

**Characterization of the Nature and Role of the Dry  
Molten Globule-like Intermediate States During Protein  
Folding and Misfolding Reactions**

Thesis submitted to AcSIR

For the Award of the Degree of

**DOCTOR OF PHILOSOPHY**

In

**BIOLOGICAL SCIENCES**



By

**NIRBHIK ACHARYA**

Registration Number: 10BB14A26044

Under the guidance of

Dr. Santosh Kumar Jha

Physical and Materials Chemistry Division

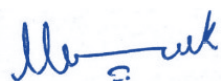
CSIR- National Chemical Laboratory

Pune-411008, India

## Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, “*Characterization of the Nature and Role of the Dry Molten Globule-like Intermediate States During Protein Folding and Misfolding Reactions*”, submitted by Mr. Nirbhik Acharya to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of *Doctor of Philosophy in Biological Sciences*, embodies original research work under my guidance. We, further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research materials obtained from other sources have been duly acknowledged in the thesis. Images, illustrations, figures, tables etc., used in the thesis from other sources, have also been duly cited and acknowledged.

It is also certified that this work done by the student under my supervision is plagiarism free.



Nirbhik Acharya

Date: 06/08/2020

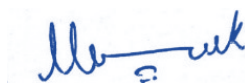


Santosh Kumar Jha

Date: 06/08/2020

## Declaration

I, Nirbhik Acharya, hereby declare that the work incorporated in the thesis and entitled “*Characterization of the Nature and Role of the Dry Molten Globule-like Intermediate States During Protein Folding and Misfolding Reactions*” submitted for the award of the Degree of Doctor of Philosophy in Biological Sciences to the Academy of Scientific and Innovative Research (AcSIR), New Delhi, has been carried out by me at Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory, Pune-411008, India, under the supervision of Dr. Santosh Kumar Jha. The work is original and has not been submitted as a part or full by me for any degree or diploma to this or any other university. I further declare that the material obtained from resources has been duly acknowledged in this thesis.



**Nirbhik Acharya**

Date: 06/08/2020

Place: Pune

## Acknowledgements

*The work presented in this thesis would not have been possible without my close association with many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who made this PhD thesis possible.*

*First and foremost, I would like to extend my sincere gratitude to my research guide Dr. Santosh Kumar Jha, for introducing me to the exciting field of protein folding and for his dedicated help, advice, inspiration, encouragement, and continuous support throughout my PhD. He has been supportive since the day I started my PhD journey in his lab. Ever since, Santosh has supported me academically and emotionally through all the ups and downs to finish this thesis. I am grateful for being the first student in Santosh's lab, where I have grown along with the lab while being treated as a colleague, not as a student. I owe a lot of gratitude to him for giving me the freedom to discuss scientific and non-scientific topics, freedom to come up with new problems, listening to smallest of my insights and for encouraging me to learn new things. His enthusiasm and integral view on research has made a deep impression on me. During our course of interaction in the last six years, I have learned extensively from him, including how to approach a problem by systematic thinking and data-driven decision making, how to raise new possibilities, how to regard an old question from a new perspective. I am really glad to be associated with a mentor like Santosh.*

*My special words of thanks should also go to my thesis committee members Dr. Mugdha Gadgil, Dr. Sayan Bagchi and Dr. Dhanasekaran Shanmugam for their continuous support, guidance, cooperation, and encouragement through all these years.*

*I am grateful to Dr. Jayant Udgaonkar from IISER, Pune, who gave me permission to work in his laboratory and use the facilities available there for mass spectrometry.*

*My heartfelt thanks to all the members of Jha lab: Anjali, Prajna, Meenakshi, Divya, Abhilasha, Sonal, Tejas, Anvesha, and Avishkar. A special thanks to: Anjali for always pushing me to achieve my goals on time, Prajna for always being available for listening silliest of my queries and for all the constructive arguments and scientific discussions. I thank them all for always standing by my side and sharing a great relationship as compassionate friends. I will always cherish the moments we have spent during all our lab trips, especially the Goa trip!*

*A special mention of thanks to my friends in Pune: Prajna, Ruchir, Parag, Priya, Kushal, Monika, Debjyoti, and Singam for their constant support and cooperation during the last six years. I will always remember Ruchir and Parag's efforts in keeping my spirits high.*

*I owe my deepest gratitude towards my parents and elder brother Chit for their eternal support and understanding of my goals and aspirations. Their love and support have always been my strength. Without their help, I would not have been able to complete much of what I have done and become who I am.*

*A special thanks to my dearest Dinesh, Ragini, and Charu for their love and affection.*

*I gratefully acknowledge Council of Scientific and Industrial Research for providing me financial support, Director, NCL for allowing me to carry out my research in this esteemed laboratory and AcSIR for my enrolment in the PhD program.*

*Finally, I am thankful to the library staff and administrative staff of NCL for cooperation.*

*Nirbhik Acharya*

## Table of Contents

| <b>S. No.</b>    | <b>Title</b>  | <b>Page No.</b> |
|------------------|---|-----------------|
|                  | <b>Table of Contents</b>  | i               |
|                  | <b>List of Figures</b>  | iv              |
|                  | <b>List of Tables</b>   | vii             |
|                  | <b>Synopsis of the Thesis</b>   | 1               |
| <hr/>            |   |                 |
| <b>Chapter 1</b> | <b>Dry Molten Globule-like Intermediates in Protein Folding and Misfolding Pathways</b>   |                 |
| 1.1              | Introduction  | 6               |
| 1.2              | Protein folding-unfolding and molten globule intermediates  | 7               |
| 1.3              | Evidence of DMG-like intermediates  | 9               |
| 1.4              | Significance of DMGs in protein folding and unfolding reaction  | 10              |
| 1.5              | DMGs in protein misfolding reaction?  | 12              |
| 1.6              | Conclusions and open questions  | 13              |
| 1.7              | References  | 14              |
| <hr/>            |   |                 |
| <b>Chapter 2</b> | <b>Evidence for Dry Molten Globule-like Domains in the pH-induced Equilibrium Folding Intermediate of a Multidomain Protein</b> |                 |
| 2.1              | Introduction  | 24              |
| 2.2              | Materials and Methodology   | 25              |
| 2.3              | Results and Discussion  | 34              |
| 2.3.1            | pH-induced structural transitions   | 34              |
| 2.3.2            | Structural expansion in the E form  | 36              |
| 2.3.3            | E form contain hydrophobic core   | 37              |
| 2.3.4            | E form retains N-like structural fluctuations and solvent accessibility   | 39              |
| 2.3.5            | E form retains native-like solvation dynamics   | 41              |
| 2.3.6            | Global structural comparison of the N and the E form  | 42              |
| 2.3.7            | vdW interactions contributes in the stability of the native state   | 44              |
| 2.4              | Conclusions   | 46              |
| 2.5              | References  | 47              |

---

**Chapter 3 A Dry Molten Globule-like Intermediate During the Base-induced Unfolding of a Multidomain Protein**

|       |  |    |
|-------|--|----|
| 3.1   | Introduction   | 55 |
| 3.2   | Materials and Methodology  | 56 |
| 3.3   | Results and Discussion   | 64 |
| 3.3.1 | Base-induced structural transition                                     | 64 |
| 3.3.2 | Structural expansion in the B form                                     | 65 |
| 3.3.3 | B form retains N-like hydrophobicity                                   | 68 |
| 3.3.4 | N-like structural fluctuations and solvent accessibility of the B form | 70 |
| 3.3.5 | B form retains N-like solvation dynamics                               | 71 |
| 3.3.6 | Comparison of the global structure of the N and the B form             | 73 |
| 3.3.7 | Comparison of thermodynamic stabilities of the N and the B form        | 75 |
| 3.4   | Conclusions  | 78 |
| 3.5   | References   | 79 |

---

**Chapter 4 Nucleation Dependent Aggregation of Nucleic Acid Binding Domains of TDP-43 Occur via Dry Molten Globule-like State**

|       |   |     |
|-------|---|-----|
| 4.1   | Introduction  | 86  |
| 4.2   | Materials and Methodology   | 87  |
| 4.3   | Results and Discussion  | 96  |
| 4.3.1 | TDP <sup>NBD</sup> is amyloidogenic under stress                                | 96  |
| 4.3.2 | Amyloid precursor A form has disrupted tertiary packing                         | 98  |
| 4.3.3 | Disruption of the tertiary structure is coupled with the structural expansion   | 100 |
| 4.3.4 | The expanded A form resembles a dry molten globule-like state                   | 101 |
| 4.3.5 | Amyloid fibril formation occurs via nucleation-elongation aggregation mechanism | 103 |
| 4.3.6 | Low-pH stress-induced TDPNBD aggregation is a three-step process                | 108 |
| 4.4   | Conclusions   | 109 |
| 4.5   | References  | 110 |

---

---

|                  |  |     |
|------------------|--|-----|
| <b>Chapter 5</b> | <b>Structural dynamics of the DMG-like amyloid precursor and the amyloid core of TDP<sup>NBD</sup> revealed by H/D exchange coupled to mass spectrometry</b> |     |
| 5.1              | Introduction   | 117 |
| 5.2              | Materials and Methodology  | 118 |
| 5.3              | Results and Discussion   | 122 |
| 5.3.1            | Elucidating the core of the amyloid fibril   | 122 |
| 5.3.2            | Conformational heterogeneity in amyloid fibril   | 127 |
| 5.3.3            | Amyloidogenicity of the amyloid precursor DMG form is a result of high structural dynamics   | 129 |
| 5.4              | Conclusions  | 141 |
| 5.5              | References   | 142 |
| <hr/>            |  |     |
| <b>Chapter 6</b> | <b>Conclusions and Future Directions</b>   |     |
| 6.1              | Purpose and findings   | 147 |
| 6.2              | Contributions to the field   | 149 |
| 6.3              | Limitations of the study   | 150 |
| 6.4              | Future directions  | 150 |
| <hr/>            |  |     |
|                  | <b>List of Publications</b>  | 152 |

---



## List of Figures

| Figure No.       | Title   | Page No. |
|------------------|---|----------|
| <b>Chapter 1</b> |   |          |
| Figure 1.1       | The stepwise folding of protein with wet molten globule (WMG) intermediate state  | 8        |
| Figure 1.2       | The multi-step folding of a protein   | 8        |
| Figure 1.3       | Understanding the initial steps and the nature of the initial structural transition during protein aggregation  | 13       |
| <b>Chapter 2</b> |   |          |
| Figure 2.1       | Calibration curve for the determination of concentration of HSA-IAEDANS using Bradford assay  | 26       |
| Figure 2.2       | The secondary structure and the thermodynamics of unfolding of HSA remain unperturbed upon IAEDANS labeling   | 27       |
| Figure 2.3       | Determination of the overlap integral, J, at pH 7.0 and at pH 2.2   | 30       |
| Figure 2.4       | FRET between W214 and tyrosine residues of HSA  | 31       |
| Figure 2.5       | Stern-Volmer plots for quenching of fluorescence of N-acetyl-L-tryptophanamide (NATA) by acrylamide   | 33       |
| Figure 2.6       | pH dependence of conformation of HSA and pH-induced expansion of HSA monitored by FRET  | 35       |
| Figure 2.7       | pH dependence of the solvation of hydrophobic core of domains I and II  | 38       |
| Figure 2.8       | Fluorescence intensity decay kinetics of C34-IAEDANS  | 40       |
| Figure 2.9       | Dependence of wavelength of maximum fluorescence emission of W214 on the wavelength of excitation as a function of pH   | 42       |
| Figure 2.10      | Changes in the secondary and the tertiary structure and the changes in the stability of HSA as a function of pH   | 43       |
| Figure 2.11      | Model energy diagram  | 45       |
| <b>Chapter 3</b> |   |          |
| Figure 3.1       | Determination of the overlap integral, J, in the N form and in the B form   | 59       |
| Figure 3.2       | Calibration curves for the determination of the molecular weight and the hydrodynamic radius of the N form and the B form from size exclusion chromatography. | 61       |
| Figure 3.3       | Fluorescence intensity decay kinetics of C34-IAEDANS  | 62       |

|                  |  |     |
|------------------|--|-----|
| Figure 3.4       | The base-induced N $\rightleftharpoons$ B structural transition is accompanied by a change in tertiary structure without alteration in secondary structure | 64  |
| Figure 3.5       | The B form has a larger inter-domain distance and is globally expanded compared to the N form  | 66  |
| Figure 3.6       | The B form retains N-like hydrophobicity in domain I and II  | 69  |
| Figure 3.7       | Effect of the change in excitation wavelength on the wavelength of maximum emission.   | 71  |
| Figure 3.8       | Dependence of wavelength of maximum emission of C34-IAEDANS and W214 on excitation wavelength  | 72  |
| Figure 3.9       | The global secondary structure and the tertiary structure of the B form is similar to the N form but their thermodynamic stability is different            | 74  |
| Figure 3.10      | Structural components of HSA   | 75  |
| Figure 3.11      | Comparison of the spectroscopic properties of the U form at pH 7 and pH 11   | 76  |
| Figure 3.12      | Model energy diagram showing the relative energies of the N form, the B form and the U form  | 77  |
| <b>Chapter 4</b> |  |     |
| Figure 4.1       | Domain arrangement of TDP-43   | 87  |
| Figure 4.2       | MaxEnt deconvoluted electrospray mass spectrum of TDP <sup>NBD</sup> and NBD <sup>Mut</sup> for purity analysis.   | 89  |
| Figure 4.3       | Structure of NBD <sup>Mut</sup>  | 92  |
| Figure 4.4       | Determination of overlap integral (J)  | 93  |
| Figure 4.5       | Low-pH stress-induced A form is amyloidogenic  | 97  |
| Figure 4.6       | Comparison of the distribution of the hydrodynamic radii of the N form and the A form measured by dynamic light scattering                                 | 98  |
| Figure 4.7       | The global secondary structure and the tertiary structure comparison of the N form and the A form measured by far-UV and near-UV circular dichroism        | 99  |
| Figure 4.8       | Urea-induced equilibrium unfolding transition of the N form and the A form   | 100 |
| Figure 4.9       | The intra-molecular distances between W172 and C198-TNB was monitored by FRET in the N form and the A form   | 101 |
| Figure 4.10      | The fluorescence emission spectra of W172 and C198-DANS  | 102 |
| Figure 4.11      | Comparison of the ANS dye fluorescence upon binding with the N form, A form and WMG form   | 103 |
| Figure 4.12      | ThT fluorescence monitored aggregation kinetics  | 104 |
| Figure 4.13      | Determination of critical aggregation concentration  | 105 |
| Figure 4.14      | Effect of seeding  | 106 |

|                  |  |     |
|------------------|--|-----|
| Figure 4.15      | The scaling exponent plot for the prediction of the plausible secondary aggregation pathway  | 107 |
| Figure 4.16      | Schematic model for the amyloid fibril formation of TDP <sup>NBD</sup>   | 108 |
| <b>Chapter 5</b> |  |     |
| Figure 5.1       | Schematic model for TDP <sup>NBD</sup> aggregation   | 117 |
| Figure 5.2       | Peptide mapping of TDP <sup>NBD</sup>  | 120 |
| Figure 5.3       | Mass spectra of a selected peptide fragment (124-131) during HDX reaction elucidate structural organization                                | 123 |
| Figure 5.4       | Mass spectra of selected peptide fragments in the N, DMG and the amyloid form  | 125 |
| Figure 5.5       | Extent of deuterium incorporation in different peptide fragments of TDP <sup>NBD</sup> in the N form, DMG form and the amyloid fibril form | 126 |
| Figure 5.6       | Structure of TDP <sup>NBD</sup> representing the peptide fragment level deuterium incorporation in the amyloid form                        | 127 |
| Figure 5.7       | Conformational heterogeneity in amyloid fibril form  | 128 |
| Figure 5.8       | EX1/EX2 hydrogen-deuterium exchange mechanism theoretical test for TDP <sup>NBD</sup> at pH 7.0 and 6.5                                    | 131 |
| Figure 5.9       | Hydrogen-deuterium exchange kinetics of TDP <sup>NBD</sup> peptide fragments at pH 6.5 and pH 7.0  | 132 |
| Figure 5.10      | Determination of the hydrogen-deuterium exchange regime  | 135 |
| Figure 5.11      | Hydrogen-deuterium exchange kinetics of TDP <sup>NBD</sup> peptide fragments at pH 3.0   | 136 |
| Figure 5.12      | Hydrogen-deuterium exchange kinetics of TDP <sup>NBD</sup> peptide fragments for the amyloid form at pH 3.0                                | 138 |
| Figure 5.13      | Heat-map of HDX kinetics derived fragment level protection factor ( $P_f$ ) compared in the N, DMG and the amyloid form                    | 141 |

## List of Tables

| Table No.        | Title  | Page No. |
|------------------|--|----------|
| <b>Chapter 2</b> |  |          |
| Table 2.1        | Values of FRET efficiency (E), quantum yield ( $Q_D$ ), overlap integral (J), Förster's distance ( $R_0$ ), and D-A distance (R) for FRET between W214 and C34-IAEDANS pair  | 39       |
| Table 2.2        | Values of Stern-Volmer constants ( $K_{sv}$ ), intensity averaged fluorescence lifetimes ( $\tau_0$ ) and bimolecular quenching rate constants ( $k_q$ ) for C34-IAEDANS and W214  | 40       |
| <b>Chapter 3</b> |  |          |
| Table 3.1        | Determination of FRET parameters and D-A distances. Values of FRET efficiency (E), quantum yield ( $Q_D$ ), overlap integral (J), Förster's distance ( $R_0$ ), and D-A distance (R) for FRET between W214 and C34-IAEDANS pair              | 59       |
| Table 3.2        | Parameters for dynamic fluorescence quenching experiments. Values of Stern-Volmer constants ( $K_{sv}$ ), intensity averaged fluorescence lifetimes ( $\tau_0$ ) and bimolecular quenching rate constants ( $k_q$ ) for C34-IAEDANS and W214 | 62       |
| <b>Chapter 4</b> |  |          |
| Table 4.1        | Determination of kinetic parameters of A form aggregation  | 106      |
| <b>Chapter 5</b> |  |          |
| Table 5.1        | Peptide fragments and corresponding secondary structural content   | 121      |
| Table 5.2        | Hydrogen-exchange kinetics parameters for the N form at pH 6.5   | 133      |
| Table 5.3        | Hydrogen-exchange kinetics parameters for the N form at pH 7.0   | 134      |
| Table 5.4        | Hydrogen-exchange kinetics parameters for the DMG form at pH 3.0.  | 137      |
| Table 5.5        | Hydrogen-exchange kinetics parameters for the amyloid form at pH 3.0   | 138      |
| Table 5.6        | Peptide specific protection level for the N form, DMG and the amyloid form   | 140      |

---

## Synopsis of the Thesis

This thesis investigates the nature and significance of intermediate states, particularly dry molten globule-like intermediate states, in protein folding-unfolding and misfolding reactions. The thesis includes six different chapters. Chapter 1 presents a brief introduction to protein folding and the status quo of the dry molten globule (DMG)-like intermediate states in protein folding and misfolding pathways. Chapter 2 provides evidence of DMG-like equilibrium intermediates during the pH-induced unfolding of a multidomain protein, human serum albumin (HSA). We delineate the early steps during protein unfolding reaction and estimate the energetic contribution of van der Waals packing interaction in protein stability. Chapter 3 demonstrate the effect of simple cosolvents, like base, on HSA protein. Interestingly, we observed base-induced expansion in the inter-domain region of the protein, with DMG-like characteristics, indicating for non-cooperativity in native-DMG transition. Chapter 4 gives the first report of the initiation of amyloid aggregation from a DMG-like state for a neurodegenerative disease linked protein, nucleic acid binding domains of TDP-43 (TDP<sup>NBD</sup>), and investigates the mechanism of aggregation. Chapter 5 presents the high-resolution characterization of the structural dynamics of amyloidogenic DMG-like state and non-amyloidogenic native state of TDP<sup>NBD</sup>, and elucidate the structural core of the amyloid fibril, using hydrogen-deuterium exchange and mass spectrometry. Chapter 6 summarizes the contributions of this thesis and discuss some important future directions of the work.

### Chapter 1. Dry Molten Globule-like Intermediates in Protein Folding and Misfolding Pathways

The correct folding of a protein is a crucial step that directs the unique native structure of proteins which governs the function of the protein. The protein folding process is a pathway driven process involving intermediate states. It is traditionally believed that during unfolding of proteins the disruption of side-chain packing interactions occurs simultaneously with the solvation of the hydrophobic core. This gives rise to the formation of wet molten globule (WMG)-like state which further unfolds to the unfolded state. In contrast to this traditional view, an alternative model based upon the dry molten globule hypothesis proposes that the first step during the unfolding of proteins is expansion and loosening of side-chain packing interactions to give rise to a dry molten globule (DMG)-like state. The penetration of water molecules in the hydrophobic core occurs in the second step to give rise to the WMG state. Hence, it envisages that the unlocking of tertiary packing interactions and hydrophobic

---

solvation are decoupled. . We discuss the current understanding of the nature and significance of DMG-like intermediates in the protein folding and misfolding reaction.

## **Chapter 2. Evidence for Dry Molten Globule-like Domains in the pH-induced Equilibrium Folding Intermediate of a Multidomain Protein**

The role of van der Waals (vdW) packing interactions compared to the hydrophobic effect in stabilizing the functional structure of proteins is poorly understood. We show, using fluorescence resonance energy transfer, dynamic fluorescence quenching, red-edge excitation shift, and near- and far-UV circular dichroism, that the pH-induced structural perturbation of a multidomain protein leads to the formation of a state in which two out of the three domains have characteristics of dry molten globules, that is, the domains are expanded compared to the native protein with disrupted packing interactions but have dry cores. We quantitatively estimate the energetic contribution of vdW interactions and show that they play an important role in the stability of the native state and cooperativity of its structural transition, in addition to the hydrophobic effect. Our results also indicate that during the pH-induced unfolding, side-chain unlocking and hydrophobic solvation occur in two distinct steps and not in a concerted manner, as commonly believed.

## **Chapter 3. A Dry Molten Globule-like Intermediate During the Base-Induced Unfolding of a Multidomain Protein**

The nature of the initial structural events during the base-induced unfolding of the native (N) state of proteins is poorly understood. Combining site-specific fluorescence resonance energy transfer, size exclusion chromatography, dynamic fluorescence quenching, red-edge excitation shift and circular dichroism spectroscopy, we show that an early intermediate during the base-induced unfolding of a multidomain protein, i.e., the B form, has features of a dry molten globule. We show that the  $N \rightleftharpoons B$  transition involves protein expansion and loosening of packing of inter-domain helices near domains I and II without the disruption of intra-domain packing or any change in hydration of the inter-domain region which resembles a molten hydrocarbon. Surprisingly, the disruption of inter-domain packing accounts for 40–45% of the total change in free energy of complete unfolding. Our results show that the disruption of van der Waals packing can be decoupled in different regions of a protein and could occur prior to hydrophobic solvation during base-induced unfolding, challenging the existing notion.

---

**Chapter 4. Nucleation Dependent Aggregation of Nucleic Acid Binding Domains of TDP-43 Occur via Dry Molten Globule-like State**

Loosely packed states of proteins with a dry core, such as dry molten globules (DMGs), have been recently shown to be the first step during the unfolding and function of a few proteins, but their role in amyloid aggregation, responsible for many neurodegenerative diseases, is poorly understood. With the aid of global and site-specific biophysical tools, such as FRET, and kinetic studies, we show for the first time that the amyloid-like aggregation of the nucleic acid binding domains of the amyotrophic lateral sclerosis-linked protein transactive response DNA binding protein (TDP-43), begins with a DMG-like state. Moreover, we unveil that the DMG to amyloid aggregation transition follows nucleation dependent aggregation mechanism. Our results indicate that side-chain unlocking without core solvation could be the earliest step during stress-induced neurodegenerative diseases and suggest an important early therapeutic target for drug design. Our results also implicate that the significance of DMGs is not limited to mere protein folding reaction but has multiple dimensions in function and diseases.

**Chapter 5. Site-specific Structural Dynamics of the DMG-like Amyloid Precursor and the Structural Organization of the Core of Amyloid Fibril of the Nucleic Acid Binding Domains of TDP-43 Revealed by Hydrogen-Deuterium Exchange Coupled to Mass-Spectrometry**

The nucleic acid binding domains of the amyotrophic lateral sclerosis-linked protein, transactive response DNA binding protein (TDP-43), forms amyloid fibrils via a dry molten globule (DMG)-like precursor. However, we do not understand why only DMG-like state is amyloidogenic but not the native state. A coherent differentiation between the amyloidogenic and non-amyloidogenic form could aid in designing small molecules for the targeted drug delivery to neutralize the amyloidogenic form. Moreover, understanding the structural organization of the amyloid fibril would be beneficial for targeting the rigid amyloid core of the fibril for defibrillation. We utilized the hydrogen-deuterium exchange coupled to mass spectrometry and site-specifically revealed that the DMG-like form is highly dynamic compared to the native state. The increased structural dynamics in the DMG-like state provide more chances for two or more interacting fragments of different monomeric units to come together and form the nucleus for amyloid fibril. By comparing the extent of deuterium

incorporation in the amyloid form and the DMG-like form, we also show that the disordered regions of the protein participates in the formation of the core of the amyloid fibrils.

## **Chapter 6. Conclusions and Future Directions**

In this concluding chapter, we briefly discuss the purpose and findings of the study, significant contributions of the study to the field of protein folding and misfolding, limitations and future directions of our work.



**Chapter 1.**  
**Dry Molten Globule-like Intermediates in Protein**  
**Folding and Misfolding Pathways**

## 1.1 Introduction

Proteins are one of the most abundant macromolecules in living systems. Although proteins are the polymers of just 20 different type of amino acids arranged in a linear sequence, their three dimensional folded structure vary greatly governing the most diverse range of functions among all macromolecules. Proteins are synthesized as linear polypeptide chains and each newly synthesized polypeptide chain must find its path to attain the functionally active conformation, and proteins do so efficiently after getting synthesized from the ribosome as well as when they repetitively unfold and refold during their lifespan. The proper functioning of a biological system majorly depends on the correct folding of proteins to the unique three-dimensional native state to perform the designated function, which otherwise leads to various protein folding diseases, the reason why understanding the protein folding problem has been of significant interest over the past six decades. A solution to the protein folding problem will provide the link between the amino acid sequence and the functional structure of a protein, which may aid in protein structure prediction and novel protein design.

The history of understanding the protein folding problem dates back to nearly six decades with Anfinsen's work on the folding of RNase A.<sup>1,2</sup> With the observation that RNase A can fold spontaneously to its lowest free-energy functional native state, Anfinsen and co-workers concluded that all the information needed for a protein to fold is encoded in protein's peptide sequence. Meanwhile, Levinthal perceived that for a small polypeptide chain, it would take an astronomical length of time to find the native state through the conformational space, yet proteins fold within seconds to minutes timescale, the statement which is now eminently known as 'Levinthal Paradox'.<sup>3,4</sup> According to Levinthal's view, protein folding cannot be a random search of the native conformation through all possible structural arrangements, hence, the folding must occur via some predetermined pathways involving intermediate states. Later, many kinetic and equilibrium studies also reported for the presence of folding intermediates supporting Levinthal's view of folding.<sup>5-11</sup> However, early protein folding studies were limited to small (<100 amino acids) single-domain proteins, where the common experimental observation suggested for simple two-state folding of proteins indicating for cooperativity of the protein folding reaction.<sup>12-14</sup> For proteins folding via two-stage mechanism, where at equilibrium the observable states are only the unfolded and the folded state, a high energy transition state may exist which can be visualized under stabilizing conditions. For large multidomain proteins, the folding proceeds in discrete steps involving intermediate species.<sup>15-</sup>

20

The intermediate states have been traditionally viewed as critical milestones assisting a protein to fold correctly to the native state, hence, these intermediates were also called as on-pathway intermediates.<sup>21</sup> However, in some cases, the off-pathway intermediate or misfolded state has been found with non-native interactions acting as kinetic traps resulting in less efficient folding.<sup>22</sup> The on-pathway intermediate states provide crucial information on delineating the subdomain architecture and stabilities of a protein and gives insight into some functionally important conformational deviations from the native structure.<sup>23</sup> However, misfolded intermediates are the critical species linking the folding and misfolding processes, which often leads to aggregation and disease.<sup>23</sup> Therefore, immense efforts have been made to characterize the intermediate species detected during the folding and misfolding reactions. Here, we discuss the current understanding of the nature and the role of an underexplored intermediate state, dry molten globule (DMG)-like intermediate, in protein folding and misfolding reactions.

## 1.2 Protein folding-unfolding and molten globule intermediates

During the 1970s, equilibrium unfolding intermediates were observed for several globular proteins, having native-like secondary structure without specific tertiary structure.<sup>5-11</sup> Later it was proposed that the observed equilibrium intermediates constitute a common physical state, the molten globule state, of globular proteins, which is a general intermediate of protein folding reaction.<sup>24-28</sup> Since all the side-chains of these molten globule intermediates are hydrated with the solvent, these intermediates are termed as the wet molten globule (WMG) intermediates. The traditional view of protein folding-unfolding transition suggests that during protein unfolding, the disruption of tertiary packing and hydrophobic solvation of the protein core are coupled and occur concomitantly.<sup>12,14,29,30</sup> In other words, during protein folding, the WMG state folds to the native state cooperatively by acquiring the tertiary packing along with the hydrophobic desolvation (Figure 1.1).

However, the idea of gaining the non-specific hydrophobic interactions and unique native state topology in a single step invites questions over the efficiency of folding.



Figure 1.1. The stepwise folding of protein with wet molten globule (WMG) intermediate state.

The traditional view of protein folding was challenged in an alternative protein folding theory, based upon the theoretical phase diagram of proteins, where it was proposed that the hydrophobic desolvation and native tertiary packing are decoupled via a transition state known as dry molten globule (DMG) state (Figure 1.2).<sup>31,32</sup> DMGs were proposed as a native-like expanded transition state where the protein retains native-like secondary structure and hydrophobicity with disrupted tertiary packing of the side-chains. DMGs were considered as a highly unstable transition state in two-state folding-unfolding reactions due to the large unfavorable enthalpy change resulting from the disruption of side-chain packing. However, later it was realized that the disrupted tertiary packing of the side-chains would consequently result in a significant gain in conformational entropy, which stabilizes the DMG state.<sup>33-36</sup>

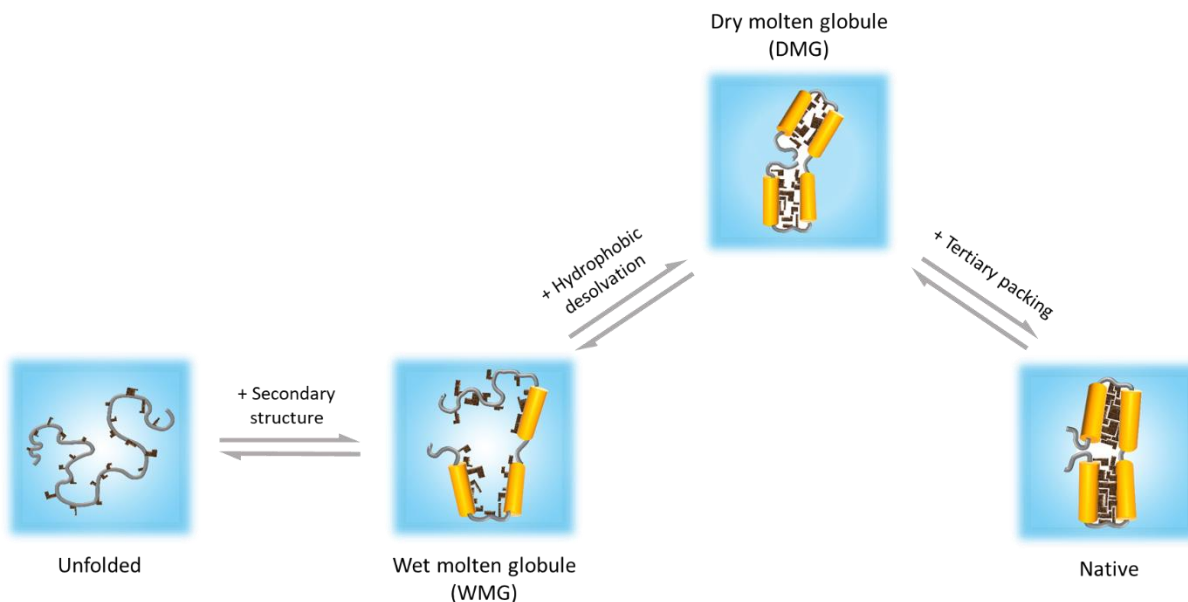


Figure 1.2. The multi-step folding of a protein. The unfolded protein first folds to a wet molten globule (WMG) intermediate state gaining some secondary structural elements followed by hydrophobic desolvation of protein forming a dry molten globule (DMG) intermediate state which further folds to the native state by tertiary packing of the side-chains.

### 1.3 Evidence of DMG-like intermediates

Kiefhaber and co-workers experimentally detected the DMG-like intermediate during the unfolding of RNase A, suggesting DMGs as an intermediate state preceding the rate-limiting step of unfolding.<sup>37,38</sup> The detection of DMG was a two-part search, where the intermediate was detected by real-time 1D H-NMR study and the dryness of the core was detected by the pulsed hydrogen exchange study during the unfolding of RNase A. In a simultaneous work, a DMG-like intermediate was observed during the unfolding of dihydrofolate reductase with the help of stopped-flow mixing and <sup>19</sup>F-NMR.<sup>39</sup> In another study, the water <sup>17</sup>O-NMR relaxation dispersion experiments revealed that the molten globule states of three structurally unrelated proteins are desolvated with native-like internal hydration.<sup>40</sup> A couple of early unfolding studies on hen lysozyme and staphylococcal nuclease indicated that the unfolding of the protein is initiated by DMG-like transition state.<sup>41,42</sup>

More direct evidence for DMG-like intermediate was reported during the initial unfolding of single-chain monellin (MNEI) protein.<sup>43</sup> The FRET measurements revealed that the intermediate state is expanded compared to the native state, consistent with the molten globule concept, and the protein interior remains dry as indicated by its failure to bind to ANS dye. The authors observe high cooperativity in the native to DMG transition and infer that the native to DMG transition could be an all-or-none transition and further unfolding from the DMG intermediate occurs non-cooperatively. Later, in a computational study, an intermediate state was observed with features of DMG-like state during the late stages of the folding of MNEI protein, consistent to the previous experimental evidence.<sup>44</sup>

For a small protein HP35, DMG-like intermediate has been found in equilibrium with the native state.<sup>45</sup> A recent study, however, shows that the transition state (for native to DMG transition) exhibits an expanded conformation and interestingly the DMG-like state of HP35 protein turned out to be compact in volume.<sup>46</sup>

In an equilibrium refolding study on high-pH-unfolded barstar protein,<sup>47</sup> it was observed that at low Na<sub>2</sub>SO<sub>4</sub> concentration, the protein folds to a pre molten globule state without any structure. The pre molten globule intermediate further folds to a DMG-like state, at higher concentration of Na<sub>2</sub>SO<sub>4</sub>, as a productive on-pathway intermediate in the folding reaction of barstar. Furthermore, during the denaturant-induced unfolding of barstar protein, an expanded early unfolding intermediate state was observed with DMG-like features,<sup>48</sup> indicating that DMGs are common intermediate states during folding and unfolding reactions.

In a combined experimental-simulation study, the SH3 domain of PI3 kinase protein was shown to form a DMG-like intermediate state before the further transition to WMG-like partially unfolded intermediate state during the denaturant-induced unfolding of protein.<sup>49</sup>

The evidence of DMG-like intermediate states are not limited to the chemical denaturant-induced folding-unfolding reaction of proteins. Recently, DMG-like intermediates with intact secondary structure and unlocked desolvated hydrophobic core were observed, using experimental and computational studies, during the pH-induced unfolding of various structurally distinct proteins (HSA, BBL, CcdB).<sup>50-54</sup> In a computational protein folding study on protein L, a late folding intermediate state was observed under low-temperature folding conditions with features of a DMG-like state.<sup>55</sup> Interestingly, some recent studies also suggest that under native-like conditions, DMG-like states exist in continuous equilibrium with the native state.<sup>45,54,56-59</sup>

The presence of DMG-like intermediates during protein folding-unfolding studies on small proteins suggests that the DMG-like intermediates are on-pathway intermediate state, not transition state species. However, the evidence for DMG-like intermediates are limited to small single-domain proteins due to the near-native characteristics of DMGs, which makes DMGs hard to detect using traditional global structural probes. We anticipate the necessity to investigate the universality of DMG-like intermediates and a unified mechanism of protein folding-unfolding of small and large proteins.

#### 1.4 Significance of DMGs in protein folding and unfolding reaction

The existence of DMGs prompts major revelations of the protein folding-unfolding mechanism:

- **Loss of close tertiary packing in the first step of unfolding**

The existence of DMGs suggests that in a protein unfolding reaction, the first step is loosening of the close tertiary packing of side-chains of the protein. In other words, close tertiary packing is present only at the end of the protein folding process. The loss of tertiary packing would also mean the loss of van der Waal's (vdW) packing interaction between the neighboring atoms of the closely packed side-chains.

- **Energetics of vdW interactions in protein stability**

The role of physicochemical forces, particularly vdW interactions, in protein folding and stability, is poorly understood.<sup>12,14,29,30,35,60-66</sup> According to the common notion, the

---

hydrophobic interactions are the primary driving force for protein folding and in stabilizing the native state.<sup>12,14,29,30</sup> The very difference between the native state and the DMG-like intermediate is the difference in vdW packing interactions, as mentioned above and DMGs would allow us to extract the complementary information regarding the significance of vdW interactions in the energetics of protein folding and stability.

- **A dynamic checkpoint during the folding reaction**

Protein folding is a pathway driven process that involves one or more intermediate states, where DMG-like intermediate states act as a final checkpoint. The higher conformational dynamics in DMG-like intermediate state aid a protein to attend multiple native conformations, depending on the environment/conditions, to perform multiple functions.

- **Rate limiting step of unfolding: Unlocking or Solvation?**

The presence of DMGs in continuous equilibrium with the native state implies that the unfolding of a protein starts from a dynamic equilibrium of the native and DMG state proceeding through a solvated WMG to the unfolded state. One important inference of this reaction is that the protein gets committed to the unfolding reaction after the solvation of the protein core, indicating that the hydration of protein core is the rate-limiting step of protein unfolding, not the tertiary structure unlocking.

- **Implication of DMGs in the mechanism of action of chemical denaturation**

Majority of the studies on protein folding predominantly focus on understanding how does a denaturant-unfolded protein fold to the native state upon dilution to the native (physiological) conditions. However, the process of protein unfolding is equally important to understand how common denaturants, like urea and guanidinium chloride, initiate the unfolding of proteins, which will consequently shine a light on the final step of the folding of proteins. Early theoretical studies on denaturant-induced unfolding suggested for the indirect unfolding mechanism, where it was proposed that the denaturant alters the structure of water, disturbing the first hydration layer and hydrophobic interactions of the protein.<sup>12,29,67,68</sup> This hypothesis advocates for a single step unfolding mechanism where the disruption of hydrophobic interactions results in the unfolding of the protein. Interestingly, however, based on the observations of the direct interaction of the denaturants with amino acids,<sup>69</sup> an alternative direct interaction model of protein unfolding was proposed.<sup>70-74</sup> In the direct interaction model, it has been proposed theoretically that denaturants unfold proteins in two

steps. Firstly, denaturants directly interact with the protein surface, replacing the water molecules from the first solvation shell of the protein. This results in the swelling of the protein to a native-like intermediate called a dry globule due to the lack of water in the core. In the second step, the protein-bound denaturant molecules, from the first solvation shell, enter the hydrophobic core of protein followed by water molecules, facilitating the global unfolding of the protein. Moreover, various evidence of DMG-like intermediates observed during protein unfolding studies also support that denaturants unfold the proteins via a two-step direct interaction unfolding mechanism and the unfolding of proteins is initiated by the formation of DMG-like intermediate states. Further, it will be interesting to understand whether other protein denaturants/cosolvents like acids and bases follow direct or indirect unfolding mechanism.

- **Functional state of protein: native or DMG?**

Since DMGs exists in equilibrium with the native state, it is crucial to venture the involvement of DMGs in function. Moreover, due to the lack of tertiary packing, DMGs have increased conformational dynamics and the increased conformational dynamics have been shown to be associated with the improved catalytic function of proteins.<sup>75-82</sup> These observations invite questions over the nature of the functional state of a protein. A careful assessment is necessary to examine the putative role of DMGs in performing functions.

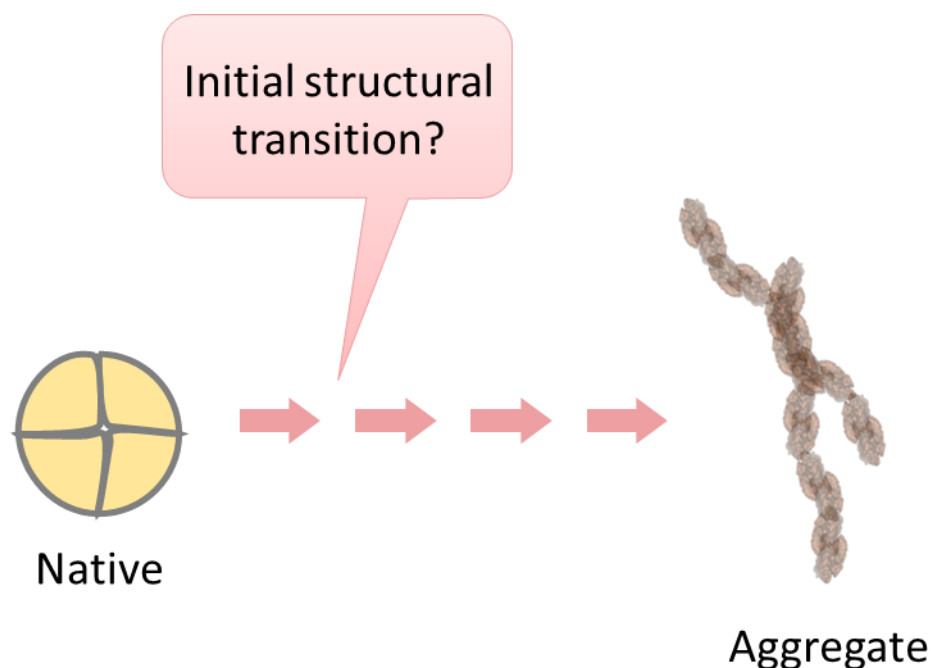
### **1.5 DMGs in protein misfolding reaction?**

The misfolding of proteins is associated with the aggregation of proteins. More than 40 human diseases are related to protein deposition in the form of aggregates.<sup>79,80</sup> Many of these diseases are fatal neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. Despite the availability of a vast amount of literature on protein aggregation, our understanding of the initial steps during protein misfolding and aggregation is inadequate (Figure 1.3). To date, there is a scarcity of literature reporting on the nature of the initial structural transition a protein encounters during the aggregation reaction.

Recently, it has been shown in proteins such as acylphosphatase, superoxide dismutase, transthyretin,  $\beta$ 2-microglobulin and insulin that the aggregation of the proteins can occur under native conditions.<sup>80,81</sup> In general, the native-like states initiate the oligomer formation, which later gets reorganized by the global conformational change to form the amyloid fibrils. A most remarkable example is insulin, which upon aggregation transforms from all helical structure to complete  $\beta$ -sheet amyloid fibril.<sup>82</sup> In case of prion protein, whose aggregation is associated



with a group of fatal neurodegenerative diseases in humans and animals, it was found that a native-like intermediate state is a branching point between the folding and aggregation pathway.<sup>83</sup> All these findings suggest the involvement of native and native-like intermediates in the aggregation and diseases.



*Figure 1.3. Understanding the initial steps and the nature of the initial structural transition during protein aggregation.*

Further, the existence of DMGs in equilibrium with the native state and the near-native characteristics of DMG state,<sup>52,60-62</sup> suggests for the reevaluation of the initial steps during the aggregation reaction for the plausible involvement of DMG-like species in aggregation. Moreover, the disrupted tertiary structure of DMGs result in increased conformational dynamics and the increased conformational dynamics, for insulin and  $\beta$ 2-microglobulin, are associated with protein aggregation.<sup>84,85</sup> Hence, the increased conformational dynamics in DMG-like intermediate states also make them a potential candidate for protein aggregation diseases. However, as yet, there is no direct evidence of the involvement of DMGs in any protein aggregation study, hence, it is necessary to examine the initial steps in the misfolding reaction, leading to the aggregation for the involvement of DMG-like intermediates.

## 1.6 Conclusions and open questions

The unfolding of the native state of small single-domain proteins begins with the disruption of tertiary packing of the side-chains without core solvation, forming a dry molten

globule-like intermediate state. The unfavorable enthalpy change resulting from the tertiary structure unlocking is compensated by the large gain in the conformational entropy during the formation of DMGs. The existence of DMG-like intermediate during the folding-unfolding reaction of proteins reveal that a protein remains flexible until the last step of folding and would also aid in the estimation of the energetic contribution of different forces, such as hydrophobic interaction and van der Waal's packing interaction, in protein stability.

However, many further underexplored questions invite study. Do DMG-like intermediates exist for large multidomain proteins? Whether DMGs universal? Is the DMG-native state transition cooperative? Can DMGs be stabilized using cosolvents? What is the energetic contribution of different forces in protein stability? What are the initial steps during the misfolding reactions? Whether DMGs are the common intermediates in the folding and misfolding pathway? How does the misfolded aggregation-prone state and the aggregated state differ from the native state? With thoughtful experimental design utilizing global and site-specific biophysical tools and careful interpretation of the results, such questions can be answered, which will increase our understanding of the significance of DMG-like intermediates in protein folding and misfolding reactions.

## 1.7 References

1. Anfinsen, C. B., Haber, E., Sela, M., and White, F. H., Jr. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A* 47, 1309-1314.
2. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* 181, 223-230.
3. Levinthal, C. (1968) Are there pathways for protein folding. *J Chim Phys* 65, 44-45.
4. Levinthal, C. (1969) How to fold graciously. *Mossbauer Spectroscopy in Biological Systems. Proceedings University of Illinois Bulletin*, 22-24.
5. Wong, K. P., and Tanford, C. (1973) Denaturation of bovine carbonic anhydrase B by guanidine hydrochloride. A process involving separable sequential conformational transitions. *J Biol Chem* 248, 8518-8523.
6. Holladay, L. A., Hammonds, R. G., Jr., and Puett, D. (1974) Growth hormone conformation and conformational equilibria. *Biochemistry* 13, 1653-1661.

7. Wong, K. P., and Hamlin, L. M. (1974) Acid denaturation of bovine carbonic anhydrase B. *Biochemistry* 13, 2678-2683.
8. Kuwajima, K., Nitta, K., Yoneyama, M., and Sugai, S. (1976) Three-state denaturation of alpha-lactalbumin by guanidine hydrochloride. *J Mol Biol* 106, 359-373.
9. Robson, B., and Pain, R. H. (1976) The mechanism of folding of globular proteins. Equilibria and kinetics of conformational transitions of penicillinase from *Staphylococcus aureus* involving a state of intermediate conformation. *Biochem J* 155, 331-344.
10. Kuwajima, K. (1977) A folding model of alpha-lactalbumin deduced from the three-state denaturation mechanism. *J Mol Biol* 114, 241-258.
11. Nozaka, M., Kuwajima, K., Nitta, K., and Sugai, S. (1978) Detection and characterization of the intermediate on the folding pathway of human alpha-lactalbumin. *Biochemistry* 17, 3753-3758.
12. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24, 1-95.
13. Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* 30, 10428-10435.
14. Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv Protein Chem* 33, 167-241.
15. Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* 70, 697-701.
16. Privalov, P. L. (1982) Stability of proteins. Proteins which do not present a single cooperative system. *Adv Protein Chem* 35, 1-104.
17. Novokhatny, V. V., Kudinov, S. A., and Privalov, P. L. (1984) Domains in human plasminogen. *J Mol Biol* 179, 215-232.
18. Jaenicke, R. (1991) Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry* 30, 3147-3161.
19. Abkevich, V. I., Gutin, A. M., and Shakhnovich, E. I. (1995) Domains in folding of model proteins. *Protein Sci* 4, 1167-1177.
20. Dobson, C. M. (1995) Finding the right fold. *Nat Struct Biol* 2, 513-517.

21. Radford, S. E. (2000) Protein folding: progress made and promises ahead. *Trends Biochem Sci* 25, 611-618.
22. Brockwell, D. J., Smith, D. A., and Radford, S. E. (2000) Protein folding mechanisms: new methods and emerging ideas. *Curr Opin Struct Biol* 10, 16-25.
23. Tsytlonok, M., and Itzhaki, L. S. (2013) The how's and why's of protein folding intermediates. *Arch Biochem Biophys* 531, 14-23.
24. Ohgushi, M., and Wada, A. (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* 164, 21-24.
25. Dolgikh, D. A., Abaturvov, L. V., Brazhnikov, Lebedev Iu, O., and Chirgadze Iu, N. (1983) [Acid form of carbonic anhydrase: "molten globule" with a secondary structure]. *Dokl Akad Nauk SSSR* 272, 1481-1484.
26. Ptitsyn, O. B. (1987) Protein folding: Hypotheses and experiments. *J Protein Chem* 6, 273-293.
27. Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6, 87-103.
28. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett* 262, 20-24.
29. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14, 1-63.
30. Tanford, C. (1968) Protein denaturation. *Adv Protein Chem* 23, 121-282.
31. Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28, 1667-1680.
32. Finkelstein, A. V., and Shakhnovich, E. I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28, 1681-1694.
33. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725-2737.

34. Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23, 4-10.
35. Bhattacharyya, S., and Varadarajan, R. (2013) Packing in molten globules and native states. *Curr Opin Struct Biol* 23, 11-21.
36. Thirumalai, D., Liu, Z., O'Brien, E. P., and Reddy, G. (2013) Protein folding: from theory to practice. *Curr Opin Struct Biol* 23, 22-29.
37. Kiefhaber, T., and Baldwin, R. L. (1995) Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc Natl Acad Sci U S A* 92, 2657-2661.
38. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
39. Hoeltzli, S. D., and Frieden, C. (1996) Real-time refolding studies of 6-19F-tryptophan labeled Escherichia coli dihydrofolate reductase using stopped-flow NMR spectroscopy. *Biochemistry* 35, 16843-16851.
40. Denisov, V. P., Jonsson, B. H., and Halle, B. (1999) Hydration of denatured and molten globule proteins. *Nat Struct Biol* 6, 253-260.
41. Segawa, S., and Sugihara, M. (1984) Characterization of the transition state of Lysozyme unfolding. I. Effect of protein-solvent interactions on the transition state. *Biopolymers* 23, 2473-2488.
42. Vidugiris, G. J. A., Markley, J. L., and Royer, C. A. (1995) Evidence for a molten globule-like transition state in protein folding from determination of activation volumes. *Biochemistry* 34, 4909-4912.
43. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* 106, 12289-12294.
44. Maity, H., and Reddy, G. (2018) Thermodynamics and Kinetics of Single-Chain Monellin Folding with Structural Insights into Specific Collapse in the Denatured State Ensemble. *J Mol Biol* 430, 465-478.

45. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* 107, 4955-4960.
46. Neumaier, S., and Kiefhaber, T. (2014) Redefining the dry molten globule state of proteins. *J Mol Biol* 426, 2520-2528.
47. Rami, B. R., and Udgaonkar, J. B. (2002) Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* 41, 1710-1716.
48. Sarkar, S. S., Udgaonkar, J. B., and Krishnamoorthy, G. (2013) Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* 105, 2392-2402.
49. Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* 118, 6380-6392.
50. Yue, Z., and Shen, J. (2018) pH-Dependent cooperativity and existence of a dry molten globule in the folding of a miniprotein BBL. *Phys Chem Chem Phys* 20, 3523-3530.
51. Baliga, C., Selmke, B., Worobiew, I., Borbat, P., Sarma, S. P., Trommer, W. E., Varadarajan, R., and Aghera, N. (2019) CcdB at pH 4 Forms a Partially Unfolded State with a Dry Core. *Biophys J* 116, 807-817.
52. Acharya, N., Mishra, P., and Jha, S. K. (2016) Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* 7, 173-179.
53. Acharya, N., Mishra, P., and Jha, S. K. (2017) A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Phys Chem Chem Phys* 19, 30207-30216.
54. Mishra, P., and Jha, S. K. (2017) An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* 121, 9336-9347.
55. Maity, H., and Reddy, G. (2018) Transient intermediates are populated in the folding pathways of single-domain two-state folding protein L. *J Chem Phys* 148, 165101.

56. Roche, J., Caro, J. A., Norberto, D. R., Barthe, P., Roumestand, C., Schlessman, J. L., Garcia, A. E., García-Moreno, B. E., and Royer, C. A. (2012) Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* 109, 6945-6950.
57. Fu, Y., Kasinath, V., Moorman, V. R., Nucci, N. V., Hilser, V. J., and Wand, A. J. (2012) Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* 134, 8543-8550.
58. Goluguri, R. R., Sen, S., and Udgaonkar, J. (2019) Microsecond sub-domain motions and the folding and misfolding of the mouse prion protein. *Elife* 8, p.e44766.
59. Mishra, P., and Jha, S. K. (2019) Slow Motion Protein Dance Visualized Using Red-Edge Excitation Shift of a Buried Fluorophore. *The Journal of Physical Chemistry B* 123, 1256-1264.
60. Sandberg, W. S., and Terwilliger, T. C. (1989) Influence of interior packing and hydrophobicity on the stability of a protein. *Science* 245, 54-57.
61. Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28, 4914-4922.
62. Behe, M. J., Lattman, E. E., and Rose, G. D. (1991) The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci U S A* 88, 4195-4199.
63. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255, 178-183.
64. Honig, B., and Yang, A. S. (1995) Free energy balance in protein folding. *Adv Protein Chem* 46, 27-58.
65. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J* 10, 75-83.
66. Baldwin, R. L. (2014) Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proc Natl Acad Sci U S A* 111, 13052-13056.

67. Frank, H. S., and Evans, M. W. (1945) Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J Chem Phys* 13, 507-532.
68. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.
69. Robinson, D. R., and Jencks, W. P. (1965) The effect of compounds of the urea-guanidinium class on the activity coefficient of acetyltetraglycine ethyl ester and related compounds. *J Am Chem Soc* 87, 2462-2470.
70. Schellman, J. A. (1978) Solvent denaturation. *Biopolymers* 17, 1305-1322.
71. Mountain, R. D., and Thirumalai, D. (2003) Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc* 125, 1950-1957.
72. O'Brien, E. P., Dima, R. I., Brooks, B., and Thirumalai, D. (2007) Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. *J Am Chem Soc* 129, 7346-7353.
73. Hua, L., Zhou, R., Thirumalai, D., and Berne, B. J. (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci U S A* 105, 16928-16933.
74. Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* 111, 4856-4861.
75. Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002) Enzyme dynamics during catalysis. *Science* 295, 1520-1523.
76. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438, 117-121.
77. Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007) Conformational entropy in molecular recognition by proteins. *Nature* 448, 325-329.



- 
78. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G., and Wand, A. J. (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6, 352-358.
  79. Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., Benkovic, S. J., and Wright, P. E. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332, 234-238.
  80. Doshi, U., McGowan, L. C., Ladani, S. T., and Hamelberg, D. (2012) Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proc Natl Acad Sci U S A* 109, 5699-5704.
  81. Tzeng, S. R., and Kalodimos, C. G. (2012) Protein activity regulation by conformational entropy. *Nature* 488, 236-240.
  82. Caro, J. A., Harpole, K. W., Kasinath, V., Lim, J., Granja, J., Valentine, K. G., Sharp, K. A., and Wand, A. J. (2017) Entropy in molecular recognition by proteins. *Proc Natl Acad Sci U S A* 114, 6563-6568.
  83. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 75, 333-366.
  84. Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5, 15-22.
  85. Jahn, T. R., Parker, M. J., Homans, S. W., and Radford, S. E. (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* 13, 195-201.
  86. Bouchard, M., Zurdo, J., Nettleton, E. J., Dobson, C. M., and Robinson, C. V. (2000) Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. *Protein Sci* 9, 1960-1967.
  87. Honda, R. P., Xu, M., Yamaguchi, K. I., Roder, H., and Kuwata, K. (2015) A Native-like Intermediate Serves as a Branching Point between the Folding and Aggregation Pathways of the Mouse Prion Protein. *Structure* 23, 1735-1742.
  88. Haas, J., Vöhringer-Martinez, E., Bögehold, A., Matthes, D., Hensen, U., Pelah, A., Abel, B., and Grubmüller, H. (2009) Primary steps of pH-dependent insulin aggregation kinetics are governed by conformational flexibility. *ChemBioChem* 10, 1816-1822.

89. Hodkinson, J. P., Radford, S. E., and Ashcroft, A. E. (2012) The role of conformational flexibility in  $\beta$ 2-microglobulin amyloid fibril formation at neutral pH. *Rapid Commun Mass Spectrom* 26, 1783-1792.

**Chapter 2.**  
**Evidence for Dry Molten Globule-like Domains in the  
pH-induced Equilibrium Folding Intermediate of a  
Multidomain Protein**

*Reproduced with the permission from Acharya, N., Mishra, P., and Jha, S. K. (2016) Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. J Phys Chem Lett 7, 173-179.*

## 2.1 Introduction

The nature of physicochemical forces that stabilize the functional native structure of proteins is poorly understood.<sup>1-12</sup> It is commonly believed that seclusion of hydrophobic amino acids from water (hydrophobic effect) is the major contributor to the protein stability.<sup>1-4</sup> The hydrophobic amino acids are also tightly packed inside the protein core, similar to the crystals of small organic molecules,<sup>13</sup> in order to maximize the strength of the van der Waals (vdW) interactions. However, the role and energetic contribution of vdW interactions relative to hydrophobic effect in stabilizing protein molecules is not yet clear.<sup>5-8,11,12</sup>

Wet molten globules (WMG) have been observed as equilibrium and kinetic intermediates during the folding of many proteins.<sup>14-20</sup> WMG possess substantial secondary structure and a fluctuating tertiary structure with perturbed side-chain packing and water solvated hydrophobic core. Theoretical studies of the thermodynamics of protein unfolding predict that a side-chain unlocking step, resulting in the formation of dry molten globules (DMG), precedes hydrophobic solvation during protein unfolding.<sup>21,22</sup> According to these theoretical studies, DMG are expanded forms of the native protein in which tight side-chain packing interactions are ruptured, but the water molecules have not penetrated the hydrophobic core. In contrast to WMG, experimental evidence for DMG have come mainly from kinetic studies where they have been observed as transient initial intermediates during protein unfolding.<sup>23-29</sup> It is important to identify them under equilibrium conditions where detailed structural characterization using high-resolution probes can be done and thermodynamic contribution of hydrophobic desolvation and side-chain locking in protein stability can be dissected. However, till date DMG have been reported at equilibrium in limited cases and involve only small single domain proteins.<sup>19,30,31</sup>

Various fundamental questions related to DMG remain unanswered.<sup>11,27,32</sup> Are DMG universal protein folding intermediates and form during all kinds of unfolding reactions? Is the core of DMG solid-like as in the native protein or like a molten liquid? What is the protection factor of DMG? When does the hydrophobic solvation occur? Are side-chain packing interactions disrupted uniformly or only in selective regions of the protein structure? In DMG whether all regions of the hydrophobic core of the protein are dry or only in patches? It becomes more important to determine the answers to these questions for multi-domain proteins because for large proteins there exists a possibility that some regions have characteristics of DMG while others behave like WMG. However, to the best of our knowledge DMG-like domains have not been identified in any multi-domain protein till date.

In this study, we have used a battery of spectroscopic probes including near- and far-UV circular dichroism (CD), fluorescence resonance energy transfer (FRET), dynamic fluorescence quenching and red-edge excitation shift (REES) to dissect the pH-induced structural perturbation of a well-studied multi-domain protein, human serum albumin (HSA), into distinct structural events characterized by changes in secondary and tertiary structure, protein expansion, side-chain unlocking, hydrophobic solvation and solvation dynamics of the protein matrix.

## 2.2 Materials and Methodology

### Spectroscopic methods and instruments

All the ultraviolet (UV) absorption spectra were collected on Perkin Elmer lambda 650 UV/Vis spectrometer using a quartz cell of path length 1 cm. All the fluorescence spectra were taken on Perkin Elmer fluorescence spectrometer LS 55 using a quartz fluorescence cell of path length 1cm. All the circular dichroism (CD) spectra were taken on Jasco J-815 CD spectrometer. For far-UV CD and near-UV CD a quartz cell of path length 0.1 cm and 1 cm, respectively, were used. The background signals due to buffers were subtracted from all the fluorescence and CD spectra. Refractive index measurements were done on an Abbe refractometer from Rajdhani Scientific Instruments Co. (Model: RSR-2). The time-resolved fluorescence decay kinetics were measured on a time-correlated single photon counting (TCSPC) set up from Horiba Scientific (Deltaflex).

### Reagents, chemicals, buffers and experimental conditions

Human Serum Albumin (HSA) (Fatty acid and Globulin free,  $\geq 99\%$  pure) was procured from Sigma and used without further purification. The concentration of HSA was determined by the measurement of absorbance at 280 nm, using an extinction coefficient ( $\epsilon_{280\text{nm}}$ ) of  $36500 \text{ M}^{-1}\text{cm}^{-1}$ .<sup>33</sup> Urea (ultra-pure grade) was purchased from Bioworld and 5-(((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) was purchased from Life technologies. Guanidine hydrochloride (GdmCl), acrylamide and all other chemicals and reagents were obtained from Sigma and were of highest purity grade.

For pH titrations, a universal buffer consisting of 20mM sodium citrate, 20mM sodium phosphate and 20mM sodium borate was used, for the pH range 3.0-11.0. For the pH range 1.0-3.0, 20 mM KCl-HCl buffer was used. For all other experiments, pH 7.0 buffer was composed of 20mM sodium phosphate, pH 2.2 buffer was composed of 20mM glycine, pH 1.0

buffer was composed of 20mM KCl-HCl, and unfolding buffer at pH 7.0 was composed of 20mM sodium phosphate and 9 M urea. All the solutions and buffers were filtered with 0.2  $\mu\text{m}$  filters before use. The concentrations of urea and GdmCl were determined by the measurement of the refractive index.<sup>34</sup>

### Preparation of 1, 5-IAEDANS labeled HSA

HSA was labeled at the free cysteine, C34, with 1,5-IAEDANS as described earlier.<sup>35</sup> In brief, GdmCl unfolded HSA was incubated with 20 fold molar excess of 1,5-IAEDANS in an unfolding buffer containing 6 M GdmCl and 20 mM Tris at pH 8.0. The reaction mixture was kept in dark with continuous stirring at room temperature for 4 hrs. After the labeling reaction was complete, the solution was diluted 10-fold with refolding buffer (20mM sodium phosphate at pH 7.0) and kept overnight at 4°C for refolding of the labeled protein. The sample was concentrated to 2.5 mL using 30 kDa centrifugal concentrator (GE Healthcare) and passed through a PD10 column (GE Healthcare) to separate the labeled protein from free dye and GdmCl. The extent of labeling was determined as described previously.<sup>36</sup> Briefly, the total protein concentration after labeling was determined by the assay of Bradford,<sup>37</sup> using unlabeled HSA to prepare the standard calibration curve (Figure 2.1). The extent of labeling was calculated by a separate determination of the concentration of HSA-IAEDANS by measuring the absorbance at 337 nm and using the extinction coefficient ( $\epsilon_{337}$ ) of 4500  $\text{M}^{-1}\text{cm}^{-1}$ .<sup>36</sup> The extent of labeling observed was > 95%. The secondary structure and thermodynamic stability of HSA-IAEDANS were observed to be similar to HSA (Figure 2.2). For all the experiments using HSA-IAEDANS, care was taken to ensure that the samples were minimally exposed to light.

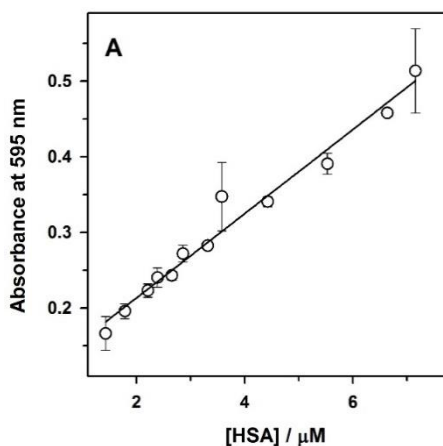


Figure 2.1. Calibration curve for the determination of concentration of HSA-IAEDANS using Bradford assay.<sup>37</sup> The x-axis represents the concentration of unlabeled HSA measured by

absorbance at 280 nm (using  $\epsilon_{280\text{nm}} = 36500 \text{ M}^{-1}\text{cm}^{-1}$ ). The y-axis represents the absorbance of the same sample at 595 nm when mixed with Bradford reagent<sup>37</sup> in 1:50 ratio. The error bars represent the standard deviation calculated from five separate experiments. The solid black line through the data points is a fit to the equation  $y = 0.0557x + 0.1019$ .

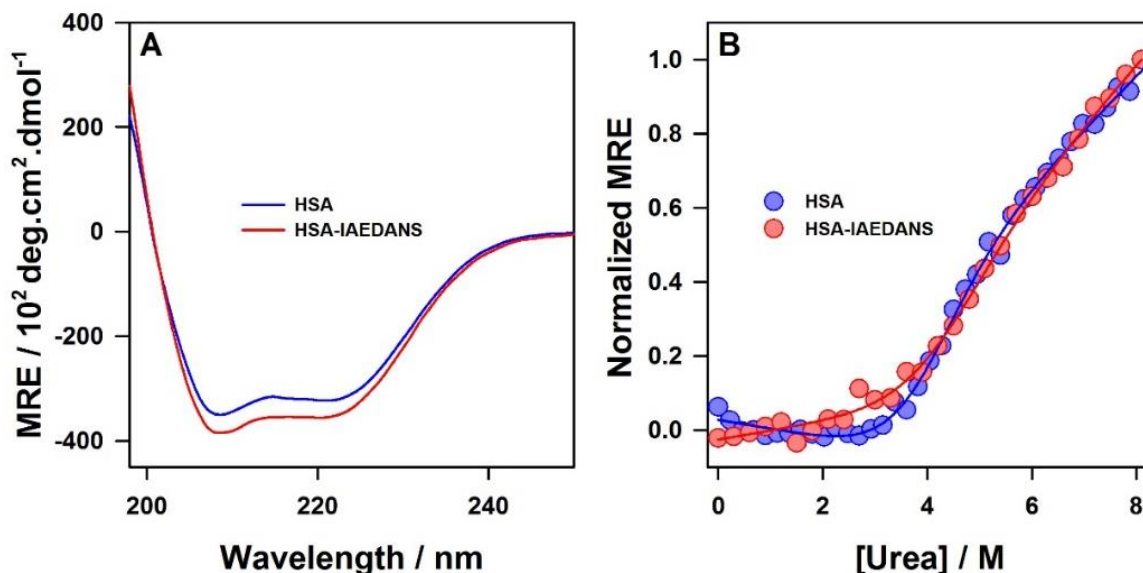


Figure 2.2. The secondary structure and the thermodynamics of unfolding of HSA remain unperturbed upon IAEDANS labeling. Panel A shows the far-UV CD spectra of the HSA and HSA-IAEDANS. Panel B shows the urea-induced equilibrium unfolding transitions of HSA and HSA-IAEDANS as monitored by the change in the far-UV CD signal at 222 nm. In panel B, for easy comparison, the data has been normalized between 0 and 1 for each curve using the equation  $\frac{S_N - S_{\text{obs}}}{S_N - S_U}$ , where  $S_{\text{obs}}$ ,  $S_N$  and  $S_U$ , respectively, denote the observed CD signal at a particular [Urea], average signal of the native baseline and the signal of the unfolded protein at 8 M urea. The red and blue solid lines through the data represent a non-linear least-squares fit to a two-state model.<sup>38</sup> The values of the midpoint of unfolding,  $C_m$ , for HSA and HSA-IAEDANS from the two-state analysis are 4.1 M and 4.2 M, respectively.

### pH titration monitored using fluorescence and CD

The protein samples at different pH were kept overnight at room temperature for equilibration. For the unlabeled protein, the fluorescence spectrum of W214 was measured using the excitation wavelength of 295 nm and collecting the emission from 310 nm to 420 nm. For pH titration with HSA-IAEDANS, the protein samples were excited at 337 nm and the emission was collected from 350 nm to 550 nm. Fluorescence spectra were collected with the

excitation slit width of 5-7 nm and the scan speed of 100 nm/min. The changes in the secondary structure were measured by monitoring the far-UV CD spectrum in the wavelength range 190-250 nm. The changes in the tertiary structure were measured by monitoring the near-UV CD spectrum in the wavelength range 250-300 nm. The protein concentrations used for the fluorescence and far-UV CD experiments were  $\sim 3-4 \mu\text{M}$  and that for the near-UV CD experiments were  $\sim 20 \mu\text{M}$ . Each CD spectrum was collected with the data pitch of 1 nm, data integration time of 1 s, bandwidth of 2 nm and scan speed of 100 nm/min. CD spectra were averaged over two scans. For comparison, the fluorescence, the far-UV and the near-UV CD spectrum of the unfolded proteins (in 9 M urea) at pH 7.0 were also recorded as described above.

### Analysis of the pH titration

For the pH titration monitored by far-UV CD, the change in CD signal at 222 nm as a function of pH was observed to occur in a single sigmoidal step. We fitted the data to equation 1, which is a transformation of the Henderson-Hasselbalch equation and is derived from a model in which the structural transition from N to either F or E form is coupled to a single protonation step:

$$Y_{obs} = \frac{Y_{P1} + Y_D 10^{(pH - pH_m)}}{1 + 10^{(pH - pH_m)}} \quad [1]$$

where,  $Y_{obs}$  corresponds to observed far-UV CD signal at a particular pH value,  $Y_D$  and  $Y_{P1}$  correspond to the signals of the deprotonated and the protonated forms, respectively, and  $pH_m$  is the midpoint of the observed titration.

For the pH titration monitored by fluorescence, the change in fluorescence signal at 340 nm as a function of pH was observed to occur in two sigmoidal steps. We fitted the data to equation 2, which is derived from a model in which the structural transition from N to F and F to E forms are each coupled to one protonation step:

$$Y_{obs} = \frac{Y_D + Y_{P1} 10^{(pH_{m1} - pH)}}{1 + 10^{(pH_{m1} - pH)}} + \frac{Y_{P2} + Y_{P1} 10^{(pH - pH_{m2})}}{1 + 10^{(pH - pH_{m2})}} \quad [2]$$

where,  $Y_{obs}$  corresponds to observed fluorescence signal at a particular pH value;  $Y_D$ ,  $Y_{P1}$ , and  $Y_{P2}$ , respectively, correspond to the signals of the deprotonated, the singly protonated and



doubly protonated forms; and  $pH_{m1}$  and  $pH_{m2}$ , respectively, correspond to the midpoints of the first and the second protonation steps.

### FRET between W214 and C34-IAEDANS and determination of the Forster's distance

For measurement of FRET between W214 and C34-IAEDANS, all the fluorescence spectra at pH 7.0 (Figure 2.6C, results) were collected in an identical manner. Similarly, all the spectra at pH 2.2 (Figure 2.6D, results) were collected in an identical manner. The excitation wavelength used in all the cases was 295 nm.

We determined the value of Forster's distance,  $R_0$ , in the N form at pH 7.0, and in the E forms at pH 2.2, as described previously<sup>25</sup> by using the equation:

$$R_0 = 0.211[Q_D J \kappa^2 n^{-4}]^{\frac{1}{6}} \quad [3]$$

In equation 3,  $Q_D$  is the quantum yield of donor,  $J$  is the overlap integral,  $\kappa^2$  is the orientation factor and  $n$  is the refractive index of the medium.<sup>39</sup> To determine the value of  $Q_D$  of W214 in the N form, we use the fact that under the conditions similar to those used in this study, the  $Q_D$  of the free tryptophan is 0.13<sup>40</sup> and the mean fluorescence lifetime ( $\tau_m$ ) is 2.7 ns.<sup>39</sup> Because  $Q_D$  is directly proportional to  $\tau_m$ ,<sup>39</sup> and  $\tau_m$  of W214 in the N state is 6.42 ns,<sup>41</sup> the value of  $Q_D$  in the N state was determined to be 0.31.  $Q_D$  is also directly proportional to the area under the fluorescence emission spectrum.<sup>39</sup> We used the value for the  $Q_D$  in N state, and the ratio of the area under the fluorescence emission spectrum of E form at pH 2.2 to the area under the fluorescence emission spectrum of N, and determined the value of  $Q_D$  in the E form to be 0.19. The overlap integral,  $J$ , is defined as the spectral overlap between the fluorescence emission spectrum of the donor and the absorbance spectrum of the acceptor, and is given by:

$$J = \frac{\int F(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda}{\int F(\lambda)d\lambda} \quad (\text{M}^{-1}\text{cm}^{-1}\text{nm}^4) \quad [4]$$

The fluorescence emission spectra,  $F(\lambda)$ , of W214 and the absorbance spectra,  $\varepsilon(\lambda)$ , of C34-IAEDANS were measured in the N and E forms (Figure 2.3), and the calculated values of  $J$  are listed in Table 2.1 (see results). The refractive index,  $n$ , of the medium was determined to be 1.333, both at pH 7.0 and at pH 2.2. The donor and acceptor are assumed to be oriented randomly with respect to each other and the value of orientation factor,  $\kappa^2$ , was taken to be 2/3. The value of  $R_0$  was calculated to be 25.8 Å in the N form at pH 7.0 and 23.8 Å in the E form at pH 2.2. The values of various FRET parameters are listed in Table 2.1 (see results). We would like to point out that due to the sixth-root dependence in the Förster's equation (equation

6), the errors in the determination of D-A distances will be small even if the errors in measurements of the values of various FRET parameters are relatively large.

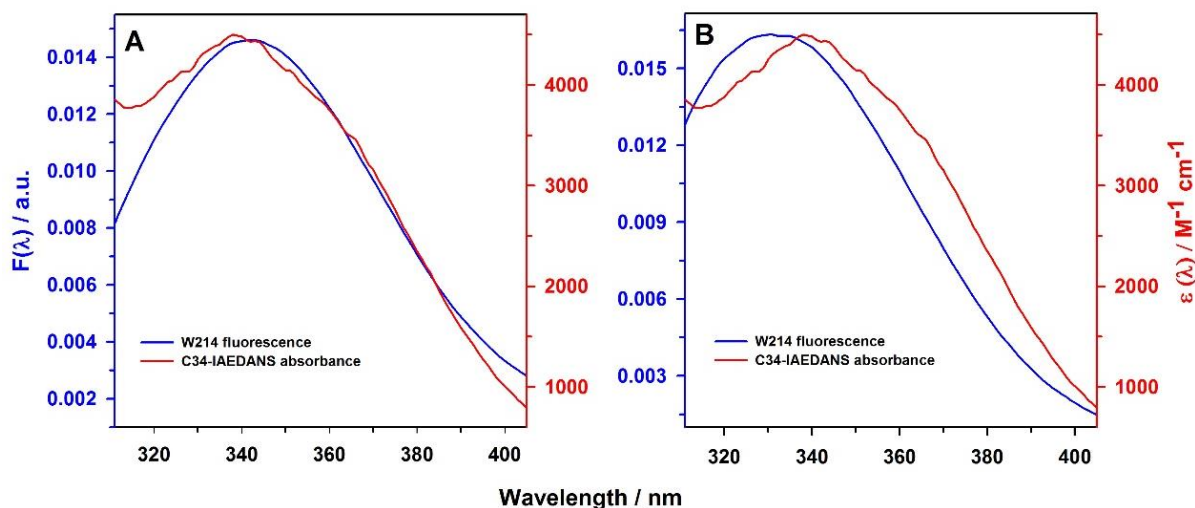


Figure 2.3. Determination of the overlap integral,  $J$ , at pH 7.0 (A) and at pH 2.2 (B). In both the panels, the fluorescence emission spectra of W214 and the absorbance spectra of C34-IAEDANS are shown according to the left y-axis and the right y-axis, respectively. Each fluorescence spectrum has been normalized so that the total area under the spectrum is unity. Each absorbance spectrum has been divided by the respective molar protein concentration to obtain  $\epsilon(\lambda)$ .

### FRET between W214 and tyrosine residues

The emission spectrum of tyrosine overlaps with the absorbance spectrum of tryptophan and this FRET pair has been used to measure changes in distances during the conformational changes in proteins.<sup>39,42</sup> HSA has 18 tyrosine residues distributed throughout its structure. To determine whether the pH-induced expansion of HSA is global in nature, we measured the extent to which the fluorescence emission of tyrosine residues is quenched due to W214 in the U form, N form and the E forms at pH 2.2 and pH 1.0 (Figure 2.4). Fluorescence excitation was done at 280 nm which excited both W214 and tyrosine residues. We observed that in the U form, where W214 and tyrosine residues were far apart and the fluorescence emission of tyrosine residues was not quenched, there is a bimodal distribution of fluorescence emission centered on  $\sim 300$  nm and  $\sim 348$  nm due to tyrosine residues and W214, respectively (Figure 2.4A, results). In the N form, tyrosine residues and W214 are spatially near to each other and we observe that the fluorescence emission due to tyrosine residues is dramatically quenched. We observed a unimodal distribution centered on  $\sim 342$  nm due to the emission of W214 (Figure 2.4B, results). In the E forms at pH 2.2 (Figure 2.4C, results) and at pH 1.0

(Figure 2.4D, results), we observe a bimodal distribution of fluorescence emission centered on  $\sim 300$  nm and  $\sim 333$  nm corresponding to tyrosine residues and W214. The fluorescence emission of tyrosine residues is not quenched significantly in the E form, similar to the U form, due to the increase in their spatial distance with W214. These results indicate that the pH-induced expansion of the protein in the E form is global in nature.

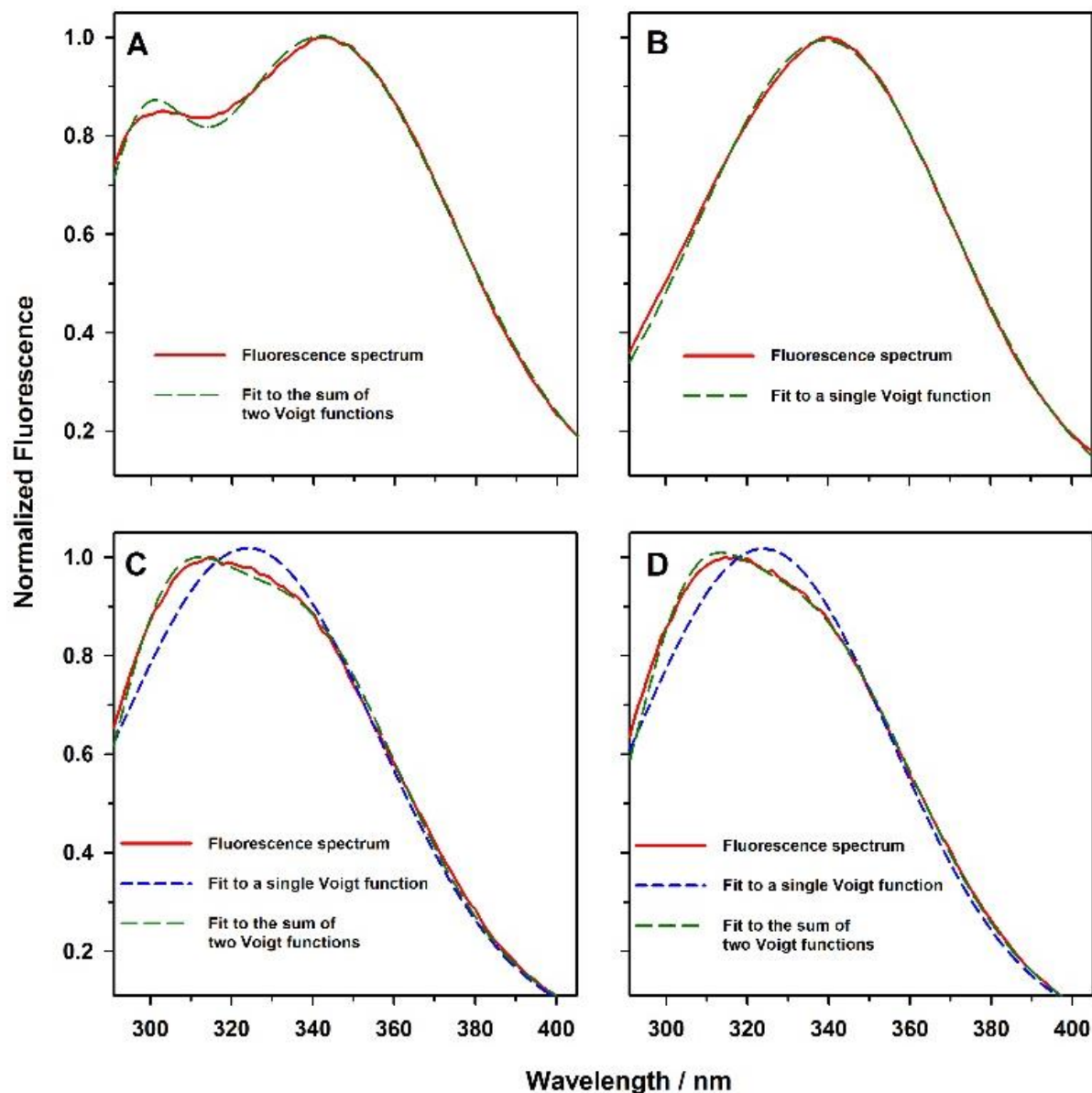


Figure 2.4. FRET between W214 and tyrosine residues of HSA. Fluorescence emission spectra in the (A) U form; (B) N form; (C) E form at pH 2.2; and (D) E form at pH 1.0; after excitation at 280 nm. The fluorescence spectra in panels A, C and D fit to the sum of two Voigt functions and the spectrum in panel B fit to a single Voigt function, indicating that tyrosine emission is quenched in the N form but not in U form and in the E forms at pH 2.2 and pH 1.0.

### Acrylamide quenching experiments

For acrylamide quenching assay, HSA and HSA-IAEDANS were incubated at pH 1.0, pH 2.2, pH 7.0 and unfolding buffer at pH 7.0. 2 M acrylamide stock solution was freshly prepared and appropriate volume for the desired acrylamide concentration was added just before taking the fluorescence spectra. For HSA, fluorescence spectra were taken by exciting W214 at 295 nm and collecting the emission from 310-420 nm. For HSA-IAEDANS, fluorescence spectra were taken by exciting C34-IAEDANS at 337 nm and collecting the emission from 350-550 nm. Quenching of fluorescence intensity at the emission maxima was analyzed according to equation 7 (see results). The spread in the values of fluorescence at the emission maximum (error bars in Figure 2.7, see results) were estimated from two independent measurements.

The values of Stern-Volmer constant,  $K_{SV}$ , were determined using equation 7, as described in the result section. Using the values of  $K_{SV}$  and separately determined values of  $\tau_0$  (intensity averaged fluorescence lifetime, see below), we calculated the values of  $k_q$ , the bimolecular quenching constant, in the E forms, the N form and the U form. It is important to note that the values of  $k_q$  in the U form were determined in the presence of 9 M urea, which was absent in the N form and the E forms. To understand the effect of viscosity of 9 M urea solution on  $k_q$ , we performed the dynamic quenching experiments on a tryptophan analog, N-acetyl-L-tryptophanamide (NATA), in the absence (at pH 2.2 and pH 7.0) and the presence of 9 M urea (pH 7.0) (Figure 2.5).

We observed that the values of  $K_{SV}$  are  $23.0 \text{ M}^{-1}$ ,  $23.3 \text{ M}^{-1}$ , and  $17.1 \text{ M}^{-1}$ , respectively, at pH 2.2, pH 7.0 and in the presence of 9 M urea. We used the previously determined values of  $\tau_0$  for NATA (2.7 ns at pH 2.2 and pH 7.0<sup>39</sup> and 2.85 ns in the presence of 9 M urea<sup>43</sup>) and determined the values of  $k_q$ . The values of  $k_q$  are  $8.5 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  at pH 2.2,  $8.6 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  at pH 7.0 and  $6.0 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  in the presence of 9 M urea. These results indicate that the presence of 9 M urea reduces the measured value of  $k_q$  by 1.4-fold. In the light of this observation, the values of  $k_q$  in the U forms of HSA and HSA-IAEDANS also need to be corrected for the presence of 9 M urea. The corrected values of  $k_q$  in the U forms are listed in Table 2.2 (see results) along with the values of  $k_q$  in the E forms and the N form.

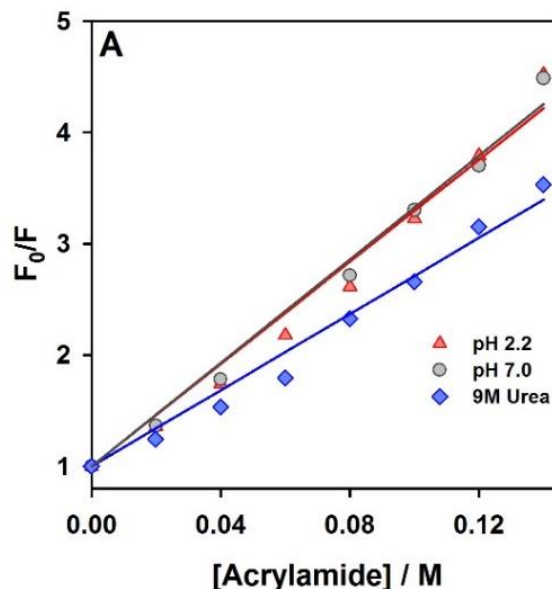


Figure 2.5. Stern-Volmer plots for quenching of fluorescence of *N*-acetyl-*L*-tryptophanamide (NATA) by acrylamide. The solid lines through the data are fits to the equation  $\frac{F_0}{F} = 1 + K_{SV}[Q]$ .

### Time-resolved fluorescence measurements

The time-resolved fluorescence decay kinetics of C34-IAEDANS was recorded using an excitation laser source of 370 nm. The instrument response function (IRF) was measured to be  $\sim 196$  ps using an aqueous solution of the milk powder. The emission decays were collected at 480 nm. A good signal-to-noise ratio was obtained by collecting 10,000 counts at the peak. The fluorescence decays were deconvoluted with the IRF and were fitted to a sum of two exponentials (using the DAS6 analysis software supplied by Horiba Scientific):

$$I(t) = \sum_i \alpha_i e^{-\frac{t}{\tau_i}}, \quad i = 1 \text{ to } 2 \quad [5]$$

where,  $I(t)$  is the fluorescence intensity at time  $t$  and  $\alpha_i$  is the amplitude of the  $i^{\text{th}}$  fluorescence lifetime  $\tau_i$  such that  $\sum_i \alpha_i = 1$ . Mean fluorescence lifetime ( $\tau_m$ ) and intensity averaged fluorescence lifetime,  $\tau_0$ , were calculated using the equations<sup>39</sup>:  $\tau_m = \sum \alpha_i \tau_i$  and  $\tau_0 = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i}$ , respectively.

The values of  $\tau_0$  for C34-IAEDANS under different conditions are listed in Table 2.2 (see results). For W214, we have used the values of  $\tau_0$  from the previous studies<sup>41,44</sup> where they have been determined under the conditions similar to those used in this study. These values are also listed in Table 2.2 (see results).

---

## REES experiments

The fluorescence emission spectra of W214 were recorded after excitation at different wavelengths ranging from 295 nm to 305 nm. Nominal slit widths were used for excitation and emission. The protein concentrations used were ~ 6-12  $\mu\text{M}$ . The fluorescence emission spectra were collected with the scan speed of 50 nm/min and averaged over three scans. Background intensities due to the solvent Raman peak were subtracted from all the samples. The spread in the values of emission maximum (error bars in Figure 2.9, see results) were estimated from two independent measurements.

## Urea-induced equilibrium unfolding transitions monitored by fluorescence and far-UV CD

The urea-induced equilibrium unfolding transitions of HSA were measured at pH 7.0 and pH 2.2. The protein samples were incubated overnight at room temperature in different concentrations of urea for equilibration. The change in fluorescence of W214 was measured using the excitation wavelength of 295 nm and collecting the emission from 310 nm to 420 nm. The changes in the secondary structure during urea-induced unfolding were measured by monitoring the far-UV CD spectrum in the wavelength range 190-250 nm. The protein concentration used for both the fluorescence and far-UV CD experiments was ~3-4  $\mu\text{M}$ .

## 2.3 Results and Discussion

### 2.3.1 pH-induced structural transitions

The structure of HSA consists of three domains, I, II and III, and each of them is divided into two subdomains, A and B (Figure 2.6A). It has been demonstrated earlier that in the low pH forms of HSA, domain III remains fully unfolded but there are only subtle structural changes in domain I and II.<sup>45</sup> Here, we report that decrease in pH of the solvent leads to the formation of a state, in which domain I and domain II of the protein have characteristics of DMG. We discuss the implications of this observation for protein stability and the mechanism of pH-induced protein denaturation.

We observed that the fluorescence of the sole tryptophan residue, W214, of HSA changes in two distinct sigmoidal steps on decreasing the pH from 7.0 to 1.0 (Figure 2.6B), indicating that the protein undergoes two structural transitions, coupled to the protonation of two ionizable groups. The mid-points of the two transitions are estimated to be at pH 3.7 and pH 2.7. It has been reported that the native form (N) of HSA transforms into a fast migrating

form (F) between pH 5.0 and 3.5 ( $N \rightleftharpoons F$  transition), and the F form converts into an acid-expanded or the extended (E) form below pH 3.5 ( $F \rightleftharpoons E$  transition).<sup>45-48</sup> Our biphasic pH titration data monitored by W214 fluorescence supports this conclusion. The pH titration of the protein was also monitored by far-UV CD signal at 222 nm (Figure 2.6B). We observed that there is no apparent change in the secondary structure during  $F \rightleftharpoons E$  transition and the change in secondary structure from the N to either F or E form occurs in a single sigmoidal step, with the mid-point of transition at pH 3.6.

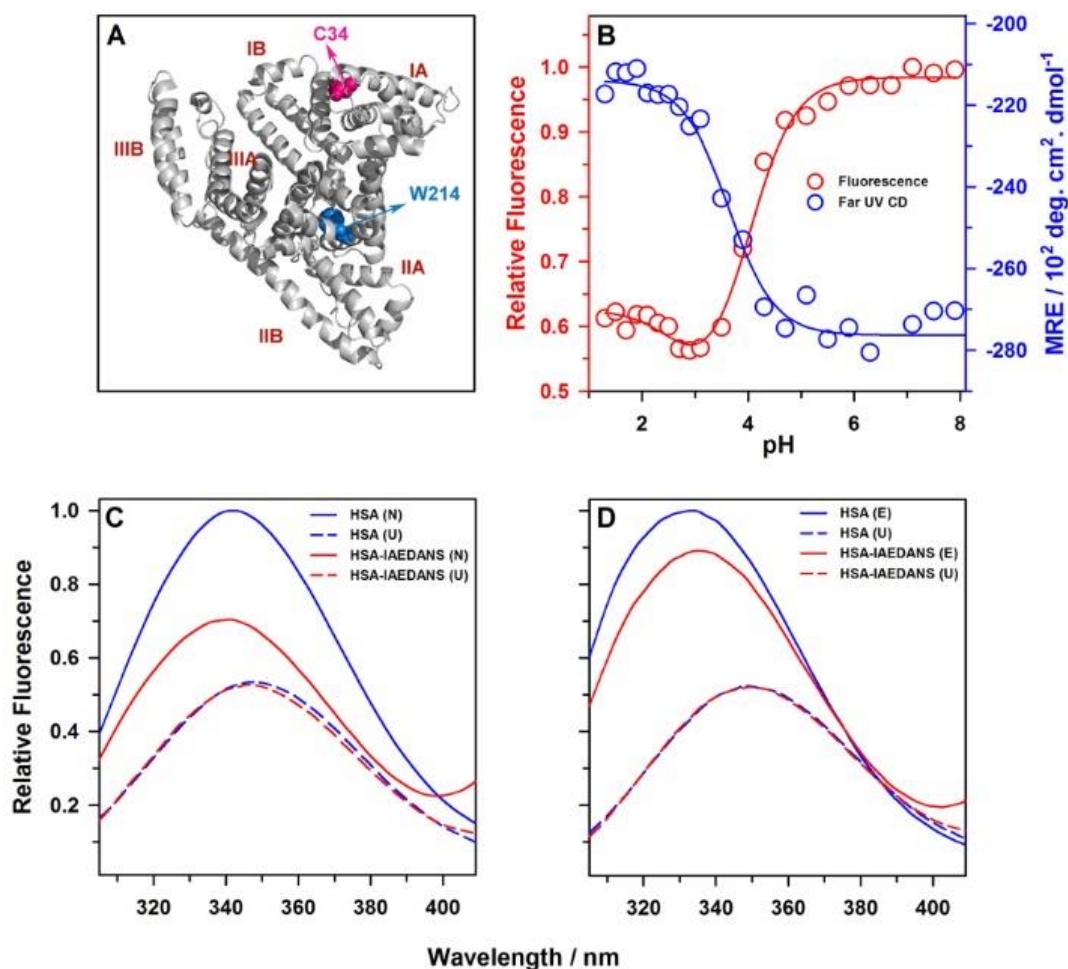


Figure 2.6. pH dependence of conformation of HSA and pH-induced expansion of HSA monitored by FRET. (A) Structure of HSA drawn from PDB file 1AO6 using the program PyMOL. (B) The change in fluorescence emission of W214 at 340 nm and the mean residue ellipticity at 222 nm are plotted against pH, according to the left and the right y-axis, respectively. The blue and red lines through the data are fits to a 2-state and a 3-state pH titration model, respectively (equation 1 and 2, SI). (C) pH 7.0 and (D) pH 2.2: Fluorescence emission spectra of HSA and HSA-IAEDANS are shown in the absence (N form and E form) and the presence of 9 M urea (U form).

### 2.3.2 Structural expansion in the E form

We explored whether the formation of the E form (at pH 2.2) from N form (at pH 7.0) is accompanied with any expansion in the dimension of the protein using FRET. W214, which is located in a helical segment of domain II, served as the donor (D) fluorophore. There are 35 cysteine residues in HSA, 34 of which form 17 disulfide bridges. The sole free thiol moiety, C34, located at the N-terminal of helix 3 in domain I, was labeled with the fluorescent dye 5-(((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS), which served as the acceptor (A) fluorophore. We observed that C34 can be quantitatively labeled with 1,5-IAEDANS (Figure 2.1), as has been observed previously for other fluorescence dyes,<sup>49</sup> and that the secondary structure and thermodynamic stability of the unlabeled (HSA) and the 1,5-IAEDANS labeled (HSA-IAEDANS) proteins are similar (Figure 2.2). The absorbance spectrum of C34-IAEDANS overlaps with the emission spectrum of W214 (Figure 2.3) (forming a FRET pair) and this pair has been used previously to monitor the change in distances during folding and unfolding of proteins.<sup>50</sup> The fluorescence of W214 is quenched significantly in the N form of HSA-IAEDANS (Figure 2.6C), because of the spatial proximity with C34-IAEDANS. However, it is not quenched in the urea unfolded protein (U form), both at pH 7.0 and at pH 2.2 (Figure 2.6C and 2.6D), where W214 and C34-IAEDANS are farther than the FRET distance. This observation indicates that C34-IAEDANS quenches the fluorescence of W214 in a distance-dependent manner. The extent of quenching of W214 by C34-IAEDANS in the E form is significantly less than that in the N form (Figure 2.6D). We used the data on HSA and HSA-IAEDANS, in Figure 2.6C and 2.6D, to quantitate the change in FRET efficiency ( $E$ ) and D-A distance ( $R$ ) during N $\rightleftharpoons$ E transition, using equation 6,<sup>39</sup>

$$E = 1 - \frac{F_{DA}}{F_D} = \left(1 + \frac{R^6}{R_0^6}\right)^{-1} \quad [6]$$

The fluorescence signal of W214 in HSA at the wavelength of the maximum emission was taken as  $F_D$  and that in the HSA-IAEDANS was taken as  $F_{DA}$ . We experimentally determined the value of Forster's distance,  $R_0$ , in the N form to be  $\sim 25.8 \text{ \AA}$ , and in the E form to be  $\sim 23.8 \text{ \AA}$ , respectively (Figure 2.3 and Table 2.1). The D-A distance in the N form was determined to be  $\sim 29.6 \text{ \AA}$  and that in the E form to be  $\sim 33.1$ . This result indicates that the protein is expanded by  $\sim 3.5 \text{ \AA}$  in the E form compared to the N form, along the axis connecting C34-IAEDANS and W214. Since, vdW interactions have a steep dependence on the interatomic distances, the movement of the helix containing C34-IAEDANS in domain I and the helix containing W214 in domain II away from each other by  $\sim 3.5 \text{ \AA}$  in the E form must be



accompanied by severe disruption of the side-chain packing interactions involving these helices. We also observed that the fluorescence of 18 tyrosine residues which are distributed throughout the structure of HSA is quenched dramatically in the N form due to FRET with W214, but the extent of quenching decreases significantly in the E form (as in the U form) (Figure 2.4), suggesting that in the E form the expansion of the protein might be global in nature.

### 2.3.3 E form contain hydrophobic core

We next investigated the water solvation of protein core in domain I and domain II. The fluorescence Stokes shift of a fluorophore is a very sensitive measure of the polarity of its surrounding medium. We monitored the changes in the hydrophobicity of domain I by monitoring the wavelength of the maximum fluorescence emission,  $\lambda_{max}^{em}$ , of C34-IAEDANS as a function of pH. In the N form, C34-IAEDANS is buried inside the hydrophobic core and its  $\lambda_{max}^{em}$  is 468 nm, which shifts towards red to 489 nm in the water-solvated U form (Figure 2.7A). We observed that in the E forms at pH 2.2 and pH 1.0,  $\lambda_{max}^{em}$  of C34-IAEDANS are 470 nm and 468 nm, respectively, indicating that inner core of domain I retains its native-like hydrophobicity and is not accessible to water.

In the N form, W214 is partially buried in the hydrophobic core of the domain II and its  $\lambda_{max}^{em}$  is ~342 nm (Figure 2.7B). In the U form, W214 is fully solvated by water and we observed that the  $\lambda_{max}^{em}$  is red-shifted to ~348 nm (Figure 2.7B). In the E forms at pH 2.2 and pH 1.0,  $\lambda_{max}^{em}$  is blue shifted to ~333 nm, indicating that the inner core of domain II is hydrophobic and not accessible to water. It appears that the loosening of side-chain packing in domain II due to protein expansion allow the movement of W214 side-chain to more hydrophobic environment. It is possible that this movement of W214 brings it in the proximity of either disulfide bonds, histidine residues or carboxyl groups, any of which can quench its fluorescence intensity (Figure 2.6B).<sup>39</sup>

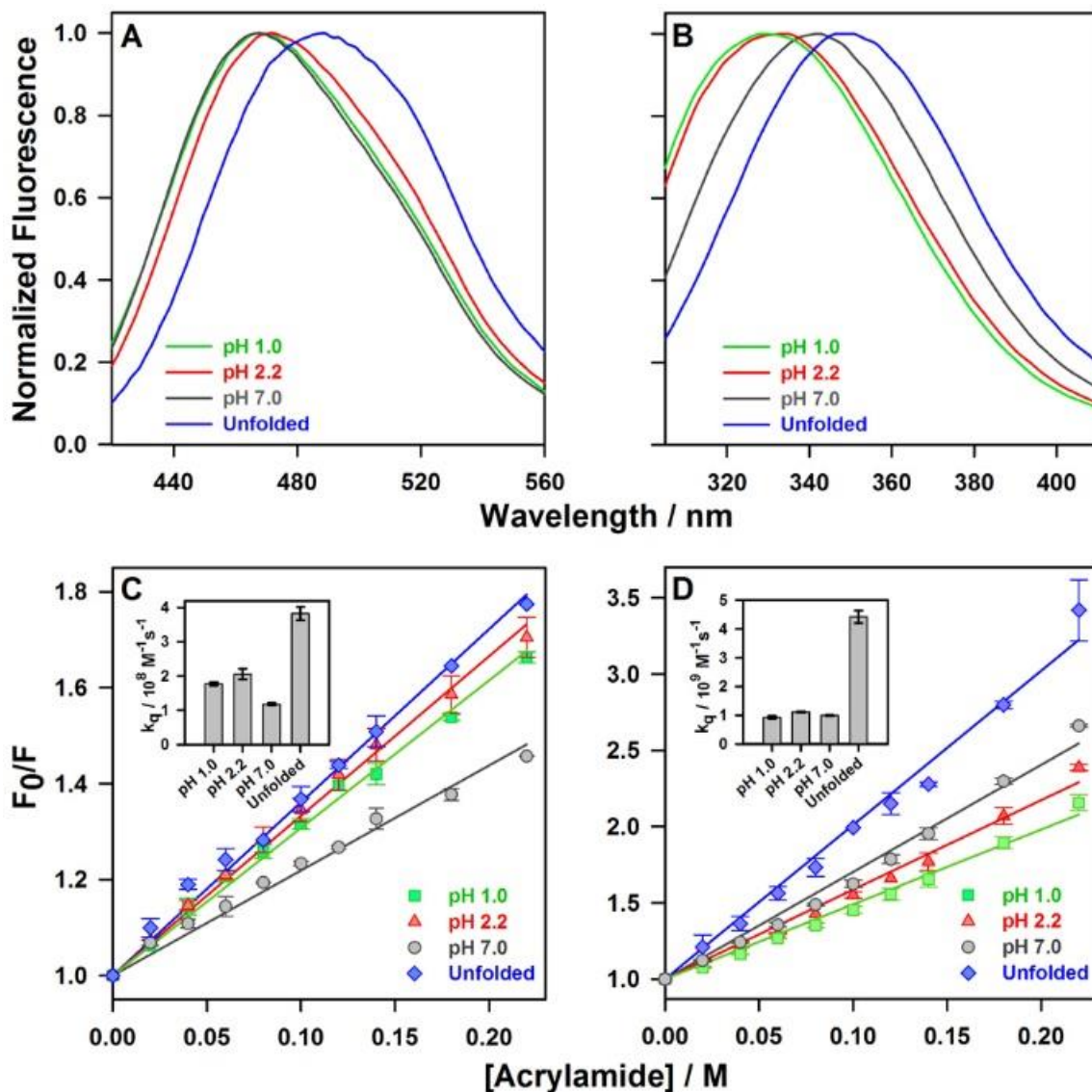


Figure 2.7. pH dependence of the solvation of hydrophobic core of domains I and II. Comparison of fluorescence spectra at different pH of (A) C34-IAEDANS and (B) W214. For comparison, each spectrum in A and B is normalized to the value of 1 at its emission maximum. Stern-Volmer plots for quenching of fluorescence of (C) C34-IAEDANS and (D) W214 by acrylamide. The solid lines through the data are fits to the equation  $\frac{F_0}{F} = 1 + K_{SV}[Q]$ . Insets in panel C (C34-IAEDANS) and in panel D (W214) compare the values of the bimolecular quenching constant,  $k_q$  (see text).

Table 2.1. Values of FRET efficiency ( $E$ ), quantum yield ( $Q_D$ ), overlap integral ( $J$ ), Förster's distance ( $R_0$ ), and D-A distance ( $R$ ) for FRET between W214 and C34-IAEDANS pair.

| Conditions    | $^*E$           | $^{\S}Q_D$ | $^{\dagger}J / 10^{13} \text{M}^{-1} \text{cm}^{-1} \text{nm}^4$ | $^{\ddagger}R_0 / \text{Å}$ | $^{\ddagger}R / \text{Å}$ |
|---------------|-----------------|------------|--|-----------------------------|---------------------------|
| <b>pH 7.0</b> | $0.30 \pm 0.01$ | 0.31       | 5.1  | 25.8                        | $29.6 \pm 0.2$            |
| <b>pH 2.2</b> | $0.12 \pm 0.02$ | 0.19       | 5.1  | 23.8                        | $33.1 \pm 0.9$            |

$^*E$  was determined using the fluorescence spectra in Figure 2.6C and 2.6D using equation 6.

$^{\S}Q_D$  was determined as described above in the 'Experimental procedures and data analysis' section.

$^{\dagger}J$  was calculated using equation 4.

$^{\ddagger}R_0$  was calculated using equation 3; with the following values:  $\kappa^2 = 2/3$ ;  $n = 1.333$ .

$^{\ddagger}R$  was calculated using equation 1. The standard deviation was estimated from three separate measurements.

Note: In the determination of the values of  $E$  and  $R$ , the errors at pH 7.0 denote the standard deviation estimated from three separate measurements and the errors at pH 2.2 denote the spread in the values from two separate measurements.

### 2.3.4 E form retains N-like structural fluctuations and solvent accessibility

Since the cores of domain I and domain II are hydrophobic but loosely packed in the E forms, we next employed dynamic fluorescence quenching experiments<sup>39,51,52</sup> and explored whether transient structural fluctuations allow the buried C34-IAEDANS and W214 side-chains to come in molecular contact with quencher molecules on the periphery of the protein structure. Dynamic fluorescence quenching is described by the Stern-Volmer equation as,<sup>39</sup>

$$\frac{F_0}{F} = 1 + K_{SV}[Q] = 1 + k_q\tau_0[Q] \quad [7]$$

We measured the fluorescence intensities of C34-IAEDANS (Figure 2.7C) and W214 (Figure 2.7D) in the absence ( $F_0$ ) and in the presence ( $F$ ) of different concentrations of acrylamide ( $Q$ ), a neutral collisional quencher of IAEDANS and tryptophan fluorescence, in the N form, E forms and the U forms.  $\frac{F_0}{F}$  was plotted against [acrylamide] (Figure 2.7C and 2.7D) and the slope of the straight line yielded the value of Stern-Volmer constant,  $K_{SV}$  (equation 7). The values of  $K_{SV}$  under different conditions are listed in Table 2.2. We observed that for C34-IAEDANS, the values of  $K_{SV}$  in the E forms at pH 2.2 and 1.0, are intermediate

between that in the N form and the U form (Figure 2.7C). For W214, the values of  $K_{SV}$  in the E forms at pH 2.2 and 1.0, are slightly lower than that in the N form (Figure 2.7D). Using the values of  $K_{SV}$ , we determined the values of  $k_q$ , the bimolecular rate constant for the formation of the molecular contact between the fluorophore and acrylamide in the photo-excited state of the fluorophore ( $K_{SV} = k_q \tau_0$ ). The values of  $\tau_0$ , the intensity averaged fluorescence lifetime, were measured separately under above conditions (Figure 2.8) and are listed in Table 2.2.

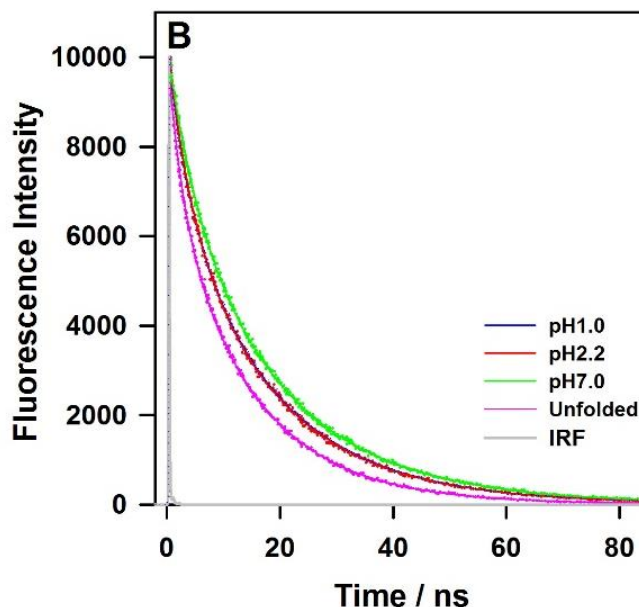


Figure 2.8. Fluorescence intensity decay kinetics of C34-IAEDANS.

Table 2.2. Values of Stern-Volmer constants ( $K_{sv}$ ), intensity averaged fluorescence lifetimes ( $\tau_0$ ) and bimolecular quenching rate constants ( $k_q$ ) for C34-IAEDANS and W214.

| Conditions      | C34-IAEDANS     |             |                         | W214            |               |                         |
|-----------------|-----------------|-------------|-------------------------|-----------------|---------------|-------------------------|
|                 | $K_{sv}/M^{-1}$ | $\tau_0/ns$ | $k_q/10^8 M^{-1}s^{-1}$ | $K_{sv}/M^{-1}$ | $^*\tau_0/ns$ | $k_q/10^9 M^{-1}s^{-1}$ |
| <b>pH 1.0</b>   | 3.07            | 17.23       | 1.8                     | 4.90            | 5.28          | 0.9                     |
| <b>pH 2.2</b>   | 3.33            | 17.31       | 1.9                     | 5.87            | 5.28          | 1.1                     |
| <b>pH 7.0</b>   | 2.19            | 18.55       | 1.2                     | 7.02            | 7.08          | 1.0                     |
| <b>Unfolded</b> | 3.61            | 14.15       | 3.7 <sup>§</sup>        | 10.11           | 3.31          | 4.4 <sup>§</sup>        |

\* These values are taken from previous studies<sup>41,44</sup>

§ These values are corrected for the effects of viscosity of 9 M urea, as described above in the methodology section.

We observed that for C34-IAEDANS, the value of  $k_q$  is  $3.7 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$  in the water-exposed U form, which decreases to  $1.2 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$  in the N form (Figure 2.7C (inset)) due to the burial of C34-IAEDANS in the protein structure. For W214, the value of  $k_q$  is  $4.4 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  in the U form, which decreases to  $1.0 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  in the N form (Figure 2.7D (inset)). The values of  $k_q$  in the N and U forms are only 3.1-fold and 4.4-fold different, respectively, for C34-IAEDANS and W214. This difference, however, is significant as the errors in the values of  $k_q$  are unlikely to be more than  $\pm 10 \%$ , considering the small errors associated with the determination of the slopes of the straight lines (values of  $K_{SV}$ ) and in the measurement of the fluorescence lifetimes (values of  $\tau_0$ ). We observed that for C34-IAEDANS, the values of  $k_q$  in the E forms at pH 2.2 and pH 1.0 are  $1.9 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$  and  $1.8 \times 10^8 \text{ M}^{-1}\text{s}^{-1}$ , respectively, which are intermediate between that in the N form and the U form (Figure 2.7C (inset)). For W214, the values of  $k_q$  in the E forms at pH 2.2 and pH 1.0 are  $1.1 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$  and  $0.9 \times 10^9 \text{ M}^{-1}\text{s}^{-1}$ , respectively, which are similar to that in the N form (Figure 2.7D (inset)). Fluorescence quenching for the fluorophores which are buried inside the hydrophobic core of proteins occurs through nanosecond (ns) structural fluctuations that expose the fluorophore to the quencher.<sup>51,52</sup> The above results suggest that in the E forms, the core of domain I exhibits faster structural fluctuations than the native protein, whereas domain II has native-like structural flexibility.

### 2.3.5 E form retains native-like solvation dynamics

It is not fully understood whether the inner core of DMG is solid-like as in the native proteins or like a molten liquid as expected by disruption of the side-packing interactions.<sup>11,27,32</sup> We explored the solvation dynamics of the protein matrix in domain II using REES experiments<sup>39,53</sup> on W214. In general,  $\lambda_{max}^{em}$  of a fluorophore does not depend on the excitation wavelength. However, we observed that the  $\lambda_{max}^{em}$  of W214 in the E forms at pH 2.2 and pH 1.0, respectively, shifts to  $\sim 339 \text{ nm}$  and  $\sim 338 \text{ nm}$  from  $\sim 333 \text{ nm}$ , when it is excited at 305 nm (at the red-edge of the excitation spectrum) compared to 295 nm (Figure 2.9A and 2.9B). In contrast, the  $\lambda_{max}^{em}$  of W214 remains constant at  $\sim 348 \text{ nm}$  in the U form (Figure 2.9A (inset) and 2.9B) and changes nominally from  $\sim 342 \text{ nm}$  to  $\sim 344 \text{ nm}$  in the N form (Figure 2.9B), when the excitation wavelength is changed from 295 nm to 305 nm.

REES is a special case of fluorescence emission which occurs for polar fluorophores when (a) there exist a heterogeneous distribution of solvent dipoles around the fluorophore resulting in a broad distribution of solvent-fluorophore interaction energies, and (b) the relaxation of solvent dipoles is slower than the fluorescence lifetime of the fluorophore, as in

a highly viscous solvation environment.<sup>39,53</sup> In the U form, no REES is observed because W214 is surrounded by highly dynamic water molecules. In the N form, W214 is located in a partially exposed hydrophobic cavity. The small magnitude ( $\sim 2$  nm) of REES observed in the N form implies that the complex solvation environment around W214 is less dynamic than the U form. In the E forms, W214 is buried in the hydrophobic core of domain II (Figures 2.7B and 2.7D). Hence, the large magnitude ( $\sim 5$ -6 nm) of REES observed in the E form indicates that in this dry globular form the solvation environment created by the dipoles of protein matrix is highly heterogeneous and is similar to a molten and viscous liquid.<sup>53</sup>

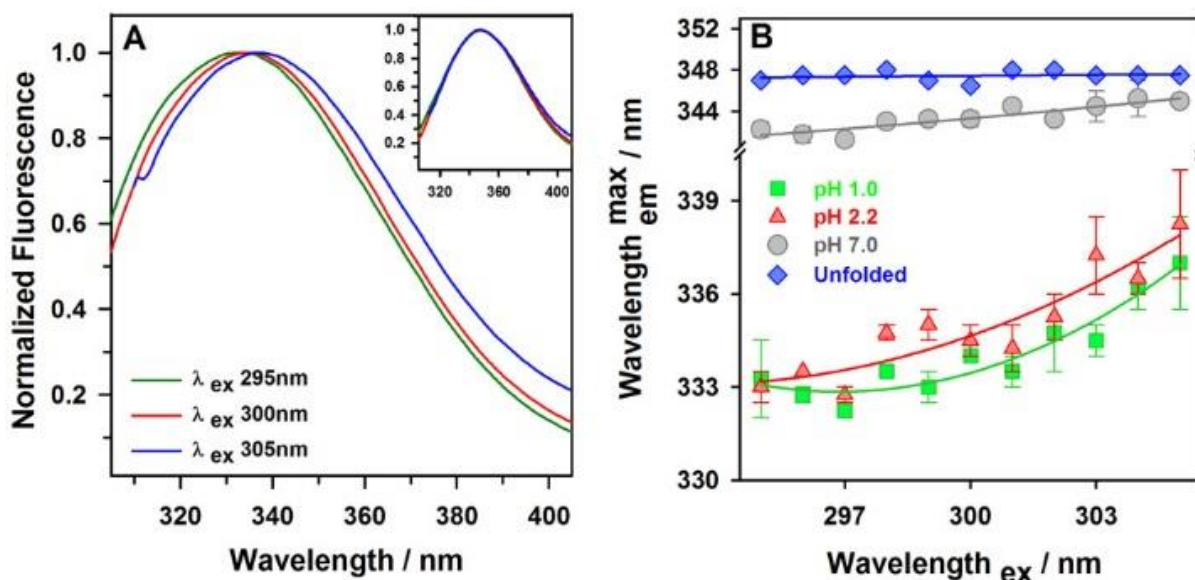


Figure 2.9. Dependence of wavelength of maximum fluorescence emission ( $\lambda_{max}^{em}$ ) of W214 on the wavelength of excitation ( $\lambda_{ex}$ ) as a function of pH. Panel A and its inset, respectively, show the representative fluorescence emission scan of W214 at pH 2.2 and in the U form when excited at different  $\lambda_{ex}$ . For comparison, each spectrum in A and its inset has been normalized to the value of 1 at its emission maximum. (B)  $\lambda_{max}^{em}$  is plotted as a function of  $\lambda_{ex}$ . In panel B, the solid lines through the data are drawn to guide the eye.

### 2.3.6 Global structural comparison of the N and the E form

We compared the global secondary structure and side-chain packing in the N form and the E form using far-UV and near-UV CD spectrum (Figure 2.10A and 2.10B). The mean residue ellipticity (MRE) at 222 nm, a measure of the  $\alpha$ -helical content in proteins, is  $-28350 \pm 3550$  deg cm<sup>2</sup> dmol<sup>-1</sup> in the N form,  $-6500 \pm 1000$  deg cm<sup>2</sup> dmol<sup>-1</sup> in the U form, and  $-21400 \pm 2000$  deg cm<sup>2</sup> dmol<sup>-1</sup> in the E forms at pH 2.2 and pH 1.0. These results indicate that E forms

retain ~68 % of the  $\alpha$ -helical content of the N form. It has been demonstrated previously that domain III gets fully unfolded in the E form,<sup>45</sup> which apparently corresponds to melting of roughly one-third of the protein structure and contributes to ~32 % change in the hydrogen-bonded secondary structure during N $\rightleftharpoons$ E transition. The changes in the fluorescence of W214 during pressure-induced denaturation also reveal that domain III of HSA unfolds prior to domain II.<sup>54</sup>

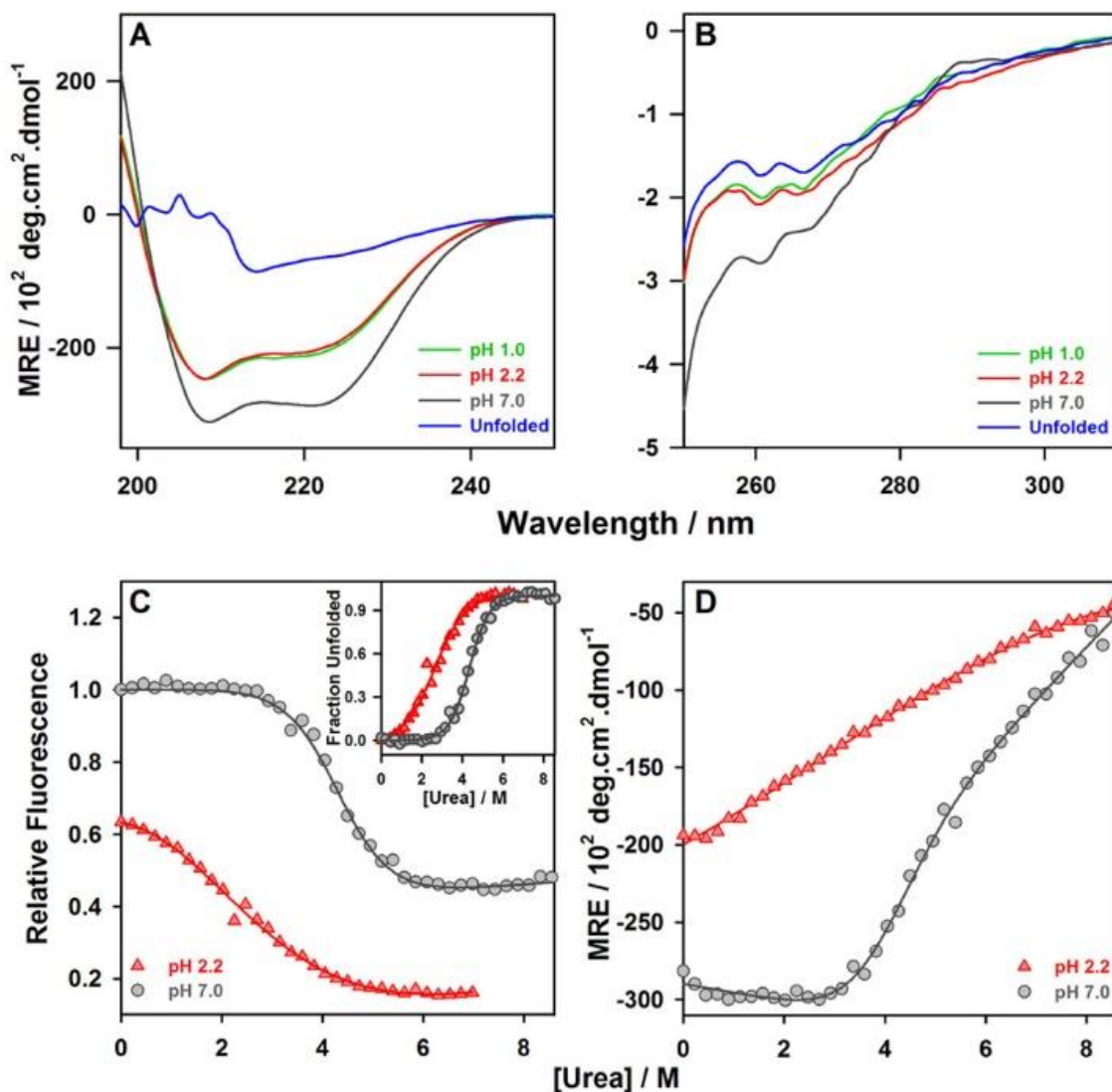


Figure 2.10. Changes in the secondary (A) and the tertiary structure (B) and the changes in the stability (C, D) of HSA as a function of pH. Urea-induced equilibrium unfolding transitions of the N form and the E form monitored by the change in fluorescence of W214 at 340 nm (C) and by the change in far-UV CD signal at 222 nm (D). In the inset of panel C, fraction of unfolded protein is plotted against [Urea]. The solid lines in panel C (and its inset) and through pH 7.0 data in panel D are fits to a two-state model.<sup>38</sup> The solid red line through pH 2.2 data in panel D is drawn to guide the eye.

There are 31 phenyl alanine residues packed tightly in the hydrophobic core of the N form and the near-UV CD spectrum appears to be dominated by their chirality. The observed near-UV CD spectrum has fine structures mainly in the 255 nm to 270 nm region (Figure 2.10B). The MRE at 261 nm in the N form is  $-275 \pm 25 \text{ deg cm}^2 \text{ dmol}^{-1}$ . The MRE at 261 nm in the E forms at pH 2.2 and pH 1.0 are  $-200 \pm 15 \text{ deg cm}^2 \text{ dmol}^{-1}$  which is very similar to the value of  $-175 \pm 15 \text{ deg cm}^2 \text{ dmol}^{-1}$  in the U form. As expected for a dry molten globular structure, these results indicate that the core packing interactions are severely disrupted in the E form compared to the N form.

### 2.3.7 vdW interactions contributes in the stability of the native state

In the E forms, domain I and II retain their secondary structure, have dry cores, but vdW packing interactions are severely disrupted. We used these observations to estimate the relative contributions of vdW interactions and hydrophobicity of the protein cores, in domain I and II, in the stability of the N form (see the model energy diagram in Figure 2.11). We measured the thermodynamic stabilities of the N form and the E form by the urea-induced equilibrium unfolding transitions at pH 7.0 and pH 2.2 which, respectively, represent  $N \rightleftharpoons U$  and  $E \rightleftharpoons U$  transitions, using fluorescence (Figure 2.10C) and far-UV CD (Figure 2.10D). For  $N \rightleftharpoons U$  transition monitored by both fluorescence and far-UV CD, the equilibrium values, respectively at 340 nm and 222 nm, show a cooperative or sigmoidal dependence on the concentrations of urea. The values of the free energy of unfolding ( $\Delta G_{N-U}$ ) and the slope of the transition ( $m_{N-U}$ ; which represents the change in the solvent accessible surface area), as extracted from fitting the data to a two-state equation<sup>38</sup> are  $5.2 \text{ kcal mol}^{-1}$  and  $1.21 \text{ kcal mol}^{-1} \text{ M}^{-1}$  for fluorescence monitored transition and  $5.1 \text{ kcal mol}^{-1}$  and  $1.24 \text{ kcal mol}^{-1} \text{ M}^{-1}$  for far-UV CD monitored transition.

The E form appears to be severely destabilized compared to the N form. The  $E \rightleftharpoons U$  transition monitored by fluorescence appears sigmoidal and the values of  $\Delta G_{E-U}$  and  $m_{E-U}$ , extracted from the two-state fit, are  $2.1 \text{ kcal mol}^{-1}$  and  $0.82 \text{ kcal mol}^{-1} \text{ M}^{-1}$  (Figure 2.10C). These values can be used to determine  $\Delta G_{N-E}$  ( $\Delta G_{N-E} = \Delta G_{N-U} - \Delta G_{E-U}$ ) and  $m_{N-E}$  ( $m_{N-E} = m_{N-U} - m_{E-U}$ ). The value of  $m_{N-E}$  is  $0.39 \text{ kcal mol}^{-1} \text{ M}^{-1}$ . It is important to note that the value of  $m_{N-E}$  corresponds to  $\sim 32\%$  of  $m_{N-U}$ , indicating that  $\sim 32\%$  of the protein core is accessible to water in the E form due to the unfolding of domain III, which is in a very good agreement with the results of far-UV CD experiments (Figure 2.10A). Hence, the change in  $m$ -value during  $N \rightleftharpoons E$  transition is not due to the water solvation of domain I and domain II. The value of  $\Delta G_{N-E} =$



3.1 kcal mol<sup>-1</sup>. Assuming an intermediate state E\* (see Figure 2.11) in which domain III is completely solvated and unfolded but side-chain packing interactions in domain I and II remain intact,  $\Delta G_{N-E}$  can be divided into two parts: (a)  $\Delta G_{N-E^*}$ , due to the unfolding of domain III during N $\rightleftharpoons$ E\* transition; and (b)  $\Delta G_{E^*-E}$ , due to the disruption of vdW packing interactions in domain I and II during E\* $\rightleftharpoons$ E transition. If we assume that N form loses roughly ~32 % of its stabilization energy ( $\Delta G_{N-U}$ ) due to the unfolding of domain III during N $\rightleftharpoons$ E\* transition, as suggested by the *m*-value and the far-UV CD measurements (see above), then the value of  $\Delta G_{N-E^*}$  will be 1.7 kcal mol<sup>-1</sup> and the value of  $\Delta G_{E^*-E}$  will be 1.4 kcal mol<sup>-1</sup>. The total contribution of domain I and II in the stability of the N form will be, then, given by the sum of  $\Delta G_{E^*-E}$  and  $\Delta G_{E-U}$  which is 3.5 kcal mol<sup>-1</sup>. A comparison of  $\Delta G_{E^*-E}$  with  $\Delta G_{E-U}$  (which is mainly due to the hydrophobic solvation of the cores of domain I and II) reveals that vdW packing and hydrophobic desolvation in domain I and II contribute in a 40:60 ratio in the thermodynamic stability of the N form.

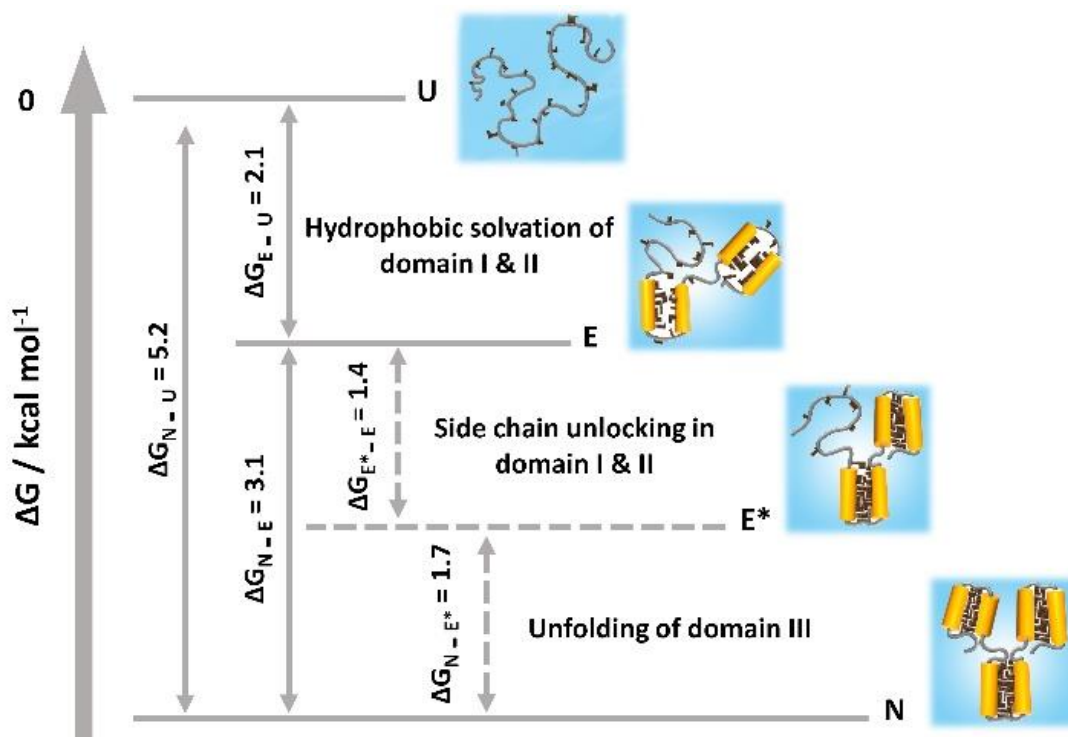


Figure 2.11. Model energy diagram.

We observed that the secondary structure in the E form unfolds in a gradual and non-cooperative fashion (Figure 2.10D). This result suggests that (a) DMG might not be discrete states but ensemble of loosely packed forms and (b) the cooperativity of the structural

transition observed during the folding reaction of proteins is closely associated with the development of the side-chain packing interactions.<sup>55,56</sup> This result also supports the phase diagram of the previous theoretical studies<sup>21,22</sup> which predicted that the transition of DMG to the fully unfolded state upon entry of water into protein interior is a continuous swelling process.

It is important to note that DMG-like expanded state was originally postulated to be the hypothetical transition state during the unfolding reaction of proteins.<sup>21,22</sup> However, here and in a few other studies they have been observed as unfolding intermediates,<sup>27,32,57</sup> which probably lie on the native-side of the free-energy barrier. DMG-like intermediates have been observed when the free-energy landscape of proteins have been modulated by denaturants,<sup>23-26,29</sup> pressure<sup>31,58</sup> and pH.<sup>19</sup> For the small protein HP35, DMG-like state has been shown to exist in equilibrium with the native state.<sup>30</sup> It has also been observed that the unlocked and conformationally flexible DMG-like state of HP35 is compact in volume than the native state but the transition state for native to DMG transition is expanded.<sup>59</sup> Recent NMR studies on a mimic of the initial kinetic intermediate of RNase H suggest that the intermediate states with dry cores can probably also form on the unfolded-side of the free-energy barrier.<sup>60</sup> DMG-like states have also been observed in molecular dynamic simulations of folding transitions of model hydrocarbon chains.<sup>61</sup> In total, above studies appear to suggest that DMG-like intermediate states might be general feature of the free-energy landscape of protein folding, as predicted in a recent theoretical study.<sup>62</sup>

## 2.4 Conclusions

We have shown that the domain I and domain II of HSA resemble a DMG in the low pH state of the protein implying that DMG-like characteristics can be localized in a multi-domain protein. We observed that (a) the domains are expanded at low pH compared to the native protein and that the side-chain packing is disrupted, but the hydrophobic core is not solvated by water; (b) the hydrophobic core of domain I exhibits ns structural fluctuation and that of domain II resembles a viscous and molten liquid in the DMG-like state; (c) the secondary structure in the DMG-like state unfolds in a gradual manner; (d) side-chain packing interactions contribute significantly to the protein folding cooperativity; and (e) the vdW interactions and hydrophobic effect in domain I and domain II relatively contribute in a 40:60 ratio towards the stability of the native state, indicating that side-chain packing plays a significantly important role in protein stability. Our observation of DMG-like intermediate

during the pH-induced unfolding has an important implication for the mechanism of action of protons in unfolding proteins. According to the traditional mechanism, the electrostatic repulsion created by the protonation of surface residues results in protein swelling, side-chain unlocking and water penetration into the hydrophobic core in a single concerted step.<sup>2,3</sup> Our results, however, indicate that rupture of tight side-chain packing interactions and hydrophobic solvation are two distinct structural events during the pH-induced unfolding of proteins. DMG-like intermediates are usually not very stable under equilibrium conditions. In the case of HSA, it appears that the loss in enthalpy due to disruption of side-chain packing interactions is fairly compensated by the gain in conformational entropy of the protein, providing enough stability to the DMG-like state to be populated at equilibrium. Finally, the results of this study have important implications for signaling and other proteins where the function is dictated by entropy-driven conformational changes.<sup>32,63-65</sup>

## 2.5 References

1. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14, 1-63.
2. Tanford, C. (1968) Protein denaturation. *Adv Protein Chem* 23, 121-282.
3. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24, 1-95.
4. Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv Protein Chem* 33, 167-241.
5. Sandberg, W. S., and Terwilliger, T. C. (1989) Influence of interior packing and hydrophobicity on the stability of a protein. *Science* 245, 54-57.
6. Kellis, J. T., Jr., Nyberg, K., and Fersht, A. R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* 28, 4914-4922.
7. Behe, M. J., Lattman, E. E., and Rose, G. D. (1991) The protein-folding problem: the native fold determines packing, but does packing determine the native fold? *Proc Natl Acad Sci USA* 88, 4195-4199.
8. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255, 178-183.

9. Honig, B., and Yang, A. S. (1995) Free energy balance in protein folding. *Adv Protein Chem* 46, 27-58.
10. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins. *FASEB J* 10, 75-83.
11. Bhattacharyya, S., and Varadarajan, R. (2013) Packing in molten globules and native states. *Curr Opin Struct Biol* 23, 11-21.
12. Baldwin, R. L. (2014) Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proc Natl Acad Sci USA* 111, 13052-13056.
13. Richards, F. M. (1977) Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 6, 151-176.
14. Ohgushi, M., and Wada, A. (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett* 164, 21-24.
15. Ikeguchi, M., Kuwajima, K., Mitani, M., and Sugai, S. (1986) Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: a comparative study of the folding reactions of alpha-lactalbumin and lysozyme. *Biochemistry* 25, 6965-6972.
16. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249, 1544-1548.
17. Colon, W., and Roder, H. (1996) Kinetic intermediates in the formation of the cytochrome c molten globule. *Nat Struct Biol* 3, 1019-1025.
18. Raschke, T. M., and Marqusee, S. (1997) The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol* 4, 298-304.
19. Rami, B. R., and Udgaonkar, J. B. (2002) Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* 41, 1710-1716.
20. Prajapati, R. S., Indu, S., and Varadarajan, R. (2007) Identification and thermodynamic characterization of molten globule states of periplasmic binding proteins. *Biochemistry* 46, 10339-10352.

21. Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28, 1667-1680.
22. Finkelstein, A. V., and Shakhnovich, E. I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28, 1681-1694.
23. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
24. Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-<sup>19</sup>F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proceedings of the National Academy of Sciences* 92, 9318-9322.
25. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci USA* 106, 12289-12294.
26. Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci USA* 111, 4856-4861.
27. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725-2737.
28. Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* 118, 6380-6392.
29. Sarkar, S. S., Udgaonkar, J. B., and Krishnamoorthy, G. (2013) Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* 105, 2392-2402.
30. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci USA* 107, 4955-4960.
31. Fu, Y., Kasinath, V., Moorman, V. R., Nucci, N. V., Hilser, V. J., and Wand, A. J. (2012) Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* 134, 8543-8550.

32. Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23, 4-10.
33. Painter, L., Harding, M. M., and Beeby, P. J. (1998) Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin. *J Chem Soc, Perkin Trans 1*, 3041-3044.
34. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.
35. Jha, A., Udgaonkar, J. B., and Krishnamoorthy, G. (2009) Characterization of the heterogeneity and specificity of interpolypeptide interactions in amyloid protofibrils by measurement of site-specific fluorescence anisotropy decay kinetics. *J Mol Biol* 393, 735-752.
36. Lillo, M. P., Beechem, J. M., Szpikowska, B. K., Sherman, M. A., and Mas, M. T. (1997) Design and characterization of a multisite fluorescence energy-transfer system for protein folding studies: a steady-state and time-resolved study of yeast phosphoglycerate kinase. *Biochemistry* 36, 11261-11272.
37. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248-254.
38. Agashe, V. R., and Udgaonkar, J. B. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* 34, 3286-3299.
39. Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, Springer, Singapore.
40. Chen, R. F. (1967) Fluorescence quantum yields of tryptophan and tyrosine. *Anal Lett* 1, 35-42.
41. Amiri, M., Jankeje, K., and Albani, J. R. (2010) Origin of fluorescence lifetimes in human serum albumin. Studies on native and denatured protein. *J Fluoresc* 20, 651-656.
42. Saito, Y., Tachibana, H., Hayashi, H., and Wada, A. (1981) Excitation-Energy Transfer between Tyrosine and Tryptophan in Proteins Evaluated by the Simultaneous Measurement of Fluorescence and Absorbance. *Photochem Photobiol* 33, 289-295.

43. Swaminathan, R., Krishnamoorthy, G., and Periasamy, N. (1994) Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state. *Biophys J* 67, 2013-2023.
44. Amiri, M., Jankeje, K., and Albani, J. R. (2010) Characterization of human serum albumin forms with pH. Fluorescence lifetime studies. *J Pharm Biomed Anal* 51, 1097-1102.
45. Dockal, M., Carter, D. C., and Ruker, F. (2000) Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *J Biol Chem* 275, 3042-3050.
46. Carter, D. C., and Ho, J. X. (1994) Structure of serum albumin. *Adv Protein Chem* 45, 153-203.
47. Era, S., and Sogami, M. (1998) <sup>1</sup>H-NMR and CD studies on the structural transition of serum albumin in the acidic region--the N-->F transition. *J Pept Res* 52, 431-442.
48. Olivieri, J. R., and Craievich, A. F. (1995) The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering. *Eur Biophys J* 24, 77-84.
49. Krishnakumar, S. S., and Panda, D. (2002) Spatial relationship between the prodan site, Trp-214, and Cys-34 residues in human serum albumin and loss of structure through incremental unfolding. *Biochemistry* 41, 7443-7452.
50. Huang, F., Settanni, G., and Fersht, A. R. (2008) Fluorescence resonance energy transfer analysis of the folding pathway of Engrailed Homeodomain. *Protein Eng Des Sel* 21, 131-146.
51. Eftink, M. R., and Ghiron, C. A. (1975) Dynamics of a protein matrix revealed by fluorescence quenching. *Proceedings of the National Academy of Sciences* 72, 3290-3294.
52. Strambini, G. B., and Gonnelli, M. (2010) Fluorescence quenching of buried Trp residues by acrylamide does not require penetration of the protein fold. *J Phys Chem B* 114, 1089-1093.
53. Demchenko, A. P. (2008) Site-selective Red-Edge effects. *Methods Enzymol* 450, 59-78.

54. Tanaka, N., Nishizawa, H., and Kunugi, S. (1997) Structure of pressure-induced denatured state of human serum albumin: a comparison with the intermediate in urea-induced denaturation. *Biochim Biophys Acta* 1338, 13-20.
55. Jha, S. K., and Udgaonkar, J. B. (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* 282, 37479-37491.
56. Jha, S. K., Dhar, D., Krishnamoorthy, G., and Udgaonkar, J. B. (2009) Continuous dissolution of structure during the unfolding of a small protein. *Proceedings of the National Academy of Sciences* 106, 11113-11118.
57. Hua, L., Zhou, R., Thirumalai, D., and Berne, B. J. (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc Natl Acad Sci USA* 105, 16928-16933.
58. de Oliveira, G. A., and Silva, J. L. (2015) A hypothesis to reconcile the physical and chemical unfolding of proteins. *Proceedings of the National Academy of Sciences* 112, E2775-2784.
59. Neumaier, S., and Kiefhaber, T. (2014) Redefining the dry molten globule state of proteins. *J Mol Biol* 426, 2520-2528.
60. Rosen, L. E., Connell, K. B., and Marqusee, S. (2014) Evidence for close side-chain packing in an early protein folding intermediate previously assumed to be a molten globule. *Proc Natl Acad Sci U S A* 111, 14746-14751.
61. Mountain, R. D., and Thirumalai, D. (2003) Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc* 125, 1950-1957.
62. Thirumalai, D., Liu, Z., O'Brien, E. P., and Reddy, G. (2013) Protein folding: from theory to practice. *Curr Opin Struct Biol* 23, 22-29.
63. Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007) Conformational entropy in molecular recognition by proteins. *Nature* 448, 325-329.
64. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G., and Wand, A. J. (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6, 352-358.



65. Tzeng, S. R., and Kalodimos, C. G. (2012) Protein activity regulation by conformational entropy. *Nature* 488, 236-240.

## Chapter 3.

# A Dry Molten Globule-like Intermediate During the Base-induced Unfolding of a Multidomain Protein

*Reproduced from the permission of Acharya, N., Mishra, P., and Jha, S. K. (2017) A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. Phys Chem Chem Phys 19, 30207-30216.*

### 3.1 Introduction

Basic pH leads to partial or complete unfolding of proteins and is an important factor in protein function and many industrial and biotechnological applications of proteins.<sup>1-9</sup> However, we do not understand the molecular mechanism and sequence of structural events that initiate the base-induced unfolding of proteins. Two different models have been proposed to describe how proteins begin to unfold. The traditional model envisages protein unfolding as a cooperative process where the dissolution of the native (N) structure occurs concomitantly with the penetration of water molecules in the hydrophobic core.<sup>10-13</sup> The alternative model, based upon the theoretical phase diagram of proteins, postulates that unfolding begins with the expansion of the protein into a dry molten globule (DMG) state in which the van der Waals (vdW) packing interactions are broken but the hydrophobic core is dehydrated and the N-like main-chain hydrogen bonding and overall tertiary structure are retained.<sup>14-19</sup> According to this model, core hydration and global structural dissolution occur in the second and final step of protein unfolding. However, experimental evidence for a DMG-like intermediate state during the base-induced unfolding of proteins has been rare.<sup>20</sup>

Human serum albumin (HSA) is the most abundant protein in human blood plasma and performs multiple functions including ligand transport<sup>21</sup> and maintenance of blood pH.<sup>22</sup> It is a well characterized three domain (I, II and III) protein and has been known to undergo a physiologically important base-induced structural transition, from the N to the basic (B) form, for the last five decades.<sup>21,23,24</sup> The B form of serum albumins impact their ligand binding properties and has physiological importance in the transport of many medicinal drugs in blood.<sup>3,24,25</sup> Surprisingly, the details of the inner molecular structure of the B form of HSA and the mechanism of its formation are very poorly understood. One major reason is that the B form is so similar in structure to the N form that the structural differences are silent to most of the global spectroscopic probes of protein structure. Various fundamental questions remain unanswered. For example, we do not understand (i) what is the nature of the structural difference between the B form and the N form? (ii) whether the B form resembles a DMG-like state? (iii) whether the structural loosening and core solvation are coupled during  $N \rightleftharpoons B$  transition? (iv) how different are the side-chain packing interactions in the core of the B form compared to the N form? (v) how the small structural differences between the N and the B form important for function?

In this study, we have used a battery of site-specific probes to investigate the base-induced partial unfolding of HSA. The simultaneous application of multiple spectroscopic probes, including fluorescence resonance energy transfer (FRET), size exclusion chromatography, dynamic fluorescence quenching, red-edge excitation shift (REES) and near- and far-UV circular dichroism (CD) allowed us to site-specifically unveil the nature of structural events that initiate the formation of the B form and its structure in terms of changes in inter-atomic distances, side-chain packing, protein expansion, core hydration, core flexibility, and global secondary and tertiary structure. Previous studies have suggested that in the B form of HSA, domain III remains fully structured, but there are subtle tertiary structural changes in domain I and II.<sup>26</sup> Here, we report that in the B form, domain I and II of HSA have DMG-like characteristics near the inter-domain interface. Our results shed important light on the mechanism of the base-induced protein unfolding and suggest that the vdW packing interactions might be an important factor in protein stability.

## 3.2 Materials and Methodology

### Reagents, chemicals, buffers and experimental conditions

Human serum albumin (99% pure, fatty acid free, globulin free) was procured from Alfa Aesar and used without further purification. The concentration of HSA was determined spectroscopically by measuring the absorbance at 280 nm, using the molar extinction coefficient of  $36500 \text{ M}^{-1}\text{cm}^{-1}$ .<sup>27</sup> Urea (ACS grade,  $\geq 99\%$  pure) was obtained from Alfa Aesar and 5-(((2-iodoacetyl)amino)ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS) from Life Technologies. Acrylamide, Guanidine hydrochloride (GdmCl) and other chemicals and reagents were purchased from Sigma.

For pH titration study, a universal buffer (20 mM sodium acetate, 20 mM sodium phosphate and 20 mM sodium borate) was used in the pH range 5 to 12. The pH was adjusted with the addition of HCl or NaOH. For all other experiments, buffers for the native (N) form at pH 7 and for the basic (B) form at pH 11 were composed of 20 mM sodium phosphate. The unfolding buffers at pH 7 and pH 11 were composed of 20 mM sodium phosphate and 9 M urea. All the buffers were filtered through 0.2  $\mu\text{m}$  filters before use. All experiments were performed at the room temperature unless otherwise specified. The concentration of urea and GdmCl solutions were determined by measuring the refractive index.<sup>28</sup>

---

## Spectroscopic instruments

All the absorption spectra were acquired on a Perkin Elmer Lambda 650 UV/Vis spectrometer using a quartz cell of 1 cm path length. All the fluorescence spectra were collected on a Perkin Elmer fluorescence spectrometer LS 55 using a fluorescence quartz cell of 1 cm path length. The collection of all the circular dichroism (CD) spectra were performed on a Jasco J-815 CD spectrometer. CD compatible quartz cells of 0.1 cm and 1 cm path length were used for far- and near-UV CD, respectively. The fluorescence lifetime decays were measured on a time-correlated single photon counting (TCSPC) setup, Deltaflex, from Horiba Scientific. The refractive index of buffers, wherever required, were measured on an Abbe refractometer from Rajdhani Scientific Instruments.

## Preparation of IAEDANS labeled HSA

HSA was labeled at the free cysteine residue, C34, with 1,5-IAEDANS as described previously<sup>29</sup>. The extent of labeling was calculated by estimating the concentration of total protein (HSA + HSA-IAEDANS) and the labeled protein (HSA-IAEDANS) spectroscopically. We determined the concentration of HSA-IAEDANS from the absorbance of C34-IAEDANS at 337 nm using a molar extinction coefficient of  $4500 \text{ M}^{-1} \text{ cm}^{-1}$ .<sup>29</sup> The total protein concentration was determined by measuring the absorbance of the protein at 280 nm, using the extinction coefficient of  $36500 \text{ M}^{-1} \text{ cm}^{-1}$ . The extinction coefficient of IAEDANS at 280 nm is  $\sim 1000 \text{ M}^{-1} \text{ cm}^{-1}$ .<sup>30</sup> Since the extinction coefficient of IAEDANS dye at 280 nm is nearly 35 times less than the extinction coefficient of HSA, the contribution of dye in the 280 nm absorbance will be negligible ( $\sim 3\%$ ) if the protein is fully labeled. The measured extent of labeling of the labeled protein was  $>95\%$ . It is important to note that our assumption is conservative and gave the lower limit of the amount of labeling and the actual percentage will be 2-3% larger than this.

## pH titration monitored by fluorescence and CD

The protein samples were incubated at different pH and equilibrated by keeping overnight at the room temperature. For the unlabeled protein, W214 was excited using the excitation wavelength of 295 nm and the emission spectra were collected from 310 to 420 nm. For HSA-IAEDANS, the protein samples were excited at 337 nm and the emission spectra were collected from 350-550 nm. All the fluorescence spectra were measured with excitation slit of 5-7 nm and scan speed of 100 nm/min. All the far-UV CD and the near-UV CD spectra were respectively measured in the 190-250 nm and 250-300 nm wavelength range. All the CD

spectra were collected with a bandwidth of 2 nm and the scan speed of 100 nm/min. The protein concentration used for the fluorescence and the far-UV CD experiments was 4-5  $\mu\text{M}$  and for the near-UV CD experiments, it was 20  $\mu\text{M}$ .

For fluorescence monitored pH titration, the change in fluorescence signal at 340 nm was observed to occur in a single sigmoidal step. The data was fitted to a model, in which the structural transition from the N form to the B form is coupled to a single deprotonation step, given by a transformed Henderson-Hasselbalch equation,<sup>31</sup>

$$Y_{obs} = \frac{Y_P + Y_D 10^{(pH-pH_m)}}{1 + 10^{(pH-pH_m)}} \quad [1]$$

where,  $Y_{obs}$  denotes the observed fluorescence signal for a given pH value,  $Y_P$  and  $Y_D$  are the signals of the protonated and the deprotonated forms and  $pH_m$  is the midpoint of the transition.

### FRET between W214 and C34-IAEDANS and the analysis of FRET data

For FRET measurements, W214 in HSA and HSA-IAEDANS was excited at 295 nm and all the emission spectra were collected identically. The FRET distance ( $R$ ) between W214 and C34-IAEDANS was calculated from FRET efficiency ( $E$ ) and Forster's distance ( $R_0$ ) as described in the results and discussion section. The Forster's distance ( $R_0$ ) between the FRET pair in each condition was calculated using the equation,<sup>32</sup>

$$R_0 = 0.211 [Q_D J \kappa^2 n^{-4}]^{\frac{1}{6}} \quad [2]$$

where,  $Q_D$  denotes for quantum yield of the donor,  $J$  represents the overlap integral between donor's emission spectrum and acceptor's excitation spectrum,  $\kappa^2$  is the orientation factor of the fluorophores and  $n$  represents the refractive index of the medium.<sup>32</sup> We experimentally determined the values of  $Q_D$  as described previously.<sup>29</sup> The value of  $Q_D$  in the N form and the B form is 0.31 and 0.23, respectively. The overlap integral  $J$  was calculated by,

$$J = \frac{\int F(\lambda) \varepsilon(\lambda) \lambda^4 d\lambda}{\int F(\lambda) d\lambda} \quad (\text{M}^{-1} \text{cm}^{-1} \text{nm}^4) \quad [3]$$

In the above equation,  $F(\lambda)$  and  $\varepsilon(\lambda)$  represents, fluorescence emission spectra of W214 and absorbance spectra of C34-IAEDANS (Figure 3.1 and Table 3.1). The orientation of the donor and the acceptor fluorophore was assumed to be random and hence the value of  $\kappa^2$  was taken as 2/3. The observed value of  $n$  for the N form at pH 7 and the B form at pH 11 was 1.332. The calculated values of Forster's distance and FRET distance between W214 and C34-IAEDANS, along with other FRET parameters are given in Table 3.1.

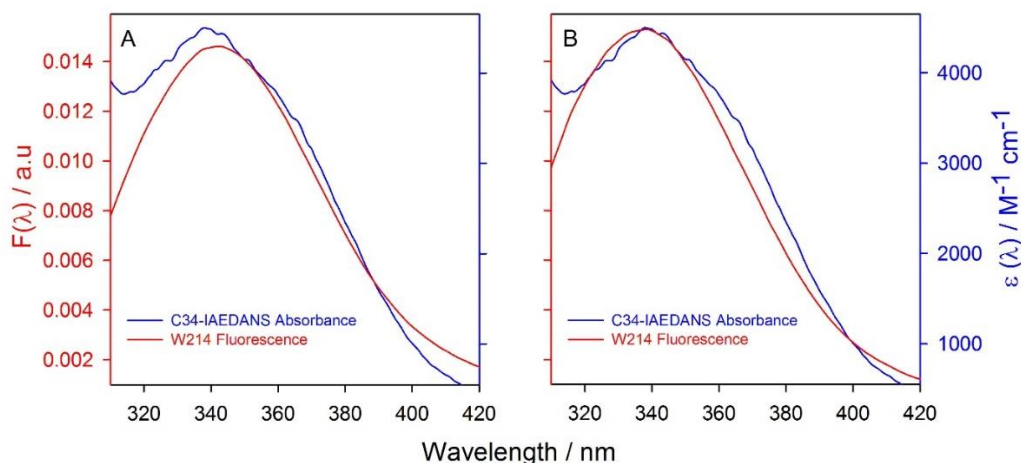


Figure 3.1. Determination of the overlap integral,  $J$ , in the N form (A) and in the B form (B). In both the panels, the left y-axis represents the fluorescence emission spectra ( $F(\lambda)$ ) of W214 and the right y-axis represents absorbance spectra ( $\epsilon(\lambda)$ ) of C34-IAEDANS, respectively. The total area under each fluorescence emission spectrum has been normalized to unity. Each absorbance spectrum has been divided by the respective molar protein concentration to obtain  $\epsilon(\lambda)$ .

Table 3.1. Determination of FRET parameters and D-A distances. Values of FRET efficiency ( $E$ ), quantum yield ( $Q_D$ ), overlap integral ( $J$ ), Forster's distance ( $R_0$ ), and D-A distance ( $R$ ) for FRET between W214 and C34-IAEDANS pair.

| Conditions    | $^*E$           | $^{\S}Q_D$ | $^{\dagger}J/10^{13}\text{M}^{-1}\text{cm}^{-1}\text{nm}^4$ | $^{\ddagger}R_0/\text{\AA}$ | $^{\ddagger}R/\text{\AA}$ |
|---------------|-----------------|------------|---|-----------------------------|---------------------------|
| <b>N form</b> | $0.31 \pm 0.01$ | 0.31       | 5.1   | 25.8                        | $29.6 \pm 0.1$            |
| <b>B form</b> | $0.20 \pm 0.01$ | 0.23       | 6.1   | 24.5                        | $31.0 \pm 0.4$            |

$^*E$  was determined using the fluorescence spectra shown in Figure 3.5A and 3.5B as described in the result section using equation 8. The standard error was estimated from three separate measurements.

$^{\S}Q_D$  was calculated as described previously.<sup>29</sup>

$^{\dagger}J$  was calculated using equation 3.

$^{\ddagger}R_0$  was calculated using equation 2; with the following values:  $\kappa^2 = 2/3$ ;  $n = 1.332$ .

$^{\ddagger}R$  was calculated using equation 9. The standard error was estimated from three separate measurements.

### Size exclusion chromatography

All the size exclusion chromatography (SEC) experiments were performed on a GE AKTA Pure FPLC system. For SEC in the N form and the B form, HSA was dissolved in pH 7 and in pH 11 buffer, respectively. SEC was performed by loading 100  $\mu$ M protein onto GE Superdex 75 10/300 GL high performance gel filtration column, pre-equilibrated at the respective pH. Due to the high concentration of protein used for SEC, we observed 5-12% of dimer in the resultant chromatograms. For better comparison between the N form and the B form, the SEC chromatograms were fitted to a sum of two Gaussians and monomer peaks were compared.

For the determination of the apparent molecular weight of the N form and the B form, we constructed a calibration curve between the partition coefficient ( $K_{av}$ ) of five different standard biomolecules (vitamin B12, aprotinin, ribonuclease A, ovalbumin and bovine serum albumin) and their known molecular weights ( $M_w$ ) (Figure 3.2A), as described in the GE size exclusion chromatography handbook provided by the manufacturer. The values of  $K_{av}$  were calculated using the following equation,

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad [4]$$

where,  $V_e$  is the elution volume for the protein,  $V_0$  is the column void volume and  $V_t$  is the total bed volume. The values of  $V_0$  and  $V_t$  were taken from the column specification sheet. The  $V_e$  values of the standard biomolecules were taken from the manufacturer provided manual of GE Superdex 75 10/300 GL high performance gel filtration column. We then calculated the  $K_{av}$  values for the N form and the B form of HSA from their corresponding experimentally measured  $V_e$  values and determined the apparent molecular weight of the N form and the B form from the calibration curve.

Similarly, for the determination of the hydrodynamic radius of the N form and the B form, we constructed a calibration curve between  $K_{av}$  values of three different standard proteins (ribonuclease A, ovalbumin and bovine serum albumin) and their previously reported hydrodynamic radii<sup>33</sup> (Figure 3.2B). Using this calibration curve and the above calculated  $K_{av}$  values of the N form and the B form, we determined the hydrodynamic radius of the N form and the B form of HSA.



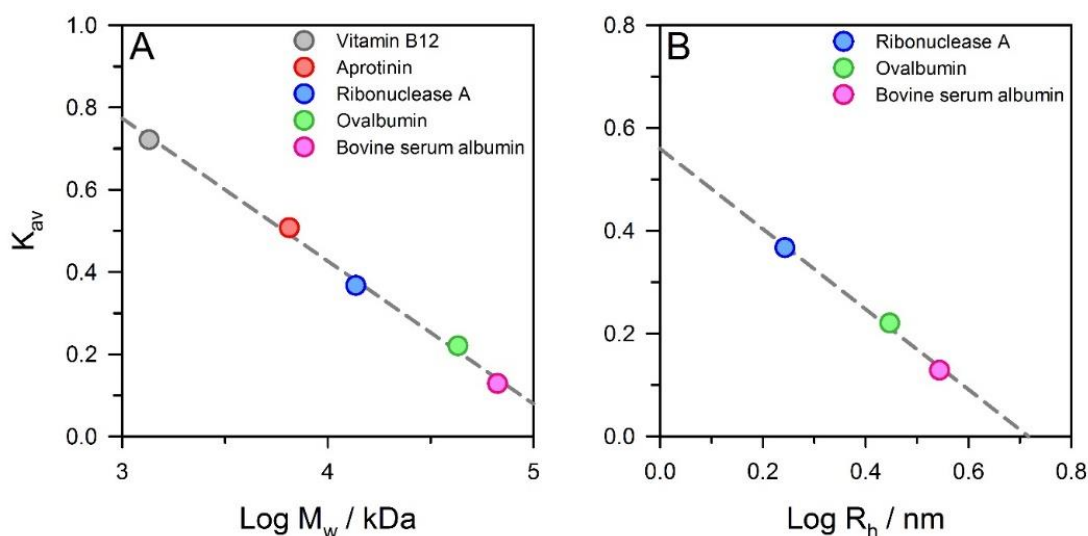


Figure 3.2. Calibration curves for the determination of the molecular weight ( $M_w$ ) and the hydrodynamic radius ( $R_h$ ) of the N form and the B form from size exclusion chromatography. Panel A shows the partition coefficient ( $K_{av}$ ) plotted against  $\log M_w$  of five different standard biomolecules. Panel B shows the  $K_{av}$  plotted against  $\log R_h$  of three different standard proteins. The  $K_{av}$  values for all the standards were calculated from the respective elution volumes, given in the manufacturer provided manual for GE Superdex 75 10/300 GL high performance gel filtration column, using equation 4. The  $R_h$  values in panel B for ribonuclease A, ovalbumin and bovine serum albumin were taken from a previous study.<sup>33</sup> The dashed lines in panel A and B are linear fits to the data and used to calculate the  $M_w$  and  $R_h$  values of the N form and the B form of HSA (Figure 3.5C (top and bottom inset), results).

### Acrylamide quenching experiments

HSA and HSA-IAEDANS were incubated at pH 7, pH 11 and their respective unfolding buffers. Appropriate volumes of acrylamide were added just before taking the fluorescence emission spectra from a freshly prepared 2 M acrylamide stock solution. For HSA, W214 was excited at 295 nm and the emission spectra were collected from 310 to 420 nm. For HSA-IAEDANS, C34-IAEDANS was excited at 337 nm and the emission spectra were collected from 350 to 550 nm.

The values of Stern-Volmer constant ( $K_{sv}$ ) and the intensity averaged fluorescence lifetime ( $\tau_0$ ) were used to determine the bimolecular quenching constant ( $k_q$ ) in each condition. For 9 M urea sample, a correction for viscosity was made, as described previously.<sup>29</sup> The values of  $K_{sv}$ ,  $\tau_0$  and  $k_q$  in different conditions are listed in Table 3.2.

Table 3.2. Parameters for dynamic fluorescence quenching experiments. Values of Stern-Volmer constants ( $K_{sv}$ ), intensity averaged fluorescence lifetimes ( $\tau_0$ ) and bimolecular quenching rate constants ( $k_q$ ) for C34-IAEDANS and W214.

| Conditions    | C34-IAEDANS     |             |                         | W214            |               |                         |
|---------------|-----------------|-------------|-------------------------|-----------------|---------------|-------------------------|
|               | $K_{sv}/M^{-1}$ | $\tau_0/ns$ | $k_q/10^8 M^{-1}s^{-1}$ | $K_{sv}/M^{-1}$ | * $\tau_0/ns$ | $k_q/10^9 M^{-1}s^{-1}$ |
| <b>N form</b> | 1.6             | 19.3        | 0.8                     | 6.1             | 7.1           | 0.9                     |
| <b>B form</b> | 1.6             | 18.9        | 0.8                     | 6.5             | 6.3           | 1.0                     |
| <b>U form</b> | 3.3             | 14.2        | 3.2 <sup>§</sup>        | 8.2             | 3.3           | 3.4 <sup>§</sup>        |

\* These values are taken from previous studies.<sup>34,35</sup>

<sup>§</sup> These values are corrected for the effects of viscosity of 9 M urea.

Note: Errors in the values of different parameters are less than 5%.

### Fluorescence lifetimes

The fluorescence lifetime decays of C34-IAEDANS in different conditions were measured using an excitation laser source of 370 nm (IRF is 196 ps) and collecting the emission decay at 480 nm (Figure 3.3).

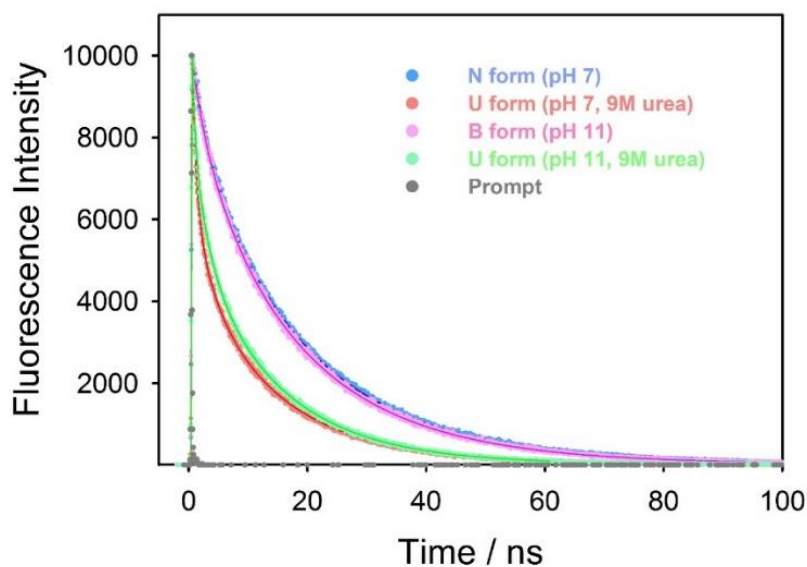


Figure 3.3. Fluorescence intensity decay kinetics of C34-IAEDANS.

The deconvolution and fitting of fluorescence decay into three exponentials were performed on DAS6 analysis software (Horiba Scientific). The intensity averaged fluorescence lifetimes ( $\tau_0$ ) (Table 3.2) and mean fluorescence lifetime ( $\tau_m$ ) were calculated using the following equations,

$$\tau_0 = \frac{\sum_i \alpha_i \tau_i^2}{\sum_i \alpha_i \tau_i} \quad [5]$$

$$\tau_m = \sum_i \alpha_i \tau_i \quad [6]$$

where,  $\alpha_i$  is the fractional amplitude of the respective fluorescence lifetime  $\tau_i$  ( $\sum_i \alpha_i = 1$ ). The values of  $\tau_0$  for W214 have been determined from the fluorescence lifetime measurements reported in previous studies.<sup>34,35</sup> These values are listed in Table 3.2.

### REES experiments

For HSA-IAEDANS, the fluorescence emission spectra of C34-IAEDANS were collected by exciting the fluorophore at different wavelengths ranging from 337 nm to 407 nm. For HSA, the fluorescence emission spectra were recorded by exciting W214 at different wavelengths ranging from 295 nm to 305 nm. For both HSA and HSA-IAEDANS, the concentration of protein used was 6-12  $\mu\text{M}$ . All the fluorescence emission spectra were recorded at a scan speed of 50 nm/min, averaged for three accumulations and subtracted for background buffer intensities.

### Urea-induced equilibrium unfolding transitions

For urea-induced equilibrium unfolding transitions, the protein samples were incubated for 3 hours at pH 7 and pH 11 in different concentration of urea. The concentration of protein used for the fluorescence and the far-UV CD experiments was 4  $\mu\text{M}$ . The fluorescence spectra of W214 were obtained by exciting W214 at 295 nm and collecting the emission spectra from 310 nm to 420 nm. The far-UV CD spectra of the same samples were measured in the wavelength range of 200 nm to 250 nm. The fluorescence and the far-UV CD monitored equilibrium unfolding transitions were analyzed using a two-state  $Z \rightleftharpoons U$  model<sup>36</sup> (Z is N at pH 7 and B at pH 11) and the data were fitted to the equation,

$$y_{obs} = \frac{y_Z + y_U e^{\frac{-\Delta G_{Z-U} + m_{Z-U}[D]}{RT}}}{1 + e^{\frac{-\Delta G_{Z-U} + m_{Z-U}[D]}{RT}}} \quad [7]$$

where,  $y_{obs}$  is the observed fluorescence/CD signal;  $y_Z$  and  $y_U$  are the signals of the native and the unfolded states, respectively;  $\Delta G_{Z-U}$  and  $m_{Z-U}$  are the standard free energy and the slope of the transition, respectively.

### 3.3 Results and Discussion

#### 3.3.1 Base-induced structural transition

We monitored the change in the tertiary and the secondary structure of the protein during the base-induced structural transition using intrinsic tryptophan fluorescence and far-UV CD. HSA has a single tryptophan residue, W214, which is located in the first helix of domain II and participates in the inter-domain side-chain packing interactions that hold the three domains together (Figure 3.4A). We observed that the fluorescence intensity of W214, reporting upon the local tertiary environment, changes in a sigmoidal fashion upon increasing the pH from 5 to 12 (Figure 3.4B), indicating that protein undergoes a base-induced structural transition coupled to the deprotonation of a single ionizable group. The midpoint of the transition is at pH 10.2. It has been reported that the N form of HSA transforms to the B form above pH 8 (N $\rightleftharpoons$ B transition)<sup>26</sup> and our sigmoidal pH transition data monitored by W214 fluorescence supports this conclusion.

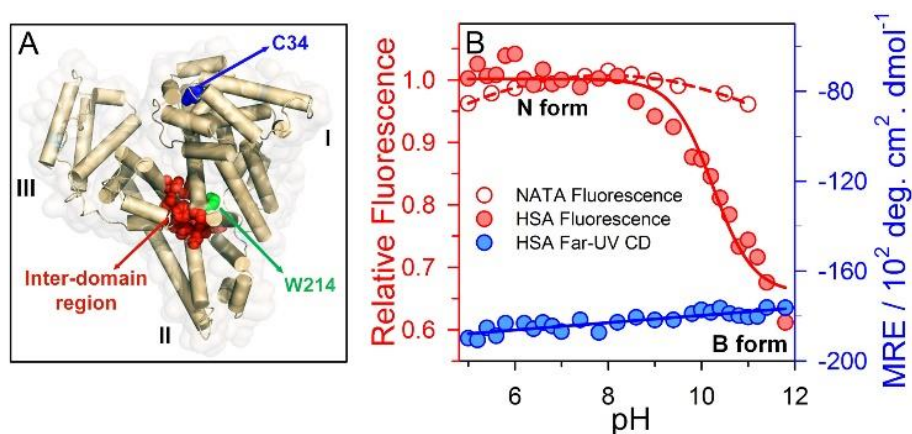


Figure 3.4. The base-induced N $\rightleftharpoons$ B structural transition is accompanied by a change in tertiary structure without alteration in secondary structure. (A) Structure of HSA, highlighting the free cysteine residue C34 (blue) in domain I, tryptophan residue W214 (green) in domain II and inter-domain region (red) drawn from PDB file 1AO6 using the program PyMOL. (B) The change in fluorescence emission of W214 and NATA at 340 nm (left y-axis) and the change in mean residual ellipticity at 222 nm (right y-axis) are plotted against pH. The solid red and blue

---

lines through the data are fits to a two-state pH transition model (equation 1) and a straight line, respectively. The dashed red line through the data is drawn to guide the eyes.

It is important to rule out that the quenching of W214 fluorescence at higher pH is not due to the deprotonation of the indole group or the solvent induced change in its fluorescence quantum yield. In order to do this, we compared the pH-dependent change in W214 fluorescence to the pH-dependent change in the fluorescence of an N-terminal and C-terminal blocked analogue of L-tryptophan, N-acetyl-L-tryptophanamide (NATA) (Figure 3.4B). We observed that the change in the fluorescence intensity of NATA upon increasing the pH from 7 to 11 is only 10-15% of the total change in W214 fluorescence in this pH range (Figure 3.4B). This comparison indicates that the observed change in W214 fluorescence upon increasing the pH is primarily due to the partial unfolding and disruption of the tertiary structure near W214. This result is in accordance with the results of a previous study on a few globular proteins, including serum albumin, which showed that the pH dependent changes in tryptophan fluorescence in the pH range of 7-12 are result of structural changes near the fluorophore.<sup>37</sup> Very interestingly, we observed that the secondary structure content of the protein, monitored by far-UV CD, does not change and remains almost constant during  $N \rightleftharpoons B$  transition (Figure 3.4B). Hence, the B form appears to be a state of HSA in which the secondary structure is similar to the N form but tertiary structure near W214 is disrupted. The experiments presented below compare the structural properties of the N form (HSA in pH 7 buffer) and the B form (HSA in pH 11 buffer).

### 3.3.2 Structural expansion in the B form

Because W214 is located in the inter-domain region at the interface of domain I and domain II, the disruption of tertiary structure near it will likely lead to the movement of domain I and domain II away from each other. To probe this further at the site-specific level, we measured the inter-domain distances between domain I and II in the N form and the B form using FRET methodology. For FRET measurements, we utilized W214 in domain II as the donor fluorophore. HSA has a single free cysteine, C34, located at the N-terminal of helix 3 in domain I (HSA contains 35 cysteine residues but 34 of them are involved in 17 disulfide bonds). We labeled C34 with the fluorescent dye 5-(((2-iodoacetyl)amino)-ethyl)amino)naphthalene-1-sulfonic acid (1,5-IAEDANS), which served as the acceptor of W214 fluorescence. We have reported earlier that C34 of HSA can be efficiently and quantitatively labeled with 1,5-IAEDANS and that the structure and thermodynamic stability

of the 1,5-IAEDANS labeled protein (HSA- IAEDANS) is very similar to the unlabeled protein (HSA).<sup>29</sup> Hence, the data on HSA and HSA- IAEDANS can be compared directly as required for the FRET measurements. The fluorescence emission spectrum of W214 overlaps with the absorbance spectrum of the C34-IAEDANS (Figure 3.1), and they form an efficient FRET pair.<sup>29</sup> We observed that in the N form of HSA-IAEDANS, where W214 and C34-IAEDANS are nearer to each other, the fluorescence of W214 is quenched significantly by C34-IAEDANS (Figure 3.5A). In contrast, the level of quenching markedly decreases in the B form than that in the N form (Figure 3.5B), indicating that the donor-acceptor (D-A) distance is larger in the B form compared to the N form.

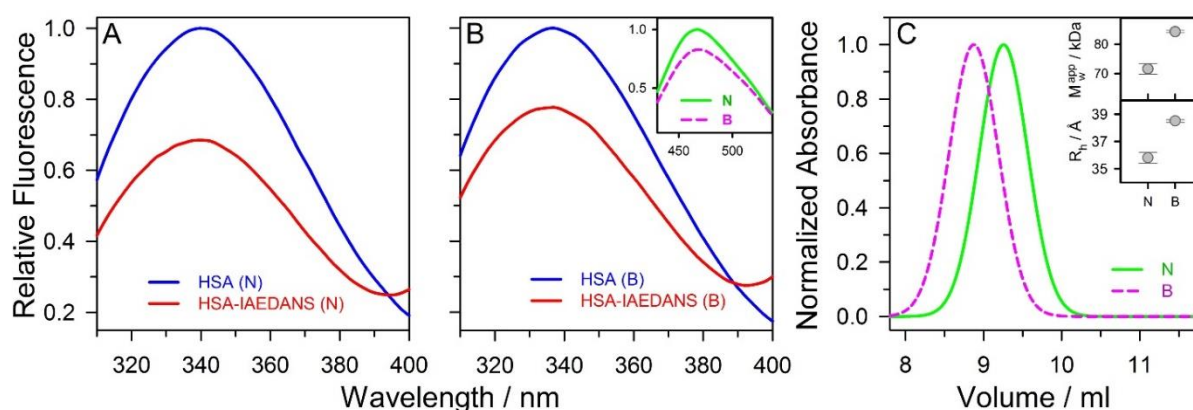


Figure 3.5. The B form has a larger inter-domain distance and is globally expanded compared to the N form. The intra-molecular distance between W214 and C34-IAEDANS is monitored by FRET in the N form (A) and in the B form (B). The change in fluorescence spectra of W214 in HSA and HSA-IAEDANS are shown. The inset in panel B compares the change in the fluorescence emission spectrum of C34-IAEDANS in the N form and the B form after excitation at 295 nm. (C) Size exclusion chromatography of the N form (green solid line) and the B form (pink dashed line). The top and the bottom inset in panel C, respectively, compare the apparent molecular weight ( $M_w^{app}$ ) and the hydrodynamic radius ( $R_h$ ) of the N form and the B form. The spread in the values of apparent molecular weight and hydrodynamic radius were estimated from two independent measurements. (N form: HSA at pH 7; B form: HSA at pH 11)

We quantified the FRET efficiency ( $E$ ) and the D-A distance ( $R$ ) in the N form and the B form using equation 8 and 9,<sup>32</sup>

$$E = 1 - \frac{F_{DA}}{F_D} \quad [8]$$

$$R = R_0 \left( \frac{1-E}{E} \right)^{\frac{1}{6}} \quad [9]$$

In equation 8,  $F_D$  and  $F_{DA}$  represent the fluorescence signal of the donor in the absence and the presence of the acceptor, respectively. In equation 9,  $R_0$  is the Forster's distance and we have experimentally determined the mean values of  $R_0$  (Figure 3.1, Table 3.1) in the N form and the B form to be 25.8 Å and 24.5 Å, respectively. We observed that the D-A distance in the N form is  $29.6 \pm 0.1$  Å and in the B form is  $31.0 \pm 0.4$  Å. The mean D-A distance in the B form is only 1.4 Å larger than that in the N form. However, this is a significant difference because of the following reasons: (i) The decreased FRET, and the increased D-A distance, in the B form compared to the N form does not only result in the decrease in the extent of quenching of the donor (W214) fluorescence (Figure 3.5A and 3.5B), but also leads to a significant decrease in the acceptor (C34-IAEDANS) fluorescence (Figure 3.5B (inset)).

These results indicate that the extent of energy transfer from the donor to the acceptor is less in the B form compared to the N form, due to the relatively large D-A distance in the B form; (ii) The measured FRET distances in the N form and the B form are very near to the experimentally determined  $R_0$  where the extent of energy transfer is most sensitive to the changes in D-A distances (between  $0.5 R_0$  to  $1.5 R_0$ ); (iii) We have determined the standard error in the measurement of D-A distances from at least three separate experiments and the standard error is limited to 0.1 Å in the N form and 0.4 Å in the B form. It is important to note that the errors in the measurement of D-A distances from FRET are typically small due to the sixth root dependence of  $R$  on  $\left(\frac{1-E}{E}\right)$  (equation 9); moreover,  $R$  also depends on  $R_0$  (equation 9) which itself has a sixth root dependence on other FRET parameters (equation 2); (iv) Most importantly, we observed that the B form also has a larger hydrodynamic radius (and hence a larger hydrodynamic volume) than the N form as determined by size exclusion chromatography (Figure 3.5C, see below). Hence, these results indicate that the D-A distance between domain I and domain II in the B form is 1.4 Å larger compared to the N form along the axis connecting W214 and C34-IAEDANS. The strength of vdW packing interactions between atoms depend steeply on inter-atomic distances. Hence, the larger inter-atomic distances between the helices of domain I and II (domain I<sub>Helix3</sub> – domain II<sub>Helix1</sub>) in the B form indicate that inter-domain helices are loosely packed and that some of the inter-domain packing interactions between domain I and II are severely disrupted, compared to the N form.

In order to investigate how the disruption of inter-domain packing interactions and the movement of domain I and domain II away from each other in the B form affects the overall dimension of the protein, we measured the apparent molecular weight and the hydrodynamic radius of the N form and the B form using analytical size exclusion chromatography on a GE Superdex 75 10/300 GL high performance gel filtration column (Figure 3.5C). We observed that the mean elution volume for the N form is  $9.25 \pm 0.05$  ml which decreases to  $8.86 \pm 0.01$  ml for the B form, indicating that the B form has a larger hydrodynamic volume and apparent molecular weight than the N form. Based upon the known elution volumes and molecular weights of five different standard biomolecules, we constructed a calibration curve (Figure 3.2A,) to determine the apparent molecular weight of the N form and the B form (Figure 3.5C (top inset)) using the data on their elution volume (Figure 3.5C). The apparent molecular weight of the N form is  $71.6 \pm 1.8$  kDa which increases considerably to  $84.3 \pm 0.4$  kDa in the B form (Figure 3.5C (top inset)). For the determination of the hydrodynamic radius, we constructed a calibration curve based upon the known hydrodynamic radius of three different standard proteins (Figure 3.2B), and determined the hydrodynamic radius of the N form to be  $35.8 \pm 0.4$  Å and the B form to be  $38.5 \pm 0.1$  Å (Figure 3.5C (bottom inset)). The B form has 15-20% larger apparent molecular weight and around 7-10% larger hydrodynamic radius than the N form. Hence, these results indicate that the disruption of inter-domain packing interactions in the B form results in the larger hydrodynamic volume and overall expansion of the protein in the B form compared to the N form.

### 3.3.3 B form retains N-like hydrophobicity

We observed that the expansion of the protein is not accompanied by any change in the hydration of the protein core in domain I and II. We investigated the core hydration by monitoring the change in fluorescence Stokes shift of C34-IAEDANS and W214 in domain I and II, respectively, which is a very sensitive indicator of the polarity of the medium surrounding the fluorophore. We compared the Stokes shift of C34-IAEDANS and W214 in the N and the B form to that in the urea unfolded U form (9 M urea, pH 7). The presence of 9 M urea in pH 7 buffer completely unfolds the protein and populates the U form in which C34-IAEDANS and W214 are completely exposed to water. The mean wavelength of maximum fluorescence emission ( $\lambda_{max}^{em}$ ) of C34-IAEDANS in the U form is 480 nm (Figure 3.6A). In the N form,  $\lambda_{max}^{em}$  of C34-IAEDANS is blue-shifted to 465 nm (Figure 3.6A), indicating that it is buried in the hydrophobic core of the domain I. We observed that the value of  $\lambda_{max}^{em}$  in the B



form is 467 nm, which is very similar to its value in the N form. This result strongly indicates that the core of domain I remains inaccessible to water and has N-like hydrophobicity.

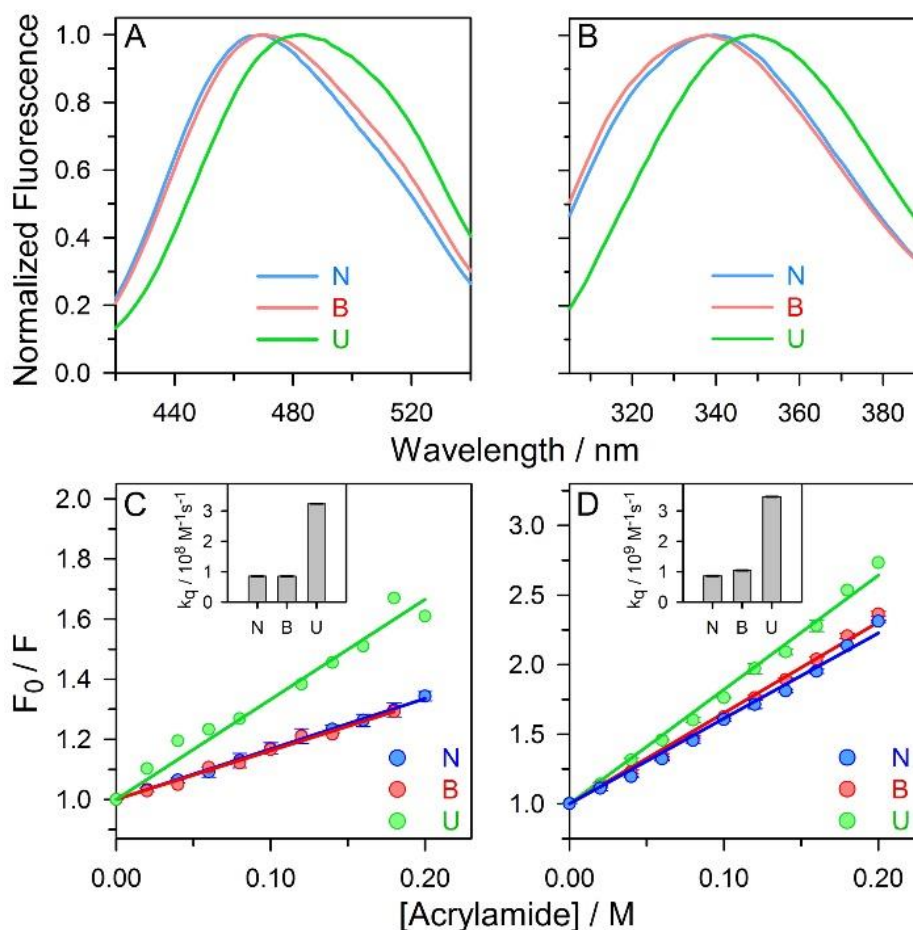


Figure 3.6. The B form retains N-like hydrophobicity in domain I and II. The fluorescence spectra of (A) C34-IAEDANS and (B) W214. The wavelength of maximum emission ( $\lambda_{max}^{em}$ ) of each spectrum was normalized to 1 for comparison. Stern-Volmer plots for dynamic quenching of (C) C34-IAEDANS and (D) W214 fluorescence. The solid lines through the data are the least-square fits to equation 10. The error bars shown in panels C and D represent the spread in the values of fluorescence at the emission maximum estimated from two or more independent measurements. The values of bimolecular quenching constant,  $k_q$ , are compared in the insets of panel C and D. The errors in the values of  $k_q$  are less than 5%. (N form: HSA at pH7; B form: HSA at pH 11; U form: HSA in 9M urea at pH 7)

The  $\lambda_{max}^{em}$  of W214 in the U form is 348 nm (Figure 3.6B). However, in the N form and the B form, the  $\lambda_{max}^{em}$  of W214 is blue-shifted to 340 nm and 338 nm, respectively (Figure 3.6B). These results indicate that in the B form, the solvation environment in the inter-domain region near W214 remains N-like or is slightly hydrophobic than the N form.

### 3.3.4 N-like structural fluctuations and solvent accessibility of the B form

We employed dynamic fluorescence quenching experiments to investigate structural fluctuations and solvent accessibility of domain I and II in the N form, the B form and the U form. In dynamic fluorescence quenching, the fluorescence of a buried fluorophore is quenched by a quencher present in the solvent due to the nanosecond fluctuations in the protein structure<sup>32,38</sup> and is described by the Stern-Volmer equation as:

$$\frac{F_0}{F} = 1 + K_{sv}[Q] \quad [10]$$

In equation 10,  $F_0$  and  $F$  are the fluorescence intensities of the fluorophore in the absence and the presence of the quencher (Q), respectively, and  $K_{sv}$  is the Stern-Volmer constant. We used acrylamide as the neutral quencher of C34-IAEDANS and W214 fluorescence, as described previously.<sup>32</sup> We observed that the mean value of  $K_{sv}$  for C34-IAEDANS in the U form is  $3.3 \text{ M}^{-1}$  (Figure 3.6C). However, the mean value of  $K_{sv}$  decreases to  $1.6 \text{ M}^{-1}$  in both the N form and the B form (Figure 3.6C). The observed mean value of  $K_{sv}$  for W214 in the U form is  $8.2 \text{ M}^{-1}$ , which decreases to  $6.1 \text{ M}^{-1}$  and  $6.5 \text{ M}^{-1}$  in the N form and the B form, respectively (Figure 3.6D). It is interesting to note that the mean values of  $K_{sv}$  remain similar in the N form and the B form for both C34-IAEDANS and W214.

The bimolecular quenching constant ( $k_q$ ) is a measure of the accessibility of fluorophore to the quencher in its photo-excited state. We measured the values of  $k_q$  using the above values of  $K_{sv}$  and the experimentally determined values of intensity averaged fluorescence lifetime ( $\tau_0$ ) (Figure 3.3, Table 3.2) ( $K_{sv} = k_q \tau_0$ ). We observed that the mean value of  $k_q$  for C34-IAEDANS in the U form is  $3.2 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ , which reduces significantly to  $0.8 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$  in both the N form and the B form (Figure 3.6C (inset)). For W214, the mean value of  $k_q$  in the N form ( $0.9 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ) and the B form ( $1.0 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ) are similar and significantly lower than the value of  $k_q$  in the U form ( $3.4 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ) (Figure 3.6D (inset)). All the values of  $K_{sv}$ ,  $\tau_0$  and  $k_q$  are listed in Table 3.2. It is important to note that the errors in estimation of  $k_q$  values are less than 5%. Hence, the similarities in the values of  $k_q$  in the N form and the B form indicate that in the B form, the domain I and domain II exhibit N-like structural fluctuations and solvent accessibilities.

### 3.3.5 B form retains N-like solvation dynamics

We investigated the solvation dynamics of the core of domain I and II around C34-IAEDANS and W214, respectively, by REES experiments.<sup>32,39</sup> In general, the  $\lambda_{max}^{em}$  of a fluorophore is independent of the excitation wavelength ( $\lambda_{ex}$ ). However, in the case of REES, the  $\lambda_{max}^{em}$  of a polar fluorophore shifts to a higher wavelength (red shift) upon exciting the fluorophore towards the red edge of its excitation spectrum. This phenomenon occurs in cases where the distribution of solvent dipoles around the fluorophore is heterogeneous and solvation dynamics is slower than the fluorescence time scale.<sup>32,39</sup> In this way, REES probes the dynamics of the immediate surroundings of the fluorophore.

We observed that in both the N form and the B form, the  $\lambda_{max}^{em}$  of W214 and C34-IAEDANS shifts towards red, when excited at the red edge of the excitation spectrum of the respective fluorophore (Figure 3.7 and Figure 3.8).

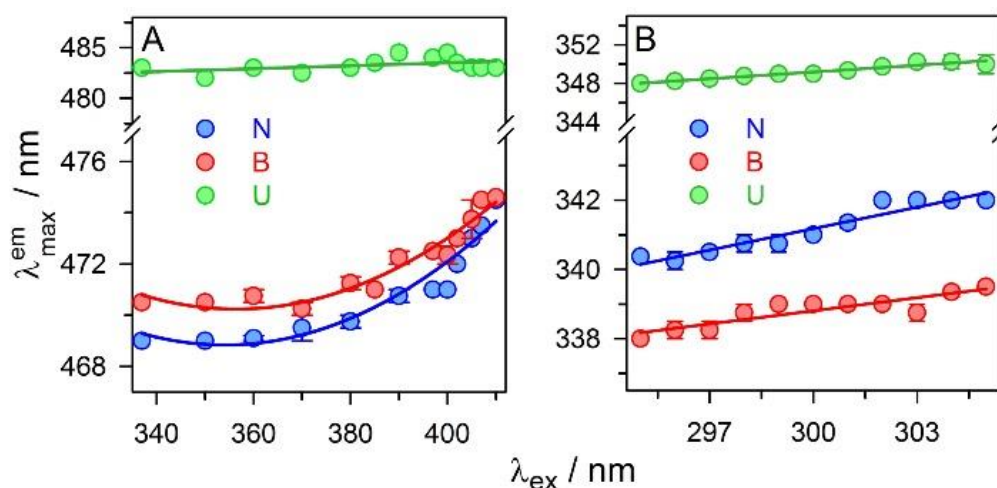


Figure 3.7. Effect of the change in excitation wavelength ( $\lambda_{ex}$ ) on the wavelength of maximum emission ( $\lambda_{max}^{em}$ ). The  $\lambda_{max}^{em}$  of (A) C34-IAEDANS and (B) W214 is plotted as a function of respective  $\lambda_{ex}$ , where the solid lines are drawn to guide the eyes. The error bars in panels A and B represent the spread in the values of emission maximum estimated from two independent measurements. (N form: HSA at pH7; B form: HSA at pH 11; U form: HSA in 9M urea at pH 7)

For C34-IAEDANS, we observed a significant amount of shift in  $\lambda_{max}^{em}$ , from 470 nm to 475 nm, in the N form and the B form, when excited from 337 nm to 407 nm (Figure 3.7A). For W214, when excited from 295 nm to 305 nm, we observed a slight shift in  $\lambda_{max}^{em}$ , from 340

nm to 342 nm, in the N form (Figure 3.7B). In the B form, the  $\lambda_{max}^{em}$  shifts from 338 nm to 340 nm (Figure 3.7B). However, for both the fluorophores, the  $\lambda_{max}^{em}$  remains constant (483 nm for C34-IAEDANS, 348 nm for W214) in the U form (Figure 3.7A and 3.7B), when excited at different  $\lambda_{ex}$  of the respective fluorophore. Both the fluorophores show no REES in the U form, which is expected due to the complete solvent exposure of the fluorophores and water molecules around the fluorophore are highly dynamic. In the N form and the B form, C34-IAEDANS shows large REES of 5 nm. Because, C34-IAEDANS is buried in domain I, it must be measuring the solvation dynamics of the side-chain packing inside the protein core.

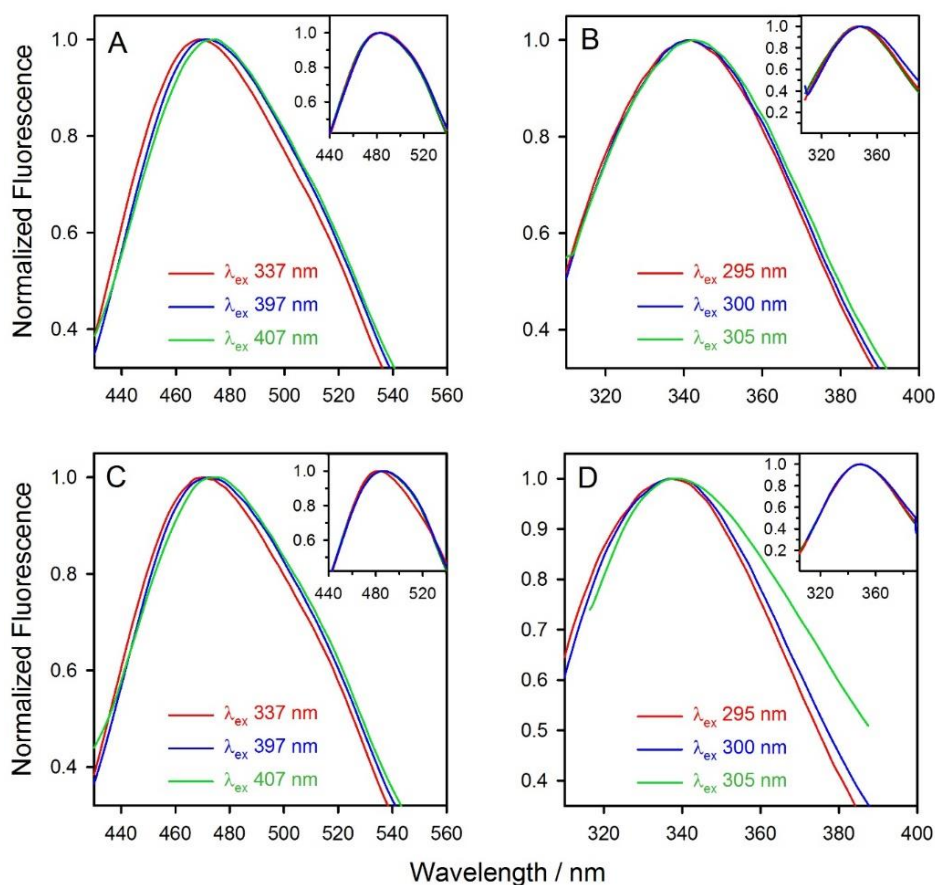


Figure 3.8. Dependence of wavelength of maximum emission,  $\lambda_{max}^{em}$ , of C34-IAEDANS and W214 on excitation wavelength,  $\lambda_{ex}$ . Panel A and B respectively show the representative fluorescence scan of C34-IAEDANS and W214 in the N form, when excited at different  $\lambda_{ex}$ . Panel C and D respectively show the representative fluorescence scan of C34-IAEDANS and W214 in the B form, when excited at different  $\lambda_{ex}$ . In each panel, the inset represents the fluorescence scans of the respective U forms when excited at different  $\lambda_{ex}$ . (N form: HSA at pH 7; B form: HSA at pH 11; U form: HSA in 9M urea accordingly at pH 7 or pH 11)

It appears that both in the N form and the B form, the protein core is heterogeneous and the side-chains solvating the C34-IAEDANS exhibit solvation dynamics on a much slower time scale than the nanosecond time scale of fluorescence. For W214, the magnitude of REES is small (2 nm) but similar in both the N form and the B form. However, it is important to note that from comparison of  $\lambda_{max}^{em}$ , W214 in the B form appears to be in slightly more hydrophobic environment than the N form. Hence, the results for W214 indicate that the solvent molecules around W214 have restricted movements and they move slower than the fluorescence time scale in both the N form and the B form.

### 3.3.6 Comparison of the global structure of the N and the B form

We compared the global secondary and tertiary structure of the N form, the B form and the U form using far- and near-UV CD spectra (Figure 3.9A and 3.9B). The observed mean residual ellipticity (MRE) at 222 nm, a measure of the global secondary structure, is  $-19365 \pm 440 \text{ deg cm}^2 \text{ dmol}^{-1}$  in the N form and  $-18805 \pm 560 \text{ deg cm}^2 \text{ dmol}^{-1}$  in the B form (Figure 3.9A). The MRE values decrease markedly to  $-5093 \pm 225 \text{ deg cm}^2 \text{ dmol}^{-1}$  in the U form (Figure 3.9A). The similar values of MRE in the N form and the B form indicate that the global secondary structure in the B form remains similar to the N form.

The intensity of the near-UV CD signal is a measure of the global tertiary structure of proteins. We observed that the near-UV CD spectra in the N form and the B form show absorption mainly in 255-270 nm wavelength range (Figure 3.9B). The MRE at 260 nm, which is the wavelength of maximum absorbance, in the N form and the B form is  $-177 \pm 8 \text{ deg cm}^2 \text{ dmol}^{-1}$  and  $-175 \pm 6 \text{ deg cm}^2 \text{ dmol}^{-1}$ , respectively (Figure 3.9B). The MRE at 260 nm in the U form is  $-92 \pm 18 \text{ deg cm}^2 \text{ dmol}^{-1}$  (Figure 3.9B). The similarity in MRE in the N form and the B form indicates that the global tertiary structure in the B form is similar to the N form. This is a surprising result in view of the fluorescence, size exclusion chromatography and the FRET experiments (Figure 3.4 and 3.5), which indicate that structurally the B form and the N form are different; specifically domain I and II move away from each other in the B form which leads to the disruption of the tertiary structure in the inter-domain region near W214 and protein expansion.

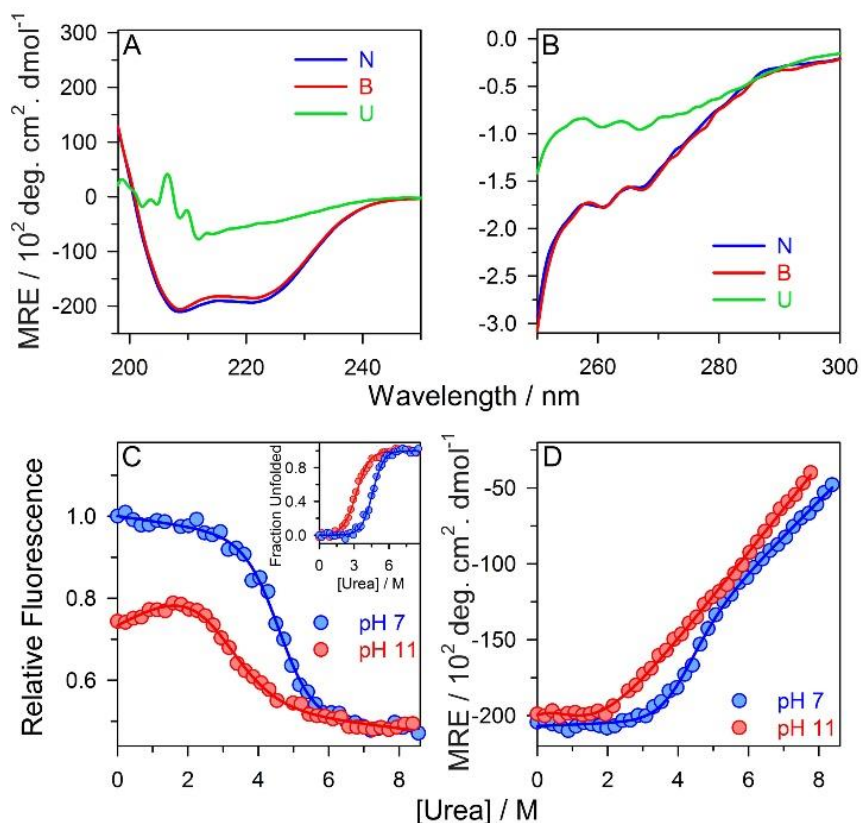
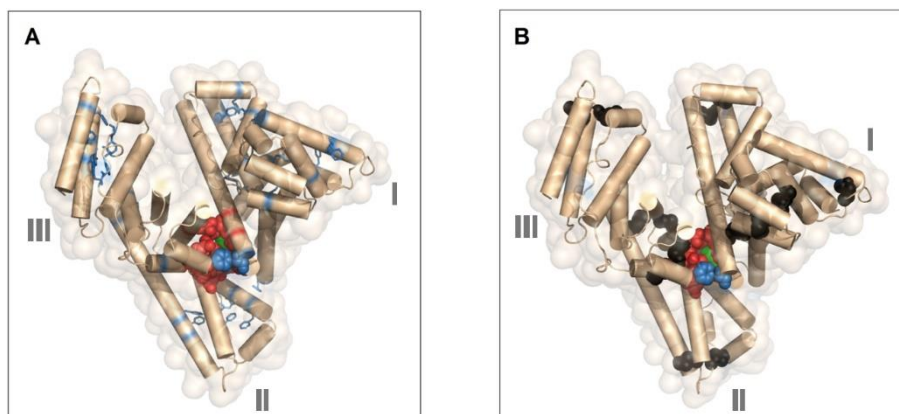


Figure 3.9. The global secondary structure (A) and the tertiary structure (B) of the B form is similar to the N form but their thermodynamic stability is different. The equilibrium unfolding transition of HSA is monitored by the change in fluorescence of W214 (C) and far-UV CD signal at 222 nm (D). The fraction of unfolded protein is plotted against [urea] in the inset of panel C. The solid lines through the data in panel C and pH 7 data in panel D are fits to a two-state model (equation 7), while for pH 11 in panel D is drawn to guide the eyes. (N form: HSA at pH 7; B form: HSA at pH 11; U form: HSA in 9M urea at pH 7)

It is important to note that near-UV CD experiments in the case of HSA, report primarily on the asymmetric burial of its 31 phenylalanine (Phe) residues. A careful examination of the crystal structure of HSA reveals that 30 out of its 31 Phe residues are distributed in the intra-domain region and only one Phe residue, F206, is located near the cluster of side-chains involved in the inter-domain side-chain packing (Figure 3.10A). It is also interesting to note that the intra-domain structure of HSA (in all the 3 domains) is stapled by a network of 17 disulfide bonds (Figure 3.10B), but not the inter-domain region. Hence, it appears that in the B form the intra-domain packing is intact resulting in the similar near-UV CD spectrum as the N form, but packing at the interface of domain I and II is broken, giving rise to the change in fluorescence, increase in inter-domain distance and protein volume. Previous studies, using nuclear magnetic resonance, ligand binding and monoclonal antibodies

binding properties, also support our conclusion that the structural fluctuation in the B form is particularly limited to domain I and II.<sup>24,40,41</sup>



*Figure 3.10. Structural components of HSA. Panel (A) shows all the Phe residues (blue), the inter-domain region (red) between domain I-II and domain II-III along with the sole tryptophan residue (green). It is important to note that although there are 31 Phe residues distributed throughout the protein, there is only single Phe residue, F206 (shown as blue sphere), in the inter-domain region. Panel (B) shows 17 disulfide bonds (black spheres) distributed in all the 3 intra-domain regions.*

### 3.3.7 Comparison of thermodynamic stabilities of the N and the B form

The base-induced B form resembles a DMG in domain I and II near the inter-domain region. We estimated the energetic contribution of vdW packing interactions near the inter-domain region of domain I and II in the stability of the N form by measuring the thermodynamic stabilities of the N form and the B form by the urea-induced equilibrium unfolding transitions using fluorescence (Figure 3.9C) and far-UV CD (Figure 3.9D). The equilibrium unfolding curves at pH 7 represent  $N \rightleftharpoons U$  transition, while at pH 11 they represent  $B \rightleftharpoons U$  transition. Here, it is important to note that the U form (unfolded in 9 M urea) at pH 7 and pH 11 are identical in terms of all the spectroscopic properties measured in this study, including Stokes shift, hydration of side-chains, solvation dynamics and secondary and tertiary structural content (Figure 3.11). All the urea-induced unfolding curves show sigmoidal transitions except far-UV CD monitored unfolding transition, which appears gradual in nature (Figure 3.9D). This result indicates that the unfolding of secondary structure in the DMG could be non-cooperative and gradual in nature.<sup>42</sup>

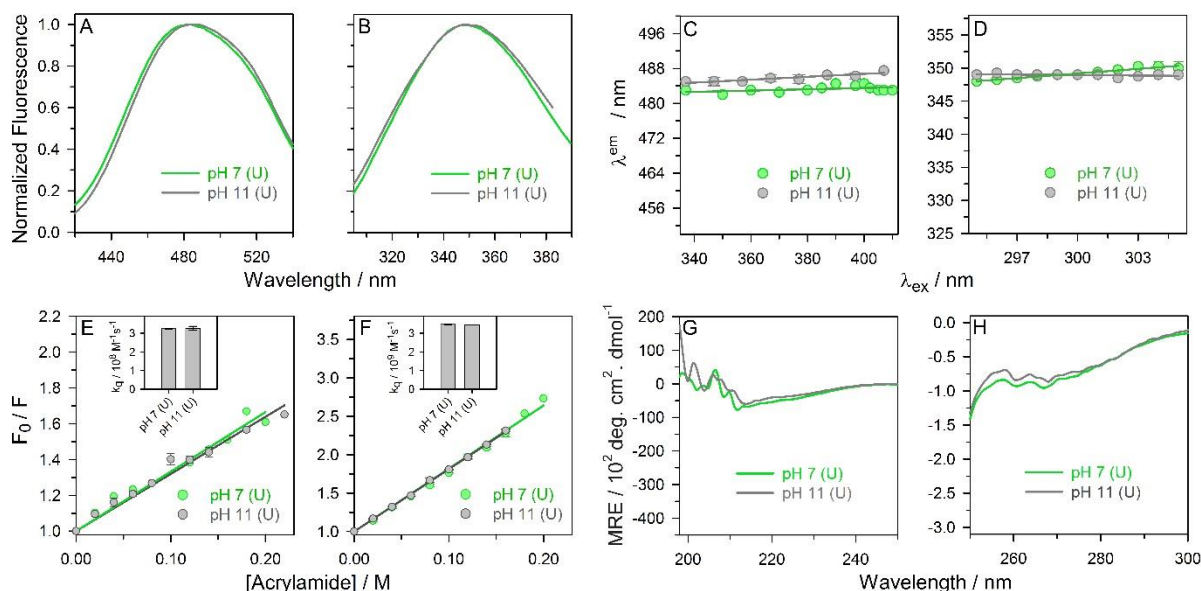


Figure 3.11. Comparison of the spectroscopic properties of the U form (unfolded in 9M urea) at pH 7 and pH 11. Panel (A) and (B) shows the fluorescence spectra of C34-IAEDANS and W214, respectively. The wavelength of maximum emission ( $\lambda_{max}^{em}$ ) of each spectrum was normalized to 1 for comparison. Panel (C) and (D) show the effect of the change in excitation wavelength ( $\lambda_{ex}$ ) on the wavelength of maximum emission ( $\lambda_{max}^{em}$ ). The  $\lambda_{max}^{em}$  of (C) C34-IAEDANS and (D) W214 is plotted as a function of respective  $\lambda_{ex}$ , where the solid lines are drawn to guide the eyes. Panel (E) and (F) show the Stern-Volmer plots for dynamic quenching of C34-IAEDANS and W214 fluorescence, respectively. The solid lines through the data are the least-square fits to equation 10 (main text). The values of bimolecular quenching constant,  $k_q$ , are compared in the insets of panel (E) and (F). The errors in the values of  $k_q$  are less than 5%. Panel (G) and (H) show the global secondary structure and the tertiary structure, respectively.

We analysed the fluorescence monitored data at pH 7 and pH 11 and CD monitored data at pH 7 using a two-state model (equation 7) to obtain the values of  $\Delta G$  (free energy of unfolding) and  $m$  (slope of the transition which represents change in the solvent accessible surface area). For  $N \rightleftharpoons U$  transition, the values of  $\Delta G_{N-U}$  and  $m_{N-U}$  are  $5.35 \pm 0.17$  kcal mol<sup>-1</sup> and  $1.17 \pm 0.03$  kcal mol<sup>-1</sup> M<sup>-1</sup>, respectively. For  $B \rightleftharpoons U$  transition, the values of  $\Delta G_{B-U}$  and  $m_{B-U}$  are  $3.0 \pm 0.06$  kcal mol<sup>-1</sup> and  $1.0 \pm 0.04$  kcal mol<sup>-1</sup> M<sup>-1</sup>. The value of  $m_{N-B}$  ( $m_{N-U} - m_{B-U}$ ), is very small ( $0.13 \pm 0.05$  kcal mol<sup>-1</sup> M<sup>-1</sup>). This result suggests that there is very little or no change in the solvent accessible surface area during  $N \rightleftharpoons B$  transition. This result is



expected if the B form is like DMG and its core is dry. The value of  $\Delta G_{N-B}$  ( $\Delta G_{N-U} - \Delta G_{B-U}$ ), is  $2.35 \pm 0.18$  kcal mol<sup>-1</sup>.

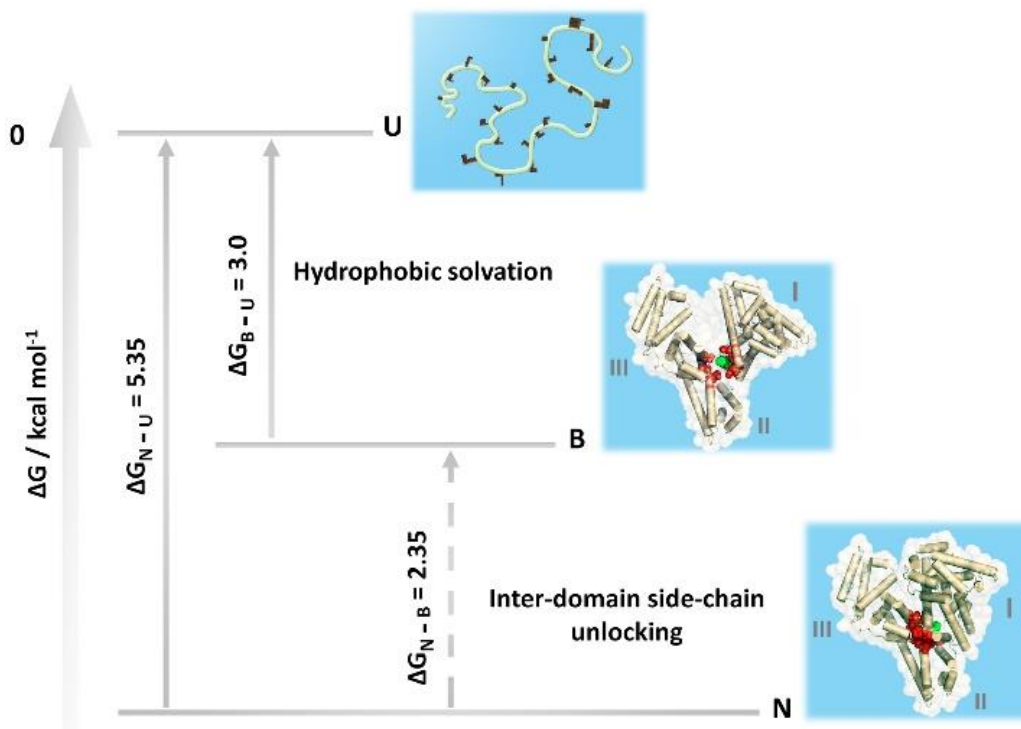


Figure 3.12. Model energy diagram showing the relative energies of the N form, the B form and the U form. In the structural models of the N form and the B form, the cylinders represent  $\alpha$ -helical secondary structures, red spheres represent the side-chains of the residues involved in the packing of the inter-domain interface, green sphere represents the side-chain of W214 and the white surface around the protein shows the solvent accessible surface area of the protein. The model of the U form shows only the peptide backbone (white thread) and the representative side-chains (black color) of the unfolded protein. The blue color represents water molecules around the protein. In the B form, the inter-domain packing is disrupted without hydrophobic solvation. The free energies of unfolding of the N form ( $\Delta G_{N-U}$ ) and the hydrophobic solvation ( $\Delta G_{B-U}$ ) were determined from equilibrium unfolding transitions of the N form ( $N \rightleftharpoons U$  transition) and the B form ( $B \rightleftharpoons U$  transition), respectively, as described in Figure 3.9C and the results and discussion section. The free energy of inter-domain side-chain unlocking ( $\Delta G_{N-B}$ ) was calculated from  $\Delta G_{N-B} = \Delta G_{N-U} - \Delta G_{B-U}$ .

A comparison of  $\Delta G_{N-U}$  and  $\Delta G_{N-B}$  indicates that 40-45% of the total change in free energy occurs during  $N \rightleftharpoons B$  transition. Hence, there is a large difference in thermodynamic

stability of the N form and the B form. One possibility for this large difference in stability is that the movement of domain I and II also disrupts some of the packing interactions between domain I and III as shown in our structural model (Figure 3.12). However, the decrease in stability is not due to the disruption of the intra-domain secondary structure and side-chain packing as shown by CD experiments. Hence, our results indicate that vdW interactions play very important role in the stability of the N form, but not the DMG. It is important to note that for a few proteins,<sup>18,43,44</sup> it has been shown that even a single cavity creating point mutation can drastically decrease the stability of the N form.

It has been postulated that vdW packing interactions develop during the last step, i.e., the DMG $\rightleftharpoons$ N transition, of protein folding.<sup>16-19,45</sup> This hypothesis has been supported by experimental observations of DMG-like intermediate states on the native-side of the free energy barrier of a few proteins.<sup>46-53</sup> However, recent findings on a few other proteins, including alpha-lactalbumin,<sup>54</sup> nuclear coactivator binding domain<sup>55</sup> and RNase H<sup>56</sup> appear to suggest that N-like packing interactions could also develop early during the folding of an unfolded polypeptide. Here, our observations that unfolding of the N form of HSA begins with the disruption of some of the inter-domain packing interactions and that intra-domain packing dissolves at a later stage indicate that vdW interactions develop in multiple stages during folding of proteins. Our results of the REES experiments (Figure 3.7) also suggest that the N state of proteins themselves might not be as tightly packed as previously thought<sup>18,57</sup> but might be an ensemble of loosely packed forms.

### 3.4 Conclusions

In summary, we have shown that the B form of HSA resembles an expanded DMG-like near native state. Fluorescence experiments show that the fluorescence of W214 in the B form has markedly decreased compared to the N form, suggesting the disruption of the tertiary structure near W214, at the interface of domain I and II. This is further confirmed by FRET and size exclusion chromatography measurements, which show that the B form is expanded than the N form. The expansion in the B form is a result of the loosening of side-chain packing interactions in domain I and II near the inter-domain region. Dynamic fluorescence quenching assays on C34-IAEDANS and W214 suggest that the expansion in the B form is not accompanied by the hydration of the protein core. REES experiments on C34-IAEDANS and W214 show that in both the N form and the B form, the solvation dynamics near respective fluorophore in domain I and II is slower than the fluorescence time scale. However, the results of the far- and near-

UV CD experiments indicate that the secondary and the tertiary structure in the B form are similar to the N form. These results suggest that in the B form, domain I and II of the protein have DMG-like characteristics near the inter-domain region. The thermodynamic stability data indicates that the B form has lost 40-45% of the total stability of the N form. Because the N form and the B form appear to be different mainly in the extent of side-chain packing at the inter-domain interface, these results suggest that the vdW packing interactions play an important role in protein stability. Our results strongly indicate that the base-induced unfolding of proteins begins with the disruption of vdW interactions and that the structural loosening and the core solvation are two distinct steps. Since the B form has physiological importance in binding to many different ligands and drugs,<sup>24</sup> it is possible that the loosening of side-chain packing and gain in conformational entropy enables the protein to undergo native state volume fluctuations in order to perform these multiple functions.<sup>17,58</sup>

### 3.5 References

1. Ross, M. H., Ely, J. O., and Archer, J. G. (1951) Alkaline phosphatase activity and pH optima. *J Biol Chem* 192, 561-568.
2. Whitaker, J. R. (1980) Changes Occuring in Proteins in Alkaline Solution. in *Chemical Deterioration of Proteins*, American Chemical Society. pp 145-163
3. Wanwimolruk, S., and Birkett, D. J. (1982) The effects of N-B transition of human serum albumin on the specific drug-binding sites. *Biochim Biophys Acta* 709, 247-255.
4. Schwass, D. E., and Finley, J. W. (1984) Heat and Alkaline Damage to Proteins - Racemization and Lysinoalanine Formation. *J Agric Food Chem* 32, 1377-1382.
5. Goto, Y., and Fink, A. L. (1989) Conformational States of Beta-Lactamase - Molten-Globule States at Acidic and Alkaline pH with High Salt. *Biochemistry* 28, 945-952.
6. Konno, T., Kamatari, Y. O., Tanaka, N., Kamikubo, H., Dobson, C. M., and Nagayama, K. (2000) A partially unfolded structure of the alkaline-denatured state of pepsin and its implication for stability of the zymogen-derived protein. *Biochemistry* 39, 4182-4190.
7. Jiang, J., Chen, J., and Xiong, Y. L. (2009) Structural and emulsifying properties of soy protein isolate subjected to acid and alkaline pH-shifting processes. *J Agric Food Chem* 57, 7576-7583.

8. Talley, K., and Alexov, E. (2010) On the pH-optimum of activity and stability of proteins. *Proteins* 78, 2699-2706.
9. Shan, Y. Y., Ma, M. H., Huang, X., Guo, Y. J., Jin, G. F., and Jin, Y. G. (2012) Simple pH Treatment as an Effective Tool to Improve the Functional Properties of Ovomucin. *J Food Sci* 77, C740-C745.
10. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14, 1-63.
11. Tanford, C. (1968) Protein denaturation. *Adv Protein Chem* 23, 121-282.
12. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24, 1-95.
13. Privalov, P. L. (1979) Stability of proteins: small globular proteins. *Adv Protein Chem* 33, 167-241.
14. Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28, 1667-1680.
15. Finkelstein, A. V., and Shakhnovich, E. I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28, 1681-1694.
16. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725-2737.
17. Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23, 4-10.
18. Bhattacharyya, S., and Varadarajan, R. (2013) Packing in molten globules and native states. *Curr Opin Struct Biol* 23, 11-21.
19. Thirumalai, D., Liu, Z., O'Brien, E. P., and Reddy, G. (2013) Protein folding: from theory to practice. *Curr Opin Struct Biol* 23, 22-29.
20. Rami, B. R., and Udgaonkar, J. B. (2002) Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* 41, 1710-1716.
21. Carter, D. C., and Ho, J. X. (1994) Structure of serum albumin. *Adv Protein Chem* 45, 153-203.

22. Figge, J., Rossing, T. H., and Fencl, V. (1991) The role of serum proteins in acid-base equilibria. *J Lab Clin Med* 117, 453-467.
23. Leonard, W. J., Kantvija, K., and Foster, J. F. (1963) A Structural Transformation in Bovine and Human Plasma Albumins in Alkaline Solution as Revealed by Rotatory Dispersion Studies. *J Biol Chem* 238, 1984-1988.
24. Peters Jr, T. (1995) *All About Albumin: Biochemistry, Genetics, and Medical Applications*, 1 ed., Academic Press, San Diego.
25. Droge, J. H., Wilting, J., and Janssen, L. H. (1982) A comparative study of some physico-chemical properties of human serum albumin samples from different sources-II. The characteristics of the N-B transition and the binding behaviour with regard to warfarin and diazepam. *Biochem Pharmacol* 31, 3781-3786.
26. Dockal, M., Carter, D. C., and Ruker, F. (2000) Conformational transitions of the three recombinant domains of human serum albumin depending on pH. *J Biol Chem* 275, 3042-3050.
27. Painter, L., M. Harding, M., and J. Beeby, P. (1998) Synthesis and interaction with human serum albumin of the first 3,18-disubstituted derivative of bilirubin. *J Chem Soc, Perkin Trans 1*, 3041-3044.
28. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.
29. Acharya, N., Mishra, P., and Jha, S. K. (2016) Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* 7, 173-179.
30. Atanasiu, C., Su, T. J., Sturrock, S. S., and Dryden, D. T. (2002) Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res* 30, 3936-3944.
31. Khurana, R., Hate, A. T., Nath, U., and Udgaonkar, J. B. (1995) pH-Dependence of the Stability of Barstar to Chemical and Thermal-Denaturation. *Protein Sci* 4, 1133-1144.
32. Lakowicz, J. R. (2006) *Principles of fluorescence spectroscopy*, Springer, Singapore.

33. Irvine, G. B. (2001) Determination of Molecular Size by Size-Exclusion Chromatography (Gel Filtration). in *Curr Protoc Cell Biol*, John Wiley & Sons, Inc. pp 5.5.1-5.5.16
34. Amiri, M., Jankeje, K., and Albani, J. R. (2010) Origin of fluorescence lifetimes in human serum albumin. Studies on native and denatured protein. *J Fluoresc* 20, 651-656.
35. Amiri, M., Jankeje, K., and Albani, J. R. (2010) Characterization of human serum albumin forms with pH. Fluorescence lifetime studies. *J Pharm Biomed Anal* 51, 1097-1102.
36. Agashe, V. R., and Udgaonkar, J. B. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* 34, 3286-3299.
37. Steiner, R. F., and Edelhoch, H. (1961) Influence of pH and urea on the ultra-violet fluorescence of several globular proteins. *Nature* 192, 873-874.
38. Eftink, M. R., and Ghiron, C. A. (1975) Dynamics of a protein matrix revealed by fluorescence quenching. *Proceedings of the National Academy of Sciences* 72, 3290-3294.
39. Demchenko, A. P. (2008) Site-selective Red-Edge effects. *Methods Enzymol* 450, 59-78.
40. Lapresle, C. (1988) Enzyme-Immunoassay Using Monoclonal-Antibodies to Study Conformational-Changes of Human-Serum Albumin. *Anal Biochem* 174, 308-312.
41. Bos, O. J. M., Labro, J. F. A., Fischer, M. J. E., Wilting, J., and Janssen, L. H. M. (1989) The Molecular Mechanism of the Neutral-to-Base Transition of Human-Serum Albumin - Acid-Base Titration and Proton Nuclear Magnetic-Resonance Studies on a Large Peptic and a Large Tryptic Fragment of Albumin. *J Biol Chem* 264, 953-959.
42. Jha, S. K., Dhar, D., Krishnamoorthy, G., and Udgaonkar, J. B. (2009) Continuous dissolution of structure during the unfolding of a small protein. *Proceedings of the National Academy of Sciences* 106, 11113-11118.
43. Sandberg, W. S., and Terwilliger, T. C. (1989) Influence of interior packing and hydrophobicity on the stability of a protein. *Science* 245, 54-57.

44. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* 255, 178-183.
45. Jha, S. K., and Udgaonkar, J. B. (2007) Exploring the cooperativity of the fast folding reaction of a small protein using pulsed thiol labeling and mass spectrometry. *J Biol Chem* 282, 37479-37491.
46. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
47. Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-<sup>19</sup>F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proceedings of the National Academy of Sciences* 92, 9318-9322.
48. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci USA* 106, 12289-12294.
49. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci USA* 107, 4955-4960.
50. Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci USA* 111, 4856-4861.
51. Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* 118, 6380-6392.
52. Neumaier, S., and Kiefhaber, T. (2014) Redefining the dry molten globule state of proteins. *J Mol Biol* 426, 2520-2528.
53. Mishra, P., and Jha, S. K. (2017) An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* 121, 9336-9347.

- 
54. Ramboarina, S., and Redfield, C. (2008) Probing the Effect of Temperature on the Backbone Dynamics of the Human alpha-Lactalbumin Molten Globule. *J Am Chem Soc* 130, 15318-15326.
  55. Kjaergaard, M., Poulsen, F. M., and Teilum, K. (2012) Is a Malleable Protein Necessarily Highly Dynamic? The Hydrophobic Core of the Nuclear Coactivator Binding Domain is Well Ordered. *Biophys J* 102, 1627-1635.
  56. Rosen, L. E., Connell, K. B., and Marqusee, S. (2014) Evidence for close side-chain packing in an early protein folding intermediate previously assumed to be a molten globule. *Proceedings of the National Academy of Sciences* 111, 14746-14751.
  57. Richards, F. M. (1977) Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng* 6, 151-176.
  58. Law, A. B., Sapienza, P. J., Zhang, J., Zuo, X., and Petit, C. M. (2017) Native State Volume Fluctuations in Proteins as a Mechanism for Dynamic Allostery. *J Am Chem Soc* 139, 3599-3602.



**Chapter 4.**  
**Nucleation Dependent Aggregation of Nucleic Acid  
Binding Domains of TDP-43 Occur *via* Dry Molten  
Globule-like State**

## 4.1 Introduction

The dry molten globule (DMG)-like intermediate states, with unlocked tertiary structure and a dry core, have been proposed as a universal protein folding intermediate.<sup>1,2</sup> DMG-like states were first proposed in theoretical studies as a hypothetical transition state during protein folding reactions.<sup>3,4</sup> However, in the past three decades, DMGs have been observed in many kinetic,<sup>5-8</sup> equilibrium<sup>9-13</sup> and computational<sup>14,15</sup> studies on protein folding. Collectively, these studies have established the role of DMGs as the last observable intermediate state formed during the folding of proteins.

Recent studies on small and large multi-domain proteins show that DMGs are in equilibrium with the native state.<sup>10,12,16,17</sup> The existence of DMGs in continuous equilibrium with the native state signifies a putative role of DMGs in performing functions. This is, however, not surprising taking into account that the formation of DMGs provides higher conformational dynamics to the protein and the increased conformational dynamics is associated with improved catalytic functions.<sup>18-21</sup>

For insulin and  $\beta$ 2-microglobulin, the increased conformational dynamics have also been shown to be associated with their aggregation.<sup>22,23</sup> In view of that, the increased conformational dynamics in DMGs also make it a potential candidate for protein aggregation anomalies such as neurodegenerative diseases. To date, there is no direct evidence of the involvement of DMGs in any protein aggregation study. Despite some studies suggest that the protein aggregation begins either from the native state or from a native-like state,<sup>24-26</sup> a consensus on what is the nature of initial structural change in the course of the aggregation reaction is yet to be reached. In this study, we report that the amyloid aggregation of the nucleic acid binding domains of the transactive response DNA binding protein (TDP-43) begins with a DMG-like intermediate state.

TDP-43 is a ubiquitously expressed nuclear protein that shuttles between the nucleus and the cytoplasm to perform various RNA processing functions.<sup>27,28</sup> Conversely, the intraneuronal TDP-43 aggregation is associated with a fatal neurodegenerative disease, amyotrophic lateral sclerosis (ALS).<sup>29-31</sup> During ALS disease, TDP-43 localizes in the cytosol, instead of the nucleus, and forms amyloid-like aggregates which possess prion-like activity.<sup>32,33</sup> Multiple factors like mutations, aberrant cleavage, and environmental stress are responsible for TDP-43 aggregation.<sup>34-36</sup> However, statistical analysis of all reported ALS linked TDP-43 mutations suggest that the mutations in TDP-43 gene contribute to less than 5% of all ALS

cases indicating that the non-genetic factors, such as stress, play a crucial role in TDP-43 aggregation.<sup>34</sup>

TDP-43 is a 414 amino acid residue, four domain protein consisting of a ubiquitin like N-terminal domain, a glycine rich disordered C-terminal domain and two structurally similar repeats of RNA recognition motif- RRM1 and RRM2 which represents the nucleic acid binding domain of TDP-43 (termed as TDP<sup>NBD</sup>, hereafter) (Figure 4.1).<sup>37-39</sup> The protein also contain a nuclear localization sequence (NLS) in N-terminal domain and a nuclear export sequence (NES) in RRM2 domain, which helps in shuttling of TDP-43 between nucleus and cytosol for its function.<sup>40</sup> The N-terminal domain facilitates the dimerization of TDP-43, C-terminal domain enables interaction with other proteins and the two RNA recognition motifs are highly conserved functional domains that binds to nucleic acids to perform a variety of RNA processing functions.<sup>41-43</sup>

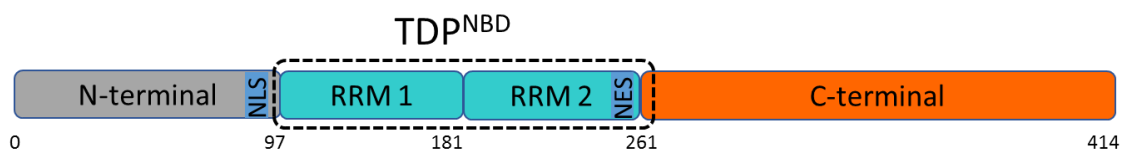


Figure 4.1. Domain arrangement of TDP-43. Four domains are shown: N-terminal domain, RNA recognition motifs (RRM 1 and RRM 2) and C-terminal domain, along with a nuclear localization sequence (NLS) and a nuclear export sequence (NES). The two RRM s are shown together as TDP<sup>NBD</sup>.

Recent studies show that TDP<sup>NBD</sup> tend to aggregate under low-pH starvation stress mimicking conditions.<sup>44,45</sup> Yet, the nature of the initial structural transition under low-pH stress and the mechanism of aggregation of TDP<sup>NBD</sup> remain unexplored. In this study, we have utilized a range of biophysical tools, including fluorescence resonance energy transfer (FRET), fluorescence spectroscopy, CD spectroscopy, dynamic light scattering, transmission electron microscopy, and fluorescence microscopy, along with kinetic studies to explore the earliest steps during low-pH stress and unravel the mechanism of TDP<sup>NBD</sup> aggregation.

## 4.2 Materials and Methodology

### Reagents, buffers, and experimental conditions

All reagents used for the experiments are of the highest purity grade procured from Sigma, Urea (ultrapure grade) from Himedia, and IAEDANS from Invitrogen unless otherwise specified.

All experiments for the N form were performed at pH 7.5, and buffers were composed of 20 mM MOPS 150 mM KCl buffer. All experiments for the A form were performed at pH 3.0 in 20 mM Glycine-HCl buffer. To perform any experiment at pH 3.0, the protein is buffer exchanged from the storage buffer (discussed below) into 20 mM Glycine-HCl pH 3.0 buffer, so that the protein is devoid of any additional salt or glycerol. The unfolded (U) form of the protein is prepared in an unfolding buffer composed of 8 M urea in 20 mM MOPS pH 7.5 buffer unless otherwise specified. The amyloid form is prepared by heating the A form in pH 3.0 buffer at 60°C for 24 hours. The urea concentration was determined by the measurement of the refractive index.<sup>46</sup> All the buffer solutions were filtered with 0.2 µm filters before use.

### Protein expression & purification

TDP<sup>NBD</sup> and NBD<sup>Mut</sup> (a mutant variant of TDP<sup>NBD</sup>, with a single tryptophan at 172 position and a single cysteine at 198 position) were expressed in *E.coli* BL21 DE3 competent cells and purified as described earlier.<sup>44</sup> Briefly, the protein expression was induced in the transformed cell by 1mM IPTG induction. After 24 hours incubation at 20°C, the transformed cells were pelleted down by centrifugation at 4500 g followed by cell lysis in lysis buffer (20 mM Tris-HCl, 30 mM imidazole, 300 mM NaCl, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 10% glycerol, pH 7.5).

For TDP<sup>NBD</sup>, the cell lysate was centrifuged at 14000 g to separate the cell debris. The supernatant fraction was filtered with a 0.2 µm PVDF membrane filter and subjected for further chromatographic purification by Ni-Sepharose His-Tag affinity purification with an increasing gradient of imidazole (30 mM to 300 mM). The purified fraction was buffer exchanged to cleavage buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 8.0) using HiPrep 26/10 desalting column (GE Healthcare) and incubated for 15 hours in the presence of prescission protease for His-tag cleavage. His-tag cleaved TDP<sup>NBD</sup> was further purified by anion-exchange purification using HiPrep Q HP 5 ml column (GE Healthcare) by increasing the gradient of NaCl (50mM to 500mM) on AKTA Pure FPLC system (GE Healthcare). Finally, the purified protein was buffer exchanged into the storage buffer (20 mM Phosphate, 150 mM KCl, 2% glycerol, 1 mM DTT, pH 7.4 buffer) and stored at -30°C till further use. All the experiments were performed with TDP<sup>NBD</sup> until otherwise specified.

NBD<sup>Mut</sup> was found to be expressed majorly in the inclusion bodies. The inclusion bodies were dissolved in 7 M urea and NBD<sup>Mut</sup> was purified by Ni-sepharose His-Tag affinity purification in the presence of 7 M urea. The purified unfolded mutant variant (in 7 M urea)

was buffer exchanged to 20 mM Tris-HCl pH 8.0 buffer with 50 mM NaCl. Further His-tag cleavage and anion-exchange purification was performed similarly as for TDP<sup>NBD</sup> and stored in 20 mM MOPS buffer at -30°C till further use.

The purity of TDP<sup>NBD</sup> and NBD<sup>Mut</sup> was confirmed ESI-MS (Figure 4.2). The calculated mass of TDP<sup>NBD</sup> is 19429 Da, and of NBD<sup>Mut</sup> is 19310 Da. The concentration of protein was determined spectroscopically from their absorbance at 280 nm, using a molar extinction coefficient of 15470 M<sup>-1</sup>cm<sup>-1</sup> for TDP<sup>NBD</sup> and 11460 M<sup>-1</sup>cm<sup>-1</sup> for NBD<sup>Mut</sup>.

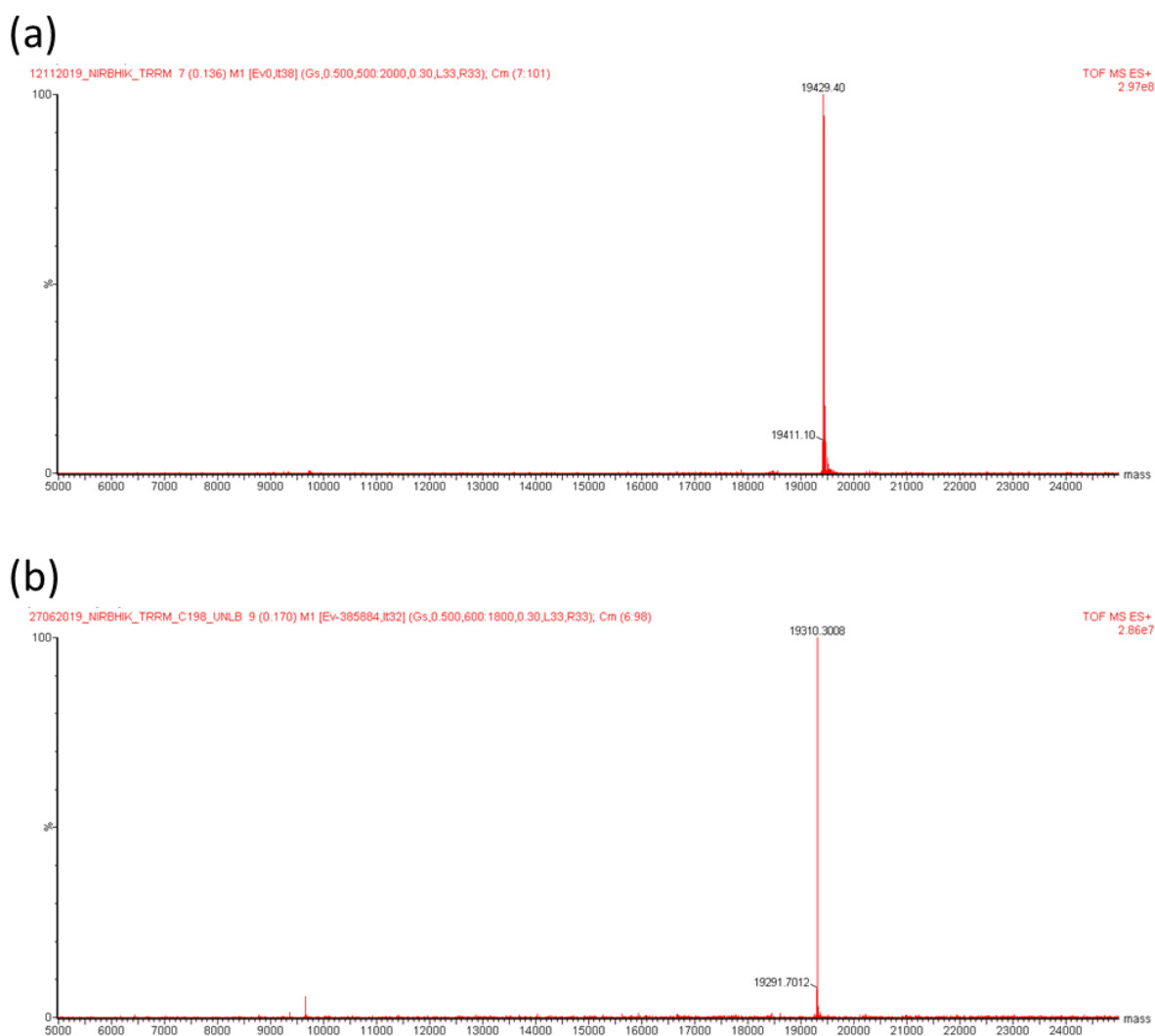


Figure 4.2. MaxEnt deconvoluted electrospray mass spectrum of (a) TDP<sup>NBD</sup> and (b) NBD<sup>Mut</sup> for purity analysis.

### Spectroscopic measurements

All the absorbance measurements were collected on a Lab India UV/Vis spectrophotometer using a quartz cell of 1 cm path-length. All the fluorescence emission spectra

were acquired on a Fluoromax-4 fluorometer from Horiba using a quartz cell of 1 cm path-length. All fluorescence spectra were collected with the excitation bandwidth of 1 nm, emission bandwidth of 10 nm and the scan speed of 100 nm/min unless otherwise specified. The far- and near-UV circular dichroism (CD) measurements were carried out on a Jasco J-815 CD spectrophotometer using 0.1cm and 1cm path-length quartz cells, respectively. The far-UV CD spectra were collected in the 200-250 nm range, whereas the near-UV CD spectra were collected in the 250-300 nm range. Each spectrum was averaged over three scans with the following parameters: a data pitch of 1 nm, a scan speed of 20 nm/min, and bandwidth of 2 nm. All the fluorescence and CD spectra were corrected for their respective buffer background signals. Refractive index measurements were performed on an RSR-2 Abbe refractometer from Rajdhani Scientific Instruments Co.

### Urea-induced equilibrium unfolding studies

For urea-induced equilibrium unfolding transition studies for the N form and the A form of TDP<sup>NBD</sup>, the samples were prepared by incubating 10  $\mu$ M protein for 3 hours in a varied concentration of urea, respectively, at pH 7.5 and pH 3.0. The fluorescence spectra of intrinsic tryptophans (Trp) were acquired by exciting the Trp at 295 nm and collecting the emission spectra from 310-400 nm. The far-UV CD spectra of the same samples were measured in the range of 200-250 nm. The change in Trp fluorescence at 340 nm and far-UV CD signals at 222 nm, as a function of urea concentration, was analyzed using a two-state  $X-U$  model<sup>47</sup> ( $X$  is N at pH 7.5 and A at pH 3.0) and fitted to the equation,

$$y_{obs} = \frac{y_X + y_U e^{\frac{-\Delta G_{X-U} + m_{X-U}[D]}{RT}}}{1 + e^{\frac{-\Delta G_{X-U} + m_{X-U}[D]}{RT}}} \quad (eq 1)$$

where,  $y_{obs}$  is the observed fluorescence change,  $y_X$  and  $y_U$  are the fluorescence signal of the native state (N form or A form) and their respective unfolded form.  $\Delta G_{X-U}$  is the standard free energy and  $m_{X-U}$  is the slope of the transition.

### Dynamic Light Scattering measurement

The dynamic light scattering (DLS) measurements were carried out on a DynaPro-99 Series dynamic light scattering module (Protein Solutions Ltd/ Wyatt). The micropipette tips and microcentrifuge tubes were thoroughly rinsed with filtered buffer before use. For the N form, A form, and the amyloid form, a 25  $\mu$ M protein concentration was used to acquire DLS reading. The samples were centrifuged at 15000 g for 15 min to remove any suspended

particles. The aliquots from the supernatant were put into a 45  $\mu$ l cuvette followed by placing the cuvette in the sample chamber maintained at 25°C. The samples were illuminated with a laser of 829.4 nm wavelength, and the scattering intensity was measured at the right angle simultaneously with the autocorrelation function. For each sample, 70 acquisitions were collected. The acquisition time and sensitivity parameters used for the DLS measurement are 5 s and 70%, respectively. The fluctuations greater than 15% of the scattering intensity were marked and excluded from the data analysis. The DynaLS program was used to resolve the accepted acquisitions into hydrodynamic radii distribution. The results were also verified by using the regularization algorithm (Protein Solutions Ltd).

### **Cys-TNB labeling and FRET distance measurement**

2-nitro-5-thiobenzoic acid (TNB) acts as a quencher of tryptophan (Trp) fluorescence, and hence Trp-TNB forms a FRET pair for intramolecular distance measurement. The wild type TDP<sup>NBD</sup> contain two Trp and four Cys, hence, for this assay, we used a mutant version of the protein, NBD<sup>Mut</sup>, which contains a single tryptophan at 172 position and a single cysteine at 198 position (Figure 4.3) and labeled the single cysteine (C198) with TNB dye for FRET distance measurements.

To label the protein, 50  $\mu$ M protein was unfolded by 10 times dilution into unfolding buffer (8 M Urea, 20 mM Tris-HCl buffer pH 8.0), followed by concentrating it back to 50  $\mu$ M. To prepare a 0.5 mM 5,5'-dithiobis(2-nitrobenzoic acid) or DTNB solution, the appropriate amount of DTNB was mixed in the pH 8.0 unfolding buffer. DTNB will make the solution acidic, and hence DTNB will dissolve only when the pH of the solution is adjusted back to pH 8 by the addition of 1 M NaOH (pH of the solution must not go above pH 8). The above step is very crucial because the thiol group of Cys will only get labeled by TNB when the dye is in DTNB form; above pH 8 the DTNB dissociates into TNB, which cannot label the thiol group of the protein. To initiate the labeling reaction, 100  $\mu$ l of 0.5 mM DTNB solution was slowly mixed with 1ml of unfolded protein. The protein-DTNB solution was kept for 30 min in the dark to complete the labeling reaction. The C198-TNB labeled unfolded protein was then desalted in pH 7.5 native buffer to separate the free TNB and DTNB using a 5 ml desalting column attached to the AKTA Pure FPLC system (GE Healthcare). The labeling of the protein was confirmed by measuring the mass of labeled and unlabeled protein using ESI-MS. The efficiency of labeling was measured to be >98%.

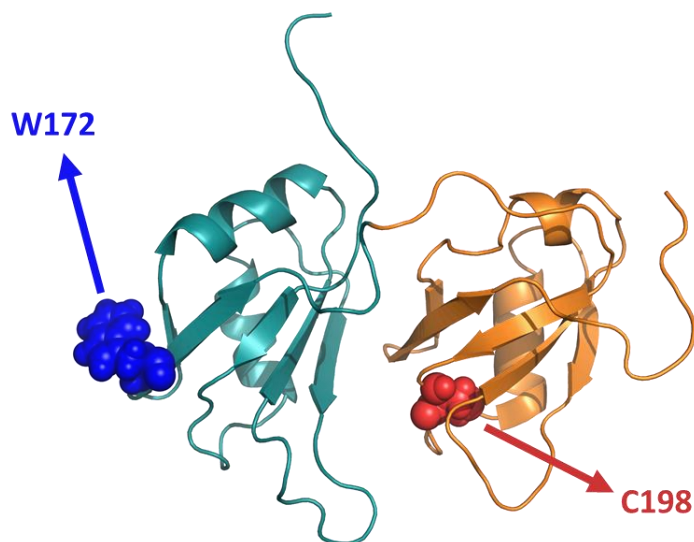


Figure 4.3. Structure of  $NBD^{Mut}$ . The two RNA recognition motifs, RRM 1 and RRM 2, are shown in cyan and orange color, respectively. The blue and red color spheres represents the sole tryptophan (W172) and cysteine (C198), respectively, in domain I and domain II of the protein.

For FRET assay, equimolar stocks of 20  $\mu\text{M}$  protein concentration were prepared for labeled and unlabeled protein, where the concentration for labeled and unlabeled protein was determined by Bradford assay. The labeled and unlabeled protein was then diluted into different buffer conditions for the N form (pH 7.5) and the A form (pH 3.0). The samples were equilibrated for 30 min, followed by acquiring the Trp fluorescence spectra of unlabeled and labeled protein identically. The Forster's distance ( $R_0$ ) between the W172 and C198-TNB was calculated using the following equation,

$$R_0 = 0.211[Q_D J \kappa^2 n^{-4}]^{\frac{1}{6}} \quad (\text{eq 2})$$

where,  $Q_D$  is the quantum yield of the donor,  $J$  is the overlap integral defined as the spectral overlap between the donor's fluorescence emission spectrum and the acceptor's absorbance spectrum,  $\kappa^2$  is the orientation factor, and its value was taken to be 2/3, assuming that the orientation of the donor and the acceptor is random with respect to each other.  $n$  is the refractive index of the medium, observed to be 1.332. The value of  $Q_D$  was determined experimentally as described previously<sup>11</sup> from the mean fluorescence lifetime and fluorescence emission spectra. The value of  $Q_D$  in the N form and the A form is 0.24 and 0.21, respectively. The value of  $J$  is determined by,



$$J = \frac{\int F(\lambda)\varepsilon(\lambda)\lambda^4 d\lambda}{\int F(\lambda)d\lambda} \quad (\text{M}^{-1}\text{cm}^{-1}\text{nm}^4) \quad (\text{eq 3})$$

where,  $F(\lambda)$  represents the fluorescence emission spectrum of W172, and  $\varepsilon(\lambda)$  represents the absorbance spectrum of C198-TNB (Figure 4.4).

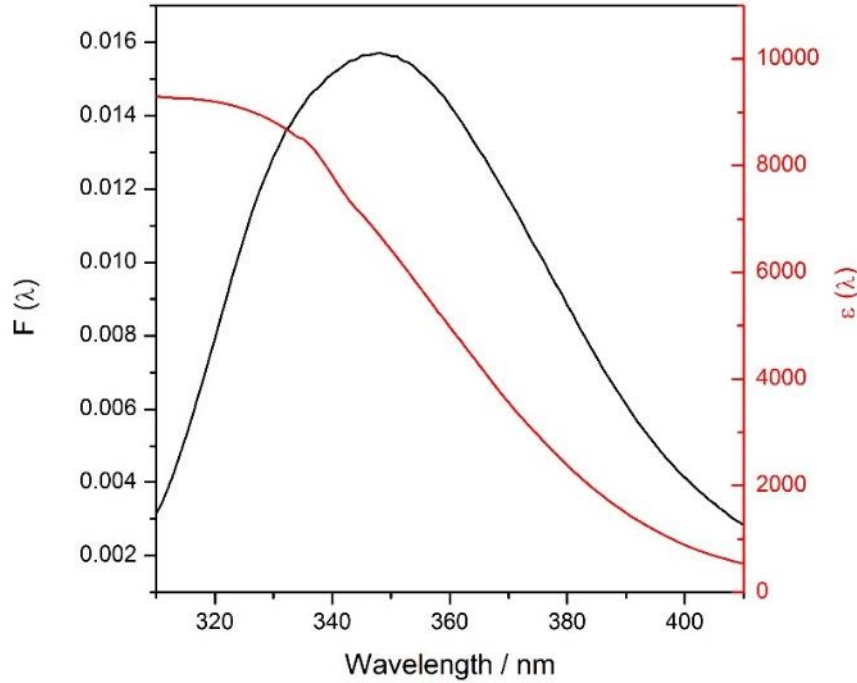


Figure 4.4. Determination of overlap integral ( $J$ ). The fluorescence emission spectra ( $F(\lambda)$ ) of W172 in black and the absorbance spectra of C198-TNB ( $\varepsilon(\lambda)$ ) in red are shown according to the left and right y-axis, respectively. The area under the fluorescence spectrum has been normalized to unity. The absorbance spectrum has been divided with the molar concentration of protein to get the  $\varepsilon(\lambda)$ .

The FRET efficiency ( $E$ ) and donor-acceptor distance ( $R$ ) in different conditions was quantified using the following equations,

$$E = 1 - \frac{F_{DA}}{F_D} \quad (\text{eq 4})$$

$$R = R_0 \left( \frac{1-E}{E} \right)^{\frac{1}{6}} \quad (\text{eq 5})$$

where,  $F_D$  is the fluorescence signal of the donor in the absence of the acceptor,  $F_{DA}$  is the fluorescence signal of the donor in the presence of the acceptor, and  $R_0$  is the Forster's distance.

### Cys-IAEDANS labeling

The sole cysteine (C198) of NBD<sup>Mut</sup> was labeled with 5-((2-[(iodoacetyl)amino]ethyl)amino)naphthalene-1-sulfonic acid or 1,5-IAEDANS fluorescent dye for Stokes shift measurement study. NBD<sup>Mut</sup> was unfolded and incubated with a 20 fold molar excess of 1,5-IAEDANS in an unfolding buffer containing 8 M urea and 20 mM Tris at pH 8.0. The reaction mixture was kept in the dark with continuous stirring at room temperature for 4 hrs. After the labeling reaction was complete, the C198-DANS labeled protein was desalted in pH 7.5 native buffer to separate the free dye using a 5 ml desalting column attached to the AKTA Pure FPLC system (GE Healthcare). The labeling of the protein was confirmed by ESI-MS. The labeled protein was concentrated to ~20  $\mu\text{M}$ , where the concentration for labeled protein was determined by Bradford assay. The labeled protein was then diluted into different buffer conditions for the N form (pH 7.5), A form (pH 3.0) and 8 M urea unfolded form (pH 7.5). The samples were equilibrated for 10 min, followed by acquiring the C198-DANS fluorescence spectra in the wavelength range of 350-450 nm upon excitation at 337 nm.

### ANS binding assay

To perform the 8-anilino-1-naphthalene-sulfonic acid (ANS) binding assay, a 10 mM ANS stock solution was prepared in dimethyl sulfoxide (DMSO) and stored at 4°C. The ANS concentration was determined spectroscopically by measuring the absorbance of ANS solution at 350 nm using the molar extinction coefficient of 5000  $\text{M}^{-1}\text{cm}^{-1}$ .<sup>48</sup> Different concentrations (0-60  $\mu\text{M}$ ) of ANS were prepared in pH 7.5 and pH 3.0 buffer for ANS binding assay. The N form, A form and the wet molten globule form<sup>44</sup> with a final concentration of 2  $\mu\text{M}$  was incubated for 15 min with varying concentrations of ANS (0-150  $\mu\text{M}$ ). For each sample, the ANS fluorescence spectrum was acquired by exciting the fluorophore at 380 nm and collecting the fluorescence emission spectra from 400 to 600 nm. The excitation and emission bandwidths used were 1 nm and 5 nm, respectively.

### Aggregation kinetics monitored by ThT fluorescence measurement

The aggregation kinetics at different concentrations (1-60  $\mu\text{M}$ ) of the A form was monitored by thioflavin T (ThT) binding assay. For the ThT binding assay, a 200  $\mu\text{M}$  ThT stock solution was prepared in deionized water. The concentration ThT solution was determined spectroscopically by measuring the absorbance of the ThT solution, diluted in ethanol, at 416 nm using the molar extinction coefficient of 26600  $\text{M}^{-1}\text{cm}^{-1}$  (Sigma product sheet). The aggregation of TDP<sup>NBD</sup> was initiated by heating a desired concentration of the A

form in pH 3.0 at 60°C. For the assay, a calculated volume of protein (from aggregation vial) and ThT (from ThT stock) was mixed into assay solution (50 mM Tris-HCL pH 8.0 buffer) so that the final concentration of protein and ThT becomes 1  $\mu$ M and 20  $\mu$ M, respectively. The ThT fluorescence spectra were collected within 1 min of the addition of the protein to the assay solution using the following experimental parameters: excitation wavelength, 440 nm; emission wavelength, 450-550 nm.

### Analysis of aggregation kinetics using two-step model

All aggregation kinetic traces were analyzed by a two-step model as described previously<sup>49</sup> using equation,

$$B_t = A_0 - \frac{\frac{k_1}{k_2} + A_0}{1 + \frac{k_1}{k_2 A_0} e^{(k_1 + k_2 A_0)t}} \quad (eq\ 6)$$

where,  $B_t$  is the observed ThT fluorescence signal,  $A_0$  is the amplitude of the signal change,  $k_1$  is the rate constant of nucleation and  $k_2$  is the rate constant for elongation. This aggregation model comprises of two steps: first is slow nucleation step followed by fast elongation step. This model is an oversimplified, minimal model that is sufficient and complex enough to explain the aggregation process that involves nucleation and elongation steps.

### Seeding assay

For the seeding assay, 20  $\mu$ M A form was incubated for 7 days at 60°C. A fresh 20  $\mu$ M A form aggregation reaction was prepared, and a relevant volume of the sample was removed and replaced by an equal volume of preformed fibril seed to obtain 30% seeding. The aggregation kinetics of the unseeded and seeded reaction was monitored using the ThT fluorescence measurements.

### Fluorescence microscopy assay

We visualize the ThT bound amyloid fibrils by fluorescence microscopy using the EVOS FL imaging system (Thermo Fisher Scientific). For the assay, 60  $\mu$ M A form was incubated for 24 hours at 60°C to obtain amyloid fibrils. The incubated sample was mixed with the ThT assay solution so that the final concentration of protein and ThT becomes 30 $\mu$ M and 100 $\mu$ M, respectively. 10 $\mu$ l of the protein-ThT solution was placed on a fluorescence microscopy compatible slide and visualized using a 100x oil emulsion objective.

### Transmission electron microscopy assay

Different forms of TDP<sup>NBD</sup> (N form, A form, and amyloid form) were visualized using transmission electron microscopy (TEM) by staining the sample with a negative stain. For the assay, 10  $\mu$ l of protein sample was carefully placed on a 300-mesh formvar carbon-coated copper grid (Electron Microscopy Science). The sample was allowed to adhere to the grid for 5 min, and excess solution was removed carefully from the edge of the grid using a lint-free paper. For negative staining of the sample, 10  $\mu$ l of freshly prepared 2% uranyl acetate solution was placed on the TEM grid for 90 s, followed by the removal of excess solution. Finally, the grid was washed with deionized water and air-dried. The sample grids were examined on the Technai-T20 transmission electron microscope at an accelerating voltage of 200 kV.

### 4.3 Results and Discussion

#### 4.3.1 TDP<sup>NBD</sup> is amyloidogenic under stress

TDP<sup>NBD</sup> at under low-pH stress (pH 3) forms an acidic form of protein (A form), which forms amyloid fibrils upon thermal stress (Figure 4.5). In the present study, we investigate the mechanism of heat induced formation of amyloid fibrils by TDP<sup>NBD</sup> at low pH in absence of any salt (protein was desalted in glycine-HCl pH 3 buffer with no additional salts). To generate the amyloid fibrils, 60  $\mu$ M A form was heated at 60°C for 24 hours. We compared the amyloidogenicity of the A form and the amyloid form by monitoring the thioflavin T (ThT) binding capacity of the A form and the amyloid form by mixing the protein in ThT assay buffer, with final protein:dye concentration of 1:20  $\mu$ M, and acquiring the fluorescence spectra of ThT dye (Figure 4.5A). ThT is a fluorescent dye, commonly used as a probe to monitor amyloid fibril formation.<sup>50</sup> ThT, in its unbound form, has a very low quantum yield with the wavelength of maximum fluorescence emission at 420 nm upon excitation at 340 nm. However, upon binding with the cross beta structure of the amyloid fibrils, the quantum yield of ThT dye increases dramatically with a blue shift in the wavelength of maximum fluorescence emission to 482 nm upon excitation at 440 nm. We observed that the A form has a flat ThT fluorescence spectrum, indicating that the ThT dye does not bind to A form. However, for the heated amyloid fibril form, the ThT fluorescence increases substantially characteristic for amyloid fibril formation.

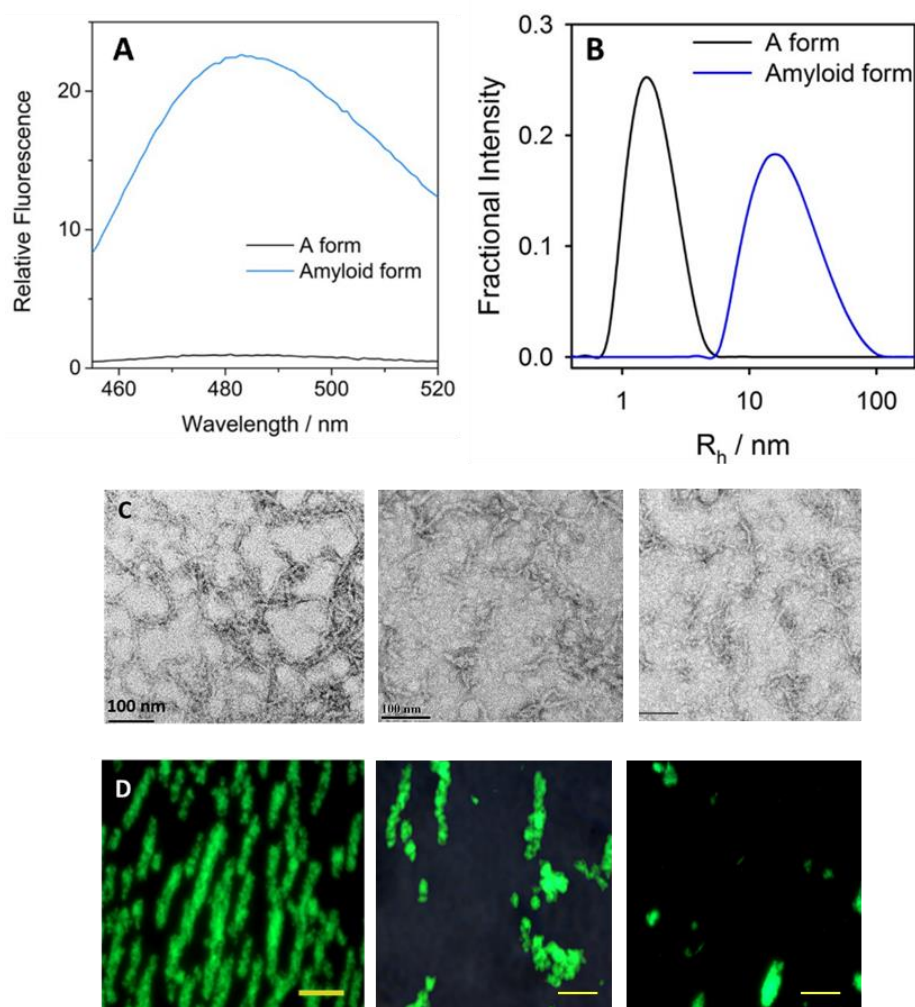


Figure 4.5. Low-pH stress-induced A form is amyloidogenic. The A form is the TDP<sup>NBD</sup> protein desalted in a glycine-HCl pH 3 buffer containing no other salt, while the amyloid form is the A form (60  $\mu$ M) heated at 60°C for 24 hours. (A) ThT fluorescence emission spectra upon binding with the A form and the amyloid form. The final protein:dye ratio is 1:20. (B) Comparison of the distribution of the hydrodynamic radii of the A form and the amyloid form. (C) Three representative TEM micrographs for the amyloid form. The scale bar corresponds to 100 nm (D) Three representative fluorescence micrographs for the amyloid form (labeled with ThT dye). The scale bar corresponds to 10  $\mu$ m.

To assess the formation of higher-order fibrillar structure, we compared the hydrodynamic radii of the A form and the amyloid fibril using dynamic light scattering measurements (Figure 4.5B). We observed that the A form has a mean hydrodynamic radius of ~2 nm. Conversely, for the amyloid form, the apparent mean hydrodynamic radius increases to ~20 nm with a broad distribution, indicating for the formation of higher-order structures with heterogeneous nature.

For understanding the morphology of the amyloid fibrils, we analyzed the micrographs obtained by transmission electron microscopy (TEM) and fluorescence microscopy (Figure 4.5C and 4.5D). For the TEM assay, we negatively stained the protein sample on formvar carbon coated copper grids and analyzed on the transmission electron microscope at an accelerating voltage of 200 kV. The analyzed TEM micrograph (Figure 4.5C) shows that the amyloid form has tangled amyloid fibril-like morphology. To visualize the amyloid fibrils using fluorescence microscopy, we labeled the protein sample with ThT dye and analyzed the sample on a fluorescence microscope. The fluorescence micrograph (Figure 4.5D) suggests that these tangled amyloid fibrils also form higher-order fibrillar structures.

### 4.3.2 Amyloid precursor A form has disrupted tertiary packing

It is essential to characterize the nature of the A form for understanding the initial molecular events in the aggregation pathway of TDP<sup>NBD</sup> before investigating the mechanism of its aggregation. Perceiving the similarities and differences between the A form and the native (N) form is crucial for deducing the initial structural transformation that makes the protein aggregation-prone. To assess the monomeric or oligomeric nature of the A form, we compared the hydrodynamic radii ( $R_h$ ) of the N form and the A form using dynamic light scattering (Figure 4.6). The A form has a mean  $R_h$  of  $\sim 2$  nm similar to the mean  $R_h$  of  $\sim 2.1$  nm of the monomeric N form. This result indicates that the A form remains in a monomeric state. Interestingly in a previous study, it was shown that the structural organization of the A form is highly sensitive to the ionic strength of the solution.<sup>44,45</sup> In as low as 3-5 mM of KCl, A form forms molten-globular oligomers and at  $\sim 20$  mM KCl it forms worm like amyloid fibrils. Here we show that in absence of any salt, A form exist as a monomer.

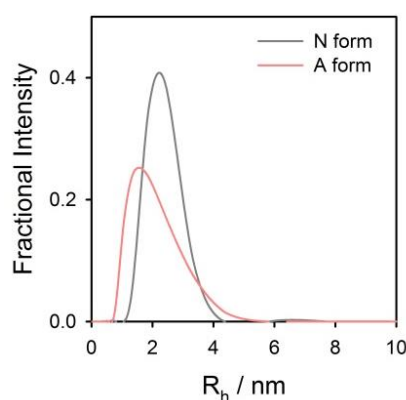


Figure 4.6. Comparison of the distribution of the hydrodynamic radii ( $R_h$ ) of the N form and the A form measured by dynamic light scattering. The A form and the N form have similar  $R_h$  of  $\sim 2$  nm.

For global structural analysis of the N form and the A form, we collected the far-UV and near-UV circular dichroism (CD) spectra to compare the global secondary and tertiary structure, respectively (Figure 4.7). The far-UV CD spectrum of the A form overlaps with the far-UV CD spectrum of the N form (Figure 4.7A). The mean residual ellipticity (MRE) of the A form at 222nm is  $-5000 \text{ deg cm}^2 \text{ dmol}^{-1}$  similar to the MRE of the N form  $-5100 \text{ deg cm}^2 \text{ dmol}^{-1}$ . However, there is an apparent disagreement between the near-UV CD spectrum of the A form and the N form (Figure 4.7B), where the MRE of the A form at 270nm is  $-25 \text{ deg cm}^2 \text{ dmol}^{-1}$  which is substantially less than the MRE of the N form ( $-65 \text{ deg cm}^2 \text{ dmol}^{-1}$ ). These results indicate that the A form retains the N-like secondary structure; however, the tertiary structure packing in the A form is markedly disrupted compared to the N form.

For a profound assessment of the secondary and tertiary structural characteristics of the A form, we analyzed the secondary and tertiary structure stabilities by monitoring the urea-induced equilibrium unfolding transition using far-UV CD and intrinsic tryptophan fluorescence, respectively (Figure 4.8A & 4.8B). All the urea-induced unfolding transitions showed sigmoidal curves and were analyzed using a two-state model (eq 1). The calculated free energy of the unfolding of the secondary structure for the N form is  $4.9 \text{ kcal mol}^{-1}$  and the A form is  $4 \text{ kcal mol}^{-1}$ . However, the calculated free energy of unfolding of the tertiary structure for the N form is  $5.3 \text{ kcal mol}^{-1}$  and the A form is  $1.5 \text{ kcal mol}^{-1}$ . These results show that the secondary structure content in the A form is marginally less stable compared to the N form, while the stability of the tertiary structure in the A form is severely compromised. These results complement well with the comparison of the global secondary and tertiary structure in the N form and the A form.

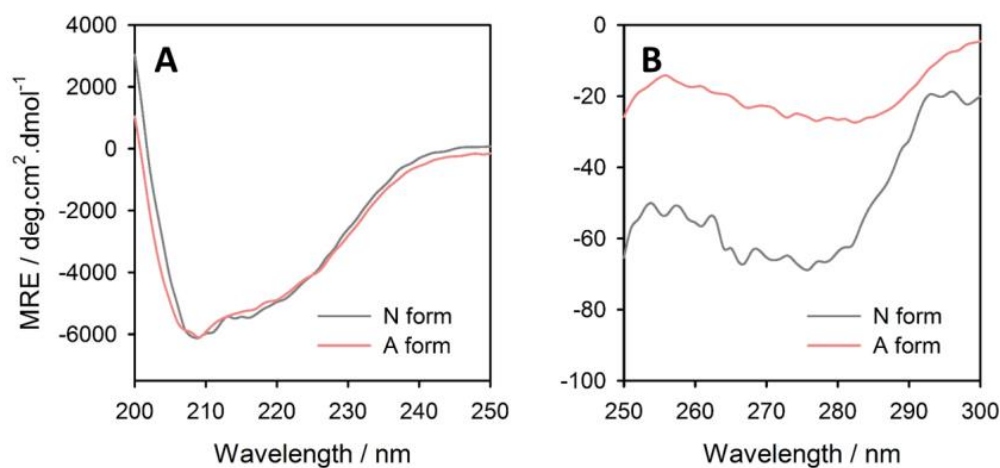


Figure 4.7. The global secondary structure (A) and the tertiary structure (B) comparison of the N form and the A form measured by far-UV and near-UV circular dichroism.

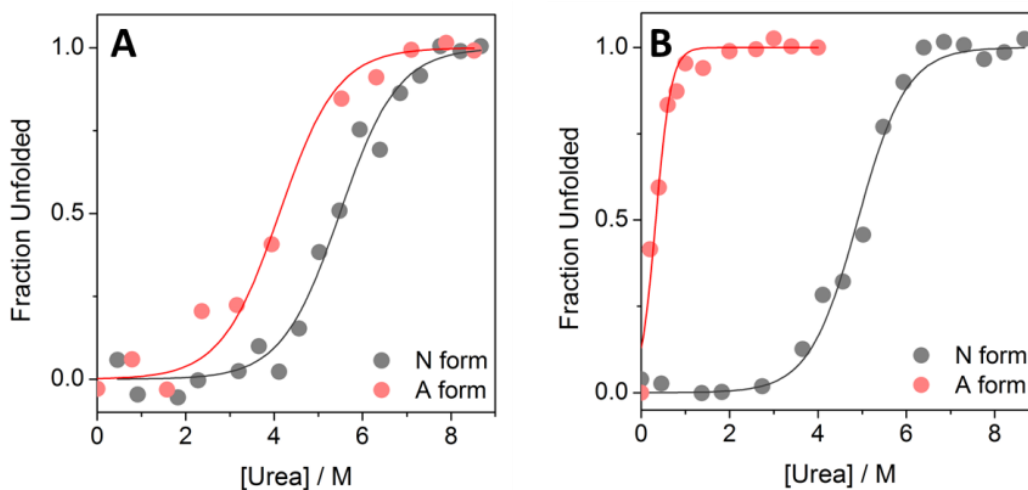


Figure 4.8. Urea-induced equilibrium unfolding transition of the N form and the A form, monitored by the change in far-UV CD signal at 222 nm for the secondary structure stability (A) and change in fluorescence signal of intrinsic tryptophan at 340 nm for the tertiary structure stability (B).

### 4.3.3 Disruption of the tertiary structure is coupled with the structural expansion

When  $\text{TDP}^{\text{NBD}}$  is exposed to low-pH stress, the earliest step is the unlocking of the tertiary structure. However, to investigate whether this structural unlocking is coupled with the structural expansion, we employed fluorescence resonance energy transfer (FRET) assay as a molecular ruler for interatomic distances. We used  $\text{NBD}^{\text{Mut}}$  for this study, which is a mutant variant of  $\text{TDP}^{\text{NBD}}$  with a single tryptophan (W172) and a single cysteine (C198). The TNB absorbance spectrum overlaps with the fluorescence spectrum of the Trp, and hence they form a FRET pair where Trp act as a donor fluorophore, and the TNB act as an acceptor (TNB act as a quencher of Trp fluorescence). We labeled the C198 of  $\text{NBD}^{\text{Mut}}$  with DTNB to make the Trp-TNB FRET pair and experimentally determined the values of Forster's distance ( $R_0$ ) between this FRET pair in the N form (26.5 Å) and the A form (25.9 Å). We monitored the W172 fluorescence of TNB unlabeled and TNB labeled protein in the N form and the A form (Figure 4.9A & 4.9B) and calculated their respective FRET efficiency using eq 4. The FRET efficiencies were further converted to the FRET distances by using eq 5. The calculated FRET distance between W172 and C198-TNB in the N form is 25.5 Å, which increases to 27.9 Å in the A form, suggesting that the A form is an expanded conformation of the  $\text{NBD}^{\text{Mut}}$  with loosened tertiary structure and N-like secondary structure.



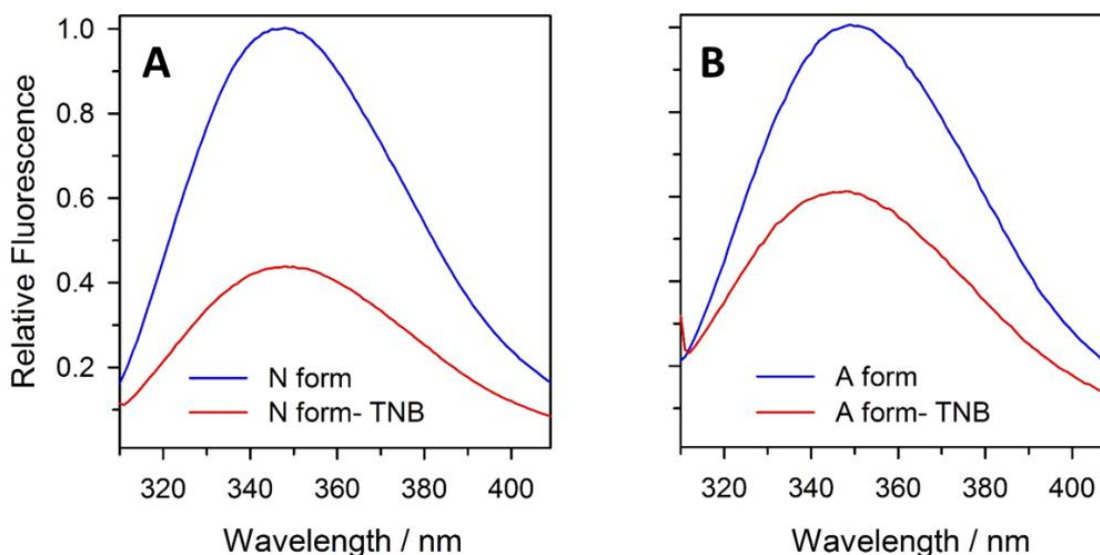


Figure 4.9. The intra-molecular distances between W172 and C198-TNB was monitored by FRET in the N form (A) and the A form (B). The respective change in the fluorescence spectra of W172 in the TNB labeled and unlabeled protein are shown.

#### 4.3.4 The expanded A form resembles a dry molten globule-like state

We compared the hydrophobicity of the protein core in the N and the A form by monitoring the change in Stokes shift of the W172 fluorescence (in domain I) and Cys198-DANS fluorescence (domain II). Stokes shift is a sensitive and excellent probe for the estimation of the polarity of the surrounding environment of the fluorophore. If a fluorophore is surrounded by a polar environment (water molecules), the wavelength of maximum fluorescence emission will be red-shifted (large Stokes shift). Conversely, if a nonpolar environment surrounds the fluorophore, then the wavelength of maximum fluorescence emission will be blue-shifted compared to the above (small Stokes shift). Hence, any deviation in the polarity of the surrounding medium of the fluorophore will result in the blue shift or red shift in wavelength of maximum fluorescence emission depending on the polarity of the new surroundings. In this study, we have used the same NBD<sup>Mut</sup> protein, which was used in the FRET measurement. To monitor the hydrophobicity in the domain I, we compared the fluorescence spectrum of W172 in the N form, A form, and the unfolded form (Figure 4.10A).

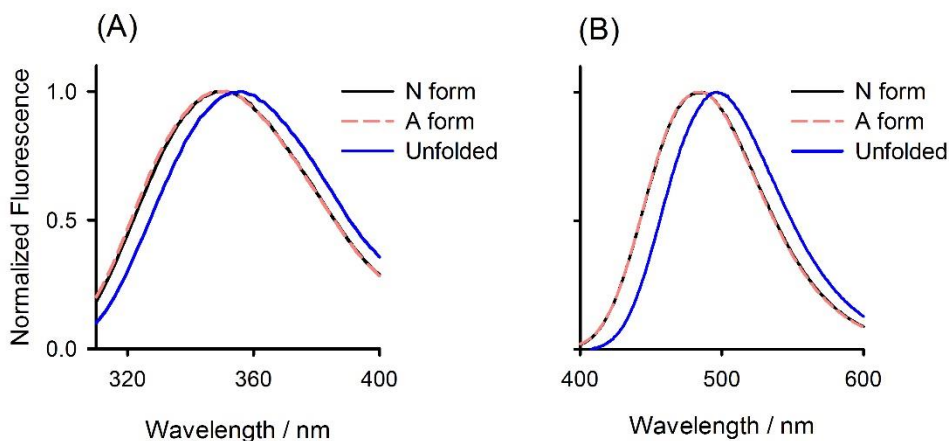


Figure 4.10. The fluorescence emission spectra of W172 (A) and C198-DANS (B). The wavelength of maximum fluorescence emission was normalized to unity for all the spectra for comparison of Stokes shift in the N form, A form and the unfolded form.

The wavelength of maximum fluorescence emission in the 8M urea unfolded form (where the W172 is completely exposed to polar water) is 356nm. However, the wavelength of maximum fluorescence emission blue shifts to 348nm in both N form and the A form indicating that the A form has N-like hydrophobicity in the domain I. To monitor the hydrophobicity in the domain II, we labeled C198 with 1,5-IAEDANS dye and compared the fluorescence spectrum of C198-DANS in the N form, A form, and the unfolded form (Figure 4.10B). The wavelength of maximum fluorescence emission remains the same (480 nm) in both N and the A form, and blue-shifted from the wavelength of maximum fluorescence emission in the unfolded form (495 nm). These observations indicate that the expansion in the A form is not coupled with the solvation of the protein core in both domains I and II, characteristic to the DMG-like state.

The hydrophobicity within the protein core can also be monitored by 8-Anilino-naphthalene-1-sulfonic acid (ANS) binding assay. ANS is a dye that binds to the exposed hydrophobic patches in the protein resulting in the blue shift and dramatic increase in the fluorescence spectra. ANS binding assay is a standard method for investigating the wet molten globule (WMG)-like intermediates, where the core of the protein becomes solvated, exposing the hydrophobic patches to the solvent. However, this is not true for the DMGs where the protein core remains dry, and hence the ANS dye and hydrophobic patches do not come into contact. Previously it was shown that at under similar conditions at low pH but in the presence of 3-5 mM of KCl, DMG is not stable, instead a WMG is populated, which forms

molten globular oligomers on the timescale of many days.<sup>45</sup> Here, we performed the ANS binding assay in the N form, A form, and WMG form of TDP<sup>NBD</sup> (Figure 4.11).

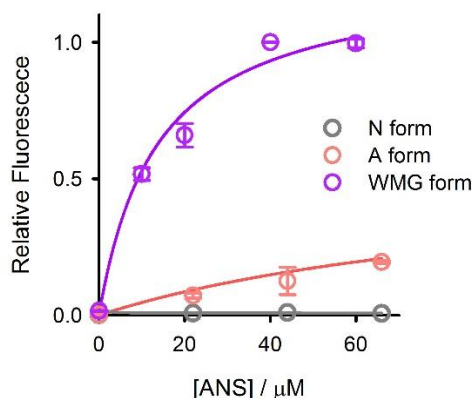


Figure 4.11. Comparison of the ANS dye fluorescence upon binding with the N form, A form and WMG form. The relative change in the intensity of ANS fluorescence at 480 nm upon excitation at 380 nm, where ANS fluorescence at 2:40  $\mu\text{M}$  protein:dye concentration was normalized to unit, is plotted against the varying concentrations of ANS.

The WMG form of TDP<sup>NBD</sup> shows maximum ANS fluorescence (normalized to the unit) at protein:dye concentration of 2:40  $\mu\text{M}$ . The relative ANS fluorescence is markedly less in the A form (0.2) and comparable to the ANS fluorescence in the N form (0). We anticipate that the slight increase in the ANS fluorescence in the A form compared to the N form could be a consequence of the exposure of minor part of the protein interior, nevertheless, a major part of the protein remains unexposed to the solvent. These results show that the A form has an expanded conformation, N-like secondary structure, disrupted tertiary structure and desolvated core, typical of a DMG-like state.

#### 4.3.5 Amyloid fibril formation occurs via nucleation-elongation aggregation mechanism

Understanding the mechanism of aggregation is essential for deciphering the molecular-level information about the temporal events taking place during the aggregation reaction. This can further aid in designing drugs that can target the crucial steps of aggregation, stalling or speeding up the reaction to a non-toxic intermediate or final product. Generally, the aggregation mechanism can be divided into two broad categories: nucleation dependent aggregation and linear aggregation. The nucleation dependent aggregation is characterized by the appearance of a lag phase, presence of a critical aggregation concentration, and the abolishment of the lag phase upon seeding the aggregation reaction.<sup>51-53</sup> However, in linear aggregation mechanism, aggregation proceeds via interaction of two monomers and

subsequent addition of monomeric proteins with no observable lag phase, no critical concentration required for initiation of aggregation and no seeding effect.<sup>54,55</sup>

To investigate the mechanism of amyloid fibril formation, we examined the dependence of the kinetics of amyloid fibril formation on protein concentration monitored by an increase in ThT fluorescence with time. For 15-30  $\mu\text{M}$  protein concentration, the aggregation kinetics proceed in a sigmoidal manner, which indicates the presence of a lag phase (Figure 4.12A). The lag phase represents the time required for the formation of the nucleus (nucleation) upon which the aggregation occurs by subsequent monomer addition. Nucleation is a thermodynamically unfavorable step, where the rate of monomer dissociation is higher than the rate of monomer association. The length of the lag phase is proportional to the energetic barrier of nucleation and depends on protein concentration.<sup>56</sup>

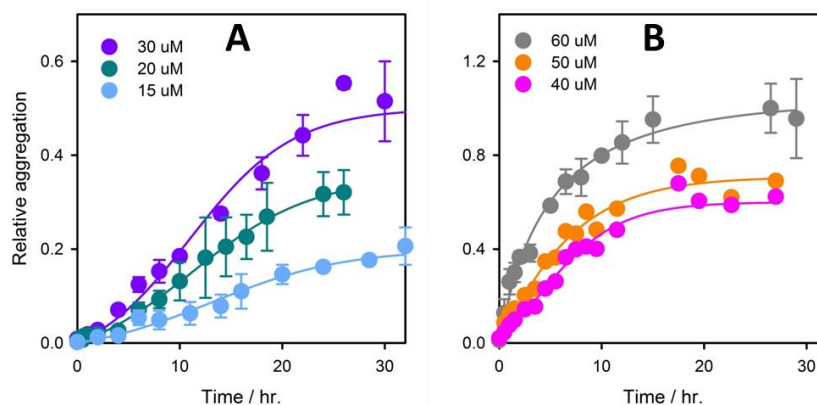
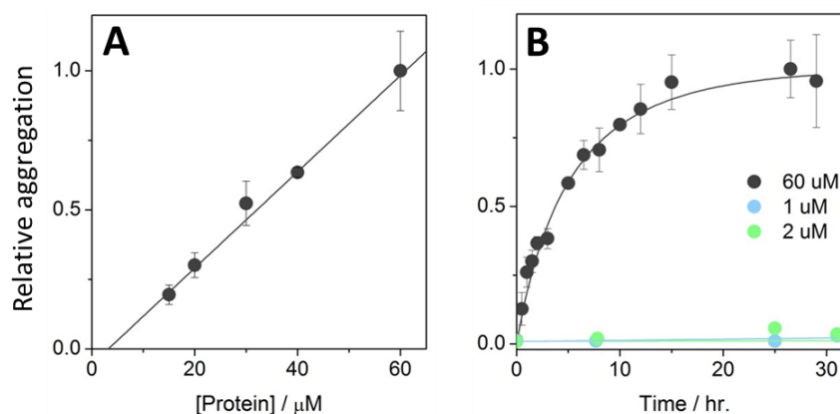


Figure 4.12. ThT fluorescence monitored aggregation kinetics. ThT fluorescence at 482 nm (wavelength of fluorescence emission maximum) have been plotted against the time of aggregation. Aggregation kinetic traces follow sigmoidal curves in the protein concentration range of 15-30  $\mu\text{M}$  (A) and hyperbolic curves in the protein concentration range of 40-60  $\mu\text{M}$  (B). All aggregation kinetic traces are normalized relative to the aggregation kinetics of the 60  $\mu\text{M}$  protein, where the maximum signal of the 60  $\mu\text{M}$  protein aggregation kinetics is normalized to the unit. The continuous lines through the data are the least square fit to the eq 6.

At higher concentrations of protein (40-60  $\mu\text{M}$ ), the aggregation kinetic traces appear as hyperbolic curves (Figure 4.12B). The disappearance of the lag phase at higher protein concentration does not indicate any change in the aggregation mechanism; instead, it is a result of higher nucleation rates.

The nucleation reaction has a dependence on protein concentration controlled by the monomer association and dissociation rate constants.<sup>53</sup> In an aggregation reaction, if the

concentration of the protein is below the critical aggregation concentration, the energetic barrier of nucleation will never be crossed, and the aggregation will not proceed. To determine the critical aggregation concentration, we utilized the amplitude of the ThT fluorescence monitored aggregation kinetics (from Figure 4.12) as a measure of the amount of aggregate formed at different protein concentrations (15-60  $\mu\text{M}$ ). The amount of aggregates increases linearly with an increase in protein concentration (Figure 4.13A). Using the linear least-square fit, we estimated that the protein would not aggregate below the critical aggregation concentration of 3  $\mu\text{M}$ . To validate this, we monitored the aggregation kinetics at 1  $\mu\text{M}$  and 2  $\mu\text{M}$  protein concentration and compared with aggregation kinetics at 60  $\mu\text{M}$  (Figure 4.13B), and as expected, we do not see any increase in ThT fluorescence in the monitored timescale. Our results show that the aggregation of the A form occurs only above a critical aggregation concentration indicative of nucleation dependent aggregation mechanism.



*Figure 4.13. Determination of critical aggregation concentration. The amplitude of total signal change obtained from ThT monitored aggregation kinetics from figure 4.12 was plotted as a function of protein concentration (A). The amplitudes are normalized relative to the amplitude observed for 60  $\mu\text{M}$  protein. The continuous line through the data is the least square fit to a straight line. The aggregation kinetic traces for 1  $\mu\text{M}$  and 2  $\mu\text{M}$  protein are compared with aggregation kinetics of 60  $\mu\text{M}$  protein (B). The continuous lines through the data are the least square fit to the eq 6.*

We further validated the aggregation mechanism by performing the seeding assay, where the preformed fibrils are provided as seeds to the aggregation reaction. A seeding assay is a key experiment for the estimation of the aggregation model. When an aggregation reaction is seeded with preformed fibril, the elongation occurs directly on the provided seeds. The seeding evades the nucleation step, and hence there is no lag phase in the aggregation kinetics. We examined the aggregation kinetics of unseeded A form and 30% seeded A form by

monitoring the increase in the ThT fluorescence as a function of time (Figure 4.14). We observed that for the unseeded protein, the aggregation kinetics proceeds with an observable lag phase. However, for the seeded aggregation reaction, the lag phase is abolished due to the presence of preformed fibrils acting as a proxy for the nucleus. These results provide conclusive evidence for the nucleation dependent aggregation mechanism.

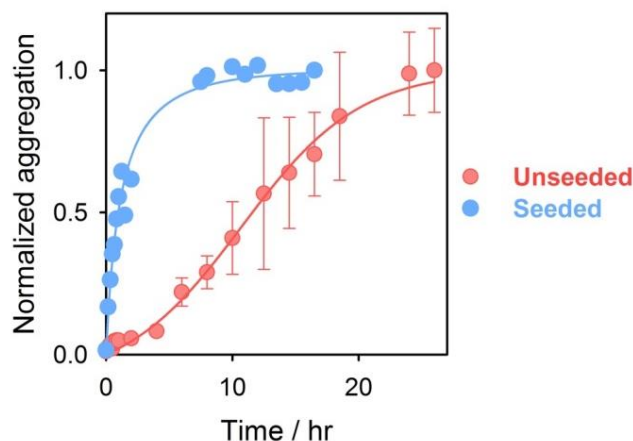


Figure 4.14. Effect of seeding. ThT fluorescence monitored aggregation kinetics of 20  $\mu\text{M}$  A form in the presence and the absence of 30% preformed fibril. The continuous lines through the data are the least square fit to the eq 6.

Table 4.1. Determination of kinetic parameters of A form aggregation. Values of nucleation rate constant ( $k_1$ ), elongation rate constant ( $k_2$ ) and the time at which 50% aggregation is completed ( $t_{50}$ ).

| [Protein] / $\mu\text{M}$ | $^\dagger k_1 / \text{h}^{-1}$ | $^\ddagger k_2 / \mu\text{M}^{-1}\text{h}^{-1}$ | $^\S t_{50} / \text{h}$ |
|---------------------------|--------------------------------|---|-------------------------|
| 15                        | 0.018                          | 0.68  | 16                      |
| 20                        | 0.022                          | 0.41  | 13                      |
| 30                        | 0.035                          | 0.18  | 11.5                    |
| 40                        | 0.079                          | 0.19  | 7                       |
| 50                        | 0.129                          | 0.15  | 5                       |
| 60                        | 0.231                          | 0.19  | 4                       |

$^\dagger k_1$  and  $^\ddagger k_2$  was determined using eq 6.

$^\S t_{50}$  was determined from Figure 4.12.

In addition to the above results, the analysis of all aggregation kinetic traces from Figure 4.12, using the two-step model (eq 6), also validates that the amyloid fibril formation of the A form occurs by nucleation-elongation aggregation mechanism. The fit generated two crucial parameters: nucleation and elongation rate constants for the aggregation kinetics at different protein concentrations (Table 4.1). We observed that the nucleation rate constant increases with an increase in the initial protein concentration. This is expected since nucleation is a pseudo-first-order and an association reaction, so higher the initial protein concentration higher the nucleation rate. However, the rate constant for elongation does not change substantially with an increase in protein concentration, suggesting that the aggregation kinetics is majorly affected by the formation of the nucleus and once the nucleus has formed, the elongation happens at similar pace regardless of the initial monomer concentration.

Furthermore, it is important to note that there can be other secondary processes involved in nucleation dependent aggregation mechanism such as secondary nucleation and fragmentation. Therefore, it is important to carefully analyze the aggregation kinetics for the presence of any other secondary pathway. For this purpose, we extracted the  $t_{50}$  (time at which 50% aggregation is completed) values of the ThT fluorescence monitored aggregation kinetics at different protein concentrations from Figure 4.12 (Table 4.1) and plotted a double log plot of the  $t_{50}$  values against the protein concentration (Figure 4.15). The slope of this double log plot is known as the scaling exponent, which is used as a measure to predict the aggregation model.<sup>57</sup> The double log plot fits a straight line with the scaling exponent of -1.0, indicating that the A form aggregation occurs via nucleation elongation mechanism with no secondary processes involved.<sup>57</sup>

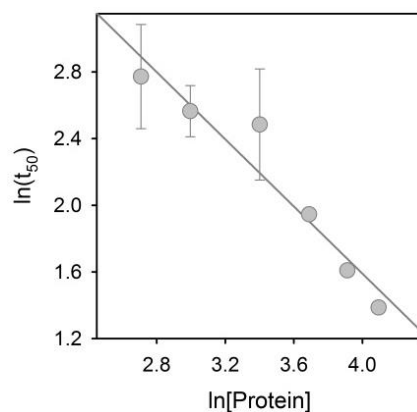


Figure 4.15. The scaling exponent plot for the prediction of the plausible secondary aggregation pathway. A double log plot is generated for the  $t_{50}$  values of the aggregation

kinetics against the protein concentration. The continuous line through the data is a linear least square fit with a slope (scaling exponent) of  $-1.0$ .

#### 4.3.6 Low-pH stress-induced TDP<sup>NBD</sup> aggregation is a three-step process

Based on our results, we propose that the low-pH stress-induced aggregation of TDP<sup>NBD</sup> occurs in three steps (Figure 4.16). Firstly, upon low-pH stress, TDP<sup>NBD</sup> undergoes early structural transformation leading to the formation of a DMG-like A form. The formation of a DMG-like state indicates that the protein has undergone partial unfolding. Therefore it is fair to assume that till this point, all the molecular events are happening in the folding-unfolding pathway of the protein. In the second step, these DMG-like states form the aggregation-prone nucleus upon heating. It appears that the presence of DMG-like states provides conformational flexibility to the protein, sufficient to interact with other partners and form the nucleus for the aggregation.

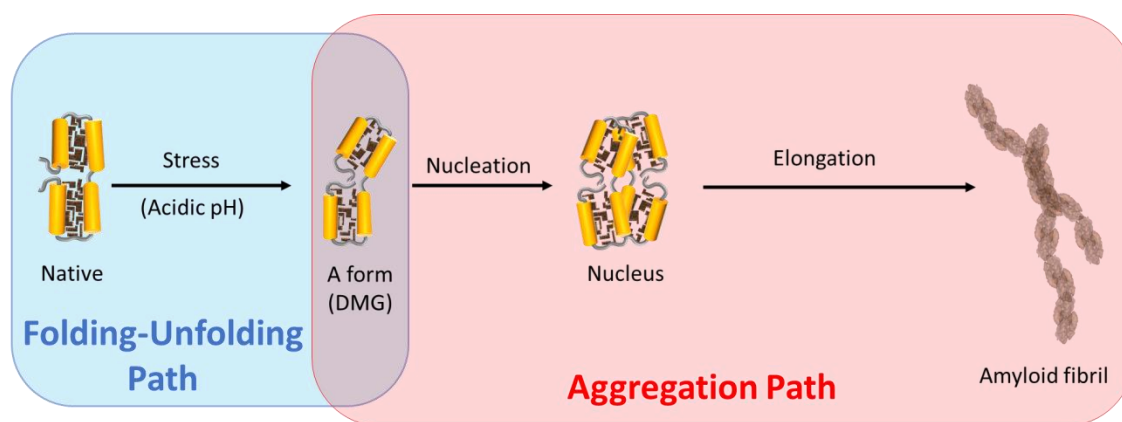


Figure 4.16. Schematic model for the amyloid fibril formation of TDP<sup>NBD</sup>. The conversion occurs in three steps. The first step is the formation of a DMG-like state, which is a part of the protein folding-unfolding pathway. From the second step, which is the nucleation step, the protein diverges to the aggregation path. In the final step, the elongation of the nucleus eventually leads to amyloid fibril formation.

However, in the case of the native protein, these interactions are disfavoured and hence the native protein does not proceed to the aggregation path. The last step is the elongation of the fibril upon the nucleus by subsequent addition of the monomeric protein on either end. The nucleation and fibrillation steps signify the aggregation pathway, where the protein experiences intermolecular non-native interactions leading to the formation of higher order amyloid aggregates. We propose that the initial step, i.e., the formation of DMG-like A form, is the



---

most significant in the aggregation of TDP<sup>NBD</sup> because it is the point where the protein is diverging from the folding-unfolding pathway to the aggregation pathway.

#### 4.4 Conclusions

In this study, we have shown that under low-pH stress, TDP<sup>NBD</sup> form amyloid aggregates with tangled amyloid fibril like morphology. We employed a range of spectroscopic tools to assess the early structural transformation of TDP<sup>NBD</sup> during the low-pH stress. We have shown that the low-pH A form retains N-like secondary structure, but most of its tertiary packing is lost, as evident from the comparison of the global secondary and tertiary structure and their respective stabilities in the N and the A form. With the aid of the FRET measurement, we show that the loosening of tertiary structure in the A form is also accompanied by an expansion in the protein. The Stokes shift analysis and the ANS binding assay, however, show that the loss of tertiary packing and protein expansion is not coupled with the solvation of the protein core. These results provide evidence that the A form is a dry molten globule-like state.

Interestingly, this is the first report where we show that DMGs can act as a starting point for the amyloid aggregation. We further utilized the kinetic tools to investigate the aggregation mechanism. We monitored the aggregation kinetics of the A form at different protein concentrations. We found that the aggregation kinetics of the A form fulfills the three crucial criteria for the nucleation dependent aggregation model: the appearance of the lag phase, presence of the critical aggregation concentration, and disappearance of the lag phase upon seeding the aggregation reaction with preformed fibrils. Moreover, all kinetic traces also fit well to a nucleation-elongation model, validating that the aggregation of the low-pH stress-induced DMG-like A form proceeds via nucleation elongation aggregation mechanism. DMGs are the near-native states with increased conformational dynamics (because of disrupted tertiary structure), this provides more chances for two or more interacting fragments, of different monomers, to come together and form the nucleus. However, for the native protein, the odds for two or more interacting fragments of different monomers, to come together is very unlikely because of the restricted dynamics. Our results indicate that DMGs are not just folding intermediates and the role of DMGs, beyond protein folding, has been underestimated. We demonstrated that DMGs could link the folding-unfolding pathway and the aggregation pathway of proteins. We suggest that the aggregation studies, where the aggregation begins with a native or native-like protein, should be revisited for a plausible involvement of DMG-like states in the initial step of the protein aggregation.

---

## 4.5 References

1. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725-2737.
2. Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23, 4-10.
3. Shakhnovich, E. I., and Finkelstein, A. V. (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28, 1667-1680.
4. Finkelstein, A. V., and Shakhnovich, E. I. (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28, 1681-1694.
5. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
6. Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-<sup>19</sup>F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A* 92, 9318-9322.
7. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* 106, 12289-12294.
8. Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* 111, 4856-4861.
9. Rami, B. R., and Udgaonkar, J. B. (2002) Mechanism of formation of a productive molten globule form of barstar. *Biochemistry* 41, 1710-1716.
10. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* 107, 4955-4960.

11. Acharya, N., Mishra, P., and Jha, S. K. (2016) Evidence for Dry Molten Globule-Like Domains in the pH-Induced Equilibrium Folding Intermediate of a Multidomain Protein. *J Phys Chem Lett* 7, 173-179.
12. Mishra, P., and Jha, S. K. (2017) An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* 121, 9336-9347.
13. Acharya, N., Mishra, P., and Jha, S. K. (2017) A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Phys Chem Chem Phys* 19, 30207-30216.
14. Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* 118, 6380-6392.
15. Yue, Z., and Shen, J. (2018) pH-Dependent cooperativity and existence of a dry molten globule in the folding of a miniprotein BBL. *Phys Chem Chem Phys* 20, 3523-3530.
16. Roche, J., Caro, J. A., Norberto, D. R., Barthe, P., Roumestand, C., Schlessman, J. L., Garcia, A. E., García-Moreno, B. E., and Royer, C. A. (2012) Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* 109, 6945-6950.
17. Fu, Y., Kasinath, V., Moorman, V. R., Nucci, N. V., Hilser, V. J., and Wand, A. J. (2012) Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* 134, 8543-8550.
18. Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002) Enzyme dynamics during catalysis. *Science* 295, 1520-1523.
19. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438, 117-121.
20. Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., Benkovic, S. J., and Wright, P. E. (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* 332, 234-238.

21. Doshi, U., McGowan, L. C., Ladani, S. T., and Hamelberg, D. (2012) Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proc Natl Acad Sci U S A* 109, 5699-5704.
22. Haas, J., Vöhringer-Martinez, E., Bögehold, A., Matthes, D., Hensen, U., Pelah, A., Abel, B., and Grubmüller, H. (2009) Primary steps of pH-dependent insulin aggregation kinetics are governed by conformational flexibility. *ChemBioChem* 10, 1816-1822.
23. Hodkinson, J. P., Radford, S. E., and Ashcroft, A. E. (2012) The role of conformational flexibility in  $\beta$ 2-microglobulin amyloid fibril formation at neutral pH. *Rapid Commun Mass Spectrom* 26, 1783-1792.
24. Jahn, T. R., Parker, M. J., Homans, S. W., and Radford, S. E. (2006) Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* 13, 195-201.
25. Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5, 15-22.
26. Honda, R. P., Xu, M., Yamaguchi, K. I., Roder, H., and Kuwata, K. (2015) A Native-like Intermediate Serves as a Branching Point between the Folding and Aggregation Pathways of the Mouse Prion Protein. *Structure* 23, 1735-1742.
27. Lee, E. B., Lee, V. M., and Trojanowski, J. Q. (2011) Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nat Rev Neurosci* 13, 38-50.
28. Buratti, E., and Baralle, F. E. (2012) TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends Biochem Sci* 37, 237-247.
29. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314, 130-133.
30. Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y., and Oda, T. (2006) TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem Biophys Res Commun* 351, 602-611.

31. Chen-Plotkin, A. S., Lee, V. M., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease. *Nat Rev Neurol* 6, 211-220.
32. Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A., and Patel, B. K. (2019) Molecular Mechanisms of TDP-43 Misfolding and Pathology in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci* 12, 25.
33. Nonaka, T., and Hasegawa, M. (2020) Prion-like properties of assembled TDP-43. *Curr Opin Neurobiol* 61, 23-28.
34. Buratti, E. (2015) Functional Significance of TDP-43 Mutations in Disease. *Adv Genet* 91, 1-53.
35. Zhang, Y. J., Xu, Y. F., Cook, C., Gendron, T. F., Roettges, P., Link, C. D., Lin, W. L., Tong, J., Castanedes-Casey, M., Ash, P., Gass, J., Rangachari, V., Buratti, E., Baralle, F., Golde, T. E., Dickson, D. W., and Petrucelli, L. (2009) Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proc Natl Acad Sci U S A* 106, 7607-7612.
36. Dewey, C. M., Cenik, B., Sephton, C. F., Johnson, B. A., Herz, J., and Yu, G. (2012) TDP-43 aggregation in neurodegeneration: are stress granules the key? *Brain Res* 1462, 16-25.
37. Qin, H., Lim, L. Z., Wei, Y., and Song, J. (2014) TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc Natl Acad Sci U S A* 111, 18619-18624.
38. Gitler, A. D., and Shorter, J. (2011) RNA-binding proteins with prion-like domains in ALS and FTL-D. *Prion* 5, 179-187.
39. Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res* 37, 1799-1808.
40. Ayala, Y. M., Zago, P., D'Ambrogio, A., Xu, Y. F., Petrucelli, L., Buratti, E., and Baralle, F. E. (2008) Structural determinants of the cellular localization and shuttling of TDP-43. *J Cell Sci* 121, 3778-3785.
41. Buratti, E., Brindisi, A., Giombi, M., Tisminetzky, S., Ayala, Y. M., and Baralle, F. E. (2005) TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-

- terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J Biol Chem* 280, 37572-37584.
42. Shiina, Y., Arima, K., Tabunoki, H., and Satoh, J. (2010) TDP-43 dimerizes in human cells in culture. *Cell Mol Neurobiol* 30, 641-652.
  43. Sun, Y., and Chakrabartty, A. (2017) Phase to Phase with TDP-43. *Biochemistry* 56, 809-823.
  44. Pillai, M., and Jha, S. K. (2019) The Folding and Aggregation Energy Landscapes of Tethered RRM Domains of Human TDP-43 Are Coupled via a Metastable Molten Globule-like Oligomer. *Biochemistry* 58, 608-620.
  45. Pillai, M., and Jha, S. K. (2020) Early Metastable Assembly during the Stress-Induced Formation of Worm-like Amyloid Fibrils of Nucleic Acid Binding Domains of TDP-43. *Biochemistry* 59, 315-328.
  46. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.
  47. Agashe, V. R., and Udgaonkar, J. B. (1995) Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* 34, 3286-3299.
  48. Stryer, L. (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J Mol Biol* 13, 482-495.
  49. Morris, A. M., Watzky, M. A., Agar, J. N., and Finke, R. G. (2008) Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry* 47, 2413-2427.
  50. LeVine, H., 3rd. (1993) Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* 2, 404-410.
  51. Ferrone, F. (1999) Analysis of protein aggregation kinetics. *Methods Enzymol* 309, 256-274.

- 
52. Harper, J. D., and Lansbury, P. T., Jr. (1997) Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem* 66, 385-407.
  53. Kodaka, M. (2004) Interpretation of concentration-dependence in aggregation kinetics. *Biophys Chem* 109, 325-332.
  54. Oosawa, F., and Kasai, M. (1962) A theory of linear and helical aggregations of macromolecules. *J Mol Biol* 4, 10-21.
  55. Frieden, C. (2007) Protein aggregation processes: In search of the mechanism. *Protein Sci* 16, 2334-2344.
  56. Kumar, S., and Udgaonkar, J. B. (2010) Mechanisms of amyloid fibril formation by proteins. *Curr Sci*, 639-656.
  57. Meisl, G., Kirkegaard, J. B., Arosio, P., Michaels, T. C., Vendruscolo, M., Dobson, C. M., Linse, S., and Knowles, T. P. (2016) Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat Protoc* 11, 252-272.

## **Chapter 5.**

# **Structural dynamics of the DMG-like amyloid precursor and the amyloid core of TDP<sup>NBD</sup> revealed by H/D exchange coupled to mass spectrometry**



## 5.1 Introduction

The neuronal aggregates of the transactive response DNA binding protein (TDP-43) are the hallmark of the amyotrophic lateral sclerosis (ALS) disease.<sup>1</sup> TDP-43 is a four domain protein consist of an N-terminal domain (0-96 amino acid), a C-terminal domain (262-414 amino acid) and two nucleic acid binding domains (97-261 amino acid). The role of C-terminal and N-terminal domains of TDP-43 have been well studied in the aggregation of TDP-43, where the N-terminal region assists in dimerization of the protein, while the disordered C-terminal region has been shown to undergo liquid-liquid phase separation and possess many pathogenic mutation sites.<sup>2-4</sup> Recent studies show the involvement of the nucleic acid binding domains (TDP<sup>NBD</sup>) in TDP-43 aggregation.<sup>5,6</sup>

Previously we have shown that the low-pH induced dry molten globule (DMG) form of TDP<sup>NBD</sup> forms amyloid fibrils upon thermal stress via nucleation-elongation aggregation mechanism (Unpublished data, Chapter 4) (Figure 5.1). Understanding the structural organization of TDP<sup>NBD</sup> amyloid fibril would be beneficial for targeting the rigid amyloid core of the TDP<sup>NBD</sup> fibril for defibrillation. Though the structural organization of the C-terminal amyloid fibrils of TDP-43 has been well studied utilizing cryo-EM,<sup>7</sup> the structural organization of TDP<sup>NBD</sup> fibrils is still unexplored.

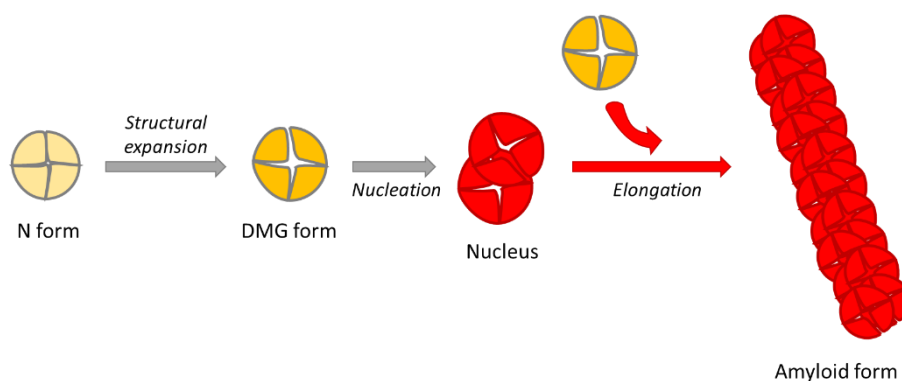


Figure 5.1. Schematic model for TDP<sup>NBD</sup> aggregation.

TDP<sup>NBD</sup> is non-amyloidogenic in its native (N) form; however, the low-pH stress triggers the formation of a structurally expanded DMG form, which is amyloidogenic (Chapter 4). The conversion of the N form to the DMG like intermediate state is a critical molecular event responsible for the nucleation step in the amyloid aggregation of TDP<sup>NBD</sup>. Although structural expansion has been identified as the initial step during the fibrillization of TDP<sup>NBD</sup> (Chapter 4), a high-resolution characterization between the amyloidogenic DMG-form and the

non-amyloidogenic N form is yet to be achieved. A meticulous understanding of the early steps during the fibrillization process and coherent differentiation between the amyloidogenic DMG form and the non-amyloidogenic N form could aid in designing small molecules for the targeted drug delivery to neutralize the amyloid fibrillization at the initial step.

Moreover, despite our understanding of the dynamic nature of DMGs, there is no direct high-resolution evidence for comparison of the dynamics of the N form and the DMG form.<sup>8,9</sup> This is mainly due to the transient nature of DMGs as they have been predominantly observed in kinetic studies and partly due to the existence of a continuous equilibrium between DMGs and the native state.<sup>10-19</sup> High-resolution characterization of the dynamics of DMG-form of TDP<sup>NBD</sup> would, in general, benefit in understanding the nature of DMG-like early unfolding intermediate states.

In this study, we elucidate the structural organization of the amyloid core of the TDP<sup>NBD</sup> amyloid fibril along with deciphering the amyloidogenic differences between the DMG form and the N form to depict different molecular events during the initiation of the aggregation reaction, utilizing a high-throughput method, hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS).

## 5.2 Materials and Methodology

### Protein expression & purification

TDP<sup>NBD</sup> was expressed in *E.coli* BL21 DE3 competent cells and purified as described earlier.<sup>20</sup> Briefly, the protein expression was induced in the transformed cell by 1mM IPTG induction. After 24 h incubation at 20°C, the transformed cells were pelleted down by centrifugation at 4500 g followed by cell lysis in lysis buffer (20 mM Tris-HCl, 30 mM imidazole, 300 mM NaCl, 1 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 10% glycerol, pH 7.5). The cell lysate was centrifuged at 14000 g to separate the cell debris. The supernatant fraction was filtered with a 0.2 µm PVDF membrane filter and subjected for further chromatographic purification by Ni-Sepharose His-Tag affinity purification with an increasing gradient of imidazole (30 mM to 300 mM). The purified fraction was buffer exchanged to the cleavage buffer (20 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, pH 8.0) using HiPrep 26/10 desalting column (GE Healthcare) and incubated for 15 h in the presence of prescission protease for His-tag cleavage. His-tag cleaved TDP<sup>NBD</sup> was further purified by anion-exchange purification using HiPrep Q HP 5 ml column (GE Healthcare) by increasing the gradient of NaCl (50mM to 500mM) on AKTA Pure FPLC system (GE

Healthcare). Finally, the purified protein was buffer exchanged into the storage buffer (20 mM Phosphate, 150 mM KCl, 2% glycerol, 1 mM DTT, pH 7.4 buffer) and stored at  $-30^{\circ}\text{C}$  till further use.

The purity of TDP<sup>NBD</sup> was confirmed by SDS-PAGE and ESI-MS. The calculated mass of TDP<sup>NBD</sup> is 19420 Da. The concentration of protein was determined spectroscopically from the absorbance at 280 nm, using a molar extinction coefficient of  $15470\text{ M}^{-1}\text{cm}^{-1}$ .

### Reagents and experimental conditions

All reagents used for the experiments are of the highest purity and LC-MS grade procured from Sigma unless otherwise specified.

All experiments for the N form were performed at pH 7.0, and buffers were composed of 20 mM MOPS 150 mM KCl buffer. All experiments for the DMG form were performed at pH 3.0 in 20 mM Glycine-HCl buffer, where the protein is buffer exchanged from the storage buffer into 20 mM Glycine-HCl pH 3.0 buffer so that the protein lack any additional salt or glycerol. For all the experiments performed at pH 6.5, the buffer was composed of 20 mM MOPS with 150 mM KCl at pH 6.5. The amyloid form is prepared by heating the DMG form in pH 3.0 buffer at  $60^{\circ}\text{C}$  for 24 hours. For the hydrogen-deuterium exchange (HDX) experiments, all buffers were prepared in deuterium oxide ( $\text{D}_2\text{O}$ ), and the pH was adjusted using sodium deuterioxide (NaOD) and deuterium chloride (DCl). For HDX experiments, the pH of all buffers and samples was corrected for isotopic effect, where a factor of 0.4 was added to the pH meter reading to get the final pH of a deuterated buffer/sample. The guanidium hydrochloride (GdmCl) concentration was determined by the measurement of the refractive index.<sup>21</sup> All the buffer solutions were filtered with  $0.2\ \mu\text{m}$  filters before use.

### Peptide mapping

Peptide map was generated for TDP<sup>NBD</sup> by controlled proteolysis at low pH and temperature, as described previously.<sup>22</sup> Briefly, the protein desalted in Glycine-HCl pH 3.0 buffer was subjected to online pepsin digestion in 0.05% formic acid using an immobilized pepsin column (Applied Biosystems) at a flow rate of  $50\ \mu\text{L}/\text{min}$  on a nanoAquity UPLC system (Waters). The eluted peptides were collected on the trap column (C18 reversed-phase chromatography column, Waters), washed to remove salts, and eluted on an analytical C18 reversed-phase chromatography column (Waters), with multiple gradients of acetonitrile (3-20% in 2 min, 20-65% in 4 min, 65-95% in 2 min, 95-3% in 1 min) with 0.1% formic acid at a flow rate of  $40\ \mu\text{L}/\text{min}$  for efficient elution. The peptides were directed to the attached Synapt

G2 HD mass spectrometer (Waters) and sequenced using MS<sup>E</sup> method, followed by analysis with ProteinLynx Global Server (Waters) and manual inspection. The following parameters were used for mass detection: Positive (sensitivity) mode, source temperature 50°C; desolvation temperature 100°C; capillary voltage 3 kV; desolvation gas 600 L/hr. Ten reproducible peptide fragments were selected, encompassing 83% of the total protein (Figure 5.2, Table 5.1). The level of coverage obtained was sufficient to map the structural dynamics and organization of the different forms of TDP<sup>NBD</sup>.

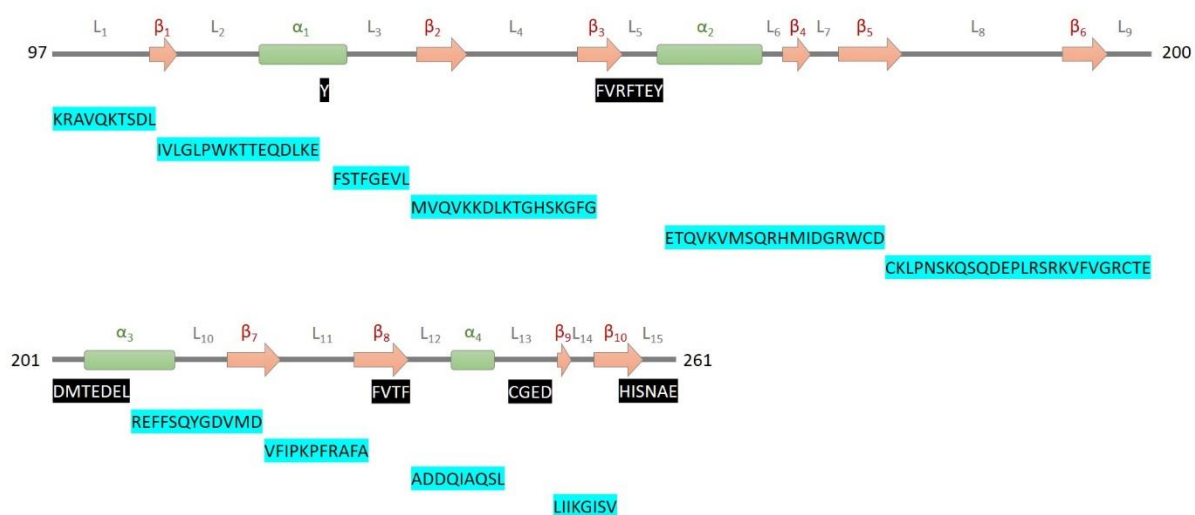


Figure 5.2. Peptide mapping of TDP<sup>NBD</sup>. Linear TDP<sup>NBD</sup> protein is shown along with all secondary structural elements (beta sheet- arrow ( $\beta_1$ - $\beta_{10}$ ), alpha helix- rectangle ( $\alpha_1$ -  $\alpha_4$ ) and loop/disordered region- line ( $L_1$ - $L_{15}$ )). 10 selected peptide fragments are shown in cyan color and the uncovered amino acid stretches are shown in black color.

Table 5.1. Peptide fragments and corresponding secondary structural content. (Beta-sheet:  $\beta$ , alpha helix:  $\alpha$  and loop/disordered region: L)

| S. No. | Peptide sequence | Secondary structural content                                 |
|--------|------------------|--|
| 1      | 97-106           | L <sub>1</sub>   |
| 2      | 107-122          | $\beta_1$ L <sub>2</sub> $\alpha_1$                          |
| 3      | 124-131          | L <sub>3</sub>   |
| 4      | 132-148          | $\beta_2$ L <sub>4</sub> $\beta_3$                           |
| 5      | 156-174          | $\alpha_2$ L <sub>6</sub> $\beta_4$ L <sub>7</sub> $\beta_5$ |
| 6      | 175-200          | L <sub>8</sub> $\beta_6$ L <sub>9</sub>                      |
| 7      | 208-219          | $\alpha_3$ L <sub>10</sub> $\beta_7$                         |
| 8      | 220-231          | L <sub>11</sub>  |
| 9      | 235-243          | L <sub>12</sub> $\alpha_4$                                   |
| 10     | 248-255          | $\beta_9$ L <sub>14</sub> $\beta_{10}$                       |

### Hydrogen-deuterium exchange measurements

For hydrogen-deuterium exchange (HDX) assay, the protein (N form, DMG form and amyloid fibril) was concentrated to 250-300  $\mu$ M by centrifugation (5000 g) through 10 kDa molecular mass cut-off concentrators. The exchange was initiated by diluting the concentrated protein sample 20 fold into the respective exchange buffer (see experimental conditions) and the exchange reaction, where the protein is now in 95% D<sub>2</sub>O, was allowed to proceed at 25°C. At different time points of exchange, an aliquot of 25  $\mu$ L was withdrawn from the exchange reaction and mixed with 175  $\mu$ L of ice-cold 7.5 M GdmCl under exchange quenched conditions (100 mM glycine-HCl, pH 2.5). After 1 min of incubation, the sample was desalted in ice-cold 100 mM glycine-HCl pH 2.5 buffer, using G-25 spin trap desalting column. The desalted sample was immediately injected into the HDX module coupled with UPLC for online pepsin digestion at a flow rate of 50  $\mu$ L/min (0.05% formic acid in water). The eluted peptides were collected on the trap column, washed to remove salt, eluted on analytical C18 RPC column, with multiple gradients of acetonitrile (3-20% in 2 min, 20-65% in 4 min, 65-95% in 2 min, 95-3% in 1min) with 0.1% formic acid at a flow rate of 40  $\mu$ L/min. The peptides were directed

to the attached Synapt G2 HD mass spectrometer (Waters). The following parameters were used for mass detection: Positive-Sensitivity mode, source temperature 50°C; desolvation temperature 100°C; capillary voltage 3 kV; desolvation gas 600 L/hr. The average masses of the peptides were determined from the centroid mass of the isotopic distribution. Control experiments were carried out for back-exchange correction, where TDP<sup>NBD</sup> was unfolded by 20 fold dilution in 6M deuterated GdmCl (95% D) and incubated for 24 h to deuterate the protein sample fully. The fully deuterated samples were processed identically as the exchange reaction sample. The extent of deuterium incorporation (%D) in each peptide was determined using the below equation,<sup>23</sup>

$$\% D = \frac{m(t) - m(0\%)}{m(95\%) - m(0\%)} \times 100 \quad (eq\ 1)$$

where,  $m(t)$  and  $m(0\%)$  represent the centroid mass of the peptide at a given time and the centroid mass of the undeuterated sample, while  $m(95\%)$  is the centroid mass of the fully deuterated peptide (in 95% D<sub>2</sub>O).

### 5.3 Results and Discussion

#### 5.3.1 Elucidating the core of the amyloid fibril

For elucidating the structural core of the amyloid fibril, we utilized HDX-MS to examine the extent of deuterium incorporation into peptide fragments (see peptide mapping in method) of TDP<sup>NBD</sup> in different forms. We performed HDX reactions for the N form, DMG form and the amyloid form as described in the method section and monitored the change in the mass spectrum of all peptide fragments of different forms of TDP<sup>NBD</sup> with the time of HDX reaction. For instance, in Figure 5.3a, we show the change in the mass spectra of the peptide fragment 124-131 with the time of HDX reaction for the N form, DMG form and the amyloid form. We observed that the mass spectra of 124-131 peptide fragment, in the N and DMG form, shifts to the mass spectrum of the 95% deuterated control peptide fragment in the observed time (Figure 5.3b), however, in the amyloid fibril form, the mass spectrum of the peptide fragment remains unchanged and similar to the 0% deuterated control. The peptide fragment 124-131 comprises the loop L<sub>3</sub> region of the protein (Figure 5.2, Figure 5.3c and Table 5.1), which is mainly disordered. Interestingly, our result shows that this L<sub>3</sub> region of the protein has become highly protected in the amyloid fibril form, building the core of the amyloid fibril, which indicates a disordered to ordered transition.

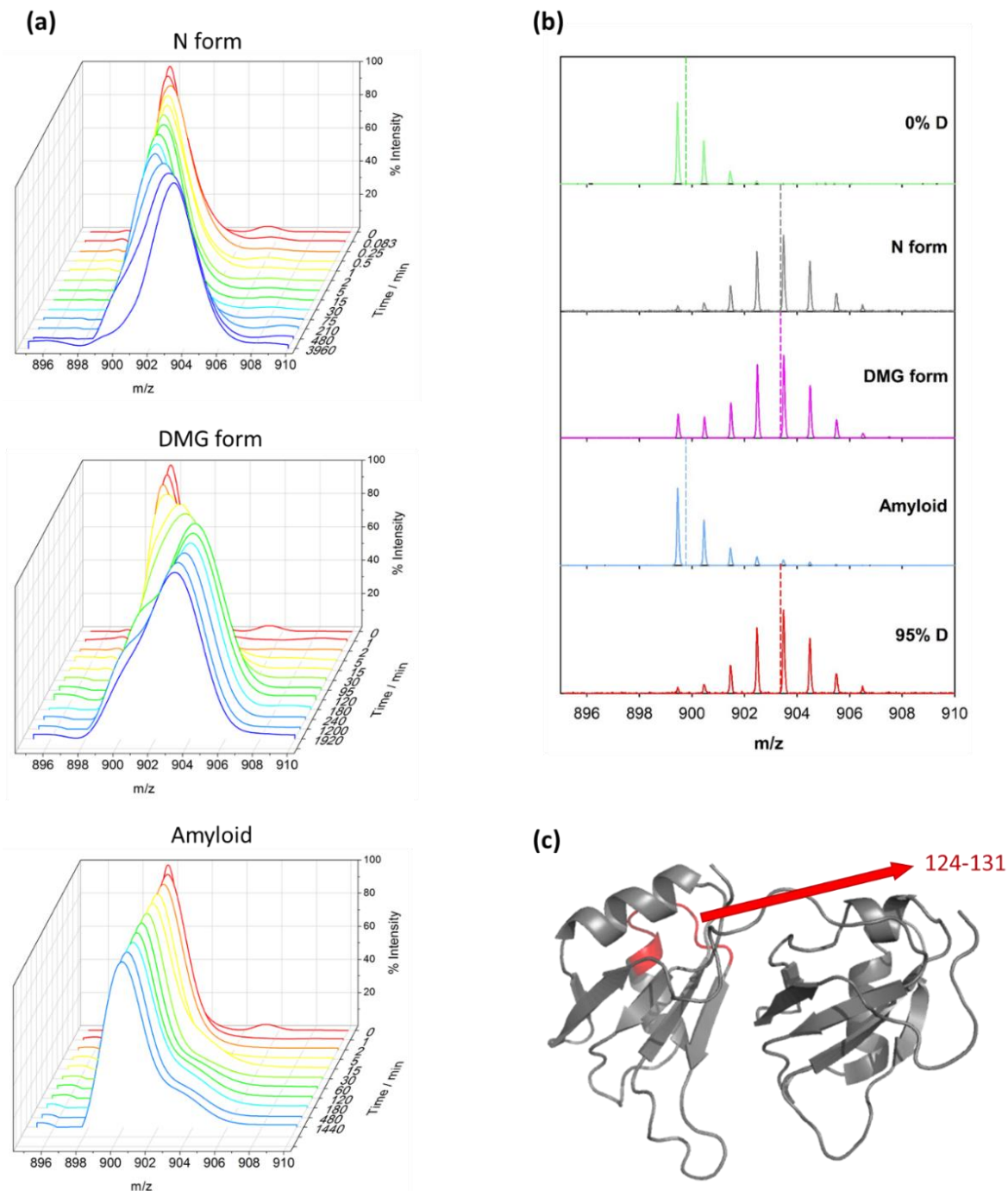


Figure 5.3. Mass spectra of a selected peptide fragment (124-131) during HDX reaction elucidate structural organization. (a) Change in mass spectra of 124-131 peptide fragment with HDX reaction time are compared for the N, DMG and the amyloid form. (b) Final mass spectra of 124-131 peptide fragment in the N, DMG and the amyloid form are compared with the undeuterated (0% D) and deuterated (95% D) controls. (c) Ribbon structure of TDP<sup>NBD</sup> showing the secondary structural content of the 124-131 peptide fragment in red color.

Similarly, we compared the mass spectra of other peptide fragments in the N form, DMG form and the amyloid fibril form (Figure 5.4). We observed that for all the peptide fragments in the N and DMG form, maximum deuterium incorporation has achieved in the observed time (24 h) of exchange reaction (Figure 5.5). However, the amyloid form show varied extent of deuterium incorporation in different peptide fragments (Figure 5.5). Our results show that out of 10 peptide fragments of TDP<sup>NBD</sup>, four peptide fragments (124-131, 156-174, 175-200 and 248-255) corresponding to different structural regions of TDP<sup>NBD</sup>: L<sub>3</sub>, α<sub>2</sub>L<sub>6</sub>β<sub>4</sub>L<sub>7</sub>β<sub>5</sub>, L<sub>8</sub>β<sub>6</sub>L<sub>9</sub> and β<sub>9</sub>L<sub>14</sub>β<sub>10</sub>, do not show any deuterium exchange. Two peptide fragments (132-148 and 235-243) with β<sub>2</sub>L<sub>4</sub>β<sub>3</sub> and L<sub>12</sub>α<sub>4</sub> secondary structure show 40% deuterium incorporation. Four peptide fragments (97-106, 107-122, 208-219 and 220-231) with L<sub>1</sub>, β<sub>1</sub>L<sub>2</sub>α<sub>1</sub>, α<sub>3</sub>L<sub>10</sub>β<sub>7</sub> and L<sub>11</sub> structural content show 70-90% deuterium incorporation within the monitored time. Using our results, we propose that the protected regions, i.e., L<sub>3</sub>, α<sub>2</sub>L<sub>6</sub>β<sub>4</sub>L<sub>7</sub>β<sub>5</sub>, L<sub>8</sub>β<sub>6</sub>L<sub>9</sub> and β<sub>9</sub>L<sub>14</sub>β<sub>10</sub>, constitute the structural core of the amyloid fibrils, where the amide protons are unavailable for hydrogen-deuterium exchange due to their burial in the rigid amyloid core. The complete protection of these fragments, encompassing significant loop/disordered regions, as seen in the ribbon structure of TDP<sup>NBD</sup> (Figure 5.6), suggests that a significant region of the protein has undergone disordered to ordered transition during the transformation of DMG form to the amyloid fibril.



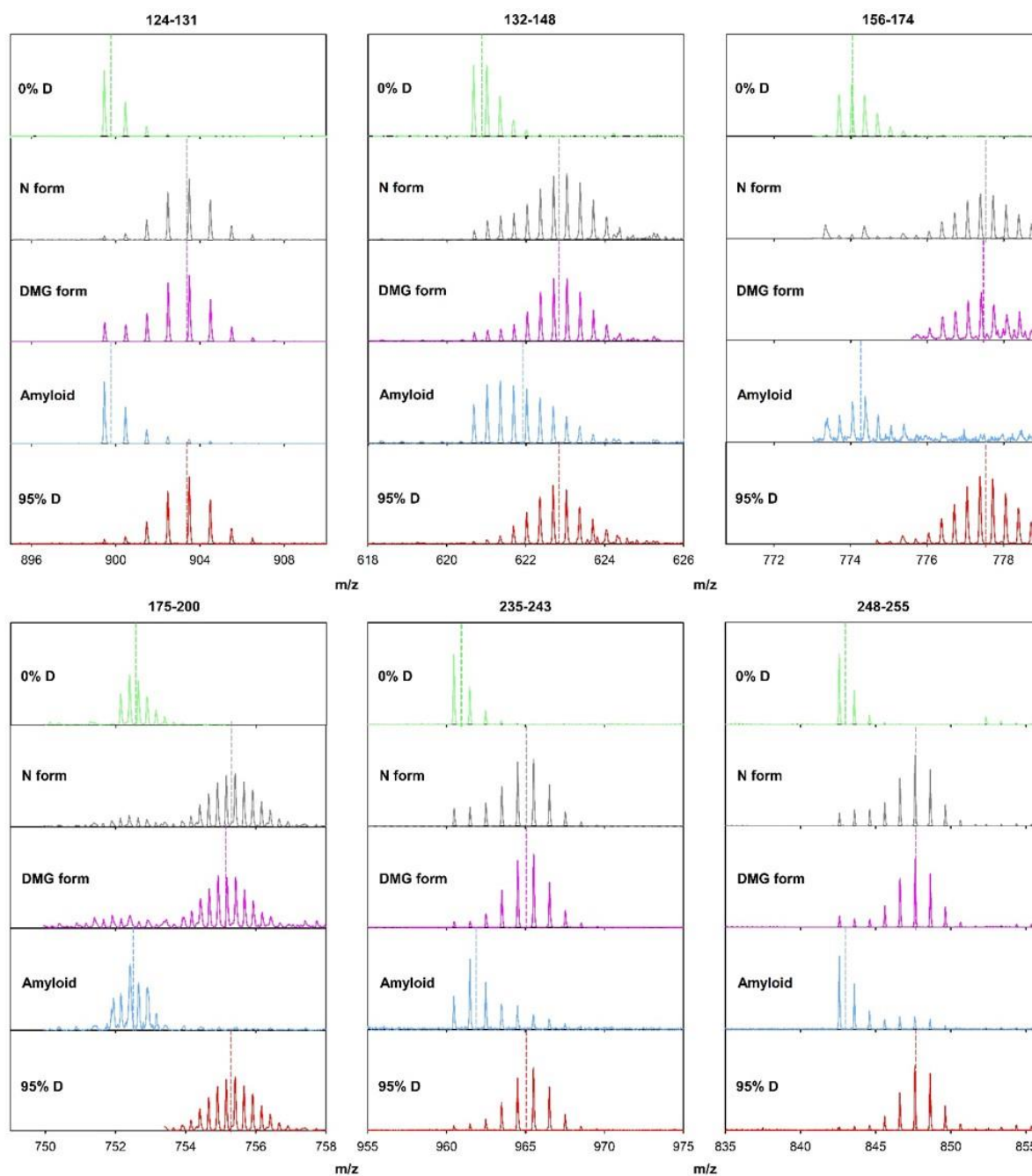


Figure 5.4. Mass spectra of selected peptide fragments in the N, DMG and the amyloid form, at the end of exchange reaction, are compared with the undeuterated (0% D) and deuterated (95% D) controls. Dashed lines represent the centroid average  $m/z$  for the given peptide.

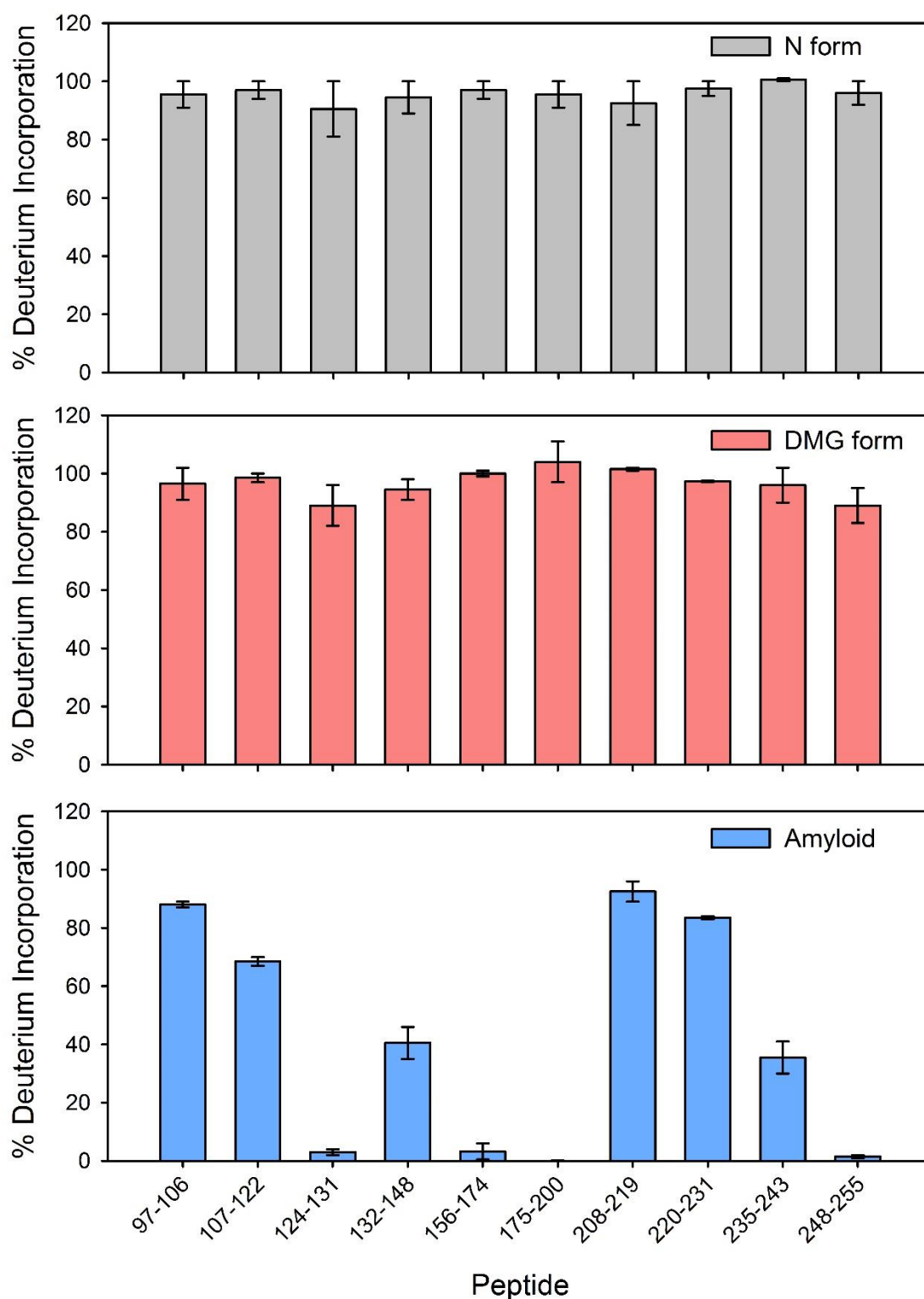
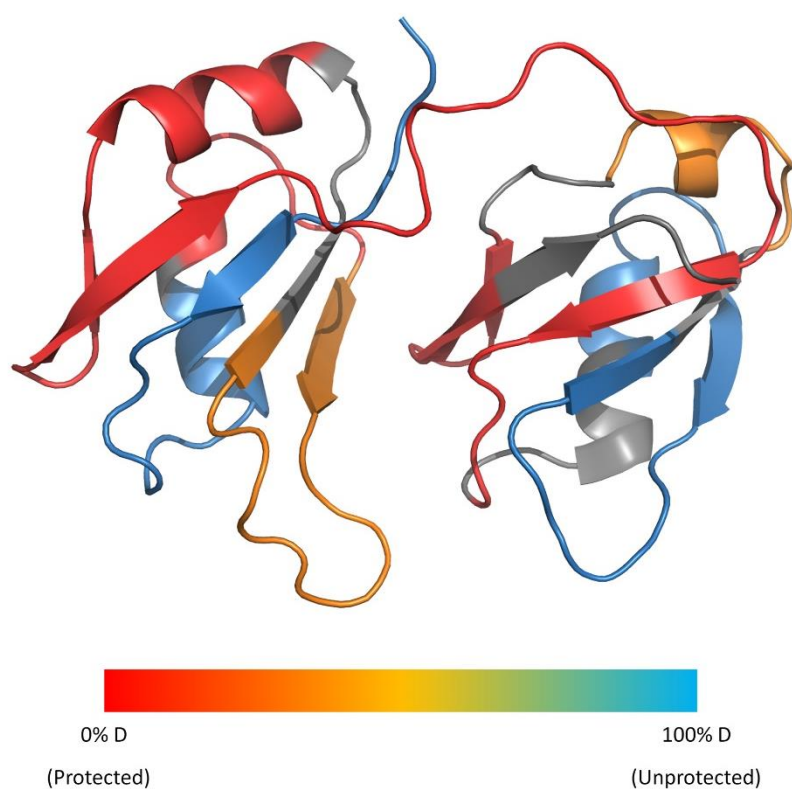


Figure 5.5. Extent of deuterium incorporation in different peptide fragments of TDP<sup>NBD</sup> in the N form (grey), DMG form (red) and the amyloid fibril form (blue) after completion of the exchange reaction. The error bars represent spread in the data from two independent experiments.

On the basis of above observations, we predict the structure of TDP<sup>NBD</sup> showing the peptide fragment level structural organization of the protein in the low-pH thermal stress induced amyloid fibrils (Figure 5.6). We show the deuterium-exchange protected regions L<sub>1</sub>, β<sub>1</sub>L<sub>2</sub>α<sub>1</sub>, α<sub>3</sub>L<sub>10</sub>β<sub>7</sub> and L<sub>11</sub> (red color), with 0% deuterium incorporation, builds the rigid core of the amyloid fibril. With 40% deuterium incorporation, β<sub>2</sub>L<sub>4</sub>β<sub>3</sub> and L<sub>12</sub>α<sub>4</sub> regions (orange color) participate in making a relatively dynamic region of the fibril and L<sub>1</sub>, β<sub>1</sub>L<sub>2</sub>α<sub>1</sub>, α<sub>3</sub>L<sub>10</sub>β<sub>7</sub> and L<sub>11</sub> regions (blue color), with 70-90% deuterium incorporation, constitute a highly dynamic region of the amyloid fibril.

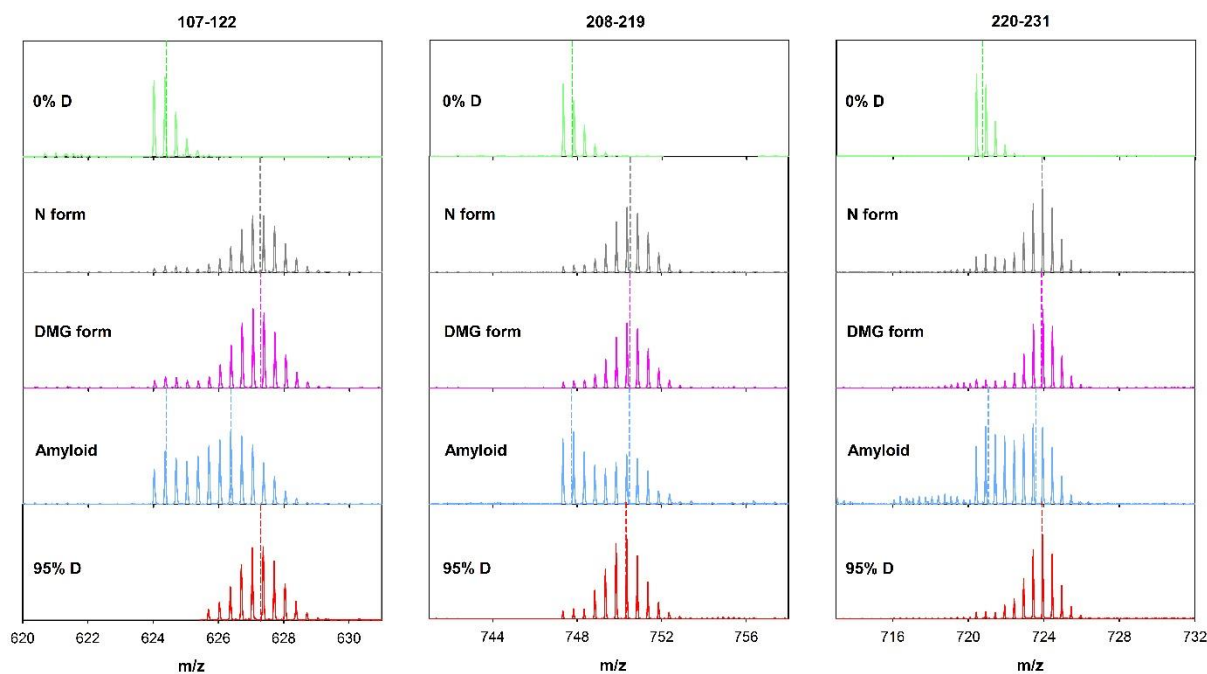


*Figure 5.6. Structure of TDP<sup>NBD</sup> representing the peptide fragment level deuterium incorporation in the amyloid form. Scale bar represents the percent deuterium incorporation. 0% D indicates peptides buried in the core of the amyloid fibril, while 100% D represents the dynamic regions of the amyloid fibril.*

### 5.3.2 Conformational heterogeneity in amyloid fibril

The amyloid fibrils are often observed to contain heterogeneity in their structure.<sup>22</sup> Using HDX-MS, we can assess the structural heterogeneity in the amyloid fibril by examining the heterogeneity in the extent of deuterium incorporation in a peptide fragment. For TDP<sup>NBD</sup>,

the majority of the peptide fragments display unimodal mass spectra at the end of the exchange reaction (Figure 5.4), suggesting that these peptide fragments exist in single conformation in the amyloid fibril form. However, three peptide fragments (107-122, 208-219 and 220-231), constituting  $\beta_1L_2\alpha_1$ ,  $\alpha_3L_{10}\beta_7$  and  $L_{11}$  regions, display bimodal mass spectra in the amyloid fibril form (Figure 5.7), indicating for the existence of heterogeneous conformations of these peptide fragments in the amyloid fibril.



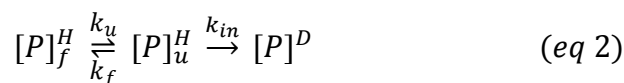
*Figure 5.7. Conformational heterogeneity in amyloid fibril form. Mass spectra of selected peptides of  $TDP^{NBD}$  are compared for the N, DMG and the amyloid fibril after the completion of the exchange reaction, along with the undeuterated (0% D) and deuterated (95% D) controls. Dashed lines represent the centroid average  $m/z$  for the given peptide.*

The heterogeneity in the conformation of these peptide fragments is a result of the presence of at least two different conformations differing greatly in the protection level of their amide protons against HDX. Our observations indicate that in the amyloid fibrils, for the  $\beta_1L_2\alpha_1$ ,  $\alpha_3L_{10}\beta_7$  and  $L_{11}$  regions, one of the population is weakly protected and gets fully deuterated within 20 min of HDX; however, the other conformation remains undeuterated at the end of HDX reaction (24 h). Thus, we propose that in some amyloid fibrils, the fibril core also comprises  $\beta_1L_2\alpha_1$ ,  $\alpha_3L_{10}\beta_7$  and  $L_{11}$  regions of the protein.

### 5.3.3 Amyloidogenicity of the amyloid precursor DMG form is a result of high structural dynamics

We further assessed why only DMG form of TDP<sup>NBD</sup> is amyloidogenic, serving as an amyloid precursor, but not the N form? Previously we have shown the molecular differences between the DMG form and the N form, where we established that the low-pH induces an expanded conformation of TDP<sup>NBD</sup> with a dry core (Chapter 4). However, fragment level information on structural dynamics is desirable for a profound understanding of the amyloidogenicity (nucleation step) of the DMG form. We utilized the power of HDX-MS to compare the fragment level structural dynamics in the N form and DMG form.

In HDX technique, the amide protons of a native protein exchange with deuterium as shown below,<sup>24</sup>



where,  $[P]_f^H$  is the folded state of protonated protein which can unfold to a  $[P]_u^H$  state with a rate constant of unfolding ( $k_u$ ) and can fold again without exchanging the proton with a rate constant of folding ( $k_f$ ). The exchange labile  $[P]_u^H$  state can exchange the proton with deuterium to form  $[P]^D$  state with a rate constant of intrinsic exchange ( $k_{in}$ ). For a given peptide fragment/ proton, the observed rate constant ( $k_{obs}$ ) for hydrogen-deuterium exchange, with an assumption that  $k_u \ll k_f$  (reasonable for a folded protein), can be given by,

$$k_{obs} = \frac{k_u k_{in}}{k_f + k_{in}} \quad (eq\ 3)$$

HDX provides several significant pieces of information, such as the rate of unfolding and the equilibrium constant of the unfolding of a peptide fragment, depending on the exchange conditions, namely, EX1 and EX2 regime. In the EX1 regime, for a peptide fragment at a given pH and temperature, the rate of intrinsic exchange is significantly higher than the rate of folding ( $k_{in} \gg k_f$ ), resulting all unfolding events into successful exchange events. Thus,  $k_{obs}$  in EX1 exchange regime only depends on  $k_u$  and expressed as,

$$k_{obs} = k_u \quad (eq\ 4)$$

where, the exchange kinetics is independent of  $k_{in}$ .

The  $k_{obs}$  in EX2 exchange regime, with an assumption that the rate of intrinsic exchange is much lower than the rate of folding ( $k_f \gg k_{in}$ ), is directly proportional to the equilibrium constant of unfolding ( $K_{eq}$ , where  $K_{eq} = \frac{k_u}{k_f}$ ) and  $k_{in}$ ,

$$k_{obs} = K_{eq}k_{in} \quad (eq\ 5)$$

Furthermore, the  $K_{eq}$  can be used to extract the protection factor ( $P_f = 1/K_{eq}$ ) for individual peptide fragments.  $P_f$  is a measure of the structural dynamics of a given peptide fragment, where, higher  $P_f$  dictates restricted dynamics of the peptide fragment, as in the tightly packed core of a protein and lower  $P_f$  signifies unimpeded dynamics of the peptide fragment, as in the solvent-exposed unfolded polypeptide chain. Hence, to elucidate the fragment level structural dynamics in the DMG form and the N form using HDX-MS, the exchange of amide protons must occur in the EX2 exchange regime.

Since,  $k_{in}$  changes with change in pH, therefore, any change in pH changes the  $k_{obs}$  in the EX2 regime but not in the EX1 regime, given that the stability of the protein is unchanged. With the knowledge that TDP<sup>NBD</sup> remains in the native ensemble between pH 6.5 to pH 7.5,<sup>20</sup> a test was performed to determine whether a peptide segment exhibits EX1 or EX2 hydrogen-deuterium exchange kinetics.<sup>25</sup> The theoretical values of  $k_{in}$  for linear peptide fragments can be determined from literature<sup>26</sup> and we calculated the values of  $k_{in}$  at pH 6.5 and pH 7.0 (Table 5.2 and 5.3) for solvent-exposed linear peptide fragments of TDP<sup>NBD</sup> using spreadsheets downloadable from Walter Englander's website (<http://hx2.med.upenn.edu>).<sup>27</sup> The calculated  $k_{in}$  values for all peptide fragments were plotted on a double log plot of  $k_{in}$  at pH 7.0 vs  $k_{in}$  at pH 6.5 and fitted to a linear least-square fit to obtain the slope of the plot pertaining to EX2 exchange kinetics (Figure 5.8). A hypothetical plot, corresponding to EX1 exchange kinetics, is added to the above double log plot (Figure 5.8), assuming that the exchange rates at pH 6.5 and pH 7.0 are unchanged and fitted to a linear least-square fit to obtain the slope of the plot pertaining to EX1 exchange kinetics. The intercepts obtained from the fits were 1.25 and 0 for EX2 and EX1 exchange kinetics, respectively, with a common slope of 1. These predictions would also hold true for the native protein, as the stability of the protein is unchanged in the pH range 6.5-7.0 (Patni and Jha unpublished data), and no conformational differences exist in the native state ensemble at these two pH values. In a double log plot of  $k_{obs}$  at pH 7.0 vs  $k_{obs}$  at pH 6.5, the log  $k_{obs}$  values of the amide protons exchanging by EX1 would fall on EX1 slope (intercept= 0 and slope= 1) and the log  $k_{obs}$  values of the amide sites exchanging purely

by EX2 are expected to fall on EX2 slope (intercept= 1.25 and slope= 1). According to these predictions, the experimental hydrogen-exchange data can be clustered into two groups: peptides with EX2 exchange kinetics and peptides with EX1 exchange kinetics (Figure 5.8).

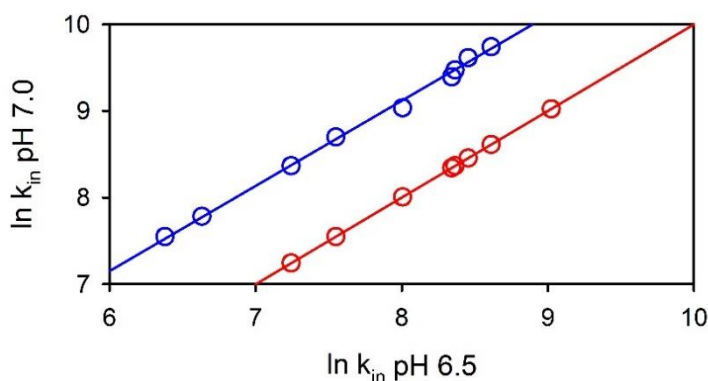


Figure 5.8. EX1/EX2 hydrogen-deuterium exchange mechanism theoretical test for  $TDP^{NBD}$  at pH 7.0 and 6.5. Blue circles denote exchange rate constants ( $k_{in}$ ) of different peptide fragments exchanging the amide protons in EX2 regime.  $k_{in}$  values were determined theoretically as given in Table 5.2 and 5.3. Red circles denote peptide fragments exchanging the amide protons in EX1 regime with the same rate constant (hypothetical assumption) at pH 7.0 and 6.5. The  $k_{in}$  values at pH 7.0 were kept identical as the  $k_{in}$  values at pH 6.5, assuming that in the EX1 regime, the observed rate constant does not change with pH. Solid blue and red lines are linear least square fit representing the slopes for EX2 and EX1 exchange regimes, respectively.

We acquired the hydrogen exchange kinetics of  $TDP^{NBD}$  at pH 6.5 and pH 7.0 (Figure 5.9a and 5.9b) and determined the respective  $k_{obs}$  values (Table 5.2 and 5.3) by fitting all the kinetic traces to bi or triexponential function. The experimentally determined  $k_{obs}$  values were plotted in a double log plot ( $k_{obs}$  at pH 7.0 vs  $k_{obs}$  at pH 6.5) and compared with the EX1 and EX2 exchange kinetics slopes (Figure 5.10). Our results suggest that all peptide fragments, at pH 7.0, exchange the amide protons predominantly via EX2 exchange kinetics. Consequently, for hydrogen exchange at pH 3.0, it is fair to assume that the exchange of amide protons would follow the EX2 exchange regime, considering that the  $k_{in}$  values at pH 3.0 (Table 5.4) are significantly lower than the  $k_{in}$  values at pH 7.0 (Table 5.3).

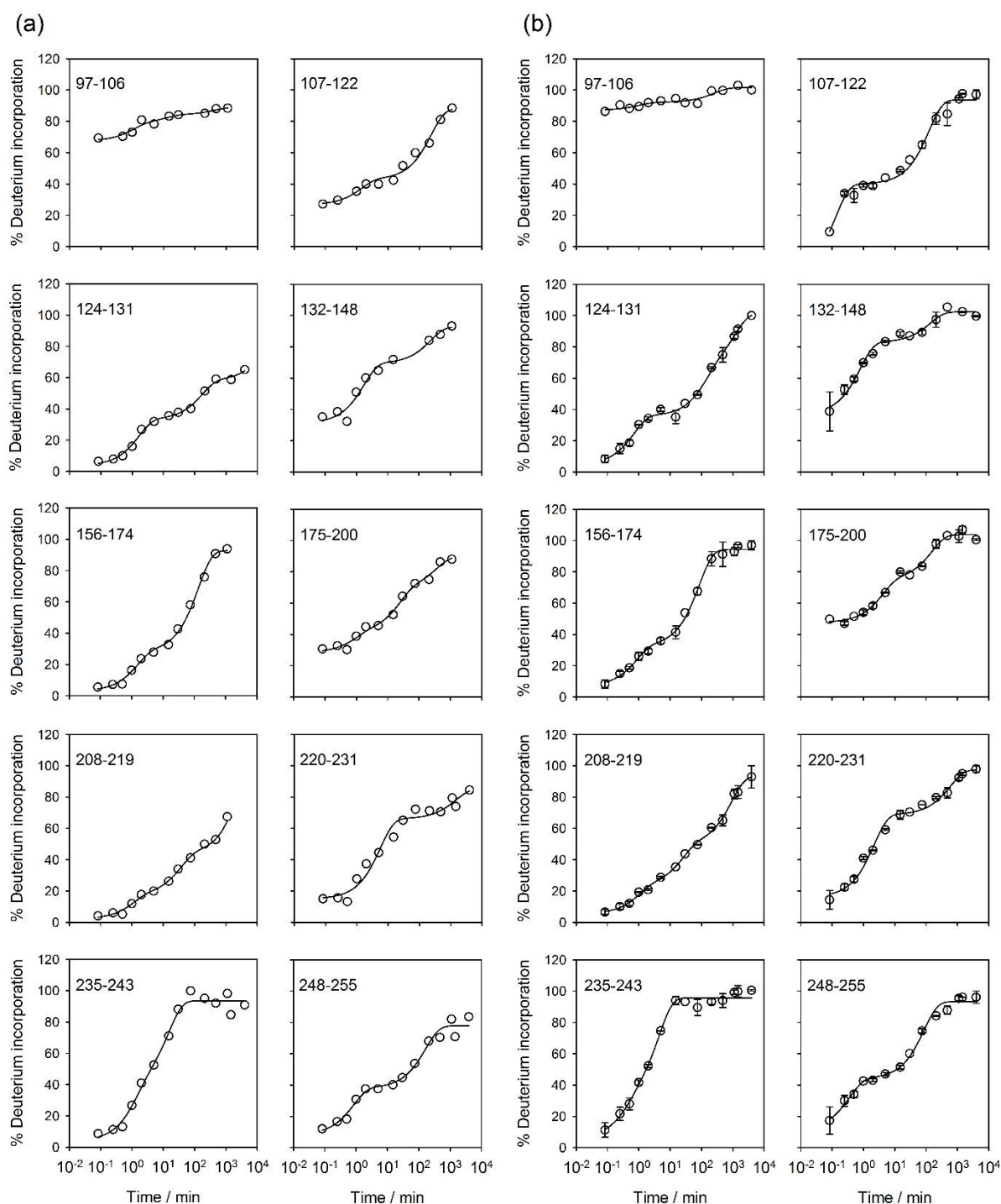


Figure 5.9. Hydrogen-deuterium exchange kinetics of  $TDP^{NBD}$  peptide fragments at pH 6.5 (a) and pH 7.0 (b). Experimentally determined percent deuterium incorporation (from eq 1) profiles are shown for ten peptide fragments. Error bar represents spread in the data from two independent experiments and the solid lines through the data represent fits to either bi or triexponential function.



Table 5.2. Hydrogen-exchange kinetics parameters for the N form at pH 6.5. The values of the intrinsic exchange rate constant ( $k_{in}$ ), observed rate constants ( $k_{obs}$ )  $k_1$ ,  $k_2$  and  $k_3$ , and their respective percent amplitudes (% Amp) along with missing amplitude are shown.

| Seq     | $k_{in} / \text{min}^{-1}$ | Burst | $k_{obs} / \text{min}^{-1}$                |
|---------|----------------------------|-------|--|
| 97-106  | 100                        | 80%   | 0.90 (10%)<br>0.09 (5%)<br>0.002 (5%)      |
| 107-122 | 73                         | 40%   | 0.70 (15%)<br>0.004 (45%)                  |
| 124-131 | 1200                       | 5%    | 0.60 (30%)<br>0.005 (25%)<br>0.00004 (40%) |
| 132-148 | 190                        | 40%   | 0.55 (30%)<br>0.004 (30%)                  |
| 156-174 | 230                        | 10%   | 0.60 (25%)<br>0.007 (65%)                  |
| 175-200 | 560                        | 40%   | 1.05 (10%)<br>0.042 (30%)<br>0.002 (20%)   |
| 208-219 | 420                        | -     | 0.88 (15%)<br>0.028 (25%)<br>0.0002 (60%)  |
| 220-231 | 200                        | 30%   | 0.18 (50%)<br>0.0007 (20%)                 |
| 235-243 | 540                        | 10%   | 0.81 (35%)<br>0.067 (55%)                  |
| 248-255 | 84                         | 30%   | 1.20 (30%)<br>0.006 (40%)                  |

Table 5.3. Hydrogen-exchange kinetics parameters for the N form at pH 7.0. The values of the intrinsic exchange rate constant ( $k_{in}$ ), observed rate constants ( $k_{obs}$ ), their respective percent amplitudes (% Amp) and protection factors ( $P_f$ ) along with burst phase amplitude are shown.

| Seq     | $k_{in} / \text{min}^{-1}$ | Burst | $k_{obs} / \text{min}^{-1}$              | $P_f$   |
|---------|----------------------------|-------|--|---|
| 97-106  | 330                        | 85%   | 0.94 (5%)<br>0.004 (10%)                 | $3.5 \times 10^2$<br>$8.2 \times 10^4$                      |
| 107-122 | 230                        | 10%   | 5.50 (30%)<br>0.008 (60%)                | $4.2 \times 10^1$<br>$2.9 \times 10^4$                      |
| 124-131 | 3800                       | 10%   | 1.50 (30%)<br>0.01 (30%)<br>0.0008 (30%) | $2.5 \times 10^3$<br>$3.8 \times 10^5$<br>$4.7 \times 10^6$ |
| 132-148 | 530                        | 35%   | 1.40 (45%)<br>0.008 (20%)                | $3.8 \times 10^2$<br>$6.6 \times 10^4$                      |
| 156-174 | 670                        | 10%   | 1.30 (25%)<br>0.012 (65%)                | $5.1 \times 10^2$<br>$5.5 \times 10^4$                      |
| 175-200 | 1800                       | 45%   | 0.25 (25%)<br>0.006 (30%)                | $7.2 \times 10^3$<br>$3.0 \times 10^5$                      |
| 208-219 | 1300                       |       | 1.20 (20%)<br>0.06 (25%)<br>0.0011 (55%) | $1.1 \times 10^3$<br>$2.1 \times 10^4$<br>$1.9 \times 10^6$ |
| 220-231 | 620                        | 20%   | 0.50 (50%)<br>0.002 (30%)                | $1.2 \times 10^3$<br>$3.1 \times 10^5$                      |
| 235-243 | 1700                       | 10%   | 2.73 (25%)<br>0.23 (65%)                 | $6.2 \times 10^2$<br>$7.0 \times 10^3$                      |
| 248-255 | 270                        | 20%   | 2.71 (30%)<br>0.011 (50%)                | $1.0 \times 10^2$<br>$2.5 \times 10^4$                      |

$P_f$ : Rounded off to 1 digit after the decimal.

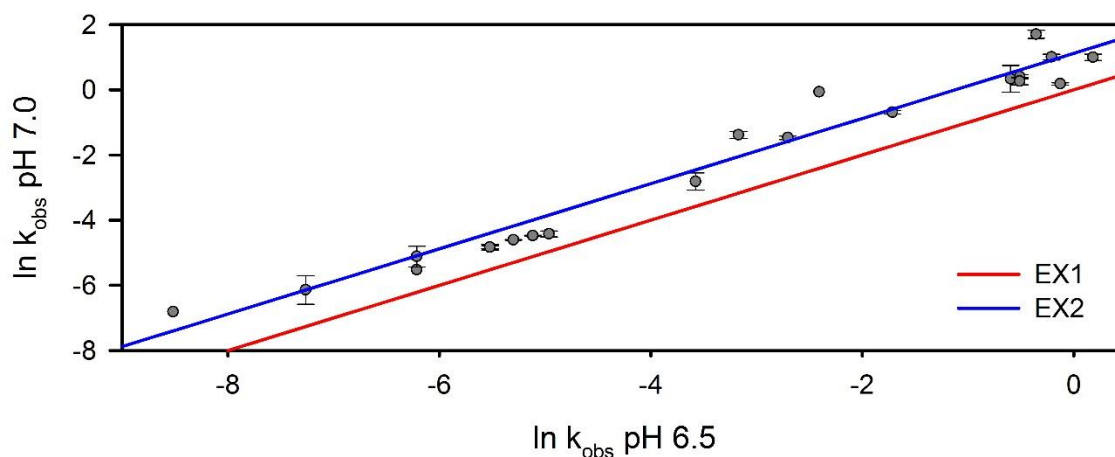


Figure 5.10. Determination of the hydrogen-deuterium exchange regime. Double log plot of the observed rate constants of hydrogen-deuterium exchange at pH 7.0 vs pH 6.5. The solid blue and red lines represent the slope of the EX2 and the EX1 exchange regimes, respectively.

After validating that all amide protons exchange with deuterium predominantly in EX2 exchange regime for both N form and the DMG form, we determined the values of  $k_{obs}$  by fitting all hydrogen-exchange kinetic traces for the N form at pH 7.0 to either bi or triexponential function (Figure 5.9b) and for the DMG (Figure 5.11) and amyloid form (Figure 5.12) at pH 3.0 to either mono, bi or triexponential function. The hydrogen-exchange kinetics with bi or tri exponential function represents the presence of multiple secondary structure components in the peptide fragment, each exchanging at different rates. The HDX kinetics of some peptide fragments includes a missing burst kinetic phase resulting from the ultra-fast exchange of amide protons within the disordered regions of the protein (loops or random coils), for which the  $k_{obs}$  cannot be determined. All  $k_{obs}$  values for the DMG form at pH 3.0 are given in Table 5.4 and for amyloid form in Table 5.5.

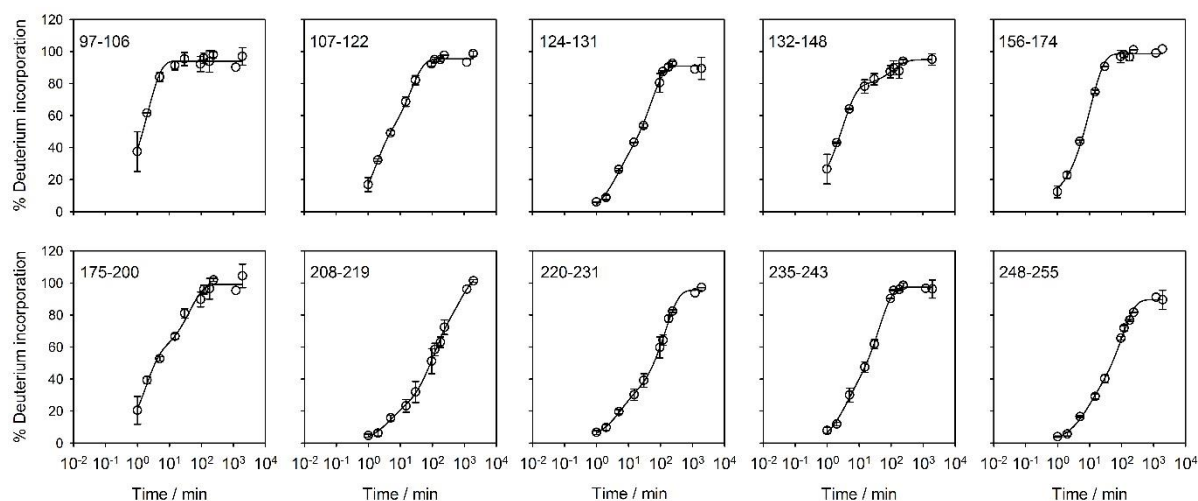


Figure 5.11. Hydrogen-deuterium exchange kinetics of  $TDP^{NBD}$  peptide fragments at pH 3.0. Experimentally determined percent deuterium incorporation (from eq 1) profiles are shown for ten peptide fragments. Error bar represents spread in the data from two independent experiments and the solid lines through the data represent fits to mono/bi/triexponential function.

It is important to note that for the N and the DMG form, the hydrogen-exchange kinetics was monitored at two different pH, where the  $k_{in}$  changes with change in pH, hence, the direct comparison of the exchange kinetics and the  $k_{obs}$  values would be inadequate. However, the  $K_{eq}$  is a measure of  $k_{obs}$  value normalized by the  $k_{in}$  value ( $K_{eq} = k_{obs}/k_{in}$ ), which can further be translated to  $P_f$  ( $P_f = 1/K_{eq}$ ), therefore, the  $P_f$  values are comparable across different pH. Previously, for myoglobin, it has been shown that the hydrogen exchange  $P_f$  value of  $10^0$  or 1 represents that the amide proton is fully exposed to the solvent, contrarily, a  $P_f$  value near  $10^7$  indicates buried amide protons.<sup>9,28</sup> For  $TDP^{NBD}$ , we determined the hydrogen exchange  $P_f$  values from the  $k_{obs}$  values and the  $k_{in}$  values for the N form (Table 5.3), DMG form (Table 5.4) and the amyloid form (Table 5.5).

Table 5.4. Hydrogen-exchange kinetics parameters for the DMG form at pH 3.0. The values of the intrinsic exchange rate constant ( $k_{in}$ ), observed rate constants ( $k_{obs}$ ), their respective percent amplitudes (% Amp) and protection factors ( $P_f$ ) along with burst phase and very slow phase amplitude are shown.

| Seq     | $k_{in} / \text{min}^{-1}$ | Burst | $k_{obs} / \text{min}^{-1}$ (%amp)        | $P_f$            | Very slow<br>(unrecorded<br>%amp) |
|---------|----------------------------|-------|---|------------------|-----------------------------------|
| 97-106  | 0.17                       | 35%   | 0.48 (65%)                                | 0.35             |                                   |
| 107-122 | 0.15                       | 20%   | 0.70 (30%)<br>0.048 (50%)                 | 0.2<br>3.1       |                                   |
| 124-131 | 1.3                        |       | 0.28 (35%)<br>0.020 (65%)                 | 4.6<br>65        |                                   |
| 132-148 | 0.27                       | 25%   | 0.33 (55%)<br>0.008 (20%)                 | 0.8<br>34        |                                   |
| 156-174 | 0.56                       | 10%   | 0.09 (90%)                                | 6.2              |                                   |
| 175-200 | 0.67                       | 20%   | 0.58 (40%)<br>0.023 (40%)                 | 1.5<br>29        |                                   |
| 208-219 | 0.98                       |       | 0.26 (20%)<br>0.011 (50%)<br>0.0014 (30%) | 3.8<br>89<br>700 |                                   |
| 220-231 | 0.23                       |       | 0.21 (40%)<br>0.007 (60%)                 | 1.1<br>33        |                                   |
| 235-243 | 2.2                        |       | 0.33 (30%)<br>0.024 (70%)                 | 6.6<br>92        |                                   |
| 248-255 | 0.1                        |       | 0.13 (30%)<br>0.010 (60%)                 | 0.8<br>10        | 10%                               |

$P_f$ : Rounded off to 1 digit after the decimal.

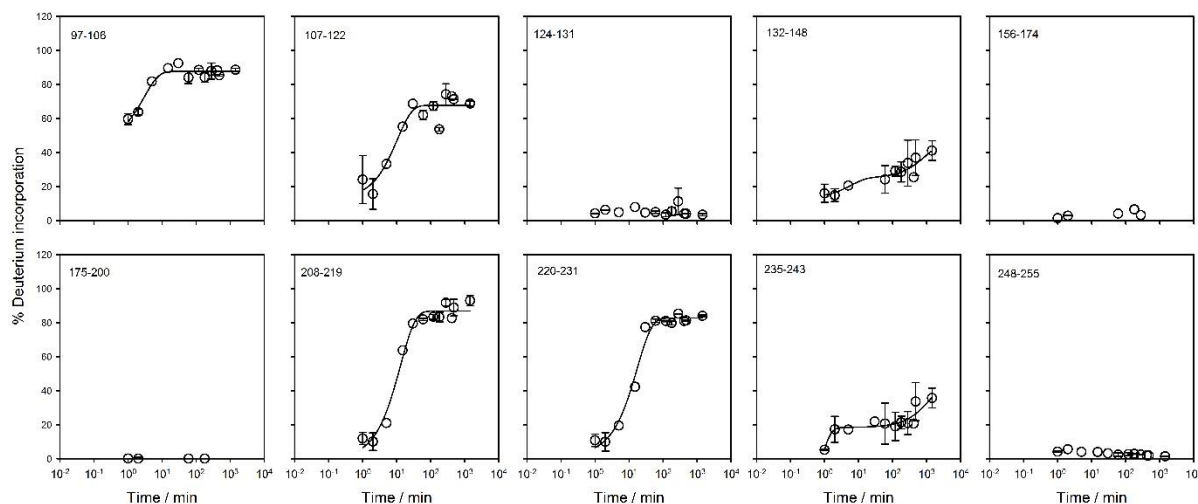


Figure 5.12. Hydrogen-deuterium exchange kinetics of  $TDP^{NBD}$  peptide fragments for the amyloid form at pH 3.0. Experimentally determined percent deuterium incorporation (from eq 1) profiles are shown for ten peptide fragments. Error bar represents spread in the data from two independent experiments and the solid lines through the data represent fits to mono/biexponential function.

Table 5.5. Hydrogen-exchange kinetics parameters for the Amyloid form at pH 3.0. The values of the intrinsic exchange rate constant ( $k_{in}$ ) and observed rate constants ( $k_{obs}$ ) with respective percent amplitudes (% amp) and protection factors ( $P_f$ ) along with % amplitude of the burst and very slow phase are shown.

| Seq     | $k_{in} / \text{min}^{-1}$ | Burst | $k_{obs} / \text{min}^{-1}$ (%amp) | $P_f$                    | Very slow |
|---------|----------------------------|-------|------------------------------------|--------------------------|-----------|
| 97-106  | 0.17                       | 60%   | 0.34 (30%)                         | 0.5                      | 10%       |
| 107-122 | 0.15                       | 15%   | 0.09 (55%)                         | 1.7                      | 30%       |
| 124-131 | 1.3                        |       | ND                                 | $>10^5 \#$               | 100%      |
| 132-148 | 0.27                       | 15%   | 0.16 (10%)<br>0.001 (20%)          | 1.7<br>$2.7 \times 10^2$ | 55%       |
| 156-174 | 0.56                       |       | ND                                 | $>10^5 \#$               | 100%      |
| 175-200 | 0.67                       |       | ND                                 | $>10^5 \#$               | 100%      |
| 208-219 | 0.98                       |       | 0.08 (90%)                         | 12                       | 10%       |
| 220-231 | 0.23                       |       | 0.06 (80%)                         | 4                        | 20%       |
| 235-243 | 2.2                        |       | 2.3 (20%)<br>0.0008 (20%)          | 0.9<br>$2.7 \times 10^3$ | 60%       |
| 248-255 | 0.1                        |       | ND                                 | $>10^4 \#$               | 100%      |

*ND: Not determined*

*P<sub>f</sub>: Rounded off to 1 digit after the decimal.*

*# (Estimated P<sub>f</sub>): Back calculations for the estimation of P<sub>f</sub> values were performed using the respective k<sub>in</sub> values with an assumption of 1% deuterium incorporation in these peptide fragments in 24 h.*

We compared the  $P_f$  across different forms and categorised different regions of the protein into four groups: highly-protected ( $P_f > 10^5$ ), moderately-protected ( $P_f = 10^3$ - $10^4$ ), weakly-protected ( $P_f = 10^1$ - $10^2$ ) and unprotected ( $P_f < 10^1$ ) (Table 5.6 and Figure 5.13). The 97-106 peptide fragment corresponding to the L<sub>1</sub> region of the protein remain unprotected across all forms. The 107-122, 208-219 and 220-231 peptide fragments, corresponding to  $\beta_1L_2\alpha_1$ ,  $\alpha_3L_{10}\beta_7$  and L<sub>11</sub> regions of the protein, show increased dynamics after N to DMG transformation with  $P_f$  values changing from highly/moderately-protected to weakly-protected or unprotected. Interestingly, we observed that the 132-148 peptide fragment, corresponding to  $\beta_2L_4\beta_3$  region, was weakly-protected in the N and unprotected in the DMG form (with >80% peptide is exchanging with a  $P_f$  of  $10^2$ ), however, the protection level of this region increases in the amyloid form (with >55% peptide remaining unexchanged, Figure 5.5). The remaining five peptide fragments, 124-131, 156-174, 175-200, 235-243 and 248-255 (corresponding to L<sub>3</sub>,  $\alpha_2L_6\beta_4L_7\beta_5$ , L<sub>8</sub> $\beta_6$ L<sub>9</sub>, L<sub>12</sub> $\alpha_4$  and  $\beta_9L_{14}\beta_{10}$  regions of protein) exhibit dramatic changes in their protection level during N to DMG to amyloid fibril transformation. These regions were highly/moderately-protected in the N form, however, the N to DMG transformation results in drastic decrease in the protection level leaving these regions weakly-protected or unprotected. Further, in the matured fibrils these regions redeem rigidity and become highly protected, with L<sub>3</sub>,  $\alpha_2L_6\beta_4L_7\beta_5$ , L<sub>8</sub> $\beta_6$ L<sub>9</sub> and  $\beta_9L_{14}\beta_{10}$  regions showing no exchange and L<sub>12</sub> $\alpha_4$  region showing only 40% exchange.

We propose that the increased structural dynamics in the DMG form, specifically in the L<sub>3</sub>,  $\alpha_2L_6\beta_4L_7\beta_5$ , L<sub>8</sub> $\beta_6$ L<sub>9</sub> and  $\beta_9L_{14}\beta_{10}$  regions of the protein, is responsible for amyloidogenic nature of the DMG form, where the increased dynamics provide more chances for these fragments from different monomeric DMG units to interact and constitute the nucleus for amyloid fibril. Further, as the amyloid fibril grow, these regions get buried in the core of the fibril resulting in highly restricted dynamics. While, due to the restricted dynamics of L<sub>3</sub>,  $\alpha_2L_6\beta_4L_7\beta_5$ , L<sub>8</sub> $\beta_6$ L<sub>9</sub> and  $\beta_9L_{14}\beta_{10}$  regions in the N form, the interaction between these fragments from different monomeric units is highly unlikely, rendering the N form non-amyloidogenic.

Table 5.6. Peptide specific protection level for the N form, DMG form and the amyloid form. The protection factor values ( $P_f$  corresponding >50% of overall exchange kinetics from Table 5.3, Table 5.4 and Table 5.5) are categorized into 4 groups: High, Moderate, Weak and unprotected.  $P_f$  values  $> 10^5$  are grouped into High,  $P_f$  values between  $10^3$ - $10^4$  are grouped into Moderate,  $P_f$  values between  $10^1$ - $10^2$  are grouped into Weak protection level and  $P_f$  values  $< 10^1$  are grouped into unprotected peptides.

| Peptide sequence | Secondary structure  | N form Protection level | DMG Protection level | Amyloid Protection level |
|------------------|--|-------------------------|----------------------|--------------------------|
| 97-106           | L <sub>1</sub>   | Unprotected             | Unprotected          | Unprotected              |
| 107-122          | $\beta_1$ L <sub>2</sub> $\alpha_1$                          | Moderate                | Unprotected          | Unprotected              |
| 124-131          | L <sub>3</sub>   | High                    | Weak                 | High                     |
| 132-148          | $\beta_2$ L <sub>4</sub> $\beta_3$                           | Weak                    | Unprotected          | Moderate                 |
| 156-174          | $\alpha_2$ L <sub>6</sub> $\beta_4$ L <sub>7</sub> $\beta_5$ | Moderate                | Unprotected          | High                     |
| 175-200          | L <sub>8</sub> $\beta_6$ L <sub>9</sub>                      | Moderate                | Weak                 | High                     |
| 208-219          | $\alpha_3$ L <sub>10</sub> $\beta_7$                         | Weak                    | Weak                 | Weak                     |
| 220-231          | L <sub>11</sub>  | Moderate                | Weak                 | Unprotected              |
| 235-243          | L <sub>12</sub> $\alpha_4$                                   | Moderate                | Weak                 | Moderate                 |
| 248-255          | $\beta_9$ L <sub>14</sub> $\beta_{10}$                       | Moderate                | Weak                 | High                     |



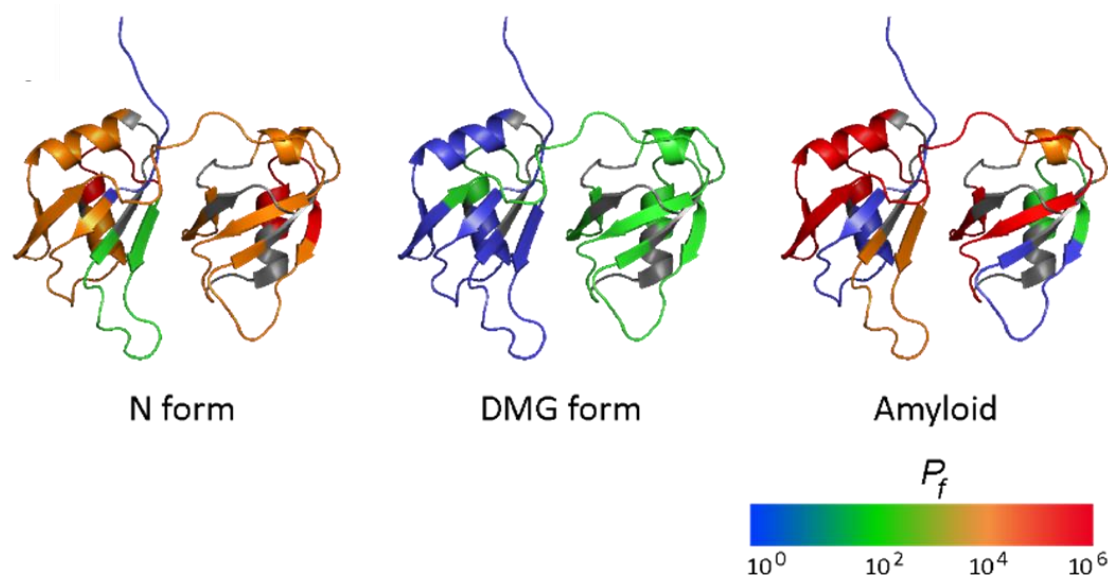


Figure 5.13. Heat-map of HDX kinetics derived fragment level protection factor ( $P_f$ ) compared in the N, DMG and the amyloid form. A  $P_f$  value of  $10^0$  represents unprotected amide protons suggesting unfolded peptide like conformation, while a  $P_f$  value of  $10^6$  denotes highly protected amide protons suggesting burial of peptide fragment as in the core of a native protein.

#### 5.4 Conclusions

In this study, we assessed the structural organization of the amyloid fibril core and performed high-resolution characterization of the amyloidogenic DMG form of TDP<sup>NBD</sup> using HDX-MS. We monitored the extent of deuterium incorporation in the N, DMG and the amyloid form of TDP<sup>NBD</sup> to gain the structural insights of the amyloid fibrils. High deuterium incorporation in a peptide fragment represents very weak protection of the amide protons against the exchange, ideal for a highly dynamic region of the protein. Whereas, low or no deuterium incorporation represents strong protection of the amide protons against the exchange, ideal for buried core region of a protein with highly restricted dynamics. We show that out of ten peptide fragments of TDP<sup>NBD</sup>, four peptide fragments participate in building the core of the amyloid fibril, two fragments have moderate protection, and the remaining four fragments are in the unprotected region. The exchange protected peptide fragments that build the core of the amyloid fibrils contain a significant amount of disordered regions. We propose that these fragments, upon transformation to amyloid fibril structure, undergo structural transformation with disordered to ordered transition. We also show that for three peptide fragments, the observed mass spectra display bimodal distribution, suggesting that these

fragments are protected in some molecules and unprotected in other molecules conclusive for the heterogeneity in the population of amyloid fibrils. We further characterized the structural dynamics of the amyloidogenic DMG form, by monitoring the hydrogen-deuterium exchange kinetics of the N form and the DMG form in the EX2 exchange regime and compared their respective protection factors to evaluate the high-resolution differences between the two forms. We observed that the N form possess restricted dynamics with the overall protection factor ranging between  $10^2$ - $10^6$ . However, for the amyloidogenic DMG form, the overall protection ranges between  $10^0$ - $10^2$ , which signifies that the DMG form is highly dynamic compared to the N form. Further, the site-specific fragment-level analysis of the structural rearrangement during N to DMG to amyloid transformation indicates that the exposure of L<sub>3</sub>, α<sub>2</sub>L<sub>6</sub>β<sub>4</sub>L<sub>7</sub>β<sub>5</sub>, L<sub>8</sub>β<sub>6</sub>L<sub>9</sub> and β<sub>9</sub>L<sub>14</sub>β<sub>10</sub> regions of TDP<sup>NBD</sup> provide a platform for the amyloidogenic nucleation step where these regions from different monomeric proteins interact and assemble to form a nucleus for the amyloid aggregation.

## 5.5 References

1. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314, 130-133.
2. Buratti, E., Brindisi, A., Giombi, M., Tisminetzky, S., Ayala, Y. M., and Baralle, F. E. (2005) TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J Biol Chem* 280, 37572-37584.
3. Shiina, Y., Arima, K., Tabunoki, H., and Satoh, J. (2010) TDP-43 dimerizes in human cells in culture. *Cell Mol Neurobiol* 30, 641-652.
4. Sun, Y., and Chakrabarty, A. (2017) Phase to Phase with TDP-43. *Biochemistry* 56, 809-823.
5. Garnier, C., Devred, F., Byrne, D., Puppo, R., Roman, A. Y., Malesinski, S., Golovin, A. V., Lebrun, R., Ninkina, N. N., and Tsvetkov, P. O. (2017) Zinc binding to RNA recognition motif of TDP-43 induces the formation of amyloid-like aggregates. *Sci Rep* 7, 6812.

6. Zacco, E., Graña-Montes, R., Martin, S. R., de Groot, N. S., Alfano, C., Tartaglia, G. G., and Pastore, A. (2019) RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *J Mol Biol* 431, 1671-1688.
7. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P., and Eisenberg, D. S. (2019) Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat Struct Mol Biol* 26, 619-627.
8. Baldwin, R. L., Frieden, C., and Rose, G. D. (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78, 2725-2737.
9. Baldwin, R. L., and Rose, G. D. (2013) Molten globules, entropy-driven conformational change and protein folding. *Curr Opin Struct Biol* 23, 4-10.
10. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375, 513-515.
11. Hoeltzli, S. D., and Frieden, C. (1995) Stopped-flow NMR spectroscopy: real-time unfolding studies of 6-<sup>19</sup>F-tryptophan-labeled Escherichia coli dihydrofolate reductase. *Proc Natl Acad Sci U S A* 92, 9318-9322.
12. Jha, S. K., and Udgaonkar, J. B. (2009) Direct evidence for a dry molten globule intermediate during the unfolding of a small protein. *Proc Natl Acad Sci U S A* 106, 12289-12294.
13. Reiner, A., Henklein, P., and Kiefhaber, T. (2010) An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proc Natl Acad Sci U S A* 107, 4955-4960.
14. Roche, J., Caro, J. A., Norberto, D. R., Barthe, P., Roumestand, C., Schlessman, J. L., Garcia, A. E., García-Moreno, B. E., and Royer, C. A. (2012) Cavities determine the pressure unfolding of proteins. *Proc Natl Acad Sci U S A* 109, 6945-6950.
15. Fu, Y., Kasinath, V., Moorman, V. R., Nucci, N. V., Hilser, V. J., and Wand, A. J. (2012) Coupled motion in proteins revealed by pressure perturbation. *J Am Chem Soc* 134, 8543-8550.

16. Sarkar, S. S., Udgaonkar, J. B., and Krishnamoorthy, G. (2013) Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* 105, 2392-2402.
17. Jha, S. K., and Marqusee, S. (2014) Kinetic evidence for a two-stage mechanism of protein denaturation by guanidinium chloride. *Proc Natl Acad Sci U S A* 111, 4856-4861.
18. Dasgupta, A., Udgaonkar, J. B., and Das, P. (2014) Multistage unfolding of an SH3 domain: an initial urea-filled dry molten globule precedes a wet molten globule with non-native structure. *J Phys Chem B* 118, 6380-6392.
19. Mishra, P., and Jha, S. K. (2017) An Alternatively Packed Dry Molten Globule-like Intermediate in the Native State Ensemble of a Multidomain Protein. *J Phys Chem B* 121, 9336-9347.
20. Pillai, M., and Jha, S. K. (2019) The Folding and Aggregation Energy Landscapes of Tethered RRM Domains of Human TDP-43 Are Coupled via a Metastable Molten Globule-like Oligomer. *Biochemistry* 58, 608-620.
21. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131, 266-280.
22. Singh, J., Sabareesan, A. T., Mathew, M. K., and Udgaonkar, J. B. (2012) Development of the structural core and of conformational heterogeneity during the conversion of oligomers of the mouse prion protein to worm-like amyloid fibrils. *J Mol Biol* 423, 217-231.
23. Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* 2, 522-531.
24. Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins. *Adv Protein Chem* 21, 287-386.
25. Qu, Y., and Bolen, D. W. (2003) Hydrogen exchange kinetics of RNase A and the urea:TMAO paradigm. *Biochemistry* 42, 5837-5849.
26. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17, 75-86.

27. Chen, G. (2014) *Characterization of protein therapeutics using mass spectrometry*, Springer.
28. Hughson, F. M., Wright, P. E., and Baldwin, R. L. (1990) Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249, 1544-1548.

## **Chapter 6.**

# **Conclusions and Future Directions**

---

In this concluding chapter, we summarize the contributions of this thesis and discuss some important future directions of the work.

## 6.1 Purpose and findings

The central problem addressed in this thesis is understanding the nature of the early intermediate states during protein unfolding and misfolding reaction, and their implication in protein folding and disease. We briefly discuss below the specific aims and respective findings of this thesis:

### *Do DMG-like intermediates exist for large multi-domain proteins?*

We studied the unfolding and misfolding reactions of multi-domain proteins, human serum albumin and nucleic acid binding domains of TDP-43 in chapter 2, chapter 3 and chapter 4. Our findings suggest that DMG-like intermediates can be stabilized in equilibrium for large multi-domain proteins at conditions favoring the partial unfolding of proteins. Moreover, we showed that the unfolding and misfolding reaction of large multi-domain proteins begins via the formation of DMG-like intermediate states.

### *Can DMGs be stabilized using cosolvents?*

In chapter 3, we studied the effect of the base, which is a simple cosolvent, on the protein structure, where we revealed that the bases induce structural expansion in the inter-domain region of human serum albumin without core solvation, characteristic to DMG-like state. Our finding indicates that DMGs-like states are not limited to folding-unfolding transitions and other cosolvents, like base, can also stabilize the DMG-like states.

### *Is the DMG-native transition cooperative?*

The DMG-native state transition has been previously shown to be a highly cooperative step for a small protein, single-chain monellin. However, our findings from chapter 3, where we observed the base-induced formation of DMG-like state, show that only the inter-domain region of human serum albumin consists of DMG-like characteristics and other regions of the protein remain native-like. These results infer that the transition between the native and DMG-like state is non-cooperative and the formation of DMG-like state can occur in parts in different regions of a protein.

---

***Whether DMGs universal?***

The evidence of DMG-like state in two functionally and structurally distinct multi-domain proteins, human serum albumin and nucleic acid binding domains of TDP-43 in varying conditions suggest for the universality of the DMG-like state (chapter 2, chapter 3 and chapter 4). Our results, in unison, infer that the formation of a DMG-like state is the earliest structural transition a protein encounter, regardless of the protein's size/function/identity.

***What is the energetic contribution of different forces in protein stability?***

With evidence of the presence of DMG-like states, we argued that the disruption of tertiary packing of the side-chains of a protein and hydrophobic solvation of the protein core, during protein unfolding, are two distinct steps. This inference allowed us to extract the contribution of different forces in protein stability (chapter 2 and chapter 3). Notably, we estimated the energetic contribution of the van der Waals interactions, which governs the tertiary packing of the side-chains. We showed that vdW interaction contributes 40% to the stability of the native state, opposing the common notion that the hydrophobic interactions are the major contributor in stabilizing a protein.

***What are the initial steps during the misfolding reactions?***

We explored, in chapter 4, the early steps in the low-pH induced misfolding of the nucleic acid binding domains of TDP-43 (TDP<sup>NBD</sup>), which eventually leads to the amyloid aggregation of the protein. Our findings suggest that the aggregation of the protein begins with the formation of a DMG-like state, making the protein amyloidogenic. We inferred that this initial native to DMG-like transition increases the conformational dynamics of the protein, crucial for the nucleation step of amyloid aggregation of TDP<sup>NBD</sup>.

***Whether DMGs are the common intermediates in the folding and misfolding pathway?***

Since we have shown that both unfolding and misfolding of proteins begin with the formation of DMG-like states (chapter 2 and chapter 4), therefore it is rational to conclude that DMGs are common intermediate states in the folding-unfolding and the misfolding pathway. However, the fate of a protein is decided by the available environmental conditions, favouring either unfolding or aggregation of the protein.

***How misfolded aggregation-prone state and the aggregated state differ from the native state?***

We explored the high-resolution differences between the amyloidogenic DMG state and non-amyloidogenic native state of TDP<sup>NBD</sup>, where we found that the structural dynamics



---

of the amyloidogenic DMG state is very high compared to the non-amyloidogenic native state with restricted dynamics. The higher structural dynamics permit two or more proteins to interact via otherwise inaccessible segments of protein and form the amyloid nucleus. Furthermore, we revealed the structural organization of the core of the amyloid fibril, where we showed that significant disordered regions of TDP<sup>NBD</sup> have undergone disordered to ordered transition while making the amyloid fibril core.

## 6.2 Contributions to the field

This study has made five significant contributions to the field of protein folding and misfolding:

- 1) Our findings strengthen the notion that a protein begins to unfold via a common DMG-like state, which in some cases can be a non-cooperative transition and that the tertiary structural unlocking and hydrophobic solvation are two distinct steps.
- 2) Our study challenges the common belief of hydrophobic interactions being the dominant force that stabilizes a protein. We inferred that van der Waals packing interaction contributes around 40% in protein stability proposing vdW packing interactions as an important energetic force in stabilizing the native state.
- 3) Our investigation of the aggregation reaction of TDP<sup>NBD</sup> showed that the first step during the aggregation of TDP<sup>NBD</sup> is the formation of a DMG-like state, which acts as an amyloidogenic precursor for nucleation dependent aggregation of TDP<sup>NBD</sup>. Our finding suggests that the role of DMGs is not limited to mere protein folding pathway and if conditions permit, they can also direct a protein towards the misfolding and aggregation pathway.
- 4) We differentiated the structural dynamics in the DMG-like state and the native state of TDP<sup>NBD</sup>, where we proposed that the increased structural dynamics in the DMG-like state is responsible for the amyloidogenicity of TDP<sup>NBD</sup>. Our finding is crucial for designing small molecules or other ligands that restrict the dynamics of TDP<sup>NBD</sup>, making it non-amyloidogenic.
- 5) Lastly, we revealed that the unstructured regions of TDP<sup>NBD</sup> constitute the core of amyloid fibril while undergoing disordered to ordered transition during the aggregation process. Our results can be utilized to impede the amyloid aggregation of TDP<sup>NBD</sup> by targeting and stabilizing the fragments that participate in amyloid core formation.

---

### 6.3 Limitations of the study

We identified the following limitation of this study:

- 1) All studies were performed in vitro under buffered conditions on purified protein, however, inside a cell, the conditions change dramatically. In a cell, proteins experience a crowded environment because of the presence of hundreds of other proteins and other cellular organelles. In our study, we have not explored the effect of crowding on protein folding-unfolding and misfolding pathway.
- 2) We interpreted the dryness of the core of the intermediate states using stokes shift analysis and ANS binding assay. However, since these are site-specific probes, reporting on particular regions of the protein, it is possible that other regions may or may not be solvent-exposed.
- 3) We have utilized highly destabilizing low-pH condition for aggregation and observed that of TDP<sup>NBD</sup> begins with the formation of a DMG-like state at low-pH (pH 3). These results suggests that extensive protonation of the buried side-chain residues could be a key trigger for initiation of misfolding and aggregation reaction. However, it remains to be determined how proteins aggregate in native-like stabilizing conditions (under physiological condition).
- 4) The structural dynamics and organization of different forms of TDP<sup>NBD</sup> were studied using hydrogen-deuterium exchange coupled to mass spectrometry (HDX-MS) for 10 peptide fragments (pepsin digests) of TDP<sup>NBD</sup>, giving us fragment level information. However, using a cocktail of different proteolytic enzymes, which will generate a range of overlapping peptide fragments of the protein, the resolution can be improved to get residue level information of the dynamics and organization.

### 6.4 Future directions

- 1) The protein folding-unfolding transition can be revisited in the presence of crowding agents, like ficol and PEG, to investigate the presence of DMG-like intermediates in conditions mimicking the cellular environment.
- 2) We have shown the involvement and significance of DMG-like states in both the folding-unfolding pathway and disease-related aggregation pathway. However, we have not explored the role of DMG-like state in the function of a protein. It will be crucial to explore the nature of the functional state of proteins. In the case of TDP<sup>NBD</sup>, various nucleotides (TG-12/UG-12) act as a natural ligand of the protein and the nucleotide bound form of the

---

protein represents the functional form. In chapter 5, we have already explored the structural dynamics differences of the native and the DMG-like form of TDP<sup>NBD</sup>, which provides us a baseline for the nature of the protein. Hence the nature of the nucleotide bound functional form of TDP<sup>NBD</sup> can be identified by monitoring the structural dynamics using HDX-MS.

- 3) In chapter 5, we have revealed the structural organization of the amyloid fibrils of TDP<sup>NBD</sup> formed at low-pH thermal stress. However, in a recent study, TDP<sup>NBD</sup> has been shown to form amyloid structure at low-pH with high salt concentration. It will be interesting to compare the two fibril forms and understand how different aggregation mechanisms impacts the overall morphology and structural organization of the amyloid fibrils.

### **List of Publications**

1. Nirbhik Acharya, Prajna Mishra and Santosh Kumar Jha. Evidence for dry molten globule-like domains in the pH-induced equilibrium folding intermediate of a multidomain protein. *The Journal of Physical Chemistry Letters*. 2016;7(1):173-9.
2. Nirbhik Acharya, Prajna Mishra and Santosh Kumar Jha. A dry molten globule-like intermediate during the base-induced unfolding of a multidomain protein. *Physical Chemistry Chemical Physics*. 2017;19(44):30207-16.