutant variants. The solid lines through the data points for pWT and P93A represent single-exponential processes with apparent rate constants of 0.0008 ± 0.0003 s^{-1} and 0.0005 ± 0.00001 s^{-1}, respectively, whereas the solid lines through the data points for P41A and P41AP93A are drawn to guide the eye. The error bars represent the standard errors obtained from at least two different experiments.

2.5 Discussion

Effect of the Pro to Ala mutations on protein structure and stability

Replacement of a *cis* Pro residue by an Ala residue that adopts the *trans* conformation might be expected to affect the structure of MNEI. While the local environment of Trp19 is affected to only a minor extent, as reflected in small differences in the fluorescence spectra between pWT and the different Pro to Ala mutant variants (Figure 2.2a), changes in secondary structure as reflected in the far-UV CD spectra (Figure 2.2b) appear to be more significant. It is possible that some of the differences in the far-UV CD spectra do not reflect changes in secondary structure but changes in the environment of aromatic residues which can also contribute to the shape of far-UV CD spectra (Woody 1978; Chakrabartty et al. 1993). Such an explanation is probable, as the near-UV CD spectra, for which the major contribution is from aromatic residues, are seen to be affected significantly (Figure 2.2c).

The stabilities of pWT, P41A, P93A and P41AP93A are, however, not very different (Figure 2.3). It is difficult to predict the effect of substituting Pro with another residue. It could be expected that because a Pro residue confers rigidity to the polypeptide chain, substituting it with an Ala residue should entropically stabilize the U state, and hence, reduce protein stability. Moreover, substituting a Pro residue with the smaller Ala should perturb packing interactions and enthalpically destabilize the N state, again reducing protein stability. In the case of MNEI, it also appears that the side-chain of Pro41, which is nearly completely buried in the core of the protein (Figure 2.1a), forms a CH/π interaction with the aromatic side-chain of Tyr64 (Figure 2.8a): the far-UV CD spectra of pWT and P93A, but not of P41A and P41AP93A in which Pro41 is substituted with Ala, show a positive plateau-like hump at 205 nm (Figure 2.2b) that is known to arise from contributions of aromatic residues (Woody 1978; Chakrabartty et al. 1993). The stabilizing CH/π interaction would therefore be absent in P41A. Pro93 is the second residue of the hydrophobic GPVPPP sequence segment at the C-terminal end of MNEI, and appears to have stabilizing interactions with hydrophobic residues, including Phe72, in the β-sheet, which are enabled by it being in the *cis* conformation (Figure 2.8b, c). The stabilizing hydrophobic interactions appear to be absent in P93A (Figure 2.8d). The observation that the Pro93 to Ala mutation barely affects the stability, and that the Pro41 to Ala mutation reduces stability by only 0.9 kcal mol^{-1}, suggests that the destabilizing effect of replacing the *cis* Pro residue must be offset by the stabilizing effect of substituting in the *trans* Ala residue. It seems that the N states of the mutant variants are stabilized by at least one additional hydrogen bond involving the main chain, due to the replacement of Pro with Ala, than is the N state of pWT.

Such stabilization has been observed when Pro has been substituted with other residues in the case of other proteins (Pal and Chakrabarti 1999; Jain et al. 2002; Bajaj et al. 2007).

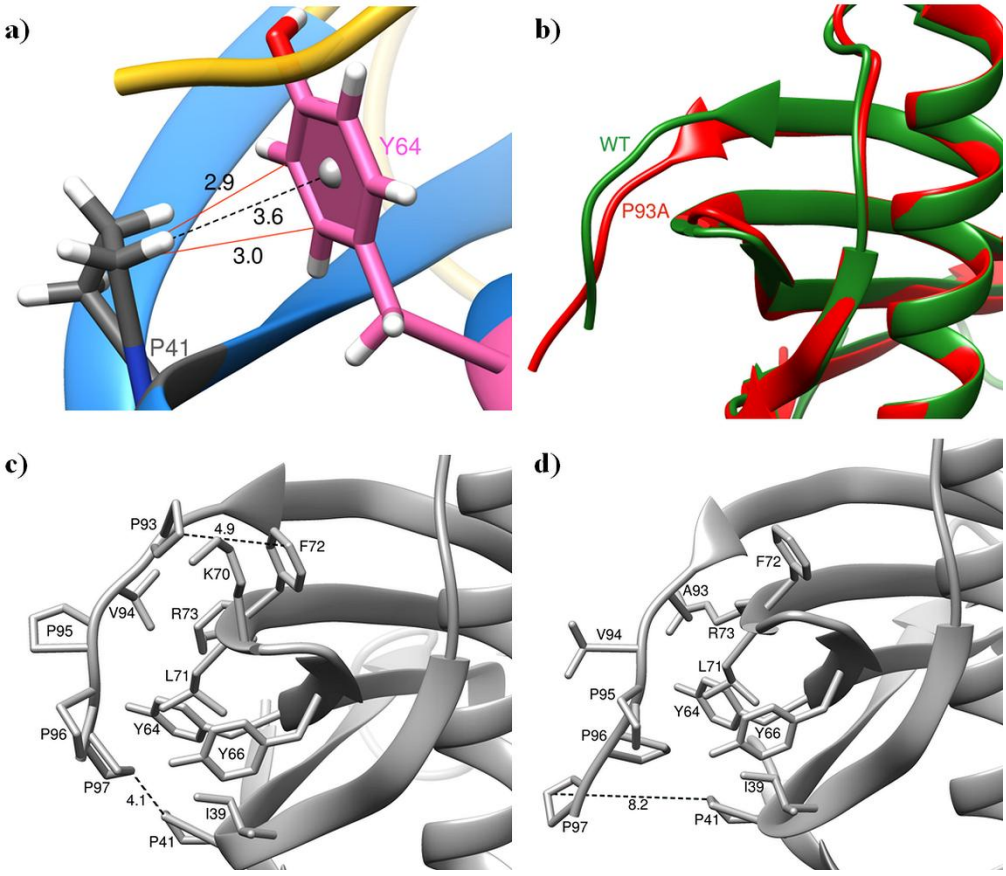


Figure 2.8. Structural interactions of Pro41 and Pro93. a) Structure showing the apparent CH/π interaction between Pro41 (in the β2-strand) and Tyr64 (in the β3-strand). The solid black lines represent the pseudo bonds given by the software (Chimera). The dashed line represents the distance separating the centroid of the Tyr ring to the δ hydrogen of Pro41. b) Structures of MNEI (PDB ID: 1IV7) and P93A (PDB ID: 7EUA) overlapped to highlight the difference at the C-terminus region. The ribbon colors represent the mutant variants as indicated. c) The WT structure highlights the hydrophobic interactions between the residues of the C-terminus 'GPVPPP' sequence with those in the β-sheet. d) Structure of the P93A showing the apparent loss of the hydrophobic interactions upon mutation of Pro93 to Ala. The distance separating Pro41 from Pro97 is seen to increase from 4.1 to 8.2 Å when Pro93 is mutated to Ala. All the distances are shown in Å.

Proline residues and heterogeneity in the unfolded state

The presence of two proline residues, Pro41 and Pro93, in their *cis* conformations in the N state suggests that *cis* to *trans* proline isomerization should lead to the population of at least four distinct conformations in slow equilibrium with each other in the U state. U_{C41C93}, U_{C41T93}, U_{T41C93} and U_{T41T93} are expected to be populated to extents of about 3 %, 14 %, 14 % and 69 %, respectively (Kiefhaber, Kohler, and Schmid 1992; Bradley and Barrick 2005). The equilibria between these four U state conformations would be slow because of the slow rate constant of proline isomerization, which typically falls in the range of 0.01 to 0.001 s^{-1} (Schmid and Baldwin 1979). Since the folding reaction starting from each of these unfolded conformations would be much faster, each conformation would fold *via* an independent pathway, and the flux on each pathway would be determined by the population of the starting unfolded conformation.

The folding of MNEI in the millisecond to second time domain occurs in three kinetic phases - very fast, fast and slow - when monitored by changes in Trp fluorescence, when a single Trp residue is present at either residue position 4 in WT MNEI (Patra and Udgaonkar 2007; Goluguri and Udgaonkar 2015) or residue position 19 in pWT MNEI (Figure 2.4). Double-jump, interrupted folding experiments indicated that the very fast phase of folding arises from a separate population of unfolded molecules (U_1 in Figure 2.1b). These experiments also showed that the fast and slow kinetic phases of folding arise from another population of unfolded molecules (U_2 in Figure 2.1b). The observation that the fraction of molecules folding *via* the very fast phase is about 14 % (Figure 2.4, 2.5), suggests that the very fast phase arises from either U_{C41T93} or U_{T41C93}. Since the very fast phase is eliminated upon mutation of Pro93 to Ala, U_1 must correspond to U_{T41C93}. In the case of other proteins too, it has been shown that mutation of Pro to other residues results in a decrease in U state heterogeneity (Nakano et al. 1993; Kiefhaber, Grunert, et al. 1990; Wu and Matthews 2003; Napolitano et al. 2021).

U_2 would therefore be expected to be a mixture of U_{C41C93}, U_{C41T93}, and U_{T41T93}. U_{C41C93} is expected to populate to too low an extent (~ 3 %) for its folding process to be separately detected. U_{C41T93} is, however, expected to be populated sufficiently for its folding to be separately observable. It is possible that U_{C41T93} is less populated in the U state ensemble, because of the presence of destabilizing interactions that are present when the Arg40-Pro41 bond is in the *cis* conformation and which are absent in the *tran*s conformation. It is known that the ratio of *cis* to *trans* isomers can depend not only on the residue preceding Pro (MacArthur and Thornton 1991; Reimer et al. 1998) and on the initial denaturing conditions

(Houry, Rothwarf, and Scheraga 1994; Pertinhez et al. 2000), but also on the local environment (Grathwohl and Wüthrich 1981). In any case, U_2 appears to consist predominantly of U_{T41T93}.

cis Pro93 enables faster folding than *trans* Pro93

Previously, the very fast phase of folding had been attributed to the folding of U_1, and the fast and slow phases of folding were attributed to the folding of U_2 (Figure 2.1b) (Patra and Udgaonkar 2007). The current study validates this aspect of the folding scheme shown in figure 2.1b: the mutation of Pro93 to Ala eliminates only the very fast phase of folding and has no effect on the fast and slow phases (Figure 2.5).

The current study also suggests that when Pro93 is in the *cis* conformation, the folding of MNEI is faster because some stabilizing interactions form early. It is possible that these stabilizing interactions are the same hydrophobic interactions that appear to be present in the N state of MNEI (see above), which appear to be enabled by the conformational rigidity provided by the *cis* Gly92-Pro93 bond. In the case of ribonuclease A, an early intermediate stabilized by hydrophobic interactions is populated on the very fast pathway originating from an unfolded form having all Pro residues in their native conformations (Houry and Scheraga 1996). In the case of yeast phosphoglycerate kinase too, a native *cis* Pro residue speeds up folding (Osváth and Gruebele 2003). A recent study has shown that Pro residues can facilitate the compaction of disordered protein chains (Mateos et al. 2020), in agreement with a much earlier study that showed that native isomers of certain Pro residues can facilitate the association of residues on folding pathways *via* hydrophobic interactions (Lewis, Momany, and Scheraga 1973). Specific hydrophobic clustering mediated by Pro during the initial folding steps is also known for CRABP1 (Eyles and Gierasch 2000). In the case of MNEI, a collapsed globule having solvent-exposed hydrophobic patches is known to form very early during folding (Goluguri and Udgaonkar 2016), and if stabilizing hydrophobic interactions are absent early during the folding of U_2, just as they appear to be absent in the N state of the P93A variant in which Ala93 is in the *trans* conformation (Figure 2.8d), that could be the reason why the folding of U_2 occurs more slowly than that of U_1 in the case of pWT. For several proteins whose folding is accompanied by *trans* to *cis* Pro isomerization, it has been shown that the unfolded form having the non-native *trans* Pro conformation folds slower than the unfolded form having the native *cis* Pro conformation (Napolitano et al. 2021; Nakano et al. 1993; Kiefhaber, Quaas, et al. 1990b), but this is not always the case (Schreiber and Fersht 1993; Shastry, Agashe, and Udgaonkar 1994; Shastry and Udgaonkar 1995).

Proline isomerization during folding

Structure formation reactions during protein folding, including chain collapse (Haran 2012; Udgaonkar 2013) and secondary structure formation (Udgaonkar 2008) are fast compared to the time-scale of proline isomerization. It is therefore not surprising that for many proteins that contain *cis* Pro residues in their native conformations, *trans* to *cis* isomerization occurs as the last step in folding (Wedemeyer, Welker, and Scheraga 2002; Wu and Matthews 2002a; Kelley and Richards 1987; Schreiber and Fersht 1993; Veeraraghavan, Nall, and Fink 1997; Pappenberger et al. 2003). In many (Schreiber and Fersht 1993; Pappenberger et al. 2003; Kelley and Richards 1987) but not all (Wu and Matthews 2002a; Schmid et al. 1986; Veeraraghavan, Nall, and Fink 1997) cases, the *trans* to *cis* isomerization step appears to be silent to the commonly used (Figure 2.4) spectroscopic probes, indicating that major structure acquisition precedes the proline isomerization step. For example, in the case of MNEI, the very fast, fast and slow phases of folding can be monitored by measuring changes in fluorescence or CD (Patra and Udgaonkar 2007; Goluguri and Udgaonkar 2015), but not the very slow phase whose rate constant corresponds to that of proline isomerization. The very slow phase can only be monitored by measurements of ANS fluorescence (Goluguri and Udgaonkar 2016; Patra and Udgaonkar 2007) and near UV CD (Figure 2.7a), as well as by double-jump, interrupted folding experiments (Patra and Udgaonkar 2007). No lag phase is observed in the formation of the N. A possible reason for this could be the presence of multiple folding pathways, with native molecules being formed more rapidly along the folding pathway arising from the U_{C41C93} sub-ensemble, which is populated to 3-5% (discussed above). Proline isomerization would not be rate-limiting during the folding of the U_{C41C93} sub-ensemble.

Isomerization of Pro41. In the case of pWT, the folding of U_2 to N must be accompanied by the *trans* to *cis* isomerization of both Pro41 and Pro93. The very slow phase of folding of pWT, which leads to the formation of N, is about 3-fold slower than that of WT MNEI. In the latter case, it had been shown previously that the very slow phase of folding is independent of denaturant concentration (Patra and Udgaonkar 2007). In the case of many other proteins, it has been shown that folding steps involving proline isomerization reactions can be independent of denaturant concentration (Shastry, Agashe, and Udgaonkar 1994; Cook, Schmid, and Baldwin 1979; Ogasahara and Yutani 1997). Hence, the very slow phase of the folding of WT MNEI had been attributed to proline isomerization (Patra and Udgaonkar 2007). The observation, from double-jump, interrupted refolding experiments, that the N state forms with a rate constant of about 0.0009 ± 0.00001 s^{-1} in the case of pWT, and that the rate constant of

the very slow phase is about the same for the P93A variant (Figure 2.6), suggests that this very slow phase of folding must correspond to the *trans* to *cis* isomerization of Pro41. This conclusion is validated by the observation that when Pro41 is mutated to Ala in P41A and P41AP93A, the very slow phase of folding is abolished (Figure 2.6), and the N state forms much more rapidly.

In the N state of pWT and WT MNEI, Pro41 introduces a sharp bend of about 90^{O} in the second β-strand, which allows tight packing between the β-sheet and the helix of MNEI (Figure 2.1a). The side-chain of Pro41 is itself nearly completely buried. It is likely that during the folding of pWT, Pro41 remains solvent-accessible in its *trans* conformation until the last step of folding, when the Arg40-Pro41 bond undergoes *trans* to *cis* isomerization, and the side-chain of the *cis* Pro41 isomer gets inserted into the hydrophobic core of the protein. This would require that at least a few of the hydrophobic core residues remain solvent-accessible until the last step. It is therefore not surprising that the hydrophobic dye ANS can remain bound to folding MNEI molecules until the commencement of the very slow phase of folding (Patra and Udgaonkar 2007; Goluguri and Udgaonkar 2016). It has been shown, both in the case of WT MNEI (Goluguri and Udgaonkar 2016; Patra and Udgaonkar 2007) and pWT (data not shown), that ANS becomes completely unbound only during the very slow phase of folding that leads directly to the formation of N. It is also not surprising that the very slow phase of the folding of both pWT and P93A, involves a final consolidation of tertiary structure measurable by a change in near-UV CD, given that it involves accommodation of the side-chain of Pro41 in the hydrophobic core of the protein. Not surprisingly, no change in near-UV circular dichroism is detectable in the last step leading to the formation of N in the case of the P41A and P41AP93A variants in which Pro41 is replaced by Ala (Figure 2.7). Furthermore, it is also not surprising to find that the *trans* to *cis* isomerization of Pro41 occurs in the compact and highly structured, native-like intermediate N* (Bhatia, Krishnamoorthy, and Udgaonkar 2021a) as it transits to N (Figure 2.1b). Indeed, proline isomerization has been found to occur in native states of proteins (Hodel et al. 1995; Andreotti 2003; Mukaiyama et al. 2013; Schmidpeter and Schmid 2015).

Isomerization of Pro93. It was possible that *trans* to *cis* isomerization of Pro93 occurs with the same very slow rate constant as that of *trans* to *cis* isomerization of Pro41 during the folding of pWT. Such a possibility could be discounted in the current study, which shows that the very slow phase of folding is absent in the case of the P41A variant which possesses *cis* Pro93 in its N state. The observed rate constant of the slow phase of folding (Figure 2.4) is within the range of rate constants with which proline isomerization can occur during folding. The observation

from double-jump, interrupted refolding experiments that the N state of P41A forms during the slow phase of folding indicates that *trans* to *cis* isomerization of Pro93 must occur during the slow phase of folding of pWT, as well as of P93A.

Hence, the *trans* to *cis* isomerization reactions of the Gly92-Pro93 and Arg40-Pro41 peptide bonds occur sequentially during the folding of pWT, with the former occurring during the slow phase, and the latter during the very slow phase of folding. It is perhaps not surprising that the *trans* to *cis* isomerization of the Arg40-Pro41 bond is slower than that of the Gly92-Pro43 bond, as it is known that when a bulky residue (such as Arg) immediately precedes a Pro residue, isomerization is retarded in comparison to when a smaller residue (such as Gly) is the preceding residue (Grathwohl and Wüthrich 1981).

More likely, the *trans* to *cis* isomerization reaction of Pro93 is speeded up significantly because it is accompanied by substantial structure formation: the slow phase is known to lead to substantial formation of secondary and tertiary structure (Patra and Udgaonkar 2007; Goluguri and Udgaonkar 2016). The formation of stabilizing structure would stabilize the transition state that slows down *trans* to *cis* isomerization of Pro93. In contrast, the *trans* to *cis* isomerization reaction of Pro41 is not accompanied by any secondary structure formation, and is accompanied by the formation of only minor tertiary structure. In the case of both staphylococcal nuclease (Maki et al. 1999) and CRABP1 (Burns-Hamuro, Dalessio, and Ropson 2004), a kinetic phase known to involve proline isomerization was not abolished upon replacing the Pro residue, indicating that the kinetic phase also involved structure formation. For other proteins too, proline isomerization has been shown to be faster when accompanied by structure formation (Jackson and Fersht 1991b; Texter et al. 1992; Plaxco et al. 1996), although it is also possible for structure in an intermediate to slow down proline isomerization (Kiefhaber, Quaas, et al. 1990b). It appears that in the case of MNEI, the slow phase of structure formation and the *trans* to *cis* isomerization reaction of Pro93 occur with about the same rate constant, as the replacement of the *cis* Gly92-Pro93 bond with the *trans* Gly92-Ala93 bond in P93A does not significantly affect the rate constant of the slow phase of folding.

Simplification of the folding mechanism upon mutation of the Pro residues

The current study indicates that in the case of pWT, *trans* to *cis* isomerization of Pro93 occurs during the slow phase leading to the formation of the native-like intermediate, N* (Figure 2.1b) and that *trans* to *cis* isomerization of Pro41 occurs during the very slow phase corresponding to the transition of N* to N. The current study also shows that the U state heterogeneity seen in the case of pWT originates in the *cis* to *trans* isomerization of Pro93.

Figure 2.1b also shows that the fast and slow phases of folding arise because of the transient population of intermediates, and it is known that these intermediates are structurally distinct in having different structural segments collapsed (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). Thus, it could be expected that mutations might differentially affect the stabilities of the folding intermediates, and hence determine which intermediates are populated sufficiently to be observable. Hence, it would not be surprising if the mutations of Pro41 and Pro93 to Ala, singly as well as together, lead to apparently simpler folding mechanisms.

Folding mechanism of the P41A variant. Since U_1 is populated, the very fast phase is observed when folding is monitored using the fluorescence of Trp19 as the probe. The fast and slow phases of folding are also observed. The Pro41 to Ala mutation appears to destabilize the native-like intermediate, N*, seen in the case of pWT so that it is populated to a negligible extent. The very slow phase of folding is absent, and the N state appears to form as a product of the slow phase of folding. The very fast, fast and slow phases of fluorescence change are observed. *Trans* to *cis* isomerization of the Gly92-Pro93 peptide bond occurs during the slow phase of folding. Thus, the folding mechanism of pWT (Figure 2.1b) simplifies upon mutation of Pro41 to Ala (Figure 2.9a).

Folding mechanism of the P93A variant. Since U_1 is absent, the very fast phase of folding is not observed. Only the fast and slow phases are observed when folding is monitored using the fluorescence of Trp19 as the probe. *Trans* to *cis* isomerization of the Arg40-Pro41 peptide bond occurs during the very slow phase of folding as N* transitions to N. Thus, the folding mechanism of pWT (Figure 2.1b) simplifies upon mutation of Pro93 to Ala (Figure 2.9b).

Folding mechanism of the P41AP93A variant. U_1 is absent; hence, the very fast phase is not observed when folding is monitored using Trp fluorescence as the probe. Only the fast and slow phases of folding are observed. Folding is no longer limited by *trans* to *cis* proline isomerization as both the *cis* Pro41 and *cis* Pro93 have been replaced by Ala residues which adopt *trans* conformations. N forms *via* two pathways on which the slow folding intermediates differ in having different structural segments compacted (Bhatia, Krishnamoorthy, and Udgaonkar 2021a), which would result in them differing in their stabilities. It appears that on one of these pathways, the slow folding intermediate is destabilized sufficiently by the mutations so that it is no longer populated significantly; hence, the N state forms during the fast phase of folding on that pathway. On the other pathway, N forms during the slow phase of folding. The observed rate constant for the formation of N, about 1.1 s^{-1}, is the sum of the

amplitude-weighted rate constants for the formation of N along the two pathways. Thus, the folding mechanism of pWT (Figure 2.1b) simplifies significantly upon replacement of both the native-cis Pro to Ala (Figure 2.9c).

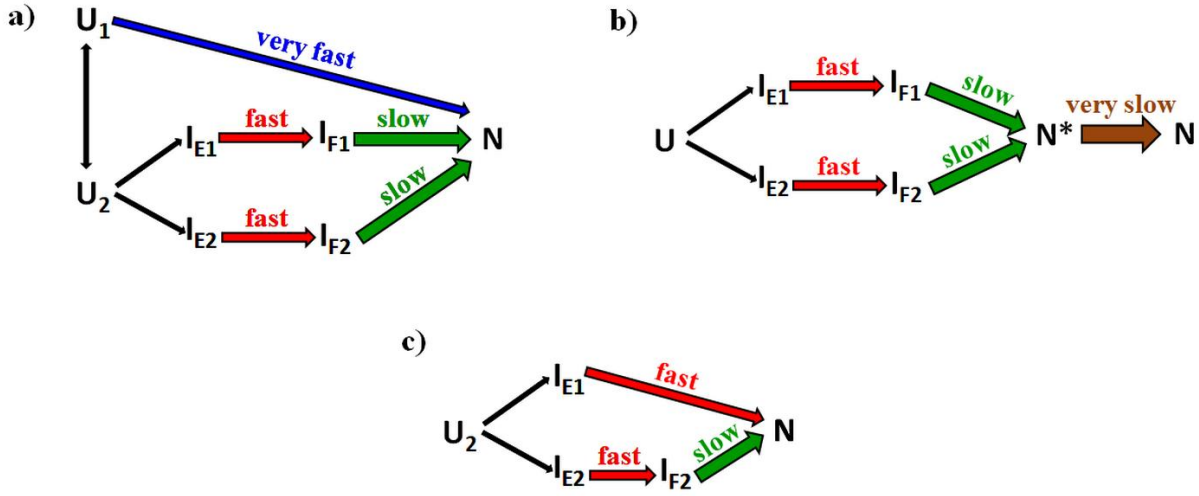


Figure 2.9. Simplification of the complex folding mechanism of MNEI achieved upon replacement of the native-*cis* Pro by Ala. The folding schemes of (a) P41A, (b) P93A, and (c) P41AP93A.

Comparison of the folding mechanisms of the single chain and double chain variants of monellin. The folding mechanism of naturally occurring double-chain monellin (dcMN) appears to be simpler than that of its artificially created single chain counterpart, MNEI (Patra and Udgaonkar 2007; Bhattacharjee and Udgaonkar 2022). While the structures of the two variants are virtually identical, it was possible that the increased complexity observed for the folding of MNEI was due to changes in enthalpy as well as lower chain entropy caused by the covalent linkage of the two chains in MNEI. The current study suggests that this is not the case. The folding mechanism of MNEI has been shown to be complex because of proline isomerization, and it reduces to a much simpler mechanism similar to that of dcMN upon mutation of the two *cis* Pro residues to Ala. What remains surprising is why the same two Pro residues, which are also *cis* in dcMN do not introduce similar complexity into the folding

mechanism of dcMN. Future studies will investigate the role of proline isomerization in the folding of dcMN.

Proline isomerization and protein folding: lessons from MNEI

The current study has highlighted several aspects of proline isomerization reactions coupled to protein folding, which need to be taken into consideration in any protein folding study: A) It is invariably assumed, when parallel pathways of folding are observed for a protein containing Pro residues, that their origin lies in the heterogeneity arising because of the presence of the Pro residues. The observation in the case of MNEI that competing pathways remain even when both Pro residues are replaced, suggests that the complexity of folding of Pro-containing proteins can arise for reasons other than proline isomerization. B) Proline isomerization is known to occur in the unfolded state (Kim and Balwin 1982; Eberhardt, Loh and Raines 1993; Wedemeyer, Welker and Scheraga 2002), and substitution of Pro with other residues invariably results in the abolishment of slow phases of folding. It seems now that fast phases of folding may also be abolished: the mutation of Pro93 to Ala leads to complete abolishment of the very fast phase of folding of MNEI (Figure 2.9b). C) Proline isomerization reactions are known to occur late during the course of protein folding (Brandts, Halvorson and Brennan 1975; Shastry, Agashe and Udgaonkar 1994; Kim and Balwin 1982; Wedemeyer, Welker and Scheraga 2002; Schreiber and Fersht 1993; Wu and Matthews 2002a), and it is therefore not surprising that in the case of MNEI, the very slow phase of folding is abolished upon mutation of Pro41 to Ala (Figure 2.9a). What is surprising is that Pro to Ala mutations may differentially affect the stabilities of folding intermediates in a manner that leads to the abolishment of a slow folding phase on one pathway but not a competing pathway (Figure 2.9c). D) When multiple *cis* Pro residues are present in a protein, their *trans* to *cis* isomerization reactions during folding may occur not concurrently but sequentially with very different rate constants. E) Pro residues may play a role in preventing aggregation of folding intermediates. In the case of MNEI, substitution of Pro41 with Ala leads to aggregation during folding at higher (> 60 μM) protein concentrations at which the wt protein is fully soluble.

2.6 Conclusion

The Pro to Ala mutations simplify the folding mechanism of MNEI in multiple ways. By reducing unfolded state heterogeneity, the Pro93 to Ala mutation results in the elimination of a folding pathway. The Pro41 to Ala mutation results in the elimination of the slowest proline isomerization step, in the destabilization of the late native like intermediate, N*, and in the

faster formation of N. Together, the Pro41 to Ala and Pro93 to Ala mutations result in a simpler folding mechanism in which elimination of the proline isomerization steps, and destabilization of folding intermediates lead to more than a thousand-fold acceleration of folding.

Chapter 3

Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding

Reproduced with permission from:

Kaushik, A. and Udgaonkar, J.B., 2026. Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding. *Protein Science,* 35(3), p.e70512.

3.1 Significance Statement

Chain compaction and structure formation are central features of protein folding, yet how these two processes are coupled remains unclear. Although early chain collapse can precede structure formation, the two often appear coupled at later stages, raising the question of whether this coupling is obligatory or tunable. Here, the coupling is shown to be tunable. Using a quantitative framework that resolves coexisting compact intermediates, this study demonstrates that proline-imposed backbone rigidity governs intermediate-state heterogeneity, which in turn determines the separability of these processes.

3.2 Introduction

The folding of a protein is a complex process (Ferguson and Fersht 2003; Udgaonkar 2008; Bhatia and Udgaonkar 2022), involving both compaction of the polypeptide chain (Agashe, Shastry, and Udgaonkar 1995; Ballew, Sabelko, and Gruebele 1996; Nath and Udgaonkar 1997; Chen et al. 1998; Hagen and Eaton 2000; Qiu, Zachariah, and Hagen 2003; Magg and Schmid 2004; Sinha and Udgaonkar 2007) and the formation of secondary and tertiary structure (Parker et al. 1997; Raschke and Marqusee 1997). The two processes are intimately linked, and seem to proceed together for the most part, but not much is known about how they are coupled to each other. Chain compaction favors the formation of secondary structure which is itself compact (Ballew, Sabelko, and Gruebele 1996; Zhou and Dill 2001; Roder 2004), and the coalescence of secondary structural units also induces compaction (Karplus and Weaver 1994; Scott et al. 2006). Different extents of chain compaction and structure formation occur during the early and late stages of the folding reaction and understanding how these changes are linked mechanistically to each other, remains an important challenge.

For many proteins, folding appears to begin with a collapse of the polypeptide chain, with (Uzawa et al. 2004; Kimura, Akiyama, et al. 2005; Bartlett and Radford 2010; Fazelinia et al. 2014; Peran et al. 2019) or without (Dasgupta and Udgaonkar 2010; Chan et al. 1997; Qi, Sosnick, and Englander 1998; Magg et al. 2006) the concurrent formation of structure. In the case of several proteins, different structural segments appear to undergo significantly different extents of initial collapse, as well as of later contraction (Sinha and Udgaonkar 2005, 2007; Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Maity and Reddy 2018). Moreover, when initial collapse occurs, it does not happen to the same extent in all molecules (Roh et al. 2010; Aghera and Udgaonkar 2013; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021b). Hence, understanding how compaction is coupled to structure formation is

complicated because the product of initial collapse is heterogeneous with sub-populations of molecules that have undergone substantial differential compaction across multiple structural segments, in dynamic equilibrium with sub-populations that have undergone only minor compaction (Bhatia, Krishnamoorthy, and Udgaonkar 2021b). Delineating the heterogeneity inherent in the initial collapse reaction, as well as in the subsequent chain contraction that occurs as structure develops, is important for understanding how folding begins and proceeds.

Here, initial chain collapse and further chain contraction of single chain monellin (MNEI), a well-established model for both experimental (Kimura, Uzawa, et al. 2005; Jha et al. 2009; Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a, 2021b; Kaushik and Udgaonkar 2023) and computational (Maity and Reddy 2018; Mascarenhas, Terse, and Gosavi 2018) studies of protein folding, have been studied. The folding of MNEI is known to begin with a non-specific chain collapse complete within 37 μs, which reduces the average diameter by approximately 30%, and leads to the formation of some non-native interactions (Goluguri and Udgaonkar 2016). Subsequent stepwise consolidation of the collapsed globule occurs during the first millisecond of folding, without a further decrease in average dimensions, and results in the formation of a kinetic molten globule intermediate. At this stage, no significant secondary structure has formed, and even by 100 ms of folding in strongly native-like conditions, only about 10 % of N-like secondary structure is observed (Goluguri and Udgaonkar 2015; Bhatia et al. 2019).

Time-resolved fluorescence resonance energy transfer (trFRET) measurements, analyzed by the Maximum Entropy Method (MEM) (Lakshmikanth et al. 2001; Jha et al. 2009; Bhatia, Krishnamoorthy, and Udgaonkar 2021a, 2021b), not only are able to distinguish but can also quantify sub-populations of molecules present in different conformations. This methodology had been used earlier to monitor distances spanning different segments of the protein: the core (C), the β-sheet (B), the end-to-end distance (E), and the sole helix (H) of the protein (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). The H segment spanning the helix contracted gradually during folding. The distances spanning the C, B and E segments exhibited bimodal distributions at 100 milliseconds of folding, corresponding to two distinct sub-populations, unfolded (U)-like and native (N)-like, representing different extents of chain compaction (Bhatia, Krishnamoorthy, and Udgaonkar 2021a, 2021b). The dependence of spatial separation on sequence separation indicated that the U-like sub-ensemble at 100 ms was random-coil-like with non-specific intra-chain interactions, while the N-like sub-ensemble had

at least some specific native-like intra-chain contacts (Bhatia, Krishnamoorthy, and Udgaonkar 2021b). Analysis of the kinetic evolution of the U-like to the N-like sub-ensemble for the B, C and E segments (Bhatia, Krishnamoorthy, and Udgaonkar 2021a) led to a mechanism for chain contraction (Figure 3.1c), which closely resembled the previously established mechanism for folding (Patra and Udgaonkar 2007). This suggested that, following the initial chain collapse at 100 ms, chain contraction and structure formation proceed concurrently during the folding of MNEI.

A key question is whether chain compaction can be separated temporally from structure formation steps at later stages of folding, just as initial chain compaction is decoupled from structure formation in the case of MNEI (Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016) as well as other proteins (Dasgupta and Udgaonkar 2010; Chan et al. 1997; Qi, Sosnick, and Englander 1998; Magg et al. 2006). It is likely that sequence-encoded structural constraints such as those imposed by proline residues, modulate the coupling of chain contraction and structure formation not only during initial collapse but also at later stages of the folding reaction. In the case of MNEI, an earlier study had shown that mutation of the two proline residues (Pro41 and Pro93), which are *cis* in the N state, to alanine, simplified the folding landscape, and had indicated that chain contraction might also be modulated (Kaushik and Udgaonkar 2023).

In the current study, site-specific Pro to Ala mutations were introduced at the two native-*cis* prolines (Pro41 and Pro93) of MNEI, and FRET pairs were designed to monitor compaction of distinct structural segments (Figure 3.1). The effect of the P41A mutation on compaction of the C segment was monitored using measurements of FRET between Trp19 and Cys42-TNB in both the WT protein (WT_C) and the P41A mutant variant (P41A_C) (Figure 3.1a). Similarly, the effect of the P93A mutation on compaction of the E segment was monitored using measurements of FRET between Trp4 and Cys97-TNB in both the WT protein (WT_E) and the P93A mutant variant (P93A_E). It should be noted that the FRET acceptor TNB is a small and charged moiety compared to the bulky and highly hydrophobic moieties (Hiranmay and Govardhan 2016; Riback et al. 2019) that have been used as probes in other FRET studies. Because Pro41 and Pro93 are located in the proximity of the core (C segment) and the end-to-end (E segment) regions, respectively, their mutation has enabled site-specific probing of how proline residues influence local chain contraction at the C and E segments during folding. Distributions of intramolecular distances spanning the C and E segments and their contraction at different times of folding were characterized using MEM-coupled tr-FRET

measurements. Folding itself was monitored using measurements of the changes in the fluorescence of Trp19 and Trp4 during the folding of the Trp19-containing and Trp4-containing protein variants, respectively. The measurements have provided mechanistic insight into the role of the proline residues in the contraction of protein regions in their proximity. Importantly, they have revealed that backbone rigidity can tune the coupling between chain compaction and structure formation during all stages of folding.

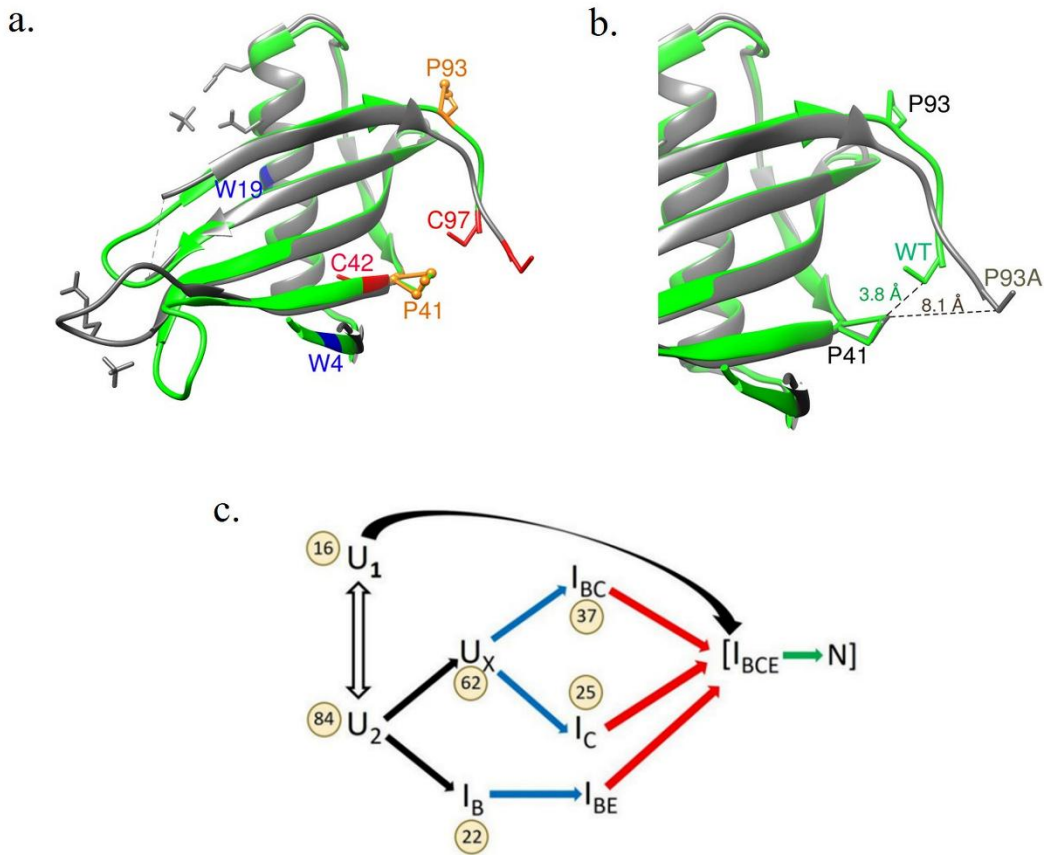


Figure 3.1. Structure and mechanism of folding/contraction of single-chain monellin (MNEI). a) The structure of WT MNEI (PDB ID: 1IV7), shown in green, has been superimposed on the structure of the P93A mutant variant (PDB ID: 7EUA), shown in grey. The residues used for monitoring FRET, and the locations of the two native *cis*-Pro residues (Pro41 and Pro93, in orange) that were mutated to Ala, are shown on the structure of the WT protein. Trp4 and Trp19, which served as FRET donors, and Cys42 and Cys97, which served as FRET acceptors, are marked in blue and red, respectively. FRET was measured for the Trp19-Cys42TNB pair, for mapping the central core (C) segment, and for the Trp4-Cys97TNB pair, to map the end-to-end distance (E) segment. Structural differences between the WT and mutant proteins can be seen at the C-terminal region and at residues 49-54, while residues 80-84 are missing in the mutant protein structure. b) Structural differences at the C-terminal region are highlighted, particularly in its proximity to Pro41 in the β2 strand. The protein structures shown in panels a and b were drawn using Chimera. The distances between the side-chains are shown in Å. c) Scheme showing chain contraction during folding (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). The black, blue, red and green arrows represent the unobservable (complete within 100 ms), fast, slow and very slow kinetic phases of folding/contraction. The subscripts on the intermediate (I) denote the segments (B, C and/or E) which have contracted to N-like dimensions in that intermediate. The numbers in the circles indicate the percentage of molecules following a given pathway.

3.3 Materials and Methods

In this study, intramolecular FRET changes at two specific segments of MNEI were investigated. For FRET measurements, a Trp residue served as the donor fluorophore, while a Cys residue was covalently linked to the thionitrobenzoate (TNB) moiety, acting as the FRET acceptor. The two segments studied were the core (C) segment, spanned by Trp19-Cys42TNB FRET pair, and the end-to-end (E) segment, spanned by Trp4-Cys97TNB FRET pair. The effect of the Pro41 to Ala mutation on changes in compaction of segment C and the effect of Pro93 to Ala mutation on changes in the end-to-end distance (segment E) were studied. Of the six Pro residues in MNEI, only Pro41 and Pro93 adopt *cis* conformations in the native (N) state (Kaushik and Udgaonkar 2023).

For each FRET pair, fluorescence measurements were carried out on both the unlabeled and the corresponding TNB-labeled variants. FRET was assessed using both steady-state and time-resolved measurements. In the case of steady-state FRET measurements, Trp fluorescence was measured in the absence (donor only) and presence of the acceptor moiety, TNB (donor–acceptor), with a dead time of approximately 10 ms. For time-resolved FRET measurements, fluorescence decay curves were collected either as a function of GdnHCl concentration or as a function of folding time using a double-kinetics setup, both in the absence and presence of the FRET acceptor. Because fluorescence intensity decays occur on a nanosecond timescale, much faster than the timescale of conformational fluctuations, time-resolved FRET enables monitoring of changes in the distribution of donor–acceptor distances within the population of molecules as a function of folding time, as demonstrated previously (Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). Information on FRET efficiency, average intramolecular distances, and fluorescence lifetime distributions was extracted from the decay curves using established methods (Jha et al. 2009; Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

Proteins and Reagents

The expression, purification, and labeling of the mutant variants used in this study have been described previously (Kaushik and Udgaonkar 2023; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). All the experiments were carried out at pH 8.0 and 25 °C. The details of the reagents used in the experiments have been reported earlier (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

Site-directed Mutagenesis

The Pro41 to Ala mutation was introduced in the mutant variant W19C42, as described in a previous study (Kaushik and Udgaonkar 2023). The Pro93 to Ala mutation was introduced in the mutant variant W4C97, using the QuickChange site-directed mutagenesis method developed by Stratagene.

Measurement of fluorescence and far UV CD spectra

Fluorescence spectra were measured using a Fluoromax 4 (Horiba) spectrofluorometer. Protein samples were excited at 295 nm, and emission spectra were recorded from 305 to 450 nm with a response time of 1 s. The excitation and emission slit widths were set to 1 nm and 5 nm, respectively. The protein concentration was 10 μM. Each spectrum represents the average of three independent emission scans.

Far-UV circular dichroism (CD) spectra were obtained using a Jasco J815 spectropolarimeter and a cuvette with a 0.1 cm path length. Data were collected with a 1 nm bandwidth, a response time of 1 s, and a scan speed of 20 nm min^{-1}. The protein concentration was 10 μM. Each spectrum was generated by averaging twenty individual wavelength scans.

Steady-state and time resolved FRET-monitored equilibrium unfolding experiments

For GdnHCl-induced equilibrium unfolding experiments, 10 μM of protein was incubated for >6 h in different concentrations of GdnHCl (0–4 M). The equilibrated unlabeled and labeled protein samples were excited at 295 nm, and the fluorescence signals were monitored on an MOS-450 optical system (Biologic), using a 360 ± 10 nm band-pass filter (Asahi spectra) for steady-state FRET measurements. For time-resolved FRET measurements, fluorescence decay curves were acquired for both the unlabeled and labeled proteins using the time correlated single photon counting system set-up as described in a previous study (Bhatia, Krishnamoorthy, and Udgaonkar 2018). A two-state, N ↔ U model was used to fit the sigmoidal equilibrium unfolding transitions (Agashe and Udgaonkar 1995) to obtain the values for the free energy of unfolding in water, ΔG_U, and the midpoint of transition (Cm).

Trp-fluorescence monitored folding kinetics experiments

The folding reaction was initiated by rapidly mixing unfolded protein (in 4 M GdnHCl) with folding buffer (containing no GdnHCl) in a 1:9 ratio using a stopped-flow cuvette (FC20)

and a stopped-flow module from Biologic (SFM-400). The protein concentration was 10 μM. The dead time of stopped-flow mixing was ~10 ms.

Double-kinetics experiment

Simultaneous acquisition of fluorescence decay traces and fluorescence intensities was performed at different time points (every 100 ms) during the folding reaction, using a pulsed laser (Mai Tai HP, Spectra-Physics) as the excitation light source (Bhatia et al. 2019). Notably, the simultaneous acquisition of steady-state fluorescence intensity served to validate the accuracy of the double-kinetics data for each mixing experiment (Figures 3.8 and 3.9). Photon arrival data were collected using a single-photon counting card module (SPC-630) from Becker and Hickl. For all fluorescence decay curves, it was ensured that the counts were ≥ 20,000 in the peak channel, and the decay was collected until the counts were < 200 in the last channel. A detailed description of the double-kinetics setup and fluorescence data collection has been reported previously (Bhatia et al. 2019). The final protein concentration used was ~20 μM.

Analysis of the fluorescence lifetime decay traces: Details of the analysis are given in an earlier study (Bhatia, Krishnamoorthy, and Udgaonkar 2018). A brief description is given below:

Discrete analysis: The decay traces were fit to a sum of exponentials,

$$F(t) = \sum_{i=1}^{n} \alpha_i e^{-\left(\frac{t}{\tau_i}\right)} \quad \textbf{(1)}$$

Here, α_i is the relative amplitude of the τ_i lifetime component, t is time and n ranges from 2 to 3.

An amplitude-weighted mean lifetime, the mean lifetime, $\boldsymbol{\tau}_m$ at every time point of the refolding reaction was determined, was caculated as:

$$\boldsymbol{\tau}_m = \frac{\sum \alpha_i \tau_i}{\sum \alpha_i};\ \sum \alpha_i = 1 \quad \textbf{(2)}$$

MEM analysis: The analysis begins with the assumption that the decay corresponds to a distribution of 100-150 lifetimes in the range 10 ps to 10 ns. The a priori distribution of lifetimes was assumed to be uniform in the logarithms of lifetimes being uniformly distributed in this range. Then, the best fit values of α_i and τ_i are determined using the Maximum Entropy Method (MEM) (Bhatia, Krishnamoorthy, and Udgaonkar 2021a, 2021b).

The a posteriori distribution of these parameters was obtained by maximizing the Shannon Jaynes entropy S, defined as

$$S = -\Sigma P_i \log(P_i) \quad \textbf{(3)}$$

P_i is the normalized amplitude of the ith lifetime.

$$P_i = \frac{\alpha_i}{\Sigma \alpha_i} \quad \textbf{(4)}$$

α_i is the amplitude of the ith lifetime.

Analysis parameters were optimized for obtaining precise and reproducible MEM distributions. The most probable (MEM peak) lifetime refers to the lifetime corresponding to the maximum amplitude in the lifetime distribution data.

FRET efficiency and distance determination

The mean FRET efficiency ($<E_{FRET}>$) was obtained from mean fluorescence lifetimes for the unlabeled ($<\tau_D>$) and the corresponding TNB-labeled ($<\tau_{DA}>$) variants using the following equation:

$$<E_{FRET}> = 1 - \frac{<\tau_{DA}>}{<\tau_D>} \quad \textbf{(5)}$$

The mean fluorescence lifetimes were determined by discrete analysis of the fluorescence decay traces as described earlier (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

The average FRET efficiency values were converted to average intramolecular distance ($<R_{DA}>$) using the Förster equation:

$$<R_{DA}> = R_0 \left(\frac{1 - <E_{FRET}>}{<E_{FRET}>}\right)^{1/6} \quad \textbf{(6)}$$

The values for the Forster's distance, R_0 used for the WT proteins and their determination has been described in an earlier study (Bhatia, Krishnamoorthy, and Udgaonkar 2018). Upon Pro mutation to Ala, the R_0 values remained similar to those of their respective background constructs (24.5 Å for P41A and 22.7 Å for P93A).

Quantification of the N-like and U-like sub-populations from lifetime distributions

Fluorescence lifetime distributions of both the unlabeled and the TNB labeled variants were determined for consecutive acquisition time windows of 100 ms by fitting the observed fluorescence decay profiles to the MEM algorithm (described above). A cut-off of 0.6 ns

(corresponding to a diatance-cutoff of ~ 20 Å, as calculated using equations 5 and 6) was used to distinguish between lifetimes in the fluorescence lifetime distributions, which arose from the N-like and U-like sub-populations. Using the cut-off of 0.6 ns, the fractions of molecules in the N-like and U-like sub-populations were quantified directly from the MEM-derived fluorescence lifetime distributions of the TNB-labeled variants (Figure 3.9), as had been done in an earlier study (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). This provided a robust and model-independent measure of population redistribution during folding.

Converting fluorescence lifetime distributions into distance distributions

The lifetime distributions of the TNB-labeled variant at different times of folding were then converted to FRET efficiency distributions by using the peak lifetime values obtained for the corresponding unlabeled variant (see Eq. 5 above). Since the peak lifetime of the dominant long-lifetime sub-population seen in the MEM lifetime distribution for the unlabeled protein did not vary significantly with time of folding (Figure 3.8), its average value was taken as the lifetime, τ_D, of the donor. The minor shorter lifetime sub-population present in the MEM distributions of the unlabeled proteins was not used to define the lifetime of the donor, τ_D, since it contributes $\leq$ ~30% throughout the folding time course (Figure 3.14).

The FRET efficiency distributions were then transformed into distance distributions using the well-known Forster relation (see Eq. 6 above), as established in previous studies (Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021b). It was important to note that the apparent widths of the derived distance distributions might reflect contributions from (i) the intrinsic width of the donor-only lifetime distribution, (ii) true donor–acceptor distance heterogeneity, and (iii) the nonlinear mapping between E_{FRET} and R_{DA}, especially in the low-FRET regime. Quantitative separation of these contributions cannot be achieved uniquely from the data alone and would require introducing additional, model-dependent assumptions. When such model-based approaches were employed in previous studies, it was found that the absolute distances obtained can differ from those estimated by direct conversion of FRET efficiencies, although the underlying qualitative trends in chain compaction remained unchanged (Sinha and Udgaonkar 2007; Zhou 2002; Goldenberg 2003; Wang, Bodunov, and Nau 2003; Laurence et al. 2005; Saxena, Udgaonkar, and Krishnamoorthy 2006). Hence, in this study, due to the nonlinear dependence of FRET efficiency on distance, the distance distributions were not used to determine the relative fractions of N-like and U-like molecules at different times of folding.

The mean distances (Figure 3.10), as well as the distance distributions (Figure 3.11) used here were interpreted primarily in a qualitative manner.

3.4 Results and Discussion

The structure and stability of MNEI are not affected significantly by Pro to Ala mutation and TNB-labeling

Far-UV circular dichroism spectra showed that both single Pro to Ala mutations, as well as TNB-labeling, had only small effects on the secondary structure of the protein (Figure 3.2). Fluorescence spectra showed that the environments of Trp19 and Trp4 remain largely unchanged upon Pro to Ala mutation (Figure 3.3). The fluorescence emission maxima of Trp19 in the N states of both WT_C and P41A_C were red-shifted compared to that of Trp4 in WT_E and P93A_E, indicating that the environment of Trp19 is more polar than that of Trp4 (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

Upon TNB-labeling, the quenching of Trp19 fluorescence because of FRET across the C segment to the TNB adduct on Cys42, as well as the quenching of Trp4 fluorescence because of FRET across the E segment to the TNB adduct on Cys97, were both reduced upon Pro to Ala mutation. The fluorescence intensities of Trp19 and Trp4 in the N states of TNB-labeled P41A_C and P93A_E were 2.5-fold and 2-fold higher than those for TNB-labeled WT_C and WT_E, respectively (Figure 3.3). The FRET efficiencies calculated for the N states of WT_C and WT_E were 0.9 and 0.8, respectively, and decreased to 0.8 and 0.6 for the N states of P41A_C and P93A_E, respectively (Figure 3.4, insets). These observations suggested that the N states of P41A_C and P93A_E have slightly expanded C and E segments, respectively. In the case of P93A_E, this was consistent with the fact that the Pro93 to Ala mutation causes the C-terminal sequence segment 93-97 to move away from the protein core, because of the Gly92-Ala93 peptide bond adopting the *trans* conformation (Figure 3.1b).

Neither the single Pro to Ala mutations nor TNB-labeling significantly affected the stability of the protein (Figure 3.4 and Table 3.1). P41A_C was marginally less stable than WT_C, and the stabilities of both proteins were not altered to significant extents upon TNB labelling (Table 3.1). TNB-labeling of Cys97 had no significant effect on the stabilities of WT_E and P93A_E. Overall, the effects of the mutations and TNB-labeling on the structure and dynamics were minor (Figures 3.2, 3.3 and 3.4). It should be noted that protein stabilities were determined by carrying out two-state analyses of the equilibrium unfolding transitions

(Bhatia, Krishnamoorthy, and Udgaonkar 2018). It had been shown previously that fluorescence and CD-monitored equilibrium unfolding transitions yield the same values for the protein stability, for these four MNEI variants (Bhatia, Krishnamoorthy, and Udgaonkar 2018; Kaushik and Udgaonkar 2023), as well as for other MNEI variants (Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016).

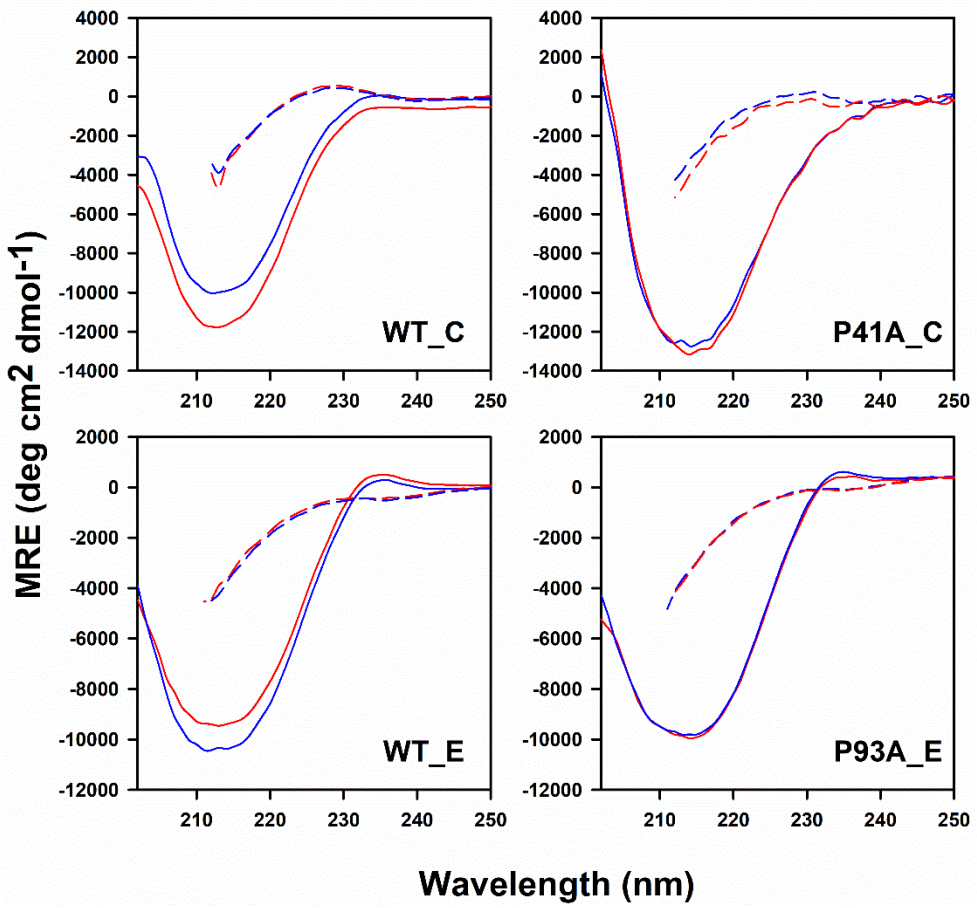


Figure 3.2. Far-UV CD spectra of the different mutant variants of MNEI. Spectra of the N (solid lines) and U (dashed lines) states are shown for the TNB-labeled (red lines) and unlabeled (blue lines) proteins. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

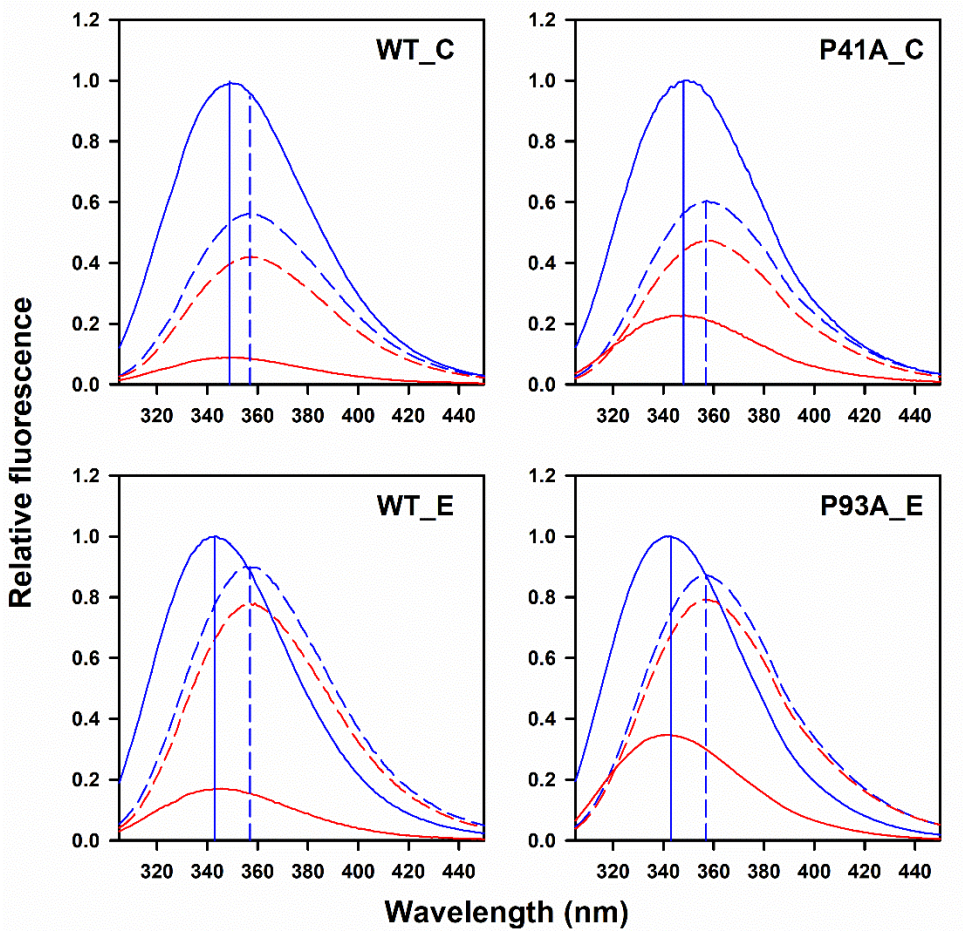


Figure 3.3. Fluorescence emission spectra of the different mutant variants of MNEI. The excitation wavelength was 295 nm. The solid and dashed curves represent the spectra of the native and unfolded states, respectively, for the TNB-labeled (red) and unlabeled (blue) proteins. The solid and dashed vertical lines indicate the peak positions of the spectra of the native and unfolded unlabeled proteins, respectively. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

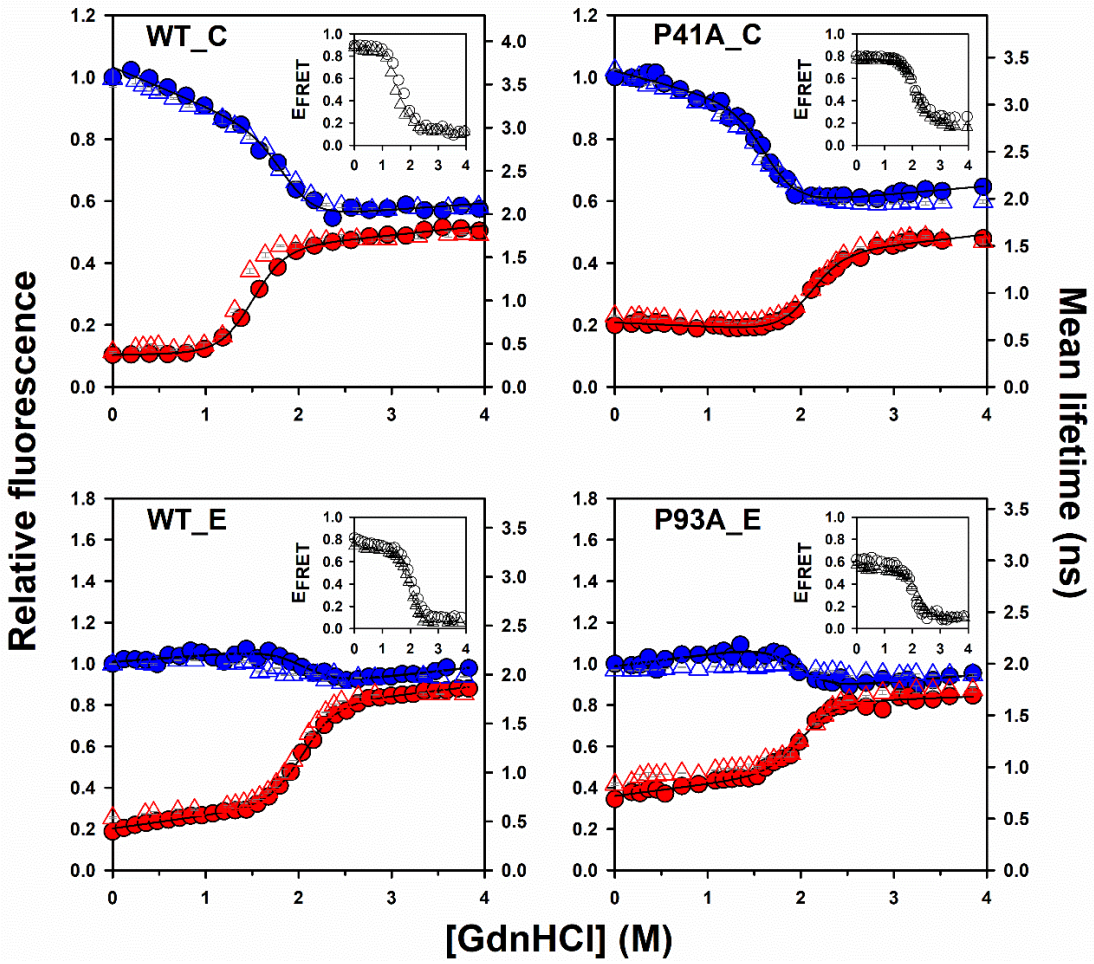


Figure 3.4. Equilibrium unfolding transitions of MNEI monitored by steady-state FRET (circles) and time-resolved FRET (triangles). Blue and red colors represent the fluorescence intensities and mean lifetimes of the unlabeled and TNB-labeled proteins, respectively. The solid lines through each dataset are non-linear least-squares fits to a two-state unfolding model (Agashe and Udgaonkar 1995). The thermodynamic parameters obtained are listed in Table 3.1. The inset in each panel shows the correspondence between FRET efficiency calculated from fluorescence intensity (circles) and fluorescence mean lifetime (triangles) measurements. The error bars represent the spread in the data obtained from two independent experiments. The data for WT_C and WT_E are from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

Table 3.1. Thermodynamic parameters obtained from fluorescence-monitored equilibrium unfolding measurements for the different mutant variants of MNEI at pH 8 and 25°C.

Protein variant	Free energy of unfolding, (ΔG_U) kcal mol^{-1}	Mid-point of unfolding,(C_m) M
WT_C	6.3 ± 0.1	1.86 ± 0.01
WT_C-TNB	4.9 ± 0.2	1.46 ± 0.06
P41A_C	5.4 ± 0.1	1.62 ± 0.02
P41A_C-TNB	6.9 ± 0.1	2.08 ± 0.07
WT_E	6.7 ± 0.4	1.97 ± 0.13
WT_E-TNB	6.9 ± 0.1	2.03 ± 0.01
P93A_E	6.6 ± 0.1	1.94 ± 0.04
P93A_E-TNB	6.8 ± 0.04	2.03 ± 0.01

Validation of the time-resolved fluorescence measurements

The equilibrium unfolding transitions monitored by measurement of the Trp fluorescence intensity at each GdnHCl concentration closely matched those monitored by measurement of the mean fluorescence lifetime determined from discrete analysis (see Materials and Methods) of the Trp fluorescence decay measured at each GdnHCl concentration (Figure 3.4). In multiexponential systems, the steady-state fluorescence intensity is proportional to $\sum \alpha_i \tau_i$, from which the amplitude-weighted mean lifetime is derived (equation 2, Materials and Methods); therefore, changes in fluorescence intensity and mean lifetime are expected to correlate closely. This was an important result because it validated the accuracy of the fluorescence lifetime decay measurements and set the stage for analysis of the fluorescence decays by the maximum entropy method (MEM).

MEM -analyzed fluorescence decays distinguish quantitatively between U-like and N-like sub-populations present together

MEM analysis (see Materials and Methods) was used to convert fluorescence decays collected at different GdnHCl concentrations to fluorescence lifetime distributions for both the unlabeled (Figure 3.5) and labeled (Figure 3.6) protein variants. The MEM-derived distributions for all the unlabeled variants were largely unimodal (Figure 3.5). In contrast, those for the labeled variants were typically bimodal (Figure 3.6), and displayed a peak for lifetimes shorter than 0.6 ns, originating from a N-like sub-population (more FRET, hence shorter lifetimes), as well as a peak for lifetimes longer than 0.6 ns, originating from an U-like sub-population (less FRET, hence longer lifetimes) (Bhatia, Krishnamoorthy, and Udgaonkar 2021a; Bhatia and Udgaonkar 2022). This assignment was based on the observation that the U state predominantly populates lifetimes > 0.6 ns (U-like) and the N state predominantly populates lifetimes < 0.6 ns (N-like) (Figure 3.6) (Bhatia, Krishnamoorthy, and Udgaonkar 2021a; Bhatia et al. 2019).

Such groupings of lifetimes allowed the fractions of molecules in the U-like and N-like sub-populations to be determined at each GdnHCl concentration. To extract these fractions from the MEM distributions, it was first necessary to account for the observation that for any segment, the equilibrium U state contained a fraction of molecules with lifetimes corresponding to the N-like distribution (< 0.6 ns), while the equilibrium N state contained a fraction of molecules with lifetimes corresponding to a U-like distribution (> 0.6 ns) (Figures 3.5 and 3.6). These fractions were comparable for each pair of unlabeled (Figure 3.5) and corresponding

labeled (Figure 3.6) unfolded proteins, indicating that they originated from differences in the electronic structure of the fluorophore or from distinct Trp rotamers (Creed 1984; Philips et al. 1988), and not from the presence of the quenching TNB moiety. The fraction of molecules expanded (U-like) at a given segment was determined using the equation: $f_U=Y_i-Y_N/Y_U-Y_N$, where Y_U represents the relative sum of amplitudes for the U-like distribution in the equilibrium U state, Y_N corresponds to the relative sum of amplitudes for the U-like distribution in the equilibrium N state, and Y_i denotes the relative sum of amplitudes for the U-like distribution at a given GdnHCl concentration. The relative sum of amplitudes was calculated as the sum of amplitudes for the U-like or N-like distributions divided by the total sum of amplitudes for both distributions. f_U derived from MEM-derived fluorescence lifetime distributions (Figure 3.6), had been shown previously to accurately estimate the relative fractions of N-like and U-like molecules present together (Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

It should be noted that while the N states of WT_C, P41A_C, and WT_E, like those of several other proteins (Bhatia, Krishnamoorthy, and Udgaonkar 2018; Kishore, Krishnamoorthy, and Udgaonkar 2013; Swaminathan, Krishnamoorthy, and Periasamy 1994), displayed bimodal MEM-derived fluorescence lifetime distributions with a very minor (~10–20%) fraction of molecules appearing to be U-like (Figure 3.6), the N state of P93A_E exhibited a broad unimodal distribution with no clear distinction between the N-like and U-like sub-ensembles. Clearly, the Pro93 to Ala mutation made the N state more heterogeneous, with molecules exhibiting a broad range of end-to-end distances. It is possible that the *cis* Pro93 introduces a kink or bend in the protein backbone in the N state (Figure 3.1b), which may contribute to the stability of a specific conformation or sub-ensemble. Thus, in WT_E-TNB, two distinct conformations (N-like and U-like) might be stabilized, leading to a bimodal distribution. When Pro93 was mutated to Ala, a more flexible amino acid, the rigidity that Pro93 provided was lost. This increased flexibility likely allows the protein to sample a broader range of conformations, thereby blurring the separation between the two sub-ensembles, and resulting in the broad unimodal distribution seen for P93A_E (Figure 3.6). Importantly, although the distribution was unimodal, the relative sum of amplitudes for lifetimes >0.6 ns (U-like) was still ~25–30%, comparable to WT_E. Normalization of the amplitudes (see above) ensured that f_U could still be accurately determined.

For each protein variant, the dependence of the MEM-derived fraction of molecules that were U-like calculated as described above (Figure 3.7) was found to match the unfolding

transition monitored by the measurement of fluorescence intensity or mean fluorescence lifetime of Trp4 or Trp19, and to yield similar values for the stability, ΔG_U and the midpoint of the transition, C_m (Table 3.1). This correspondence validates the use of the MEM-derived fluorescence lifetime distributions for determining accurately the relative fractions of N-like and U-like molecules present together, using the cut-off value of 0.6 ns for the fluorescence lifetime (see above).

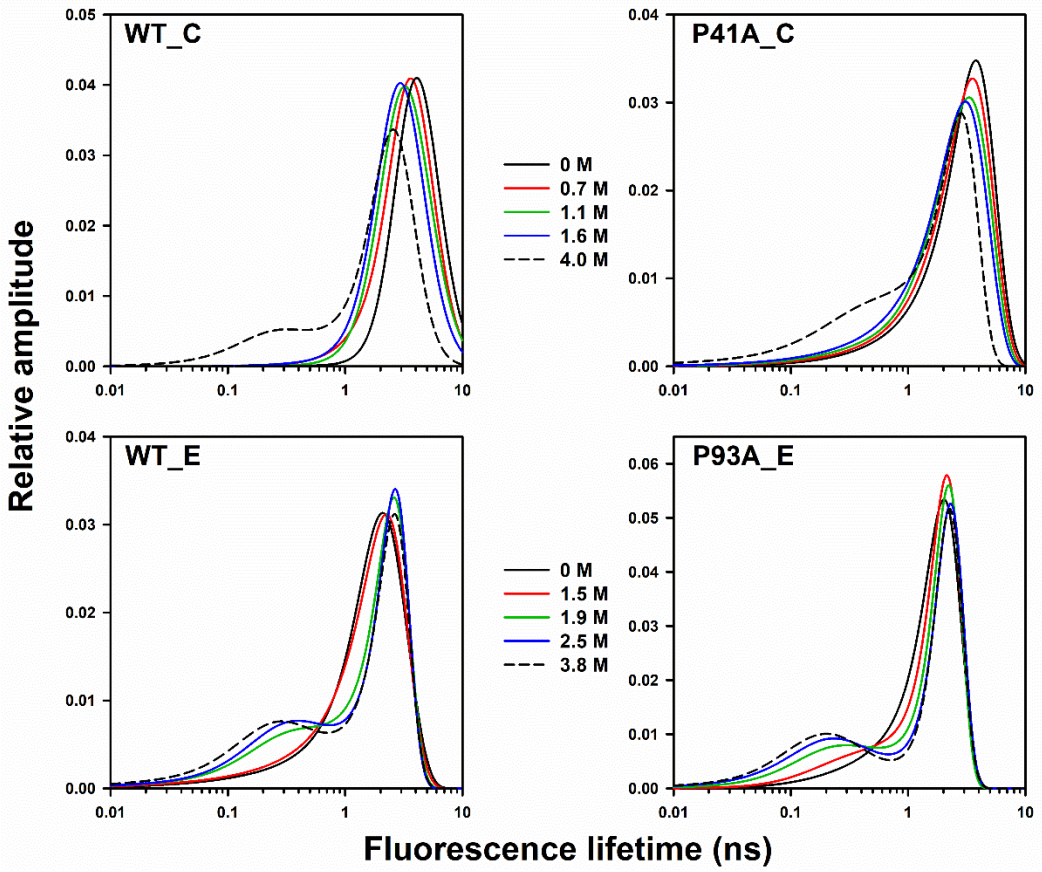


Figure 3.5. MEM-derived fluorescence lifetime distributions of the different mutant variants of MNEI at varying concentrations of GdnHCl. The colors of the lines represent the GdnHCl concentrations, as indicated. The data in panels WT_C and WT_E are from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

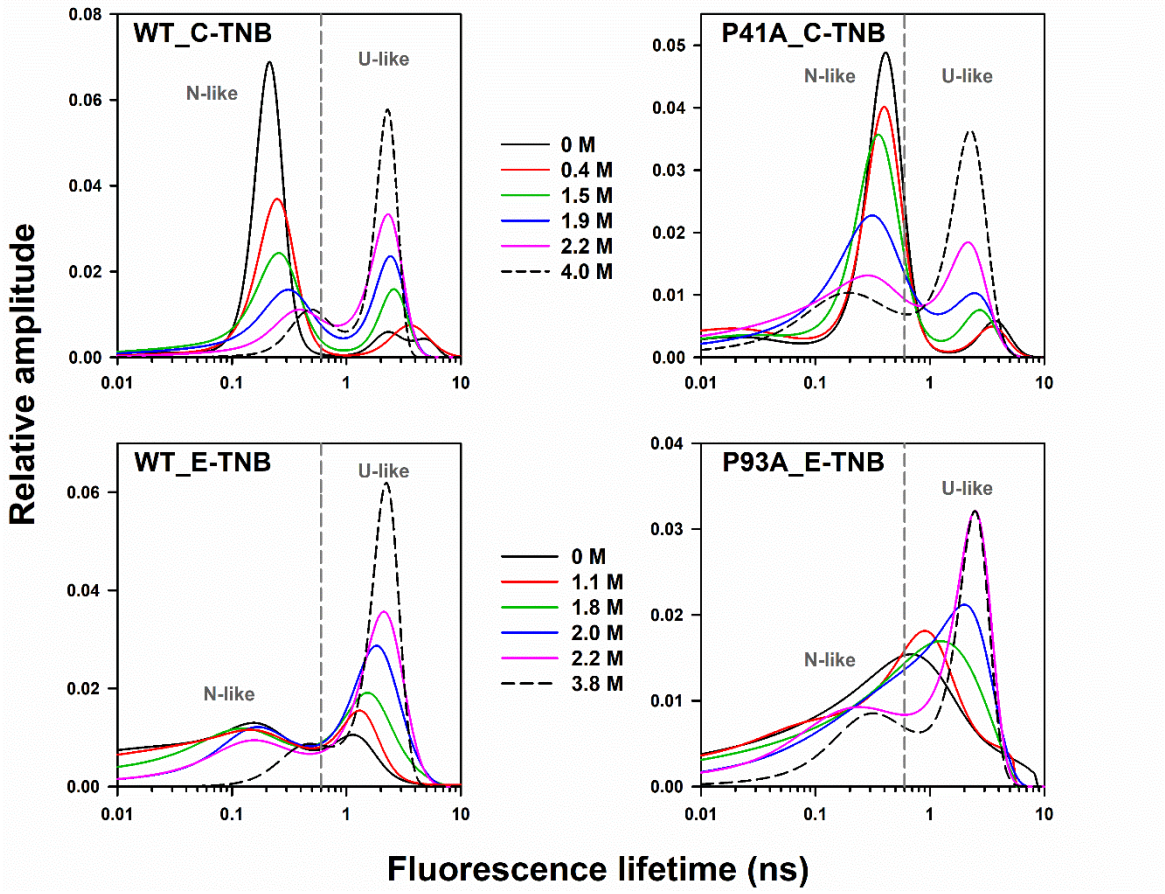


Figure 3.6. MEM-derived fluorescence lifetime distributions of different TNB-labeled variants of MNEI at varying concentrations of GdnHCl. The colors of the lines represent different GdnHCl concentrations, as indicated. The data in panels WT_C-TNB and WT_E-TNB are from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

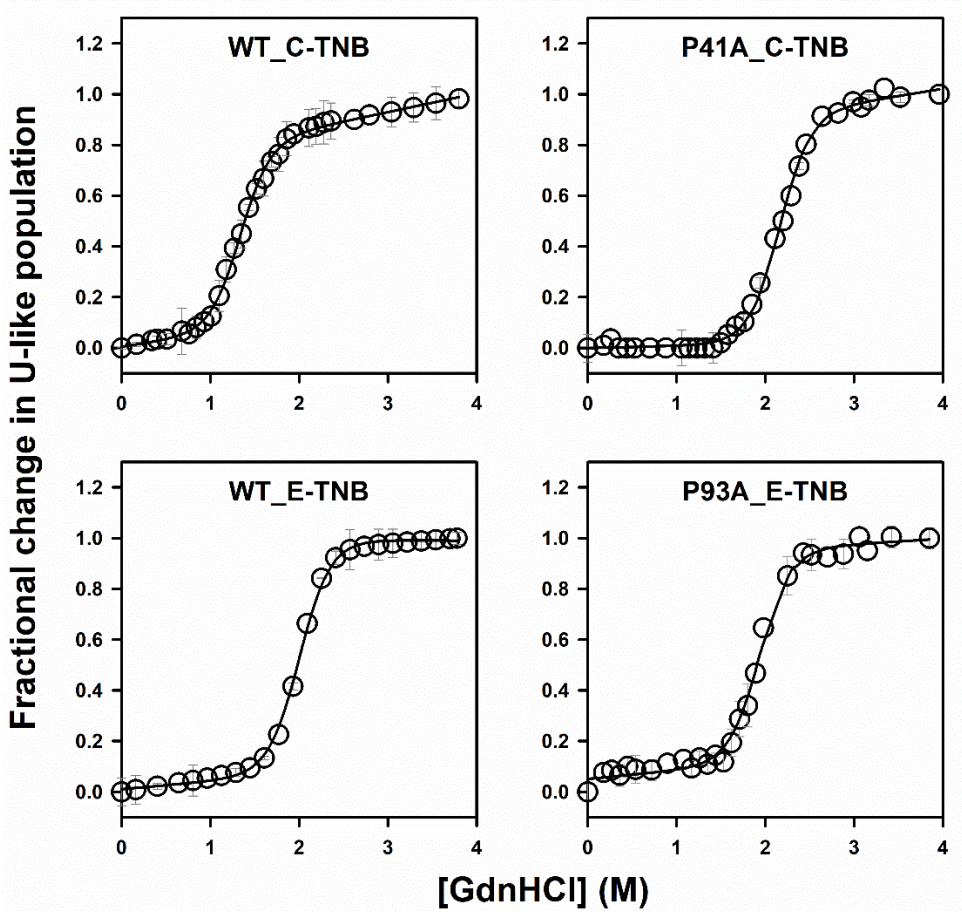


Figure 3.7. Fractional change in the U-like population calculated from the relative sum of amplitudes of the MEM distributions at different concentrations of GdnHCl. The fits to the two-state, N ↔ U unfolding model gave values for ΔG_U of 4.6 ± 0.3, 7.1 ± 0.2, 6.9 ± 0.1, and 6.7 ± 0.2 kcal mol^{-1} for WT_C-TNB, P41A_C-TNB, WT_E-TNB and P93A_E-TNB, respectively. The mid-points (C_m) of the unfolding transitions for WT_C-TNB, P41A_C-TNB, WT_E-TNB and P93A_E-TNB are at 1.41 ± 0.08, 2.10 ± 0.04, 2.03 ± 0.02, and 1.97 ± 0.04 M GdnHCl, respectively. The error bars represent the spread in the data obtained from two independent experiments. Panels WT_C-TNB and WT_E-TNB have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

The Pro41 to Ala mutation stabilizes N-like structure at the C segment in the product of the initial collapse reaction

Transformation of the kinetic data in the insets of Figures 3.8 and 3.9 into FRET efficiency (Figure 3.10; see Materials and Methods) revealed a distinct burst phase for both WT_C and P41A_C, which was more pronounced in the latter. One explanation for the observed change in FRET efficiency measured in this way could be that all molecules became compact initially, but that the extent of compaction was more in the case of P41A_C than for WT_C. In other words, it could be that the Pro41 to Ala mutation led to a greater reduction in the mean intramolecular distance at the C segment (insets, Figure 3.10). The alternative explanation could be that all molecules did not become compact, but only a sub-population of them underwent significant compaction, while another sub-population underwent little if any compaction. To distinguish between these possibilities and obtain quantitative insight into the fraction of molecules undergoing compaction, the fluorescence decays were analyzed using MEM (Figures 3.8, 3.9 and 3.11; see Materials and Methods). As in the case of the equilibrium unfolding data (Figures 3.5 and 3.6), N-like and U-like sub-populations were observed at different times of folding. The kinetics of the decrease in the fraction of U-like molecules, f_U (described above), as they contracted to become N-like during folding, was determined (Figure 3.12).

The major effect of the P41A mutation was that the fraction of molecules that become N-like at the C segment at 100 ms of folding, increased from 16 % in the case of WT_C to 63 % for P41A_C (Figure 3.12, Table 3.2). It is important to explain this observation in the context of the mechanism of chain contraction and folding proposed for the wild-type protein (Figure 3.1c), for which the collapsed intermediate ensemble populated at 100 ms consists of three sub-ensembles (Figure 3.1c). 16% are present as I_{BCE} in which segments B, C, and E have all collapsed to become N-like; 22% are present as I_B in which only segment B has collapsed to become N-like; and 62% remain U-like, as U_X, at all three segments. It appears that the P41A mutation could have resulted in two different types of effects on the heterogeneity of this collapsed ensemble formed at 100 milliseconds of folding. (1) It could have led to the induction of contraction of segment C in some of the U_X and I_B molecules. (2) It could have led to the conformational selection of very minor pre-existing sub-ensembles present in equilibrium with U_X and I_B, in which the C segment is already N-like in its dimensions.

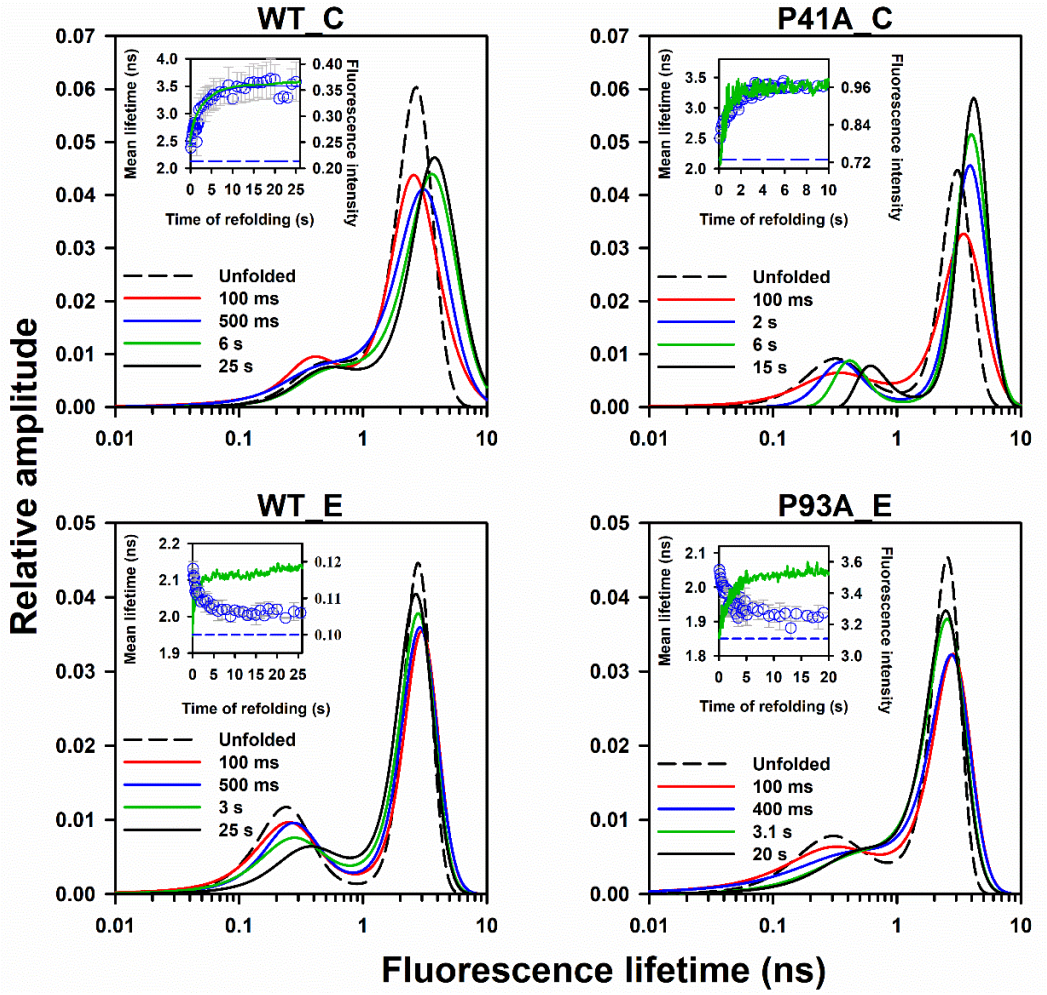


Figure 3.8. Kinetics of folding of the unlabeled MNEI variants in 0.4 M GdnHCl. The four panels correspond to different unlabeled variants (as indicated on the top of each panel). The dashed fluorescence lifetime distribution in each panel is of the U state. Distributions corresponding to different times of the folding reaction are shown in different colours as described for each panel. The inset in each panel represents the kinetics of folding monitored using simultaneous measurements of fluorescence intensity (green trace) and mean lifetime (blue circles) changes. The dashed blue line in the inset of each panel represents the mean lifetime of the U state. The error bars represent the standard errors of measurements from at least two independent double kinetics experiments. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

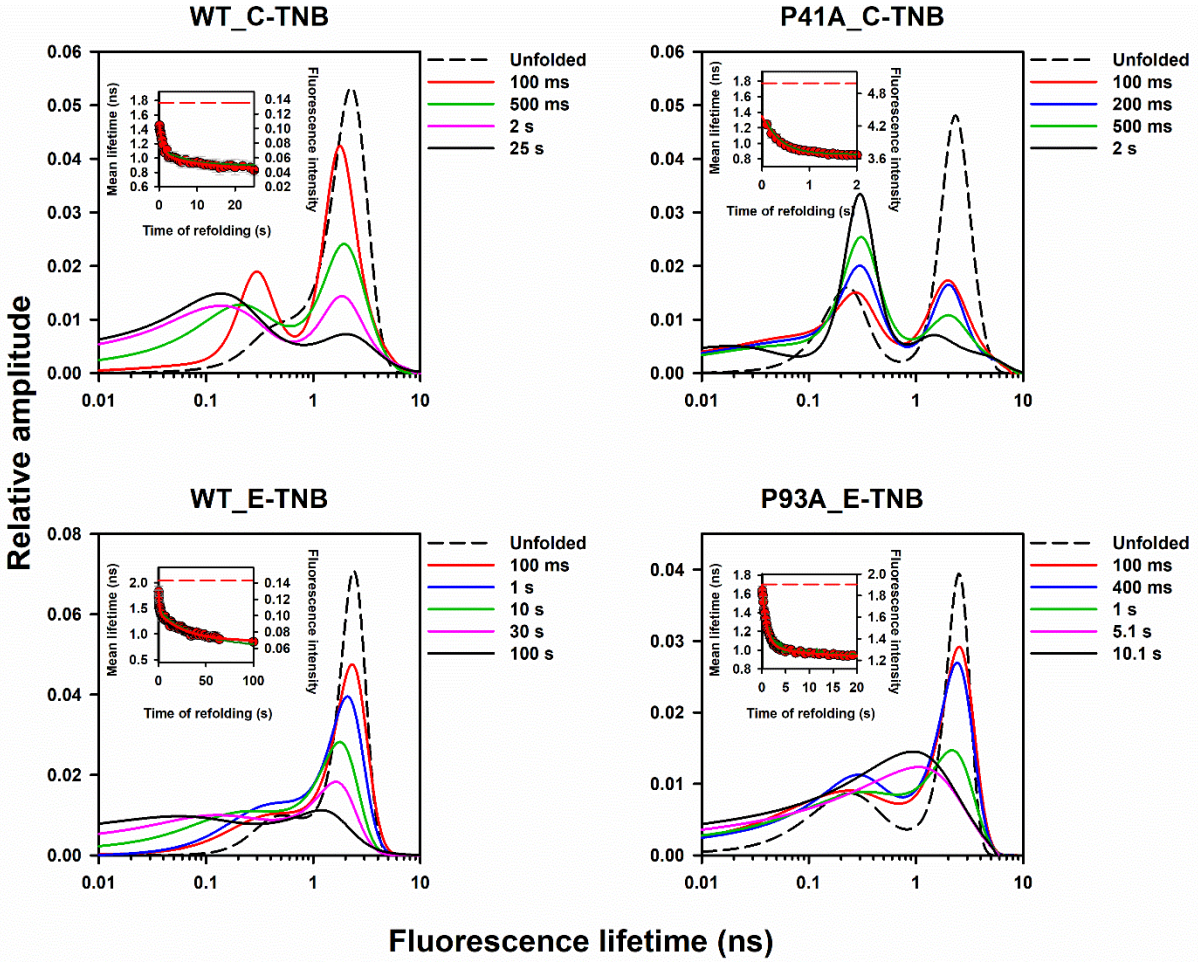


Figure 3.9. Kinetics of folding of the TNB-labeled mutant variants of MNEI in 0.4 M GdnHCl. The four panels correspond to different TNB-labeled mutant variants (as indicated on the top of each panel). The dashed fluorescence lifetime distribution in each panel is of the U state. Distributions corresponding to different times of the folding reaction are shown in different colours as described for each panel. The inset in each panel represents the kinetics of folding monitored using simultaneous measurements of fluorescence intensity (green trace) and mean lifetime (red circles) changes for different mutant variants of MNEI. The dashed red line in the inset of each panel represents the mean lifetime of the U state. The error bars represent the standard errors of measurements from at least two independent double kinetics experiments. Panels WT_C-TNB and WT_E-TNB have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

Table 3.2. Rate constants and their corresponding relative amplitudes obtained by fitting the kinetic phases of folding/contraction.

Protein	Phase	Rate constants (amplitudes) of folding[a] (s^{-1})	Rate constants (amplitudes) of the ss-FRET efficiency monitored change[b] (s^{-1})	Rate constants (amplitudes) of the tr-FRET efficiency monitored change[c] (s^{-1})	Rate constants (amplitudes) of the decrease in U-like population[d] (s^{-1})
WT_C	Burst	—	— (27%)	— (39%)	— (16%)
	Very fast	4.7 (14%)	7.5 (19%)	—	—
	Fast	0.6 (60%)	1 (40%)	1 (43%)	2 (61%)
	Slow	0.06 (26%)	0.07 (14%)	0.1 (18%)	0.17 (23%)
P41A_C	Burst	—	— (44%)	— (58%)	— (63%)
	Very fast	10.5 (12%)	34 (14%)	—	—
	Fast	1.5 (64%)	2.8 (38%)	2 (42%)	2 (37%)
	Slow	0.14 (24%)	0.16 (4%)	—	—
WT_E	Burst	—	—	—	— (11%)
	Very fast	25 (25%)	26 (21%)	—	—
	Fast	1 (48%)	1 (26%)	1 (42%)	2 (19%)
	Slow	0.1 (27%)	0.03 (53%)	0.03 (58%)	0.04 (70%)
P93A_E	Burst	—	— (14%)	— (17%)	— (15%)
	Very fast	—	—	—	—
	Fast	1 (78%)	1 (70%)	1 (69%)	1 (75%)
	Slow	0.1 (22%)	0.1 (16%)	0.1 (14%)	0.1 (10%)

[a] Measured from the kinetics of change in the fluorescence of Trp19 or Trp4 in the unlabeled proteins.

[b] Measured from the kinetics of change in fluorescence intensity of Trp19 or Trp4 in the TNB-labeled and unlabeled proteins. The burst phase was of 10 ms duration.

[c] Measured from the kinetics of change in the mean fluorescence lifetime of Trp19 or Trp4 in the TNB-labeled and unlabeled proteins, shown in Figure 3.10. The burst phase was of 100 ms duration.

[d] Measured from the kinetics of change in the U-like population seen in the MEM distributions of TNB-labeled proteins, shown in Figure 3.12. The burst phase was of 100 ms duration.

The data for WT_C and WT_E have been taken from (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). The errors in the rate constants and their corresponding relative amplitudes are within ± 10%.

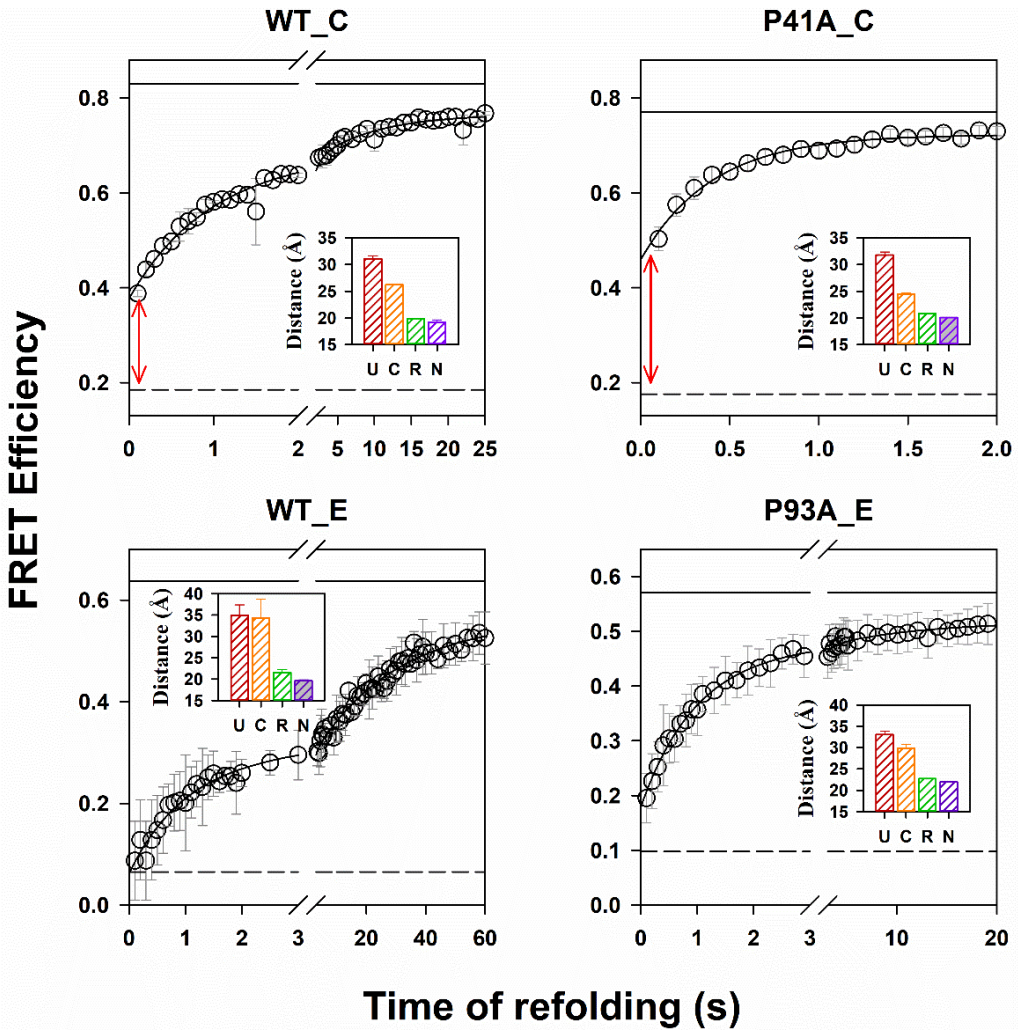


Figure 3.10. Kinetics of folding in 0.4 M GdnHCl monitored by trFRET for the different mutant variants of MNEI. FRET efficiency was calculated using the values of the mean lifetime of the unlabeled and TNB-labeled variants at each folding time point (equation 5, see Materials and Methods). In each panel, the solid and dashed black horizontal lines represent the FRET efficiencies in the N and U states, respectively. The red arrow in each panel indicates the burst-phase change in FRET efficiency from the U state to the first observable folding time point (100 ms). The data for WT_C, WT_E and P93A_E were fit to a two-exponential equation, while that for P41A_C was fit to a single exponential equation. The rate constants and relative amplitudes obtained from the fitting are reported in Table 3.2. The inset in each panel shows the average donor−acceptor pair distance ($<R_{DA}>$) calculated at different stages of folding (U, unfolded; C, collapsed intermediate at 100 ms; R, refolded; and N, native state) using FRET efficiency values and the Förster equation (equation 6, see Materials and Methods). Error bars indicate standard errors from two independent double-kinetics experiments. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

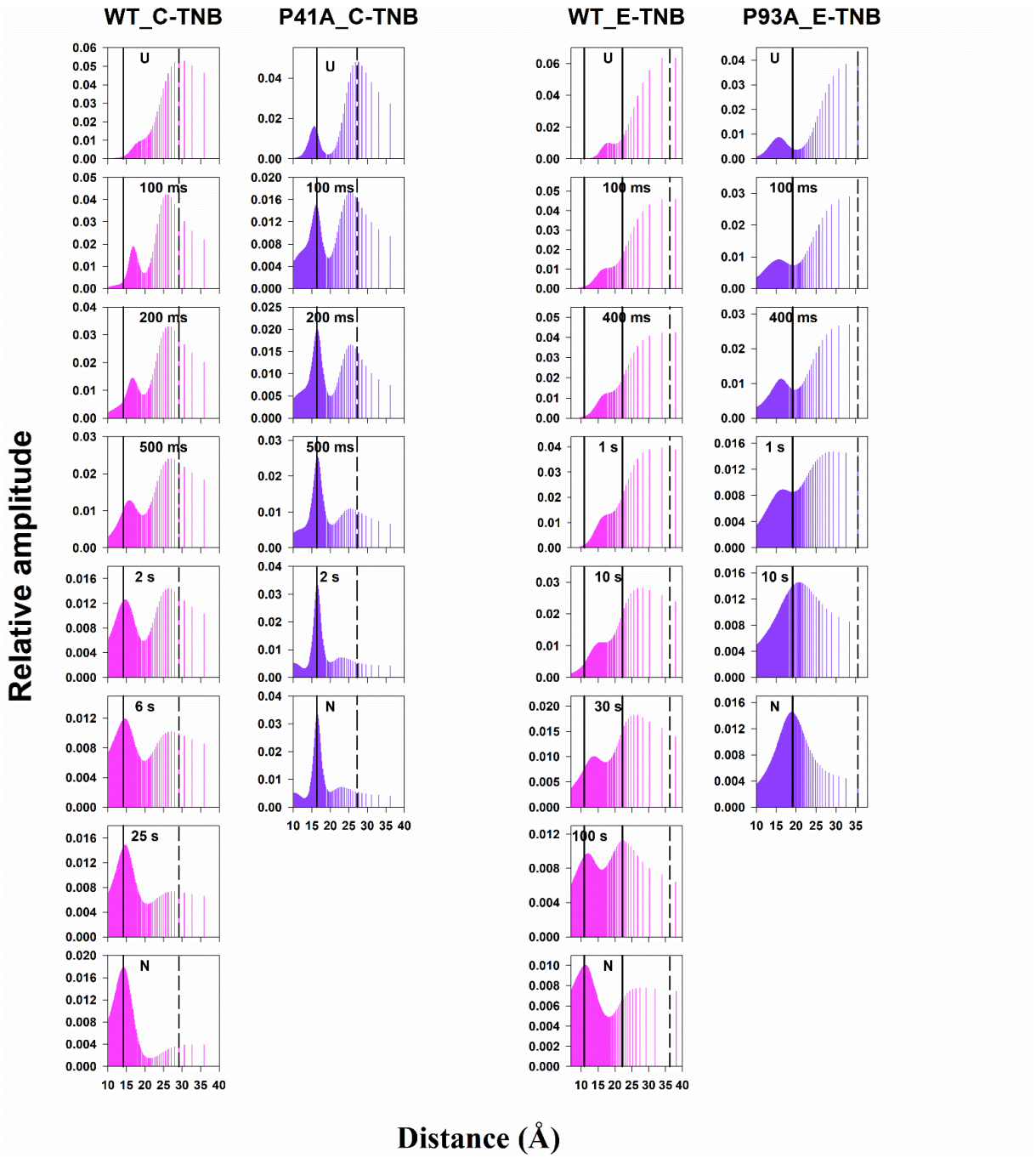


Figure 3.11. Evolution of distance distributions as a function of the time of folding following a 4 to 0.4 M GdnHCl jump. Experimentally derived distance distributions at representative time points of the folding reaction. The topmost panel is for the U state, the bottom-most panel corresponds to the distance distribution of the refolded N state, and the middle panels correspond to intermediate time points during folding as described in each panel. The vertical solid and dashed black lines indicate the peak positions of the distance distributions corresponding to the refolded N state and U state, respectively. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021b).

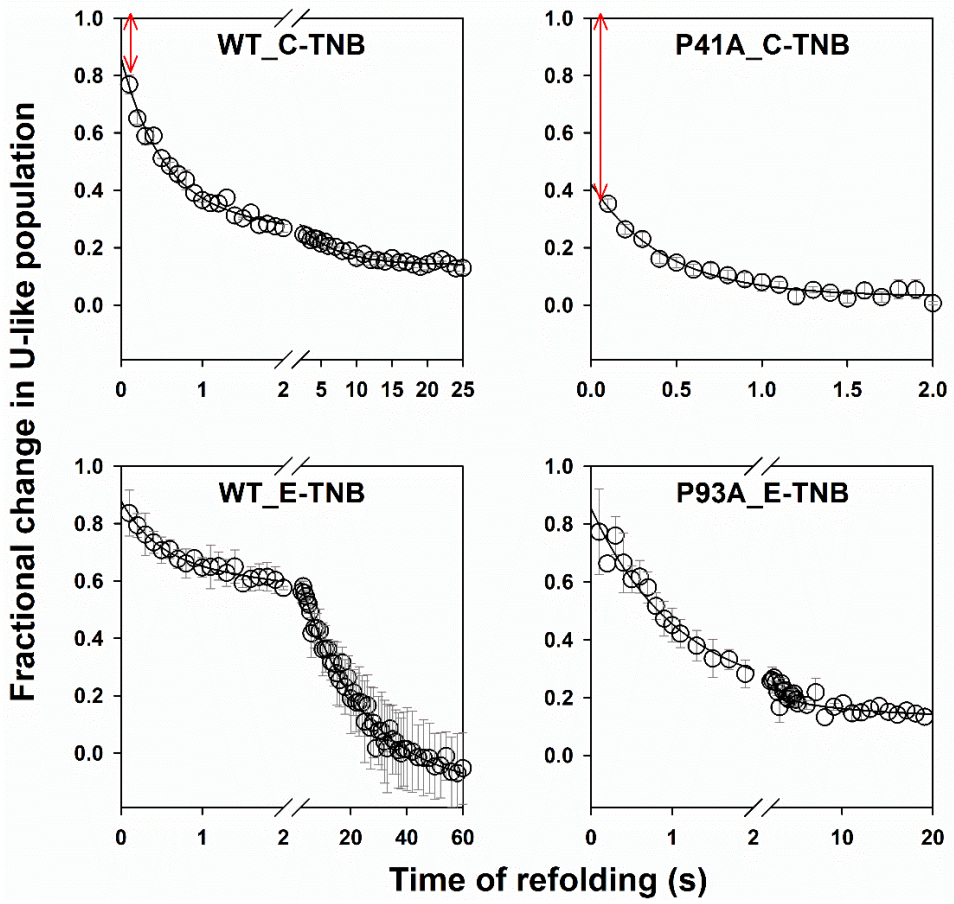


Figure 3.12. Kinetics of conversion from the expanded, U-like to the contracted, N-like population, measured at segments C and E for the different mutant variants of MNEI. The fractional population of molecules that are expanded (U-like) was determined as described in the text. The data for WT_C, WT_E and P93A_E were fit to a two-exponential equation, while that for P41A_C was fit to a single-exponential equation. The rate constants and relative amplitudes obtained from the fitting are reported in Table 3.2. The error bars represent the standard errors of measurements from at least two independent double kinetics experiments. Panels WT_C-TNB and WT_E-TNB have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

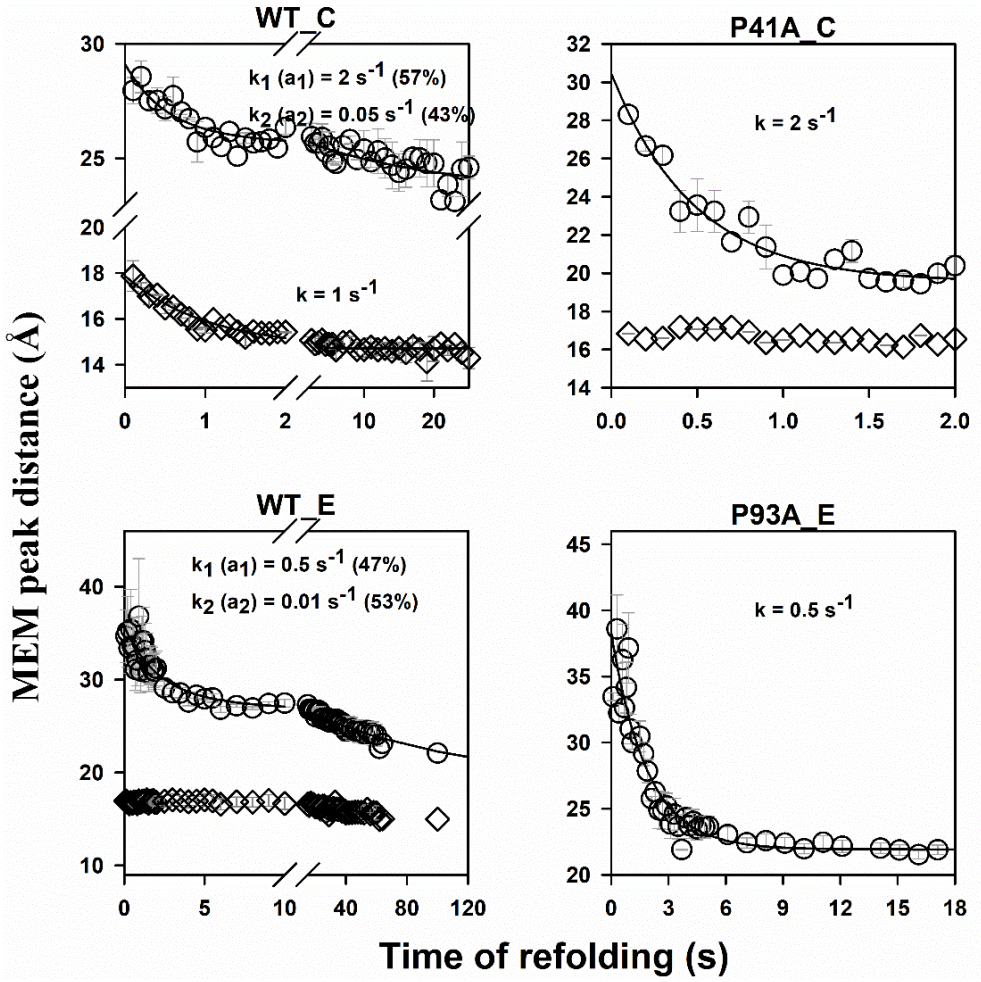


Figure 3.13. Kinetics of contraction of the U-like (circles) and N-like (diamonds) populations for different mutant variants of MNEI during folding in 0.4 M GdnHCl. For each mutant variant, the peak distance at each folding time was obtained using the MEM peak value of the U-like population and the N-like population in the MEM-derived fluorescence lifetime distribution obtained for the TNB-labeled variant (Figure 3.9) and the corresponding unlabeled counterpart (Figure 3.8) using equations 5 and 6 (see Materials and Methods). The solid lines through the kinetic data represent fits to either a single or double exponential equation. The error bars represent the standard errors of measurements from at least two independent kinetic experiments. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021b).

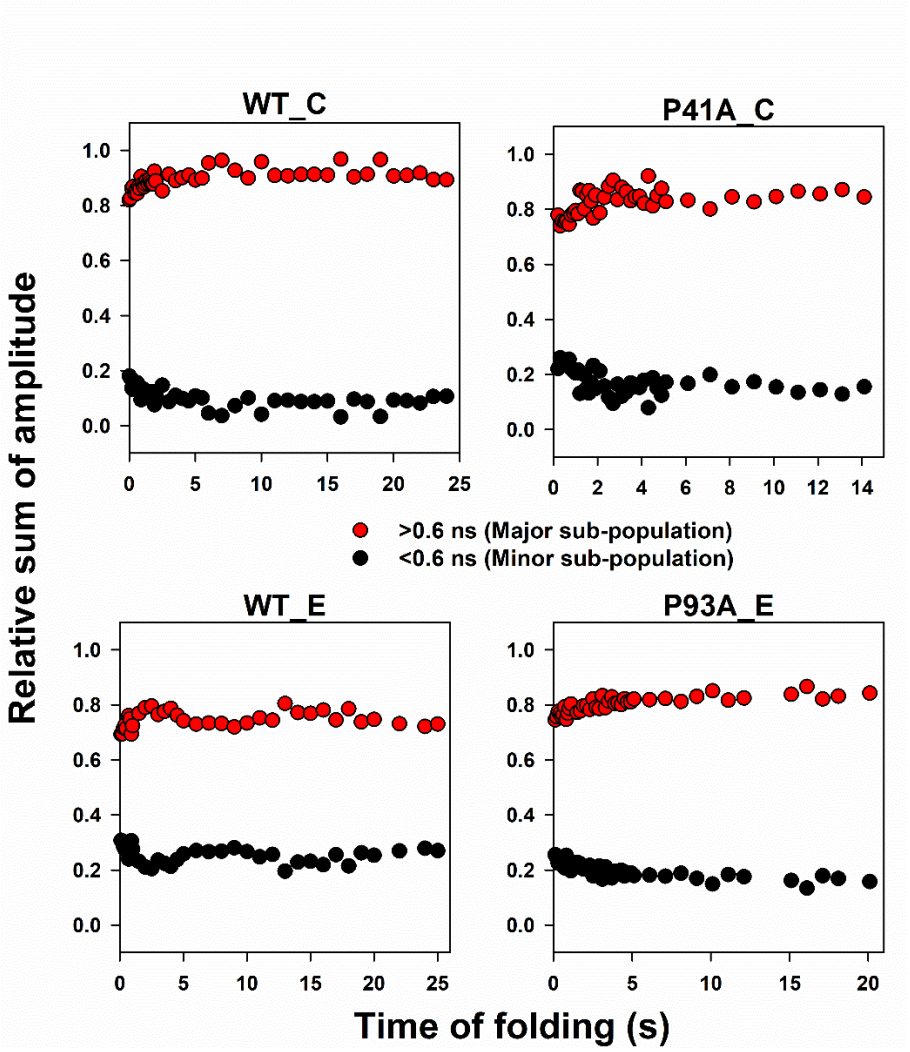
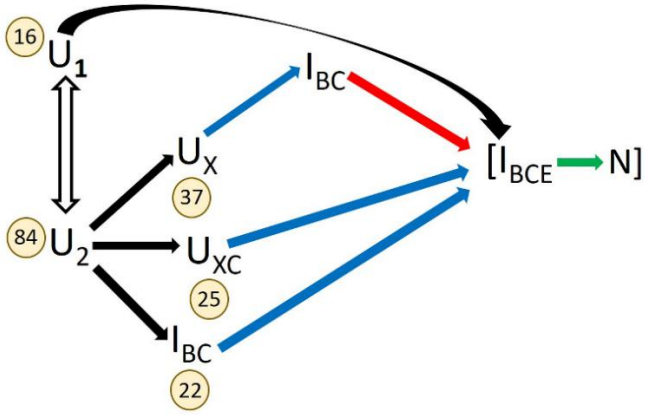


Figure 3.14. The minor short-lifetime sub-population in the MEM distributions of the unlabeled variants remains a small fraction throughout the folding reaction. The circles denote the relative sum of amplitudes of the two sub-populations obtained from MEM distributions of the unlabeled variants. The relative sum of amplitude for the long-lifetime sub-population (>0.6 ns; red circles) and the short-lifetime sub-population (<0.6 ns; black circles) was calculated as the sum of amplitudes of the respective sub-population distribution divided by the sum of amplitudes of the entire distribution. The relative sum of amplitudes corresponds to the fractional population of the respective sub-populations. Panels WT_C and WT_E have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021b).

The P41A mutation does not affect the heterogeneity in the unfolded state (Kaushik and Udgaonkar 2023); hence, the 63% molecules that have collapsed at 100 ms in the case of P41A_C would include the 16% which folded during the very fast phase (originating from U_1 sub-ensemble) leading to the formation of I_{BCE} (Figure 3.1c). Since the slow phase of contraction at the C segment was absent for P41A_C (Figure 3.12, Table 3.2), it seems likely that the 22% molecules that contract slowly in the case of WT_C (Figure 3.1c), instead collapse during the burst phase on the same pathway. To account for the remaining 25% (63-(16+22)) molecules that collapsed at the C segment during the burst phase in the case of P41A_C, it is necessary to posit that the mechanism of folding/contraction for WT_C (Figure 3.1c) has been altered by the P41A mutation to that shown below in Scheme 1.



Scheme 1

Several salient features of Scheme 1 need to be noted. (1) In the case of WT MNEI (WT_C), the kinetic data had indicated that the 100 ms burst phase led to the formation of two burst phase intermediates, U_X and I_B, from U_2, and that intermediates I_C and I_{BC} formed by kinetic partitioning from U_X (Figure 3.1c) (Bhatia, Krishnamoorthy, and Udgaonkar 2021a). It seems now, that U_X is not homogeneous, but consists of two sub-populations U_X and U_{XC}. In the case of WT_C, U_{XC} is populated to a negligible, undetectable extent, but in the case of P41A_C, it becomes populated to 25%. It appears that the P41A mutation significantly stabilizes U_{XC} by relieving the conformational strain imposed by Pro41. Stabilization may also occur *via* an intramolecular hydrogen bond, in which the introduced Ala, unlike Pro, can participate; such bonds have been suggested to facilitate initial polypeptide chain collapse (Möglich, Joder, and Kiefhaber 2006; Teufel et al. 2011; Nettels et al. 2009; Rose et al. 2006). (2) The stabilization afforded by the mutation is not large, and 37% molecules remain as U_X at the end of the burst phase. In these molecules, the C segment becomes N-like only at the end

of the fast phase of folding, along with the B segment. Hence, the fast phase of contraction at the C segment seen for WT_C is also seen for P41A_C. (3) The C segment, which was U-like in I_B in the case of WT_C, is stabilized by the P41A mutation so that it becomes N-like in the case of P41A_C; hence, I_B is shown as I_{BC}. Consequently, the slow phase of contraction at the C segment seen for WT_C (Figure 3.1c) was not seen for P41A_C. The effects of the P41A mutation described above are consistent with known effects of mutations in modulating the kinetic partitioning between competing folding pathways by stabilizing or destabilizing specific intermediates (Chiti et al. 2002; Sánchez, Ferreiro, and de Prat Gay 2011; Aghera and Udgaonkar 2013; Karamanos et al. 2016). (4) It is not known whether the P41A mutation has a similar effect on the B segment (β-sheet). It will be important to investigate this in the future, especially since residues in the β-sheet contribute to the protein core, and since studies with other proteins have reported coupling of β-sheet formation with hydrophobic core packing (Vu, Brewer, and Dyer 2012; Narayanan and Dias 2013). It is therefore not surprising that the β-sheet and core both become N-like together on the $U_2 \rightarrow I_{BC} \rightarrow I_{BCE}$ and $U_2 \rightarrow U_X \rightarrow I_{BC} \rightarrow I_{BCE}$ pathways.

The Pro93 to Ala mutation stabilizes N-like structure at the E segment in the products of the fast phase of collapse

In the case of WT_E, only 22% of the molecules were found to contract their end-to-end distance (the E segment) during the fast phase of folding when they form I_{BE} (Figure 3.1c). The contraction of the E segment in the remaining molecules was found to occur primarily during the slow phase (Figures 3.1c and 3.12; Table 3.2). However, in the case of P93A_E, 75 % molecules were found to contract during the fast phase, and only about 10% did so in the slow phase. Hence, the P93A mutation leads to the contraction of an additional 53% (75-22) of molecules at the E segment to become N-like during the fast phase of contraction. To account for these additional 53% molecules, it becomes necessary to propose that I_{BC} and I_C (Figure 3.1c) also exist in equilibrium with a so far hidden I_{BCE} and a hidden I_{CE}, respectively. Because of the strain introduced in the polypeptide backbone by Pro93, I_{BCE} and I_{CE} are unstable and hence populated to an insignificant extent in the case of WT_E. In the case of P93A_E, they became stabilized and populated to a larger extent because of the replacement of Pro93 with Ala. The P93A mutation therefore leads to conformational selection of I_{BCE} and I_{CE} such that they are together populated to about 53%, and I_{BC} and I_C are consequently together populated

to only about 10 %. Thus, the 75 % molecules contracting their E segment in the fast phase in the case of P93A_E, are those which form I_{BE}, I_{CE} and I_{BCE}.

It should be noted that *trans* to *cis* isomerization of Pro93 occurs during the slow phase of folding of WT_E (Kaushik and Udgaonkar 2023). In this context, an alternative explanation could have been that the contraction that occurs during the slow phase of folding of WT_E is only that which accompanies the *trans* to *cis* isomerization of the Gly92-Pro93 bond, which brings the FRET acceptor, Cys97TNB as well as the residues 93-97 closer to the FRET donor, Trp4 and the core of the protein (Figure 3.1b). In the case of P93A_E, the Gly92-Ala93 bond remains *trans*, and does not undergo *trans* to *cis* isomerization; consequently, the FRET acceptor does not move towards the donor (Figure 3.1b), and no slow phase contraction would be observed. While this explanation would have been difficult to rule out if only FRET measurements had been carried out, it can be ruled out now based on the MEM analysis which showed that about 10% molecules still contract during the slow phase in the case of P93A_E (Figure 3.12, Table 3.2).

Thus, the result that the P93A mutation alters the population of intermediates by stabilizing the E segment in I_{CE} and I_{BCE}, which were populated only insignificantly in the case of WT_E, mirrors the effects observed for the P41A mutation (see above). This suggests that Pro41 and Pro93 dictate the partitioning of pathways, and play a crucial role in determining the stability of specific intermediates and hence, the heterogeneity of the intermediate ensemble. Similar proline-mediated modulation of pathway partitioning has also been observed for other proteins (Cook, Schmid, and Baldwin 1979; Osváth and Gruebele 2003; Taler-Verčič et al. 2017), but has not been directly quantified. This is perhaps the first experimental study to directly quantify the redistribution of folding intermediates arising from the gain in backbone flexibility consequent to replacing rigid Pro with flexible Ala. By quantitatively tracking shifts in the distribution of coexisting sub-populations within intermediate ensembles, this study provides direct evidence that the stability and partitioning of sub-populations are tunable features encoded in the primary sequence.

In summary, the Pro41 to Ala mutation has revealed hidden heterogeneity in the size of the C segment in the product of initial collapse. Similarly, the Pro93 to Ala mutation has revealed hidden heterogeneity in the size of the E segment in a late intermediate. In previous studies of the folding of barstar, it had been shown that the product of initial collapse was also heterogenous, with different structural segments collapsed to different extents (Sinha and

Udgaonkar 2005, 2007). It was shown that this collapsed state at a few milliseconds of folding also contained sub-populations of differently structured molecules, and that different structured sub-populations could become stabilized and hence be populated to detectable extents in different folding conditions (Pradeep and Udgaonkar 2002, 2004a). Previous trFRET studies of the slow folding of barstar had shown that the structural heterogeneity of a late folding intermediate could also be modulated by a change in folding conditions (Sridevi et al. 2004). The current study demonstrates that structural heterogeneity in intermediate ensembles can be modulated by site-specific mutations that relieve local backbone rigidity, thereby redistributing molecules among distinct sub-populations (see above). Consequently, backbone rigidity emerges as a key determinant of how structural heterogeneity is modulated during protein folding.

Decoupling of structure formation and chain compaction during folding

Different segments of the polypeptide chain of WT MNEI were shown to collapse to different extents during the burst phase of folding, suggesting that the product of the burst phase is structurally heterogeneous (Goluguri and Udgaonkar 2015). It was subsequently determined that the product of folding at 100 ms is also heterogeneous, comprising sub-populations of molecules that differed in which of the B, C and E segments had become N-like in dimensions (Bhatia, Krishnamoorthy, and Udgaonkar 2021b), even though insignificant secondary structure has formed at this time (Goluguri and Udgaonkar 2015; Goluguri and Udgaonkar 2016; Bhatia et al. 2019). These results indicated that initial collapse at 100 ms of folding is decoupled from structure formation. In this study, it is seen that the Pro41 to Ala mutation further decouples initial collapse at the C segment from structure formation, by increasing the number of molecules becoming N-like in their dimensions during initial collapse (Figure 3.12). This increase in decoupling because of the P41A mutation suggests that rigid proline residues may serve as local gatekeepers of the initial collapse process. While initial collapse may be hydrophobic in nature, it also appears to be under precise, residue-level control.

Importantly, the increase in the number of molecules undergoing initial collapse occurred at the expense of fewer molecules contracting during the subsequent phases, even though the number of molecules undergoing a change in structure in each phase, measured by the change in fluorescence of Trp19, was affected only marginally (Table 3.2). In this context, it should be noted that the rate constants of the change in fluorescence were found to match those of the change in far-UV circular dichroism, for WT_C and WT_E, as well as for other

MNEI variants (Goluguri and Udgaonkar 2015). Similarly, it was also seen that the Pro93 to Ala mutation increases the fraction of molecules becoming N-like in dimensions during the fast phase of contraction, at the expense of the slow phase (Figure 3.12) even though the number of molecules undergoing a change in structure in each phase, was affected only marginally (Table 3.2). The very fast phase of folding observed for WT_E (~25% of molecules) is abolished in the case of P93A_E (Kaushik and Udgaonkar 2023), with these molecules now folding during the fast phase. Thus, while the overall number of molecules folding remains unchanged, the extent of contraction during the fast phase is clearly increased. Thus, it appears that while an intermediate ensemble may consist of sub-populations that have compacted differently in different segments, the sub-populations nevertheless do not differ in the extent of structure that has formed (schematically illustrated in Figure 3.15). Together, these results show that the Pro to Ala mutations lead to more contracted sub-populations being favored, without significantly altering the extent of structure formed. This decoupling of chain contraction from structure formation demonstrates that the two processes, though often concurrent, are separable events during folding.

Local backbone rigidity therefore appears to modulate the extent of compaction occurring at different stages of folding, without affecting the extent of structure formed. These effects, although mechanistically different for the two Pro to Ala mutations, demonstrate collectively that structure formation can be decoupled from chain contraction during all kinetic phases of folding, just as it is known to be decoupled during the initial collapse reaction. This appears to be first study that directly demonstrates the decoupling of chain contraction and structure formation during the later stages of folding. The results reveal that folding occurs through separable physical events, chain compaction and structure formation, which are not obligatorily coupled to each other, even at later stages of folding.

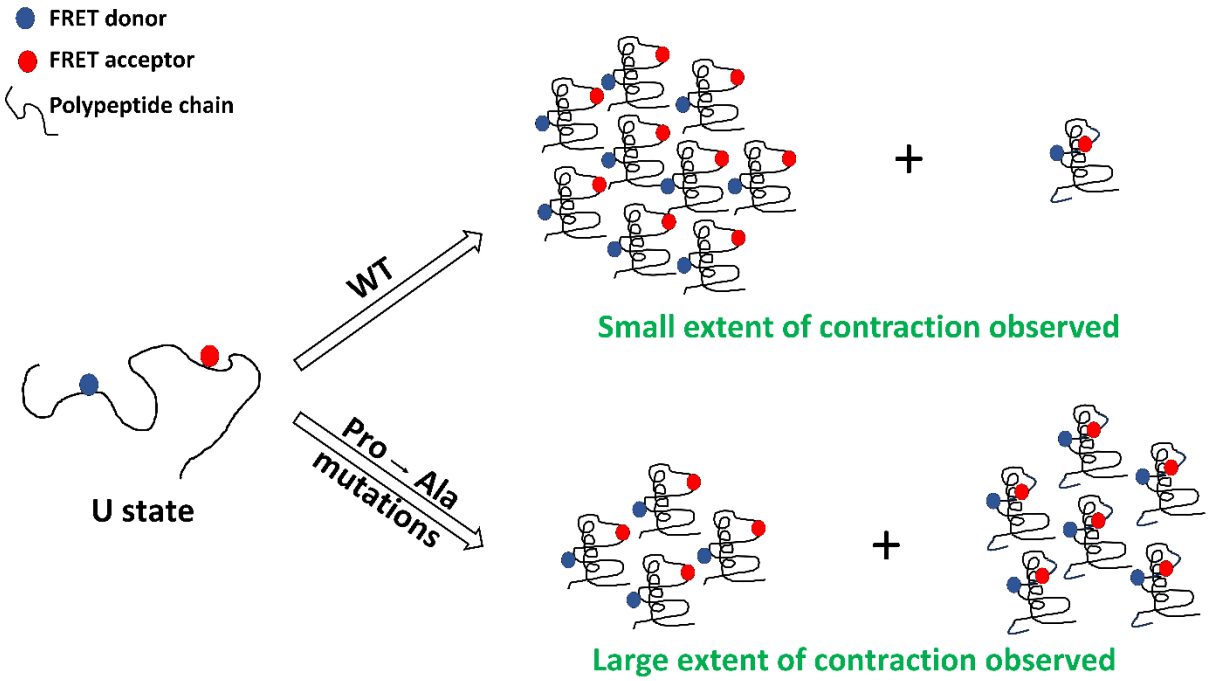


Figure 3.15. Schematic illustrating the effect of Pro to Ala mutations on the heterogeneity of an intermediate ensemble during the folding of MNEI. The mutations shift molecules toward more compact sub-populations, while the extent of structure formed remains largely unaffected. Symbols for donor, acceptor, and polypeptide chain are shown on the left. Numbers of molecules are illustrative and not exact.

Contraction occurs gradually during folding and is accelerated because of Pro to Ala mutations

The trFRET measurements showed that the U-like sub-population contracts to an N-like sub-population during the fast and slow phases of folding, as well as during an unobservable 100 millisecond burst phase in the case of WT_C and WT_E (Figure 3.12, Table 3.2). These measurements also showed that each kinetic phase does not represent a two-state transition but is characterized by molecules contracting in a gradual manner (Figure 3.13). In the case of WT_C and WT_E, both the N-like and U-like sub-populations underwent gradual contraction during folding, as evident in the gradual shift in their MEM-derived distance distributions, and their peaks, toward shorter distances (Figures 3.11 and 3.13). An earlier report (Bhatia, Krishnamoorthy, and Udgaonkar 2021b) attributed this contraction to continuous breakage of non-native interactions. Surprisingly, however, in the case of P41A_C, the N-like sub-population did not undergo any contraction during folding (Figure 3.13). It seems that the release of strain in the polypeptide backbone, and/or introduction of additional hydrogen bonding, consequent to replacing Pro41 with Ala, allowed the N-like sub-population to relax to the dimensions of the final N state as soon as it was formed. Moreover, the U-like sub-population for P41A_C undergoes contraction only during the fast phase (the slow phase was absent, see above and Figure 3.12). This indicates that compaction at the C segment is accelerated as a result of the Pro41 to Ala mutation. In the case of WT_E, the gradual nature of the slow compaction phase may be modulated by the *trans*-to-*cis* isomerization of the Gly92–Pro93 peptide bond. In the case of P93A_E, for which contraction was seen to occur only during the fast phase of folding, this contraction was also observed to be gradual in nature, and not to be all-or-none (Figure 3.16).

Nonetheless, for both Pro mutant variants, the U-like sub-population contracted gradually while at the same time undergoing first order transitions to become N-like, as also seen for the WT proteins (Figures 3.12 and 3.13). Similarly, gradual exposure of amide sites to solvent had been observed to occur in multiple kinetic phases in hydrogen exchange-mass spectrometry (HX-MS) studies of the unfolding of MNEI in native and native-like conditions (Malhotra and Udgaonkar 2015; Malhotra and Udgaonkar 2016b), in which folding would occur by the same mechanism in reverse. The HX-MS studies were complemented by SH-labeling studies which revealed that each kinetic pause during folding in native conditions occurs because a specific tight packing interaction has to form, and that interaction forms in an all-or none manner (Malhotra and Udgaonkar 2014).

The shift in the MEM peak position provides only qualitative insight into compaction (Figure 3.13). Acceleration of the kinetics of compaction is obvious in the case of the P41A_C, as a slow phase of compaction was not observed (Figure 3.12; Table 3.2). To quantify the mutation-induced acceleration of the compaction of the E segment during the fast and slow kinetic phases, the apparent time constants (determined as the amplitude-weighted average of the time constant, $\tau_{av} = (a_f\tau_f + a_s\tau_s)/(a_f + a_s)$; $\tau_i = 1/k_i$) were calculated for WT_E and P93A_E, where τ_f and τ_s are the time constants of the fast and slow kinetic phases, and a_f and a_s are their respective relative amplitudes (Figure 3.12 and Table 3.2). The apparent time constants for WT_E and P93A_E were 20 s and 2 s, respectively, indicating a ten-fold acceleration in compaction due to the P93A mutation. The acceleration at both the C and E segments may arise, in part, from the replacement of the rigid Pro residue with a more flexible Ala, facilitating faster chain contraction, as previously observed (Krieger et al. 2003; Krieger, Möglich, and Kiefhaber 2005). Moreover, early stabilization of structure in both segments (see above) by Pro to Ala mutation, may pre-organize local and/or non-local conformations, leading to a reduction in the conformational degrees of freedom and enhancement in the efficiency of subsequent conformational search.

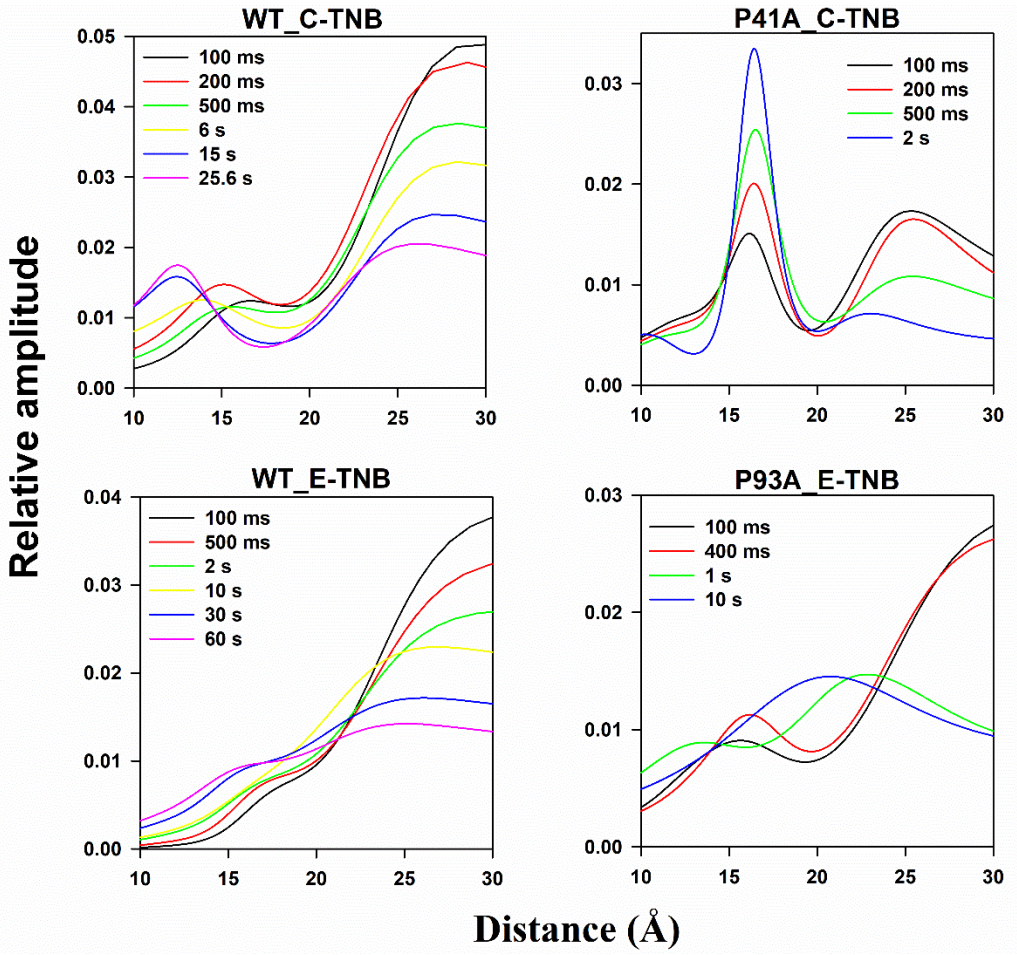


Figure 3.16. MEM-derived distance distributions at different times of folding of the different labeled variants of MNEI. In each case, the distributions have multiple cross-over points, indicative of a non-two state transition. Distributions corresponding to different times of folding are shown in different colours, as described for each panel. The y-axis in each panel represents the relative amplitude normalized to the sum of amplitudes for each distribution to make the total population fraction equal to 1. Panels WT_C-TNB and WT_E-TNB have been adapted and modified from (Bhatia, Krishnamoorthy, and Udgaonkar 2021b).

3.5 Conclusion

Site-specific Pro to Ala mutations reveal how the heterogeneity within intermediate ensembles modulates the coupling between chain compaction and structure formation during the folding of MNEI. The P41A mutation stabilized a minor sub-population of molecules during the initial collapse process, while the P93A mutation stabilized late intermediates during the fast phase of compaction at the C and E segments, respectively. Stabilization of these sub-populations led to pathway partitioning and redistribution of intermediate sub-populations, resulting in a decoupling of structure formation from chain compaction during all kinetic phases of folding. These findings highlight a general principle: sub-population heterogeneity, encoded by local backbone rigidity, is a key determinant of whether and how compaction and structure formation are dynamically coupled during folding, challenging the traditional view that these processes are obligatorily coupled during the later stages of folding.

Chapter 4

Chain entropy modulates cooperativity selectively within intermediate sub-populations during protein unfolding

Reproduced with permission from:

Kaushik, A. and Udgaonkar, J.B., 2026. Chain entropy modulates cooperativity selectively within intermediate sub-populations during protein unfolding. *Biochemistry.* DOI: 10.1021/acs.biochem.6c00188.

4.1 Introduction

The extent and origin of cooperativity in protein folding and unfolding reactions are not yet fully understood. Folding transitions are often described as "two-state" when monitored by ensemble-averaging probes, which typically report sigmoidal transitions between the native (N) and unfolded (U) states (Privalov 1979; Jackson and Fersht 1991a; Eaton et al. 2000). This seeming all-or-none behaviour led to the view that proteins have evolved to fold and unfold cooperatively, so that the population of partially folded intermediates that could be aggregation-prone, is minimal (Canet et al. 2002; Dobson 2003). A two-state description, however, does not exclude the possibility that high-energy intermediates are populated transiently during folding. When folding is probed by more sensitive techniques, partially folded intermediates have been detected and structurally characterized, and in several cases intermediates stable enough to accumulate to measurable extents, have been identified (Roder, Elöve, and Englander 1988; Bai et al. 1995; Lakshmikanth et al. 2001; Akiyama et al. 2002; Jha et al. 2009; Malhotra and Udgaonkar 2016b; Bhatia, Krishnamoorthy, and Udgaonkar 2018; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). Indeed, even when the thermodynamic and kinetic criteria for two-state folding are satisfied, partially unfolded intermediates may still be sparsely populated, and different structural segments may lose structure gradually or asynchronously.

A true two-state conformational change in a protein would be one in which different structural regions undergo synchronous change. But unfolding under equilibrium conditions can appear to be two-state even when (i) different structural regions of the protein have distinct local stabilities that are distributed around the global stability, resulting in unfolding being heterogeneous and not fully cooperative (Holtzer et al. 1997; Sadqi, Fushman, and Munoz 2006; Chatterjee et al. 2007; Julien et al. 2009; Krishna Mohan, Chakraborty, and Hosur 2009; Sborgi et al. 2015), and (ii) certain regions undergo gradual unfolding, in which structural order is lost progressively (Kuzmenkina, Heyes, and Nienhaus 2006; Jha et al. 2009; Malhotra and Udgaonkar 2015; Sabelko, Ervin, and Gruebele 1999; Song et al. 1999; Lakshmikanth et al. 2001; Garcia-Mira et al. 2002; Sinha and Udgaonkar 2008; Waldauer et al. 2008; Wani and Udgaonkar 2009). In such cases, the non-cooperative behaviour of some structural elements may remain hidden beneath the apparent global cooperativity reported by ensemble-averaging probes, and be revealed only by site-specific, high-resolution methods such as single-molecule fluorescence resonance energy transfer (sm-FRET) (Kuzmenkina, Heyes, and Nienhaus 2006), time-resolved FRET (Sridevi et al. 2004; Kishore, Krishnamoorthy, and Udgaonkar 2013; Halloran et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2018) and hydrogen exchange

coupled to mass spectrometry (HX-MS) (Malhotra and Udgaonkar 2015; Arrington, Teesch, and Robertson 1999; Malhotra and Udgaonkar 2016b; Bhattacharjee and Udgaonkar 2021). For several proteins, these methodologies have revealed that (un)folding can be slowed down not just by a dominant high free energy barrier, but also by a rugged landscape with many distributed small free energy barriers, (McCammon, Gelin, and Karplus 1977; Bryngelson et al. 1995; Henzler-Wildman and Kern 2007) leading to gradual, asynchronous structural transitions (Lakshmikanth et al. 2001; Rami and Udgaonkar 2002; Rami, Krishnamoorthy, and Udgaonkar 2003; Jha et al. 2009; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2018). What remains unclear is whether deviations from two-state behaviour originate solely within secondary structural elements or whether they also involve long-range interactions. How the intrinsic differences in cooperativity of formation of different structural motifs are influenced by their placement within the overall protein architecture remains poorly understood. Moreover, whether multi-chain proteins display distinctive cooperative or non-cooperative unfolding patterns compared to their single-chain counterparts is an open question.

The small protein monellin is an ideal model system for investigating folding cooperativity, as it exists in two closely related forms: a naturally occurring heterodimer (dcMN) (Aghera and Udgaonkar 2011; Aghera, Earanna, and Udgaonkar 2011), and an artificially created single-chain variant (MNEI) (Patra and Udgaonkar 2007) in which the C-terminus of chain B (in β2) is linked covalently to the N-terminus of chain A (in β3) *via* a short Gly-Phe peptide linker (Figure 4.1a). The native-state structures of dcMN and MNEI are nearly identical (Kim, de Vos, and Ogata 1988; Aghera, Earanna, and Udgaonkar 2011) except at the linker region (Figure 4.1a). These two variants of monellin offer a unique opportunity to directly compare the unfolding cooperativity of single-chain and double-chain architectures, while minimizing confounding differences in sequence or structure, thereby isolating the effect of chain connectivity.

It is known that MNEI exhibits deviations from two-state behaviour: kinetic and equilibrium studies using multisite tr-FRET and HX-MS have revealed continuous, segmental expansion of the β-sheet during unfolding (Jha et al. 2009; Jha and Udgaonkar 2009) and limited cooperativity (Malhotra and Udgaonkar 2016b; Malhotra and Udgaonkar 2015; Bhatia, Krishnamoorthy, and Udgaonkar 2018). Other studies have shown asynchronous folding of the helix (Goluguri and Udgaonkar 2015), and the transient population of metastable intermediates (Goluguri and Udgaonkar 2016; Bhatia et al. 2019; Bhatia, Krishnamoorthy, and Udgaonkar 2021a) that further underscore the heterogeneous and non-cooperative nature of the folding and unfolding pathways. dcMN appears to differ from MNEI in its stability (Aghera, Earanna,

and Udgaonkar 2011; Bhattacharjee and Udgaonkar 2021) as well as in its folding cooperativity (Malhotra and Udgaonkar 2015; Bhattacharjee and Udgaonkar 2021; Malhotra, Jethva, and Udgaonkar 2017), and it is not known whether these differences arise because MNEI is an artificially created protein. An HX-MS study on dcMN (Bhattacharjee and Udgaonkar 2021) showed that unfolding under native and mildly denaturing conditions involves three non-cooperative kinetic phases followed by a slower cooperative phase. Interestingly, while only β3 in chain A unfolds cooperatively under strongly native conditions (≤0.4 M GdnHCl), both β3 and β2 in chain B unfold cooperatively under mildly denaturing concentrations (0.6–1 M GdnHCl). However, HX-based methods probe only those regions that display measurable protection, and typically provide little or no information on more than half of the structure, because that offers little or no protection against HX.

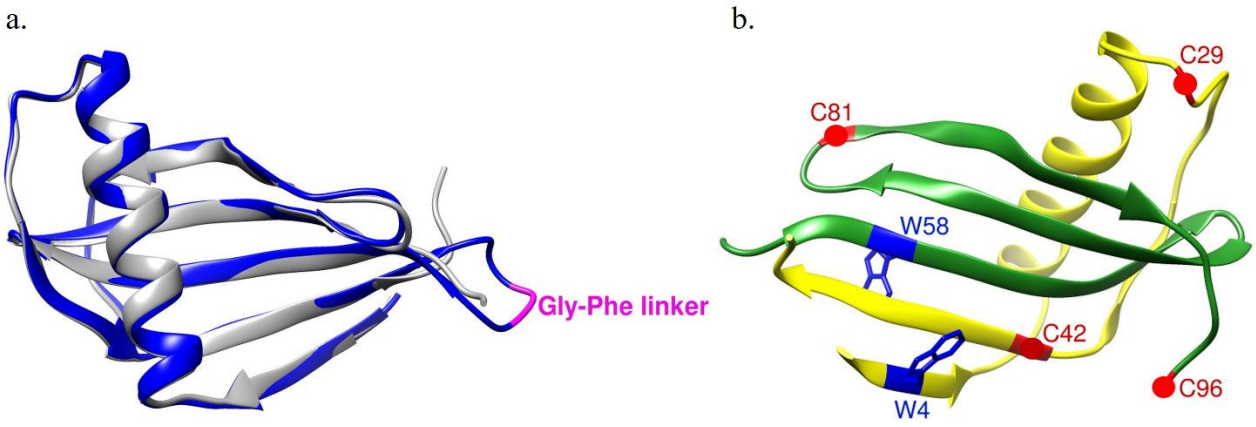


Figure 4.1. a) Structural comparison of MNEI and dcMN. The Gly–Phe linker (shown in pink) joins the two chains of dcMN (shown in grey) to yield MNEI (shown in blue). b) Structure of dcMN showing the different residues used for monitoring FRET. Chain B of dcMN is shown in yellow, while chain A is shown in green. Wild-type dcMN has one Trp residue (W4) and one Cys residue (C42). The side chains of the Trp residues (used as the FRET donor) are shown as blue rings, and the Cys residues (to which the FRET acceptor TNB was attached) are shown as red spheres. FRET was measured in single tryptophan-containing, single cysteine-containing variants. FRET changes were monitored within chain B (in W4C42), within chain A (in W58C81), between the two chains (in W4C96), and across the helix (in W4C29) present in chain B. The distances between the centre of the Trp ring and Cβ atoms of the Cys residue (attached to the acceptor) in the N state, were 13.0 Å (W4C42), 14.3 Å (W58C81), 16.0 Å (W4C96) and 22.5 Å (W4C29). The structure has been drawn using the Chimera software and the PDB IDs 1IV7 and 3MON.

As a result, intrinsically flexible segments and transiently unprotected regions, which may include parts of the inter-chain dimer interface, are often not captured, even though they may play a key role in coordinating conformational transitions during unfolding.

To understand how protein architecture and chain connectivity modulate folding cooperativity and heterogeneity, MEM-coupled time-resolved FRET measurements were used together with site-specific fluorescence anisotropy measurements to study the unfolding of dcMN under equilibrium conditions. Four FRET pairs were designed: two intra-chain pairs to map distance distributions within the structural cores of chain A (in W58C81) and chain B (in W4C42), one inter-chain pair (in W4C96) to monitor the distance distribution across the two chains, and another (in W4C29) to map the helix segment (Figure 4.1b). Measurements of the time-resolved fluorescence anisotropy decay of Trp58 in W58C81 provide complementary insights into the motional dynamics at the dimer interface and within chain A, while those of Trp4 in W4C42 report on chain B. These approaches together reveal localized deviations from two-state behaviour that remain hidden when ensemble-averaging probes are utilized to study unfolding. By following the GdnHCl concentration-dependent evolution of MEM-derived lifetime distributions and anisotropy-derived dynamics across intra-chain cores, inter-chain contacts, the helix, and the dimer interface, the structural basis of unfolding cooperativity in dcMN has been delineated.

4.2 Materials and Methods

Protein expression, purification, and TNB labelling

The wild-type protein (W4C42) has a single tryptophan (Trp4) residue located in the first β-strand and a single buried cysteine (Cys42) in the second β-strand of chain B. Site-directed mutagenesis was used to generate the single-tryptophan, single-cysteine containing mutant variants: C42AQ29C (W4C29), C42AP96C (W4C96), and W4YY58WC42AT81C (W58C81). Protein purification was carried out following an established protocol (Aghera and Udgaonkar 2011). The single cysteine residue in each protein variant was conjugated with the acceptor thionitrobenzoate (TNB) moiety. Labeling was performed as previously described (Jha et al. 2009). In brief, the protein was unfolded in 2 M GdnHCl and incubated with DTNB solution at pH 8 for ≥ 2 h; after which excess label was removed by desalting. The purity of each protein preparation was confirmed by electrospray ionization mass spectrometry (ESI-MS). The mass of the labeled protein showed the expected increase of 197 Da, corresponding to the mass of the TNB group, and the extent of labeling was > 95%. The concentrations of the unlabeled protein solutions were determined from absorbance measurements at 280 nm, using

the molar extinction coefficient value of 14 600 M^{-1} cm^{-1}. For the TNB-labeled proteins, the contribution of the TNB absorbance to the total absorbance at 280 nm was 20% for all TNB-labeled protein variants, except for W4C96-TNB, for which the contribution was 10% (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

Reagents

All the experiments were carried out at pH 8.0 and 25ºC. The reagents used in the experiments were of the highest purity grade from Sigma. Guanidine hydrochloride (GdnHCl) was purchased from Thermo Scientific, and was of the highest purity grade. The native buffer contained 50 mM Tris and 250 μM EDTA, and unfolding buffer contained, in addition, GdnHCl. 1 mM DTT was added for all experiments with unlabeled proteins. The concentrations of GdnHCl solutions were determined by the measurements of the refractive index on an Abbe 3L refractometer from Milton Roy. All buffers and solutions were passed through 0.22 μm filters before use. A protein concentration of 5 μM was used for all experiments.

Measurement of fluorescence and far UV CD spectra

Fluorescence spectra were collected on a Fluoromax 4 (Horiba) spectrofluorometer. The protein samples were excited at 295 nm, and the emission spectra were collected from 310 to 450 nm, with a response time of 1 s, and excitation and emission bandwidths of 1 nm and 5 nm, respectively. Each spectrum was recorded as the average of three fluorescence emission wavelength scans. Measurements of CD spectra were carried out on a Jasco J-815 spectropolarimeter. Far-UV CD spectra were collected using a 0.2 mm pathlength cuvette, with a bandwidth of 1 nm, a response time of 1 s, and a scan speed of 20 nm/min. Thirty wavelength scans were averaged for each spectrum.

Steady-state fluorescence and anisotropy measurements

For equilibrium unfolding studies, protein samples were incubated in different concentrations of GdnHCl (0 to 2 M) for ≥ 48 h prior to measurement. The fluorescence intensities of both unlabeled and TNB-labeled protein variants for steady-state FRET measurements were measured using the MOS-450 optical system (Biologic). The protein samples were excited at 295 nm, and the emitted fluorescence was monitored at 360 nm using a 10 nm band-pass filter (Asahi Spectra). The excitation slit width was kept at 2 nm. For each sample, the data were acquired and averaged for 30 s. The fluorescence anisotropies of Trp4 in

W4C42 and Trp58 in W58C81 were measured using a Fluoromax 4 (Horiba) spectrofluorometer by monitoring the emission at parallel and perpendicular polarizations simultaneously using the T-format optical arrangement.

Time-resolved fluorescence and anisotropy measurements

Fluorescence intensity decay curves for the four pairs of unlabeled and labeled proteins, and fluorescence anisotropy decay curves for unlabeled W4C42 and W58C81 were acquired after equilibrating the samples at different GdnHCl concentrations for ≥ 48 h prior to measurements.

Fluorescence lifetime and anisotropy measurements were carried out using an excitation wavelength of 295 nm, generated by frequency-tripling 885 nm femtosecond pulses from a Ti:sapphire laser. Emission at 360 nm was detected using hybrid PMTs (HPM-100-40 or HPM-100-07). Time-correlated single-photon counting (TCSPC) was employed to record decay profiles, and instrument response functions (IRF) were recorded for deconvolution. A polarizer (Glan-Thompson), set at 54.71° (magic angle) with respect to excitation, was kept in front of the long-pass filter to avoid polarization effects. In time-resolved anisotropy decay measurements, the emission was collected at directions parallel ($I_{\parallel}$) and perpendicular ($I_{\perp}$) to the polarization of the excitation beam, for the same duration. The IRF was measured using light scattered from a LUDOX solution. The HPM-100-07 consistently gave an IRF FWHM of ~60 ps, and the HPM-100-40 detector gave an IRF FWHM of ~100 ps. The fluorescence lifetime of NATA (2.5–2.8 ns) in water was routinely recorded as a standard. All decay transients were collected to a peak count of 20 000, and up to 99.9% of completion. Detailed descriptions of the instrumentation for lifetime (Bhatia, Krishnamoorthy, and Udgaonkar 2018) and anisotropy (Swaminathan et al. 1996; Saxena, Udgaonkar, and Krishnamoorthy 2006; Sarkar, Udgaonkar, and Krishnamoorthy 2013) measurements have been reported previously.

Data analysis

Analysis of the fluorescence intensity decay traces: Details of the analysis are given in an earlier study (Bhatia, Krishnamoorthy, and Udgaonkar 2018). A brief description is given below:

Discrete analysis: The decay traces were fit to a sum of exponentials,

$$F(t) = \sum_{i=1}^{n} \alpha_i e^{-\left(\frac{t}{\tau_i}\right)} \quad \textbf{(1)}$$

Here, α_i is the relative amplitude of the τ_i lifetime component, t is time, and n ranges from 3 to 4.

The amplitude-weighted mean lifetime, $\boldsymbol{\tau}_m$ was caculated as:

$$\boldsymbol{\tau}_m = \frac{\sum \alpha_i \tau_i}{\sum \alpha_i};\ \sum \alpha_i = 1 \quad \textbf{(2)}$$

MEM analysis: The analysis began with the assumption that the decay corresponds to a distribution of 100-150 lifetimes in the range 1 ps to 10 ns. The *a priori* distribution of lifetimes was assumed to be uniform in the logarithms of lifetimes being uniformly distributed in this range. Then, the best fit values of α_i and τ_i were determined using the Maximum Entropy Method (MEM) (Jha et al. 2009; Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

The *a posteriori* distribution of these parameters was obtained by maximizing the Shannon Jaynes entropy S, defined as

$$S = -\Sigma\, P_i \log(P_i) \quad \textbf{(3)}$$

P_i is the normalized amplitude of the ith lifetime.

$$P_i = \frac{\alpha_i}{\sum \alpha_i} \quad \textbf{(4)}$$

α_i is the amplitude of the ith lifetime.

Analysis parameters were optimized for obtaining precise and reproducible MEM distributions. The most probable (MEM peak) lifetime refers to the lifetime corresponding to the maximum amplitude in the lifetime distribution data.

Fitting of MEM analysis-derived distributions to a two-state model

Fluorescence lifetime distributions obtained from MEM analysis were fitted to a two-state model using a nonlinear least-squares curve fitting algorithm implemented in MATLAB. This algorithm used the fluorescence lifetime distributions of the native state, N(℧), and the unfolded state, U(℧), as basis distributions, and carried out a nonlinear least-squares fit across all GdnHCl concentrations according to the equation:

$$Y(\tau) = f_N \times N(\tau) + f_U \times U(\tau) \quad \textbf{(5)}$$

where f_N and f_U represent the fractions of native and unfolded protein, respectively, at each GdnHCl concentration. The root mean-squared deviation (rmsd) at a given GdnHCl concentration (Figure 4.14) was calculated as the square root of the mean of the squared residuals across all lifetime values.

FRET efficiency and distance determination

The mean FRET efficiency (<E>) was obtained from mean fluorescence lifetimes of the unlabeled (<τ_D>) and the corresponding TNB-labeled (<τ_{DA}>) variants using the following equation:

$$<E> = 1 - \frac{<\tau_{DA}>}{<\tau_D>} \quad \textbf{(6)}$$

The mean fluorescence lifetimes were determined by discrete analysis of the fluorescence decay traces as described earlier (Jha et al. 2009; Bhatia, Krishnamoorthy, and Udgaonkar 2018).

The mean FRET efficiency value was converted to the average intramolecular distance (<R_{DA}>) using the Förster equation:

$$<R_{DA}> = R_0 \left(\frac{1 - <E>}{<E>}\right)^{1/6} \quad \textbf{(7)}$$

The values for the Forster's distance, R_0 used for the different protein variants are listed in Table 4.1.

Table 4.1. Energy transfer parameters of the native and unfolded states.

Protein	FRET efficiency	Quantum yield	Overlap Integral, J*10^ (-13)	Förster distance, R_o (Å)
W4C42_N	0.84	0.1130	7.5	23.4
W4C42_U	0.10	0.0851	8.7	23.6
W58C81_N	0.82	0.1346	8.9	24.3
W58C81_U	0.18	0.0900	9.1	22.7
W4C96_N	0.50	0.0996	8.1	22.6
W4C96_U	0.02	0.0915	9.2	22.8
W4C29_N	0.31	0.1053	9.5	23.5
W4C29_U	0.12	0.0920	9.7	23.0

Determination of the fractions of molecules in the N-like and U-like subpopulations from MEM distributions

To determine the fractions of N-like and U-like molecules from the MEM distributions (see introduction) at different GdnHCl concentrations, it was essential to take into account the observation that, for any segment, the equilibrium U state contained a fraction of molecules exhibiting lifetimes corresponding to an N-like distribution (<0.6 ns), while the equilibrium N state included a fraction of molecules exhibiting lifetimes corresponding to a U-like distribution (>0.6 ns) (Figures 4.8 and 4.11). These fractions were comparable for each pair of unlabeled (Figure 4.11) and corresponding labeled (Figure 4.8) unfolded proteins, suggesting that they originate from differences in the electronic structure of the fluorophore or distinct Trp rotamers (Beierlein et al. 2006; Maglia et al. 2008; Moors et al. 2008; Gasymov, Abduragimov, and Glasgow 2012) rather than from the presence of the quenching TNB moiety. The fraction of molecules expanded (U-like) at a given segment was determined using the equation: $f_U = Y_i - Y_N / Y_U - Y_N$, where Y_U represents the relative sum of amplitudes for the U-like distribution in the equilibrium U state, Y_N corresponds to the relative sum of amplitudes for the U-like distribution in the equilibrium N state, and Y_i denotes the relative sum of amplitudes for the U-like distribution at a given GdnHCl concentration. The relative sum of amplitudes was calculated as the sum of amplitudes for the U-like or N-like distribution divided by the total sum of amplitudes for both distributions. f_U derived from MEM-derived fluorescence lifetime

distributions (Figures 4.8 and 4.9), had been shown previously to accurately estimate the relative fractions of N-like and U-like molecules present together (Bhatia, Krishnamoorthy, and Udgaonkar 2021a).

Analysis of the fluorescence anisotropy decay traces

Time-resolved anisotropy decays were analyzed on the basis of the following equations.

$$I_{\parallel}(t) = \frac{1}{3} I(t) [I + 2 r(t)] \quad \textbf{(8)}$$

$$I_{\perp}(t) = \frac{1}{3} I(t) [1 - r(t)] \quad \textbf{(9)}$$

$$I(t) = \sum_i \alpha_i \exp\left(-\frac{t}{\tau_i}\right), \quad i = 1 - 3 \quad \textbf{(10)}$$

$$r(t) = r_o \sum_j \beta_j \exp\left(-\frac{t}{\tau_{rj}}\right), \quad j = 1 \; or \; 2 \quad \textbf{(11)}$$

where $I_{\parallel}$ and $I_{\perp}$ are the emission intensities collected at polarizations parallel or perpendicular to the polarization of the excitation beam. *I*(*t*) is the total fluorescence intensity at time *t*. r_o is the initial anisotropy, and α_i and β_j are the amplitudes associated with the *i*th fluorescence lifetime and *j*th rotational correlation time such that $\sum\alpha_i = \sum\beta_j = 1$. In this model, each τ_i is associated with both τ_{r1} and τ_{r2}. $I_{\parallel}$ and $I_{\perp}$ were analyzed globally (Swaminathan, Krishnamoorthy, and Periasamy 1994). It should be noted that the values of both rotational correlation times are viscosity-uncorrected, i.e., no correction was applied for changes in solvent viscosity with GdnHCl concentration. It should also be noted that no specific molecular shape was assumed in the anisotropy decay analysis, and the rotational correlation times were used only as empirical descriptors of rotational dynamics. The accuracy of the estimated anisotropy decay parameters depended significantly on the accuracy of the geometry (G) factor of the emission monochromator. Hence, extreme care was taken in the estimation of the G-factor as described previously (Jha et al. 2009). The steady-state fluorescence anisotropy (r_{ss}) was calculated from the parameters obtained from the time domain data, using equation 12.

$$rss = \frac{ro \sum_i \sum_j \alpha_i \beta_j \left(\frac{1}{\tau_i} + \frac{1}{\tau_{rj}}\right)^{-1}}{\sum_i \alpha_i \tau_i} \quad \textbf{(12)}$$

The measurement techniques and the data analysis procedures used were as described earlier (Swaminathan, Krishnamoorthy, and Periasamy 1994; Swaminathan et al. 1996; Jha et al. 2009).

4.3 Results

The structure and stability of the protein were not significantly perturbed by mutation and labelling

To monitor distinct intramolecular distances in dcMN, four single-Trp, single-Cys variants were generated (see Materials and Methods). The tryptophan residues (Trp4 in chain B and Trp58 in chain A) served as FRET donors, and a TNB group covalently attached to a unique cysteine acted as the acceptor. dcMN contains a native Cys residue at residue position 42, and a cysteine residue was introduced at residue position 29, 81, or 96 in place of Cys42 to enable site-specific labelling (see Materials and Methods). The unfolding transitions of these protein variants were then monitored by FRET at various concentrations of GdnHCl. The fluorescence emission maximum of both Trp4 in chain B (for W4C42, W4C96 and W4C29) and Trp58 in chain A (for W58C81), upon excitation at 295 nm, was at ~ 347 nm for the native unlabeled protein (Figure 4.2) indicating that the local environments of Trp4 and Trp58 were similar. Upon unfolding, the wavelength of maximum fluorescence emission increased to 356 nm for all the protein variants, reflecting full exposure of W4 or W58 to solvent. Upon labeling with the TNB acceptor, quenching of the Trp fluorescence intensity was observed at all wavelengths, and the extent of quenching was different for the N and U states. The extent of quenching upon TNB-labeling was different for the different protein variants, indicating that quenching was distance-dependent, and hence, due to FRET. The extent of quenching was also found to be independent of protein concentration, ruling out intermolecular energy transfer (data not shown). The secondary structure, as can be seen in the far-UV CD spectra (Figure 4.3) was not perturbed significantly by mutation and labeling. The minor differences in the spectra of the variants can be attributed to differences in the contributions of Trp and Tyr at residue positions 4 and 58 (Woody 1978; Chakrabartty et al. 1993).

The stabilities of all the unlabeled proteins were not significantly different (< 1 kcal mol^{-1}) from that of wt dcMN (W4C42) (Figure 4.4 and Table 4.2), indicating that the mutations caused minimal perturbation. Labeling with TNB also had little effect on stability for the W58C81, W4C96, and W4C29 variants, but the W4C42 variant was slightly destabilized. Cys42 is completely buried in the N state, and the covalent modification of this residue by

TNB, may have disrupted packing interactions and thereby caused a minor decrease in stability. Similar destabilization upon TNB-labeling of Cys42 had also been observed for MNEI (Bhatia, Krishnamoorthy, and Udgaonkar 2018). Nevertheless, the overall folding and unfolding properties were not altered significantly: the refolding kinetics of the labeled and unlabeled variants were indistinguishable from those of wt dcMN (data not shown).

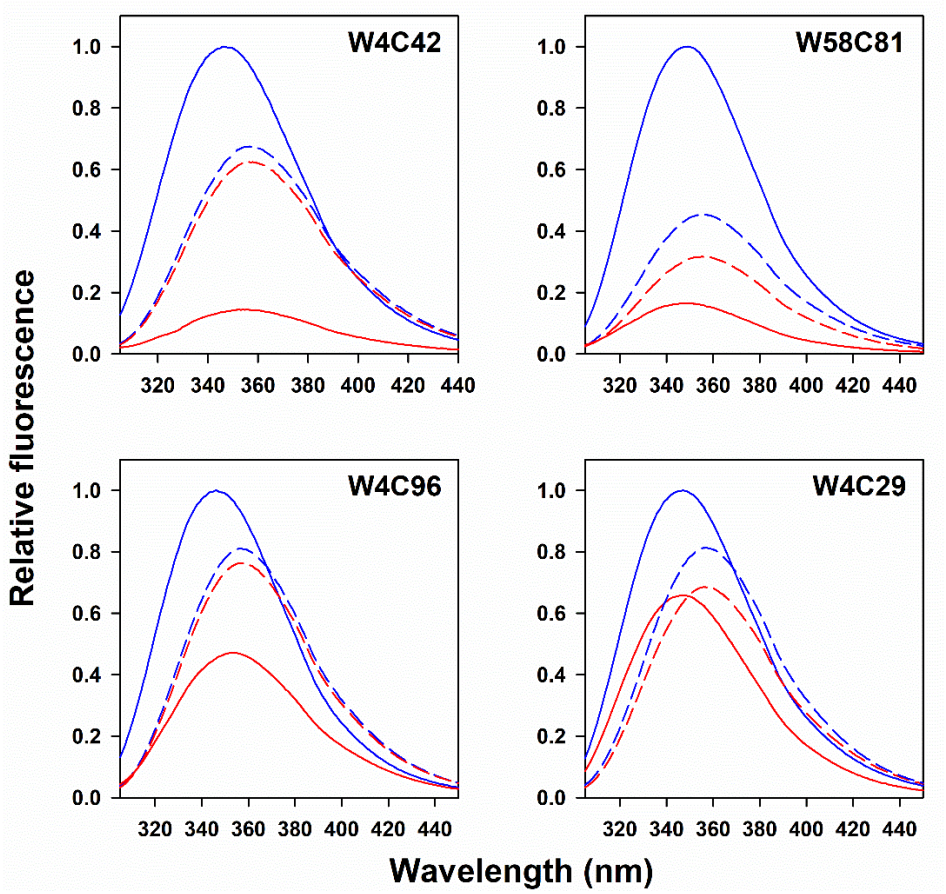


Figure 4.2. Fluorescence emission spectra of the different mutant variants of dcMN. The solid and dashed curves represent the spectra of the native and unfolded states (in 2 M GdnHCl), respectively, for the TNB-labeled (red) and unlabeled (blue) proteins. The spectra were normalized to the fluorescence value of the native unlabeled protein at 350 nm.

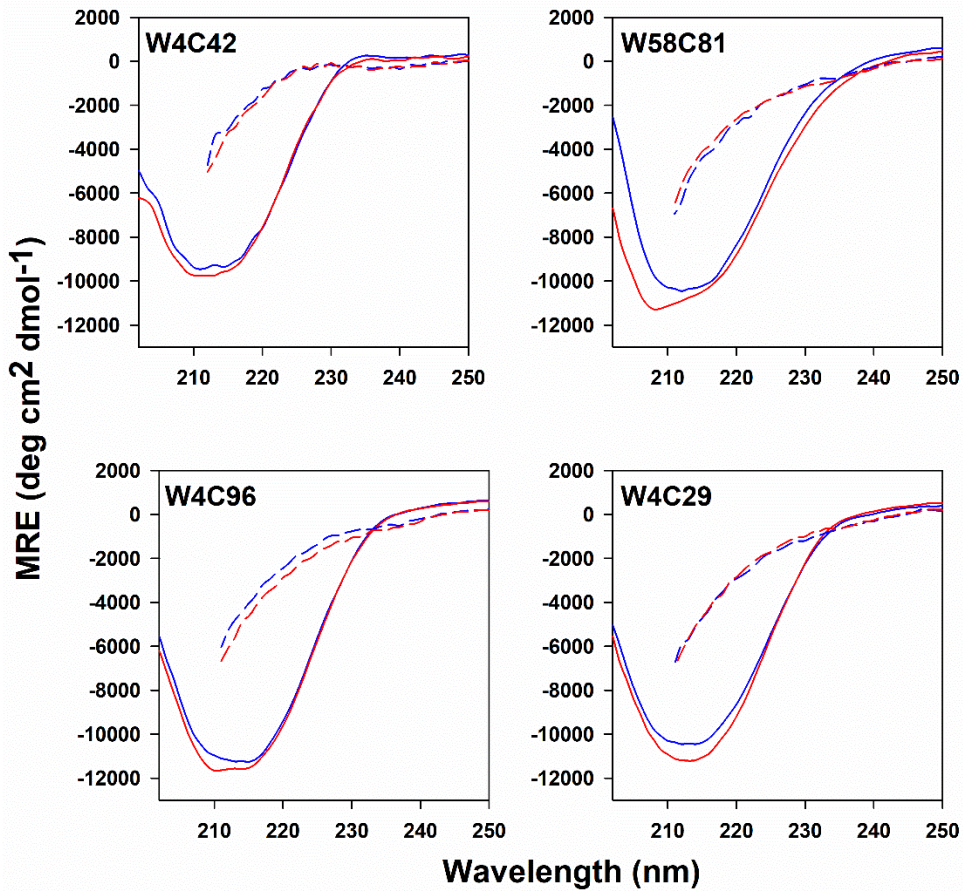


Figure 4.3. Far-UV CD spectra of the different mutant variants of dcMN. Spectra of the native (solid lines) and unfolded (dashed lines) states are shown for the TNB-labeled (red lines) and unlabeled (blue lines) proteins. Spectra of the unfolded proteins were measured in 2 M GdnHCl.

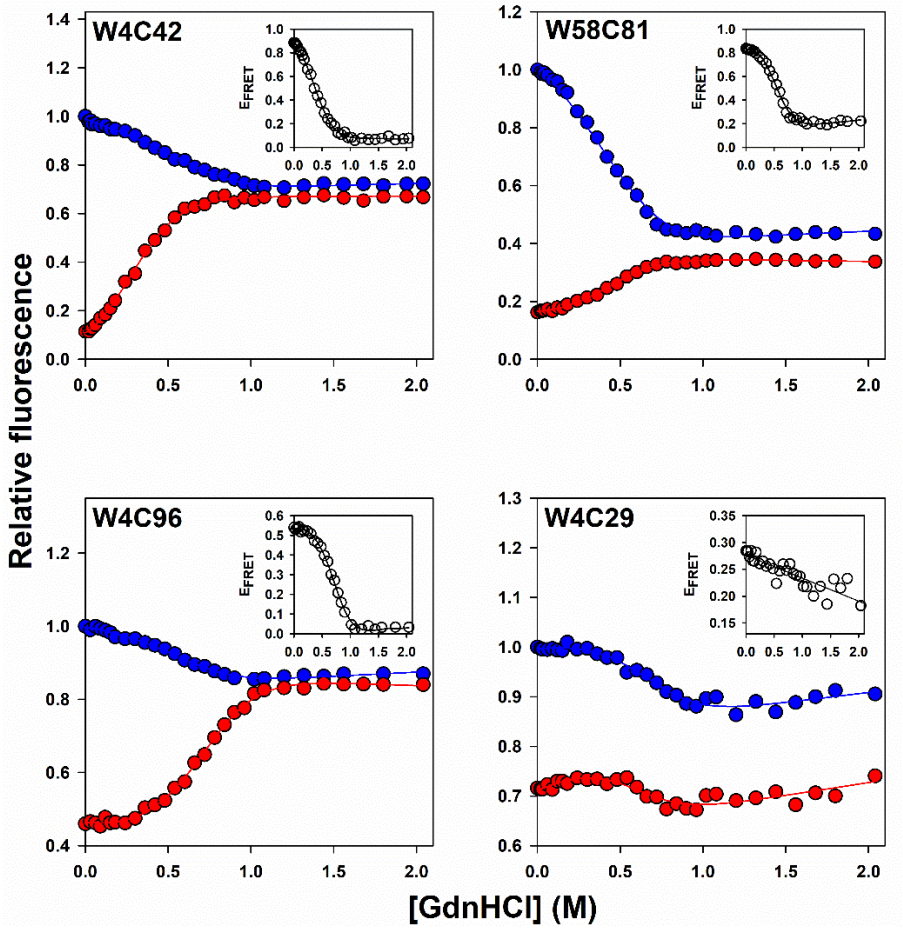


Figure 4.4. Equilibrium unfolding transitions of different variants of dcMN monitored by steady- state FRET measurements. The blue and red circles correspond to the fluorescence intensities of the unlabeled and TNB-labeled proteins, respectively. The solid line passing through each dataset is a non-linear, least-squares fit to a two-state unfolding model for a heterodimeric protein (Aghera, Earanna, and Udgaonkar 2011). The thermodynamic parameters obtained are listed in Table 4.2. The inset in each panel shows the dependence of the FRET efficiency on GdnHCl concentration. The solid line shown in the inset of the panel for W4C29 is drawn to guide the eye.

Table 4.2. Thermodynamic parameters obtained from fluorescence-monitored equilibrium unfolding measurements for the different mutant variants of dcMN at pH 8 and 25°C.

Protein	Free energy of unfolding (ΔG_U), kcal mol^{-1}	Mid-point of unfolding (C_m), M
W4C42	11.0 ± 0.1	0.69 ± 0.04
W4C42-TNB	9.2 ± 0.4	0.27 ± 0.05
W58C81	10.1 ± 0.1	0.46 ± 0.03
W58C81-TNB	9.6 ± 0.1	0.36 ± 0.04
W4C96	10.6 ± 0.2	0.58 ± 0.05
W4C96-TNB	10.8 ± 0.1	0.64 ± 0.02
W4C29	10.8 ± 0.4	0.63 ± 0.09
W4C29-TNB	10.5 ± 0.4	0.55 ± 0.08

Dependence of FRET efficiency on GdnHCl concentration appeared non-sigmoidal for the helix segment

The equilibrium unfolding transitions of W4C42, W58C81, W49C96 and W4C29 monitored by fluorescence, appeared to be two-state in nature, as indicated by the sigmoidal dependence of fluorescence intensity on GdnHCl concentration (Figure 4.4). The dependence of FRET efficiency on GdnHCl concentration (insets, Figure 4.4) also appeared to be describable as a two-state transition, for all variants except W4C29, which showed a broad non-sigmoidal transition.

The dependence of the mean lifetime on GdnHCl concentration, derived from the discrete analysis of time-resolved fluorescence intensity decay curves obtained at different GdnHCl concentrations, compared very well with that monitored by Trp fluorescence (steady-state measurements) (Figure 4.6). Moreover, for all the protein variants, there was good agreement between the FRET efficiencies obtained from fluorescence intensity and lifetime measurements (Figures 4.4 and 4.5). The FRET efficiencies at each GdnHCl concentration were converted to apparent (uncorrected) donor–acceptor distances (insets, Figure 3) using the Förster relation (equation 7, Materials and Methods), assuming dynamic isotropic averaging ($\kappa 2=2/3$). However, based on these measurements, the equilibrium unfolding transitions could not be classified as either strictly cooperative or gradual, as these measurements could not resolve distinct conformational subpopulations populated at different GdnHCl concentrations.

It should be noted that negligible FRET is observed for the segment mapping the distance across the two chains (W4C96) at GdnHCl concentrations above 1.5 M (Figures 4.4 and 4.5). If this FRET amplitude was nevertheless used to infer the distance separating Trp4 and Cys96-TNB, a distance > 40 Å was obtained, which exceeded the reliable distance-sensitive range of the FRET pair (R_0 = 22.7 Å). Hence, the distance, while exceeding 40 Å, could not be determined quantitatively. In addition, far-UV CD spectra for the segments mapping the cores of the two chains (in W4C42 and W58C81), as well as for the inter-chain segment (in W4C96), showed a complete loss of secondary structure at 2 M GdnHCl (Figure 4.3). Together, these observations indicated that at high denaturant concentrations, the two chains have separated from each other.

MEM analysis revealed the coexistence of cooperative and continuous unfolding

Fluorescence intensity decay curves were analyzed using the Maximum Entropy Method (MEM) to obtain fluorescence lifetime distributions (Figure 4.8). Representative fluorescence decay curves together with the corresponding MEM fits and residuals are shown in Figure 4.7 for the W4C96 variant under different denaturant conditions (Figure 4.7). In the case of W4C42-TNB, W58C81-TNB and W4C96-TNB, the fluorescence lifetime distributions obtained at the different GdnHCl concentrations, were bimodal (Figure 4.8). Lifetimes between 0.001 and 0.6 ns, were taken to arise from native-like (N-like) forms, and lifetimes > 0.6 ns, were taken to arise from unfolded-like (U-like) forms of the protein. The cutoff of 0.6 ns corresponds to the boundary between the two modes in the bimodal lifetime distributions, allowing clear separation of N-like (<0.6 ns) and U-like (>0.6 ns) sub-populations (Figure 4.8). This is consistent with the observation that the U state predominantly populates lifetimes > 0.6 ns (U-like) and the N state predominantly populates lifetimes < 0.6 ns (N-like) (Bhatia, Krishnamoorthy, and Udgaonkar 2018; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). This separation reflects the extent of quenching of the Trp fluorescence, where shorter lifetimes correspond to more strongly quenched, N-like sub-populations, and longer lifetimes correspond to less quenched, U-like sub-populations. Such grouping of lifetimes allowed the fractions of molecules in the U-like and N-like sub-populations to be determined at each GdnHCl concentration (see Materials and Methods). For each protein variant, the dependence of the MEM-derived fraction of molecules that were U-like (see Materials and Methods) was found to match the unfolding transition monitored by the measurement of fluorescence intensity or mean fluorescence lifetime (Figures 4.4, 4.6 and 4.9) and yielded similar values for the stability, ΔG_U and the midpoint of the transition, C_m (Figure 4.9 and Table 4.2). This was

important because it validated the grouping into N-like and U-like sub-populations. It should be noted that modest changes in the cutoff (e.g., 0.5 – 0.8 ns) resulted in only minor variations in the calculated U-like fractions or the derived thermodynamic parameters (ΔG_U and C_m), indicating that the analysis is robust to the precise choice of cutoff.

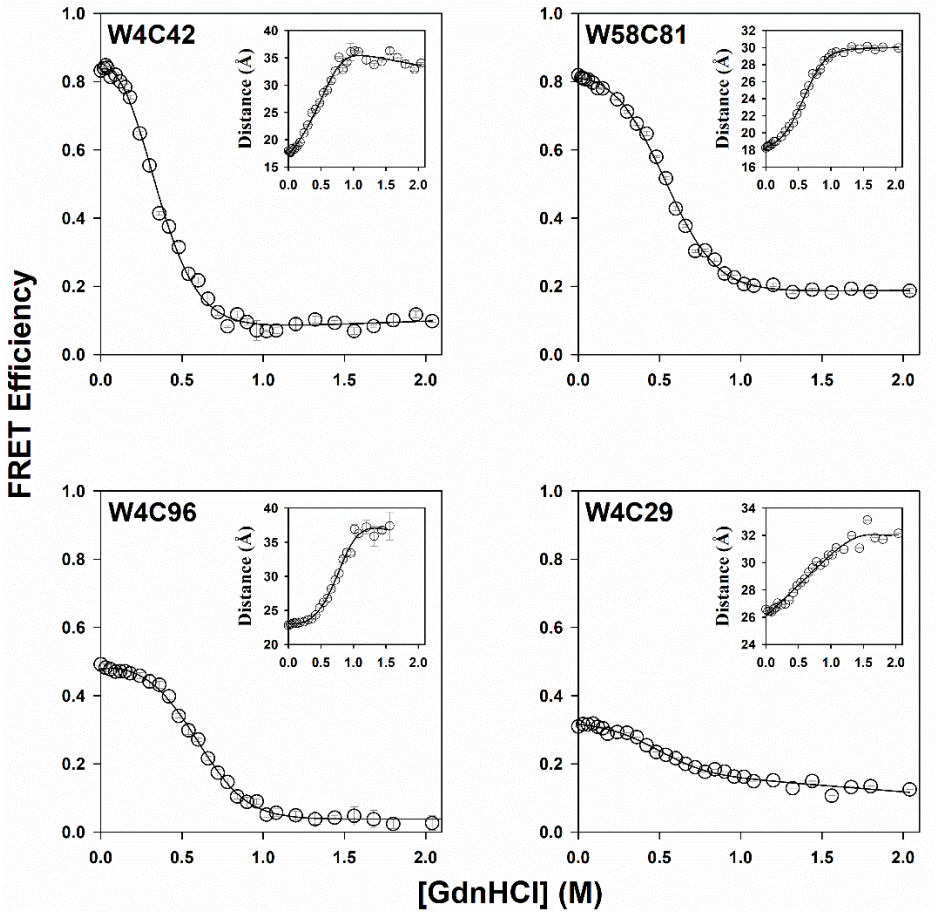


Figure 4.5. Equilibrium unfolding transitions of dcMN monitored by time-resolved FRET measurements. Mean FRET efficiency was calculated using mean lifetimes that were obtained from discrete analysis of the time-resolved data for both the unlabeled and the TNB-labeled protein variants. The inset in each panel shows the dependence of the apparent donor–acceptor distances on GdnHCl concentration. The distances represent ensemble-averaged values calculated from FRET efficiencies using R_o values (Table 4.1; Materials and Methods). The error bars represent the spread in the data obtained from two independent experiments.

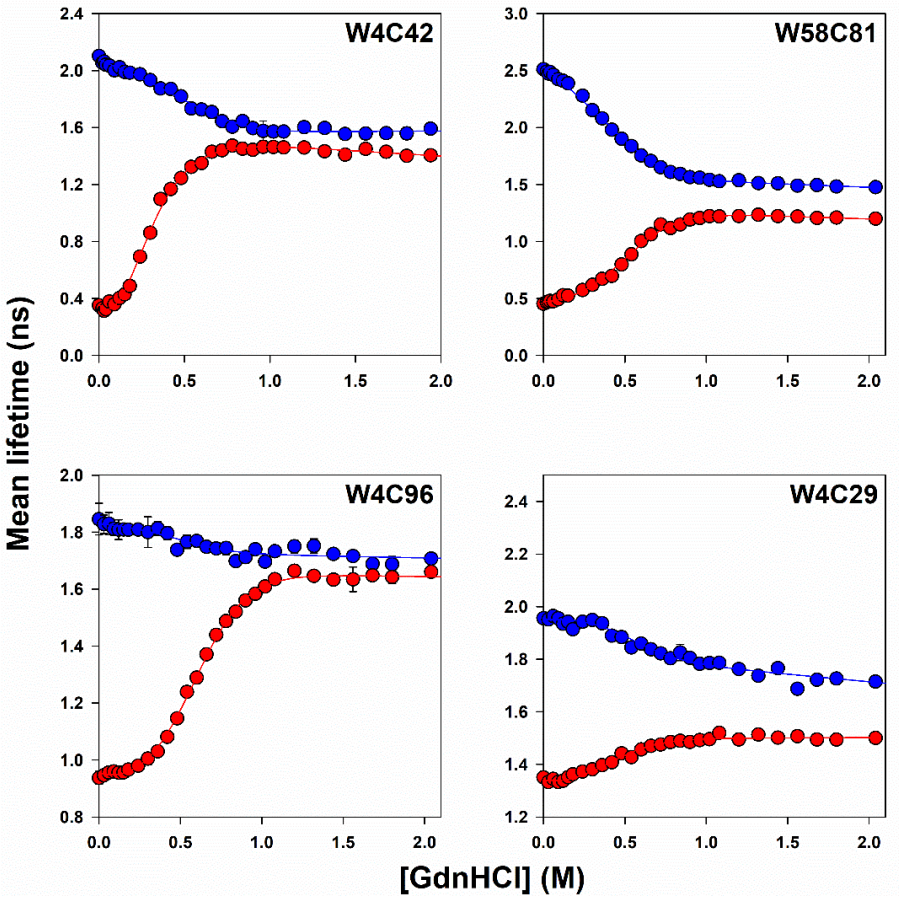


Figure 4.6. Equilibrium unfolding transitions of the different variants of dcMN monitored by time-resolved fluorescence measurements. Mean lifetimes were obtained from discrete analysis of the time-resolved fluorescence data. Blue and red circles correspond to the data obtained for the unlabeled and TNB-labeled proteins, respectively. The solid line passing through each dataset is a non-linear, least-squares fit to a two-state unfolding model for a heterodimeric protein (Aghera, Earanna, and Udgaonkar 2011). The error bars represent the spread in the data obtained from two independent experiments.

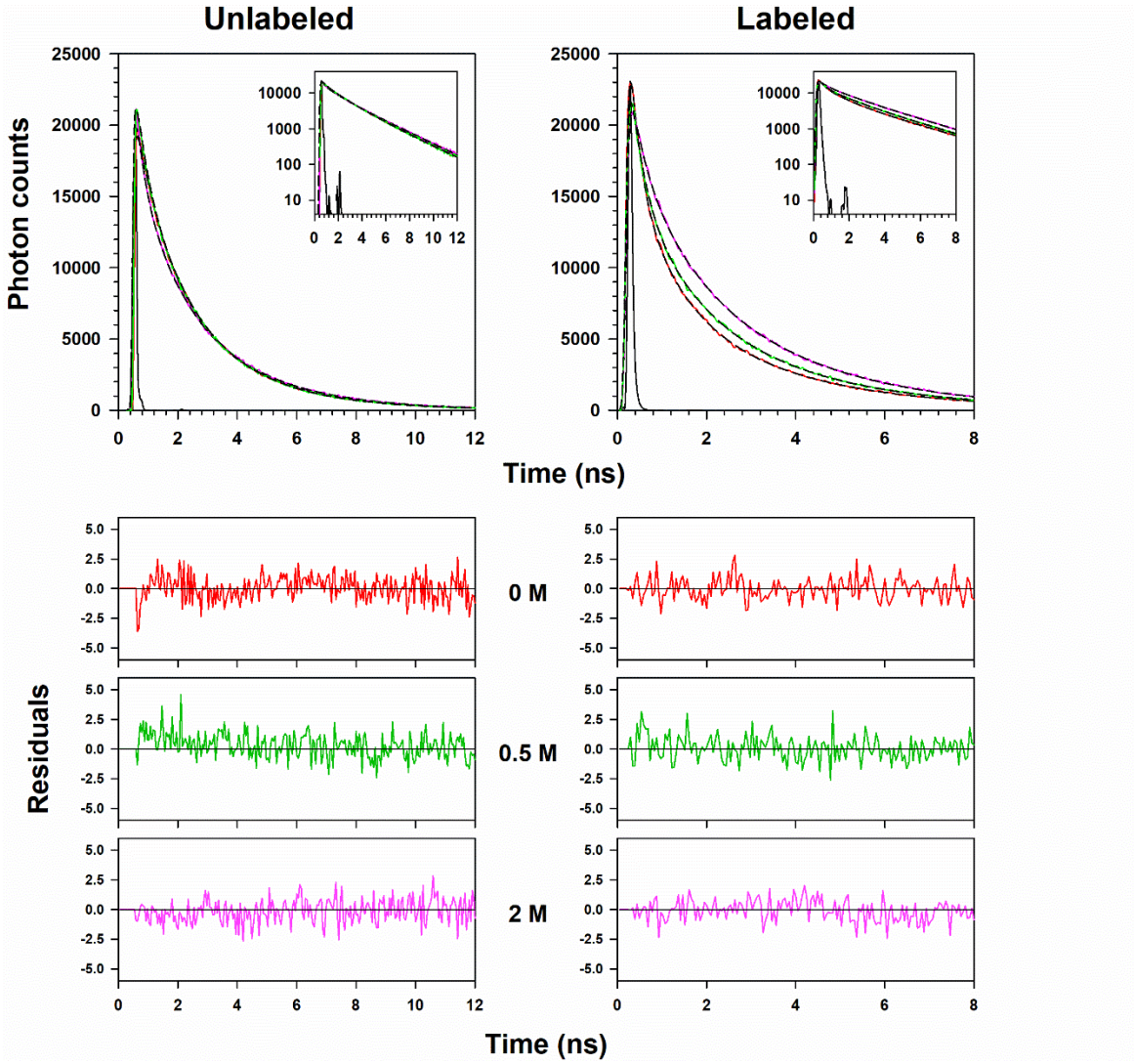


Figure 4.7. Representative fluorescence decay curves of unlabeled and labeled protein for the W4C96 variant at 0 M (red), 0.5 M (green), and 2 M (pink) denaturant concentrations. The corresponding MEM fits are shown as dashed black lines for each concentration. Insets show the decay curves with the y-axis plotted on a logarithmic scale. The corresponding residuals for the fits are shown below.

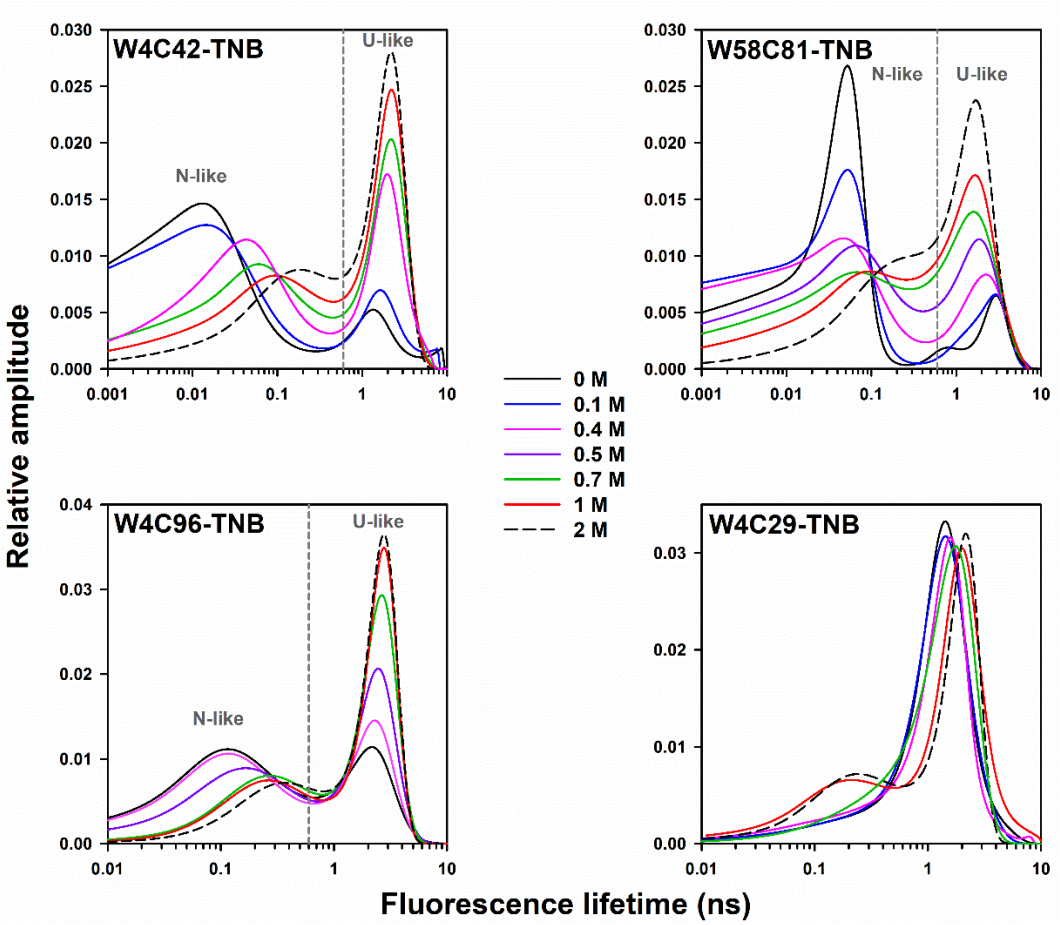


Figure 4.8. MEM-derived fluorescence lifetime distributions of the TNB-labeled mutant variants of dcMN at different GdnHCl concentrations. The differently colored lines correspond to various concentrations of GdnHCl, as described. The x-axis has been plotted on a log scale. The amplitude has been normalized to the sum of amplitudes for each distribution.

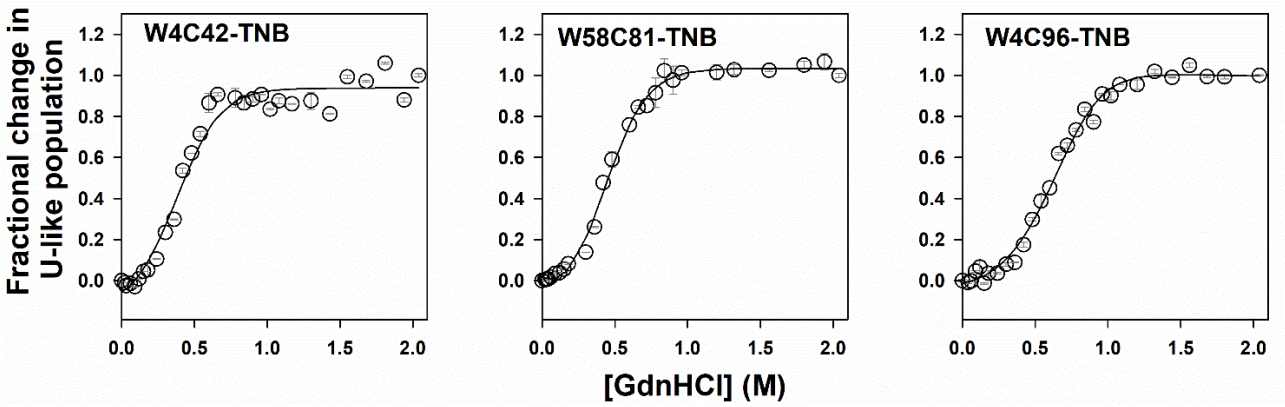


Figure 4.9. Fractional change in the U-like population calculated from the relative sum of amplitudes of the MEM distributions at different concentrations of GdnHCl. The solid line passing through each dataset is a non-linear, least-squares fit to a two-state unfolding model. The fits gave values for ΔG_U of 9.3 ± 0.2, 9.4 ± 0.3 and 10.6 ± 0.1 kcal mol^{-1} for W4C42-TNB, W58C81-TNB and W4C96-TNB, respectively. The mid-points (C_m) of the unfolding transitions for W4C42-TNB, W58C81-TNB and W4C96-TNB are at 0.28 ± 0.04, 0.33 ± 0.06 and 0.61 ± 0.02 M, respectively. The error bars represent the spread in the data obtained from two independent experiments.

The main features of the fluorescence lifetime distributions were as follows: (a) for the N-like forms, the peak position shifted gradually towards longer lifetimes with increasing GdnHCl concentration; (b) for the U-like forms, the peak shifted towards longer lifetimes in the case of W4C42-TNB, and towards shorter lifetimes in the case of W58C81-TNB; (c) the magnitude of these shifts varied among the different TNB-labeled variants (Figures 4.8 and 4.10); and (d) the relative fractions of N-like and U-like sub-populations changed in an apparently two-state manner (Figure 4.9), with an increase in the U-like fraction and a corresponding decrease in the N-like fraction as GdnHCl concentration increased. Thus, both gradual and all-or-none changes were observed within chain B, and chain A, as well as across the two chains.

In contrast, the lifetime distributions for the helix segment (W4C29-TNB) remained unimodal throughout the transition (Figure 4.8). The peak position shifted gradually towards longer lifetimes with increasing GdnHCl concentration, indicating a gradual change in the intra-helix distance during the unfolding transition (Figures 4.8 and 4.10). It should be noted that for all the unlabeled proteins, the fluorescence lifetime distributions remained unimodal throughout the unfolding transition (Figure 4.11) with the peaks not showing any significant shift upon unfolding (insets, Figure 4.11).

The dependence of the MEM peak lifetime on GdnHCl concentration was clearly sigmoidal for the N-like subensembles seen for W4C42-TNB, W58C81-TNB and W4C96-TNB (Figure 4.10). In contrast, the MEM peak lifetime for the U-like subensembles seen for these variants, as well as for W4C29-TNB, showed non-sigmoidal dependences on GdnHCl concentration (Figure 4.10), pointing to non-cooperativity. In order to interpret these changes in the MEM peak lifetimes of the TNB-labeled variants, the distances within chain B (in W4C42) and chain A (in W58C81), across both chains (in W4C96), and spanning the helix in chain B (in W4C29) were calculated using equation 7 (Materials and Methods) after correcting for the small peak shifts of the unlabeled proteins (Figures 4.10, 4.11 and 4.12).

While the MEM peak life time of the U-like subpopulation observed using the FRET pair mapping separation across the two chains (Trp4 -C96TNB) shifted slightly toward longer lifetimes, a comparable shift was also observed for the fluorescence lifetime distributions of the corresponding unlabeled protein (Figure 4.11). Consequently, the U-like sub-population exhibited negligible FRET with no discernible change over the entire range of GdnHCl concentrations. The intra-molecular distances inferred from the very small FRET values exceeded ~40 Å, at all GdnHCl concentrations, indicating that the chains had physically separated even in zero denaturant. It should be noted that the separation of Trp4 from Cys96-

TNB in MNEI at high GdnHCl concentration is less (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

The negligible FRET observed in the U-like sub-population for dcMN was consistent with the ensemble-averaged FRET efficiency being near-zero at GdnHCl concentrations > 1.5 M (Figures 4.4 and 4.5). Only at low GdnHCl concentrations, at which the U-like sub-population was negligibly populated (Figure 4.9), did the ensemble-averaged FRET efficiencies have significant values (Figures 4.4 and 4.5) because they were dominated by the contribution of the N-like sub-population. Hence, the ensemble-averaged FRET analysis was consistent with the MEM analysis which suggested that the two chains have separated even in the absence of GdnHCl in the U-like subpopulation.

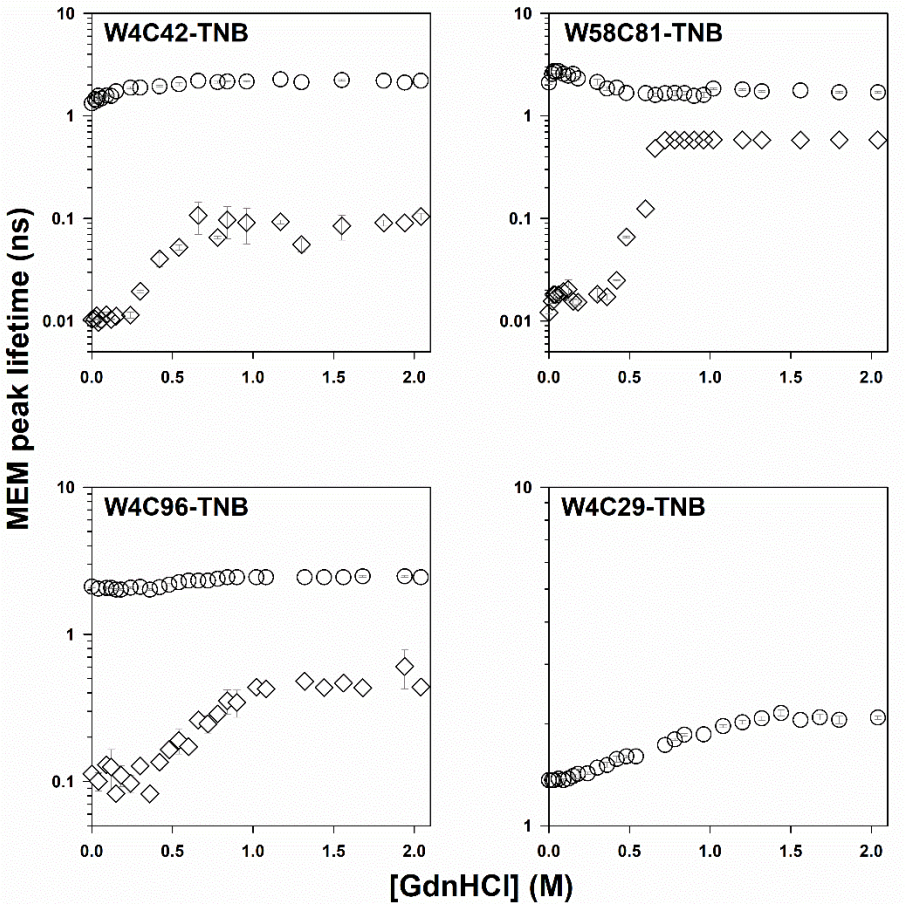


Figure 4.10. The shift in the MEM peak lifetime for the TNB-labeled variants of dcMN as a function of GdnHCl concentration. Diamonds and circles represent the lifetimes corresponding to the peaks in the MEM distributions arising from the N-like and U-like ensembles, respectively. The circles in the panel for W4C29 correspond to the peak lifetimes of the observed unimodal fluorescence lifetime distributions. The error bars represent the spread in the data obtained from two independent experiments.

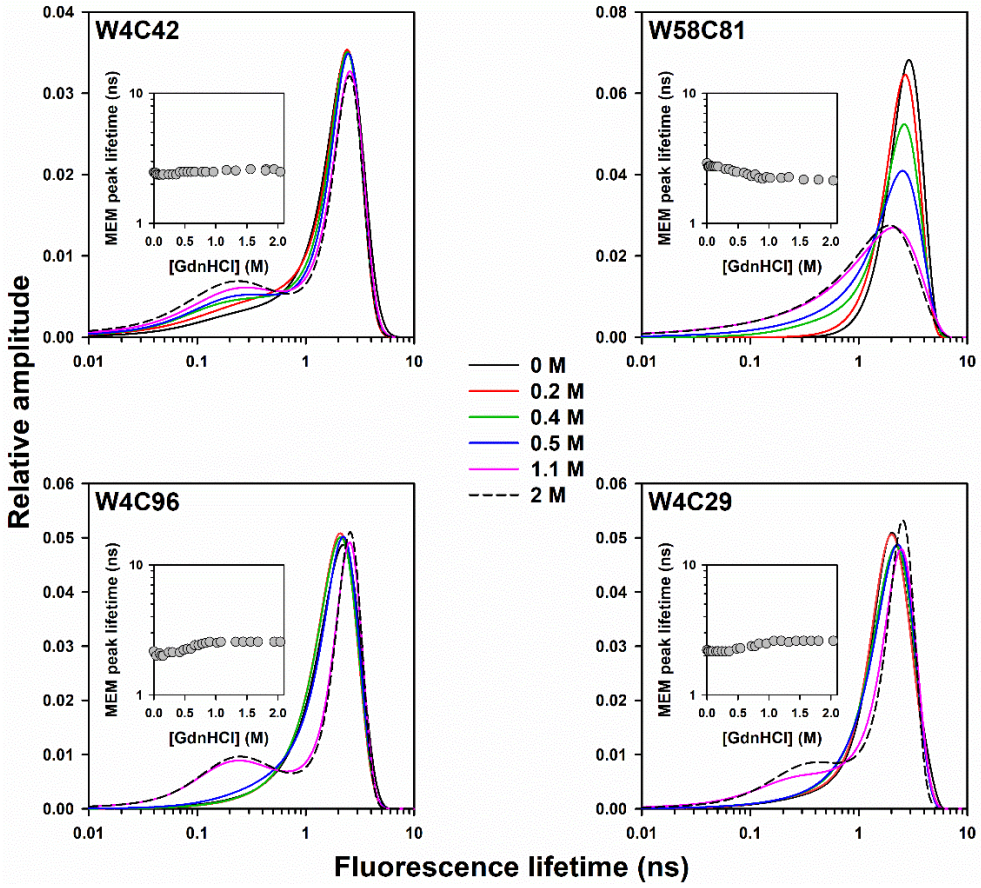


Figure 4.11. MEM-derived fluorescence lifetime distributions of the different mutant variants of dcMN at varying concentrations of GdnHCl. The colors of the curves represent the GdnHCl concentrations, as indicated. The inset in each panel shows the shift in the peak lifetime as a function of GdnHCl concentration. The error bars represent the spread in the data obtained from two independent experiments.

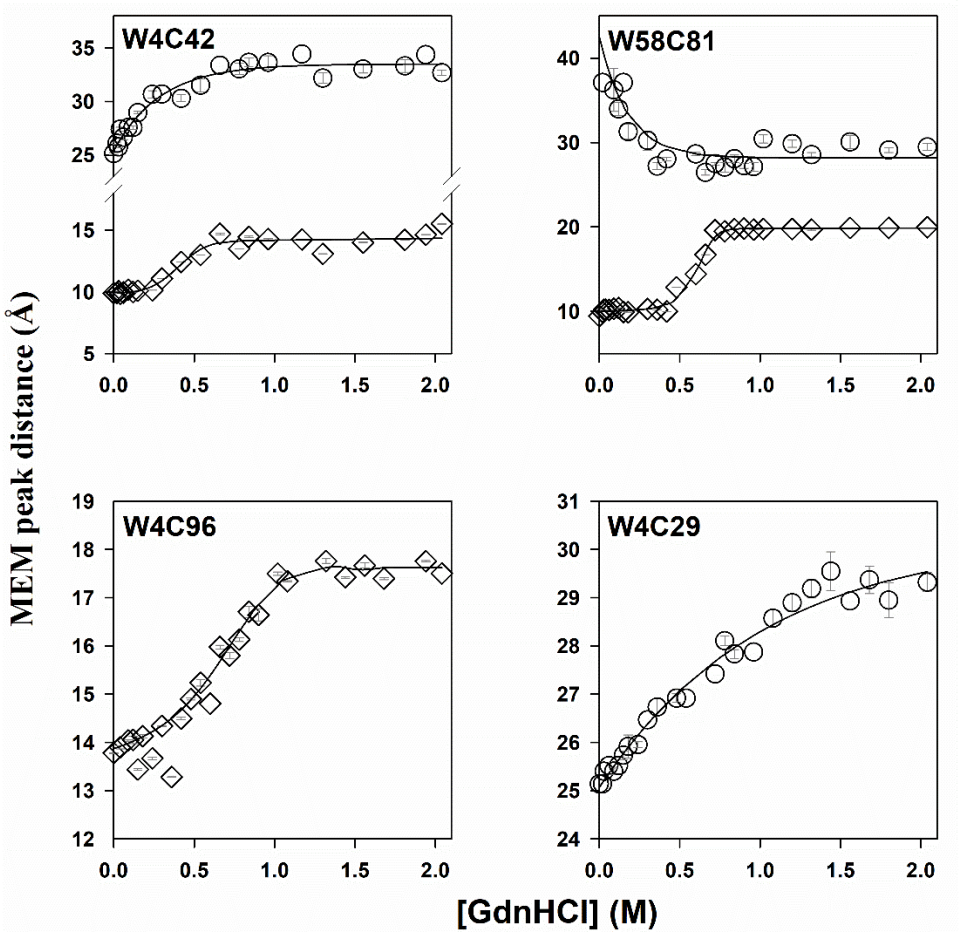


Figure 4.12. Change in the dimensions of the N-like and U-like subensembles with increasing GdnHCl concentration. The MEM peak lifetime for the TNB-labeled proteins and for the corresponding unlabeled proteins were used to calculate the MEM peak FRET efficiencies (equation 6, see Materials and Methods), which were converted to distances (equation 7, see Materials and Methods). Diamonds and circles represent the distances corresponding to the peaks of the MEM distributions of the N-like and U-like ensembles, respectively. The error bars represent the spread in the data obtained from two independent experiments.

To assess the extent of deviation of the four protein variants from two-state unfolding behaviour, the MEM distribution at each GdnHCl concentration was fitted to the weighted sum of the N- and U- state fluorescence lifetime distributions (Figure 4.13) using equation 5 (see Materials and Methods). In the case of W4C42-TNB, in which a distance change within chain B was monitored, and W58C81-TNB, in which a distance change within chain A was monitored, it was found that fluorescence lifetime distributions in the transition region did not fit to the weighted sum of the N state and U state fluorescence lifetime distributions (Figures 4.13 and 4.14). Deviations were also observed in the case of W4C96-TNB, in which a distance across the two chains was monitored and W4C29-TNB, in which a distance spanning the helix of chain B was monitored, but they were of much smaller magnitude (Figure 4.14). It should also be noted that none of the protein variants showed an iso-lifetime point, that is, a single fluorescence lifetime at which all distributions intersected. The absence of such a point, expected for a two-state transition, further confirmed that the unfolding of dcMN is not strictly two-state, but involves heterogeneous ensembles populated throughout the transition.

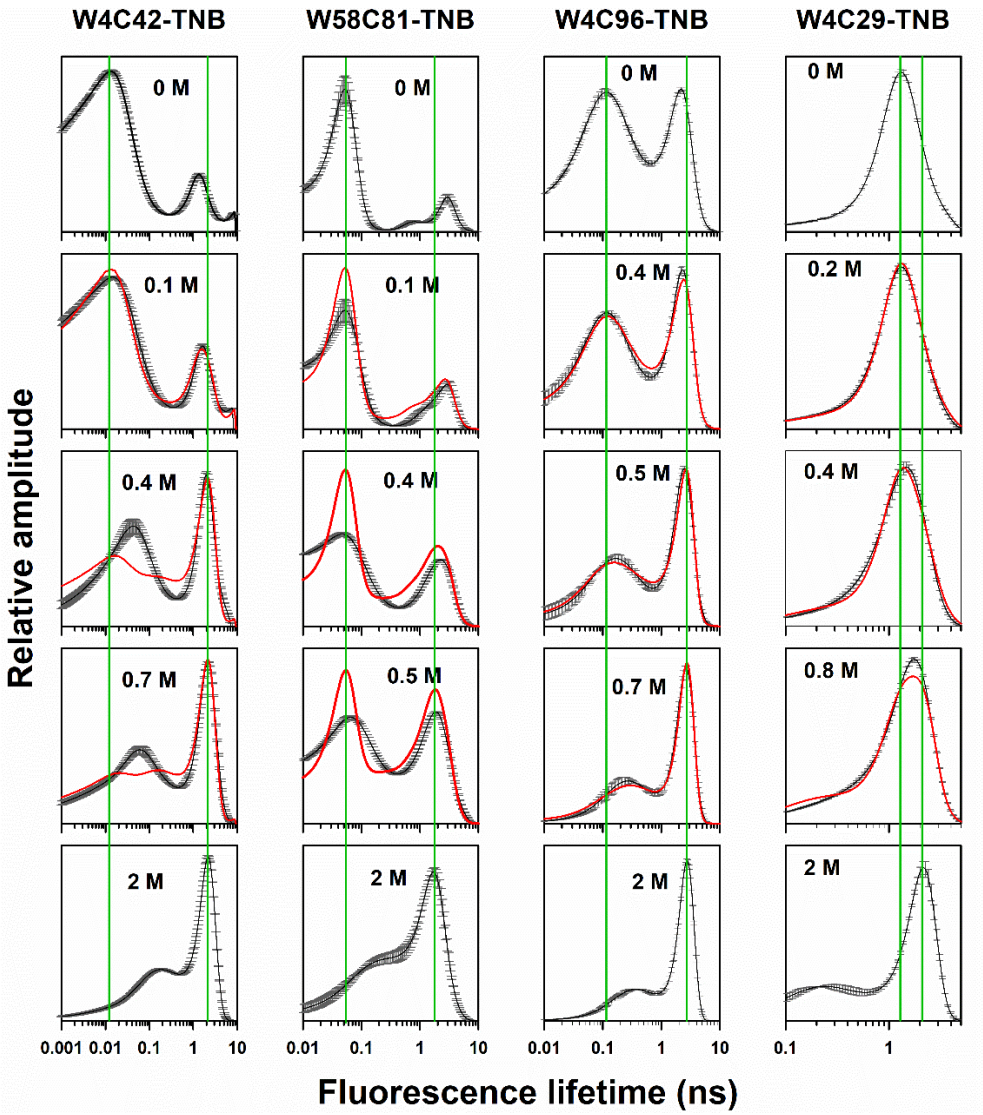


Figure 4.13. Fits of the fluorescence lifetime distributions obtained at different GdnHCl concentrations to a two-state N ↔ U model for equilibrium unfolding of the dcMN variants. The MEM-derived fluorescence lifetime distributions obtained at different GdnHCl concentrations (indicated in each panel) were fit to the weighted sum of the native state N(τ) and unfolded state U(τ) fluorescence lifetime distributions (equation 5, Materials and Methods). The black lines correspond to the experimentally determined distributions, and the red lines are fits to the two-state model. The error bars are the standard deviations of the MEM-derived fluorescence lifetimes obtained from multiple data acquisitions on the same sample. The top-most and bottom-most panels in each column show the fluorescence lifetime distributions used as the native N(t) and unfolded U(t) protein basis distributions, respectively. The vertical green lines indicate the peak lifetimes of the N-state and U-state fluorescence lifetime distributions. The x-axis has been plotted on a log scale, and the y-axis units are arbitrary.

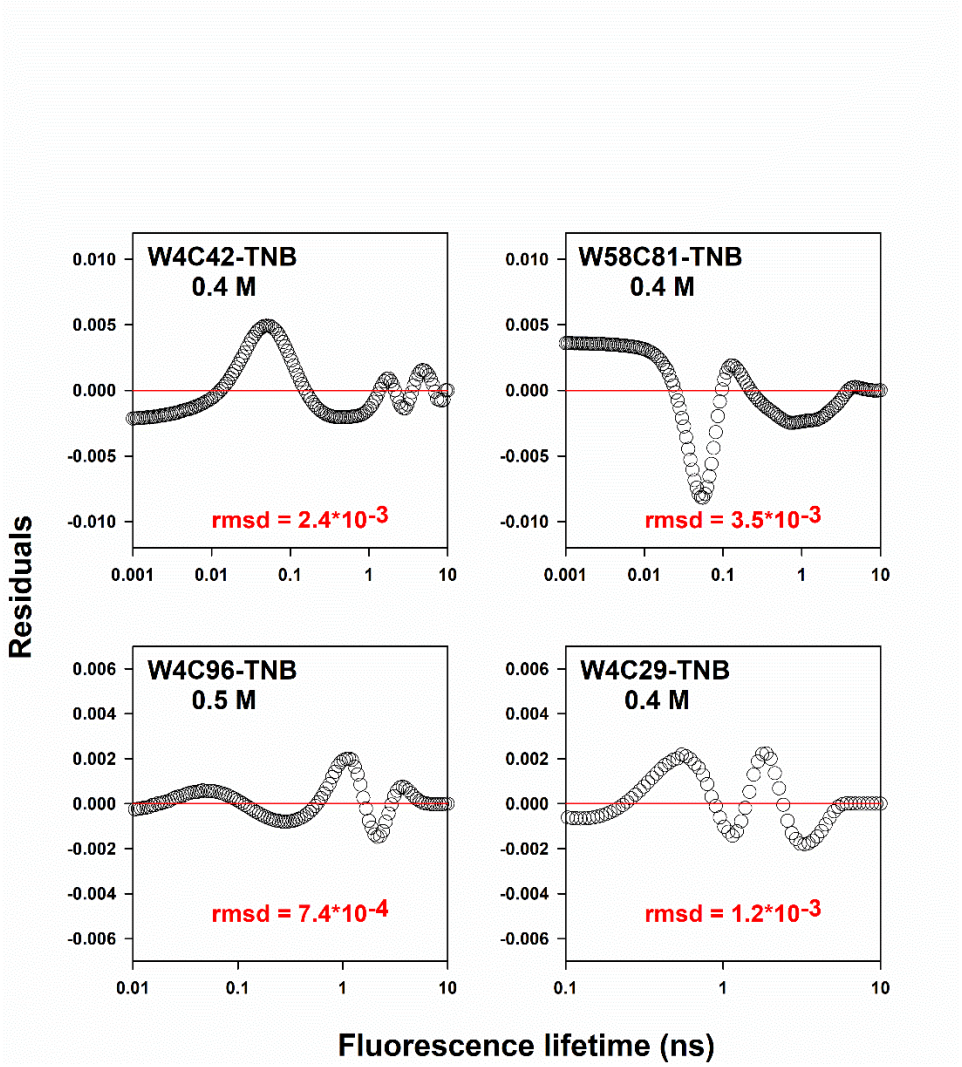


Figure 4.14. Residuals from the fits of the MEM-derived fluorescence lifetime distributions to the weighted sum of the mean of the MEM distributions determined for the N and U states (Figure 4.13). The residuals obtained for GdnHCl concentrations close to the mid-point of the unfolding transition are shown. Values in red correspond to the root mean square deviation (rmsd), which was obtained as the square root of the mean of the squares of residuals across all the lifetime values (x-axis).

Motional dynamics of Trp4 (in W4C42) and Trp58 (in W58C81) revealed heterogenous unfolding

The unfolding transition monitored by steady-state anisotropy (r_{ss}) was found to be identical to that monitored by fluorescence intensity (Figure 4.15b). Values of r_{ss} were also obtained from the time-resolved anisotropy decays at different GdnHCl concentrations (Figure 4.16) and were found to coincide with the steady-state fluorescence anisotropy measured directly (Figure 4.15b), supporting the accuracy of the time-resolved measurements.

Since r_{ss} is a function of both the rotational dynamics of the fluorophore and its excited-state lifetime (equation 12, Materials and Methods), a change in it (Figure 4.15b) may not reflect a change in rotational dynamics alone. Hence, the different rotational correlation times obtained from time-resolved fluorescence anisotropy decays of Trp4 in W4C42 and of Trp58 in W58C81 (Figure 4.16) were analyzed to obtain direct information on the rotational dynamics, which in turn relate to the structural compactness around Trp4 and Trp58 in chains B and A, respectively (Figure 4.17). Trp58 in chain A also forms a part of the core β2- β3 dimer interface (Figure 4.15a), and thus is expected to offer insights into the structural and dynamical aspects not only within chain A but at the β2- β3 dimeric interface as well.

Fluorescence anisotropy can decay relatively slowly by global rotational diffusion of the protein and/or rotational diffusion of a protein segment, defined by a slow rotational correlation time, ϕ_{slow} with corresponding relative amplitude β_{slow} as well as by independent fast rotational motion of the Trp side-chain, defined by a fast rotational correlation time, ϕ_{fast} with corresponding relative amplitude β_{fast}. The site-specific time-resolved fluorescence anisotropy decay measurements revealed distinct local dynamic environments for the two Trp residues in the native protein. In 0 M GdnHCl, the anisotropy decay curve of Trp4 (in W4C42) required a two-exponential fit, yielding ϕ_{slow} = ~ 6 ns, and ϕ_{fast} = ~ 0.2 ns, with β_{fast} = 0.25 (Figures 4.16 and 4.17). In contrast, the anisotropy decay curve of Trp58 (in W58C81) fit well to a single-exponential equation, yielding ϕ_{slow} (Figure 4.17). For both Trp residues, the value of ϕ_{slow} (6 ±1 ns) was that expected for an ~ 11 kDa protein in aqueous media at 25°C (Cantor and Schimmel 1980; Lakowicz 2006). The presence of a measurable fast component for Trp4 indicated that it undergoes local side-chain motion in the N state, which Trp58, located at the dimer interface (Figure 4.15a), does not. This indicated that Trp58 is in a conformationally restrictive environment with its side-chain immobilized in the native structure. It should be noted that the values of initial fluorescence anisotropy values (r_0) remained constant across the entire GdnHCl range for both Trp residues (Figure 4.18). This indicated that the orientations of

the absorption and emission transition dipole moments did not change appreciably during unfolding, and, consequently, variations in the fluorescence anisotropy decays arose from changes in rotational dynamics rather than changes in photo-selection or fluorophore geometry. It should be noted that despite restricted local mobility, particularly for Trp58, the observed initial anisotropy (r_0) is lower than the theoretical maximum of 0.4, consistent with the presence of multiple electronic transition dipoles in tryptophan. This supports the use of $\kappa^2=2/3$ for distance calculations using the Förster relation (see Materials and Methods).

The value of ϕ_{slow} was observed to decrease in a sigmoidal manner with increasing GdnHCl concentration for both Trp4 (in W4C42) and Trp58 (in W58C81), as rotational motion of the sequence segments harboring the Trp residues replaces global molecular tumbling as the source of slow anisotropy decay. In the U state, fluorescence anisotropy of both Trp residues decayed by segmental motion with a rotational correlation time of 1 ns (Figure 4.17), as well as the by fast rotational motions of the side-chains. The overlap in the cooperative sigmoidal transitions mapped by the values of ϕ_{slow} for Trp4 and Trp58, suggested simultaneous global loosening of the two chains.

The value of ϕ_{fast} for Trp4 remained invariant at about 0.2 ns across the entire range of GdnHCl concentration. For Trp58, ϕ_{fast} could be resolved only at GdnHCl concentrations above 0.2 M, and it too remained invariant at 0.2 ns. In the case of both Trp residues, β_{fast} was found to increase progressively in an apparently exponential manner with increasing denaturant concentration, consistent with a gradual increase in local motional freedom, before reaching a limiting value of about 0.65. This continuous redistribution between global tumbling and local flexibility for both Trp4 and Trp58 suggests a heterogeneous and distributed structural loosening, rather than an all-or-none transition. Together, these observations demonstrated that the unfolding process cannot be described by a simple two-state model but instead involves intermediate motional modes and progressive structural disruption.

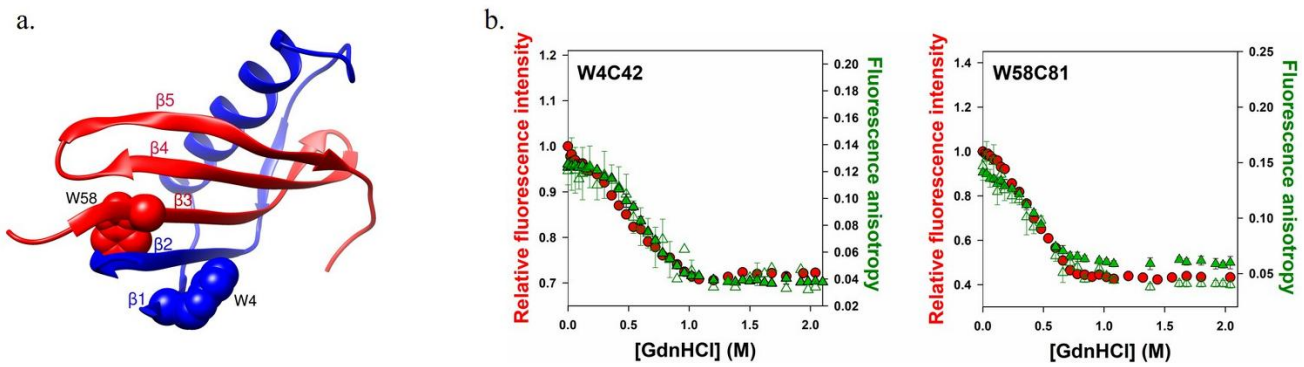


Figure 4.15. a) Structure of dcMN showing the locations of the two Trp residues. The side chains of Trp4 in chain B and Trp58 in chain A are shown in blue and red, respectively. b) Equilibrium unfolding transitions of the two variants were monitored by Trp fluorescence (red), steady-state anisotropy measured directly (filled green), and that derived from fits of time-resolved anisotropy decay traces (empty green). The fluorescence properties of Trp4 and Trp58 were monitored in W4C42 and W58C81, respectively.

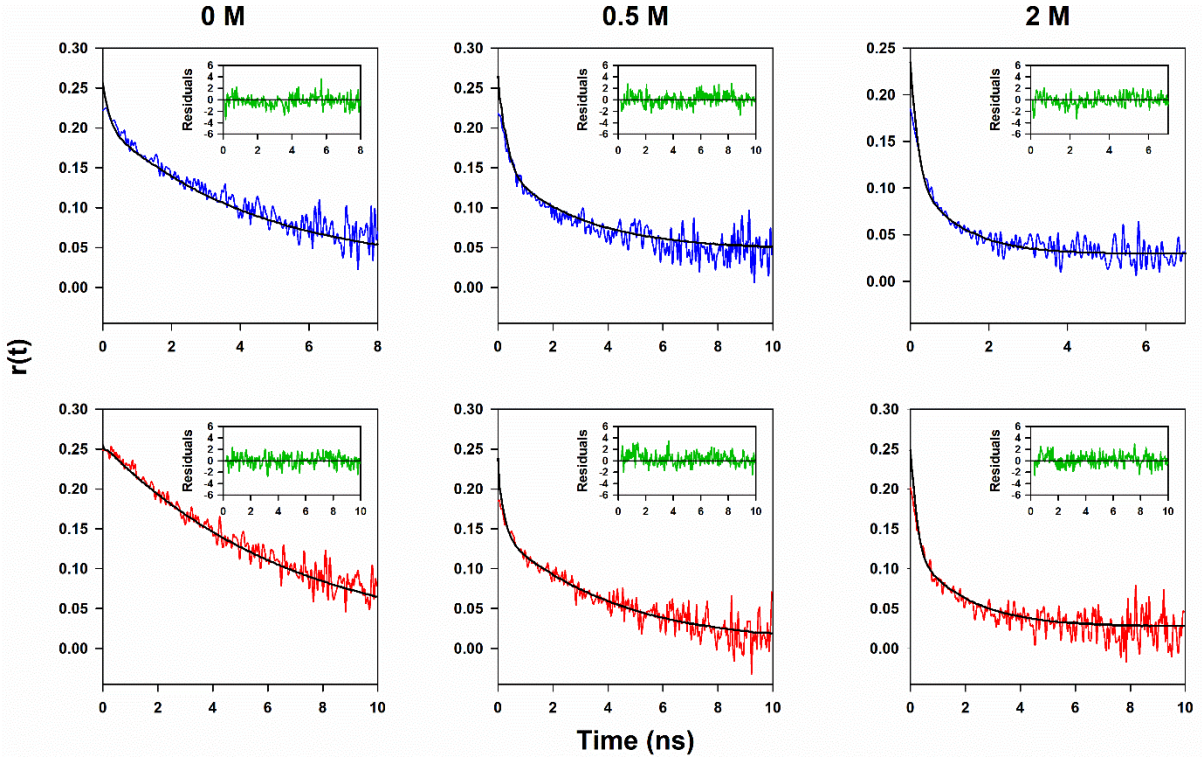


Figure 4.16. Time-resolved fluorescence anisotropy decay curves of Trp4 (blue) and Trp58 (red) monitored in W4C42 and W58C81, respectively, at the indicated concentrations of GdnHCl. The rotational correlation times and their corresponding relative amplitudes obtained from the fits (black solid lines) are shown in Figure 4.17. The decay curve of Trp58 in 0 M GdnHCl is the only one that fits well to a single-exponential equation. All other decay curves fit well only toa two-equation equation. Insets display the residuals of the fits.

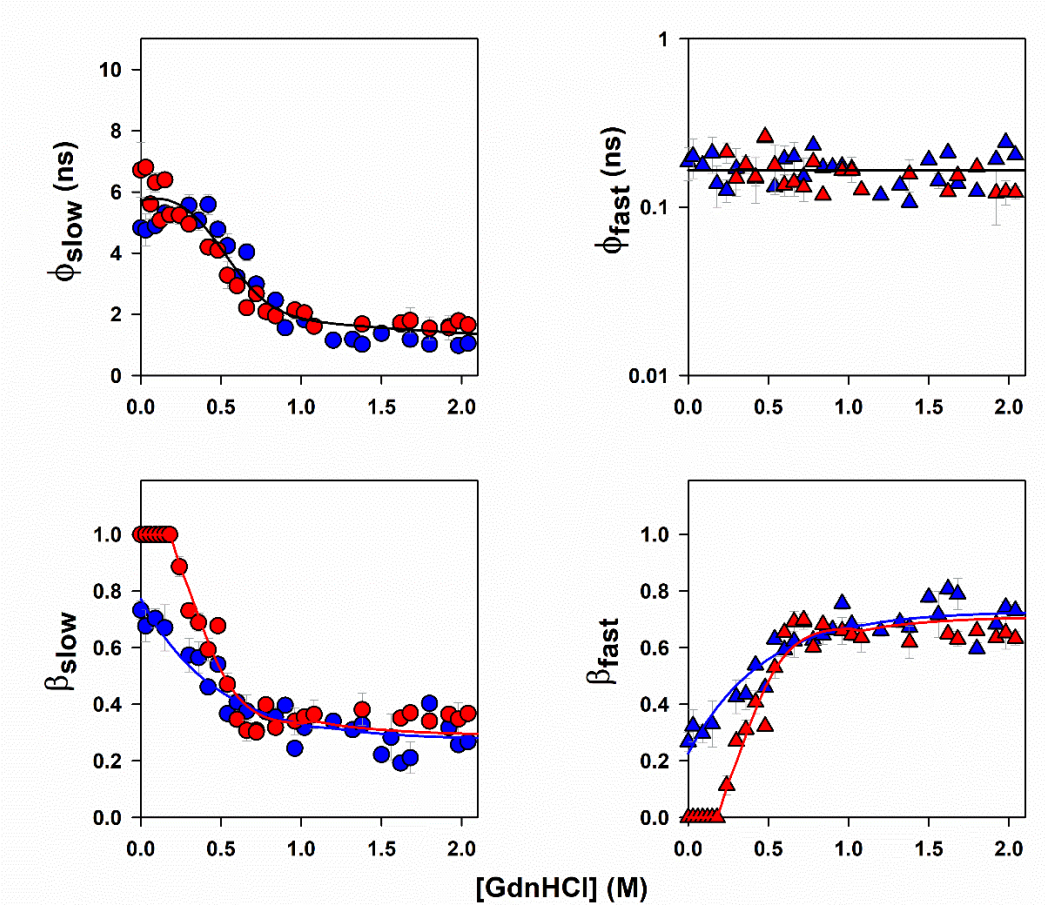


Figure 4.17. Equilibrium unfolding of W4C42 and W58C81 monitored by time-resolved anisotropy measurement. Data for Trp4 (blue) were obtained by measurements on W4C42, and data for Trp58 (red) were obtained from measurements on W58C81. Dependence of the slow (circles) and fast (triangles) rotational correlation times, along with their corresponding relative amplitudes, on GdnHCl concentration. The error bars represent the spread in the data obtained from two independent experiments.

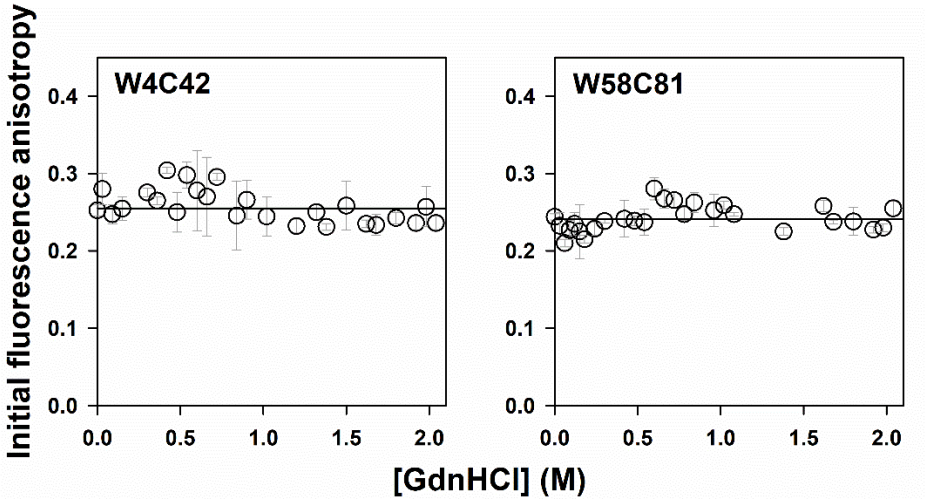


Figure 4.18. Equilibrium unfolding of W4C42 and W58C81 monitored by time-resolved fluorescence anisotropy. The dependences of the initial fluorescence anisotropy of the two variants on GdnHCl concentration are shown. The initial anisotropy of Trp4 was measured in W4C42 and that of Trp58 in W58C81. The error bars represent the spread in the data obtained from two independent experiments.

4.4 Discussion

Equilibrium unfolding of dcMN is not cooperative

Fluorescence intensity measurements of the equilibrium unfolding of all four unlabeled protein variants and their labeled counterparts, as well as steady-state FRET measurements mapping expansion of chain B (in W4C42), chain A (in W58C81), and the inter-chain distance (in W4C96), suggests that the GdnHCl-induced unfolding of dcMN is cooperative (Figure 4.4), except unfolding of the segment mapping the helix (W4C29) appears to be gradual. However, the underlying heterogeneity and lack of cooperativity in the unfolding transition become more evident in tr-FRET measurements upon a population level MEM analysis. Considerable heterogeneity is revealed. Unfolding is seen to proceed through sub-populations of N-like forms expanding continuously while transiting into U-like forms which also expand with an increase in GdnHCl concentration (Figure 4.12). The observation that the MEM-derived distributions of fluorescence lifetimes for all the protein variants cannot be described adequately as the weighted sum of the N and U state distributions (Figures 4.13 and 4.14), provides strong evidence that the cores of both chains, the inter-chain interface and the helix, all unfold in a non-cooperative manner.

Different regions of the protein display significant differences in cooperativity

The observation that the segment spanning the helix exists in a single conformational ensemble which expands continuously with an increase in GdnHCl concentration (Figure 4.12) suggests that the unfolding of the helix occurs in a completely gradual manner. In contrast, the observation of co-existing sub-populations of N-like and U-like forms, which differ in their mean intra-chain and inter-chain dimensions, for segments mapping the cores of chain A (in W58C81) and chain B (in W4C42) and across the two chains (in W4C96), suggests that they are separated by a significant free-energy barrier. A previous HX-MS study had shown that at equilibrium, multiple partially unfolded intermediate ensembles, separated by significant energy barriers, coexist with the N and U states. Moreover, the structures differed in the absence and in the presence of GdnHCl (Bhattacharjee and Udgaonkar 2021). For example, the most U-like intermediate ensemble had only β2 structured in zero denaturant, but both β2 and β3 structured in the presence of low concentrations of GdnHCl, but the HX-MS studies could not determine whether this difference arose from local or non-local interactions. In the case of adenylate kinase too, the distribution of cooperatively exchanging intermediates could be modulated by denaturant (Pirchi et al. 2011), and in the case of barstar (Lakshmikanth et al. 2001) and the SH3 domain of PI3K (Kishore, Krishnamoorthy, and Udgaonkar 2013), β-sheet

regions were found to unfold relatively more cooperatively than other structural elements under equilibrium conditions.

Nevertheless, in the case of dcMN, while different segments mapping inter-β strand distances transit from a N-like to a U-like sub-population in an apparently cooperative manner (Figure 4.9), distances within each sub-population are seen to change in an apparently continuous manner with a change in GdnHCl concentration. In the previous HX-MS study too, it was seen that each intermediate ensemble consisted of molecules sampling many different conformations that differed as little as in having or not having structure at only one amide site (Bhattacharjee and Udgaonkar 2021).

Swelling of the partially contracted N-like sub-populations is cooperative

The observation that the N-like sub-populations undergo swelling in an apparently cooperative manner with increasing GdnHCl concentration, both for the intra-chain segments in chains A (in W58C81) and B (in W4C42) and for the segment spanning both chains (inW4C96) (Figure 4.12), suggests that molecules contracted at these segments respond in a coordinated manner to the disruption of stabilizing interactions. Importantly, the denaturant dependence of swelling of the N-like ensemble differs across sites: the expansion transitions for W4C42 and W58C81 saturate at ~ 0.7 M GdnHCl, whereas the corresponding transition for W4C96 is broader and saturates only at ~1 M GdnHCl (Figure 4.12). This difference suggests that W4C42 and W58C81 report predominantly on the sub-global stabilities of the N-like sub-populations monitored in chains B and A, respectively, which are determined primarily by less stable intra-chain contacts that are disrupted at lower denaturant concentrations. On the other hand, W4C96, in which the FRET-monitored distance spans both chains, reports on the global stability of the N-like sub-population of dcMN, which is determined not only by intra-chain contacts but also by stabilizing interactions that couple the two chains together. It appears that the N-like forms preserve an inter-dependent network of non-local contacts, including β-sheet hydrogen-bonding networks that link together distant sequence positions (Abkevich, Gutin, and Shakhnovich 1995) as well as stabilizing non-covalent interactions at the inter-chain interface. Similar coupling between non-local contacts and cooperative expansion has been inferred in other β-rich proteins, where high contact order and β-sheet topology impose free-energy barriers that must be crossed collectively (Plaxco, Simons, and Baker 1998; Grantcharova et al. 1998; Roe, Hornak, and Simmerling 2005; Fenwick et al. 2014). Native-state HX-NMR studies have shown that groups of residues can lose protection in concert, defining cooperative foldon units (Englander and Mayne 2014), and a kinetic HX-MS on the

SH3 domain of PI3 kinase reveal that hydrogen bonds between adjacent strands rupture collectively rather than strand by strand (Aghera and Udgaonkar 2017).

The U-like sub-populations undergo non-cooperative unfolding

The observation that in the case of the inter-chain distance segment (W4C96), negligible FRET is observed for the U-like sub-population across the entire range of GdnHCl concentrations, suggests that the two chains have they separated even at 0 M GdnHCl (see Results). Consequently, the changes in the dimensions of individual chains A (monitored in W58C81) and B (monitored in W4C42) occur in a continuous, non-cooperative manner, with increasing denaturant concentration.

The continuous expansion of the segment mapping the core of chain B (in W4C42) with increasing GdnHCl concentration is consistent with behavior expected when solvent–chain interactions dominate over intra-chain stabilizing interactions. Such monotonic expansion has been observed for many intrinsically disordered proteins (Hofmann et al. 2012; Alessandro et al. 2016), which tend to swell with increasing denaturant concentration as solvent quality improves and chain–solvent interactions increasingly outweigh chain–chain interactions. Similar expansion of the U state ensemble with increasing denaturant concentration has been reported for a wide range of globular proteins under unfolding conditions (Lakshmikanth et al. 2001), (Merchant et al. 2007; Alessandro et al. 2016).

In contrast, the behavior of chain A (monitored in W58C81) is markedly different. Surprisingly, the segment mapping the core of chain A is seen to undergo gradual contraction upon addition of GdnHCl. In 0 M GdnHCl, this segment has a size of ~38 Å and is 35% expanded relative to its dimension in the U state at high GdnHCl concentration. This would suggest that water is a good solvent for unfolded chain A, as has been suggested for the unfolded forms of globular proteins (Bowman et al. 2020). However, a solvent quality-based interpretation is insufficient to account for this behaviour.

This observation can be rationalized by electrostatic repulsion playing a dominant role in determining the dimensions of the U-like forms of chain A at low denaturant concentrations. Screening of electrostatic interactions by Gdn^+ and Cl^- ions would reduce repulsive interactions between like-charged residues, allowing weak hydrophobic and non-local interactions to promote compaction. Similar salt-induced contraction of unfolded protein has been reported previously, including in intrinsically disordered proteins, where increasing ionic strength leads to chain compaction through electrostatic screening (Müller-Späth et al. 2010; Liu et al. 2014). Low concentrations of GdnHCl have also been shown to induce compaction of the denatured

states of apomyoglobin and cytochrome c (Hagihara et al. 1993). Furthermore, perturbation of even a single electrostatic interaction by mutation has been shown to significantly alter U state dimensions in the absence of any denaturant, and screening of charge by the addition of 0.5 M salt concentration restores the U-state dimensions to that of the wild-type U state ensemble (Pradeep and Udgaonkar 2004a).

It has been suggested that the net charge per residue (NCPR), defined as the net charge normalized by sequence length, is a key determinant of the global dimensions of unfolded and intrinsically disordered polypeptide chains (Holehouse et al. 2017; Mao et al. 2010). Both chains A and B fall within the boundary region of the diagram of states (Figure 4.19), corresponding to the weak polyampholyte regime, in which electrostatic interactions are context-dependent and do not enforce a single U state conformation but can stabilize either expanded or compact ensembles. In this regime, small imbalances between repulsive and attractive electrostatic interactions are sufficient to bias the unfolded ensemble toward expansion or compaction, as established by the classical polymer theory of polyampholytes and by studies on intrinsically disordered protein sequences (Mao et al. 2010; Srivastava and Muthukumar 1996). The sequence segment mapping the core of chain A exhibits a small but finite net charge bias (NCPR ≈ -0.08; net charge ≈ -2 at pH 8), resulting in a weak bias toward repulsive electrostatic interactions and a more expanded ensemble in 0 M GdnHCl. In contrast, the sequence segment mapping the core of chain B is charge-neutral (NCPR ≈ 0; net charge ≈ 0 at pH 8), a condition under which polyampholytes can adopt compact conformations (Bianchi et al. 2020; Srivastava and Muthukumar 1996). Screening of electrostatic interactions upon addition of GdnHCl therefore leads to contraction of chain A and expansion of chain B in the low denaturant regime.

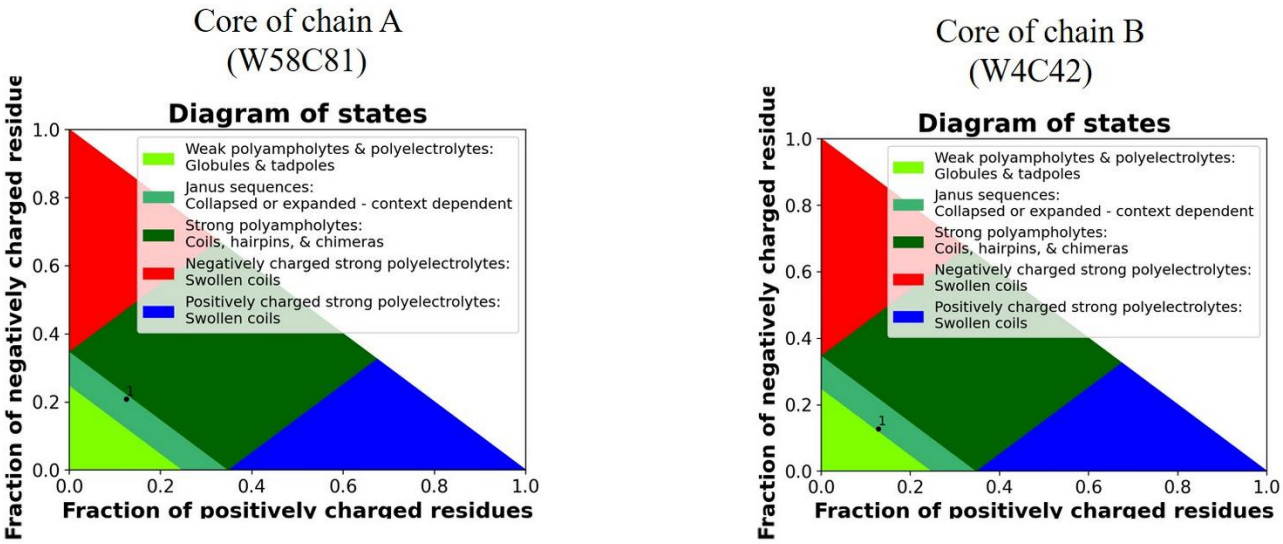


Figure 4.19. Phase diagrams for the core segments of chain A (W58C81) and chain B (W4C42), showing their placement in the weak polyampholyte boundary regime, where electrostatic interactions can support either expanded or compact unfolded-state conformations depending on context (Holehouse et al. 2017).

Together, these observations indicate that the inter-chain interface plays a central role in modulating unfolding cooperativity in dcMN. When the interface is intact (as in the N-like sub-population; see above), it promotes coordinated, cooperative responses across both chains. In contrast, in the U-like sub-population, where the inter-chain interface is already disrupted, this coupling is absent, allowing the two chains to undergo chain-specific, non-cooperative structural rearrangements. Consistent with this view, studies on coiled-coil heterodimers (O'Shea et al. 1991) and PDZ–ligand complexes (Chi et al. 2008) have shown that interfacial contacts can drive coordinated, all-or-none unfolding of coupled regions, whereas more weakly coupled segments display non-cooperative behaviour (Aronsson et al. 2015; Mendes et al. 2019).

Time-resolved fluorescence anisotropy decay measurements reveal site-dependent and asynchronous changes in local motions during unfolding.

The relative amplitudes of the fast and slow decay components (β_{fast} and β_{slow}) report the fraction of molecules that lose anisotropy *via* fast local Trp motion *versus* slower segmental or global motions. In a strictly two-state unfolding process, local probes located in different regions of the protein would therefore be expected to exhibit similar denaturant dependences of these relative amplitudes. Instead, Trp4 and Trp58 show markedly different behaviors. For Trp4, the relative contribution of fast local motion (β_{fast}) increases gradually and in an approximately exponential manner with increasing GdnHCl concentration, indicating a continuous redistribution of the ensemble toward conformations with increasing local flexibility. In contrast, for Trp58, β_{fast} remains negligibly small at low denaturant concentrations and increases gradually only beyond ~ 0.2 M GdnHCl, suggesting that local motional freedom at this site does not occur in the N state. These differences demonstrate that distinct regions of dcMN lose motional constraints at different denaturant concentrations, revealing heterogeneous and asynchronous structural loosening rather than a concerted all-or-none transition.

Moreover, if both fast and slow depolarization pathways were accessible simultaneously to all molecules, the relative amplitudes would be dictated solely by the corresponding rotational correlation times. For example, in the U state, where ϕ_{slow} is 1 ns and ϕ_{fast} is 0.2 ns, the value of β_{fast} expected for such a system would be approximately 0.83 (1/1.2). Instead, the observed value of β_{fast} is ~0.65. This discrepancy indicates that in the U state, and similarly in the N and intermediate states, there exist minor sub-populations (~15% in the U state) in which the Trp side-chain remains dynamically constrained and cannot undergo

independent local motion. Thus, the anisotropy data establish that unfolding proceeds through a progressive redistribution among sub-populations with different local motional constraints and cannot be described by a simple cooperative two-state model. Similar deviations from simple two-state behavior have been revealed by time-resolved fluorescence anisotropy decay measurements of barstar, for which the rotational dynamics in the unfolding transition region could not be described as a weighted sum of the native and unfolded state decays (Swaminathan et al. 1996). In the case of α-subunit of tryptophan synthase, non-monotonic changes in rotational correlation times were observed during folding, indicating asynchronous rather than cooperative structural rearrangements (Bilsel et al. 1999). Similarly, studies on tubulin (Sánchez et al. 2004), creatine kinase (Grossman 1994), and yeast glutathione reductase (Louzada et al. 2003) identified intermediate species and gradual structural loss, consistent with multi-step rather than cooperative unfolding.

Comparison of the unfolding cooperativity of dcMN and MNEI

Although dcMN and MNEI have nearly identical native structures (Figure 4.1a), their unfolding behavior differs markedly when resolved at the level of MEM-derived sub-populations (Figure 4.20). This provides a unique opportunity to isolate the role of chain topology and connectivity in modulating unfolding cooperativity.

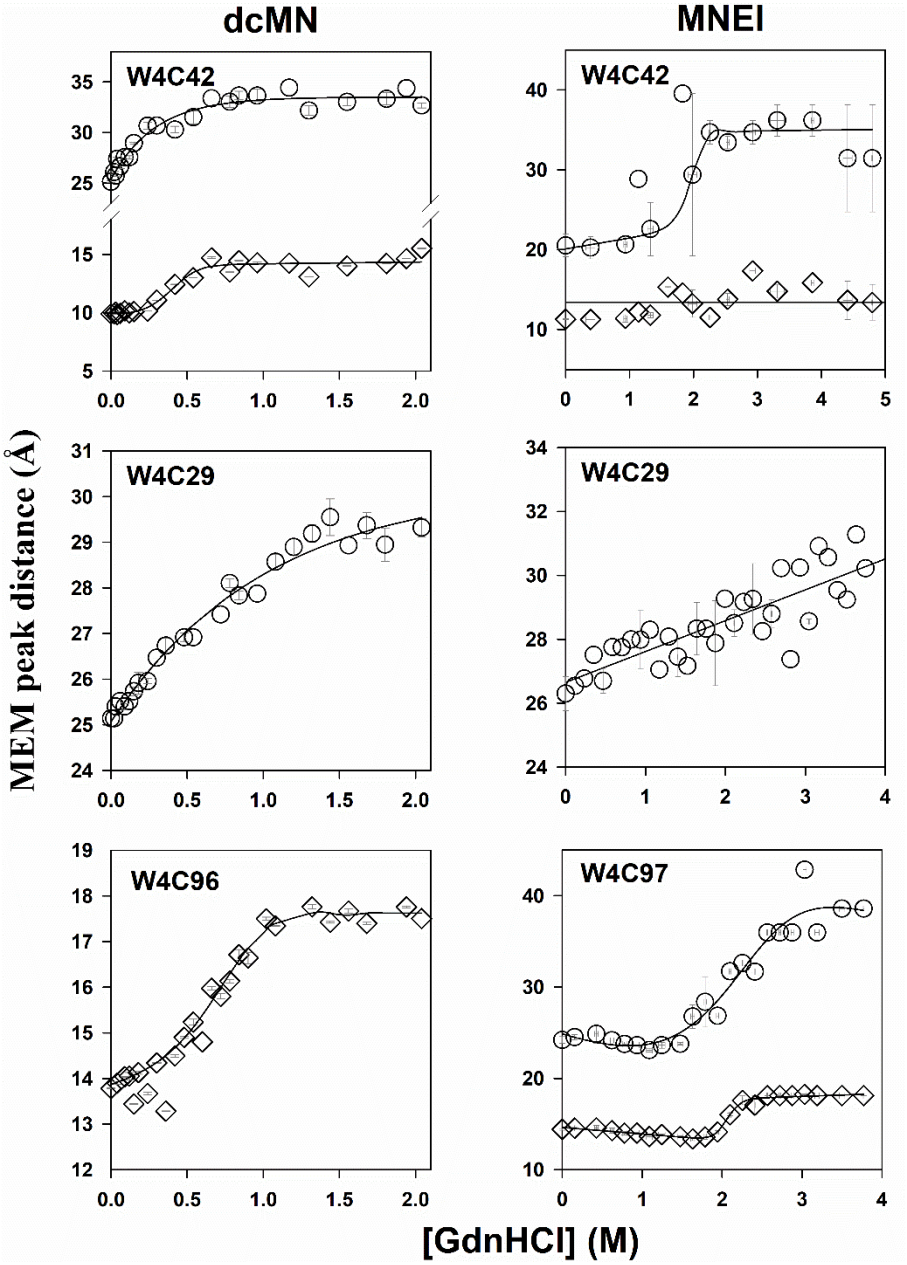


Figure 4.20. Comparison of the changes in the dimensions of the N-like and U-like sub-populations of dcMN and MNEI as a function of GdnHCl concentration. Symbol definitions are identical to those in Figure 4.12. The dcMN panels are reproduced from Figure 4.12, and the data for the corresponding segments of MNEI are reproduced from (Bhatia, Krishnamoorthy, and Udgaonkar 2018).

For the helix segment (monitored in W4C29), the unfolding transition is gradual for both MNEI and dcMN, indicating that the non-cooperative unfolding of the helix is intrinsic to the helix itself and largely insensitive to differences in chain connectivity or overall protein topology. In contrast, topology-dependent differences are observed for the β-sheet core (monitored in W4C42) and inter-chain segments (monitored inW4C96/W4C97). The N-like sub-populations swell via similar cooperative transitions, for both MNEI and dcMN. In contrast, it is seen that the U-like sub-population observed for W4C42 unfolds non-cooperatively in the case of dcMN but cooperatively in the case of MNEI. For the inter-chain segment monitored in W4C96: unfolding of the U-like sub-population is cooperative in the case of MNEI, but all the U-like sub-population is as unfolded as the U state in the case of dcMN.

Overall, these comparisons show that in the case of MNEI, unfolding transitions remain largely coupled at the segmental level, consistent with a covalently continuous polypeptide in which interacting structural elements remain effectively tethered even upon partial expansion. In sharp contrast, dcMN exhibits a clear separation of cooperative and non-cooperative behavior between sub-populations. These observations indicate that covalent continuity stabilizes cooperative responses even within partially expanded ensembles, whereas its absence permits uncoupled, non-cooperative behavior.

These contrasting behaviors can be rationalized in terms of topology and effective concentration. The β2–β3 interface is structurally equivalent in the native structures of dcMN and MNEI; however, in dcMN, the interacting β strands reside on separate polypeptide chains, whereas in MNEI they are covalently linked within a single chain (Figure 4.1a). In the case of dcMN, the effective concentrations governing various inter-chain interactions will be low because the partners are on different chains, and the high entropic cost of bringing the interacting partners together lowers the strengths of the stabilizing interactions. Consequently, unfolding is less cooperative. In the case of MNEI, effective concentrations governing the same interactions are higher because the partners are on one chain. Consequently, the stabilizing interactions are stronger and loss of structure during unfolding is cooperative. In the case of circularly permuted proteins too, the changes in covalent connectivity alone can modulate (un)folding cooperativity, even when native structure is preserved (Viguera and Serrano 1997; Otzen and Fersht 1998; Shank et al. 2010; Radou et al. 2013). It therefore seems likely that the lack of cooperativity seen for the U-like sub-populations of dcMN arises from the two-chain topology and the loss of covalent linkage between interacting β-strand elements.

Chapter 5

Concurrent continuous and activated chain collapse during folding of a small heterodimeric protein

5.1 Introduction

Protein folding often occurs on a rugged free-energy landscape in which multiple partially structured intermediates may be transiently populated. Although many small proteins display apparent two-state kinetics, such behaviour does not preclude the existence of intermediates, as ensemble-averaging probes with limited temporal or structural resolution often fail to detect short-lived or low-abundance species. High resolution approaches, including hydrogen-exchange (HX) coupled to nuclear magnetic resonance (HX-NMR) or mass spectrometry (HX-MS) (Englander and Mayne 1992; Chamberlain, Handel, and Marqusee 1996; Udgaonkar and Baldwin 1988; Hosszu et al. 1997; Baum et al. 1989; Jennings and Wright 1993; Morozova-Roche et al. 1999; Juneja and Udgaonkar 2002, 2003), single-molecule fluorescence resonance energy transfer (sm-FRET) (Schuler, Lipman, and Eaton 2002; Lipman et al. 2003; Schuler and Eaton 2008) and time-resolved FRET (Yamada et al. 2013; Bhatia and Udgaonkar 2022) have therefore been instrumental in revealing folding and unfolding intermediates, on- (Jennings and Wright 1993; Roder, Elöve, and Englander 1988; Wu et al. 2008; Bhatia et al. 2019) and off- (Bollen, Sánchez, and van Mierlo 2004; Bollen, Kamphuis, and van Mierlo 2006; Houwman et al. 2018) pathway, including molten globule-like states (Jennings and Wright 1993; Ptitsyn et al. 1990; Christensen and Pain 1991; Bhatia et al. 2019), as well as revealing multiple folding pathways (Udgaonkar 2008; Cabrita et al. 2011; Chung et al. 2012; Hu et al. 2013; Bhatia, Krishnamoorthy, and Udgaonkar 2021a). These studies have further shown that structural changes during folding can even proceed in an entirely gradual manner (Garcia-Mira et al. 2002; Sadqi, Fushman, and Munoz 2006; Fung et al. 2008; Jha et al. 2009; Bhatia et al. 2019). These observations highlight the fact that protein folding reactions are often far more heterogeneous than suggested by ensemble-averaged thermodynamic or kinetic measurements. This complexity is amplified in the case of heterodimeric proteins, where folding is coupled to chain association rather than being purely intramolecular. In heterodimeric systems whose subunits are unstructured in isolation, folding proceeds from an associated complex rather than from a single unfolded polypeptide chain. Because association enforces proximity without necessarily defining specific inter-chain contacts, folding can initiate from multiple associated conformations, resulting in increased heterogeneity and kinetic complexity.

Double-chain monellin (dcMN) is a heterodimeric protein in which the two chains are largely unstructured in isolation and associate to form a loosely organized encounter complex that subsequently acquires secondary and tertiary structure (Aghera and Udgaonkar 2012). Previous fluorescence and hydrogen-exchange mass spectrometry (HX-MS) studies have

shown that dcMN refolds under mildly denaturing conditions through multiple folding routes, including two parallel pathways: one proceeding directly to the native (N) state and another populating an on-pathway intermediate (Aghera and Udgaonkar 2012; Bhattacharjee and Udgaonkar 2022). Under refolding conditions of 0.1 M GdnHCl, a previous HX-MS study (Bhattacharjee and Udgaonkar 2022) mapped the sequence of structure formation across multiple β-strands, with stepwise acquisition of protection interpreted to arise from the formation of stabilizing tertiary interactions. However, despite its high structural resolution, HX-MS detected only a single dominant on-pathway intermediate during refolding, consistent with ensemble fluorescence measurements (Aghera and Udgaonkar 2012).

It is important to recognize that hydrogen-exchange mass spectrometry (HX-MS) inherently reports only on amide hydrogens that have acquired sufficiently stable hydrogen bonding or tertiary packing on the HX timescale. As a result, collapsed conformations that are weakly protected, or are rapidly interconverting, are difficult to distinguish from the unfolded protein ensemble. In the case of dcMN, HX-MS therefore primarily reports on the dominant, highly protected on-pathway intermediate, whereas earlier collapsed conformations populated immediately after chain association may have remained unresolved. This limitation is underscored by the observation that, although only a single intermediate is detected during refolding of dcMN, native-state HX-MS experiments have revealed multiple high-energy intermediates during unfolding under native and mildly denaturing conditions. While refolding is often assumed to proceed through a reverse sequence of events relative to unfolding, the absence of multiple detectable intermediates during refolding is therefore consistent with the presence of additional, weakly protected intermediates that escape detection by HX-MS. To directly probe such hidden intermediates during the folding of dcMN, an approach that can resolve heterogeneous conformational ensembles and quantify their relative populations, is required. Here, this challenge was addressed using time-resolved FRET analysed with the maximum entropy method (MEM), which provides intramolecular distance distributions during refolding from which distinct conformational sub-populations and their relative fractions can be determined (Bhatia and Udgaonkar 2022).

5.2 Materials and Methods

Experimental procedures were identical to those described in chapter 3.

5.3 Results and Discussion

TNB-labeling does not perturb the structure and stability of the protein significantly

Wild-type dcMN contains a single Trp and a single Cys residue at positions 4 and 42, respectively. Trp4 can serve as a FRET donor, while a TNB moiety attached to Cys42 can serve as the acceptor (Figure 5.1a). This intrinsic donor–acceptor pair (W4C42-TNB) was used to track chain contraction during refolding following dilution from 2.0 to 0.1 M GdnHCl. Previous equilibrium unfolding studies have established that the secondary structure and stability of the TNB-labeled variant is similar to that of the unlabeled protein, confirming that the TNB label does not significantly perturb the structure and stability of the protein (Chapter 4).

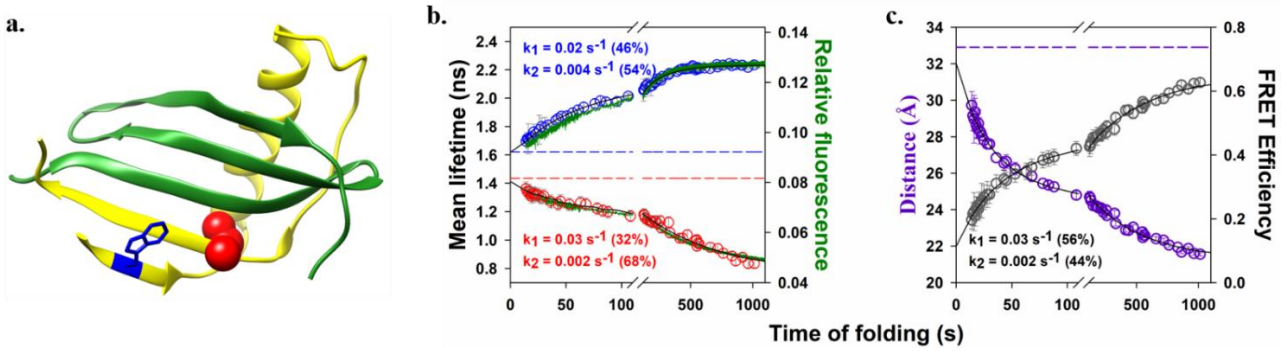


Figure 5.1. Mapping contraction of the protein during folding using FRET as a probe. a) Structure of double-chain monellin (PDB ID: 3MON) showing the donor residue Trp4 (blue sticks) and the acceptor TNB attached to Cys42 (red spheres), drawn using the program Chimera. Chain B is shown in yellow, while chain A is shown in green. b) Kinetics of folding of unlabeled and TNB-labeled W4C42 dcMN in 0.1 M GdnHCl at pH 8 and 25°C, monitored by fluorescence intensity (green traces) and time-resolved FRET (open circles). The mean fluorescence lifetimes of Trp4 during refolding for the unlabeled (blue open circles) and TNB-labeled (red open circles) variants are shown (left y axis). The dashed horizontal lines indicate the mean lifetimes of the unfolded state in 2 M GdnHCl for the unlabeled (blue) and TNB-labeled (red) protein variants. The solid black lines are fits of the mean lifetime data to a double exponential equation. The parameters obtained from the fit are shown in the panel. c) Dependence of FRET efficiency (right y axis) and corresponding distance (left y axis) between Trp4 and the TNB moiety attached to Cys42, on the folding time. The dashed horizontal line corresponds to the average distance for the unfolded protein in 2 M GdnHCl. The solid black lines are fits to a double-exponential equation. Error bars represent the standard errors of measurements from two independent double kinetics experiments.

Validation of the time-resolved measurements during refolding

Time-resolved fluorescence decays of Trp4 were analyzed using both discrete exponential fitting and the maximum entropy method (MEM) (see Materials and Methods). The mean lifetimes obtained from the discrete analysis at different times of refolding overlapped well with the simultaneously measured fluorescence intensity in the double kinetics set up (Figure 5.1b). The FRET efficiencies calculated from the mean lifetimes of the unlabeled and TNB-labeled protein variants increased with refolding time, with a corresponding decrease in donor-acceptor distance as expected during folding (Figure 5.1c). A minor burst phase change in FRET efficiency was also observed. Ensemble-averaged fluorescence lifetimes and intensities alone do not resolve population heterogeneity or the coexistence of expanded and contracted conformational sub-populations during different stages of folding. To directly probe this hidden heterogeneity, time-resolved fluorescence decays were analyzed using the maximum entropy method to obtain fluorescence lifetime distributions (see Materials and Methods).

Bimodal lifetime and distance distributions reveal U-like and N-like sub-populations during refolding

For the TNB-labeled variant, the MEM-derived lifetime distributions were clearly bimodal throughout refolding (Figure 5.2a), whereas those for the unlabeled protein remained largely unimodal (Figure 5.3a). Minor short-lifetime components observed for the unlabeled protein contributed less than 25% of the total population at all refolding times (Figure 5.3b) and were also present in the unfolded state for both unlabeled and TNB-labeled variants. These components are well established to arise from intrinsic Trp photophysics, such as different electronic states or rotameric conformations (Creed 1984; Philips et al. 1988), rather than from FRET quenching. The absence of significant time-dependent shifts in the unlabeled protein further confirmed that the bimodality observed for the labeled variant originates from distance-dependent FRET quenching by TNB and not from changes in the Trp environment. For the labeled variant, the short-lifetime peak (<0.6 ns) was centered near the native-state distribution and was therefore designated N-like (short lifetime, high FRET), whereas the long-lifetime peak (>0.6 ns) was centered near the unfolded-state distribution and was designated U-like (long lifetime, low FRET). Conversion of lifetime distributions into distance distributions, after correcting for changes observed in the unlabeled protein as described previously (Chapter 3), confirmed that the N-like and U-like sub-populations correspond to partially contracted (R_{DA} < 20 Å) and expanded (R_{DA} > 20 Å) conformations, respectively (Figure 5.4). The MEM-

derived distance distributions revealed two distinct features during refolding. First, the distance corresponding to the N-like sub-population shifted continuously toward shorter values with increasing refolding time, indicating progressive contraction within this ensemble. Second, the fraction of the U-like sub-population (R_{DA} > 20 Å) decreased concomitantly with an increase in the N-like sub-population, indicating net population transfer from expanded to contracted conformations during refolding.

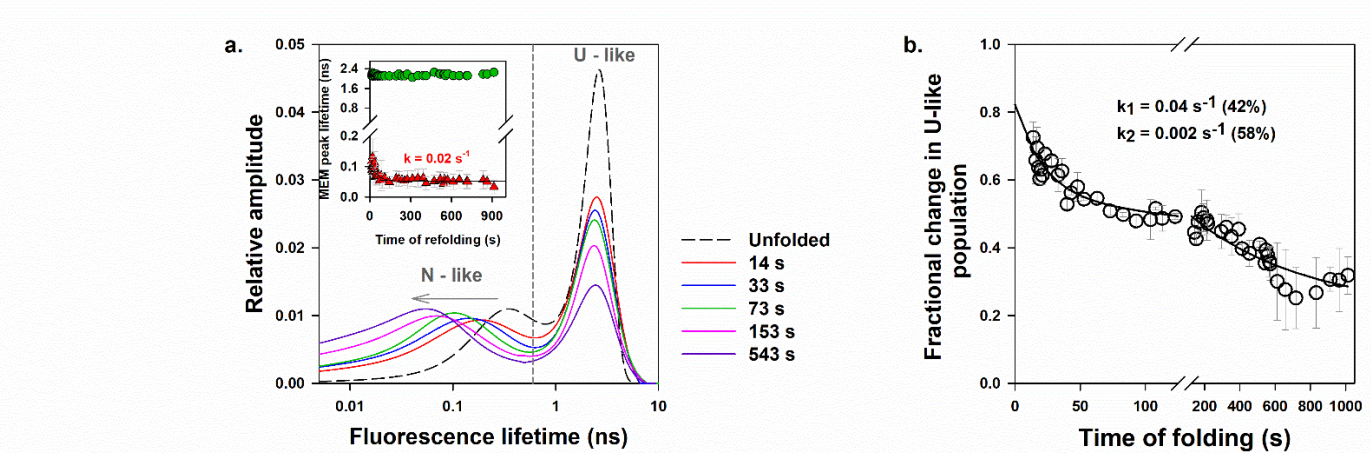


Figure 5.2. MEM-derived fluorescence lifetime distributions of TNB-labeled dcMN during folding. a) Fluorescence lifetime distributions are shown for the TNB-labeled variants at different times of folding shown with different colours, as described in the panel. The x-axis is plotted on a log scale. Each distribution was normalized to its total amplitude. The inset shows the shift in the MEM peak positions for the N-like (triangles) and U-like (circles) sub-populations with time of refolding. The solid line through the kinetic data represents a fit to a single exponential equation. b) Kinetics of conversion from the U-like to N-like sub-population. The solid line through the kinetic data represents a fit to a double exponential equation. The error bars represent the standard errors of measurements from two independent kinetic experiments.

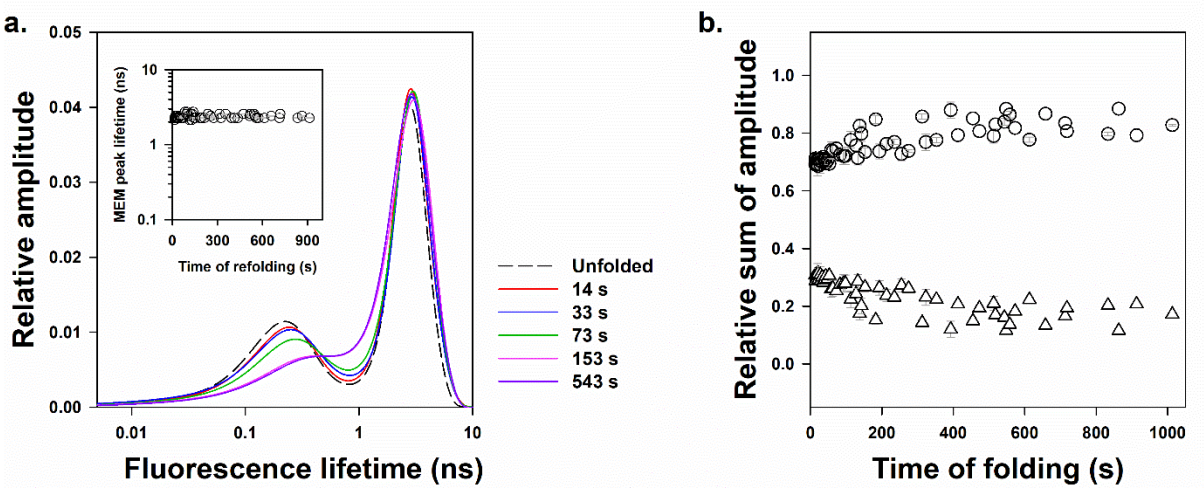


Figure 5.3. The MEM analysis derived lifetime distributions for the unlabeled protein variant. a) The lifetime distributions obtained at different time points of refolding for the unlabeled protein. The inset in each panel shows the shift in the peak lifetime as a function of refolding time. b) The relative sum of amplitudes for the short lifetime (< 0.6 ns, triangles) and the long lifetime (> 0.6 ns, circles) population distributions is the sum of amplitudes for the minor and major population distributions, respectively, divided by the sum of amplitudes for the entire distribution. The relative sum of amplitudes for any sub-population is equivalent to the fraction of that sub-population in the total population.

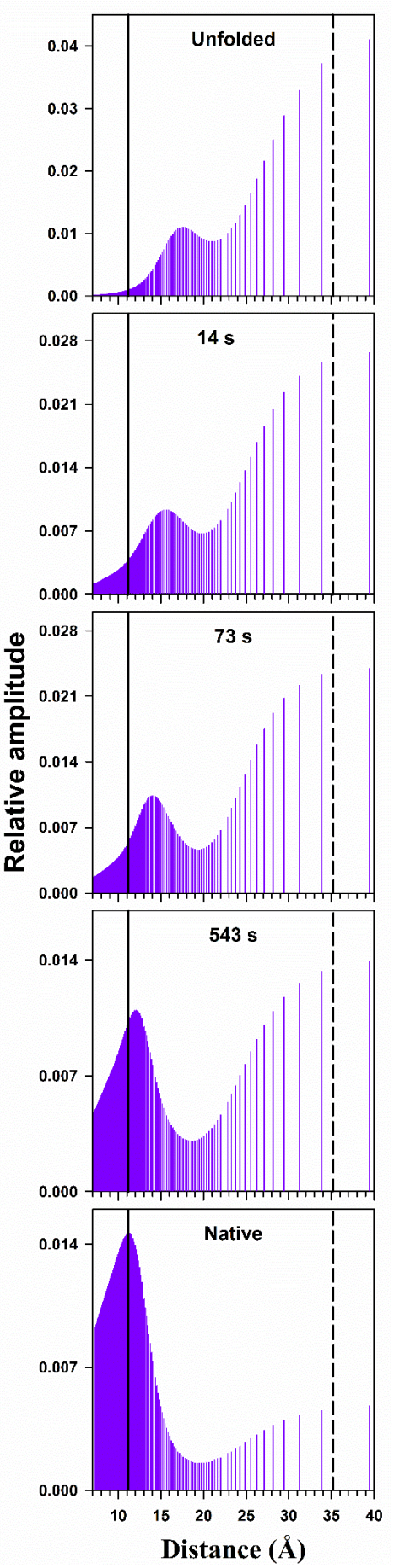


Figure 5.4. Evolution of distance distributions as a function of the time of folding following a 2 to 0.1 M GdnHCl jump. Experimentally derived distance distributions at representative time points of the folding reaction. The top panel shows the U state, the bottom panel corresponds to the distance distribution of the N state, and the middle panels correspond to intermediate time points during folding as described in each panel. The vertical solid and dashed black lines indicate the peak positions of the distance distributions corresponding to the N state and U state, respectively.

MEM reveals hidden conformational heterogeneity during folding

Quantification of sub-populations (Figure 5.2b, see Materials and Methods) showed that at the first observable time point (14 s), the distribution of intramolecular distances was already bimodal with the major component (73%) exhibiting distances comparable to that of the U state, while ~27% populated a partially compact N-like sub-population whose dimensions were ~ 50% smaller than the U state but still ~ 30% larger than the N state. As refolding progressed, the U-like sub-population decreased in a biexponential manner, comprising a fast and a slow phase (Figure 5.2b). This indicates barrier-limited conversion of the U-like molecules to the N-like molecules, and that the two sub-populations are separated by two free energy barriers. Although the structural origin of these barriers cannot yet be directly assigned, studies on the closely related single-chain variant MNEI suggest that the U-like, partially expanded conformations are stabilized by nonspecific interactions, whereas N-like, partially contracted conformations already contain a subset of native-like contacts (Bhatia, Krishnamoorthy, and Udgaonkar 2021b). By analogy, conversion from U-like to N-like molecules in dcMN likely would require the disruption of non-native contacts and/or formation of a limited number of key native interactions, giving rise to the observed activated kinetics. The slow phase most likely arises from proline isomerization, as Pro41 and Pro93 adopt *cis* conformations in the N state. This assignment is supported by the observation that the rate constant corresponding to the slow phase lies in the range expected for proline isomerization (Schmid and Baldwin 1979) and is independent of GdnHCl concentration (Aghera and Udgaonkar 2012). Consistent with this interpretation, the slowest folding phase of the single-chain variant MNEI was previously attributed to proline isomerization (Kaushik and Udgaonkar 2023).

Gradual intrachain contraction within the N-like sub-population

In contrast to this barrier-limited population transfer, contraction within the N-like sub-population itself proceeds gradually (Figure 5.4 and 5.5a). The continuous shift in the N-like distance peak during refolding is well described by an exponential relaxation with a rate constant comparable to that of the fast-folding phase (Figure 5.5a). Because the MEM peak position reflects an ensemble average over conformations, the fitted rate constant represents an effective relaxation rate rather than a true microscopic rate of contraction. Moreover, the magnitude of peak shifts is small relative to the width of the distributions, limiting a definitive assignment of a strictly barrier-free process.

Nevertheless, the presence of numerous crossover points among the distance distributions (Figure 5.6) strongly suggests that contraction within the N-like sub-population occurs through

a multitude of closely spaced states and is therefore close to being continuous in nature. Such diffusive internal rearrangements within a partially collapsed sub-population of molecules are expected on rugged free-energy landscapes and can give rise to approximately exponential relaxation behavior (Sinha and Udgaonkar 2008; Jha et al. 2009; Bhatia et al. 2019; Chung et al. 2015; Hagen 2003). Consistent with this interpretation, at no time during refolding could the experimentally derived fluorescence lifetime distributions be satisfactorily reproduced by a linear combination of the equilibrium unfolded (U) and native (N) state distributions. Fits obtained using weighted sums of the U- and N-state distributions showed significant deviations from the experimentally observed distributions at all refolding times (Figures 5.7 and 5.8), indicating that the observed N-like and U-like sub-populations cannot be described as weighted sums of the equilibrium N and U states, but instead represent structurally distinct intermediate conformations populated during folding.

To interpret the structural origin of this gradual contraction, it is important to consider the location of the W4C42 FRET pair. Although the W4C42 pair reports on an intra-chain distance within chain B, Cys42 lies in the β2–β3 region, which participates in the inter-chain interface in the N state. The observed contraction therefore reflects intrachain structural tightening within a segment that later participates in inter-chain interface formation in the N state. This continuous contraction occurs prior to the formation of stable hydrogen bonding and tertiary locking at β2, which is detected by HX-MS only at later stages when this region becomes sufficiently protected (Bhattacharjee and Udgaonkar 2022). In this context, while gradual folding has been inferred previously for coiled-coil assemblies (O'Shea et al. 1991; Steinmetz et al. 2007) and disordered binding partners (Sugase, Dyson, and Wright 2007; Rogers, Wong, and Clarke 2014), direct kinetic evidence for continuous structural evolution within a globular heterodimer has been lacking.

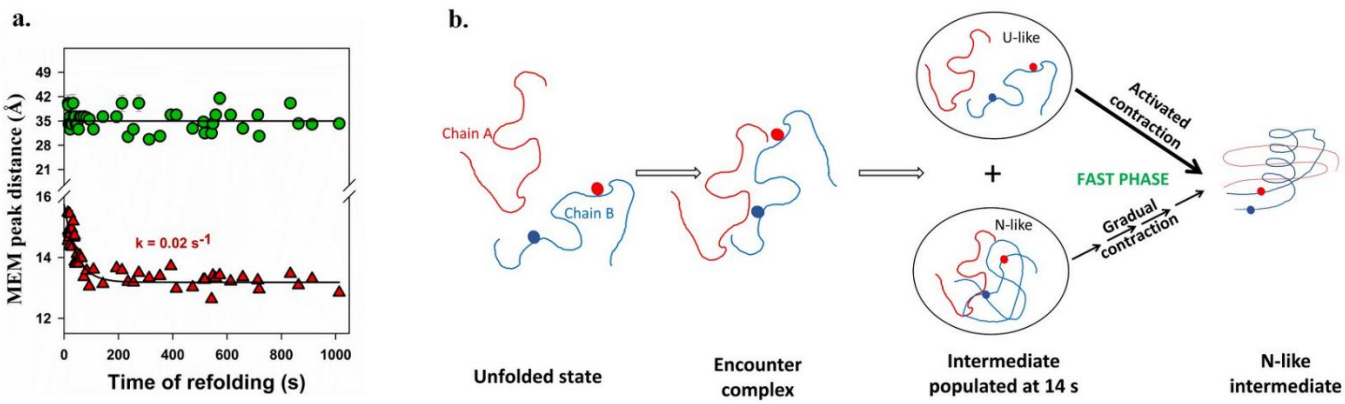


Figure 5.5. Quantification of the change in the dimensions of the N-like and U-like sub-populations with time of refolding. a) Kinetics of contraction of the U-like (green circles) and N-like (red triangles) sub-populations, obtained from the MEM-derived fluorescence lifetime distributions of the TNB-labeled variant and the corresponding unlabeled counterpart (Figures 5.2 and 5.3) using equations 6 and 7 (Chapter 4). The solid line through the N-like data represents a single-exponential fit describing continuous contraction. b) Schematic summarizing the refolding mechanism of dcMN. Following a 2 → 0.1 M GdnHCl jump, association of the two unstructured chains forms an encounter complex that evolves into a heterogeneous intermediate comprising coexisting U-like (expanded) and N-like (partially contracted) sub-populations. During the fast phase, barrier-limited population transfer from the U-like ensemble occurs in parallel with gradual, multi-step contraction within the N-like sub-population, yielding the N-like intermediate.

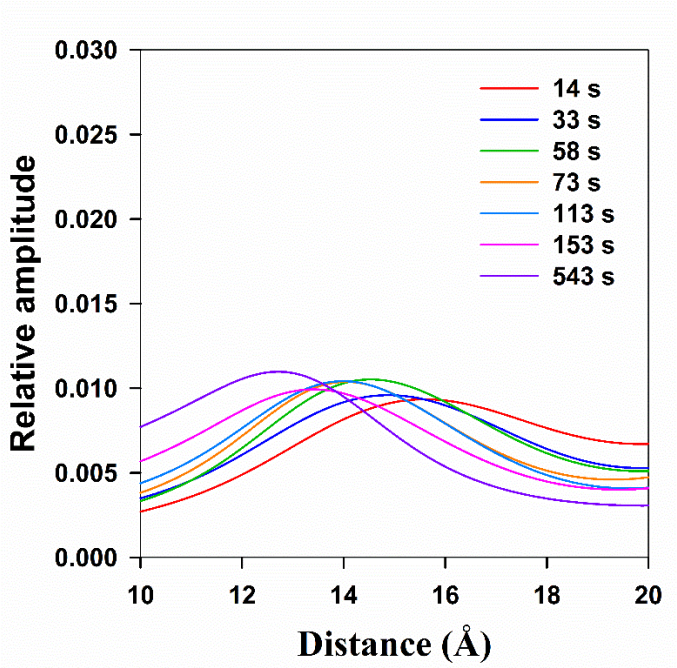


Figure 5.6. The contraction in N-like sub-ensemble involves multiple cross-over points, indicative of a nearly continuous transition. Distributions corresponding to different times of folding are shown in different colors, as described in each panel. Some of the data shown is same as that shown in Figure 5.2a. The y-axis (relative amplitude) in each panel represents the relative populations for the MEM-analysis derived fluorescence lifetime distributions as a function of logarithmically spaced fluorescence lifetimes (x-axis).

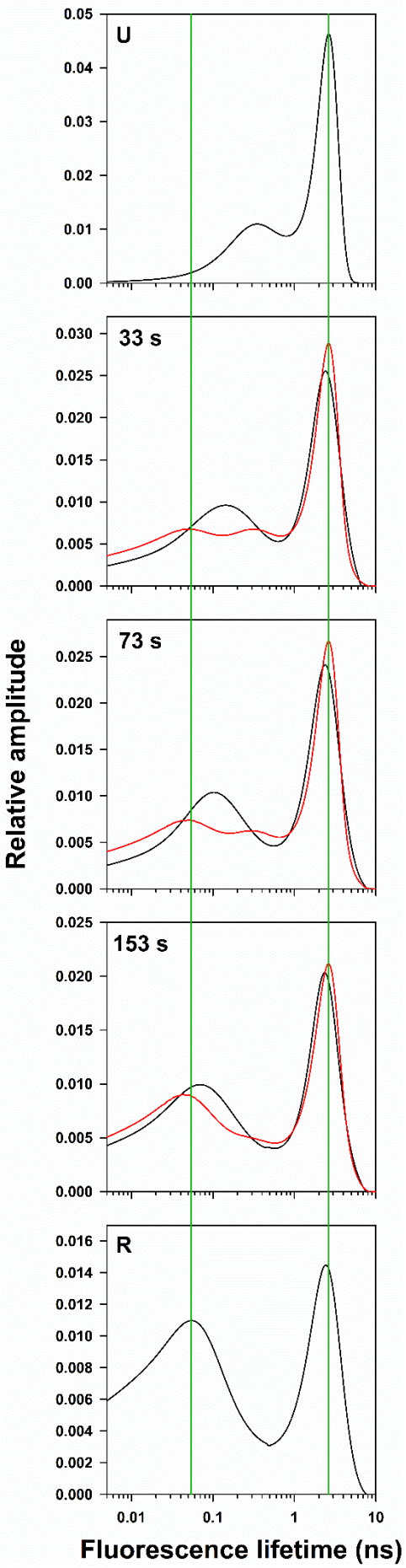


Figure 5.7. Fits of the fluorescence lifetime distributions obtained at different times of refolding to a two state N ↔ U model. The MEM-derived fluorescence lifetime distributions obtained at different times of refolding (indicated in each panel) were fit to a weighted sum of the unfolded state U (τ) (top-most panel) and refolded state R (τ) (bottom-most panel, at 1013 s) fluorescence lifetime distributions (equation 5, Chapter 4). The black lines correspond to the experimentally derived distributions, and the red lines are fits to the two-state model. The vertical green lines indicate the peak lifetimes of the R-state and U-state.

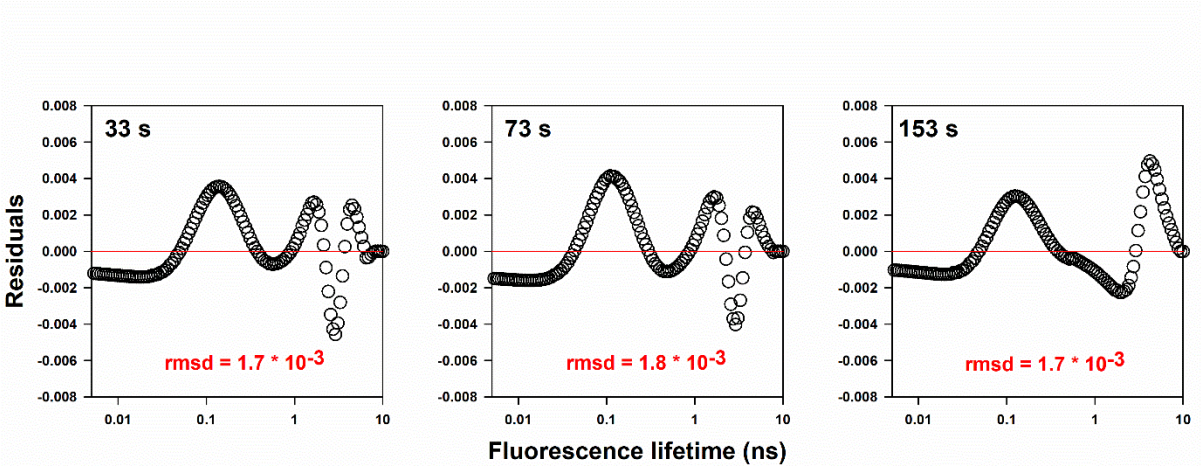


Figure 5.8. Residuals from the fits of the MEM-derived fluorescence lifetime distributions to the weighted sum of the mean of the MEM distributions determined for the R and U states (Figure 5.7). Values in red correspond to the root mean square deviation (rmsd), which was obtained as the square root of the mean of the squares of residuals across all the lifetime values (x-axis).

5.4 Conclusion

These observations reveal that the folding landscape of dcMN is more complex than can be resolved from ensemble-averaged measurements and HX-MS alone. During the fast phase of refolding, two processes occur in parallel: barrier-limited population transfer from U-like to N-like conformations, and gradual intrachain contraction within the N-like sub-population, as summarized schematically in Figure 5.5b.

5.5 Future outlook

Extending this approach to additional segments of dcMN will improve structural resolution and may reveal further levels of site-specific conformational heterogeneity. Refolding studies using the same FRET pairs employed in the unfolding experiments (Chapter 4) have already been performed. Direct comparison of these results with the corresponding single-chain monellin variants will enable a systematic assessment of how chain connectivity reshapes folding pathways and modulates site-specific conformational heterogeneity during folding.

References

Abkevich, VI, AM Gutin, and EI Shakhnovich. 1995. 'Impact of local and non-local interactions on thermodynamics and kinetics of protein folding', *Journal of Molecular Biology*, 252: 460-71.

Agashe, Vishwas R, MCR Shastry, and Jayant B Udgaonkar. 1995. 'Initial hydrophobic collapse in the folding of barstar', *Nature*, 377: 754-57.

Agashe, Vishwas R, and Jayant B Udgaonkar. 1995. 'Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride', *Biochemistry*, 34: 3286-99.

Aghera, Nilesh, Ninganna Earanna, and Jayant B Udgaonkar. 2011. 'Equilibrium unfolding studies of monellin: the double-chain variant appears to be more stable than the single-chain variant', *Biochemistry*, 50: 2434-44.

Aghera, Nilesh, and Jayant B Udgaonkar. 2017. 'Stepwise assembly of β-sheet structure during the folding of an SH3 domain revealed by a pulsed hydrogen exchange mass spectrometry study', *Biochemistry*, 56: 3754-69.

Aghera, Nilesh, and Jayant B Udgaonkar. 2011. 'Heterologous expression, purification and characterization of heterodimeric monellin', *Protein expression purification* 76: 248-53.

Aghera, Nilesh, and Jayant B Udgaonkar. 2012. 'Kinetic studies of the folding of heterodimeric monellin: Evidence for switching between alternative parallel pathways', *Journal of Molecular Biology*, 420: 235-50.

Aghera, Nilesh, and Jayant B Udgaonkar. 2013. 'The utilization of competing unfolding pathways of monellin is dictated by enthalpic barriers', *Biochemistry*, 52: 5770-79.

Aguilar, Ximena. 2014. 'Folding and interaction studies of subunits in protein complexes', Umeå University.

Akiyama, Shuji, Satoshi Takahashi, Tetsunari Kimura, Koichiro Ishimori, Isao Morishima, Yukihiro Nishikawa, and Tetsuro Fujisawa. 2002. 'Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering', *Proceedings of the National Academy of Sciences*, 99: 1329-34.

Aksel, Tural, and Doug Barrick. 2014. 'Direct observation of parallel folding pathways revealed using a symmetric repeat protein system', *Biophysical Journal*, 107: 220-32.

Alderson, T Reid, Jung Ho Lee, Cyril Charlier, Jinfa Ying, and Ad Bax. 2018. 'Propensity for cis-proline formation in unfolded proteins', *ChemBioChem*, 19: 37-42.

Alessandro, Borgia, Zheng Wenwei, Buholzer Karin, Schüler Anja, Hofmann Hagen, Soranno Andrea, Nettels Daniel, Gast Klaus, Grishaev Alexander, and Schuler Benjamin. 2016. 'Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods'.

Andreotti, Amy H 2003. 'Native state proline isomerization: an intrinsic molecular switch', *Biochemistry*, 42: 9515-24.

Anfinsen, Christian B 1973. 'Principles that govern the folding of protein chains', *Science*, 181: 223-30.

Arai, Munehito, Masahiro Iwakura, C Robert Matthews, and Osman Bilsel. 2011. 'Microsecond subdomain folding in dihydrofolate reductase', *Journal of Molecular Biology*, 410: 329-42.

Arai, Munehito, Elena Kondrashkina, Can Kayatekin, C Robert Matthews, Masahiro Iwakura, and Osman Bilsel. 2007. 'Microsecond hydrophobic collapse in the folding of Escherichia coli dihydrofolate reductase, an α/β-type protein', *Journal of Molecular Biology*, 368: 219-29.

Aronsson, Christopher, Staffan Dånmark, Feng Zhou, Per Öberg, Karin Enander, Haibin Su, and Daniel Aili. 2015. 'Self-sorting heterodimeric coiled coil peptides with defined and tuneable self-assembly properties', *Scientific Reports*, 5: 14063.

Arrhenius, Svante 1889. 'Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte', *Zeitschrift für physikalische Chemie*, 4: 96-116.

Arrington, Cammon B, Lynn M Teesch, and Andrew D Robertson. 1999. 'Defining protein ensembles with native-state NH exchange: kinetics of interconversion and cooperative units from combined NMR and MS analysis', *Journal of Molecular Biology*, 285: 1265-75.

Aznauryan, Mikayel, Leonildo Delgado, Andrea Soranno, Daniel Nettels, Jie-rong Huang, Alexander M Labhardt, Stephan Grzesiek, and Benjamin Schuler. 2016. 'Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS', *Proceedings of the National Academy of Sciences*, 113: E5389-E98.

Bai, Yawen, John S Milne, Leland Mayne, and S Walter Englander. 1994. 'Protein stability parameters measured by hydrogen exchange', *Proteins: Structure, Function, Bioinformatics* 20: 4-14.

Bai, Yawen, Tobin R Sosnick, Leland Mayne, and S Walter Englander. 1995. 'Protein folding intermediates: native-state hydrogen exchange', *Science*, 269: 192-97.

Bajaj, Kanika, Mallur S Madhusudhan, Bharat V Adkar, Purbani Chakrabarti, C Ramakrishnan, Andrej Sali, and Raghavan Varadarajan. 2007. 'Stereochemical criteria for prediction of the effects of proline mutations on protein stability', *PLoS computational biology*, 3: e241.

Baldwin, Andrew J, and Lewis E Kay. 2009. 'NMR spectroscopy brings invisible protein states into focus', *Nature chemical biology*, 5: 808-14.

Baldwin, Robert L, and George D Rose. 1999a. 'Is protein folding hierarchic? I. Local structure and peptide folding', *Trends in Biochemical Sciences*, 24: 26-33.

Baldwin, Robert L, and George D Rose. 1999b. 'Is protein folding hierarchic? II. Folding intermediates and transition states', *Trends in Biochemical Sciences*, 24: 77-83.

Ballew, RM, J Sabelko, and M Gruebele. 1996. 'Direct observation of fast protein folding: the initial collapse of apomyoglobin', *Proceedings of the National Academy of Sciences*, 93: 5759-64.

Bartlett, Alice I, and Sheena E Radford. 2010. 'Desolvation and development of specific hydrophobic core packing during Im7 folding', *Journal of molecular biology*, 396: 1329-45.

Baum, Jean, Christopher M Dobson, Philip A Evans, and Claire Hanley. 1989. 'Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig. alpha.-lactalbumin', *Biochemistry*, 28: 7-13.

Beierlein, Frank R, Olaf G Othersen, Harald Lanig, Siegfried Schneider, and Timothy Clark. 2006. 'Simulating FRET from tryptophan: is the rotamer model correct?', *Journal of the American Chemical Society*, 128: 5142-52.

Bhatia, Sandhya, G Krishnamoorthy, Deepak Dhar, and Jayant B Udgaonkar. 2019. 'Observation of continuous contraction and a metastable misfolded state during the collapse and folding of a small protein', *Journal of Molecular Biology*, 431: 3814-26.

Bhatia, Sandhya, G Krishnamoorthy, and Jayant B Udgaonkar. 2018. 'Site-specific time-resolved FRET reveals local variations in the unfolding mechanism in an apparently two-state protein unfolding transition', *Physical Chemistry Chemical Physics*, 20: 3216-32.

Bhatia, Sandhya, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. 2021a. 'Mapping distinct sequences of structure formation differentiating multiple folding pathways of a small protein', *Journal of the American Chemical Society*, 143: 1447-57.

Bhatia, Sandhya, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. 2021b. 'Resolving Site-Specific Heterogeneity of the Unfolded State under Folding Conditions', *The Journal of Physical Chemistry Letters*, 12: 3295-302.

Bhatia, Sandhya, and Jayant B Udgaonkar. 2022. 'Heterogeneity in protein folding and unfolding reactions', *Chemical Reviews*, 122: 8911-35.

Bhattacharjee, Rupam, and Jayant B Udgaonkar. 2021. 'Structural characterization of the cooperativity of unfolding of a heterodimeric protein using hydrogen exchange-mass spectrometry', *Journal of Molecular Biology*, 433: 167268.

Bhattacharjee, Rupam, and Jayant B Udgaonkar. 2022. 'Differentiating between the sequence of structural events on alternative pathways of folding of a heterodimeric protein', *Protein Science*, 31: e4513.

Bhavesh, Neel S, Juhi Juneja, Jayant B Udgaonkar, and Ramakrishna V Hosur. 2004. 'Native and nonnative conformational preferences in the urea-unfolded state of barstar', *Protein Science*, 13: 3085-91.

Bhuyan, Abani K, and Jayant B Udgaonkar. 1998. 'Multiple kinetic intermediates accumulate during the unfolding of horse cytochrome c in the oxidized state', *Biochemistry*, 37: 9147-55.

Bhuyan, Abani K, and Jayant B Udgaonkar. 1999. 'Observation of multistate kinetics during the slow folding and unfolding of barstar', *Biochemistry*, 38: 9158-68.

Bianchi, Greta, Sonia Longhi, Rita Grandori, and Stefania Brocca. 2020. 'Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins', *International journal of molecular sciences*, 21: 6208.

Bilsel, Osman, Li Yang, Jill A Zitzewitz, Joseph M Beechem, and C Robert Matthews. 1999. 'Time-resolved fluorescence anisotropy study of the refolding reaction of the α-subunit of tryptophan synthase reveals nonmonotonic behavior of the rotational correlation time', *Biochemistry*, 38: 4177-87.

Bollen, Yves JM, Monique B Kamphuis, and Carlo PM van Mierlo. 2006. 'The folding energy landscape of apoflavodoxin is rugged: Hydrogen exchange reveals nonproductive misfolded intermediates', *Proceedings of the National Academy of Sciences*, 103: 4095-100.

Bollen, Yves JM, Ignacio E Sánchez, and Carlo PM van Mierlo. 2004. 'Formation of on-and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin', *Biochemistry*, 43: 10475-89.

Borgia, Alessandro, Beth G Wensley, Andrea Soranno, Daniel Nettels, Madeleine B Borgia, Armin Hoffmann, Shawn H Pfeil, Everett A Lipman, Jane Clarke, and Benjamin Schuler. 2012. 'Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy', *Nature communications*, 3: 1195.

Borgia, Alessandro, Wenwei Zheng, Karin Buholzer, Madeleine B Borgia, Anja Schüler, Hagen Hofmann, Andrea Soranno, Daniel Nettels, Klaus Gast, and Alexander Grishaev. 2016. 'Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods', *Journal of the American Chemical Society*, 138: 11714-26.

Bower, David I. 2002. *An introduction to polymer physics* (Cambridge University Press).

Bowler, Bruce E 2012. 'Residual structure in unfolded proteins', *Current Opinion in Structural Biology*, 22: 4-13.

Bowman, Gregory R, and Phillip L Geissler. 2014. 'Extensive conformational heterogeneity within protein cores', *The Journal of Physical Chemistry B*, 118: 6417-23.

Bowman, Micayla A, Joshua A Riback, Anabel Rodriguez, Hongyu Guo, Jun Li, Tobin R Sosnick, and Patricia L Clark. 2020. 'Properties of protein unfolded states suggest broad selection for expanded conformational ensembles', *Proceedings of the National Academy of Sciences*, 117: 23356-64.

Bradley, Christina Marchetti, and Doug Barrick. 2005. 'Effect of multiple prolyl isomerization reactions on the stability and folding kinetics of the notch ankyrin domain: experiment and theory', *Journal of molecular biology*, 352: 253-65.

Brandts, John F, Maureen Brennan, and Lung-Nan Lin. 1977. 'Unfolding and refolding occur much faster for a proline-free proteins than for most proline-containing proteins', *Proceedings of the National Academy of Sciences*, 74: 4178-81.

Brandts, John F, Herbert R Halvorson, and Maureen Brennan. 1975. 'Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues', *Biochemistry*, 14: 4953-63.

Brochon, Jean-Claude. 1994. '[13] Maximum entropy method of data analysis in time-resolved spectroscopy.' in, *Methods in enzymology* (Elsevier).

Bryngelson, Joseph D, José Nelson Onuchic, Nicholas D Socci, and Peter G Wolynes. 1995. 'Funnels, pathways, and the energy landscape of protein folding: a synthesis', *Proteins: Structure, Function, Bioinformatics* 21: 167-95.

Bryngelson, Joseph D, and Peter G Wolynes. 1987. 'Spin glasses and the statistical mechanics of protein folding', *Proceedings of the National Academy of Sciences*, 84: 7524-28.

Burns-Hamuro, Lora L, Paula M Dalessio, and Ira J Ropson. 2004. 'Replacement of proline with valine does not remove an apparent proline isomerization-dependent folding event in CRABP I', *Protein Science*, 13: 1670-76.

Cabrita, Lisa D, Christopher A Waudby, Christopher M Dobson, and John Christodoulou. 2011. 'Solution-state nuclear magnetic resonance spectroscopy and protein folding.' in, *Protein Folding, Misfolding, and Disease: Methods and Protocols* (Springer).

Canet, Denis, Alexander M Last, Paula Tito, Margaret Sunde, Andrew Spencer, David B Archer, Christina Redfield, Carol V Robinson, and Christopher M Dobson. 2002. 'Local cooperativity

in the unfolding of an amyloidogenic variant of human lysozyme', *Nature Structural Biology*, 9: 308-15.

Cantor, Charles R, and Paul R Schimmel. 1980. *Biophysical chemistry: Part II: Techniques for the study of biological structure and function* (Macmillan).

Chakrabartty, Avijit, Tanja Kortemme, S Padmanabhan, and Robert L Baldwin. 1993. 'Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities', *Biochemistry*, 32: 5560-65.

Chamberlain, Aaron K, Tracy M Handel, and Susan Marqusee. 1996. 'Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH', *Nature Structural Biology*, 3: 782-87.

Chan, Chi-Kin, Yi Hu, Satoshi Takahashi, Denis L Rousseau, William A Eaton, and James Hofrichter. 1997. 'Submillisecond protein folding kinetics studied by ultrarapid mixing', *Proceedings of the National Academy of Sciences*, 94: 1779-84.

Chan, Hue Sun, and Ken A Dill. 1990. 'Origins of structure in globular proteins', *Proceedings of the National Academy of Sciences*, 87: 6388-92.

Chatterjee, Amarnath, PM Krishna Mohan, Arati Prabhu, Anindya Ghosh-Roy, and Ramakrishna V Hosur. 2007. 'Equilibrium unfolding of DLC8 monomer by urea and guanidine hydrochloride: Distinctive global and residue level features', *Biochimie*, 89: 117-34.

Chattopadhyay, Krishnananda, Elliot L Elson, and Carl Frieden. 2005. 'The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods', *Proceedings of the National Academy of Sciences*, 102: 2385-89.

Chavez, Leslie L, Shachi Gosavi, Patricia A Jennings, and José N Onuchic. 2006. 'Multiple routes lead to the native state in the energy landscape of the β-trefoil family', *Proceedings of the National Academy of Sciences*, 103: 10254-58.

Chen, Lingling, Gudrun Wildegger, Thomas Kiefhaber, Keith O Hodgson, and Sebastian Doniach. 1998. 'Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering', *Journal of Molecular Biology*, 276: 225-37.

Chi, Celestine N, Lisa Elfström, Yao Shi, Tord Snäll, Åke Engström, and Per Jemth. 2008. 'Reassessing a sparse energetic network within a single protein domain', *Proceedings of the National Academy of Sciences*, 105: 4679-84.

Chiti, Fabrizio, and Christopher M Dobson. 2017. 'Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade', *Annual review of biochemistry*, 86: 27-68.

Chiti, Fabrizio, Niccolò Taddei, Fabiana Baroni, Cristina Capanni, Massimo Stefani, Giampietro Ramponi, and Christopher M Dobson. 2002. 'Kinetic partitioning of protein folding and aggregation', *Nature Structural Biology*, 9: 137-43.

Christensen, H, and RH Pain. 1991. 'Molten globule intermediates and protein folding', *European Biophysics Journal*, 19: 221-29.

Chung, Hoi Sung, Kevin McHale, John M Louis, and William A Eaton. 2012. 'Single-molecule fluorescence experiments determine protein folding transition path times', *Science*, 335: 981-84.

Chung, Hoi Sung, Stefano Piana-Agostinetti, David E Shaw, and William A Eaton. 2015. 'Structural origin of slow diffusion in protein folding', *Science*, 349: 1504-10.

Clark, Patricia L, Kevin W Plaxco, and Tobin R Sosnick. 2020. 'Water as a good solvent for unfolded proteins: folding and collapse are fundamentally different', *Journal of Molecular Biology*, 432: 2882-89.

Cook, Kem H, Franz X Schmid, and Robert L Baldwin. 1979. 'Role of proline isomerization in folding of ribonuclease A at low temperatures', *Proceedings of the National Academy of Sciences*, 76: 6157-61.

Creed, David. 1984. 'The photophysics and photochemistry of the near-UV absorbing amino acids–i. Tryptophan and its simple derivatives', *Photochemistry Photobiology*, 39: 537-62.

Creighton, Thomas E 1990. 'Protein folding', *Biochemical journal*, 270: 1.

Daggett, Valerie, and Alan R Fersht. 2003. 'Is there a unifying mechanism for protein folding?', *Trends in Biochemical Sciences*, 28: 18-25.

Daidone, Isabella, Hannes Neuweiler, Sören Doose, Markus Sauer, and Jeremy C Smith. 2010. 'Hydrogen-bond driven loop-closure kinetics in unfolded polypeptide chains', *PLoS computational biology*, 6: e1000645.

Das, A, BK Sin, AR Mohazab, and SS Plotkin. 2013. 'Unfolded protein ensembles, folding trajectories, and refolding rate prediction', *The Journal of Chemical Physics*, 139.

Dasgupta, Amrita, and Jayant B Udgaonkar. 2010. 'Evidence for initial non-specific polypeptide chain collapse during the refolding of the SH3 domain of PI3 kinase', *Journal of Molecular Biology*, 403: 430-45.

Dasgupta, Amrita, and Jayant B Udgaonkar. 2012. 'Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state', *Biophysical Journal*, 102: 453a.

De Gennes, Pierre-Gilles 1975. 'Collapse of a polymer chain in poor solvents', *Journal de Physique Lettres*, 36: 55-57.

Deniz, Ashok A, Ted A Laurence, Maxime Dahan, Daniel S Chemla, Peter G Schultz, and Shimon Weiss. 2001. 'Ratiometric single-molecule studies of freely diffusing biomolecules', *Annual review of physical chemistry*, 52: 233-53.

Dill, Ken A, Sarina Bromberg, Kaizhi Yue, Hue Sun Chan, Klaus M Fiebig, David P Yee, and Paul D Thomas. 1995. 'Principles of protein folding—a perspective from simple exact models', *Protein Science*, 4: 561-602.

Dill, Ken A, and Hue Sun Chan. 1997. 'From Levinthal to pathways to funnels', *Nature Structural Biology*, 4: 10-19.

Dill, Ken A, and Justin L MacCallum. 2012. 'The protein-folding problem, 50 years on', *Science*, 338: 1042-46.

Dill, Ken A, S Banu Ozkan, M Scott Shell, and Thomas R Weikl. 2008. 'The protein folding problem', *Annu. Rev. Biophys.*, 37: 289-316.

Dill, Ken A, and David Shortle. 1991. 'Denatured states of proteins', *Annual review of biochemistry*, 60: 795-825.

Dobson, Christopher M 2001. 'The structural basis of protein folding and its links with human disease', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356: 133-45.

Dobson, Christopher M. 2003. 'Protein folding and misfolding', *Nature*, 426: 884-90.

Doose, Sören, Hannes Neuweiler, and Markus %J ChemPhysChem Sauer. 2009. 'Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules', *ChemPhysChem*, 10: 1389-98.

Dyson, H Jane, and Peter E Wright. 1993. 'Peptide conformation and protein folding', *Current Opinion in Structural Biology*, 3: 60-65.

Eaton, William A, Victor Munoz, Stephen J Hagen, Gouri S Jas, Lisa J Lapidus, Eric R Henry, and James Hofrichter. 2000. 'Fast kinetics and mechanisms in protein folding', *Annual review of biophysics biomolecular structure*, 29: 327-59.

Eaton, William A, and Peter G Wolynes. 2017. 'Theory, simulations, and experiments show that proteins fold by multiple pathways', *Proceedings of the National Academy of Sciences*, 114: E9759-E60.

Eberhardt, Eric S, Stewart N Loh, and Ronald T Raines. 1993. 'Thermodynamic origin of prolyl peptide bond isomers', *Tetrahedron letters*, 34: 3055.

Eftink, MR, CA Ghiron, RA Kautz, and RO Fox. 1989. 'Fluorescence lifetime studies with staphylococcal nuclease and its site-directed mutant. Test of the hypothesis that proline isomerism is the basis for nonexponential decays', *Biophysical Journal*, 55: 575-79.

Eis, Peggy S, and Joseph R Lakowicz. 1993. 'Time-resolved energy transfer measurements of donor-acceptor distance distributions and intramolecular flexibility of a CCHH zinc finger peptide', *Biochemistry*, 32: 7981-93.

Eisenberg, David, and Mathias Jucker. 2012. 'The amyloid state of proteins in human diseases', *Cell*, 148: 1188-203.

Ellison, Paul A, and Silvia Cavagnero. 2006. 'Role of unfolded state heterogeneity and en-route ruggedness in protein folding kinetics', *Protein Science*, 15: 564-82.

Elson, Elliot L, and Douglas Magde. 1974. 'Fluorescence correlation spectroscopy. I. Conceptual basis and theory', *Biopolymers: Original Research on Biomolecules*, 13: 1-27.

Englander, S Walter 2000. 'Protein folding intermediates and pathways studied by hydrogen exchange', *Annual review of biophysics biomolecular structure* 29: 213-38.

Englander, S Walter, and Neville R Kallenbach. 1983. 'Hydrogen exchange and structural dynamics of proteins and nucleic acids', *Quarterly reviews of biophysics*, 16: 521-655.

Englander, S Walter, and Leland Mayne. 1992. 'Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR', *Annual review of biophysics biomolecular structure* 21: 243-65.

Englander, S Walter, and Leland Mayne. 2014. 'The nature of protein folding pathways', *Proceedings of the National Academy of Sciences*, 111: 15873-80.

Englander, S Walter, Leland Mayne, Zhong-Yuan Kan, and Wenbing Hu. 2016. 'Protein folding—how and why: by hydrogen exchange, fragment separation, and mass spectrometry', *Annual Review of Biophysics*, 45: 135-52.

Evans, Philip A, Christopher M Dobson, Roger A Kautz, Graham Hatfull, and Robert O Fox. 1987. 'Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis', *Nature*, 329: 266-68.

Eyles, S. J., and L. M. Gierasch. 2000. Multiple roles of prolyl residues in structure and folding. J. Mol. Biol. 301:737–747.

Fazelinia, Hossein, Ming Xu, Hong Cheng, and Heinrich Roder. 2014. 'Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome c', *Journal of the American Chemical Society*, 136: 733-40.

Fenwick, R Bryn, Laura Orellana, Santi Esteban-Martín, Modesto Orozco, and Xavier Salvatella. 2014. 'Correlated motions are a fundamental property of β-sheets', *Nature communications*, 5: 4070.

Ferguson, Neil, and Alan R Fersht. 2003. 'Early events in protein folding', *Current Opinion in Structural Biology*, 13: 75-81.

Fersht, Alan R 1995. 'Characterizing transition states in protein folding: an essential step in the puzzle', *Current Opinion in Structural Biology*, 5: 79-84.

Fersht, Alan R 1997. 'Nucleation mechanisms in protein folding', *Current Opinion in Structural Biology*, 7: 3-9.

Finkelstein, AV 2018. '50+ years of protein folding', *Biochemistry*, 83: S3-S18.

Fitzkee, Nicholas C, and George D Rose. 2004. 'Reassessing random-coil statistics in unfolded proteins', *Proceedings of the National Academy of Sciences*, 101: 12497-502.

Flanagan, John M, Mikio Kataoka, David Shortle, and Donald M Engelman. 1992. 'Truncated staphylococcal nuclease is compact but disordered', *Proceedings of the National Academy of Sciences*, 89: 748-52.

Förster, Th 1948. 'Zwischenmolekulare energiewanderung und fluoreszenz', *Annalen der physik*, 437: 55-75.

Frauenfelder, Hans, Fritz Parak, and Robert D Young. 1988. 'Conformational substates in proteins', *Annual review of biophysics biophysical chemistry* 17: 451-79.

Fuertes, Gustavo, Niccolò Banterle, Kiersten M Ruff, Aritra Chowdhury, Davide Mercadante, Christine Koehler, Michael Kachala, Gemma Estrada Girona, Sigrid Milles, and Ankur Mishra. 2017. 'Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements', *Proceedings of the National Academy of Sciences*, 114: E6342-E51.

Fung, Adam, Peng Li, Raquel Godoy-Ruiz, Jose M Sanchez-Ruiz, and Victor Munoz. 2008. 'Expanding the realm of ultrafast protein folding: gpW, a midsize natural single-domain with α+ β topology that folds downhill', *Journal of the American Chemical Society*, 130: 7489-95.

Garcia-Mira, Maria M, Mourad Sadqi, Niels Fischer, Jose M Sanchez-Ruiz, and Victor Munoz. 2002. 'Experimental identification of downhill protein folding', *Science*, 298: 2191-95.

Garel, Jean-Renaud, Barry T Nall, and Robert L Baldwin. 1976. 'Guanidine-unfolded state of ribonuclease A contains both fast-and slow-refolding species', *Proceedings of the National Academy of Sciences*, 73: 1853-57.

Gasymov, Oktay K, Adil R Abduragimov, and Ben J Glasgow. 2012. 'Tryptophan rotamer distribution revealed for the α-helix in tear lipocalin by site-directed tryptophan fluorescence', *The Journal of Physical Chemistry B*, 116: 13381-88.

Gavin, Anne-Claude, Markus Bösche, Roland Krause, Paola Grandi, Martina Marzioch, Andreas Bauer, Jörg Schultz, Jens M Rick, Anne-Marie Michon, and Cristina-Maria Cruciat. 2002. 'Functional organization of the yeast proteome by systematic analysis of protein complexes', *Nature*, 415: 141-47.

Georgescu, Roxana E, Jian-Hua Li, Michel E Goldberg, Maria Luisa Tasayco, and Alain F Chaffotte. 1998. 'Proline isomerization-independent accumulation of an early intermediate and heterogeneity of the folding pathways of a mixed α/β protein, Escherichia coli thioredoxin', *Biochemistry* 37: 10286-97.

Gibbs, Josiah Willard. 1902. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics* (C. Scribner's sons).

Gō, Nobuhiro 1984. 'The consistency principle in protein structure and pathways of folding', *Advances in biophysics*, 18: 149-64.

Goldenberg, David P 2003. 'Computational simulation of the statistical properties of unfolded proteins', *Journal of Molecular Biology*, 326: 1615-33.

Goldenberg, David P, and Thomas E Creighton. 1984. 'Folding pathway of a circular form of bovine pancreatic trypsin inhibitor', *Journal of Molecular Biology*, 179: 527-45.

Goluguri, Rama Reddy, Sreemantee Sen, and Jayant Udgaonkar. 2019. 'Microsecond sub-domain motions and the folding and misfolding of the mouse prion protein', *Elife*, 8: e44766.

Goluguri, Rama Reddy, and Jayant B Udgaonkar. 2016. 'Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein', *Journal of Molecular Biology*, 428: 3102-17.

Goluguri, Rama Reddy, and Jayant B Udgaonkar. 2015. 'Rise of the Helix from a Collapsed Globule during the Folding of Monellin', *Biochemistry*, 54: 5356-65.

Goodsell, David S 1991. 'Inside a living cell', *Trends in Biochemical Sciences*, 16: 203-06.

Goodsell, David S, and Arthur J Olson. 2000. 'Structural symmetry and protein function', *Annual review of biophysics biomolecular structure* 29: 105-53.

Grantcharova, Viara P, David S Riddle, Jed V Santiago, and David Baker. 1998. 'Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain', *Nature Structural Biology*, 5: 714-20.

Grathwohl, Christoph, and Kurt Wüthrich. 1981. 'NMR studies of the rates of proline cis–trans isomerization in oligopeptides', *Biopolymers: Original Research on Biomolecules*, 20: 2623-33.

Greenfield, Norma J 2006. 'Analysis of the kinetics of folding of proteins and peptides using circular dichroism', *Nature protocols*, 1: 2891-99.

Grossman, Steven H 1994. 'An equilibrium study of the dependence of secondary and tertiary structure of creatine kinase on subunit association', *Biochimica et Biophysica Acta -Protein Structure Molecular Enzymology* 1209: 19-23.

Gruebele, Martin 2005. 'Downhill protein folding: evolution meets physics', *Comptes rendus. Biologies*, 328: 701-12.

Gruebele, Martin 2008. 'Comment on probe-dependent and nonexponential relaxation kinetics: Unreliable signatures of downhill protein folding', *Proteins: Structure, Function, Bioinformatics*, 70: 1099-102.

Gruebele, Martin 2014. "Protein dynamics in simulation and experiment." In, 16695-97. Journal of the American Chemical Society: ACS Publications.

Haas, Elisha, Meir Wilchek, Ephraim Katchalski-Katzir, and Izchak Z Steinberg. 1975. 'Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer', *Proceedings of the National Academy of Sciences*, 72: 1807-11.

Hagen, Stephen J. 2003. 'Exponential decay kinetics in "downhill" protein folding', *Proteins: Structure, Function, Bioinformatics* 50: 1-4.

Hagen, Stephen J. 2007. 'Probe-dependent and nonexponential relaxation kinetics: unreliable signatures of downhill protein folding', *Proteins: Structure, Function, Bioinformatics* 68: 205-17.

Hagen, Stephen J, and William A Eaton. 2000. 'Two-state expansion and collapse of a polypeptide', *Journal of molecular biology*, 301: 1019-27.

Hagen, Stephen J, James Hofrichter, Attila Szabo, and William A Eaton. 1996. 'Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding', *Proceedings of the National Academy of Sciences*, 93: 11615-17.

Hagihara, Yoshihisa, Saburo Aimoto, Anthony L Fink, and Yuji Goto. 1993. 'Guanidine hydrochloride-induced folding of proteins', *Journal of Molecular Biology*, 231: 180-84.

Halloran, Kevin T, Yanming Wang, Karunesh Arora, Srinivas Chakravarthy, Thomas C Irving, Osman Bilsel, Charles L Brooks III, and C Robert Matthews. 2019. 'Frustration and folding of a TIM barrel protein', *Proceedings of the National Academy of Sciences*, 116: 16378-83.

Hamid Wani, Ajazul, and Jayant B Udgaonkar. 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: 11226-38.

Haran, Gilad 2012. 'How, when and why proteins collapse: the relation to folding', *Current Opinion in Structural Biology*, 22: 14-20.

Haupts, Ulrich, Sudipta Maiti, Petra Schwille, and Watt W Webb. 1998. 'Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy', *Proceedings of the National Academy of Sciences*, 95: 13573-78.

Henry, ER, and J Hofrichter. 1992. '[8] Singular value decomposition: Application to analysis of experimental data.' in, *Methods in enzymology* (Elsevier).

Henzler-Wildman, Katherine, and Dorothee Kern. 2007. 'Dynamic personalities of proteins', *Nature*, 450: 964-72.

Hernández, Griselda, and David M LeMaster. 2008. 'NMR analysis of native-state protein conformational flexibility by hydrogen exchange.' in, *Protein Structure, Stability, and Interactions* (Springer).

Herning, Thierry, Katsuhide Yutani, Yoshio Taniyama, and Masakazu Kikuchi. 1991. 'Effects of proline mutations on the unfolding and refolding of human lysozyme: the slow refolding kinetic phase does not result from proline cis-trans isomerization', *Biochemistry*, 30: 9882-91.

Hiranmay, Maity, and Reddy Govardhan. 2016. 'Folding of Protein L with Implications for Collapse in the Denatured State Ensemble', *Journal of the American Chemical Society*.

Hodel, Alec, Luke M Rice, Thomas Simonson, Robert O Fox, and Axel T Brünger. 1995. 'Proline cis-trans isomerization in staphylococcal nuclease: Multi-substate free energy perturbation calculations', *Protein Science*, 4: 636-54.

Hoffmann, Armin, Avinash Kane, Daniel Nettels, David E Hertzog, Peter Baumgärtel, Jan Lengefeld, Gerd Reichardt, David A Horsley, Robert Seckler, and Olgica Bakajin. 2007. 'Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy', *Proceedings of the National Academy of Sciences*, 104: 105-10.

Hofmann, Hagen, Andrea Soranno, Alessandro Borgia, Klaus Gast, Daniel Nettels, and Benjamin Schuler. 2012. 'Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy', *Proceedings of the National Academy of Sciences*, 109: 16155-60.

Holehouse, Alex S, Rahul K Das, James N Ahad, Mary OG Richardson, and Rohit V Pappu. 2017. 'CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins', *Biophysical Journal*, 112: 16-21.

Holehouse, Alex S, and Rohit V Pappu. 2018. 'Collapse transitions of proteins and the interplay among backbone, sidechain, and solvent interactions', *Annual Review of Biophysics*, 47: 19-39.

Holtzer, Marilyn Emerson, Eva G Lovett, D Andre d'Avignon, and Alfred Holtzer. 1997. 'Thermal unfolding in a GCN4-like leucine zipper: 13C alpha NMR chemical shifts and local unfolding curves', *Biophysical Journal*, 73: 1031-41.

Hosszu, LL, C Jeremy Craven, Martin J Parker, Mark Lorch, James Spencer, Anthony R Clarke, and Jonathan P Waltho. 1997. 'Structure of a kinetic protein folding intermediate by equilibrium amide exchange', *Nature Structural Biology*, 4: 801-04.

Houry, Walid A, David M Rothwarf, and Harold A Scheraga. 1994. 'A very fast phase in the refolding of disulfide-intact ribonuclease A: implications for the refolding and unfolding pathways', *Biochemistry*, 33: 2516-30.

Houry, Walid A, and Harold A Scheraga. 1996. 'Structure of a Hydrophobically Collapsed Intermediate on the Conformational Folding Pathway of Ribonuclease A Probed by Hydrogen− Deuterium Exchange', *Biochemistry*, 35: 11734-46.

Houwman, Joseline A, Adrie H Westphal, Antonie JWG Visser, Jan Willem Borst, and Carlo PM van Mierlo. 2018. 'Concurrent presence of on-and off-pathway folding intermediates of apoflavodoxin at physiological ionic strength', *Physical Chemistry Chemical Physics*, 20: 7059-72.

Hu, Wenbing, Benjamin T Walters, Zhong-Yuan Kan, Leland Mayne, Laura E Rosen, Susan Marqusee, and S Walter Englander. 2013. 'Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry', *Proceedings of the National Academy of Sciences*, 110: 7684-89.

Huang, Jie-rong, and Stephan Grzesiek. 2010. 'Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin', *Journal of the American Chemical Society*, 132: 694-705.

Jackson, Sophie E, and Alan R Fersht. 1991a. 'Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition', *Biochemistry*, 30: 10428-35.

Jackson, Sophie E, and Alan R Fersht. 1991b. 'Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding', *Biochemistry*, 30: 10436-43.

Jaenicke, R, and H Lilie. 2000. 'Folding and association of oligomeric and multimeric proteins', *Advances in protein chemistry*, 53: 329-401.

Jaenicke, Rainer 1995. 'Folding and association versus misfolding and aggregation of proteins', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 348: 97-105.

Jain, Rinku, Vardhan Dani, Ashima Mitra, Sarika Srivastava, Siddhartha P Sarma, R Varadarajan, and S Ramakumar. 2002. 'Structural consequences of replacement of an α-helical Pro residue in Escherichia coli thioredoxin', *Protein engineering*, 15: 627-33.

Jamin, Marc, and Robert L Baldwin. 1998. 'Two forms of the pH 4 folding intermediate of apomyoglobin', *Journal of Molecular Biology*, 276: 491-504.

Jennings, Patricia A, and Peter E Wright. 1993. 'Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin', *Science*, 262: 892-96.

Jensen, Malene Ringkjøbing, Markus Zweckstetter, Jie-rong Huang, and Martin Blackledge. 2014. 'Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy', *Chemical Reviews*, 114: 6632-60.

Jethva, Prashant N, and Jayant B Udgaonkar. 2017. 'Modulation of the extent of cooperative structural change during protein folding by chemical denaturant', *The Journal of Physical Chemistry B*, 121: 8263-75.

Jha, Santosh Kumar, Amrita Dasgupta, Pooja Malhotra, and Jayant B Udgaonkar. 2011. 'Identification of multiple folding pathways of monellin using pulsed thiol labeling and mass spectrometry', *Biochemistry*, 50: 3062-74.

Jha, Santosh Kumar, Deepak Dhar, Guruswamy Krishnamoorthy, and Jayant B Udgaonkar. 2009. 'Continuous dissolution of structure during the unfolding of a small protein', *Biophysical Journal*, 96: 81a.

Jha, Santosh Kumar, and Jayant B Udgaonkar. 2009. 'Direct evidence for a dry molten globule intermediate during the unfolding of a small protein', *Proceedings of the National Academy of Sciences*, 106: 12289-94.

Julien, Olivier, Subhrangsu Chatterjee, Angela Thiessen, Steffen P Graether, and Brian D Sykes. 2009. 'Differential stability of the bovine prion protein upon urea unfolding', *Protein Science*, 18: 2172-82.

Juneja, Juhi, and Jayant B Udgaonkar. 2002. 'Characterization of the unfolding of ribonuclease A by a pulsed hydrogen exchange study: evidence for competing pathways for unfolding', *Biochemistry*, 41: 2641-54.

Juneja, Juhi, and Jayant B Udgaonkar. 2003. 'NMR studies of protein folding', *Current Science*: 157-72.

Karamanos, Theodoros K, Clare L Pashley, Arnout P Kalverda, Gary S Thompson, Maxim Mayzel, Vladislav Y Orekhov, and Sheena E Radford. 2016. 'A population shift between sparsely populated folding intermediates determines amyloidogenicity', *Journal of the American Chemical Society*, 138: 6271-80.

Karplus, Martin, and David L Weaver. 1976. 'Protein-folding dynamics', *Nature*, 260: 404-06.

Karplus, Martin, and David L Weaver. 1994. 'Protein folding dynamics: The diffusion-collision model and experimental data', *Protein Science*, 3: 650-68.

Kaushik, Anushka, and Jayant B Udgaonkar. 2023. 'Replacement of the native cis prolines by alanine leads to simplification of the complex folding mechanism of a small globular protein', *Biophysical Journal*, 122: 3894-908.

Kaushik, Anushka, and Jayant B Udgaonkar. 2026. 'Intermediate heterogeneity modulates coupling between chain compaction and structure formation during protein folding', *Protein Science,* 35(3), p.e70512.

Kaushik, Anushka, and Jayant B Udgaonkar. 2026. 'Chain entropy modulates cooperativity selectively within intermediate sub-populations during protein unfolding'. *Biochemistry.* DOI: 10.1021/acs.biochem.6c00188.

Kelley, Robert F, and Frederic M Richards. 1987. 'Replacement of proline-76 with alanine eliminates the slowest kinetic phase in thioredoxin folding', *Biochemistry*, 26: 6765-74.

Khurana, Ritu, and Jayant B Udgaonkar. 1994. 'Equilibrium unfolding studies of barstar: evidence for an alternative conformation which resembles a molten globule', *Biochemistry*, 33: 106-15.

Kiefhaber, Thomas, Annett Bachmann, Gudrun Wildegger, and Clemens Wagner. 1997. 'Direct measurement of nucleation and growth rates in lysozyme folding', *Biochemistry*, 36: 5108-12.

Kiefhaber, Thomas, Hans Peter Grunert, Ulrich Hahn, and Franz X Schmid. 1990. 'Replacement of a cis proline simplifies the mechanism of ribonuclease T1 folding', *Biochemistry*, 29: 6475-80.

Kiefhaber, Thomas, Hans-Helmut Kohler, and Franz X Schmid. 1992. 'Kinetic coupling between protein folding and prolyl isomerization: I. Theoretical models', *Journal of Molecular Biology*, 224: 217-29.

Kiefhaber, Thomas, Rainer Quaas, Ulrich Hahn, and Franz X Schmid. 1990a. 'Folding of ribonuclease T1. 1. Existence of multiple unfolded states created by proline isomerization', *Biochemistry*, 29: 3053-61.

Kiefhaber, Thomas, Rainer Quaas, Ulrich Hahn, and Franz X Schmid. 1990b. 'Folding of ribonuclease T1. 2. Kinetic models for the folding and unfolding reactions', *Biochemistry*, 29: 3061-70.

Kiefhaber, Thomas, and Franz X Schmid. 1992. 'Kinetic coupling between protein folding and prolyl isomerization: II. Folding of ribonuclease A and ribonuclease T1', *Journal of Molecular Biology*, 224: 231-40.

Kim, Peter S, and Robert L Baldwin. 1982. 'Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding', *Annual review of biochemistry*, 51: 459-89.

Kim, Peter S, and Robert L Baldwin. 1990. 'Intermediates in the folding reactions of small proteins', *Annual review of biochemistry*, 59: 631-60.

Kim, Sung-Hou, Abraham de Vos, and Craig Ogata. 1988. 'Crystal structures of two intensely sweet proteins', *Trends in Biochemical Sciences*, 13: 13-15.

Kimura, Tetsunari, Shuji Akiyama, Takanori Uzawa, Koichiro Ishimori, Isao Morishima, Tetsuro Fujisawa, and Satoshi Takahashi. 2005. 'Specifically collapsed intermediate in the early stage of the folding of ribonuclease A', *Journal of Molecular Biology*, 350: 349-62.

Kimura, Tetsunari, Takanori Uzawa, Koichiro Ishimori, Isao Morishima, Satoshi Takahashi, Takashi Konno, Shuji Akiyama, and Tetsuro Fujisawa. 2005. 'Specific collapse followed by slow hydrogen-bond formation of β-sheet in the folding of single-chain monellin', *Proceedings of the National Academy of Sciences*, 102: 2748-53.

Kirmizialtin, Serdal, Felicia Pitici, Alfredo E Cardenas, Ron Elber, and D Thirumalai. 2020. 'Dramatic shape changes occur as Cytochrome c folds', *The Journal of Physical Chemistry B*, 124: 8240-48.

Kishore, Megha, G Krishnamoorthy, and Jayant B Udgaonkar. 2013. 'Critical evaluation of the two-state model describing the equilibrium unfolding of the PI3K SH3 domain by time-resolved fluorescence resonance energy transfer', *Biochemistry*, 52: 9482-96.

Klein-Seetharaman, Judith, Maki Oikawa, Shaun B Grimshaw, Julia Wirmer, Elke Duchardt, Tadashi Ueda, Taiji Imoto, Lorna J Smith, Christopher M Dobson, and Harald Schwalbe. 2002. 'Long-range interactions within a nonnative protein', *Science*, 295: 1719-22.

Klimov, DK, and D Thirumalai. 1996. 'Factors governing the foldability of proteins', *Proteins: Structure, Function, Bioinformatics* 26: 411-41.

Knowles, Tuomas PJ, Michele Vendruscolo, and Christopher M Dobson. 2014. 'The amyloid state and its association with protein misfolding diseases', *Nature reviews Molecular cell biology*, 15: 384-96.

Koerdel, Johan, Sture Forsen, Torbjoern Drakenberg, and Walter J Chazin. 1990. 'The rate and structural consequences of proline cis-trans isomerization in calbindin D9k: NMR studies of the minor (cis-Pro43) isoform and the Pro43Gly mutant', *Biochemistry*, 29: 4400-09.

Kohn, Jonathan E, Ian S Millett, Jaby Jacob, Bojan Zagrovic, Thomas M Dillon, Nikolina Cingel, Robin S Dothager, Soenke Seifert, P Thiyagarajan, and Tobin R Sosnick. 2004. 'Random-coil

behavior and the dimensions of chemically unfolded proteins', *Proceedings of the National Academy of Sciences*, 101: 12491-96.

Konuma, Tsuyoshi, Kazumasa Sakurai, Masanori Yagi, Yuji Goto, Tetsuro Fujisawa, and Satoshi Takahashi. 2015. 'Highly collapsed conformation of the initial folding intermediates of β-lactoglobulin with non-native α-helix', *Journal of Molecular Biology*, 427: 3158-65.

Krieger, Florian, Beat Fierz, Oliver Bieri, Mario Drewello, and Thomas Kiefhaber. 2003. 'Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding', *Journal of Molecular Biology*, 332: 265-74.

Krieger, Florian, Andreas Möglich, and Thomas Kiefhaber. 2005. 'Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains', *Journal of the American Chemical Society*, 127: 3346-52.

Krishna, Mallela MG, Linh Hoang, Yan Lin, and S Walter Englander. 2004. 'Hydrogen exchange methods to study protein folding', *Methods*, 34: 51-64.

Krishna Mohan, PM, Swagata Chakraborty, and Ramakrishna V Hosur. 2009. 'NMR investigations on residue level unfolding thermodynamics in DLC8 dimer by temperature dependent native state hydrogen exchange', *Journal of biomolecular NMR*, 44: 1-11.

Krishnamoorthy, G 2018. 'Fluorescence spectroscopy for revealing mechanisms in biology: Strengths and pitfalls', *Journal of Biosciences*, 43: 555-67.

Krogan, Nevan J, Gerard Cagney, Haiyuan Yu, Gouqing Zhong, Xinghua Guo, Alexandr Ignatchenko, Joyce Li, Shuye Pu, Nira Datta, and Aaron P Tikuisis. 2006. 'Global landscape of protein complexes in the yeast Saccharomyces cerevisiae', *Nature*, 440: 637-43.

Kuwajima, Kunihiro. 1989. 'The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure'.

Kuzmenkina, Elza V, Colin D Heyes, and G Ulrich Nienhaus. 2005. 'Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions', *Proceedings of the National Academy of Sciences*, 102: 15471-76.

Kuzmenkina, Elza V, Colin D Heyes, and G Ulrich Nienhaus. 2006. 'Single-molecule FRET study of denaturant induced unfolding of RNase H', *Journal of Molecular Biology*, 357: 313-24.

Lakowicz, Joseph R. 2006. *Principles of fluorescence spectroscopy* (Springer).

Lakowicz, Joseph R, Józef Kuśba, Henryk Szmacinski, Ignacy Gryczynski, Peggy S Eis, Wieslaw Wiczk, and Michael L Johnson. 1991. 'Resolution of end-to-end diffusion coefficients and distance distributions of flexible molecules using fluorescent donor-acceptor and donor-quencher pairs', *Biopolymers: Original Research on Biomolecules*, 31: 1363-78.

Lakshmikanth, GS, K Sridevi, G Krishnamoorthy, and Jayant B Udgaonkar. 2001. 'Structure is lost incrementally during the unfolding of barstar', *Nature Structural Biology*, 8: 799-804.

Lapidus, Lisa J, William A Eaton, and James Hofrichter. 2000. 'Measuring the rate of intramolecular contact formation in polypeptides', *Proceedings of the National Academy of Sciences*, 97: 7220-25.

Laurence, Ted A, Xiangxu Kong, Marcus Jäger, and Shimon Weiss. 2005. 'Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins', *Proceedings of the National Academy of Sciences*, 102: 17348-53.

Lewis, Peter N, Frank A Momany, and Harold A Scheraga. 1973. 'Chain reversals in proteins', *Biochimica et Biophysica Acta -Protein Structure*, 303: 211-29.

Li, Hua, and Carl Frieden. 2005. 'NMR studies of 4-19F-phenylalanine-labeled intestinal fatty acid binding protein: evidence for conformational heterogeneity in the native state', *Biochemistry*, 44: 2369-77.

Lietzow, Michael A, Marc Jamin, H Jane Dyson, and Peter E Wright. 2002. 'Mapping long-range contacts in a highly unfolded protein', *Journal of Molecular Biology*, 322: 655-62.

Lipman, Everett A, Benjamin Schuler, Olgica Bakajin, and William A Eaton. 2003. 'Single-molecule measurement of protein folding kinetics', *Science*, 301: 1233-35.

Liu, Baoxu, Darius Chia, Veronika Csizmok, Patrick Farber, Julie D Forman-Kay, and Claudiu C Gradinaru. 2014. 'The effect of intrachain electrostatic repulsion on conformational disorder and dynamics of the Sic1 protein', *The Journal of Physical Chemistry B*, 118: 4088-97.

Liu, Feng, and Martin Gruebele. 2007. 'Tuning λ6–85 towards downhill folding at its melting temperature', *Journal of Molecular Biology*, 370: 574-84.

Livesey, AK, and JC Brochon. 1987. 'Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method', *Biophysical Journal*, 52: 693-706.

Louzada, Paulo Roberto, Adriano Sebollela, Marcelo E Scaramello, and Sérgio T Ferreira. 2003. 'Predissociated dimers and molten globule monomers in the equilibrium unfolding of yeast glutathione reductase', *Biophysical Journal*, 85: 3255-61.

MacArthur, Malcolm W, and Janet M Thornton. 1991. 'Influence of proline residues on protein conformation', *Journal of Molecular Biology*, 218: 397-412.

Magg, Christine, Jan Kubelka, Georg Holtermann, Elisha Haas, and Franz X Schmid. 2006. 'Specificity of the initial collapse in the folding of the cold shock protein', *Journal of molecular biology*, 360: 1067-80.

Magg, Christine, and Franz X Schmid. 2004. 'Rapid collapse precedes the fast two-state folding of the cold shock protein', *Journal of molecular biology*, 335: 1309-23.

Maglia, Giovanni, Abel Jonckheer, Marc De Maeyer, Jean-Marie Frère, and Yves Engelborghs. 2008. 'An unusual red-edge excitation and time-dependent Stokes shift in the single tryptophan mutant protein DD-carboxypeptidase from Streptomyces: The role of dynamics and tryptophan rotamers', *Protein Science*, 17: 352-61.

Maiti, Sudipta, Ulrich Haupts, and Watt W Webb. 1997. 'Fluorescence correlation spectroscopy: diagnostics for sparse molecules', *Proceedings of the National Academy of Sciences*, 94: 11753-57.

Maity, Haripada, Woon Ki Lim, Jon N Rumbley, and S Walter Englander. 2003. 'Protein hydrogen exchange mechanism: local fluctuations', *Protein Science*, 12: 153-60.

Maity, Hiranmay, and Govardhan Reddy. 2018. 'Thermodynamics and kinetics of single-chain monellin folding with structural insights into specific collapse in the denatured state ensemble', *Journal of Molecular Biology*, 430: 465-78.

Maki, Kosuke, Teikichi Ikura, Toshiya Hayano, Nobuhiro Takahashi, and Kunihiro Kuwajima. 1999. 'Effects of proline mutations on the folding of staphylococcal nuclease', *Biochemistry*, 38: 2213-23.

Malhotra, Pooja, Prashant N Jethva, and Jayant B Udgaonkar. 2017. 'Chemical denaturants smoothen ruggedness on the free energy landscape of protein folding', *Biochemistry*, 56: 4053-63.

Malhotra, Pooja, and Jayant B Udgaonkar. 2015. 'Tuning cooperativity on the free energy landscape of protein folding', *Biochemistry*, 54: 3431-41.

Malhotra, Pooja, and Jayant B Udgaonkar. 2014. 'High-energy intermediates in protein unfolding characterized by thiol labeling under nativelike conditions', *Biochemistry*, 53: 3608-20.

Malhotra, Pooja, and Jayant B Udgaonkar. 2016a. 'How cooperative are protein folding and unfolding transitions?', *Protein Science*, 25: 1924-41.

Malhotra, Pooja, and Jayant B Udgaonkar. 2016b. 'Secondary structural change can occur diffusely and not modularly during protein folding and unfolding reactions', *Journal of the American Chemical Society*, 138: 5866-78.

Mao, Albert H, Scott L Crick, Andreas Vitalis, Caitlin L Chicoine, and Rohit V Pappu. 2010. 'Net charge per residue modulates conformational ensembles of intrinsically disordered proteins', *Proceedings of the National Academy of Sciences*, 107: 8183-88.

Mascarenhas, Nahren Manuel, Vishram L Terse, and Shachi Gosavi. 2018. 'Intrinsic disorder in a well-folded globular protein', *The Journal of Physical Chemistry B*, 122: 1876-84.

Mateos, Borja, Clara Conrad-Billroth, Marco Schiavina, Andreas Beier, Georg Kontaxis, Robert Konrat, Isabella C Felli, and Roberta Pierattelli. 2020. 'The ambivalent role of proline residues in an intrinsically disordered protein: from disorder promoters to compaction facilitators', *Journal of Molecular Biology*, 432: 3093-111.

Matouschek, Andreas, James T Kellis Jr, Luis Serrano, and Alan R Fersht. 1989. 'Mapping the transition state and pathway of protein folding by protein engineering', *Nature*, 340: 122-26.

Mayr, Lorenz M, Dieter Willbold, Paul Rösch, and Franz X Schmid. 1994. 'Generation of a non-prolyl cis peptide bond in ribonuclease T1', *Journal of Molecular Biology*, 240: 288-93.

Mazouchi, Amir, Zhenfu Zhang, Abdullah Bahram, Gregory-Neal Gomes, Hong Lin, Jianhui Song, Hue Sun Chan, Julie D Forman-Kay, and Claudiu C Gradinaru. 2016. 'Conformations of a metastable SH3 domain characterized by smFRET and an excluded-volume polymer model', *Biophysical Journal*, 110: 1510-22.

McCammon, J Andrew, Bruce R Gelin, and Martin Karplus. 1977. 'Dynamics of folded proteins', *Nature*, 267: 585-90.

Meier, Sebastian, Mark Strohmeier, Martin Blackledge, and Stephan Grzesiek. 2007. 'Direct observation of dipolar couplings and hydrogen bonds across a β-hairpin in 8 M urea', *Journal of the American Chemical Society*, 129: 754-55.

Mendes, Luis Felipe S, Mariana RB Batista, Peter J Judge, Anthony Watts, Christina Redfield, and Antonio J Costa-Filho. 2019. 'Conformational flexibility of GRASP protein and its constituent PDZ subdomains reveals structural basis of its promiscuous interactome', *bioRxiv*: 666495.

Meng, Wenli, Nicholas Lyle, Bowu Luan, Daniel P Raleigh, and Rohit V Pappu. 2013. 'Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure', *Proceedings of the National Academy of Sciences*, 110: 2123-28.

Merchant, Kusai A, Robert B Best, John M Louis, Irina V Gopich, and William A Eaton. 2007. 'Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations', *Proceedings of the National Academy of Sciences*, 104: 1528-33.

Milla, Marcos E, and Robert T Sauer. 1994. 'P22 Arc repressor: folding kinetics of a single-domain, dimeric protein', *Biochemistry*, 33: 1125-33.

Mizukami, Takuya, Ming Xu, Hong Cheng, Heinrich Roder, and Kosuke Maki. 2013. 'Nonuniform chain collapse during early stages of staphylococcal nuclease folding detected by fluorescence resonance energy transfer and ultrarapid mixing methods', *Protein Science*, 22: 1336-48.

Möglich, Andreas, Karin Joder, and Thomas Kiefhaber. 2006. 'End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation', *Proceedings of the National Academy of Sciences*, 103: 12394-99.

Moors, Samuel LC, Mario Hellings, Marc De Maeyer, Yves Engelborghs, and Arnout Ceulemans. 2006. 'Tryptophan rotamers as evidenced by X-ray, fluorescence lifetimes, and molecular dynamics modeling', *Biophysical Journal*, 91: 816-23.

Moors, Samuel LC, Abel Jonckheer, Marc De Maeyer, Yves Engelborghs, and Arnout Ceulemans. 2008. 'How do rotameric conformations influence the time-resolved fluorescence of tryptophan in proteins? A perspective based on molecular modeling and quantum chemistry', *Current Protein Peptide Science* 9: 427-46.

Morar, Artemiza S, Alina Olteanu, Gregory B Young, and Gary J %J Protein Science Pielak. 2001. 'Solvent-induced collapse of α-synuclein and acid-denatured cytochrome c', *Protein Science*, 10: 2195-99.

Morozova-Roche, Ludmilla A, Jonathan A Jones, Wim Noppe, and Christopher M Dobson. 1999. 'Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme', *Journal of Molecular Biology*, 289: 1055-73.

Mukaiyama, Atsushi, Takashi Nakamura, Koki Makabe, Kosuke Maki, Yuji Goto, and Kunihiro Kuwajima. 2013. 'Native-state heterogeneity of β2-microglobulin as revealed by kinetic folding and real-time NMR experiments', *Journal of Molecular Biology*, 425: 257-72.

Müller-Späth, Sonja, Andrea Soranno, Verena Hirschfeld, Hagen Hofmann, Stefan Rüegger, Luc Reymond, Daniel Nettels, and Benjamin Schuler. 2010. 'Charge interactions can dominate the dimensions of intrinsically disordered proteins', *Proceedings of the National Academy of Sciences*, 107: 14609-14.

Nakano, Takayuki, Lisa C Antonino, Robert O Fox, and Anthony L Fink. 1993. 'Effect of proline mutations on the stability and kinetics of folding of staphylococcal nuclease', *Biochemistry*, 32: 2534-41.

Napolitano, Silvia, Aditya Pokharna, Rudi Glockshuber, and Alvar D Gossert. 2021. 'The trans-to-cis proline isomerization in E. coli Trx folding is accelerated by trans prolines', *Biophysical Journal*, 120: 5207-18.

Narayanan, Chitra, and Cristiano L Dias. 2013. 'Hydrophobic interactions and hydrogen bonds in β-sheet formation', *The Journal of Chemical Physics*, 139.

Nath, Utpal, and Jayant B Udgaonkar. 1995. 'Perturbation of a tertiary hydrogen bond in barstar by mutagenesis of the sole His residue to Gln leads to accumulation of at least one equilibrium folding intermediate', *Biochemistry*, 34: 1702-13.

Nath, Utpal, and Jayant B Udgaonkar. 1997. 'Folding of tryptophan mutants of barstar: evidence for an initial hydrophobic collapse on the folding pathway', *Biochemistry*, 36: 8602-10.

Navon, Amiel, Varda Ittah, Pavel Landsman, Harold A Scheraga, and Elisha Haas. 2001. 'Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein', *Biochemistry*, 40: 105-18.

Neri, Dario, Martin Billeter, Gerhard Wider, and Kurt Wüthrich. 1992. 'NMR determination of residual structure in a urea-denatured protein, the 434-repressor', *Science*, 257: 1559-63.

Nettels, Daniel, Sonja Müller-Späth, Frank Küster, Hagen Hofmann, Dominik Haenni, Stefan Rüegger, Luc Reymond, Armin Hoffmann, Jan Kubelka, and Benjamin Heinz. 2009. 'Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins', *Proceedings of the National Academy of Sciences*, 106: 20740-45.

Neuweiler, Hannes, Christopher M Johnson, and Alan R Fersht. 2009. 'Direct observation of ultrafast folding and denatured state dynamics in single protein molecules', *Proceedings of the National Academy of Sciences*, 106: 18569-74.

Noguera, Martín E, Diego S Vazquez, Gerardo Ferrer-Sueta, William A Agudelo, Eduardo Howard, Rodolfo M Rasia, Bruno Manta, Alexandra Cousido-Siah, André Mitschler, and Alberto Podjarny. 2017. 'Structural variability of E. coli thioredoxin captured in the crystal structures of single-point mutants', *Scientific Reports*, 7: 42343.

O'Shea, Erin K, Juli D Klemm, Peter S Kim, and Tom Alber. 1991. 'X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil', *Science*, 254: 539-44.

Odefey, Christian, Lorenz M Mayr, and Franz X Schmid. 1995. 'Non-prolyl cis-trans peptide bond isomerization as a rate-determining step in protein unfolding and refolding', *Journal of molecular biology*, 245: 69-78.

Ogasahara, Kyoko, and Katsuhide Yutani. 1997. 'Equilibrium and kinetic analyses of unfolding and refolding for the conserved proline mutants of tryptophan synthase α subunit', *Biochemistry*, 36: 932-40.

Oikawa, Hiroyuki, Kiyoto Kamagata, Munehito Arai, and Satoshi Takahashi. 2015. 'Complexity of the folding transition of the B domain of protein A revealed by the high-speed tracking of single-molecule fluorescence time series', *The Journal of Physical Chemistry B*, 119: 6081-91.

Onuchic, José Nelson, Zaida Luthey-Schulten, and Peter G Wolynes. 1997. 'Theory of protein folding: the energy landscape perspective', *Annual review of physical chemistry*, 48: 545-600.

Onuchic, José Nelson, and Peter G Wolynes. 2004. 'Theory of protein folding', *Current Opinion in Structural Biology*, 14: 70-75.

Osváth, Szabolcs, and Martin Gruebele. 2003. 'Proline can have opposite effects on fast and slow protein folding phases', *Biophysical Journal*, 85: 1215-22.

Otosu, Takuhiro, Kunihiko Ishii, Hiroyuki Oikawa, Munehito Arai, Satoshi Takahashi, and Tahei Tahara. 2017. 'Highly heterogeneous nature of the native and unfolded states of the B domain of protein a revealed by two-dimensional fluorescence lifetime correlation spectroscopy', *The Journal of Physical Chemistry B*, 121: 5463-73.

Otzen, Daniel E, and Alan R Fersht. 1998. 'Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus', *Biochemistry*, 37: 8139-46.

Pace, C Nick, and J Martin Scholtz. 1997. 'Measuring the conformational stability of a protein', *Protein structure: A practical approach*, 2: 299-321.

Pal, Debnath, and Pinak Chakrabarti. 1999. 'Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations', *Journal of Molecular Biology*, 294: 271-88.

Palmer III, Arthur G. 2004. 'NMR characterization of the dynamics of biomacromolecules', *Chemical Reviews*, 104: 3623-40.

Palmer III, Arthur G, Christopher D Kroenke, and J Patrick Loria. 2001. 'Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules', *Methods in enzymology*, 339: 204-38.

Pande, Vijay S, Alexander Yu Grosberg, Toyoichi Tanaka, and Daniel S Rokhsar. 1998. 'Pathways for protein folding: is a new view needed?', *Current Opinion in Structural Biology*, 8: 68-79.

Pappenberger, Günter, Hüseyin Aygün, Joachim W Engels, Ulf Reimer, Gunter Fischer, and Thomas Kiefhaber. 2001. 'Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics', *Nature Structural Biology*, 8: 452-58.

Pappenberger, Günter, Annett Bachmann, Rita Müller, Hüseyin Aygün, Joachim W Engels, and Thomas Kiefhaber. 2003. 'Kinetic mechanism and catalysis of a native-state prolyl isomerization reaction', *Journal of Molecular Biology*, 326: 235-46.

Parker, Martin J, Christopher E Dempsey, Mark Lorch, and Anthony R Clarke. 1997. 'Acquisition of native β-strand topology during the rapid collapse phase of protein folding', *Biochemistry*, 36: 13396-405.

Parker, Martin J, and Susan Marqusee. 1999. 'The cooperativity of burst phase reactions explored', *Journal of Molecular Biology*, 293: 1195-210.

Patra, Ashish K, and Jayant B Udgaonkar. 2007. 'Characterization of the folding and unfolding reactions of single-chain monellin: evidence for multiple intermediates and competing pathways', *Biochemistry*, 46: 11727-43.

Peon, Jorge, Samir Kumar Pal, and Ahmed H Zewail. 2002. 'Hydration at the surface of the protein Monellin: dynamics with femtosecond resolution', *Proceedings of the National Academy of Sciences*, 99: 10964-69.

Peran, Ivan, Alex S Holehouse, Isaac S Carrico, Rohit V Pappu, Osman Bilsel, and Daniel P Raleigh. 2019. 'Unfolded states under folding conditions accommodate sequence-specific

conformational preferences with random coil-like dimensions', *Proceedings of the National Academy of Sciences*, 116: 12301-10.

Perham, Richard N 2000. 'Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions', *Annual review of biochemistry*, 69: 961-1004.

Pertinhez, Thelma A, Daizo Hamada, Lorna J Smith, Fabrizio Chiti, Niccoló Taddei, Massimo Stefani, and Christopher M Dobson. 2000. 'Initial denaturing conditions influence the slow folding phase of acylphosphatase associated with proline isomerization', *Protein Science*, 9: 1466-73.

Philips, Laura A, SP Webb, Selso J Martinez, GR Fleming, and Donald H Levy. 1988. 'Time-resolved spectroscopy of tryptophan conformers in a supersonic jet', *Journal of the American Chemical Society*, 110: 1352-55.

Pierce, Michael M, and Barry T Nall. 2000. 'Coupled kinetic traps in cytochrome c folding: His-heme misligation and proline isomerization', *Journal of Molecular Biology*, 298: 955-69.

Pirchi, Menahem, Guy Ziv, Inbal Riven, Sharona Sedghani Cohen, Nir Zohar, Yoav Barak, and Gilad Haran. 2011. 'Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein', *Nature communications*, 2: 493.

Plaxco, Kevin W, Ian S Millett, Daniel J Segel, Sebastian Doniach, and David Baker. 1999. 'Chain collapse can occur concomitantly with the rate-limiting step in protein folding', *Nature Structural Biology*, 6: 554-56.

Plaxco, Kevin W, Kim T Simons, and David Baker. 1998. 'Contact order, transition state placement and the refolding rates of single domain proteins', *Journal of Molecular Biology*, 277: 985-94.

Plaxco, Kevin W, Claus Spitzfaden, Iain D Campbell, and Christopher M Dobson. 1996. 'Rapid refolding of a proline-rich all-beta-sheet fibronectin type III module', *Proceedings of the National Academy of Sciences*, 93: 10703-06.

Pletneva, Ekaterina V, Harry B Gray, and Jay R Winkler. 2005. 'Snapshots of cytochrome c folding', *Proceedings of the National Academy of Sciences*, 102: 18397-402.

Portman, John J, Shoji Takada, and Peter G Wolynes. 2001. 'Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics', *The Journal of Chemical Physics*, 114: 5082-96.

Pradeep, Lovy, and Jayant B Udgaonkar. 2002. 'Differential salt-induced stabilization of structure in the initial folding intermediate ensemble of barstar', *Journal of Molecular Biology*, 324: 331-47.

Pradeep, Lovy, and Jayant B Udgaonkar. 2004a. 'Effect of salt on the urea-unfolded form of barstar probed by m value measurements', *Biochemistry* 43: 11393-402.

Pradeep, Lovy, and Jayant B Udgaonkar. 2004b. 'Osmolytes induce structure in an early intermediate on the folding pathway of barstar', *Journal of Biological Chemistry*, 279: 40303-13.

Privalov, Peter L 1979. 'Stability of proteins small globular proteins', *J Advances in protein chemistry*, 33: 167-241.

Ptitsyn, OB. 1973. 'Stages in the mechanism of self-organization of protein molecules', *Doklady Akademii Nauk SSSR*, 210: 1213-15.

Ptitsyn, OB, Roger H Pain, GV Semisotnov, E Zerovnik, and OI Razgulyaev. 1990. 'Evidence for a molten globule state as a general intermediate in protein folding', *FEBS letters*, 262: 20-24.

Ptitsyn, Oleg B 1996. 'A determinable but unresolved problem', *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 10: 3-4.

Ptitsyn, Oleg B, and Vladimir N Uversky. 1994. 'The molten globule is a third thermodynamical state of protein molecules', *FEBS letters*, 341: 15-18.

Qi, Phoebe X, Tobin R Sosnick, and S Walter Englander. 1998. 'The burst phase in ribonuclease A folding and solvent dependence of the unfolded state', *Nature Structural Biology*, 5: 882-84.

Qiu, Linlin, Cherian Zachariah, and Stephen J Hagen. 2003. 'Fast chain contraction during protein folding:“foldability” and collapse dynamics', *Physical review letters*, 90: 168103.

Radford, Sheena E, Christopher M Dobson, and Philip A Evans. 1992. 'The folding of hen lysozyme involves partially structured intermediates and multiple pathways', *Nature*, 358: 302-07.

Radou, Gaël, Marta Enciso, Sergei Krivov, and Emanuele Paci. 2013. 'Modulation of a protein free-energy landscape by circular permutation', *The Journal of Physical Chemistry B*, 117: 13743-47.

Rami, Bhadresh R, G Krishnamoorthy, and Jayant B Udgaonkar. 2003. 'Dynamics of the core tryptophan during the formation of a productive molten globule intermediate of barstar', *Biochemistry*, 42: 7986-8000.

Rami, Bhadresh R, and Jayant B Udgaonkar. 2002. 'Mechanism of formation of a productive molten globule form of barstar', *Biochemistry*, 41: 1710-16.

Raschke, Tanya M, and Susan Marqusee. 1997. 'The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions', *Nature Structural Biology*, 4: 298-304.

Ratner, V, D Amir, E Kahana, and E Haas. 2005. 'Fast collapse but slow formation of secondary structure elements in the refolding transition of E. coli adenylate kinase', *Journal of Molecular Biology*, 352: 683-99.

Reddy, Govardhan, and D Thirumalai. 2017. 'Collapse precedes folding in denaturant-dependent assembly of ubiquitin', *The Journal of Physical Chemistry B*, 121: 995-1009.

Reimer, Ulf, Gerd Scherer, Mario Drewello, Susanne Kruber, Mike Schutkowski, and Gunter Fischer. 1998. 'Side-chain effects on peptidyl-prolyl cis/trans isomerisation', *Journal of Molecular Biology*, 279: 449-60.

Riback, Joshua A, Micayla A Bowman, Adam M Zmyslowski, Catherine R Knoverek, John M Jumper, James R Hinshaw, Emily B Kaye, Karl F Freed, Patricia L Clark, and Tobin R Sosnick. 2017. 'Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water', *Science*, 358: 238-41.

Riback, Joshua A, Micayla A Bowman, Adam M Zmyslowski, Kevin W Plaxco, Patricia L Clark, and Tobin R Sosnick. 2019. 'Commonly used FRET fluorophores promote collapse of an otherwise disordered protein', *Proceedings of the National Academy of Sciences*, 116: 8889-94.

Ries, Julia, Simone Schwarze, Christopher M Johnson, and Hannes Neuweiler. 2014. 'Microsecond folding and domain motions of a spider silk protein structural switch', *Journal of the American Chemical Society*, 136: 17136-44.

Roder, Heinrich 2004. 'Stepwise helix formation and chain compaction during protein folding', *Proceedings of the National Academy of Sciences*, 101: 1793-94.

Roder, Heinrich, Gülnur A Elöve, and S Walter Englander. 1988. 'Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR', *Nature*, 335: 700-04.

Roderer, Daniel JA, Martin A Schärer, Marina Rubini, and Rudi Glockshuber. 2015. 'Acceleration of protein folding by four orders of magnitude through a single amino acid substitution', *Scientific Reports*, 5: 1-16.

Roe, Daniel R, Viktor Hornak, and Carlos Simmerling. 2005. 'Folding cooperativity in a three-stranded β-sheet model', *Journal of Molecular Biology*, 352: 370-81.

Rogers, Joseph M, Chi T Wong, and Jane Clarke. 2014. 'Coupled folding and binding of the disordered protein PUMA does not require particular residual structure', *Journal of the American Chemical Society*, 136: 5197-200.

Roh, Joon Ho, Liang Guo, J Duncan Kilburn, Robert M Briber, Thomas Irving, and Sarah A Woodson. 2010. 'Multistage collapse of a bacterial ribozyme observed by time-resolved small-angle X-ray scattering', *Journal of the American Chemical Society*, 132: 10148-54.

Rose, George D, Patrick J Fleming, Jayanth R Banavar, and Amos Maritan. 2006. 'A backbone-based theory of protein folding', *Proceedings of the National Academy of Sciences*, 103: 16623-33.

Rubinstein, Michael, and Ralph H Colby. 2003. *Polymer physics* (Oxford university press).

Ruff, Kiersten M, and Alex S Holehouse. 2017. 'SAXS versus FRET: A matter of heterogeneity?', *Biophysical Journal*, 113: 971-73.

Sabelko, J, J Ervin, and Martin Gruebele. 1999. 'Observation of strange kinetics in protein folding', *Proceedings of the National Academy of Sciences*, 96: 6031-36.

Sadqi, Mourad, David Fushman, and Victor Munoz. 2006. 'Atom-by-atom analysis of global downhill protein folding', *Nature*, 442: 317-21.

Saito, Masataka, Supawich Kamonprasertsuk, Satomi Suzuki, Kei Nanatani, Hiroyuki Oikawa, Keiichiro Kushiro, Madoka Takai, Po-ting Chen, Eric H-L Chen, and Rita P-Y Chen. 2016. 'Significant heterogeneity and slow dynamics of the unfolded ubiquitin detected by the line confocal method of single-molecule fluorescence spectroscopy', *The Journal of Physical Chemistry B*, 120: 8818-29.

Sánchez, Ignacio E, Diego U Ferreiro, and Gonzalo de Prat Gay. 2011. 'Mutational analysis of kinetic partitioning in protein folding and protein–DNA binding', *Protein Engineering, Design Selection*, 24: 179-84.

Sánchez, Susana A, Juan E Brunet, David M Jameson, Rosalba Lagos, and Octavio Monasterio. 2004. 'Tubulin equilibrium unfolding followed by time-resolved fluorescence and fluorescence correlation spectroscopy', *Protein Science*, 13: 81-88.

Sarkar, Saswata Sankar, Jayant B Udgaonkar, and Guruswamy Krishnamoorthy. 2013. 'Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state', *Biophysical Journal*, 105: 2392-402.

Saxena, Anoop M, Jayant B Udgaonkar, and G Krishnamoorthy. 2006. 'Characterization of intra-molecular distances and site-specific dynamics in chemically unfolded barstar: Evidence for denaturant-dependent non-random structure', *Journal of Molecular Biology*, 359: 174-89.

Sborgi, Lorenzo, Abhinav Verma, Stefano Piana, Kresten Lindorff-Larsen, Michele Cerminara, Clara M Santiveri, David E Shaw, Eva de Alba, and Victor Munoz. 2015. 'Interaction networks in protein folding via atomic-resolution experiments and long-time-scale molecular dynamics simulations', *Journal of the American Chemical Society*, 137: 6506-16.

Schachman, HK 1988. 'Can a simple model account for the allosteric transition of aspartate transcarbamoylase?', *Journal of Biological Chemistry*, 263: 18583-86.

Schmid, Franz X, and Robert L Baldwin. 1979. 'The rate of interconversion between the two unfolded forms of ribonuclease A does not depend on guanidinium chloride concentration', *Journal of Molecular Biology*, 133: 285-87.

Schmid, Franz X, Reinhard Grafl, Alex Wrba, and Jaap J Beintema. 1986. 'Role of proline peptide bond isomerization in unfolding and refolding of ribonuclease', *Proceedings of the National Academy of Sciences* 83: 872-76.

Schmidpeter, Philipp AM, and Franz X Schmid. 2015. 'Prolyl isomerization and its catalysis in protein folding and protein function', *Journal of Molecular Biology*, 427: 1609-31.

Schönbrunner, Nancy, Klaus-Peter Koller, and Thomas Kiefhaber. 1997. 'Folding of the disulfide-bonded β-sheet protein tendamistat: rapid two-state folding without hydrophobic collapse', *Journal of Molecular Biology*, 268: 526-38.

Schreiber, Gideon, and Alan R Fersht. 1993. 'The refolding of cis-and trans-peptidylprolyl isomers of barstar', *Biochemistry*, 32: 11195-203.

Schuler, Benjamin. 2007. 'Application of single molecule Förster resonance energy transfer to protein folding.' in, *Protein Folding Protocols* (Springer).

Schuler, Benjamin, and William A Eaton. 2008. 'Protein folding studied by single-molecule FRET', *Current Opinion in Structural Biology*, 18: 16-26.

Schuler, Benjamin, Everett A Lipman, and William A Eaton. 2002. 'Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy', *Nature*, 419: 743-47.

Schuler, Benjamin, Andrea Soranno, Hagen Hofmann, and Daniel Nettels. 2016. 'Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins', *Annual Review of Biophysics*, 45: 207-31.

Schwalbe, Harald, Klaus M Fiebig, Matthias Buck, Jonathan A Jones, Shaun B Grimshaw, Andrew Spencer, Steffen J Glaser, Lorna J Smith, and Christopher M Dobson. 1997. 'Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea', *Biochemistry*, 36: 8977-91.

Scott, Kathryn A, Darwin OV Alonso, Yongping Pan, and Valerie Daggett. 2006. 'Importance of context in protein folding: secondary structural propensities versus tertiary contact-assisted secondary structure formation', *Biochemistry*, 45: 4153-63.

Serrano, Arnaldo L, Osman Bilsel, and Feng Gai. 2012. 'Native state conformational heterogeneity of HP35 revealed by time-resolved FRET', *The Journal of Physical Chemistry B*, 116: 10631-38.

Shank, Elizabeth A, Ciro Cecconi, Jesse W Dill, Susan Marqusee, and Carlos Bustamante. 2010. 'The folding cooperativity of a protein is controlled by its chain topology', *Nature*, 465: 637-40.

Shastry, MC Ramachandra, and Heinrich Roder. 1998. 'Evidence for barrier-limited protein folding kinetics on the microsecond time scale', *Nature Structural Biology*, 5: 385-92.

Shastry, MCR, Vishwas R Agashe, and Jayant B Udgaonkar. 1994. 'Quantitative analysis of the kinetics of denaturation and renaturation of barstar in the folding transition zone', *Protein Science*, 3: 1409-17.

Shastry, MCR, and Jayant B Udgaonkar. 1995. 'The folding mechanism of barstar: evidence for multiple pathways and multiple intermediates', *Journal of Molecular Biology*, 247: 1013-27.

Sherman, Eilon, and Gilad Haran. 2006. 'Coil–globule transition in the denatured state of a small protein', *Proceedings of the National Academy of Sciences*, 103: 11539-43.

Sherman, Eilon, and Gilad Haran. 2011. 'Fluorescence correlation spectroscopy of fast chain dynamics within denatured protein L', *ChemPhysChem*, 12: 696-703.

Shortle, David, and Michael S Ackerman. 2001. 'Persistence of native-like topology in a denatured protein in 8 M urea', *Science*, 293: 487-89.

Sinha, Kalyan K, and Jayant B Udgaonkar. 2009. 'Early events in protein folding', *Current Science*: 1053-70.

Sinha, Kalyan K, and Jayant B Udgaonkar. 2005. 'Dependence of the size of the initially collapsed form during the refolding of barstar on denaturant concentration: evidence for a continuous transition', *Journal of Molecular Biology*, 353: 704-18.

Sinha, Kalyan K, and Jayant B Udgaonkar. 2007. 'Dissecting the non-specific and specific components of the initial folding reaction of barstar by multi-site FRET measurements', *Journal of molecular biology* 370: 385-405.

Sinha, Kalyan K, and Jayant B Udgaonkar. 2008. 'Barrierless evolution of structure during the submillisecond refolding reaction of a small protein', *Proceedings of the National Academy of Sciences*, 105: 7998-8003.

Skinner, John J, Woon K Lim, Sabrina Bédard, Ben E Black, and S Walter Englander. 2012. 'Protein dynamics viewed by hydrogen exchange', *Protein Science*, 21: 996-1005.

Song, Jianhui, Gregory-Neal Gomes, Tongfei Shi, Claudiu C Gradinaru, and Hue Sun Chan. 2017. 'Conformational heterogeneity and FRET data interpretation for dimensions of unfolded proteins', *Biophysical Journal*, 113: 1012-24.

Song, Jianxing, Nadège Jamin, Bernard Gilquin, Claudio Vita, and André Ménez. 1999. 'A gradual disruption of tight side–chain packing: 2D 1H–NMR characterization of acid–induced unfolding of CHABII', *Nature Structural Biology*, 6: 129-34.

Soranno, Andrea, Andrea Holla, Fabian Dingfelder, Daniel Nettels, Dmitrii E Makarov, and Benjamin Schuler. 2017. 'Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations', *Biophysical Journal*, 112: 318a.

Soranno, Andrea, Renato Longhi, Tommaso Bellini, and Marco Buscaglia. 2009. 'Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides', *Biophysical Journal*, 96: 1515-28.

Sosnick, Tobin R 2011. 'The folding of single domain proteins-have we reached a consensus?', *Biophysical Journal*, 100: 373a.

Sosnick, TR, L Mayne, R Hiller, and SW Englander. 1994. 'The barriers in protein folding', *Nature Structural Biology*, 1: 149-56.

Spudich, Giulietta M, Erik J Miller, and Susan Marqusee. 2004. 'Destabilization of the Escherichia coli RNase H kinetic intermediate: switching between a two-state and three-state folding mechanism', *Journal of Molecular Biology*, 335: 609-18.

Sridevi, K, GS Lakshmikanth, G Krishnamoorthy, and Jayant B Udgaonkar. 2004. 'Increasing stability reduces conformational heterogeneity in a protein folding intermediate ensemble', *Journal of Molecular Biology*, 337: 699-711.

Srivastava, Devesh, and Murugappan Muthukumar. 1996. 'Sequence dependence of conformations of polyampholytes', *Macromolecules*, 29: 2324-26.

Steinmetz, Michel O, Ilian Jelesarov, William M Matousek, Srinivas Honnappa, Wolfgang Jahnke, John H Missimer, Sabine Frank, Andrei T Alexandrescu, and Richard A Kammerer. 2007. 'Molecular basis of coiled-coil formation', *Proceedings of the National Academy of Sciences*, 104: 7062-67.

Stewart, David E, Atom Sarkar, and John E Wampler. 1990. 'Occurrence and role ofcis peptide bonds in protein structures', *Journal of Molecular Biology*, 214: 253-60.

Sugase, Kenji, H Jane Dyson, and Peter E Wright. 2007. 'Mechanism of coupled folding and binding of an intrinsically disordered protein', *Nature*, 447: 1021-25.

Sung, Yoon-Hui, Hee-Deok Hong, Chaejoon Cheong, Jin Hwan Kim, Joong Myung Cho, Yong-Rok Kim, and Weontae Lee. 2001. 'Folding and stability of sweet protein single-chain monellin: An insight to protein engineering', *Journal of Biological Chemistry*, 276: 44229-38.

Svensson, L Anders, Eva Thulin, and Sture Forsén. 1992. 'Proline cis-trans isomers in calbindin D9k observed by X-ray crystallography', *Journal of Molecular Biology*, 223: 601-06.

Swaminathan, R, G Krishnamoorthy, and N Periasamy. 1994. 'Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state', *Biophysical Journal*, 67: 2013-23.

Swaminathan, R, Utpal Nath, Jayant B Udgaonkar, N Periasamy, and G Krishnamoorthy. 1996. 'Motional dynamics of a buried tryptophan reveals the presence of partially structured forms during denaturation of barstar', *Biochemistry*, 35: 9150-57.

Takahashi, Satoshi, Aya Yoshida, and Hiroyuki Oikawa. 2018. 'Hypothesis: structural heterogeneity of the unfolded proteins originating from the coupling of the local clusters and the long-range distance distribution', *Biophysical Reviews*, 10: 363-73.

Taler-Verčič, Ajda, Samra Hasanbašić, Selma Berbić, Veronika Stoka, Dušan Turk, and Eva Žerovnik. 2017. 'Proline residues as switches in conformational changes leading to amyloid fibril formation', *International journal of molecular sciences*, 18: 549.

Teufel, Daniel P, Christopher M Johnson, Jenifer K Lum, and Hannes Neuweiler. 2011. 'Backbone-driven collapse in unfolded protein chains', *Journal of molecular biology*, 409: 250-62.

Texter, Frieda L, Donald B Spencer, Richard Rosenstein, and C Robert Matthews. 1992. 'Intramolecular catalysis of a proline isomerization reaction in the folding of dihydrofolate reductase', *Biochemistry*, 31: 5687-91.

Thirumalai, D, DK Klimov, and SA Woodson. 1997. 'Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules', *Theoretical Chemistry Accounts*, 96: 14-22.

Thirumalai, Dave, Himadri S Samanta, Hiranmay Maity, and Govardhan Reddy. 2019. 'Universal nature of collapsibility in the context of protein folding and evolution', *Trends in Biochemical Sciences*, 44: 675-87.

Tomishige, Michio, Dieter R Klopfenstein, and Ronald D Vale. 2002. 'Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization', *Science*, 297: 2263-67.

Tsai, Chung-Jung, Dong Xu, and Ruth Nussinov. 1998. 'Protein folding via binding and vice versa', *Folding Design*, 3: R71-R80.

Tweedy, Neil B, Satish K Nair, Steven A Paterno, Carol A Fierke, and David W Christianson. 1993. 'Structure and energetics of a non-proline cis-peptidyl linkage in a proline-202. fwdarw. alanine carbonic anhydrase II variant', *Biochemistry*, 32: 10944-49.

Udgaonkar, Jayant B 2008. 'Multiple routes and structural heterogeneity in protein folding', *Annu. Rev. Biophys.*, 37: 489-510.

Udgaonkar, Jayant B 2013. 'Polypeptide chain collapse and protein folding', *Archives of biochemistry biophysics* 531: 24-33.

Udgaonkar, Jayant B, and Robert L Baldwin. 1988. 'NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A', *Nature*, 335: 694-99.

Uzawa, Takanori, Shuji Akiyama, Tetsunari Kimura, Satoshi Takahashi, Koichiro Ishimori, Isao Morishima, and Tetsuro Fujisawa. 2004. 'Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness', *Proceedings of the National Academy of Sciences*, 101: 1171-76.

Vanhove, Marc, Xavier Raquet, Timothy Palzkill, Roger H Pain, and Jean-Marie Frère. 1996. 'The rate-limiting step in the folding of the cis-Pro167Thr mutant of TEM-1 β-lactamase is the trans to cis isomerization of a non-proline peptide bond', *Proteins: Structure, Function,Bioinformatics* 25: 104-11.

Veeraraghavan, Sudha, Barry T Nall, and Anthony L Fink. 1997. 'Effect of prolyl isomerase on the folding reactions of staphylococcal nuclease', *Biochemistry*, 36: 15134-39.

Viguera, Ana-Rosa, and Luis Serrano. 1997. 'Loop length, intramolecular diffusion and protein folding', *Nature Structural Biology*, 4: 939-46.

Voelz, Vincent A, Vijay R Singh, William J Wedemeyer, Lisa J Lapidus, and Vijay S Pande. 2010. 'Unfolded-state dynamics and structure of protein L characterized by simulation and experiment', *Journal of the American Chemical Society*, 132: 4702-09.

Vu, Dung M, Scott H Brewer, and R Brian Dyer. 2012. 'Early turn formation and chain collapse drive fast folding of the major cold shock protein CspA of Escherichia coli', *Biochemistry*, 51: 9104-11.

Waldauer, Steven A, Olgica Bakajin, Terry Ball, Yujie Chen, Stephen J DeCamp, Michaela Kopka, Marcus Jäger, Vijay R Singh, William J Wedemeyer, and Shimon Weiss. 2008. 'Ruggedness in the folding landscape of protein L', *HFSP journal*, 2: 388-95.

Waldauer, Steven A, Olgica Bakajin, and Lisa J Lapidus. 2010. 'Extremely slow intramolecular diffusion in unfolded protein L', *Proceedings of the National Academy of Sciences*, 107: 13713-17.

Walkenhorst, William F, Susan M Green, and Heinrich Roder. 1997. 'Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease', *Biochemistry*, 36: 5795-805.

Wallace, Louise A, and C Robert Matthews. 2002. 'Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions', *Biophysical chemistry*, 101: 113-31.

Wang, X, EN Bodunov, and WM Nau. 2003. 'Fluorescence quenching kinetics in short polymer chains: dependence on chain length', *Optics Spectroscopy* 95: 560-70.

Wani, Ajazul Hamid, and Jayant B Udgaonkar. 2009. 'Native state dynamics drive the unfolding of the SH3 domain of PI3 kinase at high denaturant concentration', *Proceedings of the National Academy of Sciences*, 106: 20711-16.

Waudby, Christopher A, Tomasz Wlodarski, Maria-Evangelia Karyadi, Anaïs ME Cassaignau, Sammy HS Chan, Anne S Wentink, Julian M Schmidt-Engler, Carlo Camilloni, Michele Vendruscolo, and Lisa D Cabrita. 2018. 'Systematic mapping of free energy landscapes of a growing filamin domain during biosynthesis', *Proceedings of the National Academy of Sciences*, 115: 9744-49.

Wedemeyer, William J, Ervin Welker, and Harold A Scheraga. 2002. 'Proline cis− trans isomerization and protein folding', *Biochemistry*, 41: 14637-44.

Weissman, Jonathan S 1995. 'All roads lead to Rome? The multiple pathways of protein folding', *Chemistry biology* 2: 255-60.

Weissman, Jonathan S, and Peter S Kim. 1991. 'Reexamination of the folding of BPTI: predominance of native intermediates', *Science*, 253: 1386-93.

Weissman, Jonathan S, and Peter S Kim. 1992. 'Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor', *Proceedings of the National Academy of Sciences*, 89: 9900-04.

Welker, Ervin, Kosuke Maki, MC Ramachandra Shastry, Darmawi Juminaga, Rajiv Bhat, Harold A Scheraga, and Heinrich Roder. 2004. 'Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A', *Proceedings of the National Academy of Sciences*, 101: 17681-86.

Welker, Ervin, William J Wedemeyer, Mahesh Narayan, and Harold A Scheraga. 2001. 'Coupling of conformational folding and disulfide-bond reactions in oxidative folding of proteins', *Biochemistry*, 40: 9059-64.

Wetlaufer, Donald B 1973. 'Nucleation, rapid folding, and globular intrachain regions in proteins', *Proceedings of the National Academy of Sciences*, 70: 697-701.

Wirmer, Julia, Holger Berk, Raffaella Ugolini, Christina Redfield, and Harald Schwalbe. 2006. 'Characterization of the unfolded state of bovine α-lactalbumin and comparison with unfolded states of homologous proteins', *Protein Science*, 15: 1397-407.

Wodak, Shoshana J, Anatoly Malevanets, and Stephen S MacKinnon. 2015. 'The landscape of intertwined associations in homooligomeric proteins', *Biophysical Journal*, 109: 1087-100.

Wolynes, Peter G, Jose N Onuchic, and Dave Thirumalai. 1995. 'Navigating the folding routes', *Science*, 267: 1619-20.

Woody, Robert W. 2004. 'Circular dichroism of protein-folding intermediates.' in, *Methods in enzymology* (Elsevier).

Woody, Robert W 1978. 'Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins', *Biopolymers: Original Research on Biomolecules*, 17: 1451-67.

Wright, Caroline F, Kresten Lindorff-Larsen, Lucy G Randles, and Jane Clarke. 2003. 'Parallel protein-unfolding pathways revealed and mapped', *Nature Structural Molecular Biology* 10: 658-62.

Wu, Ying, Elena Kondrashkina, Can Kayatekin, C Robert Matthews, and Osman Bilsel. 2008. 'Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein', *Proceedings of the National Academy of Sciences*, 105: 13367-72.

Wu, Ying, and C Robert Matthews. 2002a. 'A cis-prolyl peptide bond isomerization dominates the folding of the alpha subunit of Trp synthase, a TIM barrel protein', *Journal of molecular biology*, 322: 7-13.

Wu, Ying, and C Robert Matthews. 2002b. 'Parallel channels and rate-limiting steps in complex protein folding reactions: prolyl isomerization and the alpha subunit of Trp synthase, a TIM barrel protein', *Journal of molecular biology*, 323: 309-25.

Wu, Ying, and C Robert Matthews. 2003. 'Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of Trp synthase, a TIM barrel protein', *Journal of molecular biology*, 330: 1131-44.

Yagi-Utsumi, Maho, Mahesh S Chandak, Saeko Yanaka, Methanee Hiranyakorn, Takashi Nakamura, Koichi Kato, and Kunihiro Kuwajima. 2020. 'Residual structure of unfolded ubiquitin as revealed by hydrogen/deuterium-exchange 2D NMR', *Biophysical Journal*, 119: 2029-38.

Yamada, Seiji, Nicole D Bouley Ford, Gretchen E Keller, William C Ford, Harry B Gray, and Jay R Winkler. 2013. 'Snapshots of a protein folding intermediate', *Proceedings of the National Academy of Sciences*, 110: 1606-10.

Yang, Wei Yuan, and Martin Gruebele. 2004. 'Detection-dependent kinetics as a probe of folding landscape microstructure', *Journal of the American Chemical Society*, 126: 7758-59.

Yengo, Christopher M, and Christopher L Berger. 2010. 'Fluorescence anisotropy and resonance energy transfer: powerful tools for measuring real time protein dynamics in a physiological environment', *Current opinion in pharmacology*, 10: 731-37.

Yoo, Tae Yeon, Steve P Meisburger, James Hinshaw, Lois Pollack, Gilad Haran, Tobin R Sosnick, and Kevin Plaxco. 2012. 'Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state', *Journal of Molecular Biology*, 418: 226-36.

Zaidi, Faisal N, Utpal Nath, and Jayant B Udgaonkar. 1997. 'Multiple intermediates and transition states during protein unfolding', *Nature Structural Biology*, 4: 1016-24.

Zeeb, Markus, Georg Lipps, Hauke Lilie, and Jochen Balbach. 2004. 'Folding and association of an extremely stable dimeric protein from Sulfolobus islandicus', *Journal of Molecular Biology*, 336: 227-40.

Zheng, Wenwei, Alessandro Borgia, Karin Buholzer, Alexander Grishaev, Benjamin Schuler, and Robert B Best. 2016. 'Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment', *Journal of the American Chemical Society*, 138: 11702-13.

Zhou, Huan-Xiang 2002. 'Dimensions of denatured protein chains from hydrodynamic data', *The Journal of Physical Chemistry B*, 106: 5769-75.

Zhou, Huan-Xiang, and Ken A Dill. 2001. 'Stabilization of proteins in confined spaces', *Biochemistry*, 40: 11289-93.

Ziv, Guy, D Thirumalai, and Gilad Haran. 2009. 'Collapse transition in proteins', *Physical Chemistry Chemical Physics*, 11: 83-93.