Mapping the energy landscape of amyloid-like aggregation of the nucleic acid binding domain of TDP-43

by

MEENAKSHI PILLAI

10BB16A26037

A thesis submitted to the

Academy of Scientific & Innovative Research

for the award of the degree of

DOCTOR OF PHILOSOPHY

in

SCIENCE

Under the supervision of
Dr. Santosh Kumar Jha

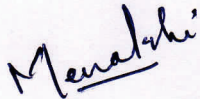
CSIR- National Chemical Laboratory, Pune

Academy of Scientific and Innovative Research

AcSIR Headquarters, CSIR-HRDC campus
Sector 19, Kamla Nehru Nagar,
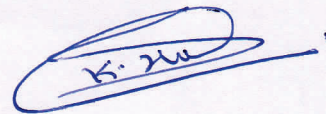Ghaziabad, U.P. – 201 002, India

JUNE 2022

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, "*Mapping the energy landscape of amyloid-like aggregation of the nucleic acid binding domain of TDP-43*", submitted by *Meenakshi Pillai* 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 Science*, embodies original research work carried-out by the student. 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 material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.

(Signature of Student)
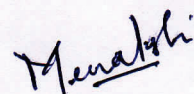Meenakshi Pillai
Date: 21/06/2022

(Signature of Supervisor)
Santosh Kumar Jha
Date: 21/06/2022

Statements of Academic Integrity

I, Meenakshi Pillai, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10BB16A26037 hereby undertake that, the thesis entitled *"Mapping the energy landscape of amyloid-like aggregation of the nucleic acid binding domain of TDP-43"* has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for *"Ethics in Research and in Governance (2020)"*.

Signature of the Student

Date : 21/06/2022

Place : Pune

It is hereby certified that the work done by the student, under my/our supervision, is plagiarism-free in accordance with the UGC Regulations on *"Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)"* and the CSIR Guidelines for *"Ethics in Research and in Governance (2020)"*.
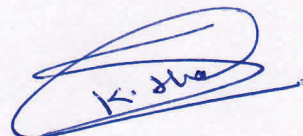
Signature of the Supervisor

Name: Santosh Kumar Jha

Date: 21/06/2022

Place: Pune

Acknowledgement

The work that I present in my thesis would not have been possible without the support and help of many people. I am happy to write a few lines about each of these fantastic people who have been my pillar of strength throughout my tenure!

I owe my deepest sense of gratitude to my mentor, Dr. Santosh K. Jha. I am extremely grateful to have been his student. He has always been an enthusiastic optimistic mentor who has moulded my scientific thought process and has given me absolute freedom to work. He has always believed in me, encouraged me, and pushed me to achieve my best. There are so many things that I have learned from him, his non-biased approach toward any scientific problem, his analysis of experimental data, and his self-critical approach to inferring the data. I would always remember his saying- there is no such thing as a negative result; every result says something! Beyond the scientific research and experimental results, he has set a remarkable example of how an ideal work culture should be. I am incredibly grateful to have been treated as a colleague by him. I still have many things to learn from him, and I am really fortunate to have had an opportunity to work with him.

I would extend my sincere gratitude to my doctoral advisory committee members, Dr. Kadhiravan Shanmuganathan, Dr. Chaitanya Athale, and Dr. N.P. Argade for their constructive feedback, guidance, and support throughout my Ph.D. tenure. Their questions motivated me and helped me gain a very different perspective about my research work.

I am grateful to Dr. Jayant Udgaonkar for giving me constant access to the mass spectrometry and DLS instrument. I would also extend my gratitude to his lab members- Harish, Rupam and Suman, who have helped me navigate through different challenges while learning the instrument and analyzing the data.

I am thrilled to have had an opportunity to work with Dr. Atanu Das. Within the short time that we have collaborated I have learnt a great many things from him.

I have received relentless academic support from my labmates, who have given me critical suggestions and support on various occasions. My labmates- Nirbhik, Prajna, Divya, Abhilasha and Sonal have constantly been supporting me, and have been an audience to my countless presentations, and I really appreciate their patience towards me. I especially have to thank Divya and Sonal for keeping the lab environment so entertaining and lively! I am extremely happy to have had constant support from Anjali Jha, who has supported me both academically and non-academically. Her confidence in me keeps my spirit high. And I thank

List of Figures

List of Tables

Synopsis

Introduction

Aggregation and deposition of TDP-43 have been shown to be a major causal factor in amyotrophic lateral sclerosis and frontotemporal lobar degeneration.[1, 2] Almost 90% of the diseased condition is due to the stress that the cell experiences over its lifetime, while only 10 % is contributed by the familial inheritance, suggesting that stress plays a very dominant role in triggering the diseased condition.[3] TDP-43 has been shown to undergo the formation of different kinds of assemblies under the diseased condition.[4, 5] Different studies have suggested the role of the different structural elements of the protein (N-terminal domain, RRM domains, and the C-terminal domain) to be responsible for the aggregate formation.[3] However, a recent study used ss-NMR to understand the core of the aggregates using these different variants of TDP-43.[6] The study suggested that the RRM domain forms the core of the aggregate while the presence of the C-terminal domain simply increases the polymorphism that exists within the amyloid structure.[6] It thus becomes key to understand how the RRM domain of TDP-43 (TDP-43^{tRRM}) can detect the stress and undergo the formation of different assemblies. In this context, it is also essential to address the sequential order of events that accompanies the protein as it undergoes a structural transformation to form the amyloid-like aggregates.

Statement of Problem

The aggregation process has been a common denominator across all the different neurodegenerative diseases. It has been observed that the primary sequence of the protein involved in different diseases are different, but the nature of the final aggregates remains similar, suggesting a common pathway for the aggregation reaction.[7] Additionally, we also do not understand the nature of the different intermediates observed during the aggregation process and thus, targeting them for therapeutic purposes becomes challenging. In this context, it becomes essential to understand how the aggregation process begins, what are the different structural level changes that the protein molecules undergo as a result of aggregation and what are the factors that modulate these different processes.

Objectives

The following open questions were answered in the scope of this thesis work-

- How does TDP-43^{tRRM} sense stress, and what are the consequences of the stress sensing at the molecular scale?

- How is the folding and the aggregation energy landscape coupled?
- What are the conformational changes associated with the protein as it undergoes change from the native state to amyloid-like state?
- How does the internal structure of the amyloid-like aggregate look like at the site-specific level?
- How do the disease-relevant mutations, D169G and P112H, modulate the aggregation process?
- What role does ionic strength play in the aggregation process?

Methodology

The RRM domain of TDP-43 protein has been shown to be forming the core of the aggregates.[6] However, we do not understand how this domain senses stress and undergo conformational changes to form these aggregates. To understand this process, we exposed the protein to different pH and ionic strength stress and studied the formation of both equilibrium and kinetic intermediates formed as a result of the site-specific modification in the protein molecules. Using a battery of spectroscopic techniques such as fluorescence spectroscopy (Trp fluorescence, Thioflavin T fluorescence, and ANS fluorescence), anisotropy, and circular dichroism (CD), we characterized the local and the global conformational changes within the protein molecules. Additionally, the change in the size was also monitored by size exclusion chromatography and dynamic light scattering (DLS). Further, we used a site-specific and high-resolution method called hydrogen-deuterium exchange coupled to mass spectrometry to characterize the internal structure of the aggregates. This allowed us to understand and identify the polymorphism that exists within the internal structure of the different aggregates formed under different stress-like conditions.

Conclusion

We have characterized the structural changes associated with TDP-43[tRRM] as the protein senses pH and ionic strength stress. Using a range of different spectroscopic probes, we identified the presence of two different intermediates-A form and the L form. These intermediates have disrupted side-chain packing but a remarkably intact native-like secondary structure. Additionally, the size of the A form and the L form measures around 10-12 nm and 12-15 nm, respectively. The A form and the L form shows an increased ANS binding capacity suggesting that the protein molecules that form these intermediates are wet molten globule in nature. Structural organization of the A form is different as compared to the L form. These

intermediates have been shown to undergo a reversible transition to the native state upon removing the pH stress, while in the presence of persistent stress in the form of heat/salt/denaturant, the intermediates undergo a structural transformation to amyloid-like aggregate called the β form. Thus these metastable intermediates act as a bridge to link the folding and the aggregation energy landscape of TDP-43^{tRRM}.

The transition from the native state to the amyloid-like aggregate, β form has been studied with time, which allowed us to delineate the temporal order of events that accompany this transition. There are significant structural events that are associated with this transition – disruption in side-chain packing, conformational change, disorder-to-order transition, and change in the hydrodynamic radii of the protein molecules. The process of elongation monitored by ThT undergoes a mono-exponential change while the process of conformational change monitored by CD and change in the hydrodynamic radii monitored by DLS follows a bi-exponential change. The fastest change that occurs is the disruption in side-chain packing monitored by the Trp fluorescence. These different changes reflect the four different phases of the model- very fast, fast, slow, and very slow. The apparent rates calculated as a function of different protein concentrations suggested that the process of aggregation is a multi-step sequential process dependent on the protein concentration. At low protein concentration, the process of conformational change precedes the ThT elongation, while at high protein concentration, the elongation precedes the conformational conversion. These changes are only observed at the initial phase of the reaction, while the final products formed at the end of the CD and DLS remains same at both low and high protein concentration. Our results suggest a multi-step sequential model that forms the basis of the aggregation kinetics of TDP-43^{tRRM}.

We next tried to understand the role of the disease-associated mutations, D169G and P112H, in the aggregation process. We observed that the mutants D169G and P112H, does not affect the global structure of the protein molecules but however increases the stability of the protein as compared to the wild type. Our results on the kinetic analysis of the aggregation suggest that the mutant D169G behaves very similarly to wild type while the mutant, P112H shows almost a four-fold faster rate of aggregation when monitored using CD. Additionally, 20% of the signal is missing suggesting the presence of a very fast phase in the aggregation kinetics. Overall our data indicate that the electrostatic interactions modulated by the charge mutations play a key role in lowering the activation barrier associated with the aggregation reaction.

The role of ionic strength and pH in modulating the structure of the TDP-43^{tRRM} has been very prominent. All our studies suggest that TDP-43^{tRRM} can undergo the formation of

different kinds of species such as – monomeric native, dry molten globule, native-like oligomers (with wet molten globule-like characteristics), amyloid aggregates and amorphous aggregates depending upon the pH and the ionic strength of the solution. We have experimentally measured the amount of salt required for the protein molecules to transform into an amyloid-like state. We observe that the molecules remain in a native-like oligomer in low salt concentration at low pH while it forms β form at high salt concentration. However, the aggregation rate increases with increasing salt concentration, suggesting that the salt affects the transition state and decreases the activation barrier associated with the amyloid-like transformation process. Our data on different salt suggests that electroselectivity plays an important role in modulating the aggregation process.

The structure of the internal core of the aggregates is not fully understood, and there is debate in the literature about the polymorphism that exists among the different aggregates. The structural basis of the disease is often correlated with the different aggregates isolated from different patients. To understand which region of the protein molecule is involved in the aggregate formation and address the polymorphism, we mapped the aggregates using hydrogen-deuterium exchange coupled to mass spectrometry. Our results indicated that the peptide 124-131 and 248-255 remains protected and does not undergo exchange suggesting that these regions form the core of the amyloid-like aggregate. We observed that the RRM1 region of the protein remains significantly more protected in the aggregates as compared to the RRM2 domain.

References

(1) 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.
(2) 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.
(3) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.
(4) Fang, Y. S., Tsai, K. J., Chang, Y. J., Kao, P., Woods, R., Kuo, P. H., Wu, C. C., Liao, J. Y., Chou, S. C., Lin, V., Jin, L. W., Yuan, H. S., Cheng, I. H., Tu, P. H., and Chen, Y. R. (2014) Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* 5, Article 4824.

(5) Kao, P. F., Chen, Y. R., Liu, X. B., DeCarli, C., Seeley, W. W., and Jin, L. W. (2015) Detection of TDP-43 oligomers in frontotemporal lobar degeneration-TDP. *Ann. Neurol.* 78, 211-221.

(6) Shenoy, J., El Mammeri, N., Dutour, A., Berbon, M., Saad, A., Lends, A., Morvan, E., Grelard, A., Lecomte, S., Kauffmann, B., Theillet, F. X., Habenstein, B., and Loquet, A. (2020) Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. *FEBS J.* 287, 2449-2467.

(7) Glabe, C. G. (2006) Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiol. Aging* 27, 570-575.

Chapter 1

Navigating through the aggregation energy landscape of RRM domain of TDP-43 (TDP-43^{tRRM})

1.1 Neurodegeneration and protein aggregation

All cells, including neurons, experience death. In particular, neurodegeneration involves a gradual loss of specific neurons, which coincides with the clinical syndromes witnessed typically in a rapidly aging population. Neurodegeneration is an umbrella term used to define a variety of diseases, some of which include Alzheimer's,[1, 2] Parkinson's,[3] Huntington's,[4] Guam-Parkinsonism,[5] Amyotrophic lateral sclerosis (ALS),[6] Fronto-temporal lobar degeneration (FTLD).[7] Each of these distinct diseases involves deposition or aggregation of distinct proteins in different anatomical regions of the brain, often reflecting the symptoms of the diseases. For example, in Alzheimer's', we observe major deposition of Aβ protein in the medial temporal lobe and neocortical structures, and non-functionality of this region results in cognitive impairment, including memory loss, difficulty in reading and writing, loss of impulse control, etc. [8, 9] In contrast, for ALS, the deposition of TAR DNA binding protein is primarily seen in the upper and lower motor neurons. The degeneration of the upper motor neuron gives rise to hyper-excitability and spasticity, while the degeneration of the lower motor neuron results in muscle weakness, fasciculation and muscular atrophy, etc.[10] Among the causative factors that have been consistently observed in neurodegenerative diseases are chronic environmental stress, deposition of protein aggregates, and inefficient clearance of the protein aggregates from the cell due to dysfunctional ubiquitin-proteasome system, non-functional autophagosomal/lysosomal system, excitotoxic insult, synaptic failure, mitochondrial dysfunction and neuroinflammation.[11] Compelling evidence suggests that oligomerization of the proteins leading to the deposition of amyloids is the fundamental cause of neuronal loss and degeneration.[12, 13] Different diseases show deposition of different proteins such as Aβ in Alzheimer's, α-synuclein in Parkinson's, polyglutamine protein in Huntington's disease, TDP-43 in ALS and FTLD. Despite the differences in their primary sequence and overall structural fold of the soluble precursor protein, the aggregates are characterized by highly ordered cross-β sheet rich structure.[14, 15] Therefore, there appears to be common underlying mechanism that might populate generic conformational states along the energy landscape of aggregation-prone proteins. Characterization of these conformational states and understanding the kinetic barriers between them could shed light on new avenues to detect and target the pathological protein.

One poorly understood area of research in neurodegeneration is the role of nucleic acid binding domains or RRM domains in aggregation. Nearly 50% of the protein involved in

mammalian stress granules are RNA binding proteins (RBP). TDP-43 is one such RBP which contains two RRM domains, implicated in neurodegeneration. Although numerous studies on TDP-43 is focused on the role of N-terminal and C-terminal domains, the role of functional RRM domain is only now beginning to be explored. In this review, we first define the general folding and aggregation energy landscape and focus on understanding the conformational constraints for the amyloid formation. We discuss how changes in the environmental conditions modulate the energy landscape and the heterogeneity that exists among the different conformers. We then discuss the existing knowledge on the role of RRM domains of TDP-43 (TDP-43^{tRRM}) in aggregation while examining the gaps in our understanding that still need to be addressed.

1.2 Free energy landscape of folding and aggregation

The concept of the energy landscape was introduced to illustrate the conformational space available for a particular polypeptide sequence for structure formation. [16, 17] The evolution of the unfolded polypeptide along the topography of the free energy landscape to fold into a compact native structure appears to be funnel-like (Figure 1). The ruggedness of this funnel-shaped landscape depends on several different factors. It has been observed that large polypeptide sequences usually have a rugged energy landscape, which allows the protein to take a multitude of different conformational states (on- or off-pathway) before reaching the native fold. [18, 19] It becomes essential to characterize the different conformational states of the polypeptide populated along the energy landscape to understand the factors that control folding.[20] Additionally, it also provides information on the structural characteristics of the different states that bypasses the folding landscape to traverse into the aggregation energy landscape. Principally, any polypeptide sequence can adopt an amyloid-like structure irrespective of its sequence, suggesting that the amyloid fold is unanimously the global free-energy minimum for any polypeptide sequence.[13, 21] The aggregation energy landscape is quite complex owing to the competition between intramolecular and intermolecular interactions. A wide variety of conformational and structural states are possible within the aggregation energy landscape. [22] The relative depth of the different energy wells in the aggregation landscape depends not only on the sequence of the polypeptide but also on the environmental condition and the concentration of the protein, and the intermolecular interactions between the protein molecules.[22]
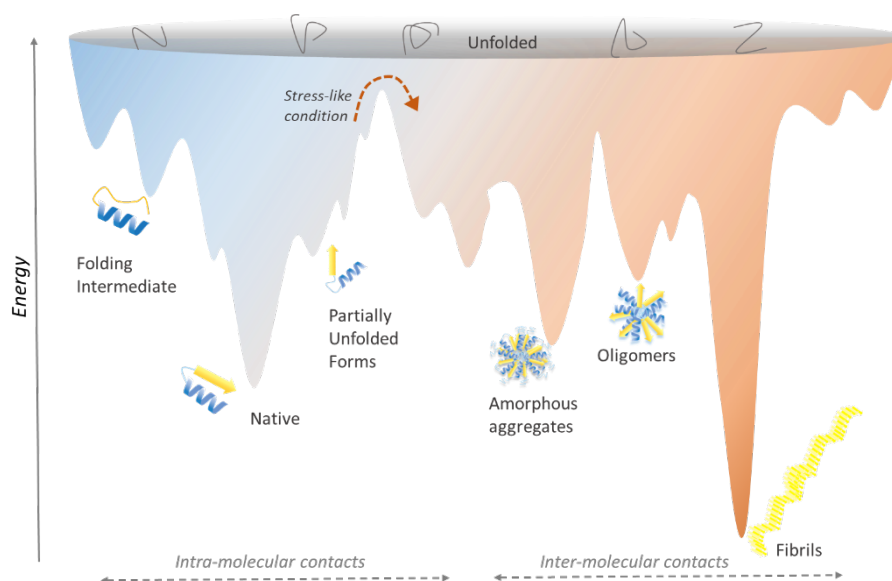
Figure 1.1: *The folding and aggregation energy landscape of protein. The folding energy landscape indicates the search of a polypeptide to achieve its native state. The aggregation energy landscape illustrated by the dark brown colour represents the different kinds of intermediates and assembly formation. The various energy minima on the landscape represents the possibility of formation of different intermediates and their respective stability.*

1.3 The requirement for partial unfolding and molten globule formation

Experimental evidence suggests that for a well-folded globular protein to undergo the formation of amyloid or amorphous aggregates, it requires to partially unfold and destabilize its native structure. [12, 13, 23-27] However, a protein in its native fold is not a static structure but is an ensemble of conformationally heterogeneous populations which undergoes fluctuations in various degrees and timescales.[28] Due to this process of conformational breathing, there is always an existence of partially unfolded conformers among the well-folded molecules, the population of the latter predominating under the native condition.[28] This equilibrium, however, shifts when there is a destabilizing condition. The partially unfolded intermediates or pre-molten globules or molten globule conformations formed under the aggregation condition are usually metastable, and therefore characterizing such species becomes extremely challenging. Formation of such molten globule precursor states has been observed along the energy landscape of many proteins such as β2-microglobulin,[29, 30] prion,[31] p53[32, 33] and TDP-43.[34, 35]

Dynamic molten globule states have been preferred by the protein as a means to survive in the various compartments of the cell, which possess distinct chemical environments and pH ranges.[36] The exposed hydrophobic surface and the increased plasticity of the molten globule allow it to function in signalling, molecular recognition and assembly formation.[36] In contrast, the same exposed hydrophobic surface might also promote amyloid-like aggregation. Molten globules of p53 protein have been shown inside the living cells, thus providing evidence for the existence of these dynamic structures in vivo. [32] Many aggregation-prone proteins show the formation of partially unfolded structures under destabilizing conditions such as β2 microglobulin,[37] prion,[38, 39] SOD1,[40] and transthyretin.[41] Transient formations of such partially unfolded species allow specific electrostatic, hydrophobic, and intermolecular interactions that enable the protein to undertake key structural re-arrangements to undergo fibril formation. [24]

1.4 Conformational polymorphism observed in oligomers, pre-fibrillar aggregates and amyloid fibrils

The generic nature of amyloid as a cross-β structure was first defined by Eanes and Glenner in 1968.[42] A protein undergoing aggregation reaction can form several different kinds of amyloid assemblies differing in their structural arrangement upon changing the solution condition[43-46] or even by keeping the condition constant.[47-49] The polymorphism observed in the morphologies of the final fibrils can often be attributed to the heterogeneity that exists among the pre-fibrillar aggregates or the intermediates. Therefore, there exists two kinds of polymorphism- polymorphism of the initial assemblies (oligomers, pre-fibrillar intermediates, protofilaments) and the polymorphism of the final aggregates at the end of the reaction. The heterogeneity observed in the initial species is often a result of the solvation condition around the protein molecules. Different environmental conditions such as mechanical agitation or different ionic strength can give rise to different types of assemblies. However, once a scaffold is formed, the growth continues, keeping the atomistic order same even when the solution condition is changed. A given type of polymorph is able to seed the same type of architecture even under different solution conditions, similar to those observed in crystalline systems.[50] This has been studied in the case of Aβ40, where the formation of striated ribbons and twisted fibrils formed under agitation and quiescent condition, respectively, when used to act as a seed, propagated similar structures under different conditions.[46] This might be due to the intrinsic stability of the extremely ordered architecture and the high energy barriers between the different polymorphs.
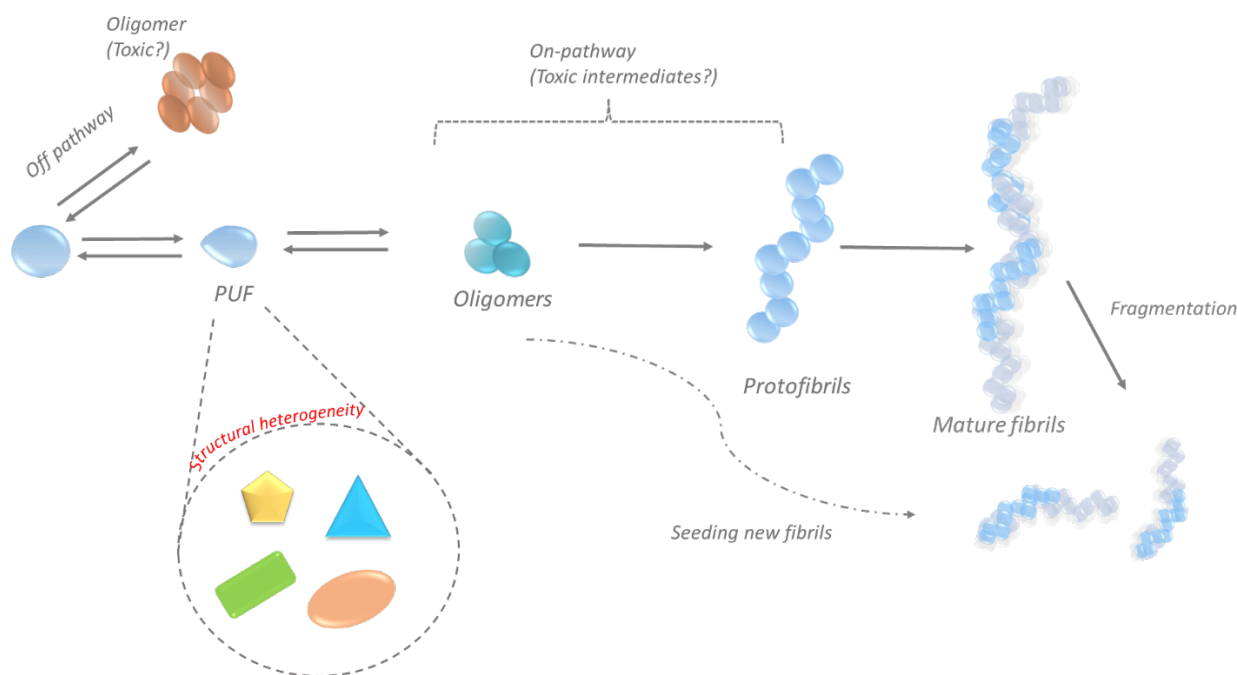
Figure 1.2: *Aggregation pathway and its complexity. The complex multi-step aggregation process and the heterogeneity observed during the aggregation reaction. The pathway represent the possibilities of forming different intermediates- on-pathway or off-pathway or different assemblies such as oligomers and protofibrils before the formation of mature fibrils.*

Across the different amyloid structures, the core of the fibrils is rich in β-sheets. Within the β-sheet bilayer, the side chains tightly interdigitate, giving rise to a "steric-zipper" interface. Three models have been proposed to form the basis of the amyloid polymorphs- packing polymorphism (amyloid segments can pack in different ways giving rise to amyloid fibrils with distinct properties), segmental polymorphism (two or more different amyloid segments of a protein can form different steric-zipper spines giving rise to different fibrils and heterosteric zippers (Non-identical β sheets interdigitate to form steric-zipper spines of distinct properties).[51] Thus, the polymorphism seen in the amyloid can be contributed by the side-chain conformation, backbone conformation or the supramolecular assembly. A number of different proteins such as Aβ40,[52-55] α-synuclein,[49, 56, 57] IAPP,[58] tau,[59, 60] and prion[61] show polymorphic amyloid fibrils. A range of different conformations adopted by the protein can often contribute distinctly to its toxicity, and transmission.[62, 63] Unlike a native fold which is evolutionarily selected, the formation of an amyloid fold is not optimized by evolutionary pressure, and hence there is a multitude of different arrangements that becomes accessible to the polypeptide sequence.[50]

1.5 Redesigning the aggregation energy landscape- Effect of cellular stress on the protein

Cells contain a complex environment fuelled by the number of different components existing inside them. In a complex microenvironment, the energy landscape will typically be rough for a multifunctional protein, which can often be correlated to frustration in the folding process.[64] A rough energy landscape also means that a protein will have several energy minima available depending upon its immediate micro-environment. This also suggests that as the protein traverses from one energy minima to the other, it is able to adopt alternative conformations depending upon the binding partners present. Subtle changes in the structure, dynamics of the protein or exposure of certain regions to allow new interactions can be a manifestation of this alternative conformation. However, such exposure or altered dynamics might also increase the likelihood of creating insignificant and disastrous interactions leading to pathological diseases.[64]

The relative probability of different conformational states (thermodynamics) and the energy barrier (kinetics) between them is highly sensitive to mutation, post-translational modifications and the environmental conditions such as pH, temperature, and presence of crowding agents.[65] Different stresses can modulate the cellular condition and reduce the energy barriers allowing easy kinetic access to the aggregation energy landscape in biologically relevant time scale. With a small change in the energy, there is a huge change in the population of the conformations that may drive the equilibrium towards misfolding and aggregation. As the cells age, the cellular proteome becomes an essential target for different stress-like conditions such as oxidative and thermal stress.[66] Studies have examined that almost 40-50% of the proteins undergo some form of oxidative damage, and nearly 30% of proteome becomes carbonylated, resulting in decreased functional activity.[67-69] In aged organisms, the thermal stability of many proteins is greatly diminished.[70] Loss of proteome-wide stability disturbs the proteostatic balance in the cell and is often accompanied by aggregation of a vast number of proteins. [71, 72] The damage to the protein occurs in the form of breakage of the protein backbone or modification of side-chain residues; the latter was found to be more abundant than other types of damage. [73-75]

1.6 Assembly-Disassembly as an energy-efficient response to stress

Some proteins or protein domains inside the cell can act as biosensors and sense environmental changes. The different types of cellular stresses experienced by the cell can change the physicochemical environment of the cell differently, causing the biosensor protein to respond differently. In particular, under oxidation stress, the most commonly oxidized side-chains are either charged amino acids (lysine and arginine) [73, 76] or neutral amino acids (cysteine, methionine, proline, histidine, and threonine).[77, 78] Such modifications often result in changes in the net charge of the protein. A similar change in electrostatics is observed when there is a change in pH. Studies have pointed out that the pH varies between compartments and is more significantly observed during aging and disease. Kroschwald and co-workers have shown acidification of the cytoplasm upon glucose starvation stress.[79] It is known that the production of ATP is hugely declined under starvation conditions.[80, 81] With a decline in the ATP level, the ATPase pump is no longer able to function and pump protons from the cytoplasm to the outside of the cell, resulting in a decrease in the cytosolic pH.[81] Under a stressed acidic condition, the proteins have been shown to undergo demixing and universally assemble to form higher-order structures, which allows the yeast cells to adapt to stress-like conditions (stress granules).[82, 83] Several in-vitro studies have demonstrated that a change in the pH destabilizes the protein molecules and induces a conformational change to form aggregates under prolong stress condition.[35, 84-87] In some proteins, protonation has been shown to act as a molecular switch to control the equilibrium between native and the partially unfolded molecules that can prime misfolded structures.[35, 37, 88-90] Therefore, cells mitigate starvation stress by coupling the protonation-deprotonation equilibrium of the biosensor protein with the assembly-disassembly equilibrium. However, prolonged stress condition shifts the equilibrium toward the aggregated state. Consistent with this idea, it has been shown that assembly-disassembly is one of the most energy-efficient ways to mitigate stress, unlike post-translational modification or degradation.[91] Additionally, the timescales of formation of these higher-order assemblies are also less, allowing the cells to quickly respond to stress-like condition.[91]

One of the widely studied higher-order assemblies in the cell are stress granules (SG) that are biomolecular condensates lacking fixed stoichiometry formed in response to environmental stress experienced by the cell.[92] SG contains stalled translation initiation factors, non-translating polyadenylated mRNA and RBP's.[93] They are highly regulated assemblies whose composition and dynamics depend upon the type of stress experienced by the cell. Under normal cellular metabolism, the disassembly of stress granules upon stress removal is essential

to maintain homeostasis.[92] The potential disassembly of SG occurs due to the weak multiple transient interactions that exist among the low complexity domains and the RNA binding domains with their target RNA.[94, 95] Any factors affecting these intermolecular interactions such as post-translation modification of the protein, pathological mutation or depleting cellular ATP might affect the residence time of the proteins in SG resulting in their altered dynamics.[96-99] Many neurodegenerative disorders are proposed to be a result of the misregulated SG pathway.[100] However, the detailed mechanism of the dynamic intermolecular interactions governing the initial reversible assembly formation and conversion to mature amyloid or amorphous aggregate has been very challenging.

1.7 Kinetic models to describe the molecular mechanism of the aggregation pathway

In order to understand the nature of transition states and kinetic barriers between different conformational states on the energy landscape, kinetic studies has proved to be an elegant tool.[27, 101] The process of protein aggregation involving interactions between an ensemble of protein molecules can be quite complex. There is a multitude of different species which can be populated as the protein undergoes conformation and structural transition to form aggregates. Several models have been proposed, [101], which are built by monitoring the process of fibrillation using a range of spectroscopic, microscopic and amyloid-specific dyes.[102-104] Based on the polymerization theory, which is often used to evaluate the aggregation features, we can define two basic models-

1) Nucleation-dependent polymerization

A nucleation-dependent polymerization (NDP) reaction involves the formation of a nucleus which is the rate-limiting step of the reaction. The formation of the nucleus involves high activation barrier and is thermodynamically unfavorable. The process of NDP reaction involves measuring both the forward and the reverse rate constants.[105, 106] Once nucleation formation occurs, the subsequent steps in the aggregation reaction occur spontaneously. NDP reaction is associated with the following characteristics[106]-

    a. The kinetics of an NDP reaction shows a lag phase. The lag time is highly dependent on the protein concentration, size of the nucleus and the association and the dissociation rate constant.[107] This weak lag phase is described by τ^2 function, where τ represents time constant of the reaction. [101]

b. The extent of the lag phase depends on the concentration of the protein; below a specific concentration, no nucleus formation occurs. This threshold concentration required for nucleus formation is called critical concentration, and it varies from polymer to polymer.

c. The addition of pre-formed nuclei (seeding) can abolish the lag phase of the reaction.

2) Linear (Isodesmic) polymerization

An isodesmic polymerization mechanism does not show the presence of any lag phase.[108-110] It can be considered to be similar to the elongation step of an NDP reaction. In the case of the linear polymerization reaction, no critical concentration of protein is required. The rate of the aggregation reaction is fastest at the start when the concentration of monomers is highest; thereafter, the reaction reaches equilibrium, and the rate subsequently decreases. It is important to note that NDP and linear polymerization are the two extreme cases that can be used to represent the aggregation mechanism. However, it is equally possible that a particular aggregation kinetic might involve both NDP and linear polymerization mechanisms. [110]

Based on the different kinetic models adopted by the protein during aggregation, few other mechanisms have been proposed. These models have been developed considering the structural features of the species undergoing aggregation and the rate-limiting step of the aggregation reaction. In the 'Templated Assembly' (TA model), the nucleus can act as a template provided the monomers attached to the nucleus undergo a simultaneous conformational conversion such that the rate-limiting step of conversion is coincidental with that of the aggregate formation. [111] In the case of the monomer-directed conversion model, conformational conversion acts as a rate-limiting step, but the process of conformational conversion of monomer occurs in the solution and not as a part of the growing aggregates.[112] In the nucleated conformational conversion model, we observe the formation of oligomers (unlike the nucleation polymerization mechanism) which assemble to undergo the conformational conversion process.[113, 114]

1.8 TDP-43 is a major protein aggregate in ALS and FTLD

Gene coding for TAR DNA binding protein is well conserved among Drosophila, human, mouse and *Caenorhabditis elegans*.[115] The 43 kDa protein, TDP-43, was first discovered in 1996 for its role in binding to the TAR DNA of HIV and repressing the transcription process. [116] TDP-43 was later studied for its role as an exon-skipping promoter

during splicing of apolipoprotein A-II (apoA-II) and cystic fibrosis transmembrane regulator (CFTR) transcripts. Despite its various vital functions, TDP-43 protein was found to be a major protein deposited in the spinal cord and brain of patients suffering from ALS and FTLD. [117, 118] Since then, several neurodegenerative diseases such as Alzheimer's, Parkinson's, Huntington's, and Guam-Parkinsonism dementia have shown intra-neuronal deposition of TDP-43 aggregates.[119-121] Together, these diseases showing deposition of TDP-43 protein were termed TDP-43 proteinopathies.[122]
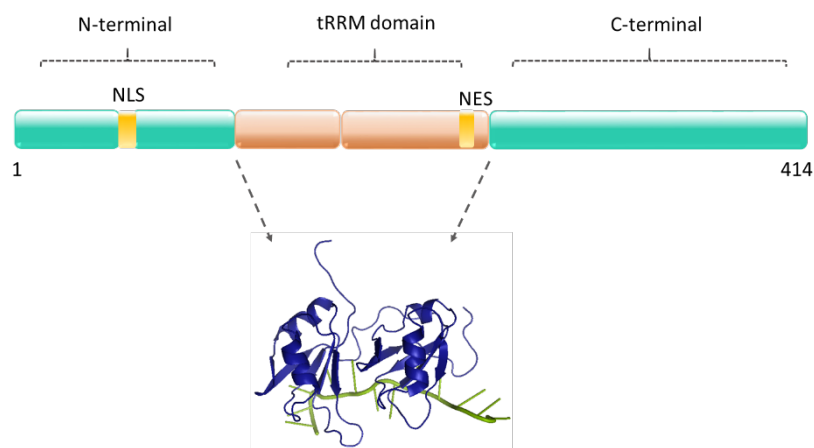
Figure 1.3: *Schematic representation of TDP-43 protein. The protein shows four domains- N-terminal domain, the two RRM domains (tRRM domain) and the C-terminal domain. 3D representation of tRRM domain with the nucleic acid is taken from PDB4BS2.*

Structurally, TDP-43 protein has four major domains- an N-terminal domain (aa, 1-102), two -RNA Recognition motif domains: RRM1 (aa 104-176) and RRM2 (aa 192-262) linked via a linker region and a C-terminal region (aa 274-414).[123-126] The protein has two signal sequences –nuclear localization signal (NLS; aa 82-98) and a nuclear export signal (NES; aa 239-250), through which it shuttles between the nucleus and cytoplasm to perform its function. [127] In the case of ALS, however, the protein is redistributed, showing an increased deposition in the form of inclusions in the cytoplasm.[128] Due to high aggregation propensity and poor solubility, the complete structure of TDP-43 remains unresolved until now, but the role of individual domains in aggregation is being addressed.

Experimental evidence suggests that TDP-43 protein remains in a monomer-dimer equilibrium inside the cell under normal conditions.[129, 130] The process of dimerization is shown to be facilitated by the N-terminal domain of TDP-43. Arguably, studies have suggested that

dimerization may be involved in the aggregation of the protein.[129-131] In contrast to this idea, it has been reasoned that the dimeric NTD enhances the solubility of the protein through its pre-mRNA splicing activity and protects against inclusion formation. [132] Taken together, these data imply that the N-terminal region may increase the local concentration of the protein and can serve as a prerequisite for aggregation reaction.[133] The C-terminal domain of TDP-43 consists of a glycine-rich region and a segment containing polar uncharged amino acids such as glutamine and asparagine (Q/N). This domain architecture is highly comparable to the prion-like domains of several proteins such as Sup35, Fused-In-Sarcoma (FUS), TATA-box binding protein associated factor 15 (TAF-15) and Ewing sarcoma breakpoint region 1 (EWSR1) and heterogeneous nuclear ribonucleoproteins (hnRNPs) family of proteins.[134] A majority of studies in the literature focus on the contribution of the C-terminal region toward the aggregation of the protein. A chief reason for using the C-terminal region as a model domain is because it harbors the majority of the mutations and phosphorylation sites.[135] However, the C-terminal region being intrinsically disordered, is better able to tolerate the mutation than the highly conserved RRM domain. Additionally, it has also been seen that C-terminal fragments of sizes ~25-35 kDa are observed in the inclusion bodies.[136, 137] These fragments are cleaved along the caspase site located on the RRM2 domain of TDP-43 protein.[136]

1.9 Role of RRM domains of TDP-43 in function and aggregation

RNA Recognition Motifs present in the RNA binding proteins (RBPs) are amongst the most abundant and conserved domains in the eukaryotes.[138-140] The RRM domains of TDP-43 contains a conserved set of amino acid that binds to nucleic acids and perform a host of different functions such as mRNA transcription and splicing, mRNA transport, mRNA maturation and stability, and mRNA translation. [141-144] TDP-43 has two RRM domains separated by 15 amino acid residue linker; folded into 5 β-stranded sheet and 2 α-helices arranged in β1-α1-β2-β3-α2-β4-β5 pattern.[123] Unlike the typical 4 strands present in an RRM domain,[144] there is an additional β strand present in TDP-43, which is believed to give stability to the protein.[145] The two RRM domains can bind to UG and TG rich sequences.[123, 124] RRM1 binds to UG_6 with K_d of 65.2 nM while RRM2 binds to UG_3 with K_d of 379 nm. RRM1 contains a longer loop3 region as compared to RRM2, which is believed to contribute to RRM1's higher affinity for nucleic acid interaction.[146] However, when both the RRM domains bind to UG_6, there is a synergistic binding effect that drastically increases binding affinity and reduces the K_d to 14.2 nM.[124] Interestingly, the TDP-43 RRM domains bind to RNA in a reverse manner, unlike the

typical RRM domains that usually bind in 3'-to-5' direction.[123] This reverse interaction allows the 15 residue linker to participate more extensively with the nucleic acid targets.[123] Apart from playing an important role in function, binding to nucleic acid has also been shown to rescue the protein from undergoing aggregation.[147-149]

A recent NMR study coupled with X-ray fiber diffraction data points out a strong similarity in the structure of amyloid fibrils formed by full-length TDP-43 and TDP-35 (90-414 residue).[150] The C-terminal showed polymorphic structure upon forming fibrils alone or in combination with other domains. This suggests that both the N-terminal and RRM domains are crucial in forming the core structure of TDP-43.[150] The RRM domains of TDP-43 have been studied individually[151] and in tandem[34, 35] and shown to undergo aggregation. Two regions within the RRM domain have been identified to play a vital role in the aggregation process, 166-173 in RRM1 and 246-255 in RRM2.[145, 152] There are two important electrostatic mutations present within RRM domains – P112H and D169G.[153, 154] The mutation D169G in the RRM1 domain does not affect the aggregation propensity but induces a local conformational change due to the loss of hydrogen bond with T115 residue.[155] It also increases the caspase cleavage activity, a process known to trigger aggregate formation.[155] Interestingly, these regions are also shown to undergo oxidation and result in oligomerization of the protein.[156] Cys173 of the RRM1 domain gets preferentially oxidized, resulting in a conformational change, and further oxidation of Cys175 results in the formation of the insoluble RRM1 dimer aggregate. Oxidation of a buried cysteine residue might change the local electrostatics, resulting in partial unfolding and exposure of hydrophobic patches, causing destabilization.[156] The two cysteines, Cys173 and Cys175, make contact with residues in the RRM domain, and it has been hypothesized that the loss of these contacts due to oxidation may have triggered the aggregation reaction. Oxidation of cysteine (Cys198 and Cys244) in RRM2 has also been shown to cause disulphide-driven dimer formation, which consequently associates to form aggregates.[157] Structural studies have revealed that the RRM2 domain of TDP-43 populates an unfolding intermediate at 4M GdmCl.[158] Multiple high-energy microstates were characterized at 6M urea concentration using an all-atom replica-averaged metadynamics (RAM) simulations coupled to chemical restraints from NMR.[159] Although these states were categorized into 3 regions depending on the type of structure present, but a common feature among all these states showed the exposure of hydrophobic region on strands β3, β4, β5 that are aggregation-prone to the solvent. These results suggest that RRM2 unfolding may drive the protein to undergo aggregate formation.[159] In line with the above results, the MD

stimulation study also revealed that RRM2 forms multiple metastable states governed by non-native contact and hydrogen bond formation.[160] Experimental data showing partial destabilization of RRM domain due to zinc binding results in the formation of ThT–positive aggregates.[161]

RRM domains, in general, have been shown to be part of many aggregation-prone proteins such as FUS, hnRNPs, TAF-15 and EWSR1, which are associated with various neurodegenerative diseases.[134, 162] Remarkably, around ~240 proteins within the human proteome contain prion-like low complexity domain, many of which also harbor a canonical RNA Recognition Motif.[10, 163] Numerous studies have focused on the role of low complexity domain in misfolding and aggregation, but investigations with emphasis on the role of RRM domain in aggregation is only now evolving.[164]

1.10 Conclusion

Multiple studies have emphasized the effect of different stress-like conditions on the RRM domain and showed the transformation from monomer to aggregate-like structures. However, we do not yet understand how the RRM domain detects the different stress-like conditions. It remains unanswered if different stress results in different types of initial conformations? If the protein is able to detect the stress, what are the initial molecular level changes that occur in the protein? Do metastable assemblies form in response to stress? Are these changes localized to a specific region of the protein structure? What is the temporal order of events upon stress detection? What is the nature of the kinetic barriers? Does the initial structural conformation dictate the structure of the final aggregate? How does the disease-relevant mutation regulate the kinetics of the aggregation process? How do the structural dynamics of the protein evolve as it undergoes to form an ordered aggregate? Which regions form the core of the aggregate? This thesis aims to address some of these exciting questions. In this study, we have attempted to construct the aggregation energy landscape of TDP-43^{tRRM}, decode its internal core and structural dynamics and understand the effect of environmental conditions including the effect of disease-relevant electrostatic mutations on the kinetic barriers of the aggregation reaction.

1.11 References

(1) Selkoe, D. J. (1991) The molecular pathology of alzheimer's disease. *Neuron* 6, 487-498.

(2) Hardy, J. (1997) The alzheimer family of diseases: Many etiologies, one pathogenesis? *Proc. Natl. Acad. Sci. U. S. A.* 94, 2095-2097.

(3) Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A. E., and Lang, A. E. (2017) Parkinson disease. *Nat. Rev. Dis. Primers* 3, 17013.

(4) Walker, F. O. (2007) Huntington's disease. *Lancet* 369, 218-228.

(5) Steele, J. C. (2005) Parkinsonism-dementia complex of guam. *Mov. Disord.* 20 Suppl 12, S99-S107.

(6) Rowland, L. P., and Shneider, N. A. (2001) Amyotrophic lateral sclerosis. *N. Engl. J. Med.* 344, 1688-1700.

(7) Seltman, R. E., and Matthews, B. R. (2012) Frontotemporal lobar degeneration: Epidemiology, pathology, diagnosis and management. *CNS Drugs* 26, 841-870.

(8) Breijyeh, Z., and Karaman, R. (2020) Comprehensive review on alzheimer's disease: Causes and treatment. *Molecules* 25.

(9) De-Paula, V. J., Radanovic, M., Diniz, B. S., and Forlenza, O. V. (2012) Alzheimer's disease. *Subcell. Biochem.* 65, 329-352.

(10) 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, Article 25.

(11) Bossy-Wetzel, E., Schwarzenbacher, R., and Lipton, S. A. (2004) Molecular pathways to neurodegeneration. *Nat. Med.* 10 Suppl, S2-9.

(12) Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* 8, 101-106.

(13) Dobson, C. M. (2001) The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 133-145.

(14) Sunde, M., and Blake, C. (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein. Chem.* 50, 123-159.

(15) Makin, O. S., and Serpell, L. C. (2002) Examining the structure of the mature amyloid fibril. *Biochem. Soc. Trans.* 30, 521-525.

(16) Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. (1995) Navigating the folding routes. *Science* 267, 1619-1620.

(17) Dill, K. A., and Chan, H. S. (1997) From levinthal to pathways to funnels. *Nat. Struct. Biol.* 4, 10-9.

(18) Vendruscolo, M., Paci, E., Karplus, M., and Dobson, C. M. (2003) Structures and relative free energies of partially folded states of proteins. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14817-14821.

(19) Brockwell, D. J., and Radford, S. E. (2007) Intermediates: Ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* 17, 30-37.

(20) Jahn, T. R., and Radford, S. E. (2005) The yin and yang of protein folding. *FEBS J.* 272, 5962-5970.

(21) Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D., and Dobson, C. M. (1998) Amyloid fibril formation by an sh3 domain. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4224-4228.

(22) Jahn, T. R., and Radford, S. E. (2008) Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* 469, 100-117.

(23) Dobson, C. M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* 24, 329-332.

(24) Uversky, V. N., and Fink, A. L. (2004) Conformational constraints for amyloid fibrillation: The importance of being unfolded. *Biochim. Biophys. Acta* 1698, 131-153.

(25) Lansbury, P. T., Jr. (1999) Evolution of amyloid: What normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3342-3344.

(26) Fink, A. L. (1998) Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold Des* 3, R9-23.

(27) Zerovnik, E. (2002) Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. *Eur. J. Biochem.* 269, 3362-3371.

(28) Uversky, V. N., and Finkelstein, A. V. (2019) Life in phases: Intra- and inter- molecular phase transitions in protein solutions. *Biomolecules* 9.

(29) Skora, L., Becker, S., and Zweckstetter, M. (2010) Molten globule precursor states are conformationally correlated to amyloid fibrils of human beta-2-microglobulin. *J. Am. Chem. Soc.* 132, 9223-5.

(30) Carrotta, R., Bauer, R., Waninge, R., and Rischel, C. (2001) Conformational characterization of oligomeric intermediates and aggregates in beta-lactoglobulin heat aggregation. *Protein Sci.* 10, 1312-1318.

(31) Honda, R. P., Yamaguchi, K. I., and Kuwata, K. (2014) Acid-induced molten globule state of a prion protein: Crucial role of strand 1-helix 1-strand 2 segment. *J. Biol. Chem.* 289, 30355-30363.

(32) Bom, A. P., Freitas, M. S., Moreira, F. S., Ferraz, D., Sanches, D., Gomes, A. M., Valente, A. P., Cordeiro, Y., and Silva, J. L. (2010) The p53 core domain is a molten globule at low pH: Functional implications of a partially unfolded structure. *J. Biol. Chem.* 285, 2857-2866.

(33) Pedrote, M. M., de Oliveira, G. A. P., Felix, A. L., Mota, M. F., Marques, M. A., Soares, I. N., Iqbal, A., Norberto, D. R., Gomes, A. M. O., Gratton, E., Cino, E. A., and Silva, J. L. (2018) Aggregation-primed molten globule conformers of the p53 core domain provide potential tools for studying p53c aggregation in cancer. *J. Biol. Chem.* 293, 11374-11387.

(34) 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.

(35) 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.

(36) de Oliveira, G. A. P., and Silva, J. L. (2017) The push-and-pull hypothesis in protein unfolding, misfolding and aggregation. *Biophys. Chem.* 231, 20-26.

(37) McParland, V. J., Kad, N. M., Kalverda, A. P., Brown, A., Kirwin-Jones, P., Hunter, M. G., Sunde, M., and Radford, S. E. (2000) Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry* 39, 8735-46.

(38) Jain, S., and Udgaonkar, J. B. (2008) Evidence for stepwise formation of amyloid fibrils by the mouse prion protein. *J. Mol. Biol.* 382, 1228-1241.

(39) Moulick, R., Das, R., and Udgaonkar, J. B. (2015) Partially unfolded forms of the prion protein populated under misfolding-promoting conditions: Characterization by hydrogen exchange mass spectrometry and NMR. *J. Biol. Chem.* 290, 25227-25240.

(40) Niu, B., Mackness, B. C., Zitzewitz, J. A., Matthews, C. R., and Gross, M. L. (2020) Trifluoroethanol partially unfolds g93a SOD1 leading to protein aggregation: A study by native mass spectrometry and fpop protein footprinting. *Biochemistry* 59, 3650-3659.

(41) Quintas, A., Vaz, D. C., Cardoso, I., Saraiva, M. J., and Brito, R. M. (2001) Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. *J. Biol. Chem.* 276, 27207-27213.

(42) Eanes, E. D., and Glenner, G. G. (1968) X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* 16, 673-677.

(43) Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005) Competing pathways determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol.* 351, 850-864.

(44) Makarava, N., and Baskakov, I. V. (2008) The same primary structure of the prion protein yields two distinct self-propagating states. *J. Biol. Chem.* 283, 15988-15996.

(45) Kad, N. M., Thomson, N. H., Smith, D. P., Smith, D. A., and Radford, S. E. (2001) Beta(2)-microglobulin and its deamidated variant, n17d form amyloid fibrils with a range of morphologies in vitro. *J. Mol. Biol.* 313, 559-571.

(46) Petkova, A. T., Leapman, R. D., Guo, Z., Yau, W. M., Mattson, M. P., and Tycko, R. (2005) Self-propagating, molecular-level polymorphism in alzheimer's beta-amyloid fibrils. *Science* 307, 262-5.

(47) Goldsbury, C., Frey, P., Olivieri, V., Aebi, U., and Muller, S. A. (2005) Multiple assembly pathways underlie amyloid-beta fibril polymorphisms. *J. Mol. Biol.* 352, 282-298.

(48) Kad, N. M., Myers, S. L., Smith, D. P., Smith, D. A., Radford, S. E., and Thomson, N. H. (2003) Hierarchical assembly of beta2-microglobulin amyloid in vitro revealed by atomic force microscopy. *J. Mol. Biol.* 330, 785-797.

(49) Heise, H., Hoyer, W., Becker, S., Andronesi, O. C., Riedel, D., and Baldus, M. (2005) Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15871-15876.

(50) Chiti, F., and Dobson, C. M. (2017) Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* 86, 27-68.

(51) Eisenberg, D., and Jucker, M. (2012) The amyloid state of proteins in human diseases. *Cell* 148, 1188-1203.

(52) Bertini, I., Gonnelli, L., Luchinat, C., Mao, J., and Nesi, A. (2011) A new structural model of abeta40 fibrils. *J. Am. Chem. Soc.* 133, 16013-16022.

(53) Niu, Z., Zhao, W., Zhang, Z., Xiao, F., Tang, X., and Yang, J. (2014) The molecular structure of alzheimer beta-amyloid fibrils formed in the presence of phospholipid vesicles. *Angew Chem.* 53, 9294-9297.

(54) Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N., and Fandrich, M. (2009) Abeta(1-40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils. *J. Mol. Biol.* 386, 869-877.

(55) Lu, J. X., Qiang, W., Yau, W. M., Schwieters, C. D., Meredith, S. C., and Tycko, R. (2013) Molecular structure of beta-amyloid fibrils in alzheimer's disease brain tissue. *Cell* 154, 1257-1268.

(56) Bousset, L., Pieri, L., Ruiz-Arlandis, G., Gath, J., Jensen, P. H., Habenstein, B., Madiona, K., Olieric, V., Bockmann, A., Meier, B. H., and Melki, R. (2013) Structural and functional characterization of two alpha-synuclein strains. *Nat. Commun.* 4, 2575.

(57) Lemkau, L. R., Comellas, G., Kloepper, K. D., Woods, W. S., George, J. M., and Rienstra, C. M. (2012) Mutant protein a30p alpha-synuclein adopts wild-type fibril structure, despite slower fibrillation kinetics. *J. Biol. Chem.* 287, 11526-11532.

(58) Luca, S., Yau, W. M., Leapman, R., and Tycko, R. (2007) Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. *Biochemistry* 46, 13505-13522.

(59) Andronesi, O. C., von Bergen, M., Biernat, J., Seidel, K., Griesinger, C., Mandelkow, E., and Baldus, M. (2008) Characterization of alzheimer's-like paired helical filaments from the core domain of tau protein using solid-state NMR spectroscopy. *J. Am. Chem. Soc.* 130, 5922-5928.

(60) Daebel, V., Chinnathambi, S., Biernat, J., Schwalbe, M., Habenstein, B., Loquet, A., Akoury, E., Tepper, K., Muller, H., Baldus, M., Griesinger, C., Zweckstetter, M., Mandelkow, E., Vijayan, V., and Lange, A. (2012) Beta-sheet core of tau paired helical filaments revealed by solid-state NMR. *J. Am. Chem. Soc.* 134, 13982-13989.

(61) Theint, T., Nadaud, P. S., Aucoin, D., Helmus, J. J., Pondaven, S. P., Surewicz, K., Surewicz, W. K., and Jaroniec, C. P. (2017) Species-dependent structural polymorphism of y145stop prion protein amyloid revealed by solid-state NMR spectroscopy. *Nat. Commun.* 8, 753.

(62) DePace, A. H., and Weissman, J. S. (2002) Origins and kinetic consequences of diversity in sup35 yeast prion fibers. *Nat. Struct. Biol.* 9, 389-396.

(63) Tanaka, M., Chien, P., Naber, N., Cooke, R., and Weissman, J. S. (2004) Conformational variations in an infectious protein determine prion strain differences. *Nature* 428, 323-328.

(64) Gershenson, A., Gierasch, L. M., Pastore, A., and Radford, S. E. (2014) Energy landscapes of functional proteins are inherently risky. *Nat. Chem. Biol.* 10, 884-891.

(65) Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* 5, 15-22.

(66) Smith, C. D., Carney, J. M., Starke-Reed, P. E., Oliver, C. N., Stadtman, E. R., Floyd, R. A., and Markesbery, W. R. (1991) Excess brain protein oxidation and enzyme dysfunction in normal aging and in alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10540-10543.

(67) Starke-Reed, P. E., and Oliver, C. N. (1989) Protein oxidation and proteolysis during aging and oxidative stress. *Arch. Biochem. Biophys.* 275, 559-567.

(68) Sharma, H. K., and Rothstein, M. (1980) Altered enolase in aged turbatrix aceti results from conformational changes in the enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 77, 5865-5868.

(69) Carney, J. M., Starke-Reed, P. E., Oliver, C. N., Landum, R. W., Cheng, M. S., Wu, J. F., and Floyd, R. A. (1991) Reversal of age-related increase in brain protein oxidation, decrease in enzyme activity, and loss in temporal and spatial memory by chronic administration of the spin-trapping compound N-tert-butyl-alpha-phenylnitrone. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3633-3636.

(70) Oliver, C. N., Ahn, B. W., Moerman, E. J., Goldstein, S., and Stadtman, E. R. (1987) Age-related changes in oxidized proteins. *J. Biol. Chem.* 262, 5488-5491.

(71) Balch, W. E., Morimoto, R. I., Dillin, A., and Kelly, J. W. (2008) Adapting proteostasis for disease intervention. *Science* 319, 916-919.

(72) Ben-Zvi, A., Miller, E. A., and Morimoto, R. I. (2009) Collapse of proteostasis represents an early molecular event in caenorhabditis elegans aging. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14914-14919.

(73) Stadtman, E. R. (2006) Protein oxidation and aging. *Free Radic. Res.* 40, 1250-1258.

(74) de Graff, A. M., Hazoglou, M. J., and Dill, K. A. (2016) Highly charged proteins: The achilles' heel of aging proteomes. *Structure* 24, 329-336.

(75) Shacter, E. (2000) Quantification and significance of protein oxidation in biological samples. *Drug Metab. Rev.* 32, 307-326.

(76) Petrov, D., and Zagrovic, B. (2011) Microscopic analysis of protein oxidative damage: Effect of carbonylation on structure, dynamics, and aggregability of villin headpiece. *J. Am. Chem. Soc.* 133, 7016-7024.

(77) Hipkiss, A. R. (2006) Accumulation of altered proteins and ageing: Causes and effects. *Exp. Gerontol.* 41, 464-473.

(78) Rao, R. S., and Moller, I. M. (2011) Pattern of occurrence and occupancy of carbonylation sites in proteins. *Proteomics* 11, 4166-4173.

(79) Kroschwald, S., Munder, M. C., Maharana, S., Franzmann, T. M., Richter, D., Ruer, M., Hyman, A. A., and Alberti, S. (2018) Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery. *Cell Rep.* 23, 3327-3339.

(80) Dechant, R., Binda, M., Lee, S. S., Pelet, S., Winderickx, J., and Peter, M. (2010) Cytosolic pH is a second messenger for glucose and regulates the pka pathway through v-atpase. *EMBO J.* 29, 2515-2526.

(81) Orij, R., Brul, S., and Smits, G. J. (2011) Intracellular pH is a tightly controlled signal in yeast. *Biochim. Biophys. Acta* 1810, 933-944.

(82) Munder, M. C., Midtvedt, D., Franzmann, T., Nuske, E., Otto, O., Herbig, M., Ulbricht, E., Muller, P., Taubenberger, A., Maharana, S., Malinovska, L., Richter, D., Guck, J., Zaburdaev, V., and Alberti, S. (2016) A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *Elife* 5.

(83) Petrovska, I., Nuske, E., Munder, M. C., Kulasegaran, G., Malinovska, L., Kroschwald, S., Richter, D., Fahmy, K., Gibson, K., Verbavatz, J. M., and Alberti, S. (2014) Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *Elife*.

(84) Singh, J., and Udgaonkar, J. B. (2016) Unraveling the molecular mechanism of pH-induced misfolding and oligomerization of the prion protein. *J. Mol. Biol.* 428, 1345-1355.

(85) Colon, W., and Kelly, J. W. (1992) Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* 31, 8654-8660.

(86) Pasquato, N., Berni, R., Folli, C., Alfieri, B., Cendron, L., and Zanotti, G. (2007) Acidic pH-induced conformational changes in amyloidogenic mutant transthyretin. *J. Mol. Biol.* 366, 711-9.

(87) Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Ionescu-Zanetti, C., Millett, I., and Fink, A. L. (2001) Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* 40, 3525-3535.

(88) Patni, D., and Jha, S. K. (2021) Protonation-deprotonation switch controls the amyloid-like misfolding of nucleic-acid-binding domains of TDP-43. *J. Phys. Chem. B* 125, 8383-8394.

(89) Hornemann, S., and Glockshuber, R. (1998) A scrapie-like unfolding intermediate of the prion protein domain prp(121-231) induced by acidic pH. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6010-6014.

(90) Peralvarez-Marin, A., Barth, A., and Graslund, A. (2008) Time-resolved infrared spectroscopy of pH-induced aggregation of the alzheimer abeta(1-28) peptide. *J. Mol. Biol.* 379, 589-596.

(91) Rabouille, C., and Alberti, S. (2017) Cell adaptation upon stress: The emerging role of membrane-less compartments. *Curr. Opin. Cell Biol.* 47, 34-42.

(92) Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016) Distinct stages in stress granule assembly and disassembly. *Elife* 5, No. e18413.

(93) Li, Y. R., King, O. D., Shorter, J., and Gitler, A. D. (2013) Stress granules as crucibles of ALS pathogenesis. *J Cell Biol* 201, 361-372.

(94) Guo, L., and Shorter, J. (2015) It's raining liquids: RNA tunes viscoelasticity and dynamics of membraneless organelles. *Mol. Cell* 60, 189-192.

(95) Wolozin, B. (2012) Regulated protein aggregation: Stress granules and neurodegeneration. *Mol. Neurodegener.* 7, 56.

(96) Mackenzie, I. R., Nicholson, A. M., Sarkar, M., Messing, J., Purice, M. D., Pottier, C., Annu, K., Baker, M., Perkerson, R. B., Kurti, A., Matchett, B. J., Mittag, T., Temirov, J., Hsiung, G. R., Krieger, C., Murray, M. E., Kato, M., Fryer, J. D., Petrucelli, L., Zinman, L., Weintraub, S., Mesulam, M., Keith, J., Zivkovic, S. A., Hirsch-Reinshagen, V., Roos, R. P., Zuchner, S., Graff-Radford, N. R., Petersen, R. C., Caselli, R. J., Wszolek, Z. K., Finger, E., Lippa, C., Lacomis, D., Stewart, H., Dickson, D. W., Kim, H. J., Rogaeva, E., Bigio, E., Boylan, K. B., Taylor, J. P., and Rademakers, R. (2017) Tia1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* 95, 808-816 e9.

(97) Kim, H. J., Kim, N. C., Wang, Y. D., Scarborough, E. A., Moore, J., Diaz, Z., MacLea, K. S., Freibaum, B., Li, S., Molliex, A., Kanagaraj, A. P., Carter, R., Boylan, K. B., Wojtas, A. M., Rademakers, R., Pinkus, J. L., Greenberg, S. A., Trojanowski, J. Q., Traynor, B. J., Smith, B. N., Topp, S., Gkazi, A. S., Miller, J., Shaw, C. E., Kottlors, M., Kirschner, J., Pestronk, A., Li, Y. R., Ford, A. F., Gitler, A. D., Benatar, M., King, O. D., Kimonis, V. E., Ross, E. D., Weihl, C. C., Shorter, J., and Taylor, J. P. (2013) Mutations in prion-like domains in hnrnpa2b1 and hnrnpa1 cause multisystem proteinopathy and ALS. *Nature* 495, 467-473.

(98) Panas, M. D., Ivanov, P., and Anderson, P. (2016) Mechanistic insights into mammalian stress granule dynamics. *J. Cell Biol.* 215, 313-323.

(99) Jain, S., Wheeler, J. R., Walters, R. W., Agrawal, A., Barsic, A., and Parker, R. (2016) Atpase-modulated stress granules contain a diverse proteome and substructure. *Cell* 164, 487-498.

(100) Advani, V. M., and Ivanov, P. (2020) Stress granule subtypes: An emerging link to neurodegeneration. *Cell Mol. Life Sci.* 77, 4827-4845.

(101) Kumar, S., and Udgaonkar, J. B. (2010) Mechanisms of amyloid fibril formation by proteins. *Current Science* 98, 639-656.

(102) Lomakin, A., Teplow, D. B., Kirschner, D. A., and Benedek, G. B. (1997) Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7942-7947.

(103) Modler, A. J., Gast, K., Lutsch, G., and Damaschun, G. (2003) Assembly of amyloid protofibrils via critical oligomers--a novel pathway of amyloid formation. *J. Mol. Biol.* 325, 135-148.

(104) Scheibel, T., Bloom, J., and Lindquist, S. L. (2004) The elongation of yeast prion fibers involves separable steps of association and conversion. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2287-2292.

(105) Bishop, M. F., and Ferrone, F. A. (1984) Kinetics of nucleation-controlled polymerization. A perturbation treatment for use with a secondary pathway. *Biophys. J.* 46, 631-644.

(106) Ferrone, F. (1999) Analysis of protein aggregation kinetics. *Methods Enzymol.* 309, 256-274.

(107) Kodaka, M. (2004) Interpretation of concentration-dependence in aggregation kinetics. *Biophys Chem* 109, 325-332.

(108) Oosawa, F., and Kasai, M. (1962) A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol.* 4, 10-21.

(109) Romberg, L., Simon, M., and Erickson, H. P. (2001) Polymerization of ftsz, a bacterial homolog of tubulin. Is assembly cooperative? *J. Biol. Chem.* 276, 11743-11753.

(110) Frieden, C. (2007) Protein aggregation processes: In search of the mechanism. *Protein Sci.* 16, 2334-2344.

(111) Uratani, Y., Asakura, S., and Imahori, K. (1972) A circular dichroism study of salmonella flagellin: Evidence for conformational change on polymerization. *J. Mol. Biol.* 67, 85-98.

(112) Prusiner, S. B. (1982) Novel proteinaceous infectious particles cause scrapie. *Science* 216, 136-144.

(113) Fu, Z., Aucoin, D., Davis, J., Van Nostrand, W. E., and Smith, S. O. (2015) Mechanism of nucleated conformational conversion of abeta42. *Biochemistry* 54, 4197-4207.

(114) Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F., and Lindquist, S. L. (2000) Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science* 289, 1317-1321.

(115) Wang, H. Y., Wang, I. F., Bose, J., and Shen, C. K. (2004) Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics* 83, 130-139.

(116) Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F., and Gaynor, R. B. (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* 69, 3584-3596.

(117) 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.

(118) 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.

(119) 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.

(120) Hasegawa, M., Arai, T., Akiyama, H., Nonaka, T., Mori, H., Hashimoto, T., Yamazaki, M., and Oyanagi, K. (2007) TDP-43 is deposited in the guam parkinsonism-dementia complex brains. *Brain* 130, 1386-1394.

(121) Higashi, S., Iseki, E., Yamamoto, R., Minegishi, M., Hino, H., Fujisawa, K., Togo, T., Katsuse, O., Uchikado, H., Furukawa, Y., Kosaka, K., and Arai, H. (2007) Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of alzheimer's disease and dementia with lewy bodies. *Brain Res.* 1184, 284-294.

(122) Gao, J., Wang, L., Huntley, M. L., Perry, G., and Wang, X. (2018) Pathomechanisms of TDP-43 in neurodegeneration. *J. Neurochem.* 146, 7-20.

(123) Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ue, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. T. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* 20, 1443-1449.

(124) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* 37, 1799-1808.

(125) Qin, H. N., Lim, L. Z., Wei, Y. Y., and Song, J. X. (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. *Proceedings of the National Academy of Sciences of the United States of America* 111, 18619-18624.

(126) Mompean, M., Romano, V., Pantoja-Uceda, D., Stuani, C., Baralle, F. E., Buratti, E., and Laurents, D. V. (2016) The TDP-43 N-terminal domain structure at high resolution. *FEBS J* 283, 1242-60.

(127) 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.

(128) Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., and Lee, V. M. (2008) Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *J. Biol. Chem.* 283, 13302-13309.

(129) Shiina, Y., Arima, K., Tabunoki, H., and Satoh, J. (2010) TDP-43 dimerizes in human cells in culture. *Cell. Mol. Neurobiol.* 30, 641-652.

(130) Zhang, Y. J., Caulfield, T., Xu, Y. F., Gendron, T. F., Hubbard, J., Stetler, C., Sasaguri, H., Whitelaw, E. C., Cai, S., Lee, W. C., and Petrucelli, L. (2013) The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet.* 22, 3112-3122.

(131) Afroz, T., Hock, E. M., Ernst, P., Foglieni, C., Jambeau, M., Gilhespy, L. A. B., Laferriere, F., Maniecka, Z., Pluckthun, A., Mittl, P., Paganetti, P., Allain, F. H. T., and Polymenidou, M. (2017) Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat. Commun.* 8, 45.

(132) Jiang, L. L., Xue, W., Hong, J. Y., Zhang, J. T., Li, M. J., Yu, S. N., He, J. H., and Hu, H. Y. (2017) The N-terminal dimerization is required for TDP-43 splicing activity. *Sci. Rep.* 7, 6196.

(133) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.

(134) King, O. D., Gitler, A. D., and Shorter, J. (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* 1462, 61-80.

(135) Pesiridis, G. S., Lee, V. M., and Trojanowski, J. Q. (2009) Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 18, R156-62.

(136) Zhang, Y. J., Xu, Y. F., Dickey, C. A., Buratti, E., Baralle, F., Bailey, R., Pickering-Brown, S., Dickson, D., and Petrucelli, L. (2007) Progranulin mediates caspase-dependent cleavage of TAR DNA binding protein-43. *J. Neurosci.* 27, 10530-10534.

(137) 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.

(138) Gerstberger, S., Hafner, M., and Tuschl, T. (2014) A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829-845.

(139) Marchese, D., de Groot, N. S., Lorenzo Gotor, N., Livi, C. M., and Tartaglia, G. G. (2016) Advances in the characterization of RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* 7, 793-810.

(140) Conlon, E. G., and Manley, J. L. (2017) RNA-binding proteins in neurodegeneration: Mechanisms in aggregate. *Genes Dev.* 31, 1509-1528.

(141) Buratti, E., and Baralle, F. E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of cftr exon 9. *J. Biol. Chem.* 276, 36337-36343.

(142) Casafont, I., Bengoechea, R., Tapia, O., Berciano, M. T., and Lafarga, M. (2009) TDP-43 localizes in mRNA transcription and processing sites in mammalian neurons. *J. Struct. Biol.* 167, 235-241.

(143) Tollervey, J. R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., Konig, J., Hortobagyi, T., Nishimura, A. L., Zupunski, V., Patani, R., Chandran, S., Rot, G., Zupan, B., Shaw, C. E., and Ule, J. (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* 14, 452-458.

(144) Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 272, 2118-2131.

(145) Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takahashi, R., Kitahara, R.,

Ito, H., Fujiwara, N., and Urushitani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* 288, 14886-14905.

(146) Kuo, P. H., Chiang, C. H., Wang, Y. T., Doudeva, L. G., and Yuan, H. S. (2014) The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res* 42, 4712-4722.

(147) Huang, Y. C., Lin, K. F., He, R. Y., Tu, P. H., Koubek, J., Hsu, Y. C., and Huang, J. J. (2013) Inhibition of TDP-43 aggregation by nucleic acid binding. *Plos One* 8, No. e64002.

(148) Sun, Y., Arslan, P. E., Won, A., Yip, C. M., and Chakrabartty, A. (2014) Binding of TDP-43 to the 3'utr of its cognate mRNA enhances its solubility. *Biochemistry* 53, 5885-5894.

(149) Zacco, E., Grana-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.

(150) Shenoy, J., El Mammeri, N., Dutour, A., Berbon, M., Saad, A., Lends, A., Morvan, E., Grelard, A., Lecomte, S., Kauffmann, B., Theillet, F. X., Habenstein, B., and Loquet, A. (2020) Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. *FEBS J.* 287, 2449-2467.

(151) Zacco, E., Martina, S. R., Thorogate, R., and Pastore, A. (2018) The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.* 11, Article 372.

(152) Shodai, A., Ido, A., Fujiwara, N., Ayaki, T., Morimura, T., Oono, M., Uchida, T., Takahashi, R., Ito, H., and Urushitani, M. (2012) Conserved acidic amino acid residues in a second RNA recognition motif regulate assembly and function of TDP-43. *Plos One* 7, Article e52776.

(153) Austin, J. A., Wright, G. S. A., Watanabe, S., Grossmann, J. G., Antonyuk, S. V., Yamanaka, K., and Hasnain, S. S. (2014) Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4309-4314.

(154) Moreno, F., Rabinovici, G. D., Karydas, A., Miller, Z., Hsu, S. C., Legati, A., Fong, J., Schonhaut, D., Esselmann, H., Watson, C., Stephens, M. L., Kramer, J., Wiltfang, J., Seeley, W. W., Miller, B. L., Coppola, G., and Grinberg, L. T. (2015) A novel mutation P112H in the TARDBP gene associated with frontotemporal lobar degeneration without motor neuron disease and abundant neuritic amyloid plaques. *Acta Neuropathol. Commun.* 3, Article 19.

(155) Chiang, C. H., Grauffel, C., Wu, L. S., Kuo, P. H., Doudeva, L. G., Lim, C., Shen, C. K., and Yuan, H. S. (2016) Structural analysis of disease-related TDP-43 D169G mutation: Linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Sci. Rep.* 6, 21581.

(156) Chang, C. K., Chiang, M. H., Toh, E. K., Chang, C. F., and Huang, T. H. (2013) Molecular mechanism of oxidation-induced TDP-43 RRM1 aggregation and loss of function. *FEBS Lett.* 587, 575-582.

(157) Rabdano, S. O., Izmailov, S. A., Luzik, D. A., Groves, A., Podkorytov, I. S., and Skrynnikov, N. R. (2017) Onset of disorder and protein aggregation due to oxidation-induced intermolecular disulfide bonds: Case study of RRM2 domain from TDP-43. *Sci. Rep.* 7, 11161.

(158) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem.* 289, 8264-8276.

(159) Tavella, D., Zitzewitz, J. A., and Massi, F. (2018) Characterization of TDP-43 RRM2 partially folded states and their significance to ALS pathogenesis. *Biophys. J.* 115, 1673-1680.

(160) Prakash, A., Kumar, V., Meena, N. K., Hassan, M. I., and Lynn, A. M. (2019) Comparative analysis of thermal unfolding simulations of RNA recognition motifs (RRMs) of TAR DNA-binding protein 43 (TDP-43). *J. Biomol. Struct. Dyn.* 37, 178-194.

(161) 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.

(162) Lukong, K. E., Chang, K. W., Khandjian, E. W., and Richard, S. (2008) RNA-binding proteins in human genetic disease. *Trends Genet.* 24, 416-425.

(163) Harrison, A. F., and Shorter, J. (2017) RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417-1438.

(164) Agrawal, S., Kuo, P. H., Chu, L. Y., Golzarroshan, B., Jain, M., and Yuan, H. S. (2019) RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Sci. Rep.* 9, 6171.

Chapter 2

The folding and aggregation energy landscapes of tethered RRM domains of human TDP-43 are coupled *via* a metastable molten globule-like oligomer

2.1 Introduction

TDP-43 (transactive response DNA binding protein 43 kDa) is a highly conserved and ubiquitously expressed nucleic acid binding protein that participates in vital cellular processes in the nucleus and the cytoplasm, including gene transcription, translation and mRNA processing.[1, 2] Misfolding and prion-like aggregation of TDP-43 and its fragments in neurons into a toxic conformation, primarily induced by chronic stress, have been implicated in the pathology of a wide spectrum of fatal neurodegenerative diseases including amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD), Alzheimer's disease, Parkinson's disease, Huntington's disease and Guam-Parkinsonism dementia, collectively known as TDP-43 proteinopathies.[1, 3-5] Most importantly, in the case of the incurable motor neuron disease ALS, around 97% of the patients show deposition of the TDP-43 and its fragments in motor neurons in the brain and the spinal cord.[6] Despite its central role in multiple neurodegenerative diseases, the molecular mechanism of misfolding and aggregation of TDP-43 from its functional native form is very poorly understood. In particular, we do not understand how the protein begins to misfold. It is important to understand the nature and structure of early intermediates during the misfolding process because for many neurodegenerative diseases the oligomeric intermediates en route to aggregation have been shown to be much more toxic compared to the final fibrillar state and hence early oligomeric intermediates are important targets for therapeutic intervention.[7, 8]

TDP-43 is a 414 amino acid residue protein consisting of four different domains (Figure 1A). The two tandem RNA recognition motifs, RRM1 and RRM2, of the protein are naturally tethered by a 15-residue linker (called hereafter tethered RRM domain of TDP-43 or TDP-43^{tRRM}, Figure 1A) and are flanked by an N-terminal and a C-terminal domain (accession no. Q13148 UniProtKB/Swiss-Prot). The N-terminal domain has a structural fold similar to axin 1 protein[9] and the C-terminal is largely disordered[10]. The aggregation prone regions of TDP-43 are primarily located in the TDP-43^{tRRM} and the C-terminal domain.[11-15] However, the mechanistic role and contribution of different regions of the protein in the aggregation process have remained elusive. TDP-43^{tRRM} is the principal functional region of TDP-43 which performs the essential nucleic acid binding functions of the protein.[15-18] The binding of nucleic acids to TDP-43^{tRRM} region increases the solubility and thermodynamic stability of the protein and reduces the protein aggregation.[19, 20] TDP-43^{tRRM} also contains two disease-associated mutations,[21] P112H and D169G, and multiple disease-relevant caspase and calpain cleavage sites.[22] The highly conserved mutation in ALS, D169G, has been shown to increase the

thermodynamic stability of both the apo and the DNA-bound TDP-43^{tRRM}.[14] Surprisingly, very little is known about the initial structural events and the nature of aggregation-primed early intermediates that couple the folding and aggregation energy landscape of TDP-43^{tRRM}. For example, we do not understand whether the aggregation of TDP-43^{tRRM} begins with the formation of protein droplet-like intermediates as has been observed for the N-terminal and the C-terminal domain of TDP-43?[10, 23] It is also not understood when and how the conformation of the protein molecule changes from the $\alpha+\beta$ containing structure to the predominantly β-sheet rich toxic assembly and whether the aggregated state resemble amyloids.[17, 24, 25]

One of the prime reasons for pathological assembly of TDP-43 is environmental stress.[17, 26] In one of the models of its pathological aggregation, during stress, native TDP-43 reversibly associates to form stress granules primarily *via* TDP-43^{tRRM} region, that could undergo irreversible aggregation upon prolonged stress.[26] It has been shown recently that some stress-like conditions like starvation acidify the cytoplasm,[27-29] that could perturb the electrostatic interactions and promote disease prone self-assembly of proteins.[30, 31] In this study, using an array of spectroscopic methods and tools of thermodynamics and kinetics, we show for the first time that TDP-43^{tRRM} remains in a pH-dependent reversible equilibrium with an oligomeric state in which protein molecules have native-like secondary structure but a disrupted tertiary structure and that is fully populated at pH 3 (referred as the acidic or the A form hereafter). The A form is highly metastable and readily undergoes an amyloid-like misfolding to a predominantly β-sheet rich aggregated state, upon small environmental perturbations. Our results shed important light on the earliest step of misfolding of TDP-43^{tRRM} from its folded state.

2.2 Materials and Methods

Protein expression and purification

The pGEX-6p1 plasmid containing the gene for the human TDP-43^{tRRM} (amino acid residues 97-261 from TDP-43 that has accession no. Q13148 UniProtKB/Swiss-Prot) was a generous gift from Jill A. Zitzewitz.[15] The details of the plasmid are described elsewhere.[15] The gene contains a His$_6$ tag and a PreScission protease cleavage site for the removal of the His$_6$ tag. The plasmid was transformed into BL21(DE3) *E. coli* cells. The cells were grown up to an OD$_{600}$ of 0.8 in Luria Bertani (LB) broth containing 100 µg/mL of ampicillin. Following this, the cells were induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20 ^{0}C for 24 h to overexpress the protein. The cells were lysed by sonication in lysis buffer (20 mM NaP$_i$,

pH 7.4, 300 mM NaCl, 30 mM imidazole, DNaseI). After pelleting the cell debris, the supernatant was passed through Ni Sepharose™ 6- Fast Flow beads (GE Healthcare) pre-equilibrated with the lysis buffer. The beads were washed with lysis buffer and the protein was eluted with the elution buffer containing 300 mM imidazole. The eluted protein was buffer exchanged in the protease cleavage buffer (50 mM Tris buffer, pH 8.0, 1 mM DTT) using HiPrep™ 26/10 desalting column (GE Healthcare) followed by overnight cleavage of His$_6$ tag at 15^0C using PreScission protease. Anion exchange chromatography using HiPrep™ Q FF 16/10 column (GE Healthcare) was performed to obtain pure His$_6$ tag cleaved protein. The pure protein was buffer exchanged in the storage buffer (10 mM KP$_i$, pH 7.2, 150 mM KCl, 1 mM DTT). The protein was >98% pure as determined by SDS PAGE (Figure S2.1A). Electrospray ionization-mass spectrometry showed the presence of a single protein species with a molecular weight of 19,429 Da corresponding to the mass of the His$_6$ tag cleaved TDP-43^{tRRM} protein (Figure S2.1B).

Buffer, solutions and experimental conditions

All the chemicals used are of highest purity grade obtained from Sigma and Sisco Research Laboratories (SRL). Buffers used at different pH are as follows: 20 mM glycine for the pH range 2.9-3.5, 20 mM sodium acetate for the pH range 3.7-5.3, 20 mM 2-(N-Morpholino)ethanesulfonic acid (MES) for the pH range 5.5-6.7, 20 mM 3-(N-Morpholino)propanesulfonic acid (MOPS) for the pH range 6.9-7.5. All the buffers contained 1 mM DTT. All the buffers and solutions were filtered with 0.2 μm filter before use. The protein concentration was determined by measuring the absorbance at 280 nm using a molar extinction coefficient of 15,470 M^{-1} cm^{-1}.

The buffer for the N form was composed of 20 mM MOPS, 1 mM DTT at pH 7. The buffer for the A form was composed of 20 mM glycine, 1 mM DTT at pH 3. The N form and the A form were equilibrated for 3 h at room temperature before performing any spectroscopic measurement. The unfolding buffer at pH 7 (pH 7U) was composed of 20 mM MOPS, 6 M guanidinium chloride (GdmCl) and 1 mM DTT at pH 7. The unfolding buffer at pH 3 (pH 3U) was composed of 20 mM glycine, 6 M GdmCl and 1 mM DTT at pH 3. The concentration of GdmCl was determined by measuring the refractive index on an Abbe type refractometer from Rajdhani Scientific Instruments Co. (Model: RSR-2).

pH-induced equilibrium structural transition monitored by fluorescence

All the fluorescence measurements were carried out on a FluoroMax-4® spectrofluorometer (HORIBA Scientific). The protein samples (8 µM) at different pH were equilibrated for 4-5 h at room temperature before performing the fluorescence measurements. The fluorescence spectrum of each sample was acquired by exciting the tryptophan residues at 295 nm and collecting the emission from 310-400 nm. The excitation and the emission slit widths were set to 1 nm and 8 nm, respectively. Data were averaged over 3 acquisitions. The fluorescence spectra of the native and the GdmCl-unfolded protein were also acquired with the excitation and the emission slit widths set to 1.5 nm and 3 nm, respectively. The wavelength of maximum fluorescence emission (λ_{max}) in the native (350 nm) and the unfolded (357 nm) forms remains invariant to the slit width settings.

Circular Dichroism

Circular Dichroism (CD) measurements were carried out on a Jasco J-815 spectropolarimeter. The spectra were collected using a data pitch of 1 nm, data integration time of 4 s, bandwidth of 2 nm and a scan speed of 20 nm/min. Each CD spectrum was the average of 3 scans. For the far-UV CD measurements, the spectra were collected in the wavelength range of 200-250 nm using a quartz cuvette of path length 0.1 cm. For the near-UV CD measurements, the range was set to 250-300 nm and the measurements were made using 1 cm path length cuvette. The concentration of the protein used for the far-UV CD measurements was 25 µM and for the near-UV CD measurements was 60 µM. The CD spectrum of the buffer was subtracted from the CD spectrum of each sample.

Difference UV absorbance spectroscopy

A double-beam UV-Visible spectrophotometer (UV 3200) from LABINDIA® analytical was used for measuring the absorbance of all the protein samples. Initially, a stock solution of protein (600 µM) was prepared in the pH 7 buffer. From the above stock, an equal amount of protein sample was added to the following four buffers- pH 3, pH 3U, pH 7, pH 7U (see above) so that the final concentration of the protein remains same in all the samples (15 µM). The absorbance of all the samples was measured over a wavelength range of 250-320 nm in a 1 cm quartz cuvette. Each sample was blanked with its respective buffer. The absorbance values were divided by the protein concentration to obtain a spectrum of molar extinction coefficient for each sample. The difference absorbance spectrum at pH 3 was obtained by

subtracting the extinction coefficient spectrum of pH 3U from pH 3 protein sample. A similar spectrum at pH 7 was obtained by subtracting the extinction coefficient spectrum of pH 7U from pH 7 protein sample.

Steady-state fluorescence anisotropy measurements

Steady-state fluorescence anisotropy was measured on a FluoroMax-4® spectrofluorometer (HORIBA Scientific) that is coupled to an FM4-Pol polarization accessory in the L-format configuration. Protein samples (10 μM) at different pH values varying from pH 3 to pH 7 were equilibrated at room temperature for 4-5 h. The anisotropy data were collected by exciting the tryptophan residues at 280 nm and collecting emission at 340 nm for 300 s, using anti photobleaching setting. The excitation and the emission slit width were set to 2 nm and 10 nm, respectively, while the integration time of 1 s was used. Steady-state fluorescence anisotropy (r) is related to the emission intensity I as follows:

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G} \tag{1}$$

where the subscripts V (vertical) and H (horizontal) refer to the position of polarizers in the excitation beam (first subscript) and the emission beam (second subscript) for every measurement. The instrumental correction factor, G, is equal to the ratio of I_{HV} to I_{HH}.

Dynamic Light Scattering (DLS)

The DLS measurements were carried out in a DynaPro 99 unit (Wyatt). The buffers used for DLS were filtered thrice with a 0.2 μm filter. All the tips and centrifuge tubes were washed with 0.2 μm filtered Milli Q water. The N form (25 μM) and the A form (25 μM) were centrifuged at 14,000 rpm for 15 min just before the DLS measurement. The protein sample was placed in a 60 μL cuvette maintained at 25^0 C inside the sample chamber of the instrument. The sample was illuminated with a laser of wavelength 829.4 nm and the scattering intensity at 90^0 and the autocorrelation function were measured simultaneously. Following settings were used, acquisition time, 3s; signal-to-noise threshold, 2.5; and sensitivity 90%. The mean autocorrelation function was obtained by averaging over 70 acquisitions and was used to obtain the distribution of hydrodynamic radii (R_H) using DynaLS software (Wyatt).

Size measurement using high molecular weight cut-off centrifugal concentrator

The N form (5 µM) and the A form (5 µM) were spun in a Vivaspin 20 (100 kDa molecular weight cut off) centrifugal concentrator from GE Healthcare. The protein sample was centrifuged at 1000 rpm for 5 min and the absorbance of the inside and the outside fraction were measured over a wavelength range of 250 – 300 nm.

8-Anilino-1-naphthalenesulfonic acid (ANS) fluorescence assay

A stock solution of 10 mM ANS was prepared in dimethyl sulphoxide (DMSO). The concentration of ANS was determined by measuring the absorbance at 350 nm and using the extinction coefficient of 5000 $M^{-1}cm^{-1}$.[32] The N form (2 µM) and the A form (2 µM) were incubated with varying concentration of ANS, ranging from 0 to 100 µM, for 15 min. After incubation, the fluorescence spectrum of each sample was measured by excitation at 380 nm and monitoring the emission from 400-600 nm. The excitation and the emission slits were 1 nm and 5 nm, respectively. For fluorescence resonance energy transfer (FRET) measurements, the protein was excited at 295 nm and the emission was monitored from 320-580 nm.

Reversibility of N ⇌ A transition

The A form (40 µM) was prepared from the N form by a pH jump from 7 to 3. Following this, the A form was diluted 10 fold into the pH 7 buffer and the pH 3 buffer (A form control) by manual mixing (dead time ~15 s) and the temporal change in fluorescence was monitored for each sample immediately after mixing. A sample of 40 µM N form was also diluted 10 fold into the pH 3 buffer and pH 7 buffer (N form control) by manual mixing and the temporal change in fluorescence was monitored for each sample immediately after mixing. The intrinsic tryptophan fluorescence of each sample was recorded by exciting at 295 nm and monitoring the emission from 310–400 nm. The excitation and the emission slit widths were set to 1.1 nm and 10 nm, respectively. The final protein concentration in all the samples was 4 µM. Each sample was also equilibrated at the room temperature for 2 h and the fluorescence signal, the ANS binding, and the steady-state fluorescence anisotropy were measured.

Characterization of the A ⇌ β transition

The A form (25 µM) transforms to a β-sheet rich oligomeric form (β form) upon heating in pH 3 buffer for 2 h at 40 ^{0}C or by incubating in a pH 3 buffer containing KCl or GdmCl. The A ⇌ β transition was characterized spectroscopically using far-UV CD and Thioflavin T (ThT) fluorescence. The far-UV CD spectrum of the β form was measured as

described above. In order to monitor the kinetics of formation of the β form, the equilibrated A form was heated in a 0.1 cm cuvette in a PTC-424S/15 Peltier system attached to the CD instrument. The change in ellipticity at 216 nm was measured at different times during the structural transition. For the kinetic experiment, a bandwidth of 2 nm and a data integration time of 4 s were used.

ThT fluorescence assay

The ThT fluorescence assay was performed in 50 mM Tris buffer at pH 8. In the assay solution, the concentration of protein was 2 µM and the concentration of ThT was 40 µM. The required amount of protein was withdrawn from the heated protein sample (25 µM) and cooled to 25^0 C before adding to the assay buffer. ThT fluorescence was measured within 1 min of protein addition. The experimental settings used were as follows- excitation wavelength, 440 nm; emission wavelength, 482 nm; excitation slit, 1 nm; and emission slit, 10 nm. Identical setting was used to measure the fluorescence of the buffer containing the free dye. All the data has been subtracted for the background fluorescence contributed by the free dye.

Size Exclusion Chromatography

The hydrodynamic properties of the N form and the salt-induced β form were studied using size exclusion chromatography (SEC) on an AKTA Pure M FPLC system (GE Healthcare). SEC was performed on a HiLoad™ 16/600 Superdex 200 pg column having a fractionation range of 10 to 600 kDa and a void volume of 39.8 mL. SEC was performed by loading 25 µM of protein sample on to the column, pre-equilibrated at the respective pH. All the buffers used for the SEC contained 150 mM KCl. A flow rate of 0.8 mL/min was used. SEC was also performed for the N form on a HiLoad™ 16/600 Superdex 75 pg column (Figure S2.1C).

For the determination of the apparent molecular weight (M_w^{app}) of the N form and the salt-induced β form, we first created a calibration curve between the partition coefficient (K_{av}) of four standard biomolecules (myoglobin, ovalbumin, albumin, IgG) and their respective molecular weights, as described in the GE Healthcare manual. The value of K_{av} was calculated from the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \tag{2}$$

where V_e is the elution volume of the protein, V_0 is the void volume of the column and V_t is the bed volume. The values of V_0 and V_t were taken from column specification sheet while the values for V_e for all the known standard biomolecules were taken from the manufacturer provided manual of HiLoad Superdex 16/600 200 pg gel filtration column. We used the experimentally measured elution volumes of the N and the salt-induced β form to calculate their partition coefficient and in turn determined their apparent molecular weight from the calibration curve.

GdmCl-induced equilibrium unfolding transitions

For equilibrium unfolding transition, protein samples (4 µM) were incubated with varying concentration of GdmCl at room temperature for 3 h. The unfolding transition was measured by exciting the tryptophan residues at 295 nm and collecting the fluorescence spectrum from 310–400 nm. Buffer corrected data was plotted at 340 nm as a function of GdmCl concentration. It is important to note that the A form transforms itself into a β form even with a minute amount of GdmCl. Hence for pH 3, the equilibrium unfolding transition represents β ⇌ U transition.

Experiments with DNA-bound TDP-43^{tRRM}

TDP-43^{tRRM} was incubated in the presence of (TG)$_6$ single-stranded DNA for 2 h at pH 7.5 to allow the formation of protein-DNA complex. In order to check the effect of DNA binding on the formation of the A form, pH 3 buffer was added to the protein-DNA complex. Additionally, 150 mM KCl was added in the pH 3 sample to examine the formation of the β form. The samples were analyzed using far-UV CD and ANS binding assay as described above. The final concentrations of protein and DNA in the far-UV CD experiments were 15 µM and 18 µM and in the ANS binding assay were 2 µM and 5 µM, respectively.

2.3 Results and Discussion

2.3.1 pH-induced structural transition of TDP-43^{tRRM}

We recombinantly expressed and purified human TDP-43^{tRRM} from *E. coli* (Materials and Methods). Under physiological conditions at pH 7, TDP-43^{tRRM} is monomeric as determined by size-exclusion chromatography (Figure S2.1C) and is hereafter referred to as the native (N) form. In order to understand the coupling between the thermodynamic energy landscape of folding, misfolding and aggregation, we perturbed the electrostatic interactions of

TDP-43[tRRM] by modulating the pH of the protein environment. TDP-43[tRRM] contains two tryptophan residues, W113 and W172 which are located in the RRM1 domain (Figure 2.1A). We observed that the intrinsic tryptophan fluorescence of TDP-43[tRRM] decreases in a sigmoidal fashion when the pH of the environment is decreased from 7.5 to 2.9 (Figure S2.2A and Figure 2.1B). We fitted the pH-dependent fluorescence data to a model, in which the structural transition from the N form to the A form is coupled to the protonation of a single ionizable group, given by a transformed Henderson-Hasselbalch equation,[33]

$$Y_{obs} = \frac{Y_A + Y_N 10^{(pH - pH_m)}}{1 + 10^{(pH - pH_m)}} \qquad (3)$$

where, Y_{obs} denotes the observed fluorescence signal for a given pH value, Y_A and Y_N are the signals of the protonated A form and the deprotonated N form, respectively, and pH_m is the midpoint of the transition. The value of pH_m was determined to be 5.6. The observation that a structural change is required for protonation suggests that the ionizable group is buried inside the protein structure.

2.3.2 Tryptophan residues in the A form have N-like solvation

We observed that the solvation environment of the two tryptophan residues remains N-like during N⇌A transition, as measured by fluorescence Stokes shift (Figure S2.2B and Figure 2.1B: inset). The λ_{max} in the N form is 350 nm that shifts to 357 nm upon unfolding the protein in 6 M guanidinium chloride (GdmCl) at pH 7 (U form) (Figure S2.2B and Figure 2.1B: inset). We observed that there is no pH-dependent change in the λ_{max} of the tryptophan residues between pH 7 to 3, both in the absence and the presence of 6 M GdmCl. The λ_{max} in the A form is 350 nm and in the GdmCl-unfolded U form at pH 3 is 357 nm (Figure 2.1B: inset). In addition to the similar λ_{max}, the unfolded forms at pH 7 and pH 3 also have near identical far-UV CD spectrum (Figure S2.2C). Together, these results suggest that tryptophan residues have similar solvation in the N form and the A form and that the unfolded forms at pH 7 and pH 3 are similar.
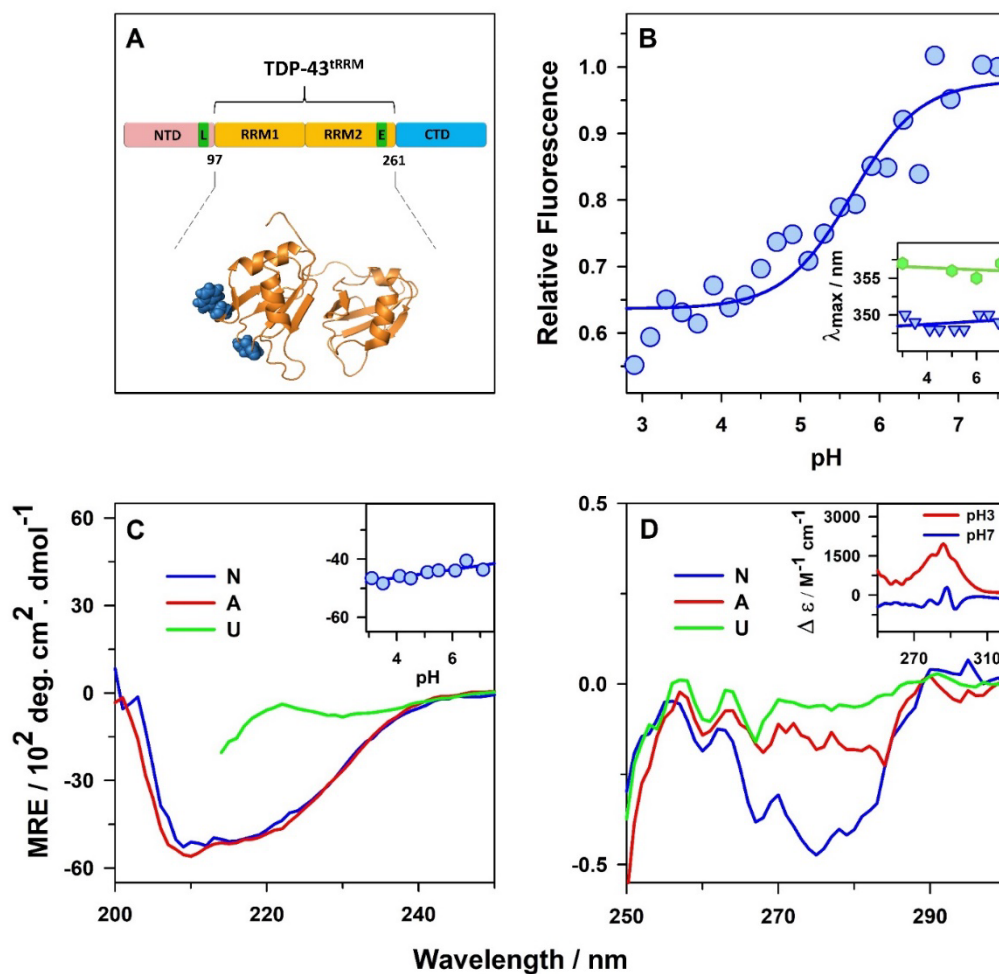
Figure 2.1: *pH-induced formation of the A form of TDP-43^{tRRM} which has N-like hydration and secondary structure but a disrupted tertiary structure. (A) The schematic representation of different domains of human TDP-43 is shown on the top. NTD: N-terminal domain; RRM1 and RRM2: two tandem RNA recognition motifs (TDP-43^{tRRM}); CTD: C-terminal domain. The nuclear localization sequence (L) in NTD and the nuclear export sequence (E) in the RRM2 domain are also shown. TDP-43^{tRRM} (amino acid residues 97-261) is studied in this work and its solution structure (Protein Data Bank entry 4BS2) is shown at the bottom. The side-chains of the two tryptophan residues, W113 and W172, of TDP-43^{tRRM} which are located in the RRM1 domain are shown by blue spheres. (B) The change in the fluorescence emission of the protein at 340 nm is plotted as a function of pH. The solid blue line through the data is a non-linear least square fit to equation 3. (Inset: panel B) The λ_{max} of the tryptophan residues is shown as a function of pH, when incubated in the absence (blue inverted triangles) and the presence of 6 M GdmCl (green hexagons). (C) The far-UV CD and (D) the near-UV CD spectra of TDP-43^{tRRM} are shown in the N form, the A form, and the U form at pH 7. (Inset: panel C) The change in mean residue ellipticity (MRE) at 222 nm is shown as a function of pH. (Inset: panel D) The protein was incubated at pH 7 and pH 3 in the absence and the presence of 6M GdmCl and the difference in the UV absorbance spectrum between the two conditions at pH 7 and pH 3 is shown.*

2.3.3 The protein molecules in the A form resemble a molten globule

We measured the global secondary and tertiary structure of the N form and the A form using far-UV and near-UV CD spectroscopy (Figure 2.1C and Figure 2.1D). The mean residue ellipticity (MRE) at 222 nm, which is a measure of total secondary structure (α-helix+β-sheet) of proteins, in the N form is -4559 ± 403 deg cm^2 dmol^{-1} and the U form is -548 ± 239 deg cm^2 dmol^{-1}. In comparison, the value of MRE at 222 nm in the A form is -5063 ± 437 deg cm^2 dmol^{-1}, that is very similar to the value in the N form. There is almost no pH-dependent change in the secondary structure during N$\rightleftharpoons$A transition (Figure 2.1C: inset). In contrast, we observed that the A form has a highly disrupted tertiary structure compared to the N form. The near-UV CD spectrum of TDP-43^{tRRM} is dominated by the chirality of its thirteen phenylalanine (wiggles in the spectrum in the 255 to 270 nm region), three tyrosine and two tryptophan residues (giving rise to a broad band around 278 nm). The MRE at 278 nm in the N form is -40 deg cm^2 dmol^{-1}, and in the U form is -5 deg cm^2 dmol^{-1}. The MRE at 278 nm in the A form is -16 deg cm^2 dmol^{-1} that is very similar to the value of MRE in the U form. These results indicate that in the A form, the asymmetric packing of the hydrophobic residues in the core is disrupted, but the secondary structure remains N-like. Hence, the structure of the protein molecules in the A form is similar to a molten globule (MG).

2.3.4 Differential molecular arrangement of the aromatic amino acid residues in the N form and the A form

The change in the UV-absorbance spectrum of buried aromatic residues upon unfolding is a sensitive indicator of their molecular environment. The exposure of aromatic residues to the polar aqueous solvent upon partial or complete unfolding leads to a blue shift in the absorbance wavelength and overall changes in the intensity.[34] The spectral shift manifests itself in the difference spectrum between the folded and the unfolded forms as peaks in 285-288 nm region for tyrosine residues and around 291-294 nm for tryptophan residues.[34] We measured the difference in UV-absorbance spectrum of the N form and the U form at pH 7 and the A form and the U form at pH 3 (Figure 2.1D: inset). The difference spectrum at pH 7 has very small peaks at 279, 286 and 292 nm while in contrast, intense peaks are observed for the same wavelengths at pH 3. These results indicate that there is a large change in the extinction coefficient of the tyrosine and tryptophan residues upon unfolding of the A form, compared to the unfolding of the N form. We observed that this difference in the extinction coefficient is primarily due to the different molecular arrangement of aromatic amino acids in the A form

compared to the N form and not due to the change in the polarity of the solvent. The difference in the UV-absorbance spectrum of the unfolded forms at pH 3 and pH 7 is a good measure of the change in the absorbance spectrum due to the change in the polarity of the solvent. We observed that the difference in the absorbance spectrum of the unfolded forms at pH 3 and pH 7 has small and broad peaks (Figure S2.2D). In contrast, the difference in absorbance spectrum of the N form and the A form has sharp and highly intense peaks at 284 nm (with a shoulder at 276 nm) and 292 nm (Figure S2.2D). The highly intense peaks compared to the small and broad peaks of the difference spectrum of the unfolded forms indicate that the difference in the molar extinction coefficient during the unfolding of the N form and the A form is not due to the changes in the polarity of the solvent but majorly due to the different molecular environments of the tyrosine and tryptophan residues in the N form and the A form.

2.3.5 Tryptophan residues have restricted mobility in the A form

Surprisingly, we observed that the local dynamics of the two tryptophan residues is highly restricted in the A form compared to the N form (Figure 2.2A). We measured the pH-dependent changes in the steady-state fluorescence anisotropy of the tryptophan residues in order to understand the changes in their local dynamics during N$\rightleftharpoons$A transition. The mean value of the steady-state anisotropy at 340 nm in the N form is 0.079 ± 0.002 that increases in an apparently sigmoidal fashion as a function of pH, to the value of 0.095 ± 0.002 in the A form. In order to rule out that the increase in anisotropy in the A form is not due to the decrease in the quantum yield of the tryptophan residues (Figure 2.1B and Figure S2.2A), we measured their steady-state anisotropy in the U form. Although the U form has a smaller quantum yield than the N form, the steady-state anisotropy remains considerably smaller (Figure 2.2A: inset) because of the higher degree of flexibility accessible to the tryptophan side-chain in the U form. This result indicates that the anisotropy experiment is primarily reporting on the flexibility of the fluorophore. Hence, the increase in anisotropy during N$\rightleftharpoons$A transition is due to the restricted dynamics of the tryptophan residues in the A form compared to the N form.

2.3.6 The A form is oligomeric

The near-UV CD and absorbance experiments suggest that the packing in the core of the A form is disrupted and hence the side-chain residues are expected to be more flexible and dynamic in the A form, but the steady-state anisotropy experiments suggest that the side-chains of the tryptophan residues are dynamically restricted. The two tryptophan residues of TDP-

43^{tRRM} are located very near to the surface of the protein (Figure 2.1A). We explored whether the restriction in the dynamics of the surface residues in the A form is due to the self-assembly of the protein into an oligomeric structure. We measured the R_H of the N form and the A form using DLS (Figure 2.2B). We observed that the mean R_H of the A form is 8.5 nm which is significantly higher than the mean R_H of the N form (2.3 nm). The width of the distribution of the R_H is larger in the A form compared to the N form, indicating that the formation of the A form is accompanied by an increase in heterogeneity in size of the population of protein molecules. These results indicate that the A form is oligomeric in nature.
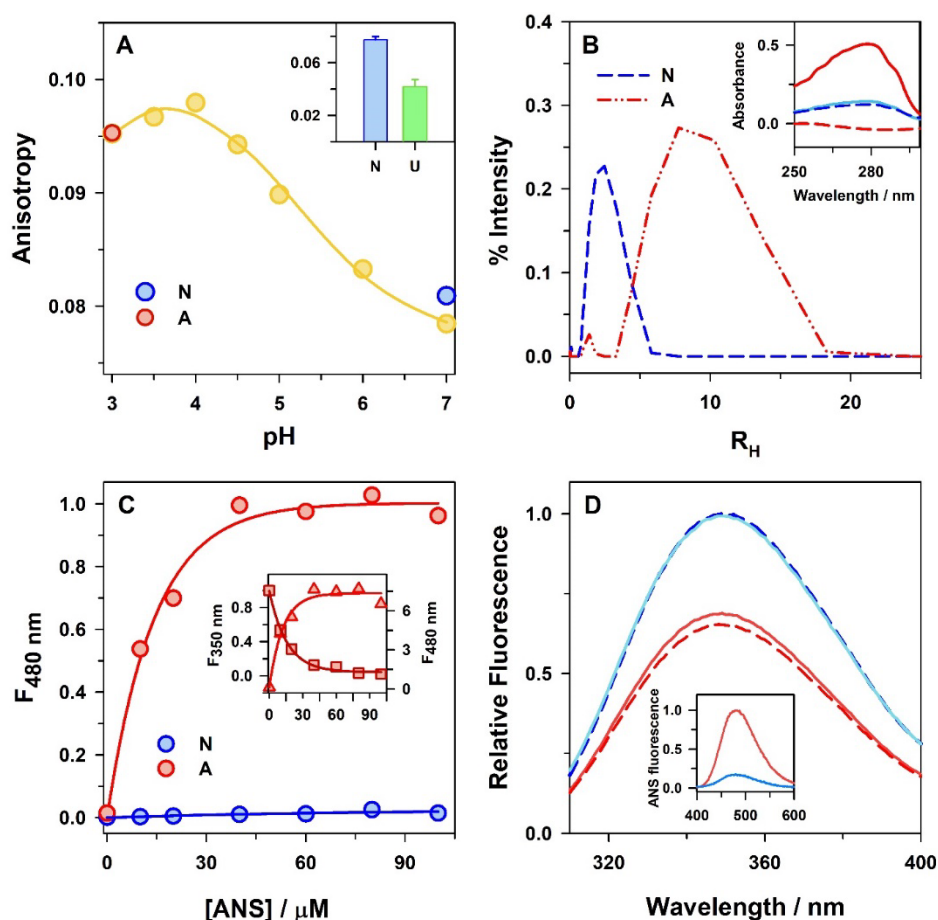
Figure 2.2: *The A form has restricted motion of tryptophan side-chains, a larger size, a single ANS binding site, and forms reversibly from the N form. (A) The change in fluorescence anisotropy of tryptophan residues at 340 nm is shown as a function of pH (yellow circles). The solid yellow line through the data is drawn to guide the eyes. The red and the blue circles, respectively, represent the signal of the A form before and after transferring to the native condition at pH 7. (Inset: panel A) Comparison of the fluorescence anisotropy of tryptophan residues in the N form and U form. The error bar represents the spread in the data from three independent experiments. (B) Distributions of R_H as monitored by DLS. (Inset: panel B) The N form (blue line) and the A form (red line) were spun in a Vivaspin 20 centrifugal filter (MWCO: 100 kDa). The UV absorbance spectrum of the solution inside (solid line) and outside (dashed*

line) the centrifugal filter is shown when the volume of the solution remaining inside the filter was one-fifth of the original volume. (C) Comparison of the binding of ANS dye with the N form and the A form. The relative change in the intensity of ANS fluorescence at 480 nm upon excitation at 380 nm is plotted against the varying concentrations of ANS. (Inset: panel C) The FRET between the tryptophan residues of TDP-43^{tRRM} and the protein bound ANS dye in the A form. The change in fluorescence intensity of tryptophan residues at 350 nm (red square) and ANS dye at 480 nm (red triangle) upon excitation at 295 nm are plotted against the concentration of ANS, according to the left and the right y-axis, respectively. The solid red lines through the data in panel C and its inset are non-linear least square fits to equation 4, and the solid blue line through the N form data in panel C is drawn to guide the eyes. (D) Reversibility of N⇌A transition. Fluorescence spectrum of the N form (blue solid line), formed after a pH 3 to pH 7 jump. The red solid line indicates the fluorescence spectrum of the A form formed after a pH 7 to pH 3 jump. As a control, the blue and red dashed lines represent the fluorescence spectrum of the N form and the A form, respectively. (Inset: panel D) The red and the blue lines, respectively, show the ANS binding of the A form before and after transferring to the native condition at pH 7. The fluorescence emission spectra of ANS is shown after excitation at 380 nm.

In previous studies, data on the correlation between the molecular weight (M_w), the radius of gyration and the R_H have been tabulated on a very large number of globular proteins.[35, 36] We constructed a calibration curve using these previously reported data to estimate the M_w of the N form and the A form from their mean R_H values. The estimated M_w of the N form was around 20 kDa, which is very similar to the mass measured using mass-spectrometry (Figure S2.1B). The M_w of the A form was estimated to be around 800 kDa. These results indicate that the A form is roughly a 40-mer.

We also confirmed the larger size of the A form compared to the N form by passing them through a 100 kDa molecular weight cut off (MWCO) centrifugal filter. The inset of figure 2.2B shows the absorbance of the protein retained and passed by the filter in the N form and the A form. We observed that the N form passed through the filter almost completely. In contrast, no protein molecule passed through the filter in the A form and almost all of the protein molecule is retained by the filter. These results indicate that almost the entire population of the protein molecules have formed the oligomeric A form at pH 3.

2.3.7 The A form has a loosely packed hydrophobic core

We observed that the A form binds to ANS, a dye that is known to bind loosely packed hydrophobic core of MG.[37] The N form does not bind to ANS, but the fluorescence intensity of ANS increases dramatically in the presence of the A form (Figure S2.3A). We measured the

binding of ANS with the A form and the N form as a function of [ANS]. The N form does not bind to ANS even when the [ANS] is as high as ~100 μM (Figure 2.2C). In contrast, the binding of the A form increases hyperbolically with an increase in the [ANS] (Figure 2.2C). In order to further characterize the binding of ANS to the A form, we performed FRET experiments (Figure 2.2C: inset). The absorbance spectrum of ANS overlaps with the fluorescence emission spectrum of tryptophan and they form a FRET pair.[37] We observed that when the tryptophan residues of TDP-43[tRRM] in the A form are excited at 295 nm, their fluorescence emission intensity decreases with an increase in the [ANS] which is accompanied by a concomitant increase in ANS fluorescence (Figure 2.2C: inset). These results suggest that ANS is binding to the A form near the tryptophan residues. We analyzed the binding of ANS with the A form using the following ligand-binding equation:[37]

$$\Delta F(L) = n\Delta F_{max} \frac{[L]}{[L] + K_D}$$
(4)

where, $\Delta F(L)$ is the change in fluorescence on the addition of ANS at a concentration [L], n is the number of identical, non-interacting binding sites for ANS, ΔF_{max} is the change in fluorescence when all the molecules of the A form are bound to ANS, and K_D is the dissociation constant for the binding of ANS to the A form. All the data set in figure 2C and its inset fitted well to equation 4. The mean values of n and K_D obtained from fits are 1.1 and 11 μM, respectively. These results indicate that the A form has a loosely packed hydrophobic core where ANS binds with a dissociation constant of around 11 μM. The observation that the 40-meric A form has only one binding site for ANS indicate that the A form has a symmetrical and spherical micelle-like structure with a centrally located hydrophobic core, as has been observed previously for the low pH intermediate of a small protein barstar.[38]

2.3.8 N⇌A transition is reversible

Although the A form is significantly distinct from the N form with respect to its side chain packing, local dynamics and size, it is interesting to note that the N⇌A transition is fully reversible. The fluorescence intensity of the tryptophan residues of TDP-43[tRRM] is significantly quenched upon formation of the A form from the N form. We observed that the fluorescence intensity of tryptophan residues is fully recovered and becomes similar in value to that of the N form when the A form is transferred to the native conditions at pH 7 (Figure 2.2D). The reversible change in fluorescence during N⇌A transition (during both the forward and the reverse reactions) occurs within the dead-time (~15 s) of the manual-mixing experiment

(Figure S2.4). Upon transfer to the native condition at pH 7, the A form also lose their ability to bind ANS (Figure 2D: inset) and regain the flexible local dynamics of the tryptophan residues (Figure 2.2A), like the N form. These results suggest that N⇌A transition is fast and completely reversible in nature.
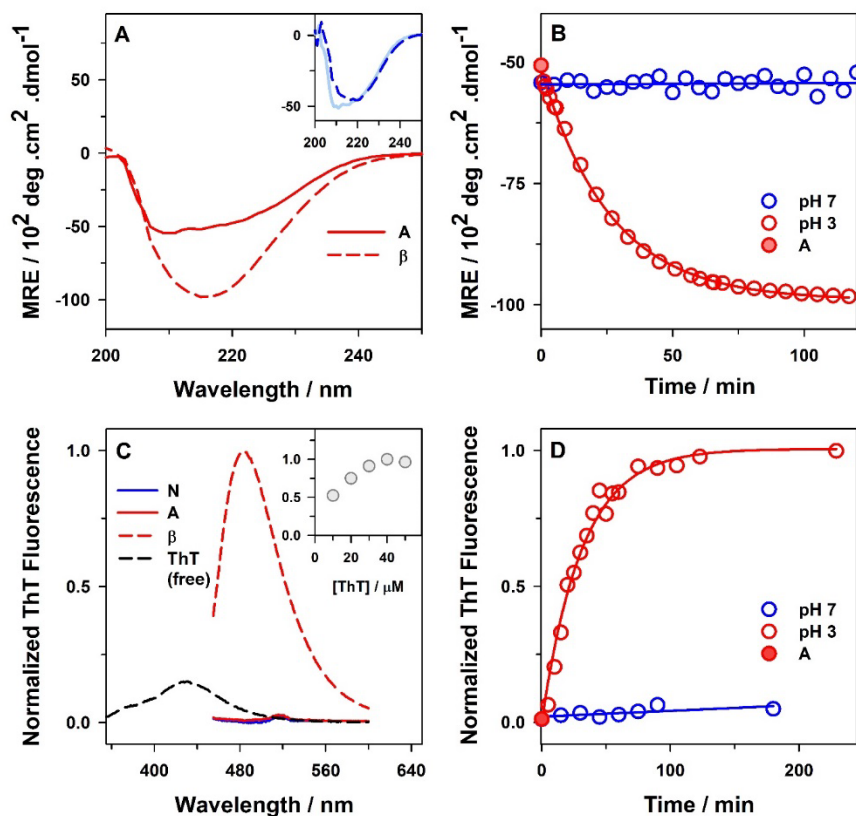
Figure 2.3: *The A form is metastable, amyloidogenic and undergoes an amyloid-like misfolding to a β-sheet rich, β form, upon heating. (A) Far-UV CD spectra of 25 μM of the A form before and after heating (β form) at 40 °C for 2 hours. The inset shows the far-UV CD spectra of 25 μM of the N form before (solid blue line) and after heating (dashed blue line) at 40 °C for 2 hours. (B) Comparison of the kinetics of conformational change of 25 μM of TDP-43^{tRRM} at pH 7 and pH 3 at 40 °C, as monitored by the change in mean residue ellipticity at 216 nm. (C) Changes in the fluorescence emission spectra of ThT dye upon binding to the N form, the A form and the β form. The excitation wavelength in the case of free dye was 340 nm and in all other cases, it was 440 nm. The inset shows the change in the fluorescence intensity of ThT measured at 482 nm in the presence of β form (2 μM) as a function of [ThT]. (D) Comparison of the kinetics of conformational change of 25 μM of TDP-43^{tRRM} at pH 7 and pH 3 at 40 °C, as monitored by the change in ThT fluorescence at 482 nm. In panels C and D, all the data has been normalized to unity to the value of ThT signal of the β form at 482 nm. In all the above panels, the β form is formed after heating 25 μM of the A form at 40 °C for 2 h. In panels B and D, the solid red lines through the data are fits to a single exponential equation and the solid blue lines are drawn to guide the eye.*

2.3.9 The A form is an aggregation-primed intermediate and is highly amyloidogenic

Intriguingly, we observed that the A form is amyloidogenic in nature. The far-UV CD spectrum of the A form is almost identical to the N form (Figure 2.1C) and shows minima at 209 nm and 222 nm, a characteristic signature of α+β protein structure (Figure 2.1A). We observed that upon heating the A form at 40 ºC for two hours, it undergoes a structural change to an amyloid-like conformation which contains predominantly β-sheet rich structure (β form) as evident from the huge dip in the far-UV CD spectrum at 216 nm (Figure 2.3A). The value of MRE at 216 nm increase from -5645 ± 611 deg cm^2 dmol^{-1} to -10365 ± 815 deg cm^2 dmol^{-1} during this structural change. Interestingly, the N form does not transform to the β form upon heating for two hours at 40 ºC (Figure 2.3A: inset). There is only a marginal change in the far-UV CD signal at 210 nm upon heating which might be due to the partial unfolding of the protein as 40 ºC is very close to the beginning of its thermal unfolding transition.[14] Figure 2.3B compares the kinetics of formation of the β form from the A form and the N form at 40 ºC as determined by the measurement of the change in MRE at 216 nm. The transformation in the secondary structure during A⇌β transition occurred in a single exponential phase with a mean time constant of 26 min (Figure 2.3B). There is no observable lag phase. In contrast, for the N form, the MRE at 216 nm remains almost constant as a function of time of heating.

The β form binds to the amyloid staining dye ThT,[39] but not the A form and the N form (Figure 2.3C). The λ_{max} of the free ThT dye (in the absence of any protein) is 440 nm which shifts to 482 nm in the presence of the β form (Figure 2.3C). The shift in λ_{max} was also accompanied by a dramatic increase in the fluorescence emission intensity. These results indicate that the β form strongly binds to ThT. In contrast, the A form and the N form do not show any binding to ThT, indicating the absence of any amyloid-like ordered β–sheet structure. Combining with the results of the far-UV CD experiments (Figure 2.3A and 2.3B), these findings strongly indicate that ordered β–sheet structure, akin to what has been observed in amyloids or soluble precursor protofibrils, is present in the β form but not in the A form and the N form.

We observed that the fluorescence intensity of ThT depends upon the [dye]:[protein] in the assay solution (Figure 2.3C: inset). The inset in figure 2.3C shows the variation of ThT emission intensity as a function of [ThT] when the concentration of the β form was kept fixed at 2 μM. When the dye concentration is 10 μM ([dye]:[protein] is 5:1), the ThT signal is only 50 % of the highest saturating signal indicating that only half of the available binding sites are

occupied by ThT. The ThT emission intensity did not increase and remained constant when the [ThT] is larger than 30 µM (Figure 2.3C: inset), indicating that all the available binding sites are occupied by ThT when the [dye]:[protein] ratio is larger than 15:1. In all the ThT experiments in this study, the concentration of ThT (40 µM) was in 20-fold excess of the protein concentration (2 µM) in the assay solution.

We monitored the kinetics of formation of the β form from the A form at 40 °C using the change in ThT fluorescence intensity (Figure 2.3D). ThT fluorescence measures the elongation of the ordered β–sheet structure during the process of amyloid-like assembly.[39] We observed that the increase in ThT fluorescence occurred in a single exponential phase with a mean time constant of 31 min, without any lag phase (Figure 2.3D). The kinetics and the rate constants of formation of the β form are similar when monitored by ThT fluorescence and the far-UV CD (Figure 2.3D and Figure 2.3B). The absence of the initial lag phase during the formation of the β form suggests that the A form is an aggregation-primed intermediate. No change in ThT fluorescence signal was observed when the N form was heated at 40 °C for two hours, indicating the absence of formation of the ordered β–sheet containing structure.
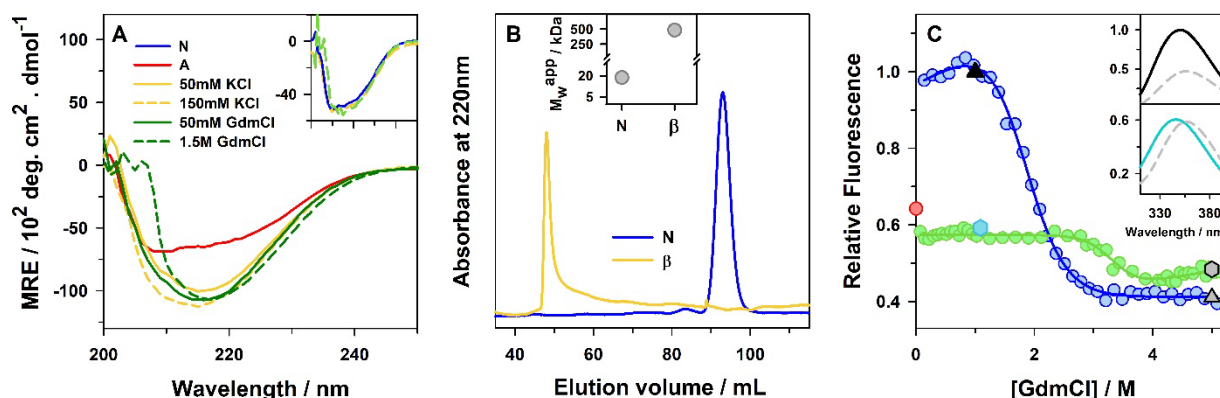
Figure 2.4: *Increase in ionic strength or partial destabilization of the A form leads to the formation of large and highly stable misfolded β form. (A) Panel A and its inset, respectively, show the far-UV CD spectra of TDP-43^{tRRM} at pH 3 and pH 7 in the absence and the presence of KCl and GdmCl. (B) Size exclusion chromatography of the N form and the β form. (Inset: panel B) Comparison of the M_w^{app} of the N form and the β form. (C) GdmCl-induced equilibrium unfolding transitions of the N form (blue circles) and the β form (green circles) monitored by the change in tryptophan fluorescence at 340 nm. The filled red circle represents the signal of the A form. The solid lines through the data in panel C are drawn to guide the eye. The gray and the black triangles, respectively, represent the signal of the U form (at 5 M GdmCl, pH 7) before and after transferring to a buffer containing 1.1 M GdmCl at pH 7. The gray and cyan hexagons, respectively, denote the signal of the U form (at 5 M GdmCl, pH 3)*

before and after transferring to a buffer containing 1.1 M GdmCl at pH 3. (Insets: panel C) The N⇌U and the β⇌U transitions are reversible. The top inset shows fluorescence spectrum of the protein in the U form (at 5 M GdmCl, pH 7) before (gray dashed line) and after (black solid line) transferring to a buffer containing 1.1 M GdmCl at pH 7, where the N form is predominantly populated. The bottom inset shows the fluorescence spectrum of the protein in the U form (at 5 M GdmCl, pH 3) before (gray dashed line) and after (cyan solid line) transferring to a buffer containing 1.1 M GdmCl at pH 3, where the β form is predominantly populated.

2.3.10 The A form is metastable

Apart from the temperature perturbation, the β form is also formed from the A form upon partial destabilization in the presence of a small amount of denaturant or an increase in the ionic strength of the solution. Figure 2.4A shows the far-UV CD spectra of the A form in the presence of two different concentrations of the denaturant, GdmCl (50 mM and 1.5 M GdmCl), and two different concentrations of the salt, KCl (50 mM and 150 mM KCl). The N-like far-UV CD spectrum of the A form changes dramatically in the presence of GdmCl or KCl, with the disappearance of two minima at 209 nm and 222 nm and appearance of a huge minima at 216 nm, indicating the formation of the β–sheet rich structure. The far-UV CD spectra in all the four conditions appear to be very similar to each other and to the heat-induced β form (Figure 2.3A). The value of MRE at 216 nm, in all the four conditions, is around -10,000 deg cm^2 dmol^{-1}, which is similar to the value of the heat-induced β form. These results suggest that the β form formed in the presence of GdmCl or KCl is similar in structure to the heat-induced β form. In contrast, no changes in the far-UV CD spectrum of the N form were observed in the presence of 1.5 M GdmCl or 150 mM KCl (Figure 2.4A: inset). The result that the A form transforms to the β form in the presence of a minute amount of salt or destabilizing agent, suggests that the A form is highly metastable.

2.3.11 The misfolded β form has a large size and higher resistance to chemical denaturation

We compared the M_w^{app} of the β form to that of the N form using size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg gel filtration column (Figure 2.4B; Materials and Methods). The column has a fractionation range of 10 to 600 kDa and a void volume of 39.8 mL. We observed that the elution volume for the N form is 92.7 mL and for the β form is 48 mL (Figure 2.4B). Using the data on elution volumes, we determined the

M_w^{app} (Figure 2.4B: inset) of the N form and the β form. We first constructed a calibration curve between the known elution volumes and M_w^{app} of a few standard biomolecules (Figure S2.5, Materials and Methods). Using the calibration curve (Figure S2.5), we determined the M_w^{app} of the N form to be 18.9 kDa and for the β form to be 485 kDa. The M_w^{app} of the N form as determined from size exclusion chromatography is similar to the mass measured using mass spectrometry (Figure S2.1B). However, the β form elutes at the low-resolution regime of the column (Figure S2.5) and the calculated mass is only a lower limit of its size. Nevertheless, these results indicate that the β form is a β–sheet containing soluble protein aggregate that is much larger in size than the N form.

The A form is highly metastable and transforms to the β form upon either heating or addition of minute amounts of chemical denaturant (Figure 2.3A and Figure 2.4A). The A⇌β transition occurs at a concentration as low as 2 mM of GdmCl, and the midpoint of the transition is 3.5 mM (Figure S2.6). The steep dependence at such a low concentration of GdmCl reflects the weak and transient nature of interactions that govern the self-assembled A form. In contrast, we observed that the β form is highly stable. Figure 2.4C compares the chemical stability of the β form and the N form. We observed that the fluorescence intensity of the tryptophan residues in the N form and the β form changes in a sigmoidal fashion, in an apparently two-state manner, as a function of [GdmCl]. It is unlikely that being oligomeric in nature, all the elements of the structure of the β form will unfold in a single step. Similarly, the N form of TDP-43[tRRM] has also been proposed to unfold *via* two intermediate states, in a previous study.[15] Due to the potential multi-step nature of both the N⇌U and the β⇌U transitions, we used the mid-point of the unfolding transition as a proxy for the chemical stability. The mid-point of the unfolding of the N form is 1.9 M of GdmCl, which is very similar to that observed previously.[15] The mid-point of the unfolding of the β form is 3.3 M, indicating that it is highly resistant to chemical denaturation. Such a high resistance to unfolding reflects the strong nature of interactions in the β form.

It is important to note that both the N⇌U and the β⇌U transitions are reversible in nature (Figure 2.4C and its insets, Figure S2.7). Upon dilution of the denaturant at pH 7, the U form fully regains the fluorescence spectrum of the N form (top inset: Figure 2.4C). Similarly, upon dilution of the denaturant at pH 3, the U form completely regains the fluorescence spectrum (bottom inset: Figure 2.4C) and the far-UV CD spectrum of the β form (Figure S2.7).

Above results indicate that the β form is highly thermodynamically stable and that the structure in the β form is held by much stronger interactions compared to the A form and the N form.

2.3.12 The disordered regions acquire ordered β-sheet structure during A⇌β transition

All the variants of the β form appear to have similar β-sheet content as evident from their similar far-UV CD spectrum (Figure 2.3A and Figure 2.4A). We observed that the values of MRE at 216 nm in the β forms are very high compared to the A form and the N form. This result is highly surprising. The value of MRE at 216 nm for the β-sheet is usually lower than the α-helix and least for the random coil or the disordered segment.[40] Accordingly, if only the α-helical segments of the A form transform to β–sheet during A⇌β transition, the β form would have a lesser value of MRE at 216 nm. However, this is not observed to be the case. It is important to note that a substantial part of the three-dimensional structure of TDP-43^{tRRM} (Figure 2.1A) is disordered.[15] The partially disordered nature of the structure of TDP-43^{tRRM} in both the N form and the A form is also evident in the smaller values of the MRE in their near-UV and far-UV CD spectra (Figure 2.1C and Figure 2.1D) compared to the other structured α+β proteins.[37, 40] TDP-43^{tRRM} has a polyampholyte-like primary sequence that is rich in both positively and negatively charged amino acid residues. It contains 26 basic residues (14 lysine, 9 arginine, 3 histidine) and 25 acidic residues (14 aspartic acid and 13 glutamic acid residues) that accounts for around one-third of the total number of amino acids in its sequence. Such polyampholyte-like sequences have been theoretically predicted[41] and experimentally observed[42] to behave like intrinsically disordered regions (IDRs). If the disordered regions of TDP-43^{tRRM} convert to β–sheet during A⇌β transition, as during the aggregation of intrinsically disordered proteins like α–synuclein and amyloid-β,[43, 44] that will dramatically increase the value of MRE at 216 nm. Hence, our results indicate that the IDRs of TDP-43^{tRRM} have converted to β–sheet in the β forms, in addition to the α-helical segments, if any.

2.3.13 The β form resembles amyloid-like protofibril

The β form of TDP-43^{tRRM} strongly binds to ThT, a dye that is known to bind amyloid-fibril like structures (Figure 2.3C and Figure 2.3D). It is important to note that the ability of the oligomers of different variants and fragments of TDP-43 to bind ThT and their resemblance to amyloids is highly debated.[17, 24, 25, 45] Aggregates formed by some short peptide fragments of

TDP-43 have been shown to bind ThT and resemble amyloids, but many other TDP-43 aggregates do not appear to bind to ThT.[11, 24, 45, 46] In a few cases, for example for the full-length TDP-43, the aggregates that do not bind ThT, do not appear to contain β–sheet rich structure as evident from the far-UV CD spectrum.[47] In contrast, in a recent study[48], some fragments of TDP-43[tRRM] have been shown to form cross-β structure similar to amyloids. We observed that the β forms of TDP-43[tRRM] strongly bind to ThT (Figure 2.3C). The ability of the β forms to bind ThT is strongly dependent on [dye]:[protein] and at least 15 fold excess concentration of the dye is required for complete binding (Figure 2.3C: inset). The β forms have very high β–sheet content (Figure 2.3A and Figure 2.4A). They are oligomeric and much larger in size than the N form (Figure 2.4B) but are soluble. Soluble oligomers with β–sheet rich structure often referred to as protofibrils, are observed frequently as the precursor to amyloid fibrils during their assembly.[49, 50] It is possible that the β-forms observed in this study are similar to amyloid protofibrils. This is an important result because recent studies have shown that soluble oligomers or protofibrils, rather than insoluble fibrils, may be the toxic species in protein aggregation related neurodegenerative diseases.[51, 52] We observed that the kinetics of A⇌β transition studied by far-UV CD and ThT fluorescence (Figure 2.3B and Figure 2.3D) yielded similar rate constants, indicating that the change in the secondary structure (conformational conversion) and elongation of the protofibrillar structure occur simultaneously. However, it is likely that the A⇌β inter-conversion occur in more than one step. A detailed investigation of the step-wise mechanism of the process is currently underway in our laboratory.
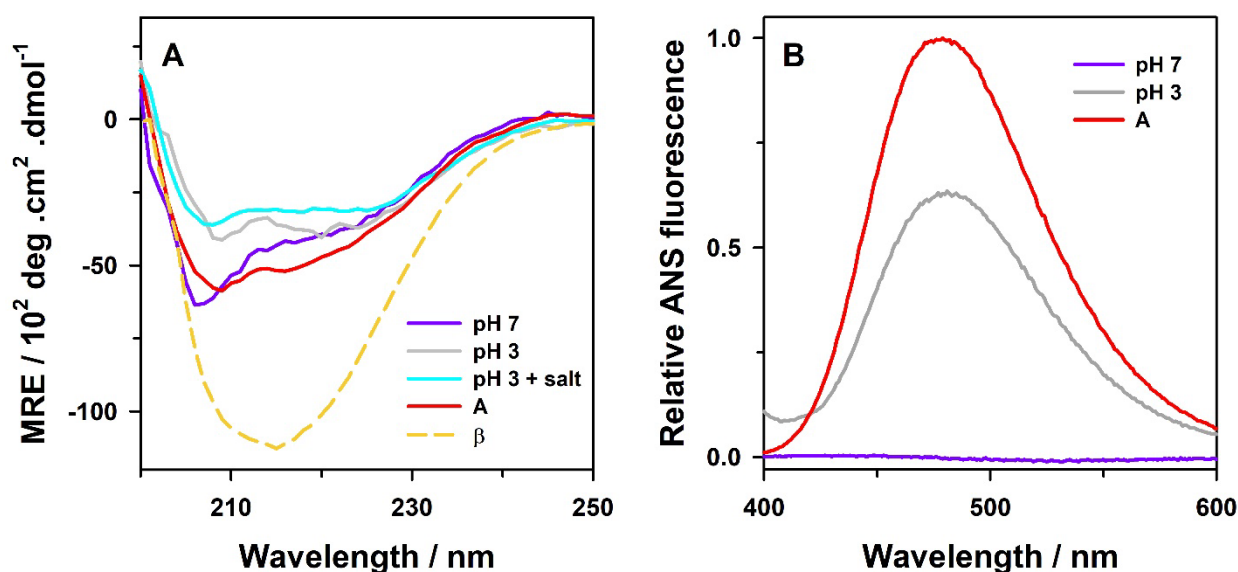
Figure 2.5: *Effect of DNA binding on the formation of the A form and the β form. (A) Far-UV CD spectra of TDP-43^{tRRM} bound to single-stranded (TG)$_6$ DNA at pH 7, pH 3 and pH 3 + 150 mM KCl. For comparison, far-UV CD spectra of the A form and the salt-induced β form (150 mM KCl) at pH 3 formed in the absence of DNA are shown. (B) Comparison of the extent of ANS binding of the protein-DNA complex at pH 7, pH 3 with the ANS binding of the A form formed in the absence of DNA.*

2.3.14 The A form and the β-form are not formed for the DNA-bound TDP-43^{tRRM}

We examined the effect of DNA-binding on the formation of the A form and the β form (Figure 2.5). The protein-DNA complex at pH 7 shows a similar far-UV CD spectrum as the unbound protein (Figure 2.5A and Figure 2.1C). This result indicates that the global secondary structure of the N form remains similar upon DNA binding. The far-UV CD spectrum of the protein-DNA complex at pH 3, however, showed a smaller MRE signal compared to that of the A form (formed in the absence of the DNA), indicating a partially disrupted secondary structure similar to an MG. Interestingly, the far-UV CD spectrum of the protein-DNA complex at pH 3 does not change upon addition of as high as 150 mM KCl. These results imply that the protein-DNA complex does not undergo a disorder-to-order transition as observed for A⇌β conversion for the free protein (Figure 2.5). Hence, the protein-DNA complex is not metastable at pH 3 and the β form is not formed. We observed that the protein-DNA complex at pH 7 does not bind ANS (Figure 2.5B), similar to the free protein (Figure S2.3A, Figure 2.2C). In contrast, the ANS binding capacity of the protein-DNA complex at pH 3 is substantially decreased compared to the A form (Figure 2.5B). The reduced ANS binding of the protein-DNA complex at pH 3 shows that the A form is not completely formed but the protein molecules retain the MG character. Combined together, these results indicate that the A form and the β form are not formed for the DNA-bound TDP-43^{tRRM}.

Above results suggest that the DNA-binding regions of TDP-43^{tRRM} are involved in the self-assembly of the A form and that the formation of the A form is an important step towards the formation of the amyloid-like β form. It is possible that RRM1 domain plays a major role in the formation of the A form because nucleic acids bind strongly to RRM1 domain compared to the RRM2 domain.[16, 18] RRM1 domain is also less stable than RRM2 domain to chemical and thermal denaturation[15, 18], that might make it more prone to participate in the oligomerization process. The high efficiency of FRET between tryptophan residues located in the disordered regions of the RRM1 domain and ANS also suggests that the RRM1 domain is involved in the formation of the self-assembled A form.
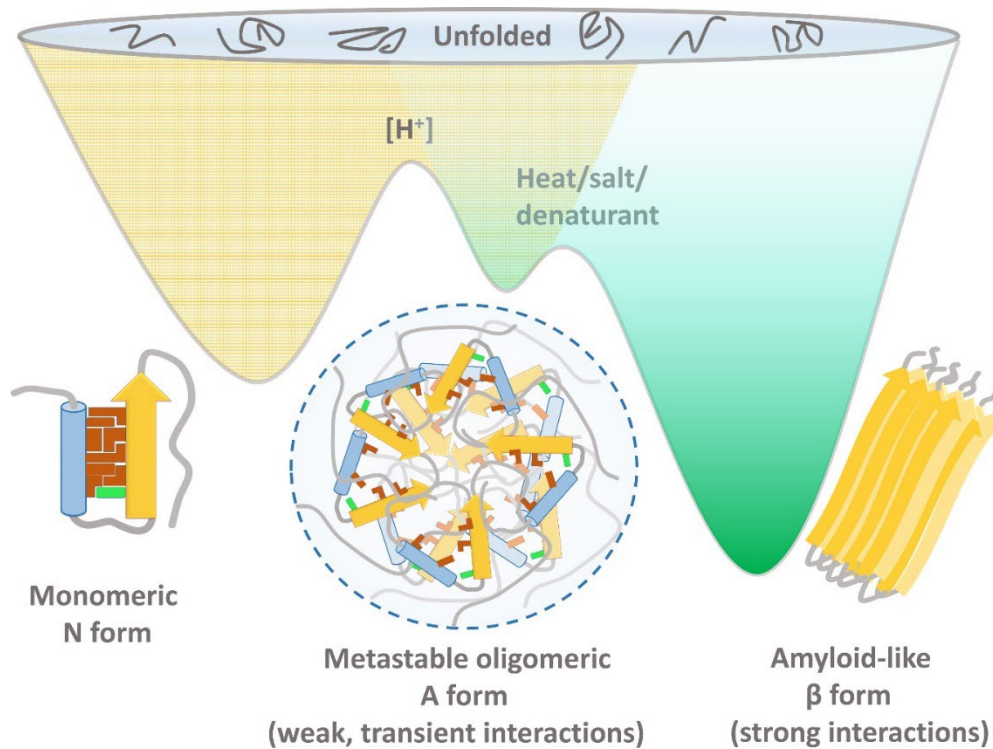
Figure 2.6: *Model energy diagram for the coupling of the folding and misfolding energy landscape of TDP-43^{tRRM}. The schematic also shows how the changes in environmental conditions like pH, co-solutes and temperature modulate the energy landscape. Under native-like conditions at pH 7, the yellow patterned portion of the landscape is operational where the monomeric N form is the most stable state. It remains in a pH-dependent thermodynamic equilibrium with a high-energy nanosized 40-meric A form. The protein molecules in the A form retain the secondary structure of the N form but have loose side-chain packing (brown bars). The N⇌A transition is coupled to the protonation of at least one buried ionizable side-chain residue (green bar) of the protein. The overall structure of the A form is like a micelle with a loosely packed central hydrophobic core in which the intrinsically disordered and aggregation-prone regions of the protein are in close proximity. The loosely packed structure of the A form is held by weak and transient interactions. At pH 3, the protonated A form is highly stabilized, the N form is destabilized, and the N⇌A equilibrium shifts to predominantly populate the A form. Hence, the green portion of the energy landscape becomes operational at pH 3. The metastable A form readily undergoes an amyloid-like conformational conversion and amyloidogenic assembly to a β-sheet rich β form upon partial destabilization (either by addition of GdmCl or heating) or an increase in the ionic strength of the medium. The IDRs of the protein acquire ordered β-sheet structure during A⇌β transition. The structure in the β form is amyloid-like and appears to be held by strong interactions. The A form thus couples the folding and aggregation energy landscape of TDP-43^{tRRM}.*

2.3.15 The metastable A form couples the folding and the misfolding energy landscapes

One of the most important results of this study is that the A form is the common link between the energy landscape of folding, misfolding and aggregation of TDP-43[tRRM] (Figure 2.6). The A form is formed by the self-assembly of the monomeric protein without misfolding into an oligomeric structure consisting of around 40 molecules of TDP-43[tRRM]. The protein molecules in the A form have N-like secondary structure but a disrupted tertiary structure (Figure 2.1 and Figure 2.2). The A form readily misfolds and aggregates into the β–sheet rich β form (Figure 2.3 and Figure 2.4) and the formation of the A form is crucial for the formation of the β form (Figure 2.5). Hence, the A form couples the folding and the aggregation energy landscapes of TDP-43[tRRM].

2.3.16 The N⇌A transition has some similarities to monomer ⇌ droplet transition of proteins

It has been recently reported that proteins containing low complexity IDRs, including the C-terminal and the N-terminal domain of TDP-43,[10, 23] begin to misfold by condensation of protein molecules into nanometer to micrometer-sized protein droplets, akin to the vapour to liquid phase transition of matter.[10, 17, 53-56] The 40-meric A form is only 8.5 nm in size, soluble and not a protein droplet. However, the N⇌A transition appears to have some similarities to monomer ⇌ droplet transition. For example, like the monomer ⇌ droplet transition,[55] a thermodynamic equilibrium exists between the monomeric N form and the self-assembled A form that is highly reversible and dynamic (Figure 2.2 and Figure S2.4). The A form is metastable and primed to undergo misfolding and aggregation reaction to the amyloid-like β form on small environmental perturbation (Figure 2.3 and Figure 2.4), akin to the liquid to solid phase transition of protein droplets.[55, 57, 58] The mechanisms and the reasons behind the formation of protein droplets are poorly understood, but one important determinant is the multivalency of the protein molecules, in terms of the presence of the positive, negative and the hydrophobic centers, like Janus particles,[55] for self-assembly.[59, 60] The protein molecules in the A form are protonated and have molten globular structure that can provide multiple charged and hydrophobic centers required for assembly. This is in line with the recent proposal that globules (like proteins in theta solvent) can self-assemble to form large structures easily compared to the fully structured native proteins.[55, 57, 60] The self-assembled A form binds to a single molecule of ANS near the tryptophan residues (Figure 2.2C and its inset) and has a centrally located loosely packed hydrophobic core similar to a micelle. As tryptophan residues

are located in the IDRs of the RRM1 (Figure 2.1A), the loosely packed hydrophobic core of the A form must involve IDRs of the RRM1, in addition to the possible involvement of the IDRs of RRM2, if any. The IDRs of the A form undergo disorder to order transition during the A⇌β conversion in the presence of cosolutes like KCl and GdmCl (Figure 2.4), similar to the liquid to solid transition of the droplets of other proteins.[57, 58, 61] It appears that the high local concentration of the protein molecules in the A form brings the aggregation-prone and IDRs of TDP-43[tRRM] in close proximity (Figure 2.6) and makes it prime to undergo misfolding and aggregation reaction on small environmental perturbations.

2.3.17 pH modulates the N⇌A transition

The partitioning of TDP-43[tRRM] into the monomeric N form *versus* oligomeric A form is highly dependent on the pH of the environment (Figure 2.1 and Figure 2.2). The population of the A form increases as the pH decreases from 7 to 3 (Figure 2.1 and Figure 2.2). The midpoint of the N⇌A transition is at pH 5.6 (Figure 2.1B). The formation of the A form is coupled to the protonation of at least one buried charged residue (Figure 2.1B). Among the amino acid residues that could titrate near pH 5.6, one histidine (H166) and three aspartic acids (D105, D138, D247) are almost fully buried inside the protein structure. It is possible that one of these residues might have a slightly modified pKa due to the burial and its protonation is coupled to N⇌A transition. It is important to note that many diseased related mutations of TDP-43,[21] like in the case of many variants of prions,[62] perturb the electrostatic interactions of the protein.

2.3.18 Crossing of the major unfolding barrier is not required for misfolding

Although the structure of the A form resembles an MG, we do not observe the population of a monomeric MG during N⇌A transition. Nevertheless, the MG-like structure of the protein molecules in the highly amyloidogenic A form suggests that the misfolding and aggregation of TDP-43[tRRM] could begin from the partially unfolded states having N-like secondary structure and that the complete crossover of the unfolding barrier is not required for misfolding, as has been proposed for a few proteins, including acylphosphatase, human β2 microglobulin, and human transthyretin.[63]

2.3.19 Modulation of the folding and the misfolding energy landscapes by stress-like conditions

The results of the current study only report on the RRM domains of TDP-43 and do not reveal the structural changes that occur in the N-terminal and the C-terminal domains of the protein. Previous studies have shown that both the N-terminal and the disordered C-terminal domains play an important role during the aggregation of TDP-43 and assemble through liquid-liquid phase separation driven droplet formation.[10, 17] A majority of the disease-associated mutations have also been observed in the C-terminal domain of the protein.[21] It is important to note, however, that the genetic mutations contribute only in ~5% of the ALS cases (including familial as well as sporadic) and in the rest ~95 % of the cases the disease is caused by environmental stress.[6, 17] An attractive alternative model of the pathological aggregation of TDP-43 is that environmental stress induces the reversible association of native TDP-43 into granules primarily *via* its TDP-43^{tRRM} region.[26] Upon chronic exposure to stress, the stress-granules associated TDP-43 undergo irreversible aggregation into diseased prone assemblies.[26, 64, 65]. It is important to note that some stress-like conditions like starvation and energy depletion decrease the pH of the cellular environment.[27-29] It has been reported recently that acidification of cellular environment by stress leads to the perturbation of electrostatic interactions and formation of disease-prone large assemblies of proteins like yeast prion Sup35[31] and poly(A)-binding protein.[30] In the case of stress-granule associated poly(A)-binding protein higher-order metastable assemblies form *via* its RRM domain.[30] It has also been suggested that at the molecular level, proteins might sense stress using the protonation-deprotonation of the protein side-chains.[29, 31] In the current study, we observed that the monomeric N form of TDP-43^{tRRM} exists in a pH-dependent reversible thermodynamic equilibrium with a metastable molten globule-like oligomeric A form. The A form is fully populated only below pH 4.5 (Figure 2.1B), in response to the low pH stress-like conditions and transforms to the β form on further stress elevation. Our results shed light on how stress-like conditions might modulate the energy landscape of TDP-43^{tRRM} inside the cell. Our data suggests (see the model in Figure 2.6) that under physiological conditions at neutral pH, the fully folded form of the protein is populated predominantly. Upon environmental stress that perturbs the electrostatic interactions of the protein, the protein self-assembles into the aggregation-primed intermediate A form. Upon prolonged heat or salt stress or in the presence of destabilizing conditions, the equilibrium shifts to populate the disease-relevant β form. Hence, the results of this study reveal a physico-chemical and thermodynamic mechanism of coupling and modulation of the folding and the aggregation energy landscapes of TDP-43^{tRRM}.

2.4 References

(1) Lee, E. B., Lee, V. M. Y., and Trojanowski, J. Q. (2012) Gains or losses: Molecular mechanisms of TDP43-mediated neurodegeneration, *Nat. Rev. Neurosci. 13*, 38-50.

(2) 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.

(3) 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. Y. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis, *Science 314*, 130-133.

(4) 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.

(5) Chen-Plotkin, A. S., Lee, V. M. Y., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease, *Nat. Rev. Neurol. 6*, 211-220.

(6) Ling, S. C., Polymenidou, M., and Cleveland, D. W. (2013) Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis, *Neuron 79*, 416-438.

(7) Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis, *Science 300*, 486-489.

(8) Caughey, B., and Lansbury, P. T. (2003) Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders, *Annu. Rev. Neurosci. 26*, 267-298.

(9) Mompean, M., Romano, V., Pantoja-Uceda, D., Stuani, C., Baralle, F. E., Buratti, E., and Laurents, D. V. (2016) The TDP-43 N-terminal domain structure at high resolution, *FEBS J. 283*, 1242-1260.

(10) Conicella, A. E., Zerze, G. H., Mittal, J., and Fawzi, N. L. (2016) ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain, *Structure 24*, 1537-1549.

(11) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis, *J. Am. Chem. Soc. 132*, 1186-1187.

(12) Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takahashi, R., Kitahara, R., Ito, H., Fujiwara, N., and Urushitani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kda (TDP-43), *J. Biol. Chem. 288*, 14886-14905.

(13) Wang, Y. T., Kuo, P. H., Chiang, C. H., Liang, J. R., Chen, Y. R., Wang, S., Shen, J. C., and Yuan, H. S. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates, *J. Biol. Chem. 288*, 9049-9057.

(14) Austin, J. A., Wright, G. S. A., Watanabe, S., Grossmann, J. G., Antonyuk, S. V., Yamanaka, K., and Hasnain, S. S. (2014) Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life, *Proc. Natl. Acad. Sci. U.S.A. 111*, 4309-4314.

(15) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state, *J. Biol. Chem. 289*, 8264-8276.

(16) Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ule, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43, *Nat. Struct. Mol. Biol. 20*, 1443-1449.

(17) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43, *Biochemistry 56*, 809-823.

(18) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions, *Nucleic Acids Res. 37*, 1799-1808.

(19) Huang, Y. C., Lin, K. F., He, R. Y., Tu, P. H., Koubek, J., Hsu, Y. C., and Huang, J. J. T. (2013) Inhibition of TDP-43 aggregation by nucleic acid binding, *PLoS One 8*, e64002.

(20) Sun, Y. L., Arslan, P. E., Won, A., Yip, C. M., and Chakrabartty, A. (2014) Binding of TDP-43 to the 3 ' UTR of its cognate mRNA enhances its solubility, *Biochemistry 53*, 5885-5894.

(21) Buratti, E. (2015) Functional significance of TDP-43 mutations in disease, *Adv. Genet. 91*, 1-53.

(22) Wobst, H. J., Delsing, L., Brandon, N. J., and Moss, S. J. (2017) Truncation of the TAR DNA-binding protein 43 is not a prerequisite for cytoplasmic relocalization, and is suppressed by caspase inhibition and by introduction of the A90V sequence variant, *PLoS One 12*, e0177181.

(23) Wang, L., Kang, J., Lim, L. Z., Wei, Y. Y., and Song, J. X. (2018) TDP-43 NTD can be induced while CTD is significantly enhanced by ssDNA to undergo liquid-liquid phase separation, *Biochem. Biophys. Res. Commun. 499*, 189-195.

(24) Mompean, M., Hervas, R., Xu, Y. Y., Tran, T. H., Guarnaccia, C., Buratti, E., Baralle, F., Tong, L., Carrion-Vazquez, M., McDermott, A. E., and Laurents, D. V. (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43, *J. Phys. Chem. Lett. 6*, 2608-2615.

(25) Mompean, M., Baralle, M., Buratti, E., and Laurents, D. V. (2016) An amyloid-like pathological conformation of TDP-43 is stabilized by hypercooperative hydrogen bonds, *Front. Mol. Neurosci. 9*, 125.

(26) 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.

(27) Petrovska, I., Nuske, E., Munder, M. C., Kulasegaran, G., Malinovska, L., Kroschwald, S., Richter, D., Fahmy, K., Gibson, K., Verbavatz, J. M., and Alberti, S. (2014) Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation, *eLife 3*, e02409.

(28) Munder, M. C., Midtvedt, D., Franzmann, T., Nuske, E., Otto, O., Herbig, M., Ulbricht, E., Muller, P., Taubenberger, A., Maharana, S., Malinovska, L., Richter, D., Guck, J., Zaburdaev, V., and Alberti, S. (2016) A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy, *eLife 5*, e09347.

(29) Kroschwald, S., and Alberti, S. (2017) Gel or die: Phase separation as a survival strategy, *Cell 168*, 947-948.

(30) Riback, J. A., Katanski, C. D., Kear-Scott, J. L., Pilipenko, E. V., Rojek, A. E., Sosnick, T. R., and Drummond, D. A. (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response, *Cell 168*, 1028-1040.

(31) Franzmann, T. M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A. S., Nuske, E., Richter, D., Baumeister, W., Grill, S. W., Pappu, R. V., Hyman, A. A., and Alberti, S. (2018) Phase separation of a yeast prion protein promotes cellular fitness, *Science 359*, eaao5654.

(32) 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.

(33) 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.

(34) Creighton, T. E. (1997) Protein structure : A practical approach, IRL Press at Oxford University Press, Oxford.

(35) Tyn, M. T., and Gusek, T. W. (1990) Prediction of diffusion coefficients of proteins, *Biotechnol. Bioeng. 35*, 327-338.

(36) Smilgies, D. M., and Folta-Stogniew, E. (2015) Molecular weight-gyration radius relation of globular proteins: A comparison of light scattering, small-angle X-ray scattering and structure-based data, *J. Appl. Crystallogr. 48*, 1604-1606.

(37) Khurana, R., and Udgaonkar, J. B. (1994) Equilibrium unfolding studies of barstar - evidence for an alternative conformation which resembles a molten globule, *Biochemistry 33*, 106-115.

(38) Juneja, J., Bhavesh, N. S., Udgaonkar, J. B., and Hosur, R. V. (2002) NMR identification and characterization of the flexible regions in the 160 kda molten globule-like aggregate of barstar at low pH, *Biochemistry 41*, 9885-9899.

(39) Naiki, H., and Gejyo, F. (1999) Kinetic analysis of amyloid fibril formation, *Meth. Enzymol. 309*, 305-318.

(40) Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc. 1*, 2876-2890.

(41) Das, R. K., and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proc. Natl. Acad. Sci. U.S.A. 110*, 13392-13397.

(42) Muller-Spath, S., Soranno, A., Hirschfeld, V., Hofmann, H., Ruegger, S., Reymond, L., Nettels, D., and Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins, *Proc. Natl. Acad. Sci. U.S.A. 107*, 14609-14614.

(43) Nath, S., Meuvis, J., Hendrix, J., Carl, S. A., and Engelborghs, Y. (2010) Early aggregation steps in α-synuclein as measured by FCS and FRET: Evidence for a contagious conformational change, *Biophys. J. 98*, 1302-1311.

(44) Jha, A., Kumar, M. G., Gopi, H. N., and Paknikar, K. M. (2018) Inhibition of β-amyloid aggregation through a designed β-hairpin peptide, *Langmuir 34*, 1591-1600.

(45) Neumann, M., Kwong, L. K., Sampathu, D. M., Trojanowski, J. Q., and Lee, V. M. (2007) TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: Protein misfolding diseases without amyloidosis, *Arch. Neurol. 64*, 1388-1394.

(46) Saini, A., and Chauhan, V. S. (2014) Self-assembling properties of peptides derived from TDP-43 C-terminal fragment, *Langmuir 30*, 3845-3856.

(47) Fang, Y. S., Tsai, K. J., Chang, Y. J., Kao, P., Woods, R., Kuo, P. H., Wu, C. C., Liao, J. Y., Chou, S. C., Lin, V., Jin, L. W., Yuan, H. S., Cheng, I. H., Tu, P. H., and Chen, Y. R. (2014) Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients, *Nat. Commun. 5*, 4824.

(48) Guenther, E. L., Ge, P., Trinh, H., Sawaya, M. R., Cascio, D., Boyer, D. R., Gonen, T., Zhou, Z. H., and Eisenberg, D. S. (2018) Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2, *Nat. Struct. Mol. Biol. 25*, 311-319.

(49) Ionescu-Zanetti, C., Khurana, R., Gillespie, J. R., Petrick, J. S., Trabachino, L. C., Minert, L. J., Carter, S. A., and Fink, A. L. (1999) Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy, *Proc. Natl. Acad. Sci. U.S.A. 96*, 13175-13179.

(50) Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B., and Teplow, D. B. (2003) Amyloid β -protein (Aβ) assembly: Aβ 40 and Aβ 42 oligomerize through distinct pathways, *Proc. Natl. Acad. Sci. U.S.A 100*, 330-335.

(51) Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campioni, S., Tzitzilonis, C., Soragni, A., Jessberger, S., Mira, H., Consiglio, A., Pham, E., Masliah, E., Gage, F. H., and Riek, R. (2011) In vivo demonstration that α-synuclein oligomers are toxic, *Proc. Natl. Acad. Sci. U.S.A. 108*, 4194-4199.

(52) He, Y., Zheng, M. M., Ma, Y., Han, X. J., Ma, X. Q., Qu, C. Q., and Du, Y. F. (2012) Soluble oligomers and fibrillar species of amyloid β-peptide differentially affect cognitive functions and hippocampal inflammatory response, *Biochem. Biophys. Res. Commun. 429*, 125-130.

(53) Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization, *Cell 163*, 123-133.

(54) Mitrea, D. M., and Kriwacki, R. W. (2016) Phase separation in biology; functional organization of a higher order, *Cell Commun. Signal 14*, 1.

(55) Bergeron-Sandoval, L. P., Safaee, N., and Michnick, S. W. (2016) Mechanisms and consequences of macromolecular phase separation, *Cell 165*, 1067-1079.

(56) Li, H. R., Chiang, W. C., Chou, P. C., Wang, W. J., and Huang, J. R. (2018) TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues, *J. Biol. Chem. 293*, 6090-6098.

(57) Pappu, R. V., Wang, X., Vitalis, A., and Crick, S. L. (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation, *Arch. Biochem. Biophys. 469*, 132-141.

(58) Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation, *Cell 162*, 1066-1077.

(59) Li, P. L., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q. X., Nixon, B. T., and Rosen, M. K. (2012) Phase transitions in the assembly of multivalent signalling proteins, *Nature 483*, 336-340.

(60) Brangwynne, C. P., Tompa, P., and Pappu, R. V. (2015) Polymer physics of intracellular phase transitions, *Nat. Phys. 11*, 899-904.

(61) Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017) Liquid-liquid phase separation of the microtubule-binding repeats of the alzheimer-related protein tau, *Nat. Commun. 8*, 275.

(62) Singh, J., and Udgaonkar, J. B. (2016) Unraveling the molecular mechanism of pH-induced misfolding and oligomerization of the prion protein, *J. Mol. Biol. 428*, 1345-1355.

(63) Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions, *Nat. chem. biol. 5*, 15-22.

(64) Parker, S. J., Meyerowitz, J., James, J. L., Liddell, J. R., Crouch, P. J., Kanninen, K. M., and White, A. R. (2012) Endogenous TDP-43 localized to stress granules can subsequently form protein aggregates, *Neurochem. Int. 60*, 415-424.

(65) Aulas, A., and Vande Velde, C. (2015) Alterations in stress granule dynamics driven by TDP-43 and FUS: A link to pathological inclusions in ALS?, *Front. Cell. Neurosci. 9*, 423.

Supporting Figures
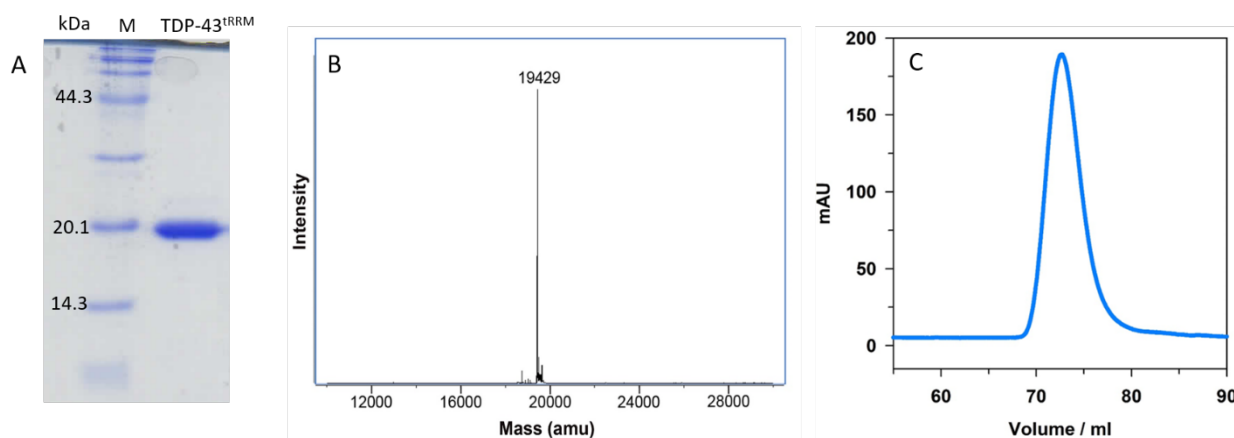
Figure S2.1: *Recombinantly purified TDP-43^{tRRM} is pure and monomeric. Panel A shows TDP-43^{tRRM} protein as a single band near 20.1 kDa marker (M) in 20% SDS-PAGE. Panel B shows the electrospray ionization mass spectrum of TDP-43^{tRRM}. The average molecular mass of the protein is 19,429 Da corresponding to the mass of the His$_6$ tag cleaved TDP-43^{tRRM}. Panel C shows the size exclusion chromatography of the N form on a HiLoad™ 16/600 Superdex 75 pg gel filtration column. The protein elutes at around 72 mL, the mass corresponding to approximately 19 kDa.*
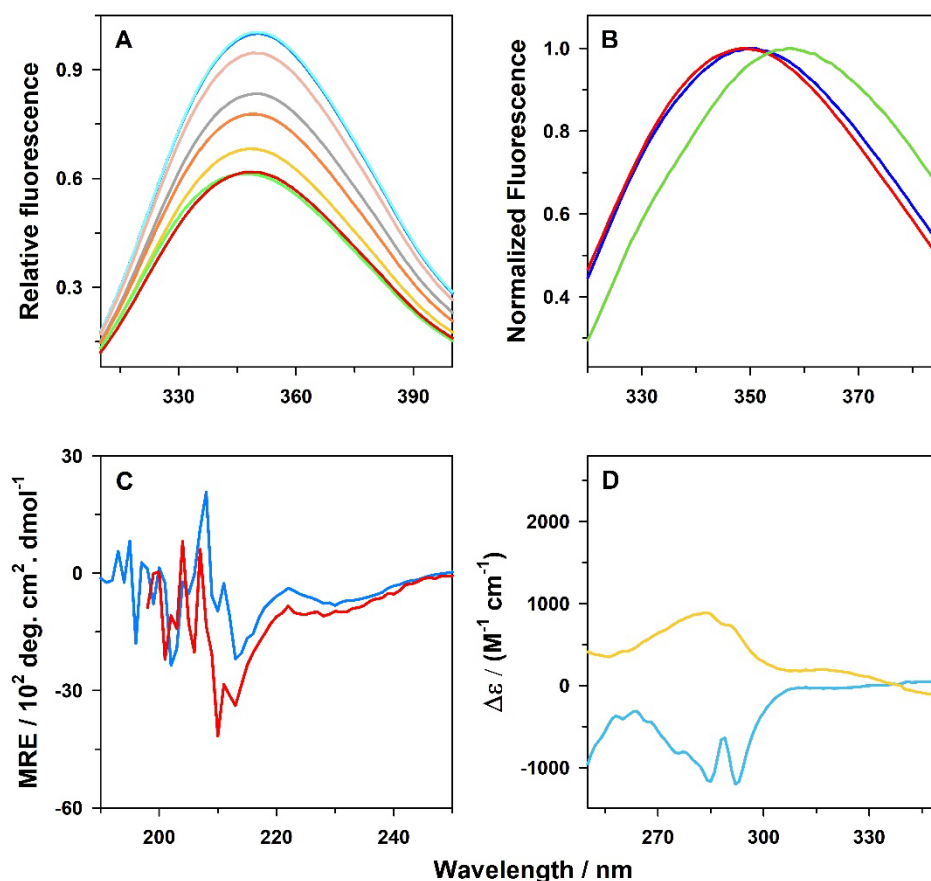
Figure S2.2: *A form has a lower quantum yield, an N-like solvation and a different molecular arrangement of aromatic amino acids compared to the N form. Panel A shows the fluorescence emission spectra of tryptophan residues at different pH: 7.5 (blue), 7.3 (cyan), 6.9 (peach), 6.5 (grey), 5.5 (orange), 4.5 (yellow), 4.1 (green), 3.5 (red). Panel B shows the fluorescence emission spectra of the N form (blue), the A form (red) and the U form (green) at pH 7. The fluorescence values at the wavelength of the maximum emission have been normalized to 1 for comparison. Panel C shows the far-UV CD spectra of the unfolded forms at pH 7 (blue) and pH 3 (red). In panel D, the blue line shows the difference in the absorbance spectrum of the N form and the A form. The yellow line shows the difference in the absorbance spectrum of the unfolded forms at pH 7 and pH 3.*
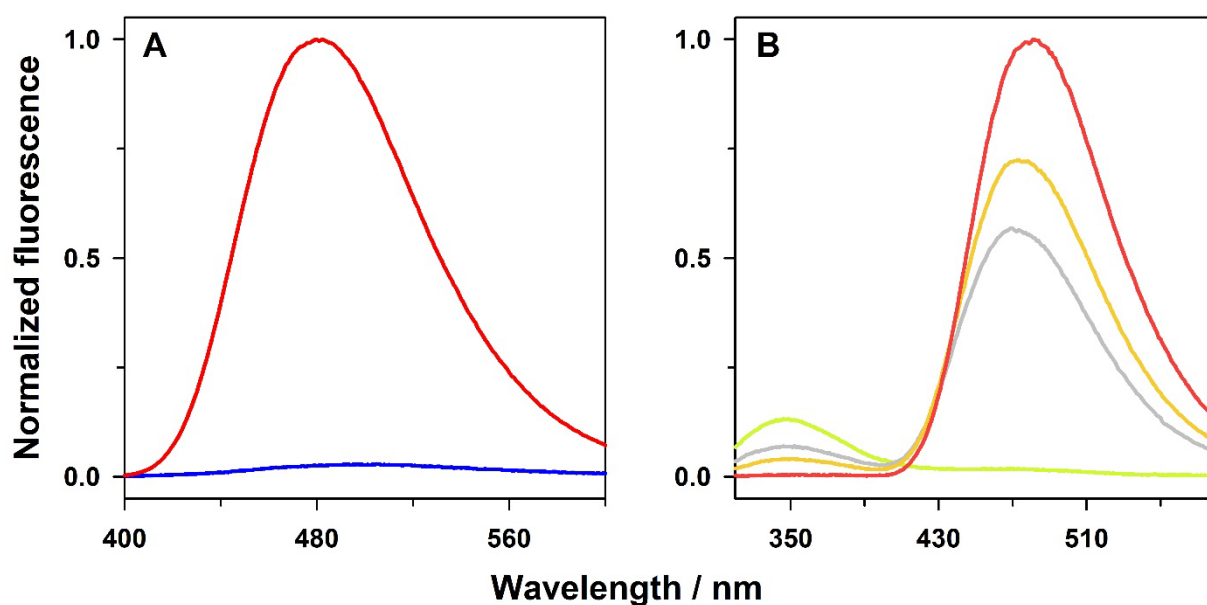
Figure S2.3. *The A form has a loosely packed hydrophobic core and the ANS binding site is located close to the tryptophan residues. Panel A shows the fluorescence emission spectra of ANS upon excitation at 380 nm when bound to the A form (red) as compared to the N form (blue). Panel B shows FRET between the tryptophan residues of TDP-43^{tRRM} and ANS bound to the A form upon excitation at 295 nm. The green line indicates the fluorescence spectrum of the A form in the absence of ANS. The grey, yellow and red lines show the fluorescence spectra of the A form (2 μM) in the presence of 10 μM, 20 μM and 80 μM ANS, respectively.*
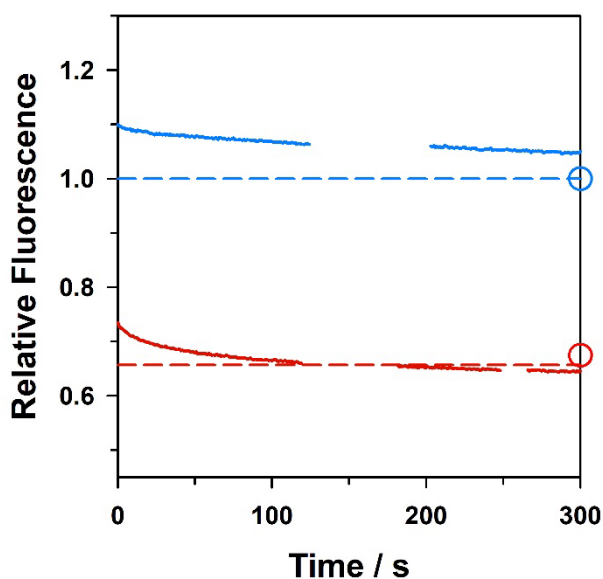
Figure S2.4. *The N⇌A transition is fast and reversible. The dashed lines represent the reference fluorescence signal for the N form (blue) and the A form (red). The solid lines represent the kinetic signal upon transferring the A form to pH 7 (blue) and the N form to pH 3 (red). The circles are equilibrium fluorescence signals of the N form (blue) and the A form (red) taken after 1 h of incubation.*
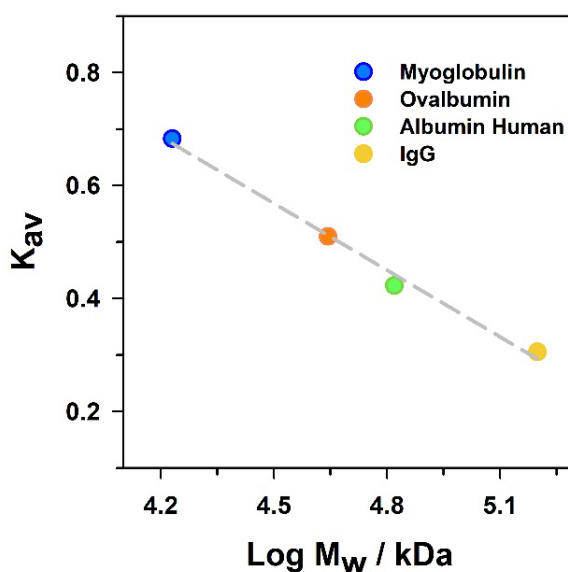
Figure S2.5. *Calibration curve for the determination of the apparent molecular weight (M_w^{app}) of the N form and the β form from the size exclusion chromatography. Log of molecular weight of four standard biomolecules is plotted against their respective partition coefficient (K_{av}). The K_{av} for all the standard biomolecules is calculated from their respective elution volumes noted from the manufacturer provided manual for HiLoad™ 16/600 Superdex 200 pg column using*

equation S2. The dashed line is a linear fit to the data and used to calculate the (M_w^{app}) of the N form and the β form in figure 2.4B.
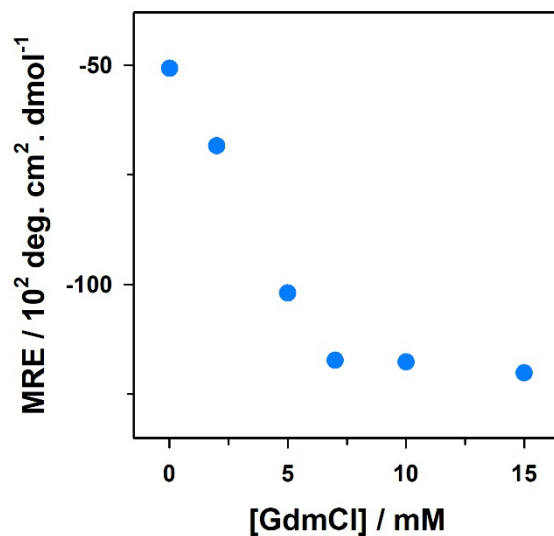
Figure S2.6. *The A ⇌ β transition occurs at a very low concentration of chemical denaturant. The transition is shown between 0 to 15 mM of [GdmCl], as monitored by far-UV CD upon incubation for 5 days.*
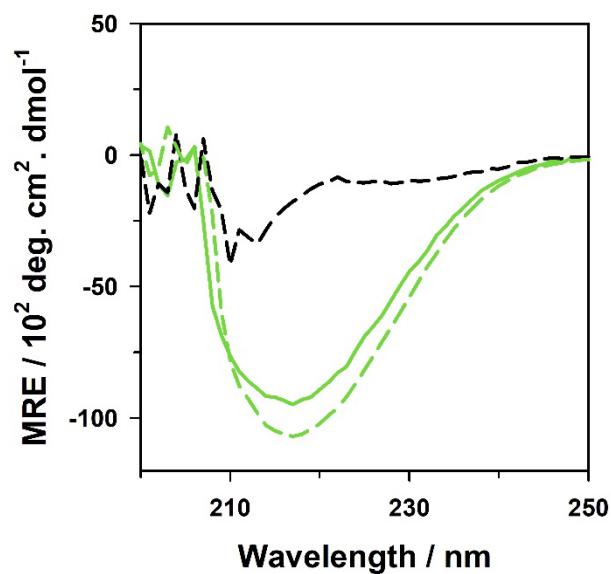
Figure S2.7. *The β⇌U transition is reversible as monitored by far-UV CD. The fluorescence spectrum of the protein in the U form (5 M GdmCl, pH 3) before (black dashed line) and after (green solid line) transferring to a buffer containing 1.1 M GdmCl at pH 3, where the β form is predominantly populated. The green dashed line represents the fluorescence signal of the control β form at 1.1 M GdmCl at pH 3.*

Chapter 3

Early metastable assembly during the stress-induced formation of worm-like amyloid fibrils of nucleic acid binding domains of TDP-43

3.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a lethal disease that leads to progressive impairment of motor neurons resulting in weakening of muscles, paralysis and eventually death of the patient.[1] A defining feature of ALS, in almost all the patients examined till date, is the presence of abnormal aggregates of the TDP-43 (transactive response element DNA binding protein of 43 kDa) protein in the cytosol of motor neurons in the central motor cortex and glial cells.[2-5] The intra-neuronal deposits of TDP-43 are also observed in multiple other TDP-43 proteinopathies.[6, 7] TDP-43 is a highly conserved nucleic acid binding protein that has an important role in many vital cellular processes like synthesis of RNA and proteins and processing of mRNA.[8] It is known that sporadic factors, like chronic environmental stress,[5, 6] are responsible for its aggregation in the majority (90-95%) of diseased conditions but the underlying molecular mechanism of the self-assembly process is poorly understood. It has been postulated that under stress-like conditions TDP-43 participates in the formation of large protein assemblies, such as stress-granules.[9, 10] These assemblies are highly dynamic and metastable and form irreversible aggregates upon chronic stress.[9, 11, 12] However, very little is known about the nature of initial conformational changes that begin the stress-induced self-assembly and misfolding of TDP-43.

The formation of stress granules, which are large assemblies consisting majorly of proteins and translationally stalled mRNA, has been reported in a variety of mortal environmental stress conditions, such as nutrient starvation,[13] osmotic,[14] thermal,[15] and oxidative stress[16] as a mean of stress adaptation. However, the mechanism by which cells detect stress at the molecular level is not well understood. In one of the models, it has been hypothesized that some proteins in the cell could function as biosensors and sense stress by site-specific protein modifications.[13, 17, 18] For example, when the cells are exposed to oxidative insults, the cysteine and the methionine residues of proteins get oxidized.[19] A change in cytosolic pH due to starvation stress is detected by protonation-deprotonation of charged residues.[13, 15, 17, 18, 20] Thermal stress is detected by the partial unfolding of proteins.[13, 15, 18] These site-specific molecular changes trigger the self-assembly of proteins into reversible oligomeric structures with largely native-like (N-like) secondary structure.[13, 17, 18] It has been postulated that coupling of assembly-disassembly of these oligomers with sensitization-desensitization of environmental stress could be a rapid and energy-efficient method of regulating stress.[13] The initial oligomeric assemblies grow further in size[21] and eventually demix and phase separate into a variety of structures with distinct morphologies and material

properties (liquids, gels, filaments, colloids, dynamic fibers, etc.), depending upon the stress.[15, 17, 18] Upon persistent stress, these assemblies irreversibly change their structure to form abnormal solid aggregates,[11, 12] implicated in various neurodegenerative diseases like prion disease,[20] ALS,[22, 23] and FTLD.[23] In order to understand the early stages of these neurodegenerative diseases including ALS, it is critical to understand the nature of initial conformational changes that allow stress-sensing and formation of N-like oligomers. It is also essential to understand how the characteristics of the initial assemblies formed under various stress conditions differ and how they might lead to polymorphism of the diseased condition.

Structurally, TDP-43 protein has four distinct domains (Figure 1). The N-terminal domain is involved in dimerization,[24] has a ubiquitin like fold,[25] and has been recently reported to be undergoing aggregation without getting misfolded.[26] The C-terminal domain is mostly involved in protein-protein interactions[27] and carries a large number of mutations, that are implicated in the diseased condition.[28] However, mutations are responsible for only 5-10% of the diseased cases[5] and thus are marginal contributors towards the diseased state. It is possible that a large number of mutations occur in the C-terminus because it is intrinsically disordered and hence mutations could be better tolerated than in the highly conserved N-terminus or the nucleic acid binding domains. It is important to note that vastly diverse C-terminus sequence are present across the different species in otherwise highly conserved TDP-43 protein.[28, 29] This also points towards the possibility that mutations may simply be disease modifiers but not the major causal factor as the same mutation has been shown to have different pharmacological/pathological effects in different patients.[28] TDP-43 also contains two RNA recognition motif domains (RRM1 and RRM2) in tandem that are linked by a 15-residue natural linker (called together TDP-43[tRRM], hereafter). TDP-43[tRRM] region is highly conserved and functionally important and binds to nucleic acids in a sequence-specific manner to perform a gamut of nucleic acid binding functions in the cell.[6, 30, 31] TDP-43[tRRM] forms abnormal aggregates in various stress-like conditions,[32-34] which is modulated by the presence of nucleic acids.[33-36] TDP-43 co-deposits in the stress-granules with other RNA binding proteins (eg-PABP, TIA, eIF3)[37, 38] possibly *via* its TDP-43[tRRM] region.[35, 39] In this work we show that under low pH conditions mimicking starvation stress TDP-43[tRRM] undergoes a partial unfolding reaction that is linked to the protonation of buried ionizable residues and grows into an oligomeric assembly (called hereafter the 'low pH form' or the 'L form') in which the protein molecules have a unfolded-like (U-like) tertiary structure and solvation environment but an N-like secondary structure. The L form is extremely metastable and swiftly misfolds into a β-

sheet rich amyloid-like aggregated state (the β form) upon a minute change in the ionic strength. The flexible loops of TDP-43^{tRRM} gains order during L to β transition. Our results presented here give a fresh perspective to the initial intermediates formed during the aggregation of TDP-43^{tRRM}.
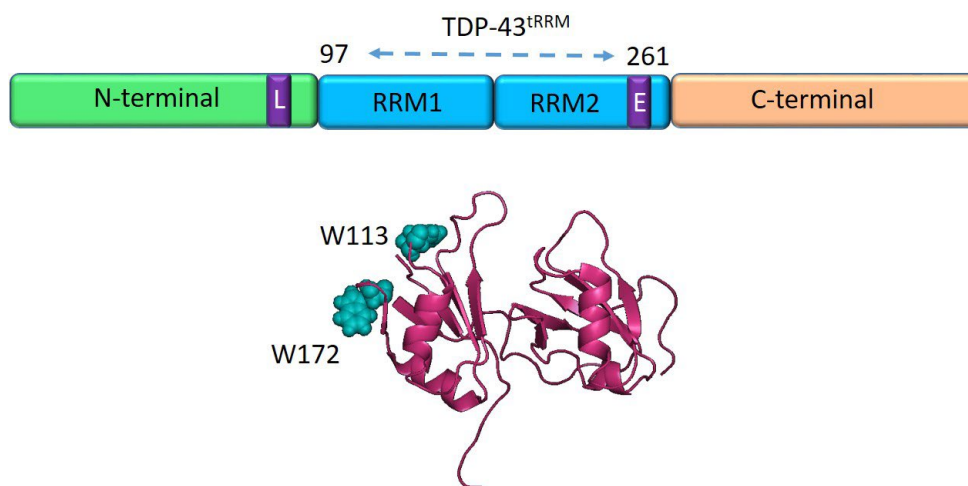
Figure 3.1: *Diagrammatic representation of TDP-43 domains. There are four distinct domains comprising the human TDP-43. The N-terminal domain, two RNA Recognition Motifs (RRM1 and RRM2) referred together as TDP-43^{tRRM} and the C-terminal domain. The letters L and E represent the nuclear localization signal sequence and nuclear export signal sequence, respectively. The bottom part of the figure shows the solution structure of TDP-43^{tRRM} (97-261 amino acid residues) taken from Protein Data Bank entry 4BS2. The tryptophan residues, W113 and W172, situated in the RRM1 domain are represented by blue spheres.*

3.2 Materials and Methods

Protein expression and purification

The expression and the purification protocol of the TDP-43^{tRRM} (UniProtKB/Swiss-Prot entry Q13148) have been described previously.[34] The pure protein was stored in a buffer containing 10 mM potassium phosphate, 150 mM KCl, 1 mM DTT at pH 7.2. SDS PAGE revealed that the protein was highly pure. The molecular weight of the His$_6$ tag cleaved TDP-43^{tRRM} protein as determined by electrospray ionization-mass spectrometry was 19,429 Da.

Buffer, solutions and experimental conditions

All the reagents are of highest purity grade procured from Sisco Research Laboratories (SRL) and Sigma. The buffers used at different pH have been reported previously.[34] Molar

extinction coefficient of 15,470 M^{-1} cm^{-1} at 280 nm was used to determine the protein concentration.

The N form and the L form were prepared by jumping required concentration of protein from a concentrated stock solution of around 600-800 μM protein which is in the storage buffer. The buffer used for the N form contained 20 mM MOPS, 1 mM DTT at pH 7.5. The buffer used for the L form contained 20 mM glycine, 1 mM DTT at pH 2.5. The unfolding buffer contained 20 mM MOPS, 6 M guanidinium chloride (GdmCl) and 1 mM DTT at pH 7.5. The concentration of GdmCl was measured as described earlier.[34, 40] Before the usage, all the buffers were filtered with 0.2 μM filter. The final concentration of residual KCl in all the experiments was around 3 mM – 5 mM unless otherwise mentioned.

pH-induced equilibrium structural transition monitored by fluorescence

For the pH-induced fluorescence measurements, the protein samples (8 μM) were equilibrated at different pH at the room temperature before acquiring fluorescence spectrum on FluoroMax-4® spectrofluorometer (HORIBA Scientific). The excitation wavelength was set to 295 nm and the emission was collected from 310- 400 nm. The excitation slit width was 1 nm while the emission slit width was 8 nm. Each fluorescence spectrum was averaged over three scans. The fluorescence spectrum collected after 5 h and 24 h of incubation remained identical.

The data on the pH dependence of the change in fluorescence of the protein is fitted to a model in which the structural change from pH 7 to pH 2.5 occurs in two steps (N⇌A⇌L) corresponding to the protonation of at least two ionizable groups, given by a modified Henderson-Hasselbalch equation:

$$Y_{obs} = \frac{Y_N + Y_A \, 10^{pH_{m1}-pH}}{1 + 10^{pH_{m1}-pH}} + \frac{Y_L + Y_A \, 10^{pH-pH_{m2}}}{1 + 10^{pH-pH_{m2}}} \qquad (1)$$

where Y$_{obs}$ denotes the observed fluorescence signal for a particular pH value, Y$_N$, Y$_A$ and Y$_L$ are the signals of the N form, the A form (acidic form) and the L form (low pH form), respectively; pH$_{m1}$ and pH$_{m2}$, respectively, are the midpoints of the N to A and A to L transitions.

Dynamic Light Scattering (DLS)

A DynaPro 99 unit (Wyatt) was used for the DLS measurements as reported previously.[34] In brief, all the buffers used for the experiment were filtered three times with a

0.2 µm filter to remove dust particles. The centrifuge tubes and the tips were rinsed with 0.2 µm filtered MilliQ water just before use. The scattering intensity at 90^0 and the autocorrelation function were acquired at the same time by illuminating the samples with a laser of wavelength 829.4 nm. The experimental setting used were as reported previously.[34] DynaLS software (Wyatt) was used to obtain the mean autocorrelation function by taking an average of 50 acquisitions and resolving into Gaussian distributions of hydrodynamic radii (R_H). The mean of all the acquisitions from the cumulant analysis was used to calculate the total light scattering intensity.

Steady-state fluorescence anisotropy measurements

Protein samples (10 µM) were equilibrated in buffers of different pH ranging from pH 7 to pH 2 and steady-state fluorescence anisotropy was measured as reported previously.[34] The tryptophan residues were excited at 280 nm and the emission was collected for 300 s at 340 nm, using an anti-photobleaching setting. The excitation and the emission slit widths were set to 2 nm and 10 nm, respectively. The integration time was set to 1 s. Steady-state fluorescence anisotropy (r) is related to the emission intensity I as follows:

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G} \tag{2}$$

where the subscript *V* (vertical) refers to the position of polarizers in the excitation beam and the subscript *H* (horizontal) to the emission beam. G refers to the instrumental correction factor and is equal to the ratio of I_{HV} to I_{HH}. The measured values of anisotropy remained identical after 5 h and 24 h of incubation.

Circular Dichroism

A Jasco J-815 spectropolarimeter was used for the circular dichroism (CD) measurements and experiments were performed using the same parameters as reported previously.[34] For far-UV CD and near-UV CD measurements, the spectra were collected in the wavelength range of 200-250 nm and 250-300 nm using a quartz cuvette of path length 0.1 cm and 1 cm, respectively. The near-UV CD measurements were carried out at 60 µM (1.1 mg/mL) protein concentration while 25 µM protein (0.48 mg/mL) was used for far-UV CD measurements.

8-Anilino-1-naphthalenesulfonic acid (ANS) fluorescence assay

For the ANS assay, the N form (2 µM) and the L form (2 µM) was incubated for 15 min in dark with different ANS concentration varying from 0 to 100 µM. The ANS binding was determined by excitation of the dye at 380 nm and acquiring the emission spectrum from 400-600 nm. For FRET measurements, the tryptophan residues were excited at 295 nm and the emission was monitored from 320-580 nm. The excitation slit was set to 1 nm and the emission slit was set to 5 nm. The binding affinity of the ANS dye to the L form is determined by fitting the data in figure 4D and its inset to equation 3:

$$\Delta F \ (ANS) = n\Delta F_{max} \frac{[ANS]}{[ANS] + K_D} \tag{3}$$

where $\Delta F(ANS)$ is the fluorescence signal change in the presence of varying concentration of ANS and ΔF_{max} is the maximum fluorescence when all of the ANS is bound to the L form. In the equation, K_D is the dissociation constant and n is the number of binding sites.

The kinetics of formation of the L form was monitored using ANS binding on a Perkin Elmer fluorescence spectrometer LS 55. After transferring the protein at pH 2.5, at different times of the formation of the L form, the binding of the ANS to the protein was measured. The excitation wavelength was set to 295 nm and the emission was measured from 310 to 570 nm. The excitation slit width was 7 nm and the emission slit width was 6.5 nm, respectively.

Salt induced formation of the β form

For the salt-dependent assay, the protein (10 µM) was incubated with varying salt concentration ranging from 2.5 mM to 200 mM, for 24 h. Following the incubation, far-UV CD spectrum of all the samples were collected. All the samples were also analyzed by Thioflavin T (ThT) fluorescence assay as described below.

For the measurement of the kinetics of formation of the β form, 10 µM of the protein was jumped into pH 2.5 buffer containing 150 mM KCl at 25˚C. The kinetics of conformational change was monitored by measuring the ellipticity change at 216 nm by CD. The bandwidth was set to 2 nm and the data integration time to 4 s for all the kinetic experiments. The elongation kinetics was measured by monitoring the ThT fluorescence change as mentioned below.

ThT fluorescence assay

For performing ThT assay, protein sample (1 µM) was added to ThT buffer (50 mM Tris, pH 8 and 20 µM ThT) and the ThT fluorescence was acquired within one minute. ThT dye was excited at 440 nm and the emission was collected from 455- 535 nm. The excitation and the emission slit width was set to 1 nm and 10 nm, respectively. The emission spectra of the buffer having the dye was measured under similar setting. For the L⇌ β transition kinetics, the emission was collected at 482 nm for 20 s. All the data has been subtracted for the background fluorescence contributed by the free dye.

Size Exclusion Chromatography

Size exclusion chromatography (SEC) was performed for the N form and the β form using a HiLoad™ 16/600 Superdex 200 pg column on an AKTA Pure M FPLC system (GE Healthcare). The fractionation range of this column is 10 to 600 kDa and it has a void volume of 39.8 mL. The column was first equilibrated with buffer of respective pH before loading 25 µM of the protein sample on to the column. The flow rate was set to 0.8 mL/min. The buffers contained 150 mM KCl to avoid protein sticking to the column. The apparent molecular weight of the N form and the β form were determined from the known molecular weights of standard biomolecules as described previously.[34]

Transmission electron microscopy (TEM)

The L form and the β form (10 µL of 25 µM) was placed on 300-mesh Formvar carbon-coated copper grid (Electron Microscopy Science) for 5 minutes. After removing off the excess solution, 2% uranyl acetate solution was used to negatively stain the grid for 1.5 minutes. The grid was then washed with filtered MilliQ for a min and allowed to air dry. The samples were examined with transmission electron microscopy (Technai-T20) at an accelerating voltage of 200 kV.

GdmCl-induced equilibrium unfolding transitions

The protein samples (4 µM) were first equilibrated at room temperature for 3 h with different concentration of GdmCl. The fluorescence spectrum of tryptophan residues were measured as described previously,[34] with the excitation slit set to 1.2 nm and the emission slit set to 10 nm. The fluorescence spectrum of the buffer was collected under identical settings. Buffer subtracted data is plotted at 340 nm as a function of GdmCl concentration.

3.3 Results

3.3.1 pH- dependent biphasic structural change monitored by fluorescence

Figure 2A displays the variation in the fluorescence spectrum of the protein (Figure 3.1) as a function of pH. The fluorescence intensity of the protein decreases in two discrete sigmoidal steps upon reducing the pH from 7.5 to 2 (Figure 3.2B). This result implies that the local tertiary structure around the tryptophan side-chains disrupts in two steps and the protein undergoes two structural transitions upon decrease in pH, linked to the protonation of at least two buried ionizable groups (Materials and Methods). The first transition has a midpoint around pH 5.6 and plateaus between pH 4.5 and pH 3. The product of the first transition is referred to as the 'acidic form' or the 'A form' of the protein (Figure 3.2B). The midpoint of the second transition is assessed to be at pH 2.75 and it plateaus around pH 2.5. The product of the second transition is referred to as the 'low pH form' or the 'L form' of the protein (Figure 3.2B). The structural properties of the A form have been described previously.[34] In this study, we dissect the structural characteristics of the L form and report its amyloidogenic nature and metastable behavior.

3.3.2 U-like solvation of the tryptophan in the L form

In addition to the change in the fluorescence intensity, the wavelength of the maximum fluorescence emission (λ_{max}^{em}) of the tryptophan residues also changes in two steps during the N to L transition (Figure 3.2B, inset). The change in the fluorescence Stokes shift reports on the extent of the polar environment around the fluorophore. We observed that the λ_{max}^{em} in the N form is 348 nm that shifts slightly to the blue to 346 nm in the A form,[34] indicating that the solvation environment of the tryptophan residues in the A form is either similar or slightly hydrophobic in comparison with the N form. Upon further reducing the pH, the λ_{max}^{em} changes to 351 nm in the L form reflecting a highly solvated polar environment around the tryptophan residues in the L form. In contrast, the GdmCl induced U form has λ_{max}^{em} of 356 nm across the different pH. These results indicate that the L form is not completely unfolded but the local tryptophan environment remains U-like.
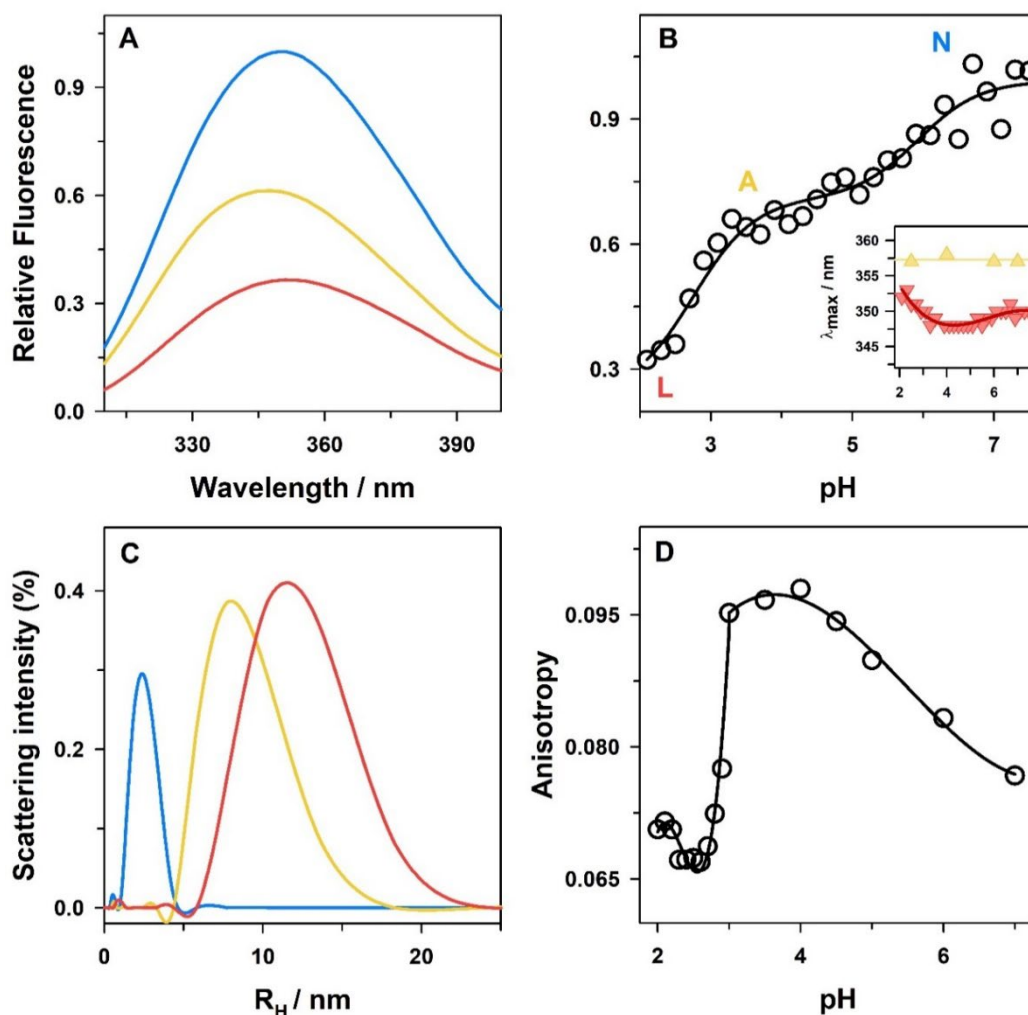
Figure 3.2: *pH-induced monomer to oligomer transition of TDP-43^{tRRM}. A) Relative fluorescence spectrum plotted at pH 7 (blue), pH 4.1 (yellow) and pH 2.5 (red). B) The fluorescence intensity of TDP-43^{tRRM} (black circles) at different pH is plotted at 340 nm. The black line through the data is a non-linear least square fit to equation 1. The relative fluorescence of the N, A and L forms are indicated. The inset of panel B displays the λ_{max}^{em} of the tryptophan residues at different pH in the presence (yellow triangles) and absence (red inverted triangle) of 6 M GdmCl. The red and yellow lines across the data are drawn for visual guidance. C) R$_H$ distributions of the N form (blue), the A form (yellow) and the L form (red) as monitored by DLS. D) The value of fluorescence anisotropy (tryptophan residues) at 340 nm is plotted at different pH (black circles). The black line across the data is shown for visual guidance.*

3.3.3 The N⇌L transition is a monomer ⇌ oligomer transition

The disrupted packing and the U-like solvation environment around the tryptophan residues suggested a partially folded structure for the L form. However, the hydrodynamic measurements using DLS revealed that the L form is an oligomer in addition to being partially folded. The N form is monomeric in nature,[34] and has a R$_H$ of around 2.3 nm (Figure 3.2C).

The R_H of the L form was observed to be around 12 nm. Intriguingly, the N to L transition is accompanied with an increase in the width of the distribution that suggests an increased heterogeneity of the population during the transition. The L form with a mean R_H of 12 nm is a higher sized species as compared to the A form whose R_H is 8.5 nm.[34] These results indicate that the transition from the N form to the L form occurs along with an increased self-association of partially folded molecules resulting in the formation of an intermediate with a higher hydrodynamic radius.

3.3.4 Change in the local dynamics of tryptophan side-chains during N ⇌ L transition

Interestingly, the local dynamics of the tryptophan residues changes in a two-step manner during the monomer ⇌ oligomer transition. We observed that the mean value of the steady-state fluorescence anisotropy increases from 0.0767 to 0.0952 as we moved from the N form to the A form, indicating an increase in the rigidity of the tryptophan residues in the A form.[34] However, the A ⇌ L transition shows a very steep decrease in the anisotropy value, that changes to 0.0674 in the L form. A low value of anisotropy reflects an unfolded-like flexible environment around the tryptophan residues in the L form,[34] which is in line with the U-like solvation for the tryptophan residues as seen by the λ_{max}^{em} (Figure 3.2B, inset). Increased dynamics might also afford multiple relative orientations to the tryptophan residues, resulting in lower anisotropy value. These results indicate that TDP-43^{tRRM} has a completely different structural organization in the L form, the A form and the N form.

3.3.5 DLS monitored kinetics of formation of the oligomeric L form

The L form is oligomeric in nature and has a larger R_H than the monomeric N form (Figure 3.2C). Figure 3A shows the temporal evolution of the distribution of R_H of 25 μM protein at room temperature as obtained from DLS experiments during the formation of the L form. The mean and the width of the distribution of R_H increases progressively with time during the N⇌L transition. This result indicates that both the size and heterogeneity of the protein molecules increases during the formation of the L form. Figure 3.3B shows the increase in total light scattering intensity with time during the formation of the L form. The increase in scattering intensity occurs in two steps. The increase is slow at the beginning with a prominent rise around 5 h. The scattering intensity achieves a plateau after 15 h. The solution of the L form remains considerably clear during the entire transition and we do not see species higher

than ~30 nm anytime during the formation of the L form (Figure S3.1, Supporting Information (SI)).

3.3.6 The molten globular nature of protein molecules in the L form
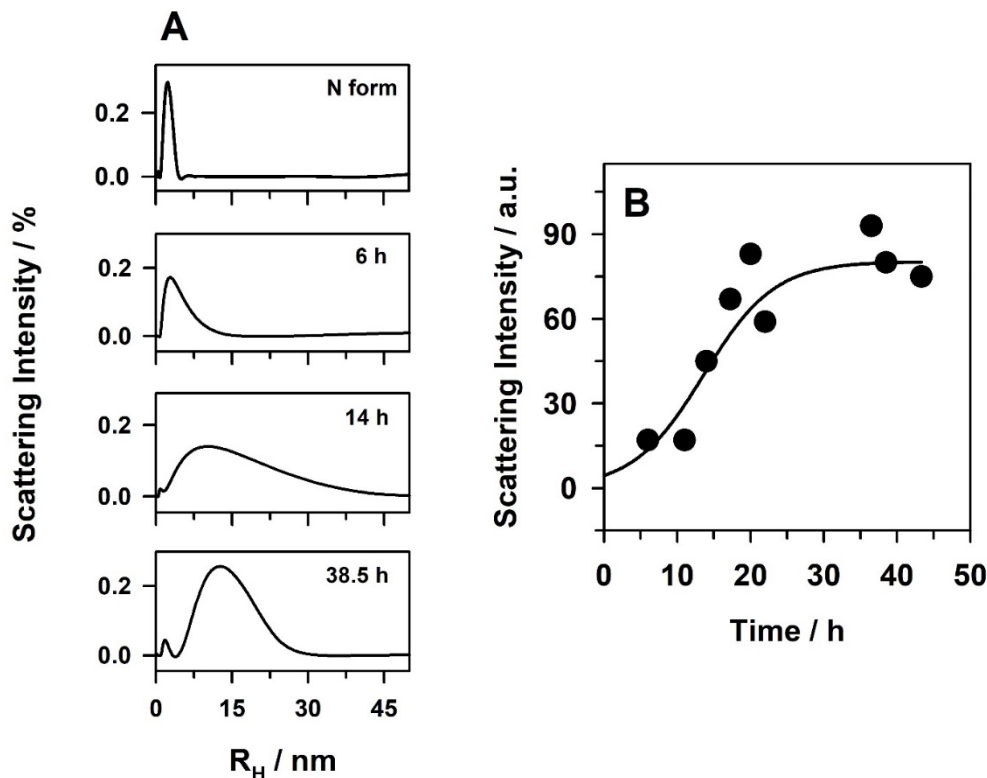
Figure 3.3: *Kinetics of formation of L form as measured by DLS. (A) The DLS distributions of the N form and at three different times after transferring the protein at pH 2.5 at the room temperature, are shown. (B) The change in the intensity of the scattered light (black circles) is plotted with respect to time during the formation of the L form. The black solid line across the data is shown for visual guidance. In both the panels, the concentration of the protein and KCl was 25 µM and 3 mM, respectively.*

We next used far-UV and near-UV CD spectroscopy for analyzing the global secondary and tertiary structure of the protein, respectively. Figure 4A compares the far-UV CD spectra of the L form, the N form and the U form of the protein. The far-UV CD spectrum of the L form is very similar to the N form reflecting an intact secondary structure of the protein molecules in the L form. The mean residue ellipticity (MRE) at 222 nm shows only a small change across the pH range suggesting a very little change in the secondary structure of the protein during the N ⇌ L conversion (Figure 3.4A, inset). The MRE of the L form at 222 nm is -4961 ± 124 which is comparable to the MRE value of the N form (-4664 ± 156). Figure 4B compares the near-UV CD spectrum of the L form, the N form, and the U form. The N form

shows weak but distinct absorbance in the 270 – 285 nm region due to the asymmetric packing of its three tyrosine (Y123, Y155, and Y214) and two tryptophan residues in the protein structure. In the U form, the protein shows near-zero absorbance in the near-UV CD spectrum due to the loss of asymmetric packing of aromatic residues. The near-UV CD spectrum of the L form is very similar to that of the U form. These results suggests that the global tertiary structure of the protein molecules in the L form is dramatically disrupted (Figure 4B), but the secondary structure is intact (Figure 3.4A). Thus the protein molecules that make up the L form are molten globular in structure.

3.3.7 The L form has exposed hydrophobic patches

The hydrophobic dye, ANS shows a huge increase in the fluorescence upon binding to the L form, indicating that the L form contains loosely packed solvent-exposed hydrophobic patch. In contrast, the ANS dye does not show any fluorescence in the presence of the N form, indicating that the N form fails to bind to ANS (Figure 3.4C). We observed that there is a transfer of energy from tryptophan residues of the protein to ANS upon photo-excitation of tryptophan. The inset of figure 4C shows that when the tryptophan is excited at 295 nm in the presence of ANS, there is a decrease in tryptophan fluorescence emission with a simultaneous increase in the ANS fluorescence. The result that the tryptophan and ANS act as a FRET pair indicates that the tryptophan residues are located close to the hydrophobic patch.
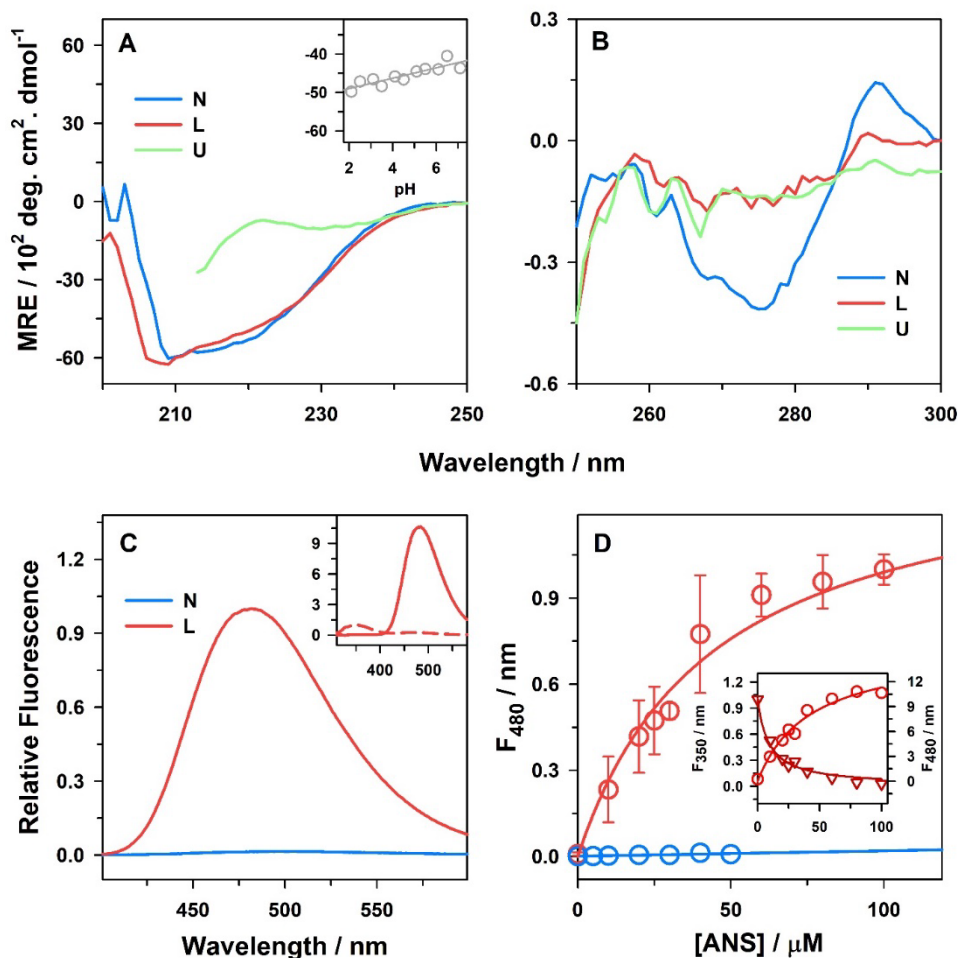
Figure 3.4: *The L form is made up of protein molecules that are molten globular in nature with exposed hydrophobic patches. A) Far-UV CD and, B) Near-UV CD spectra of the N form, the L form, and the U form. The MRE at 222 nm at different pH is plotted in the inset of panel A. C) Fluorescence spectrum of the ANS dye in the presence of the N form and the L form. The tryptophan residues in the L form exhibit FRET with protein-bound ANS. The inset in panel C displays the fluorescence spectrum of the L form in the absence and presence of ANS after excitation at 295 nm. The fluorescence of tryptophan is quenched in the presence of ANS but there is a huge increase in the fluorescence of ANS indicating energy transfer. D) Measurement of the ANS binding to the N form (blue circles) and the L form (red circles). Upon exciting the ANS dye at 380 nm, the relative change in the fluorescence intensity is plotted at 480 nm as a function of ANS concentration. FRET between tryptophan residues and the bound ANS dye in the L form is shown in the inset of panel D. The left y-axis shows the tryptophan fluorescence at 350 nm (inverted red triangles) while the right y-axis shows the ANS fluorescence at 480 nm (red circles) upon excitation at 295 nm in presence of increasing concentration of ANS. The solid red lines across the data in panel D and its inset are non-linear least square fits to equation 3, and the solid blue line across the N form data is shown for visual guidance.*

Figure 4D shows the changes in the fluorescence of ANS when the L form and the N form are incubated with increasing concentration of ANS. The inset of figure 4D shows the

reduction in the fluorescence emission of tryptophan residues and the gain in the emission of the ANS dye upon excitation at 295 nm in the L form with increasing concentration of ANS. We estimated the binding affinity of the ANS dye to the L form by fitting the figure 3.4D data and its inset to a ligand-binding model (Materials and Methods). The value of K_D for the binding of ANS to the L form is 40 µM. The binding of ANS dye to the L form is weak relative to the binding of ANS to the A form (K_D =11 µM).[34] The weak binding capacity of the hydrophobic dye hints at the possibility of solvation of the hydrophobic patches in the L form.

3.3.8 A conformational opening reaction precedes oligomerization

The results of the DLS data shows that the oligomerization of the protein during the formation of the L form occurs in two steps (Figure 3.3). In order to understand the nature of the conformational changes during the formation of the L form, we monitored the L form kinetics using far-UV CD and ANS binding (Figure 3.5). We observed that the secondary structure remains native-like during the formation of the L form (Figure 3.5A). We also examined the kinetics of formation of the L form using ANS binding and observed that the change during the transition occurs in two steps (Figure 3.5B, 3.5C and 3.5D). The major decrease in the tryptophan emission and the consequent increase in the ANS fluorescence occurs very fast during the first step of the formation of the L form (Figure 3.5B, 3.5C and 3.5D). Further, there is a very small change in the fluorescence during the second step of the formation of the L form on a slower timescale (Figure 3.5B, 3.5C and 3.5D). These experiments suggest that when the N form is jumped to low pH, it first undergoes a very fast transition into a partially unfolded conformation in which the hydrophobic patches are exposed but the secondary structure remains N-like. Combined with the DLS data (Figure 3.3), these results indicate that the partially unfolded conformation oligomerizes in the second slower step. Hence, a partial unfolding reaction and exposure of hydrophobic patches precedes the process of oligomerization during the formation of the L form. The small change in the ANS

fluorescence in the second slower step might be due to the creation of small amount of hydrophobic patches during the process of oligomerization.
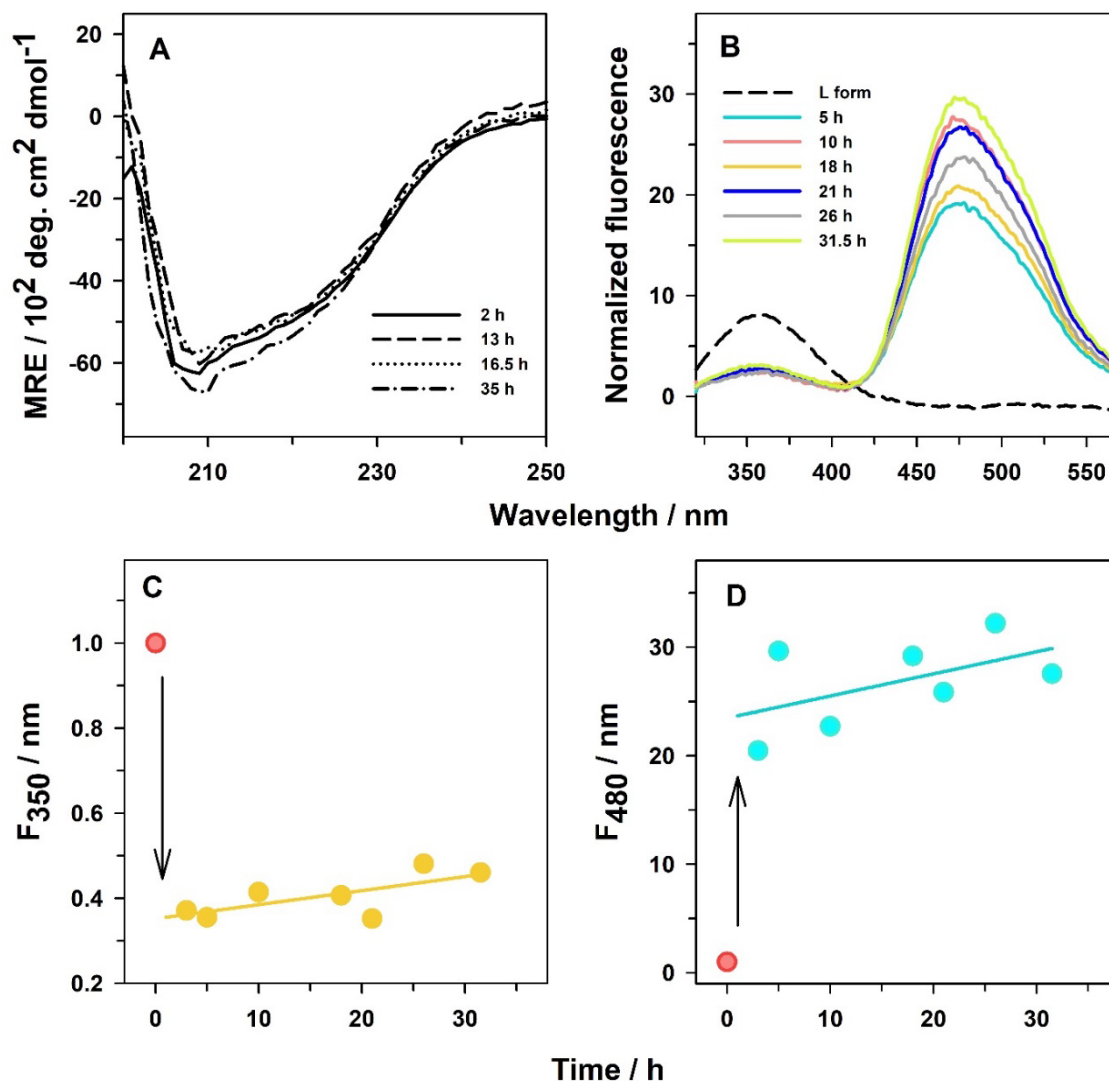
Figure 3.5: *Kinetics of the formation of the L form as measured by far-UV CD and ANS binding. The formation of the L form is initiated by transferring the protein at pH 2.5. A) The far-UV CD spectrum of the protein at different times during the formation of L form. B) The binding of the ANS to the protein is monitored at different times during the formation of L form and the panel B shows the fluorescence spectrum upon excitation at 295 nm. As a control, fluorescence spectrum of the L form in absence of ANS is also plotted (black dashed line). C) The relative change in the tryptophan fluorescence intensity (yellow circles) at 350 nm upon excitation at 295 nm in the presence of ANS is plotted at different times during the formation of L form. D) The relative change in the ANS fluorescence intensity (blue circles) at 480 nm upon excitation at 295 nm is plotted at different times during the formation of L form. In panels C and D, the red circle indicates the fluorescence emission of tryptophan residues at 350 nm and ANS emission at 480 nm upon exciting at 295 nm in absence of ANS, respectively, at pH 2.5. The yellow and the blue lines in panel C and D are shown to guide the eye.*

3.3.9 The structural organization of the L form is highly sensitive to the ionic strength of the solution

Interestingly, we observed that the structure of the L form is very sensitive to the concentration of salt in the solution. The L form is an oligomer (Figure 3.2C) made up of protein molecules that have a broken tertiary structure but an intact N-like secondary structure (Figure 3.4). Figure 3.6A and 3.6B show that the L form retains its N-like secondary structure only when the concentration of KCl in the solution is below 5-6 mM. The L form also does not bind to the amyloid-staining dye ThT (Figure 3.6A, inset), indicating it does not contain an amyloid-like structure. Upon increasing the concentration of salt there is an increase in the value of MRE in the far-UV CD spectrum accompanied by a switch in the signal maximum to 216 nm (Figure 3.6A). This result indicates that at high salt concentration the L form transforms to a state that is highly rich in β sheet structure (β form). The structural transformation of the L form to the β form is complete by 15 mM of KCl, above which the β form is predominantly populated (Figure 3.6A and 3.6B). The β form also binds strongly to ThT dye (Figure 3.6A, inset and Figure 3.6B) revealing the presence of cross β sheet structure similar to amyloids. The structural transition of the L form to the β form is quite steep and the mid-point of the transition, monitored by both far-UV CD and ThT fluorescence, occurs at around 10 mM of KCl (Figure 3.6B). The structural transformation of the L form in the presence of very low concentration of salt indicates the weak nature of the force that governs the structure of the L form. Interestingly, the conformational conversion of the L form to the β form is accompanied by an increased size, as monitored by DLS (Figure 3.6C). The mean value of R_H for the L form and the β form is 12 nm and 36 nm, respectively. The width of the DLS distribution is higher for the β form compared to the L form, indicating that the β form is more heterogeneous than the L form. These results suggest that the oligomerization status and the structural organization of the L form is highly sensitive to the ionic strength of the solution.
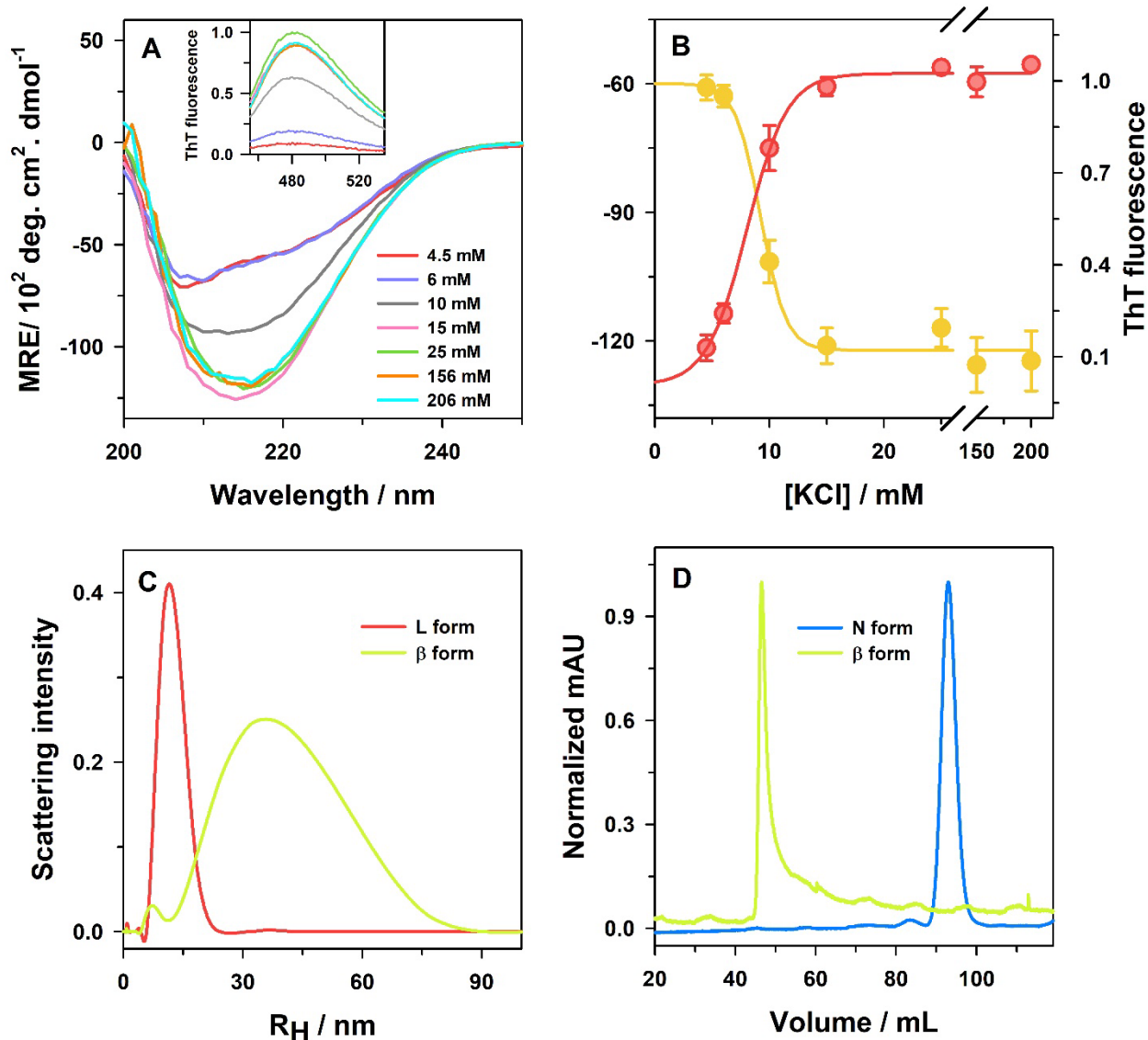
Figure 3.6: *The L form is extremely metastable that undergoes conformational conversion coupled with oligomerization in the presence of salt to form amyloid-like misfolded β form. A) Far-UV CD spectra of the protein at pH 2.5 in the presence of an increasing concentration of KCl. The inset of panel A shows ThT fluorescence spectra of corresponding samples when excited at 440 nm. B) The yellow circles show the change in MRE as a function of salt concentration (left-hand y-axis) at 216 nm while the red circles show the change in the fluorescence of ThT at 482 nm (right-hand y-axis). The solid lines across the data in panel B are shown to guide the eyes. C) DLS distributions of the L form and the β form. D) Size of the N form and the β form as monitored by SEC. The data for the N form has been taken from a previous study.[34]*

3.3.10 The β form is a higher sized species

We estimated the molecular weight of the β form using SEC. Size exclusion chromatogram of the L form could not be obtained as the protein sticks to the column at low salt concentration (less than 5 mM). The N form and the β form were passed through HiLoad 200pg Superdex 16/600 column. The N form eluted at 92.8 mL while the elution of β form was found to be at 46.7 mL (Figure 3.6D). A small trail is seen for the β form which implies the presence of heterogeneous nature of the soluble oligomers. A decrease in the elution volume for the β form reflects a higher hydrodynamic radii and apparent molecular weight as compared to the N form. The apparent molecular weight of the N form was calculated to be 18.8 kDa[34] (consistent with a monomer) which increased considerably for the β form and was found to be ~530 kDa. The elution volume of the β form (46.7 mL), however, is near to the void volume (39.8 mL) of the column. Hence, the determined molecular weight is only a lower limit and the molecular weight of the β form is expected to be higher than 530 kDa, signifying that the β form is at least a 25-mer.

3.3.11 The β form has a morphology similar to worm-like amyloid fibrils

We examined the external morphology of the L form and the β form using TEM (Figure 3.7). In the TEM micrograph, the L form appears to be oligomers of roughly ten nanometer in size (Figure 3.7A). In contrast, upon prolonged incubation, the β form showed twisted worm-like morphology which are a few hundred nanometers long (Figure 3.7B). The worm-like appearance of the β form is similar to the worm-like amyloid fibrils seen for multiple proteins including, amyloid-β,[41] β₂-microglobulin[42] and mouse prion protein.[43]
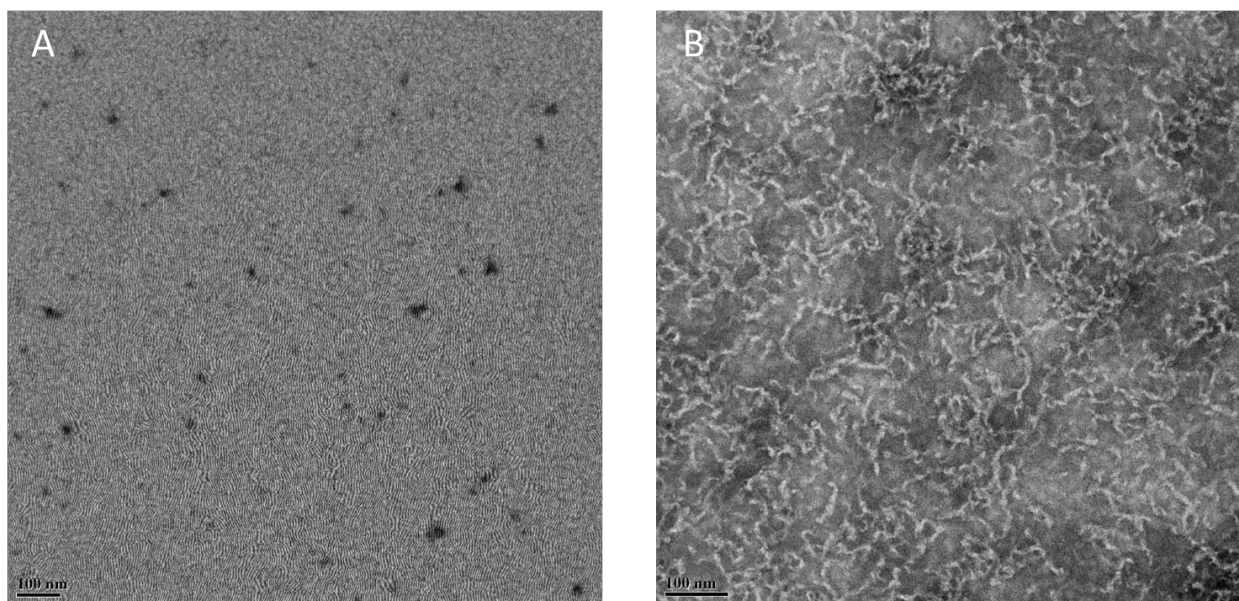
Figure 3.7: *External morphology of the L form and the β form. Transmission electron microscopy (TEM) image of (A) the L form and (B) the β form (150 mM KCl). The scale bar in the bottom left corner of the panel A and B is 100 nm.*

3.3.12 Kinetics of the L ⇌ β transition

We monitored the kinetics of L ⇌ β transition by measuring the change in ThT fluorescence and far-UV CD signal at 216 nm (Figure 3.8A). ThT is an amyloid staining dye that has been used to measure the formation and elongation of cross β sheet rich structure.[44] The L form does not bind to ThT but the β form effectively binds to ThT (Figure 3.6A, inset). In the presence of 150 mM KCl where the β form is fully populated (Figure 3.6B), we observed that the kinetics of elongation of cross β sheet structure during the formation of the β form as measured by ThT fluorescence is characterized by a single exponential change with a rate constant of 0.12 min^{-1} (Figure 3.8A). No lag phase is observed in the ThT monitored kinetics.

Interestingly, the kinetics of L ⇌ β transition as measured by far-UV CD occurs in two exponential phases (Figure 3.8A). The rate constant of the fast phase was 0.20 min^{-1} while the slow phase had a rate constant of 0.032 min^{-1} (in the presence of 150 mM KCl). The 50 % of the total change in far-UV CD signal occurred in the fast phase and the rest 50 % occurred in the slow phase of the kinetics. These results imply that the conformational conversion of the N-like secondary structure of the protein molecules in the L form to the β-sheet rich β form occurs in multiple steps. No lag phase was observed in far-UV CD monitored kinetics, similar to the ThT monitored kinetics (Figure 3.8A). The absence of lag phase in the process of conformational conversion and elongation during L ⇌ β transition implies that the L form acts as a nucleus in the formation of the β form.
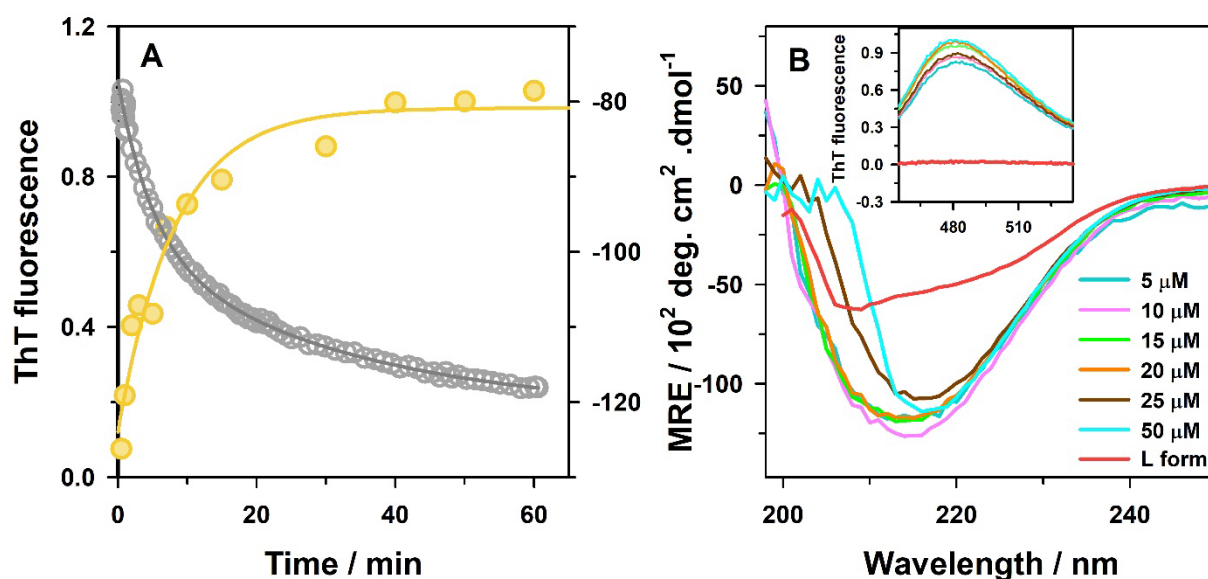
Figure 3.8: *Kinetics of L ⇌ β transition. (A) Panel A displays the change in the kinetics monitored through ThT fluorescence (yellow circles, left-hand y-axis) and far-UV CD (grey circles, right-hand y-axis) when 10 μM protein was jumped in pH 2.5 buffer containing 150 mM KCl. The yellow line through the ThT data is a fit to a single exponential equation while the black line through the far-UV CD data is a fit to a double exponential equation. (B) Far-UV CD spectra of different concentrations of the protein upon incubating at pH 2.5 with 150 mM KCl. The red line indicates the far-UV CD spectrum of the L form. The inset shows ThT fluorescence spectra of the corresponding protein samples.*

The kinetics of the L⇌β transition in figure 8A has been monitored at a monomeric protein concentration of 10 μM. However, we observed that the formation of the β form is independent of the protein concentration in the concentration range 5 μM to 50 μM, in the presence of 150 mM KCl (Figure 3.8B). The β forms formed at different protein concentrations have similar far-UV CD spectra (Figure 3.8B) and they bind to ThT dye to a similar extent (Figure 3.8B, inset). These results indicate that the β forms formed at different protein concentrations have a similar structure.
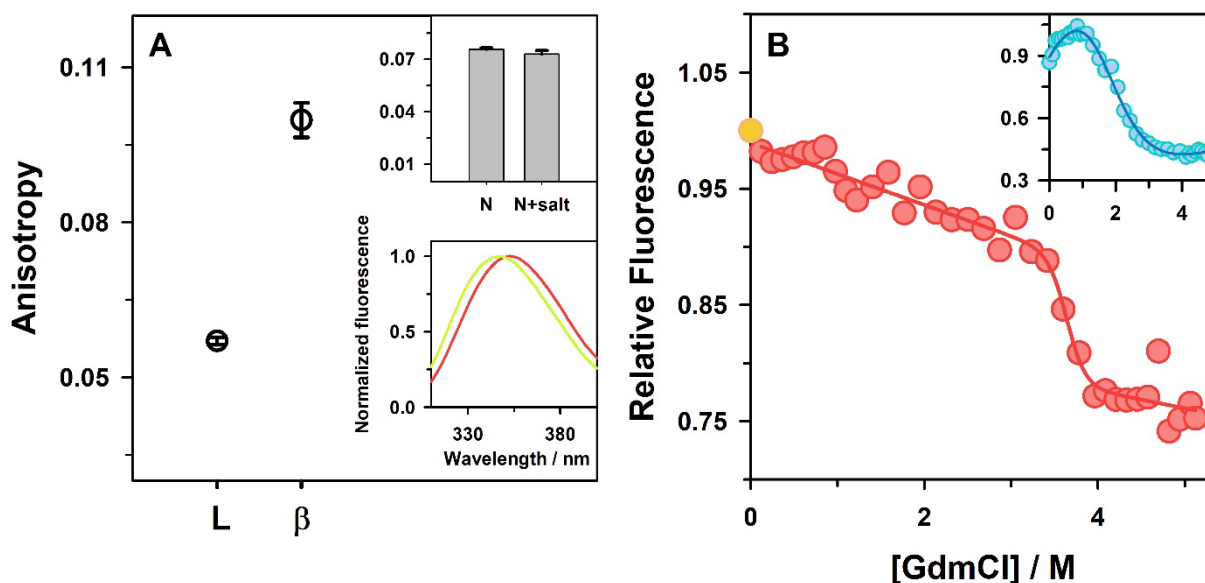
Figure 3.9: *The β form is structurally ordered and stable. A) Steady-state fluorescence anisotropy (tryptophan residues) in the L form and the β form (150 mM KCl). The top inset shows the steady-state fluorescence anisotropy (tryptophan residues) in the N form in the absence and the presence of 150 mM KCl. The bottom inset compares the normalized fluorescence spectrum of the L form (red line) and the β form (yellow line). B) Change in the fluorescence of the protein at pH 2.5 (main panel, red circles) and pH 7.5 (inset, blue circles) with varying concentration of GdmCl. The L form signal is indicated by filled yellow circle. The solid lines in the main panel and its inset are shown to guide the eyes.*

3.3.13 The β form is structurally rigid and highly stable in comparison to the L form and the N form

The increase in the MRE value in the β form in comparison to the L form (Figure 3.6A) may be due to the conversion of the disordered structure of the protein (Figure 3.1) into a more ordered structure in the β form. This random coil to β-sheet transition, however, does not occur in the N form whose CD structure remains largely unchanged with the increase in ionic strength of the solution.[34] We observed that the L⇌β transition is accompanied by an increased rigidity of protein side-chains (Figure 3.9A). The tryptophan residues in the L form are located in a region which has a U-like solvation environment (Figure 3.2B and its inset) and is highly dynamic as indicated by a low value of steady-state fluorescence anisotropy (Figure 3.2D). The mean value of steady-state fluorescence anisotropy of tryptophan residues in the L form is 0.057 which increases to 0.099 in the β form (Figure 3.9A). The higher value of anisotropy in the β form indicates a relatively ordered structure around the tryptophan residues in the β form.

In contrast, the value of fluorescence anisotropy of tryptophan residues in the N form does not change in the absence or presence of 150 mM KCl (Figure 3.9A, top inset).

The L $\rightleftharpoons$ β conversion is also marked by a blue-shift in the λ_{max}^{em} of tryptophan residues. The mean value of λ_{max}^{em} in the L form is found to be 351 nm $\pm$ 1 nm, which shows a considerable blue shift to 345 nm $\pm$ 1 nm in the β form (Figure 3.9A, bottom inset). These results indicate that tryptophan residues are in a more hydrophobic environment in the β form compared to the L form due to the exclusion of water molecules and formation of ordered protein structure.

We examined the structural stability of the L form, the β form and the N form using a denaturant (GdmCl) induced unfolding transitions (Figure 3.9B). We observed that the L form is extremely metastable and converts to the β form with an addition of as low as 120 mM of GdmCl (Figure S3.2), as inferred from changes in far-UV CD spectrum (Figure S3.2A), increase in fluorescence anisotropy of tryptophan residues (Figure S3.2B), and a blue-shift of ~ 6 nm in their λ_{max}^{em} (Figure S3.2C), and the elution volume in size exclusion chromatography (Figure S3.2D). The β form, however, showed an apparently sigmoidal unfolding transition with the mid-point at 3.6 M of GdmCl (Figure 3.9B). In comparison, the monomeric N form also showed an apparently sigmoidal transition but with the mid-point at 1.9 M GdmCl (Figure 3.9B, inset). The observation that the β form has a relatively high resistance to chemical denaturation attributes to the fact that the protein molecules in the β form are held by strong interactions.

3.4 Discussion

The aggregation of TDP-43 in the cytosol of neurons is the hallmark of ALS and multiple other TDP-43 proteinopathies.[6] A majority of the cases of ALS (90-95%) are sporadic in nature[5] and chronic environmental stress in the form of osmotic stress, oxidative insults, and change in pH or temperature has been suggested to cause the disease.[9] TDP-43 has been reported to be recruited in the stress granules in response to heat stress, oxidative stress and chemical stressors and co-deposit with the other stress-granule proteins (eg- PABP, TIA, eIF3) in the diseased brain patients.[16, 38, 39] Stress granules are dynamic biomolecular condensates consisting of proteins and nucleic acids that largely retains the native structure of proteins and protect the cell from environmental stress but under prolonged stress they transform into abnormal aggregates.[9] Starvation stress causes a change in pH of the cytosol from pH 7.4 to pH 5.5 resulting in approximately 100 fold change in proton ion concentration.[13, 15] It has been

proposed that some proteins in the cell could function as biosensors and sense stress by site-specific protein modifications.[13] For example, such a drastic change in the concentration of proton ions is efficiently sequestered by biosensor proteins by protonation of their amino acid residues. [13, 15] The process of protonation triggers the formation of metastable assemblies and the coupling of protonation-deprotonation equilibria to assembly-disassembly of proteins has been proposed to be a fast and energy-efficient method to mitigate a decrease in pH of the cytosol due to starvation stress.[13] We, however, very poorly understand how the changes in proton ion concentration affects the assembly-disassembly equilibria and in turn modulates the nature of the initial assemblies formed in the case of TDP-43 protein during its aggregation. In our quest to understand the steps involved in the formation of different initial assemblies of TDP-43 that could form in response to a change in proton ion concentration, we employ a range of pH in this study and deduce the conformational change that accompanies the metastable assembly formation.

In this study, we observed that under low pH conditions, that mimic starvation stress, TDP-43^{tRRM} undergoes a conformational opening reaction (Figure 3.5) to form a molten globule-like structure that allows the protonation of the side-chains of buried ionizable groups (Figure 3.2B). There are three aspartate residues (D105, D138, D247) and one histidine residue (H166) that are buried in the hydrophobic core of TDP-43^{tRRM}, and it is possible that the structural change protonates any of these critical residues. These events trigger the assembly of protein molecules to form an oligomeric L form (Figure 3.3). In the L form, the protein molecules possess secondary structure similar to the native protein but the tertiary structure is disrupted (Figure 3.4A and 3.4B) and the hydrophobic core is wet (Figure 3.4C and 3.4D). The L form does not show any binding to ThT indicating that amyloidogenic structure is absent in the L form. At different times of its formation, the protein molecules maintain the native-like secondary structure (Figure 3.5A). There is no presence of higher sized amorphous aggregate,[45] at any time during the oligomeric transition (Figure S3.1).

The L form is highly metastable and upon small changes in ionic strength of the solution, it further aggregates to a β sheet rich form (the β form) (Figure 3.6). As low as 15 mM of KCl is enough to fully convert the L form into the β form, indicating that the L form is held together by very weak interactions (Figure 3.6B). In contrast, the β form is highly resistant to chemical unfolding (Figure 3.9B) indicating that the β form is structurally held by very strong interactions. The TEM experiments revealed that the external morphology of the β form resembles worm-like amyloid fibrils (Figure 3.7). The β form also binds to ThT dye. In the

literature, there has been a debate on whether the aggregates of TDP-43 and its fragments resemble amyloids.[6, 46-49] Our results show that TDP-43[tRRM] forms worm-like amyloid fibril under starvation stress-like conditions.

The L to β transition measured by far-UV CD (Figure 3.6A) is accompanied by an increase in the value of MRE, in addition to the shift in maxima to 216 nm. This result suggests that the unstructured and disordered regions of the L form assemble into ordered β sheet during this conversion. The flexible side-chains of W113 and W172 located in the RRM1 domain become rigid during L to β transition, as indicated by steady-state fluorescence anisotropy experiments (Figure 3.9A), indicating that RRM1 domain gains structure in the β form. Previous studies have also suggested that the RRM1 domain is less stable compared to the RRM2 domain which might render it more aggregation-prone. [31, 50] The β form forms in two-exponential phases (Figure 3.8A), suggesting that the ordering of structure occurs in at least two distinct steps. In totality, our results indicate that the L to β conversion is a disorder-to-order transition, akin to the 'liquid to solid' transition of stress granules to amyloids.

We observed that under the conditions of mild stress, the oligomeric A form is formed (Figure 3.2), that has been characterized in detail previously.[34] The A form has a different structural organization than the L form. Contrary to the L form, in the A form tryptophan residues have restricted mobility[34] (Figure 3.2D) and are relatively buried in the hydrophobic protein environment (Figure 2B, inset). It is also smaller in size than the L form (Figure 3.2C). These results indicate that different metastable assemblies can form under the different stress-like condition as has been proposed recently for Pub1 protein.[15] This result is important because TDP-43 forms stable assemblies of discrete densities and morphology under different disease subtypes.[28, 51] It is possible that the alternate forms of disease condition occur due to the alternate pathological conformations of initial oligomeric assemblies of TDP-43.[51]

One very important suggestion from our results is that TDP-43[tRRM] region can function as a stress-sensor as reported for Sup35,[20] Pab1,[18] and Pub1,[15] in addition to performing its nucleic acid binding functions. The sensing of pH stress begins by initial conformational change and protonation that allows the protein to oligomerize into the L form. The L form is structurally held by weak interactions and is metastable. Upon small changes in ionic strength of the solution, it transforms into amyloid-like β form and disordered regions of TDP-43[tRRM] gain structure during this conversion. In the oligomeric L form, the local concentration of the protein will be high that can bring the disordered region of TDP-43[tRRM] in close proximity making it highly prone to undergo higher-order assembly upon small fluctuation in the solution

condition. This behavior of the L form is similar to the behavior of large stress-granule like assemblies that are held by weak interactions and are metastable and form abnormal, stable, solid-like aggregates under persistent stress.[9] Stress-granules are recently reported to form in multiple steps with the initial formation of the small soluble oligomeric core[21] and it is possible that the L form resembles the initial stages of stress-granule like assemblies. In this way, the results of this study shine new light on the initial stages of the stress-induced aggregation of TDP-43 and pathogenesis of ALS.

3.5 Reference

(1) Rowland, L. P., and Shneider, N. A. (2001) Amyotrophic lateral sclerosis. *N. Engl. J. Med. 344*, 1688-1700.

(2) 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.

(3) 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. Y. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science 314*, 130-133.

(4) 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.

(5) Ling, S. C., Polymenidou, M., and Cleveland, D. W. (2013) Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron 79*, 416-438.

(6) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry 56*, 809-823.

(7) Gao, J., Wang, L., Huntley, M. L., Perry, G., and Wang, X. (2018) Pathomechanisms of TDP-43 in neurodegeneration. *J. Neurochem. 146*, 7-20.

(8) 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.

(9) 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.

(10) McGurk, L., Gomes, E., Guo, L., Mojsilovic-Petrovic, J., Tran, V., Kalb, R. G., Shorter, J., and Bonini, N. M. (2018) Poly(ADP-Ribose) prevents pathological phase separation of TDP-43 by promoting liquid demixing and stress granule localization. *Mol. Cell 71*, 703-717.e9.

(11) Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell 163*, 123-133.

(12) Lin, Y., Protter, D. S., Rosen, M. K., and Parker, R. (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell 60*, 208-219.

(13) Rabouille, C., and Alberti, S. (2017) Cell adaptation upon stress: The emerging role of membrane-less compartments. *Curr. Opin. Cell Biol. 47*, 34-42.

(14) Dewey, C. M., Cenik, B., Sephton, C. F., Dries, D. R., Mayer, P., Good, S. K., Johnson, B. A., Herz, J., and Yu, G. (2011) TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Mol. Cell. Biol. 31*, 1098-1108.

(15) Kroschwald, S., Munder, M. C., Maharana, S., Franzmann, T. M., Richter, D., Ruer, M., Hyman, A. A., and Alberti, S. (2018) Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery. *Cell Rep 23*, 3327-3339.

(16) McDonald, K. K., Aulas, A., Destroismaisons, L., Pickles, S., Beleac, E., Camu, W., Rouleau, G. A., and Vande Velde, C. (2011) TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum. Mol. Genet. 20*, 1400-1410.

(17) Kroschwald, S., and Alberti, S. (2017) Gel or Die: Phase separation as a survival strategy. *Cell 168*, 947-948.

(18) Riback, J. A., Katanski, C. D., Kear-Scott, J. L., Pilipenko, E. V., Rojek, A. E., Sosnick, T. R., and Drummond, D. A. (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell 168*, 1028-1040.

(19) Cohen, T. J., Hwang, A. W., Unger, T., Trojanowski, J. Q., and Lee, V. M. (2012) Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. *EMBO J. 31*, 1241-1252.

(20) Franzmann, T. M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A. S., Nuske, E., Richter, D., Baumeister, W., Grill, S. W., Pappu, R. V., Hyman, A. A., and Alberti, S. (2018) Phase separation of a yeast prion protein promotes cellular fitness. *Science 359*, No. eaao5654.

(21) Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016) Distinct stages in stress granule assembly and disassembly. *Elife 5*, No. e18413.

(22) Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell 162*, 1066-1077.

(23) Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S. K., Rees, E., Miyashita, A., Costa, A. R., Dodd, R. B., Chan, F. T. S., Michel, C. H., Kronenberg-Versteeg, D., Li, Y., Yang, S. P., Wakutani, Y., Meadows, W., Ferry, R. R., Dong, L., Tartaglia, G. G., Favrin, G., Lin, W. L., Dickson, D. W., Zhen, M., Ron, D., Schmitt-Ulms, G., Fraser, P. E., Shneider, N. A., Holt, C., Vendruscolo, M., Kaminski, C. F., and St George-Hyslop, P. (2015) ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron 88*, 678-690.

(24) Shiina, Y., Arima, K., Tabunoki, H., and Satoh, J. (2010) TDP-43 dimerizes in human cells in culture. *Cell. Mol. Neurobiol. 30*, 641-652.

(25) Qin, H. N., Lim, L. Z., Wei, Y. Y., and Song, J. X. (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.

(26) Tsoi, P. S., Choi, K. J., Leonard, P. G., Sizovs, A., Moosa, M. M., MacKenzie, K. R., Ferreon, J. C., and Ferreon, A. C. M. (2017) The N-terminal domain of ALS-linked TDP-43 assembles without misfolding. *Angew. Chem. 56*, 12590-12593.

(27) 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.

(28) Buratti, E. (2015) Functional significance of TDP-43 mutations in disease. *Adv. Genet. 91*, 1-53.

(29) Ayala, Y. M., Pantano, S., D'Ambrogio, A., Buratti, E., Brindisi, A., Marchetti, C., Romano, M., and Baralle, F. E. (2005) Human, Drosophila, and C.elegans TDP43: Nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol. 348*, 575-588.

(30) Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ue, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. T. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol. 20*, 1443-1449.

(31) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res. 37*, 1799-1808.

(32) 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.

(33) Zacco, E., Grana-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.

(34) 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.

(35) Mann, J. R., Gleixner, A. M., Mauna, J. C., Gomes, E., DeChellis-Marks, M. R., Needham, P. G., Copley, K. E., Hurtle, B., Portz, B., Pyles, N. J., Guo, L., Calder, C. B., Wills, Z. P., Pandey, U. B., Kofler, J. K., Brodsky, J. L., Thathiah, A., Shorter, J., and Donnelly, C. J. (2019) RNA binding antagonizes neurotoxic phase transitions of TDP-43. *Neuron 102*, 321-338.e8.

(36) Huang, Y. C., Lin, K. F., He, R. Y., Tu, P. H., Koubek, J., Hsu, Y. C., and Huang, J. J. (2013) Inhibition of TDP-43 aggregation by nucleic acid binding. *Plos One 8*, No. e64002.

(37) Volkening, K., Leystra-Lantz, C., Yang, W. C., Jaffee, H., and Strong, M. J. (2009) Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). *Brain Res. 1305*, 168-182.

(38) Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y. J., Vanderwyde, T., Citro, A., Mehta, T., Zaarur, N., McKee, A., Bowser, R., Sherman, M., Petrucelli, L., and Wolozin, B. (2010) Tar DNA binding protein-43 (TDP-43) associates with stress granules: Analysis of cultured cells and pathological brain tissue. *Plos One 5*, No. e13250.

(39) Colombrita, C., Zennaro, E., Fallini, C., Weber, M., Sommacal, A., Buratti, E., Silani, V., and Ratti, A. (2009) TDP-43 is recruited to stress granules in conditions of oxidative insult. *J. Neurochem. 111*, 1051-1061.

(40) Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266-280.

(41) Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J., and Teplow, D. B. (1999) Aβ-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem. 274*, 25945-25952.

(42) Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005) Competing pathways determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol. 351*, 850-864.

(43) Jain, S., and Udgaonkar, J. B. (2008) Evidence for stepwise formation of amyloid fibrils by the mouse prion protein. *J. Mol. Biol. 382*, 1228-1241.

(44) Naiki, H., and Gejyo, F. (1999) Kinetic analysis of amyloid fibril formation. *Methods Enzymol. 309*, 305-318.

(45) Adachi, M., Noji, M., So, M., Sasahara, K., Kardos, J., Naiki, H., and Goto, Y. (2018) Aggregation-phase diagrams of β2-microglobulin reveal temperature and salt effects on competitive formation of amyloids versus amorphous aggregates. *J. Biol. Chem. 293*, 14775-14785.

(46) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc. 132*, 1186-1187.

(47) Neumann, M., Kwong, L. K., Sampathu, D. M., Trojanowski, J. Q., and Lee, V. M. (2007) TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: Protein misfolding diseases without amyloidosis. *Arch. Neurol. 64*, 1388-1394.

(48) Mompean, M., Hervas, R., Xu, Y., Tran, T. H., Guarnaccia, C., Buratti, E., Baralle, F., Tong, L., Carrion-Vazquez, M., McDermott, A. E., and Laurents, D. V. (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett. 6*, 2608-2615.

(49) Saini, A., and Chauhan, V. S. (2014) Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir 30*, 3845-3856.

(50) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem. 289*, 8264-8276.

(51) Laferriere, F., Maniecka, Z., Perez-Berlanga, M., Hruska-Plochan, M., Gilhespy, L., Hock, E. M., Wagner, U., Afroz, T., Boersema, P. J., Barmettler, G., Foti, S. C., Asi, Y. T., Isaacs, A. M., Al-Amoudi, A., Lewis, A., Stahlberg, H., Ravits, J., De Giorgi, F., Ichas, F., Bezard, E., Picotti, P., Lashley, T., and Polymenidou, M. (2019) TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci. 22*, 65-77.

Supporting Figures
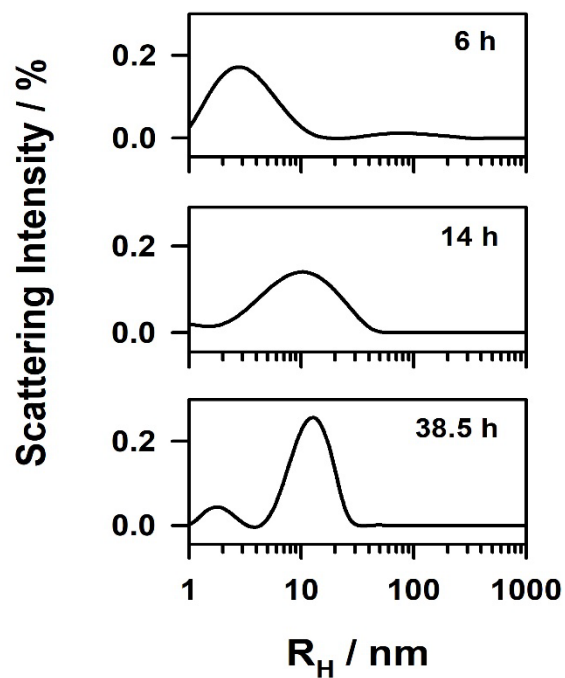
Figure S3.1. *DLS data upto 1000 nm shows absence of higher sized species at anytime during the formation of the L form. The x-axis is shown in the log scale.*
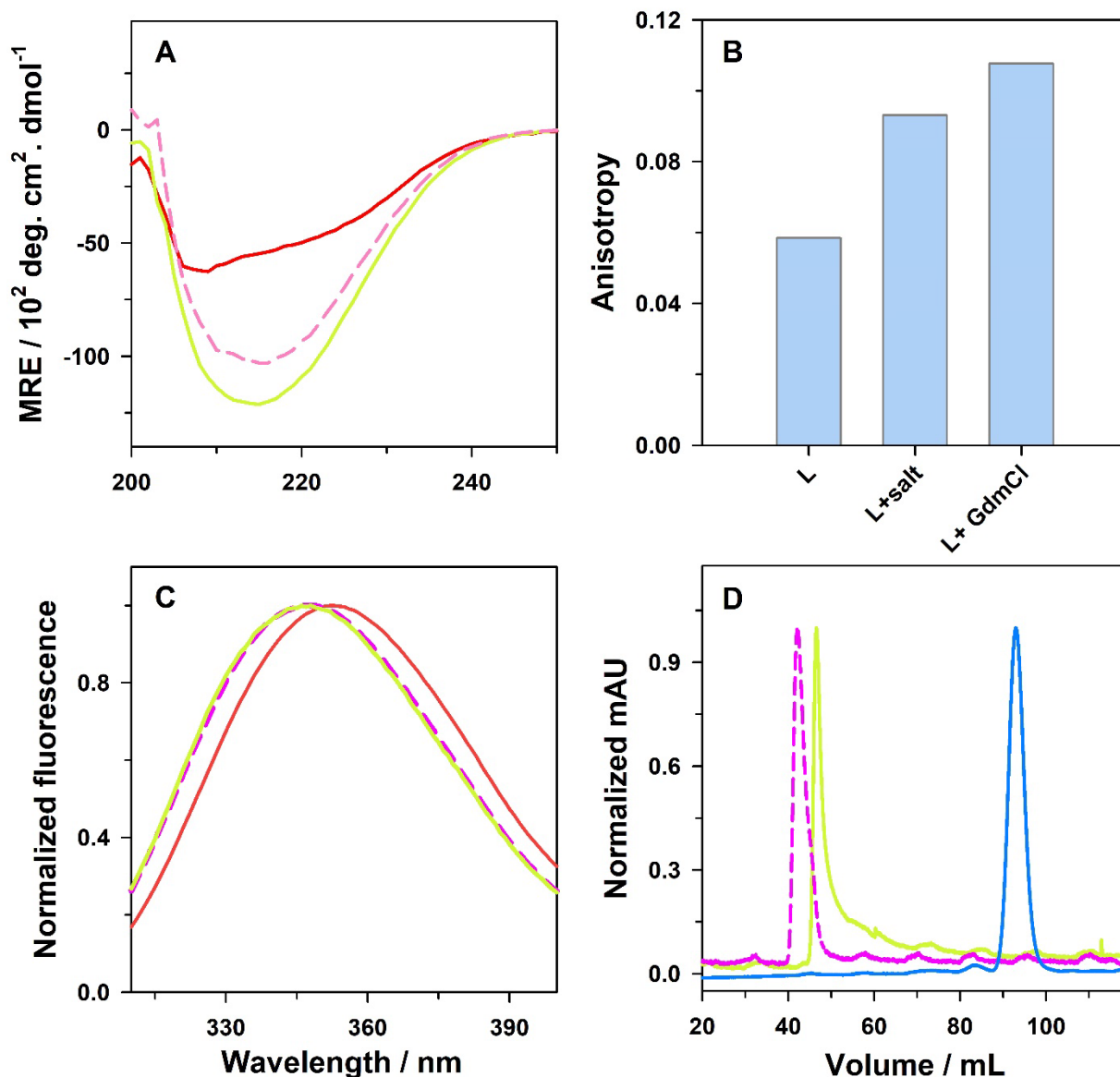
Figure S3.2. *The L form transforms into the β form in the presence of minute amounts of GdmCl. A) Far-UV CD spectrum of the L form in the presence of 3 mM KCl (solid red), L form in the presence of 150 mM KCl (β form) (solid green) and the L form in the presence of 120 mM GdmCl (dash pink). B) Steady-state fluorescence anisotropy of the tryptophan residues of the protein in the L form in the presence of 3 mM KCl, L form in the presence of 150 mM KCl (β form) and the L form in the presence of 120 mM GdmCl. C) Normalized fluorescence spectrum showing the Stokes shift of the tryptophan residues in the L form in the presence of 3*

mM KCl (solid red), L form in the presence of 150 mM KCl (β form) (solid green) and the L form in the presence of 120 mM GdmCl (dash pink). D) Comparison of the elution volumes of the N form (solid blue), the KCl induced β form (solid green) and the GdmCl induced β form (dash pink) as monitored by size exclusion chromatography.

Chapter 4

Multi-Step Molecular Mechanism of Amyloid-Like Aggregation of Nucleic Acid-Binding Domain of TDP-43

4.1 Introduction

Misfolding of proteins have been the underlying cause of many deadly human neurodegenerative disorders.[1, 2] Over 50 different proteins have been studied for their role in different neurodegenerative diseases.[1, 2] One common feature across all these disease spectrum is the deposition of intracellular inclusions or extracellular plaques, irrespective of the different primary sequence of the protein undergoing aggregation.[2] Intracellular aggregation of TDP-43 is the hallmark feature of 97% of the familial amyotrophic lateral sclerosis (ALS) and 45% of the frontotemporal lobar degeneration (FTLD) cases.[3-5] Inclusion bodies of TDP-43 have been primarily found to be deposited in the cytoplasm of the neuron and glial cells under the diseased state.[3, 5] However, recent studies have also shown the presence of monomer, heterogeneous oligomers and fibrils of TDP-43 isolated from the diseased FTD patients.[6, 7] A similar trend has also been observed for α-synuclein, tau, prion and amyloid-β-peptide (Aβ-42) which has been shown to populate a variety of intermediates prior to the formation of amyloid fibrils.[8-12] The oligomers or intermediates seen across the different diseases have been shown to share common epitopes, implying the formation of similar types of intermediates.[13, 14] Moreover, the characteristic cross-β sheet structure is also similar despite the differences in the native fold of the amyloidogenic protein.[15, 16] Many studies across different neurodegenerative disease models have shown that the early oligomeric species are toxic while the deposited plaques or the protein fibrils are usually inert.[9, 11] All these characteristic features signify a general progressive pathogenic disease mechanism that might be common to different neurodegenerative diseases.[13, 17, 18] Unfortunately, not much is known about the structural characteristics of the different intermediate states and their rate of inter-conversion during the aggregation process of TDP-43 protein.

TDP-43 is a multi-domain nucleic acid-binding protein, expressed ubiquitously, that plays a significant role in mRNA and gene regulation.[19, 20] The four domains of TDP-43 – N-terminal, tandem RNA Recognition Motifs (RRM1 and RRM2), and the C-terminal domain have been predominantly explored independently for their role in aggregation.[21] A considerable number of studies on the C-terminal domain have emphasized its role in phase separation and aggregate formation.[22-24] However, a recent study using ss-NMR on the fibrillation of different domains of TDP-43 has suggested that the N-terminal domain and the tandem RRM domain (TDP-43[tRRM]) form the core of the amyloid fibrils while the C-terminal domain contributes towards the heterogeneity in the structure of the fibril.[25] TDP-43[tRRM] has been proven to undergo oligomerization under different stress-like conditions such as oxidation,[26] pH,[27, 28] and

temperature,[29] linking stress to the aggregate formation. Furthermore, mutations within this domain have also been shown to modulate the aggregation propensity of the protein.[30-34] TDP-43[tRRM] has been shown to act as a bio-sensor and sense stress.[28] Our previous investigation has illustrated the effect of protonation-deprotonation equilibria and ionic strength on the assembly-disassembly equilibria, thus allowing the protein molecules to sense and respond to stress.[28] In our earlier study, we have demonstrated the formation of native-like oligomers at low pH and low salt concentration (~5 mM KCl), showing characteristics of wet molten globule, while at a moderate salt concentration (20-150 mM KCl), we observe the formation of curly amyloid-like assembly known as the β form.[28] We observed that by modulating the ionic strength of the solution at low pH, we could populate molecules with distinct spectroscopic properties. Preliminary kinetic experiments have shown that the conformational changes measured by circular dichroism (CD) is biphasic, indicating that the structural changes occur in multiple steps.[28] However, it is still unclear how these species are dynamically related and how the step-wise conversion of the different species occurs during the aggregation process. It is thus crucial to understand the structural features of the intermediate species and study the rate-limiting step that defines the mechanism of the aggregation to design small therapeutic molecules to target these species.[35, 36]

In the present work, we have studied the step-wise mechanism of the aggregation of TDP-43[tRRM] by monitoring the evolution of the aggregation process using a range of spectroscopic probes. Interestingly, we observe that the ThT kinetics follow a mono-exponential change while the CD and hydrodynamic radius (R_H) follows a bi-exponential change. Our kinetic data implies four distinct steps during the aggregation process- very fast, fast, slow and very slow. We show that the aggregation begins with a partial unfolding reaction in the 'very fast' step which is followed by the formation of higher order intermediates (HOI) of size 10-12 nm in the 'fast' step. These HOI's can bind to the Thioflavin T (ThT) dye and show some degree of change in its secondary structure as monitored by CD. The HOI's assemble and undergo structural rearrangement to form protofibrils which further assemble to form curly amyloid-like assembly, the β form. The protein concentration dependence of the kinetics shows additional steps at the early stage of aggregation. The amyloid-like assembly formed across different protein concentration remain similar and is found to be around 20 nm in size. The temperature dependence of the aggregation kinetics shows that there are multiple transition states accompanying the aggregation process. Our data indicate that the aggregation of TDP-43[tRRM] is a multi-step reaction whereby the protein concentration defines the structural

characteristics of the initial species resulting in multiple ways the protein molecules take to assemble in a β sheet rich amyloid-like assembly.

4.2 Materials and methods

Protein purification

TDP-43[tRRM] protein (UniProtKB/Swiss-Prot entry Q13148) has been expressed and purified as described previously.[27] The purity of the protein was confirmed by SDS-PAGE, and the mass was determined to be 19,429 Da using electrospray ionization-mass spectrometry.

Aggregation condition

The buffers and reagents used in this study were of high purity grade obtained from Sigma and HiMedia. The protein was stored in 10 mM KPi buffer containing 150 mM KCl (pH 7.2). The aggregation was carried out by concentrating the protein up to 600-800 µM and then transferring it into 20 mM Gly-HCl buffer containing 150 mM KCl (pH 3) at the required protein concentration. All the buffers contained 1 mM DTT and were filtered through 0.2 µm filter before use.

pH-induced equilibrium experiments

For the pH-induced experiments, buffers used at different pH are described in the previous work.[27] All the equilibrium experiments were carried out as described previously.[27]

Circular Dichroism

Far-UV CD measurements were carried out in Jasco J-815 spectropolarimeter. For the kinetic experiment, the required concentration of protein was transferred to the aggregation condition in 0.1 cm cuvette, and the ellipticity was measured at 216 nm at regular intervals. The deadtime for the experiment was calculated by noting the difference in the time between the protein transfer to aggregation buffer and the first reading measured by the instrument. Equilibrium spectra were collected in the 200-250 nm range using the following settings - 2 nm, bandwidth; 4 s, DIT, and 20 nm/min scan speed. Each spectrum was averaged over 3 scans.

Thioflavin T assay

For performing ThT kinetics, 1 µM protein sample was transferred from the aggregation sample at different time points and added to the ThT buffer (50 mM Tris, pH 8)

containing 20 μM ThT. The kinetics was monitored by exciting the dye at 440 nm and monitoring the emission at 482 nm for 20 s. The spectra were collected between 450- 600 nm for the equilibrium measurement with a step resolution of 1 nm. The excitation and the emission slit width were set at 1 nm and 10 nm, respectively. The emission of the buffer containing the ThT dye alone was captured under similar settings. The emission value for the kinetic experiment was plotted by calculating the average of all the measurements taken for 20 s and subtracting for the buffer contribution.

Trp Fluorescence and anisotropy

The intrinsic Trp fluorescence was obtained by exciting the Trp residues at 295 nm and collecting the emission spectra from 310- 400 nm in FluoroMax-4 spectrofluorometer (Horiba Scientific). The excitation and emission slit width was set to 1 nm and 10 nm, respectively. For the kinetic experiment, protein from the concentrated stock was transferred to both the N form and aggregation condition, and the fluorescence was measured at 340 nm, immediately after the transfer. The dead time was calculated as the time between the protein transfer and the time when the first signal was recorded.

The Trp anisotropy was measured by exciting the Trp residues at 295 nm and measuring the anisotropy at 340 nm for 120 s using an anti-photobleaching setting. The excitation slit width was set to 1 nm, while the emission slit width was set to 10 nm. An integration time of 1 s was used.

Dynamic Light Scattering

DLS was carried out on a DynaPro 99 unit (Wyatt). All the buffers used were filtered with a 0.2 μm filter, and the centrifuge tubes and tips used were also rinsed with 0.2 μm filtered MilliQ water just before use. The data was collected using a similar setting, as reported previously.[28] An average of 50 acquisitions was used to obtain the mean autocorrelation function, which was then resolved into Gaussian distributions of hydrodynamic radii.

Size Exclusion Chromatography

SEC was performed for the N form and the β form using a Superdex 75 10/300 GL column and HiLoad 16/600 Superdex 200 pg column[27] on an AKTA Pure M FPLC system (GE Healthcare). The fractionation range of this column is 3-70 kDa, and it has a void volume of

7.2 mL. The column was equilibrated with the respective buffer before loading 10 µM of the protein sample. A flow rate of 0.8 mL/min was used. The apparent molecular weight of the N form and the β form was determined from the known molecular weights of standard biomolecules as described previously.[27]

Transmission electron microscopy

The sample for TEM was prepared by placing 10 µL of 25 µM for 5 min on a 300-mesh Formvar carbon-coated copper grid (Electron Microscopy Science). The excess sample was removed, and the grid was stained negatively using 2% uranyl acetate solution for 1.5 min. This was followed by washing the grid with filtered MilliQ for a min and allowing it to air dry. The samples were examined with transmission electron microscopy (Technai-T20) at an accelerating voltage of 200 kV.

4.3 Results

4.3.1 Formation of β-sheet rich amyloid-like assembly

TDP-43^{tRRM} shows a structure rich in α-helix, β-sheet, and the disordered region, as seen in figure 4.1A. The far-UV CD spectrum of the N form shows signal minima at 208 and 222 nm (Figure 4.1B), which is characteristic of its native structure (Figure 4.1A). When the N form is transferred to pH 3 containing 150 mM KCl (aggregation buffer), there is a transition from its native structure to a structure rich in β-sheet content, as seen by the shift in the minima to 216 nm. The transition is also associated with an increase in ellipticity (Figure 4.1B). The increased ellipticity suggests a gain in the structure as the disordered regions in the protein have become more ordered, resulting in the formation of a β-sheet rich conformation, β form. The N to β transition, when measured using far-UV CD at 216 nm under equilibrium condition, as a function of pH, shows an intact secondary structure of the protein between pH 7.5 and pH 5 (Figure 4.1B; inset). However, when the pH is decreased below pH 5, ellipticity increases in a sigmoidal fashion whose value saturates at around pH 3. The value of mean residue ellipticity (MRE) at 216 nm when measured at pH 7.5 and pH 3 is -5833 ± 180 deg cm^2 dmol^{-1} and -12890 ± 700 deg cm^2 dmol^{-1}, respectively (Figure 4.1B; inset).

Interestingly, the β form binds to the ThT dye while the N form does not bind to the dye (Figure 4.1C), indicating an amyloid-like characteristic to the β form. The N to β transition was also measured using ThT fluorescence at 482 nm under equilibrium conditions as a

function of pH (Figure 4.1C; inset). The protein does not bind to ThT dye between pH 7.5 and pH 5. However, as the pH is decreased below pH 5, there is an increase in the ThT binding resulting in the formation of amyloid-like molecules at the low pH. The increase in the ThT fluorescence occurs concurrently with the secondary structure change, indicating that the regions undergoing a change in the secondary structure conformation create binding pockets for the ThT dye.
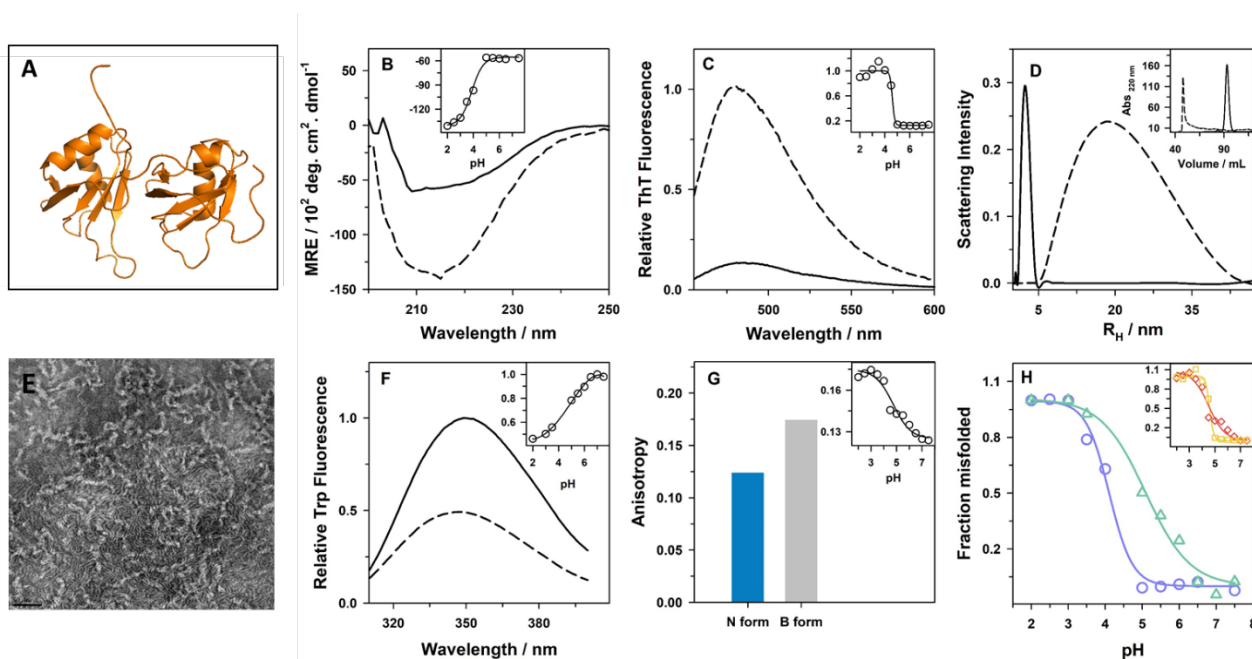
Figure 4.1: *pH-dependent change in the conformation of TDP-43^{tRRM}. (A) Cartoon representation of the nucleic acid-binding domain of TDP-43 (PDB file: 4BS2), termed here as TDP-43^{tRRM}. The protein was transferred into different pH buffers containing 150 mM KCl and incubated to achieve equilibrium before taking the measurement. In panels B, C, D and F, the equilibrium signal of the N form at pH 7.5 (solid line), and the β form at pH 3 (dash line) is plotted. (B) Far-UV CD scan of the N form and the β form. The inset shows the change in the secondary structure measured at 216 nm as a function of pH. (C) Fluorescence of the ThT dye upon excitation at 442 nm when it remains unbound to the N form and bound to the β form. The inset shows the change in the ThT fluorescence plotted at 482 nm as a function of pH. (D) Hydrodynamic radius (R_H) distributions of the N form and the β form as measured by dynamic light scattering (DLS). The inset shows the size as measured by size exclusion chromatography (SEC) adapted from reference 27. (E) Transmission electron microscope (TEM) image of the β form shows small curly assemblies. The black scale bar on the lower-left corner of the image represents 50 nm. (F) Relative fluorescence of Trp residues (W113 and W172) in the N form and the β form upon excitation at 295 nm. The inset shows the change in the fluorescence of the Trp residues at 340 nm as a function of pH. (G) The anisotropy value of the Trp residues in the N form and the β form upon excitation at 295 nm. The inset shows the change in the fluorescence anisotropy of the Trp residues at 340 nm as a function of pH. The data points in the inset of panels B, C, F, and G are fitted to a sigmoidal curve (black line). (H) The inset of*

the panels B, C, F and G were used to calculate the fraction of misfolded molecules as a function of pH: CD- blue circles, Trp fluorescence- green triangles represented in the main panel while ThT- yellow squares, anisotropy- red diamonds are represented in the inset. The lines through the data points in the main panel and the inset are fit to a sigmoidal curve.

4.3.2 The β form is curly amyloid-like assembly

The formation of the β form is accompanied by an increase in the size of the molecules, as indicated by the measurement of the R_H of the β form. The mean R_H of the N form is around 2.3 nm,[27] while that of the β form is approximately 20 nm (Figure 4.1D). The increase in the size can also be seen using size exclusion chromatography (SEC) (Figure 4.1D; inset).[27] The N form was estimated to be around 18.9 kDa, while the β form measures around ~485 kDa. However, the β form elutes near the column's void volume, and hence the calculated molecular weight is only a lower limit of its size.[27] The increase in the size of the β form was also imaged using a transmission electron microscope (TEM) (Figure 4.1E). The TEM image shows that the β form appears similar to small curly assemblies measuring around 50 nm in size. These assemblies are comparable to small worm-like fibrils seen for other proteins such as prion[37], β2 microglobulin[38] and amyloid-β.[39]

4.3.3 Altered packing in the β form results in rigid movement of Trp residues

There are two Trp residues located at 113 and 172 positions in the RRM1 domain of TDP-43[tRRM]. Upon measuring the fluorescence spectrum by exciting the Trp residues in the N form and the β form, we observed that the fluorescence is decreased in the β form compared to the N form. This indicates that the tertiary structure around the Trp residues in the β form is highly altered (Figure 4.1F). Interestingly, the λ_{max} in the N form is 349 nm, while it is slightly blue-shifted to 347 nm in the β form. Our data suggest that the solvation environment around the Trp residues does not change significantly as the N form undergoes a transition to β form. The change in the packing monitored through Trp fluorescence at 340 nm was measured during N to β transition as a function of pH under equilibrium conditions. Between pH 7.5 and pH 6, the Trp fluorescence remains similar. However, the Trp fluorescence gradually decreases beyond pH 6 and shows a sigmoidal change (Figure 4.1F; inset). There is almost a two-fold relative decrease in the Trp fluorescence during the N to β transition.

Interestingly, the altered packing of the side chain residues results in a very rigid environment around the Trp residues in the β form, as shown by the higher anisotropy value of

the β form compared to the N form (Figure 4.1G). The anisotropy value of the Trp residues in the N form is around 0.12 and the β form is 0.17 (Figure 4.1G). This increase in the anisotropy value indicates a rigid environment around the Trp residues in the β form. This rigidity may stem from the ordered structure of the β form (Figure 4.1B, 4.1D, and 4.1E). The anisotropy change as a function of N to β transition also occurs in a sigmoidal fashion with the increase in the anisotropy beginning from pH 6 (Figure 4.1G; inset). This gradual increase in the anisotropy value indicates the gradual change in the packing that accompanies the increase in the rigidity of the environment around the Trp residues.

4.3.4 Decoupling of the tertiary and the secondary structural changes during the N to β transition at equilibrium

The N to β transition as monitored by far-UV CD, ThT, Trp fluorescence, and anisotropy was used to calculate the fraction of the misfolded molecules as a function of pH. The midpoint of the transition for the Trp fluorescence is estimated to be at pH 5, while the midpoint of the transition for the far-UV CD change is pH 4 (Figure 4.1H). Although the process of ThT fluorescence and anisotropy shows a difference in the fraction of misfolded molecules during the transition, they seem to have a similar midpoint of transition estimated to be around pH 4.5 (Figure 4.1H; inset). The equilibrium misfolding curves of N to β transition monitored by different probes are non-coincident, suggesting that the protein undergoes a multi-step structural change in the conformation when exposed to increasing proton ion concentration and physiological ionic strength (150 mM) solution. The data indicate that the least stable structural element is the side-chain packing, whose disruption affects the local dynamics around the Trp residues. In contrast, the secondary structure is the most stable element whose transformation results in an increase in ThT fluorescence. These results show that the different structural events accompanying the N to β transition during the pH-induced aggregation of TDP-43^{tRRM} are decoupled under equilibrium conditions.

4.3.5 Kinetics of the amyloid fibril formation as monitored by different probes

Figure 4.2 shows the kinetics of the formation of the amyloid-like assembly upon transferring the N form into the aggregation condition. The transition was monitored using far-UV CD, Trp and ThT fluorescence and DLS. Figure 4.2A shows the change in the secondary structure monitored at 216 nm as the protein molecules undergo structural transformation. We observe that the change occurs in two stages, resulting in a biphasic conversion of the N form to the β form. Upon monitoring the far-UV CD scans at different time points, we noticed a shift in the signal minima to 216 nm accompanied by an increase in the ellipticity in the first stage of the transition. In the second stage of the transition, we see a further rise in the ellipticity giving rise to the second phase of the biphasic transition. Interestingly when we monitored the N to β transition using ThT fluorescence, we observed that the increase in the ThT fluorescence occurs in a single step resulting in a monophasic transition (Figure 4.2B).
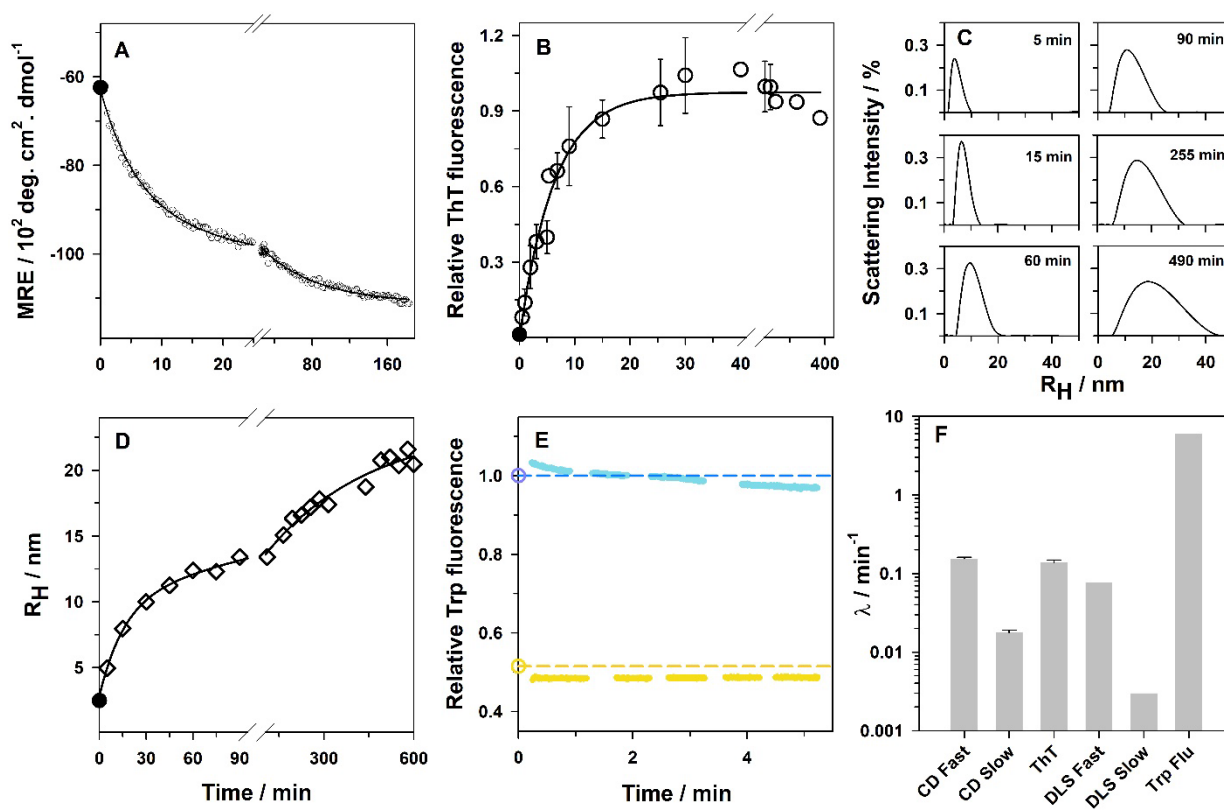
Figure 4.2: *Kinetics of the formation of the β form upon transferring 10 μM protein to the aggregation buffer. (A) Kinetics of the secondary structural change at 216 nm. The continuous line through the data points is a fit to a double exponential equation. (B) ThT fluorescence monitored kinetics plotted at 482 nm. The continuous line through the data points is a fit to a single exponential equation. (C) R_H distributions of the protein at six different times during the aggregation process are shown. (D) Mean R_H -monitored kinetics. The continuous line through*

the data points is a fit to a double exponential equation. (E) Fluorescence of the Trp residues at 340 nm upon transferring the protein to the native buffer (blue trace) and aggregation buffer (yellow trace). The blue dashed, and yellow dashed lines indicate the equilibrium signals of the N form and the β form, respectively. (F) Bar graph showing apparent rate constants on a logarithmic scale as monitored by CD, ThT, DLS, and Trp fluorescence. The error bar represents the standard deviation observed by three independent experiments. The black-filled circle in panels A, B, and D represents the signal of the N form before transferring into the aggregation condition.

Figure 4.2C shows a change in the distribution of the R_H at different times during the aggregation process. The transition into the β form is accompanied by an increase in the mean R_H of the population, as indicated by the shift in the peak with time. Upon plotting the mean R_H as a function of time (Figure 4.2D), we observed that the change in the size of the population occurs in a biphasic manner, involving a fast step that results in the formation of species of around 10-12 nm size and a very slow change that further increases the size of the population to about 20 nm. As the aggregation reaction progressed with time, we also observed an increase in the heterogeneity of the protein molecules as indicated by the width of the R_H distribution (Figure 4.2C).

Figure 4.2E shows the fluorescence of Trp residues in the N form and immediately upon transferring into the aggregation condition. The decrease in the fluorescence occurs within the dead time of the measurement (~ 14 s) and achieves equilibrium. This very fast reaction suggests that the first step during the N to β transition is the loosening of side-chain packing around the Trp residues in the protein, resulting in the transient formation of partially unfolded molecules. Interestingly, the process of CD and mean R_H monitored kinetics occurs on a much slower timescale in a biphasic fashion (Figure 4.2A, 4.2D), while the process of elongation shows a monophasic change (Figure 4.2B). In each of the kinetics, the signal of t=0 coincides with the signal of the N form. No lag phase was observed when monitored using any of the probes.

The apparent rate constants for the formation of the β form at 10 μM protein concentration, as monitored by the four structural probes, are plotted in Figure 4.2F. A minimum rate for the Trp fluorescence was estimated by assuming the dead time of the experiment as the time when ~99% of the change has occurred. The fastest change that arises is the change in the Trp fluorescence that is almost 40 fold faster than the ThT and the first phase of the CD. The change in the conformation of the protein observed in the first phase of CD and the elongation process monitored by ThT has a similar rate. However, the rate constant

of the first phase of DLS is 2.5 fold slower. Interestingly, we observed a second slow phase that involves a change in the conformation of the protein molecules monitored by CD, and this change occurs at a rate that is 10 fold slower than the first phase of the CD process. The slowest change monitored was the increase in the mean R_H of the population as observed by the DLS. The rate constant of the slowest phase is approximately 20 fold slower than the first phase of DLS and 50 fold slower than the first phase of CD and ThT. A significant difference in the apparent rate constants of all the four structural probes implies that the course of aggregation and the formation of the β form occurs in a multi-step manner that can be delineated by the structural probes examined in this study.
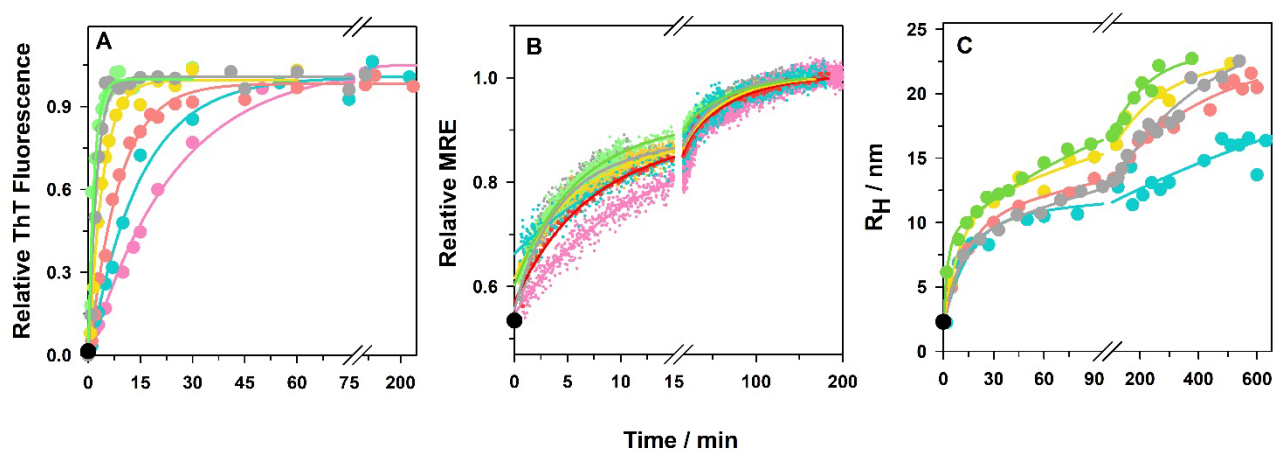
Figure 4.3: *Protein concentration dependence of the kinetics of the formation of the β form. Protein at different concentrations- 3 μM (pink), 5 μM (blue), 10 μM (red), 15 μM (yellow), 20 μM (gray) and 25 μM (green) were transferred to the aggregation buffer, and the kinetics were measured using three different probes. A representative kinetic for each of the protein concentrations are shown. (A) Elongation of the protein as monitored by ThT fluorescence at 482 nm. The continuous line through the data points is a fit to a single exponential equation. (B) Conformational change in the protein monitored using CD at 216 nm. (C) Change in the mean R_H of the protein as measured by DLS. The continuous line through the data points in panels B and C are fits to a double exponential equation. The filled black circle in all the panels is the signal of the N form before transferring the protein into the aggregation condition.*

4.3.6 Protein concentration dependence of the kinetics of the formation of β form

The kinetics of the formation of the β form was studied over a range of protein concentrations – 3 μM to 25 μM using ThT fluorescence, CD, and DLS (Figure 4.3). The kinetics of the aggregation process monitored using different probes across all the concentrations show no lag phase. Across the protein concentration range studied, the ThT

process shows monophasic change (Figure 4.3A), while the CD and DLS monitored kinetics show a biphasic change (Figure 4.3B and 4.3C). Upon monitoring the rate of change in the ThT fluorescence, CD and DLS, we observed that the apparent rate constants of the ThT and the DLS process are highly dependent on the protein concentration, but the CD seems to be fairly independent of the protein concentration. Interestingly, we observe that the total change in the signal remains similar across all the protein concentrations in different probes. A similar final signal indicates a similar type of β form is formed across all the concentration range studied. The size of the β form monitored by DLS across the protein concentration remains at around 20 nm. The kinetics of all the probes extrapolate at t=0 to the signal of the N form.
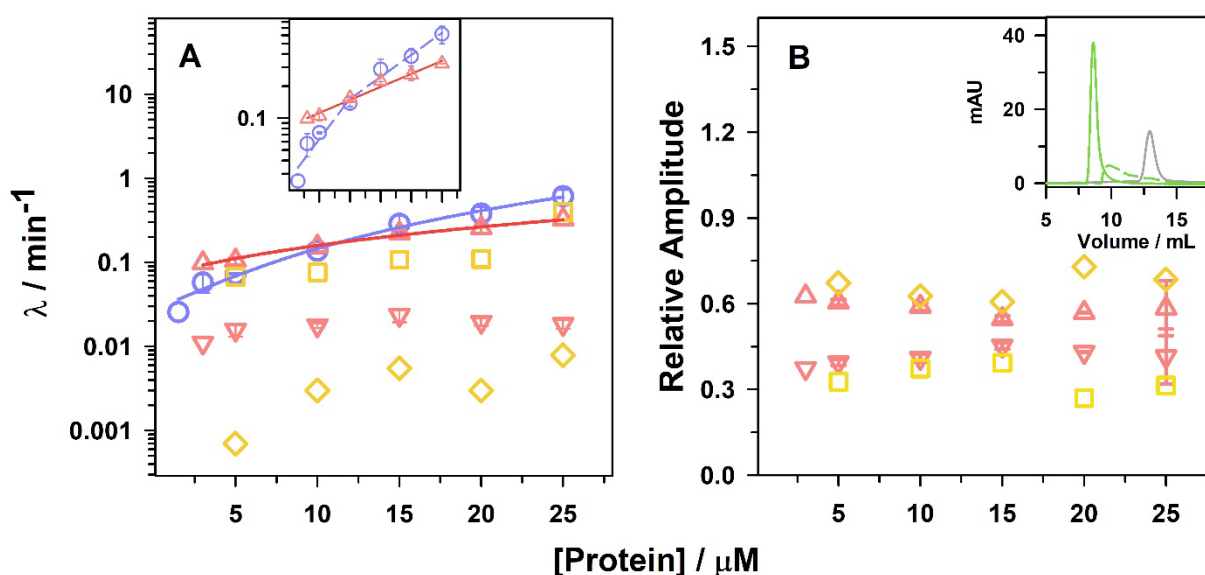
Figure 4.4: *Dependence of the apparent rate constants and relative amplitude of different phases is plotted as a function of the protein concentration. (A) The apparent rate constants at different protein concentrations, obtained from figure 3, are plotted on a logarithmic scale for the different probes. Rate constants of the ThT fluorescence monitored kinetics (blue circles), the first phase (red triangles) and the second phase (inverted red triangles) of the CD monitored kinetics and first phase (yellow squares) and the second phase (yellow diamonds) of the mean R_H monitored kinetics is plotted. The line through the data in panel A is to guide the eyes. The inset shows a zoomed plot of panel A comparing the protein concentration dependence of the rate constants of ThT and the first phase of CD. The straight line fit to the ThT data represents two slopes obtained at low protein concentration (< 10 μM) and high protein concentration (>15 μM) with a value of 0.076 $min^{-1}/μM$ and 0.041 $min^{-1}/μM$, respectively. The CD data fits to a straight line with a slope of 0.02 $min^{-1}/μM$. (B) The relative amplitude of the first phase (red triangles) and the second phase (inverted red triangles) of CD and the first phase (yellow squares) and the second phase (yellow diamonds) of DLS are plotted as a function of the concentration of protein. The inset of panel B shows size exclusion*

chromatogram of 10 µM protein under native condition (solid grey line) and under aggregation condition at the end of the first (dashed green line) and the second phase (solid green line) of the double exponential kinetics. The error bar represents the spread in data points from two separate experiments.

4.3.7 Dependence of the apparent rate constants and relative amplitudes on the protein concentration

Figure 4.4A shows the apparent rate constants obtained from ThT-, CD- and R_H-monitored kinetics plotted against the protein concentration. The process of elongation monitored by ThT shows maximum dependence on the protein concentration resulting in a 24 fold change in the rate constant when the protein concentration is changed from 1.5 µM to 25 µM (Figure 4.4A). This change in the rate constant as a function of protein concentration shows two slopes- one at low and one at high protein concentration (Figure 4.4A; inset), indicating the formation of different transition states at low and high protein concentration. The secondary structural transformation measured by the first phase of CD depends marginally upon the protein concentration (Figure 4.4A) and changes only 3 fold in the rate constant when the protein concentration is changed from 3 µM to 25 µM. We observe that the first phase of CD occurs faster than ThT at low protein concentration (<10 µM), while at high protein concentration (>10 µM), the process of ThT precedes the first phase of CD. Therefore, at low protein concentration, elongation is rate-limiting, while at high protein concentration, the process of conformational conversion becomes rate-limiting. The second phase of CD seems to be independent of the change in the protein concentration with negligible change in the rate constant. The apparent rate constant of the R_H- monitored kinetics shows the first phase to be changing along with the change in the first phase of CD or ThT monitored kinetics. The slowest change in the entire process is the second phase of the R_H- monitored kinetics, which occurs slower than the second phase of the CD. The apparent rate constant of the slow phase of R_H-monitored kinetics changes 10 fold when the protein concentration is changed from 5 µM to 25 µM. However, the rate of the slow phase of R_H- monitored kinetics seems to saturate at high protein concentrations.

Interestingly, we observed that the relative amplitude of the first and the second phases of CD- and R_H- monitored kinetics does not change with the protein concentration (Figure 4.4B). In the CD- monitored kinetics, we observe that the amplitude of the first phase is 60%, while the amplitude of the second phase is 40%. However, when monitored using DLS, the amplitude of the first phase is 40%, while that of the second phase is 60%. The amplitude of

the CD and DLS remains more or less similar and does not change significantly with the change in the protein concentration. A similarity in the relative amplitude across the protein concentration indicates that a common species, termed higher order intermediates (HOI's), is formed at the end of the first phase at different protein concentrations, in which 60 % of the total expected change in CD and 40 % of the total expected change in R_H has occurred. The ThT monitored kinetics (Figure 4.3A) suggest that 100% of the total expected change has occurred in the HOI's formed at the end of the first phase. To characterize the size of the species formed during the biphasic transition, we performed SEC at the end of the first and the second phase using 10 µM protein concentration. Interestingly we observed a small peak at the end of the first phase (~ 18 minutes) whose elution volume corresponded to a size between the N form (12 mL) and the β form (8 mL). Although the peak seems broad, indicating heterogeneity, we approximated the size of the protein molecules at the maximum height and found it to correspond to ~70 kDa. At the end of the second step, we observed a complete shift in the peak to higher-sized β form eluting near the column's void volume.
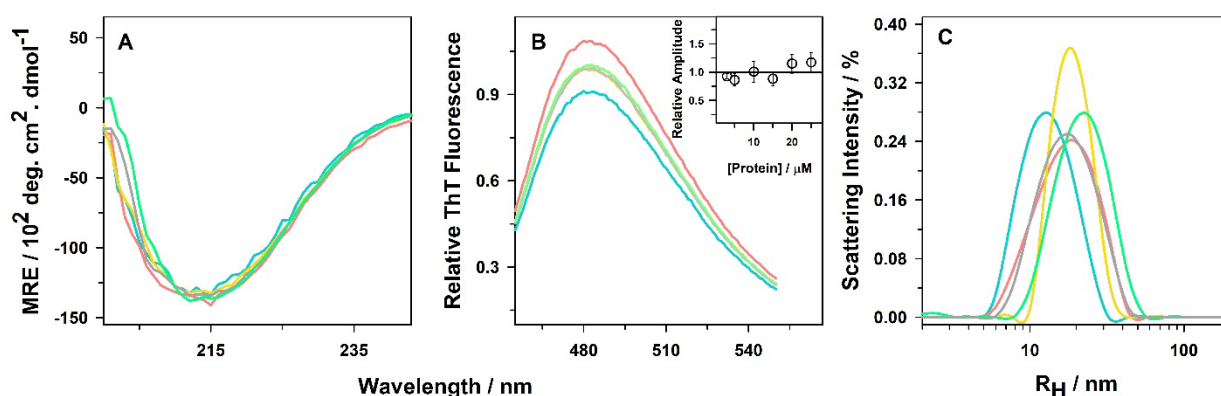
Figure 4.5: *Formation of similar type of β form at protein concentrations of 5 µM (blue), 10 µM (red), 15 µM (yellow), 20 µM (gray) and 25 µM (green). (A) The secondary structure of the β form at different protein concentrations was monitored using the far-UV CD. (B) ThT fluorescence scan of the β forms at different protein concentrations. The inset of panel B shows the relative amplitude of the ThT signal at 482 nm across different protein concentrations. (C) R_H distribution of the β forms across different protein concentrations.*

4.3.8 β form formed at different protein concentrations are similar

In order to study whether the structure of the β form is similar across the protein concentration, we monitored the equilibrium structure of the β form using CD, ThT and DLS

(Figure 4.5). We observed that the final structure of amyloid-like assembly is very similar at all the protein concentrations with respect to its secondary structure conformation measured using CD and ThT binding (Figure 4.5A, 4.5B and its inset). The R_H distribution of the β form across the protein concentration appears comparable. However, the β form is heterogeneous, as confirmed by the width of the R_H distribution (Figure 4.5C).
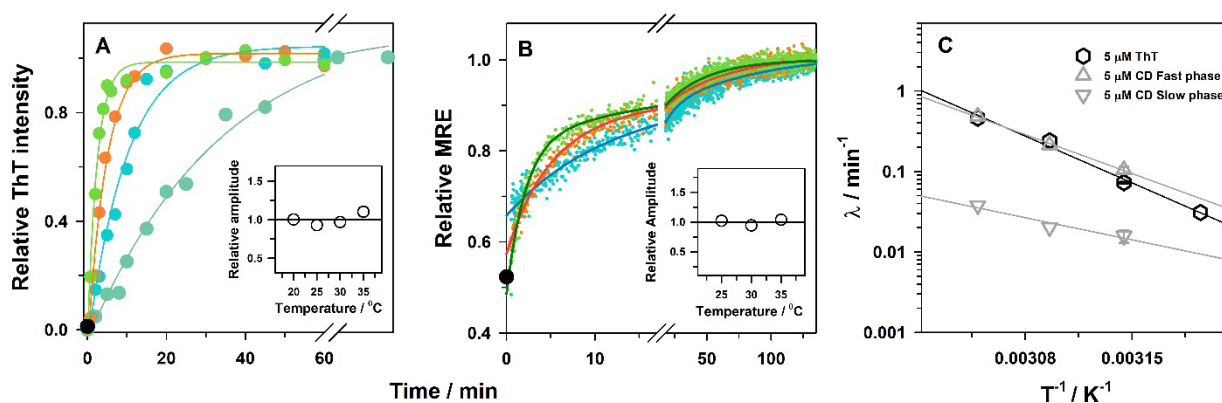
Figure 4.6: *Dependence on the temperature of the kinetics of the formation of the β form. Protein (5 µM) was heated in aggregation buffer at four temperatures – 20˚C (teal), 25 ˚C (blue), 30 ˚C (orange), 35 ˚C (green) to monitor the change in the elongation and conformational conversion process. (A) Relative change in the elongation monitored by ThT fluorescence at 482 nm. The inset of panel A shows the relative amplitude of the ThT signal across the different temperatures. (B) Relative change in the conformation of the protein as monitored by CD kinetics at 216 nm. The inset of panel B shows the relative change in the CD signal of the β form across different temperatures. (C) Arrhenius plot: the observed rate constants k obtained from panels A and B are plotted against 1/T. The linear fit to the data gives a slope representing the activation energy of the elongation process and the conformational conversion process. The continuous line through the data points in panel A fits a single exponential equation, and panel B fits a double exponential equation. The filled circle in panels A and B is the signal of the N form before transferring the protein into aggregation condition.*

4.3.9 Temperature dependence of the kinetics of the formation of β form

Figure 4.6A shows the kinetics of the formation of β form at 5 µM protein concentration using ThT fluorescence at four different temperatures, ranging from 20˚C-35˚C. We observed that at each temperature, the ThT monitored kinetics fits best to a single-exponential equation. The rate of elongation increases with increasing temperature. The amplitude achieved is similar across the temperature range, indicating the formation of β forms with similar ThT binding capacity (Figure 4.6A; inset). The change in the conformation of the protein was also monitored

through CD kinetics, and we observed that the change in the conformation occurs in a biphasic manner (Figure 4.6B). The total amplitude at different temperatures, however, remains similar (Figure 4.6B; inset). The plot of rate constant against 1/T yielded a linear Arrhenius plot (Figure 4.6C) with a slope corresponding to the activation energy of the process being monitored. The process of elongation has an activation energy of 158 kJ/mol⁻¹. In comparison, the first phase of conformational conversion has an activation energy of 132 kJ/mol⁻¹, and the second phase has an activation energy of 220 kJ/mol⁻¹. These data suggest that the N to β transformation involves multiple transition states with different barrier heights.
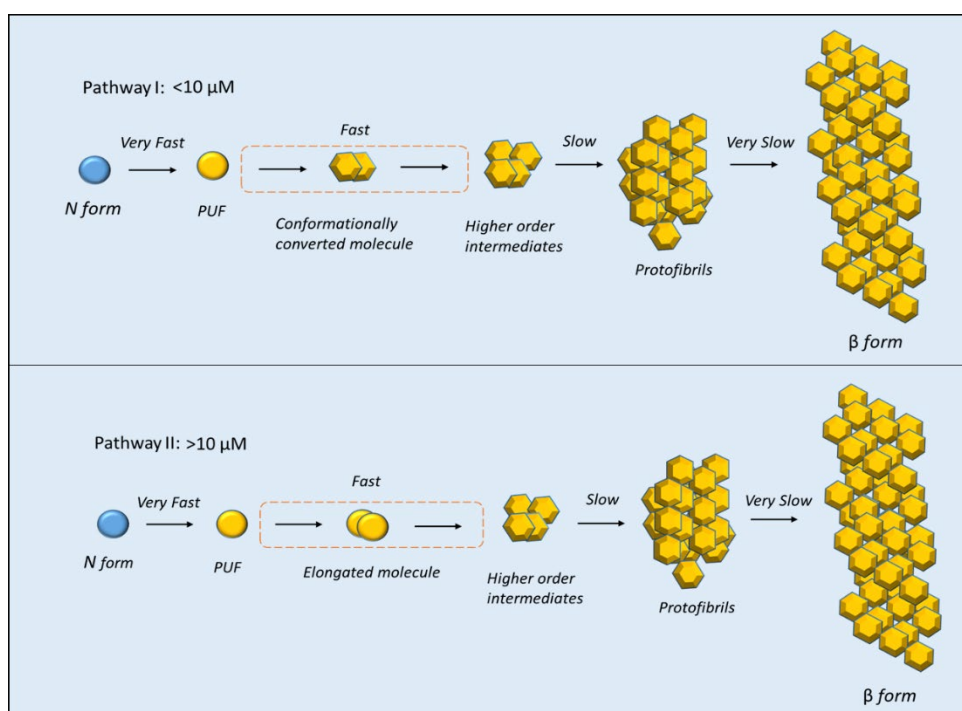
Figure 4.7: *Mechanism of the transformation of the N form to β form at low and high protein concentration. The aggregation process occurs in four distinct steps- very fast, fast, slow and very slow. In the 'very fast' step, the N form partially unfolds to form PUF. At low protein concentration, these molecules undergo a conformational conversion reaction to form conformationally converted molecules, followed by elongation and oligomerization, resulting in HOI formation in the 'fast' step. However, at high protein concentration, the PUF first undergoes an elongation process to form elongated molecules, followed by conformational conversion and oligomerization to form HOI in the 'fast' step. The evolution of the molecules from HOI to protofibrils in the 'slow' step and from the protofibrils to the β form in the 'very slow' step remains the same under both low and high protein concentrations.*

4.4 Discussion

In this study, we have employed a range of probes to kinetically map the sequence of events that characterize the TDP-43[tRRM] amyloidogenesis. We found that the process of TDP-43[tRRM] aggregation is a multi-step transition from a monomer to an amyloid-like fibrillar species. The transition has been monitored by measuring ellipticity at 216 nm, apparent hydrodynamic radius using DLS, thioflavin T and Trp fluorescence. Monitoring the same transition using multiple probes helped us to delineate the different steps involved during the aggregation process as has been characterized for various proteins.[40, 41]

Equilibrium experiments reveal a multi-step structural change associated with N to β transition

TDP-43[tRRM] undergoes a transition from a native monomeric structure to a β sheet-rich amyloid-like assembly at low pH and physiological ionic strength. When monitored under equilibrium conditions, the transition curve reveals the stability of the different structural elements of the protein. As the protein undergoes a structural transition, the very first signal change is observed in the Trp fluorescence, which decreases as the pH is decreased (Figure 4.1F; inset). The midpoint of the Trp fluorescence transition is around pH 5 (Figure 4.1H). The decrease in the Trp fluorescence might result from altered side-chain packing coupled with quenching by nearby charged residues upon protonation. Within 3 Å distance from W113, there are H143, K145, and K114 situated, while from W172, there is an H166 located, whose protonation might result in quenching of the Trp fluorescence as the pH is decreased.[42] We also observe a 2 nm blue shift in the λ_{max} of the Trp residues in the β form (Figure 4.1F), indicating that the Trp residues are not exposed to the polar environment. Additionally, we see an increased anisotropy in the β form, indicating restricted movement of the Trp residues in the β form (Figure 4.1G). These observations suggest that the Trp residues are located in a relatively hydrophobic environment and may form a part of the core residues in the β form. Although the change in the anisotropy and the ThT fluorescence shows a similar midpoint of transition (pH 4.5) but the sigmoidal curves representing these processes are non-synchronous in nature (Figure 4.1H; inset). The N to β transition monitored using CD and Trp fluorescence are also non-coincident (Figure 4.1H). Between pH 7.5 and pH 5, the protein undergoes a tertiary structure change resulting in decreased fluorescence and increased anisotropy; however, the secondary structure remains native-like with no presence of ThT binding pockets (Figure 4.1H). This suggests destabilization of the native state and the formation of partially unfolded-

like structures within the pH range of 7.5 to 5. As we increase the proton concentration further, the protein undergoes a change in its secondary structure and transitions to form a cross-β sheet-rich structure as indicated by the ThT binding (Figure 4.1C; inset). Different structural elements of the protein thus have different stability and therefore undergo structural transformation asynchronously. The non-coincidental sigmoidal curves of the different processes indicate that the transition occurs in a multiphasic manner under equilibrium conditions. Therefore, it would be interesting to study the process kinetically and delineate the temporal order of events under the aggregation condition.

Monomeric to amyloid-like transition occurs in four distinct stages

The aggregation process was monitored kinetically using different probes such as Trp and ThT fluorescence, CD and DLS (Figure 4.2) at a defined protein concentration (10 µM). When probed with time, the data reveals that the process of aggregation occurs in four distinct steps – very fast, fast, slow and very slow step. The very fast step corresponds to the decrease in the Trp fluorescence that occurs within the dead time of the experimental measurement (Figure 4.2E). This decrease in the fluorescence due to the formation of the partially unfolded form (PUF) occurs as a result of the rearrangement of the side-chain residues giving rise to the first step of the amyloid-like transition (Figure 4.7). Interestingly, we observe that the process of elongation monitored by ThT fluorescence occurs in single step while the process of conformational change monitored by CD and mean R_H change monitored by DLS occurs in two steps (Figure 4.2A, 4.2B, and 4.2D). At 10 µM, the first step of the CD, the change in the ThT fluorescence and the first step of the DLS occurs at a comparable rate (0.1229 ± 0.0414 min^{-1}) and therefore can be grouped to account for the fast step of the aggregation kinetics (Figure 4.2F). Accordingly, the PUF's undergo a transformation in its secondary structure, resulting in increased ThT binding and is accompanied by the association of the protein molecules. This results in the formation of oligomers with higher mean R_H called the HOI's (Figure 4.7). Further reorganization of the HOI's brings about a change in the chirality of the protein assembly without creating any new ThT binding pockets. These changes contribute to the slow step of the overall transition resulting in the formation of protofibrils (Figure 4.7). The slowest step of the aggregation process is the increase in the mean R_H of the protein molecules without any further change in the chirality of the structure or formation of any new ThT binding sites (Figure 4.7). This suggests that the slowest step might simply be a lateral association of

protofibrils leading to a gradual increase in the mean R_H of the protein assembly. This state of the assembly finally formed as a result of lateral association is called as the β form. The rate of the slowest step is approximately 2000 fold slower than the fastest step of the aggregation process (Figure 4.2F). Kinetic analysis thus reveals that the monomeric to amyloid-like transition can be grouped into four distinct steps helping to build a simple mechanistic model for the N to β transition (Figure 4.7).

Initial species are different at low and high protein concentration

The dependence of the apparent rate constant as a function of protein concentration suggested some additional features to the model. The monophasic nature of the ThT kinetics was maintained across the protein concentration range studied (Figure 4.3A). The rate of the elongation appeared to be strongly dependent on the protein concentration. When the rate was plotted as a function of protein concentration, there appeared to be two slopes – one at < 10 µM and the other at > 10 µM (Figure 4.4A; inset). The strong dependence of the rate is manifested by the steep nature of the two slopes when plotted as a function of the protein concentration (Figure 4.4A; inset). The two slopes also point out the different transition state corresponding to the different energy barrier associated with the process.

The conformational change probed by the CD shows biphasic behavior across the protein concentration range studied (Figure 4.3B). In contrast to the ThT rate, the rates of the first step of CD show a very weak dependence on the protein concentration. This is reflected by the shallow nature of the slope (Figure 4.4A; inset). It, therefore, appears that at low protein concentration, the process of conformational conversion precedes the elongation process (Figure 4.4A), giving rise to conformationally converted molecules (Figure 4.7). While at high protein concentration, the process of elongation precedes conformational conversion (Figure 4.4A; inset), giving rise to the elongated molecules (Figure 4.7). Consequently, the elongation process becomes rate-limiting at low protein concentration, while the process of conformational conversion becomes rate-limiting at high protein concentration (Figure 4.4A). Such a strong and a weak dependence of the apparent rate constants of ThT and the first step of CD, respectively, on the protein concentration, modulates the structural characteristics of the protein molecules formed at an early stage. Consequently, our data indicate that there are two pathways to reach the β form and is illustrated in two separate panels of the proposed model (Figure 4.7).

Common species- HOI's, protofibrils and the β form

The relative amplitude of the two phases of the CD and the DLS does not change significantly with the change in the protein concentration (Figure 4.4B), suggesting that the spectroscopic properties of the HOI's formed at the end of the first phase of CD and DLS are similar. Upon studying the structural characteristics of these species, the size of HOI appears to be around 10-12 nm (Figure 4.2D and 4.3C). The peak seems broad at the end of the first phase of DLS (Figure 4.2C), signifying heterogeneity in the size of the species. This heterogeneity can also be seen in SEC, where the elution of the protein at the end of the first phase occurs at a volume whose maximum absorbance corresponds to trimer-tetramer species (~70 kDa) (Figure 4.4B; inset). The size of these species is almost 4-fold higher than the monomeric N form, which measures around 19 kDa.[27] The HOI's further undergo structural change as monitored by the second phase of CD, giving rise to protofibrils (Figure 4.4A). The association of protofibrils to form a larger structure termed the β form results in the elution of the protein assemblies near the void volume in SEC at the end of the second phase (Figure 4.4B; inset). However, the β form has an R_H of around 20 nm when monitored at the end of 8 h using DLS (Figure 4.5C). Although the protein concentration dictates the type of species formed initially but the structural characteristics of the final fibril, β form, remain similar across the protein concentration (Figure 4.5). The β form has a similar ThT binding capacity and secondary structure indicative of a similar structure at the molecular level (Figure 4.5). When imaged, the β form appears to be small curly amyloid-like assemblies (Figure 4.1E).

Biphasic CD implies a linear structural change

The biphasic nature of the CD invokes an alternative possibility of two different species undergoing a conformational change in the two phases of 60:40 amplitude. For example, we can assume that there are two types of protein molecules in PUF following the very fast phase to begin with – one which undergoes the formation of β form at a fast rate resulting in the first step of the CD and the other which undergoes the formation of β form at a slow rate resulting in the second step of the CD. The likelihood of this possibility was considered by passing the protein through SEC column at the end of the first and the second phase (Figure 4.4B; inset). If this argument holds, then we should have observed two peaks at the end of the first phase of the CD- 40% of the population should correspond to a partially unfolded molecule whose size would be somewhere between the monomeric and the β form and 60% of the population should

correspond to size of the β form. However, we detect only one peak at the end of the first phase, which completely shifts to the higher volume eluting near the void volume at the end of the second phase (Figure 4.4B; inset). Consequently, the possibility of the presence of two different species can be ruled out. The shift in the peak accordingly suggests that the population of the protein molecules changes in a linear fashion during the biphasic transition. A series of progressive reactions monitored using different probes thus helps understand the molecular-level changes that occur in the protein as they form the curly amyloid-like assembly. We have learned that the assembly reaction occurs in a progressive and linear fashion as has been studied for other proteins.[38, 41, 43]

Different transition states are observed along the reaction co-ordinate

The differences in the rate of the different processes are represented in the form of an activation barrier calculated for various steps (Figure 4.6). We do not see a significant difference in the barrier associated with the ThT and first step of the CD process, which sounds counter-intuitive as the different steps are expected to have different energies. But the initial steps are quite complex, and the overlapping rates of the initial process give rise to a similar activation barrier. However, the second phase of CD shows an energy barrier that is almost 100 kJ greater than that of the first phase of the CD. A higher activation barrier for the second step points to a slower reaction rate and validates the biphasic conversion reaction. These different energy barriers associated with different steps during the conversion points to a multi-step reaction process.

The N form transforms into the β form in a linear isodesmic polymerization mechanism

The transformation of N form into the β form does not show the presence of the lag phase when monitored using different probes (Figure 4.2). The absence of lag phase across the protein concentration range studied (Figure 4.3) suggests a linear isodesmic polymerization mechanism.[44, 45] Although many proteins have been shown to undergo a lag in their aggregation process, signifying high energy barriers in nucleus formation,[46-48], we do not see any nucleation reaction in our process. But protonation of critical residues and enough charge shield in the form of ionic strength may overcome the high energy barrier and evade the lag phase and, hence, the nucleus formation. This could mean that the initial partial unfolding reaction (Figure 4.2E) that occurs due to side-chain disruption resulting from the protonation of critical residues might be enough to overcome the energy barrier and trigger the process of aggregation. In our

previous study, we have shown how the ionic strength at low pH dictates the structural characteristics of the protein molecules resulting in the formation of distinct species. We see that at low pH and low salt concentration (< 5 mM KCl), partial unfolding occurs prior to the formation of native-like oligomers that show characteristics of wet molten globule.[27, 28] However, at a moderate salt concentration (~20-150 mM), these native-like oligomers are unstable and transform into amyloid-like species.[28] All these observations suggest that the ionic strength modulates the thermodynamic barrier of the early reaction and thus dictates the fate of the protein molecules. In this study, we observe a very fast partial unfolding reaction, but the presence of 150 mM KCl in the solution does not stabilize the formation of native-like oligomers and instead shifts the reaction equilibrium to form amyloid-like species in an ordered progressive manner. Such an isodesmic polymerization mechanism of aggregation formation has also been studied for other proteins such as β2-microglobulin, transthyretin, and amyloid-β. [38, 43, 49]

4.5 Conclusion

The present study provides an important insight into the process of formation of amyloid–like assembly for TDP-43[tRRM] protein. Our model suggests that there are four major steps characterized by different barrier heights- very fast, fast, slow and very slow. These different steps correspond to the formation of different species- PUF, HOI, protofibril and the amyloid-like assembly formed during the aggregation process. We do not observe any lag phase in any of the probes studied, suggesting that the reaction proceeds in a linear isodesmic fashion as observed for other proteins. Our experimental data recapitulate some of the vital characteristic features of different kinds of prefibrillar species formed during the aggregation process. An in vivo study isolating oligomers and aggregates from diseased brain tissues also showed the presence of heterogeneous oligomers with sizes ranging from 40-400 nm.[6] Notably, similar kinds of oligomers and intermediate species has been observed for various amyloidogenic proteins such as barstar, amyloid-β, tau, β2-microglobulin, α-synuclein.[40, 50-52] A comprehensive kinetic understanding thus reveals the structural and the conformational evolution of the protein molecules associated with the aggregation process and might facilitate to develop site-specific molecules to target the disease-relevant assemblies for therapeutic intervention.

4.6 References

(1) Chiti, F., and Dobson, C. M. (2017) Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Annu. Rev. Biochem.* 86, 27-68.

(2) Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A., and Radford, S. E. (2018) A new era for understanding amyloid structures and disease. *Nat. Rev. Mol. Cell Biol.* 19, 755-773.

(3) 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.

(4) 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, Article 25.

(5) 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.

(6) Fang, Y. S., Tsai, K. J., Chang, Y. J., Kao, P., Woods, R., Kuo, P. H., Wu, C. C., Liao, J. Y., Chou, S. C., Lin, V., Jin, L. W., Yuan, H. S., Cheng, I. H., Tu, P. H., and Chen, Y. R. (2014) Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* 5, Article 4824.

(7) Kao, P. F., Chen, Y. R., Liu, X. B., DeCarli, C., Seeley, W. W., and Jin, L. W. (2015) Detection of TDP-43 oligomers in frontotemporal lobar degeneration-TDP. *Ann. Neurol.* 78, 211-221.

(8) Maeda, S., Sahara, N., Saito, Y., Murayama, M., Yoshiike, Y., Kim, H., Miyasaka, T., Murayama, S., Ikai, A., and Takashima, A. (2007) Granular tau oligomers as intermediates of tau filaments. *Biochemistry* 46, 3856-3861.

(9) Caughey, B., and Lansbury, P. T. (2003) Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* 26, 267-298.

(10) Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campioni, S., Tzitzilonis, C., Soragni, A., Jessberger, S., Mira, H., Consiglio, A., Pham, E., Masliah, E., Gage, F. H., and Riek, R. (2011) In vivo demonstration that α-synuclein oligomers are toxic. *Proc. Natl. Acad. Sci. USA* 108, 4194-4199.

(11) Kayed, R., and Lasagna-Reeves, C. A. (2013) Molecular mechanisms of amyloid oligomers toxicity. *J. Alzheimers Dis.* 33, S67-S78.

(12) Guerrero-Munoz, M. J., Gerson, J., and Castillo-Carranza, D. L. (2015) Tau oligomers: The toxic player at synapses in alzheimer's disease. *Front. Cell. Neurosci.* 9, Article 464.

(13) Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* 300, 486-489.

(14) Kayed, R., Head, E., Sarsoza, F., Saing, T., Cotman, C. W., Necula, M., Margol, L., Wu, J., Breydo, L., Thompson, J. L., Rasool, S., Gurlo, T., Butler, P., and Glabe, C. G. (2007) Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. *Mol. Neurodegener.* 2, Article 18.

(15) Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C. M. (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. USA* 96, 3590-3594.

(16) Jahn, T. R., Makin, O. S., Morris, K. L., Marshall, K. E., Tian, P., Sikorski, P., and Serpell, L. C. (2010) The common architecture of cross-β amyloid. *J. Mol. Biol.* 395, 717-727.

(17) O'Nuallain, B., and Wetzel, R. (2002) Conformational Abs recognizing a generic amyloid fibril epitope. *Proc. Natl. Acad. Sci. USA* 99, 1485-1490.

(18) Glabe, C. G. (2006) Common mechanisms of amyloid oligomer pathogenesis in degenerative disease. *Neurobiol. Aging* 27, 570-575.

(19) Fiesel, F. C., and Kahle, P. J. (2011) TDP-43 and FUS/TLS: Cellular functions and implications for neurodegeneration. *FEBS J.* 278, 3550-3568.

(20) 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.

(21) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.

(22) Saini, A., and Chauhan, V. S. (2011) Delineation of the core aggregation sequences of TDP-43 C-terminal fragment. *ChemBioChem* 12, 2495-2501.

(23) Li, H. R., Chen, T. C., Hsiao, C. L., Shi, L., Chou, C. Y., and Huang, J. R. (2018) The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43. *Biochim. Biophys. Acta Proteins Proteom.* 1866, 214-223.

(24) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* 132, 1186-1187.

(25) Shenoy, J., El Mammeri, N., Dutour, A., Berbon, M., Saad, A., Lends, A., Morvan, E., Grelard, A., Lecomte, S., Kauffmann, B., Theillet, F. X., Habenstein, B., and Loquet, A. (2020) Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. *FEBS J.* 287, 2449-2467.

(26) Chang, C. K., Chiang, M. H., Toh, E. K., Chang, C. F., and Huang, T. H. (2013) Molecular mechanism of oxidation-induced TDP-43 RRM1 aggregation and loss of function. *FEBS Lett.* 587, 575-582.

(27) 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.

(28) 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.

(29) Zacco, E., Martina, S. R., Thorogate, R., and Pastore, A. (2018) The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.* 11, Article 372.

(30) Moreno, F., Rabinovici, G. D., Karydas, A., Miller, Z., Hsu, S. C., Legati, A., Fong, J., Schonhaut, D., Esselmann, H., Watson, C., Stephens, M. L., Kramer, J., Wiltfang, J., Seeley, W. W., Miller, B. L., Coppola, G., and Grinberg, L. T. (2015) A novel mutation P112H in the TARDBP gene associated with frontotemporal lobar degeneration without motor neuron disease and abundant neuritic amyloid plaques. *Acta Neuropathol. Commun.* 3, Article 19.

(31) Kabashi, E., Valdmanis, P. N., Dion, P., Spiegelman, D., McConkey, B. J., Velde, C. V., Bouchard, J. P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P. F., Camu, W., Meininger, V., Dupre, N., and Rouleau, G. A. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* 40, 572-574.

(32) Shodai, A., Ido, A., Fujiwara, N., Ayaki, T., Morimura, T., Oono, M., Uchida, T., Takahashi, R., Ito, H., and Urushitani, M. (2012) Conserved acidic amino acid residues in a second RNA recognition motif regulate assembly and function of TDP-43. *Plos One* 7, Article e52776.

(33) Kovacs, G. G., Murrell, J. R., Horvath, S., Haraszti, L., Majtenyi, K., Molnar, M. J., Budka, H., Ghetti, B., and Spina, S. (2009) TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea. *Mov. Disord.* 24, 1843-1847.

(34) Austin, J. A., Wright, G. S. A., Watanabe, S., Grossmann, J. G., Antonyuk, S. V., Yamanaka, K., and Hasnain, S. S. (2014) Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc. Natl. Acad. Sci. USA* 111, 4309-4314.

(35) Bulawa, C. E., Connelly, S., Devit, M., Wang, L., Weigel, C., Fleming, J. A., Packman, J., Powers, E. T., Wiseman, R. L., Foss, T. R., Wilson, I. A., Kelly, J. W., and Labaudiniere, R. (2012) Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl. Acad. Sci. USA* 109, 9629-9634.

(36) Guerrero-Munoz, M. J., Castillo-Carranza, D. L., and Kayed, R. (2014) Therapeutic approaches against common structural features of toxic oligomers shared by multiple amyloidogenic proteins. *Biochem. Pharmacol.* 88, 468-478.

(37) Jain, S., and Udgaonkar, J. B. (2008) Evidence for stepwise formation of amyloid fibrils by the mouse prion protein. *J. Mol. Biol.* 382, 1228-1241.

(38) Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005) Competing pathways determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol.* 351, 850-864.

(39) Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J., and Teplow, D. B. (1999) Aβ-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* 274, 25945-25952.

(40) Kumar, S., Mohanty, S. K., and Udgaonkar, J. B. (2007) Mechanism of formation of amyloid protofibrils of barstar from soluble oligomers: Evidence for multiple steps and lateral association coupled to conformational conversion. *J. Mol. Biol.* 367, 1186-1204.

(41) Faria, T. Q., Almeida, Z. L., Cruz, P. F., Jesus, C. S. H., Castanheira, P., and Brito, R. M. M. (2015) A look into amyloid formation by transthyretin: Aggregation pathway and a novel kinetic model. *Phys. Chem. Chem. Phys.* 17, 7255-7263.

(42) Patni, D., and Jha, S. K. (2021) Protonation-deprotonation switch controls the amyloid-like misfolding of nucleic-acid-binding domains of TDP-43. *J. Phys. Chem. B* 125, 8383-8394.

(43) Morel, B., Carrasco, M. P., Jurado, S., Marco, C., and Conejero-Lara, F. (2018) Dynamic micellar oligomers of amyloid beta peptides play a crucial role in their aggregation mechanisms. *Phys. Chem. Chem. Phys.* 20, 20597-20614.

(44) Frieden, C. (2007) Protein aggregation processes: In search of the mechanism. *Protein Sci.* 16, 2334-2344.

(45) Kumar, S., and Udgaonkar, J. B. (2010) Mechanisms of amyloid fibril formation by proteins. *Current Science* 98, 639-656.

(46) Wood, S. J., Wypych, J., Steavenson, S., Louis, J. C., Citron, M., and Biere, A. L. (1999) α-synuclein fibrillogenesis is nucleation-dependent - implications for the pathogenesis of parkinson's disease. *J. Biol. Chem.* 274, 19509-19512.

(47) Xue, W. F., Homans, S. W., and Radford, S. E. (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl. Acad. Sci. USA* 105, 8926-8931.

(48) Sabareesan, A. T., and Udgaonkar, J. B. (2014) Amyloid fibril formation by the chain b subunit of monellin occurs by a nucleation-dependent polymerization mechanism. *Biochemistry* 53, 1206-1217.

(49) Hurshman, A. R., White, J. T., Powers, E. T., and Kelly, J. W. (2004) Transthyretin aggregation under partially denaturing conditions is a downhill polymerization. *Biochemistry* 43, 7365-7381.

(50) Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., and Teplow, D. B. (1996) On the nucleation and growth of amyloid β-protein fibrils: Detection of nuclei and quantitation of rate constants. *Proc. Natl. Acad. Sci. USA* 93, 1125-1129.

(51) Alam, P., Bousset, L., Melki, R., and Otzen, D. E. (2019) α-synuclein oligomers and fibrils: A spectrum of species, a spectrum of toxicities. *J. Neurochem.* 150, 522-534.

(52) Ren, Y., and Sahara, N. (2013) Characteristics of tau oligomers. *Front. Neurol.* 4, Article 102.

Chapter 5

Electrostatic Effects of pathological mutation D169G and P112H on the amyloid-like assembly of nucleic acid binding domain of TDP-43

5.1 Introduction

Pathogenesis of TDP-43 protein is known to cause many neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS),[1, 2] frontotemporal lobar degeneration (FTLD),[1, 2] and Guam-Parkinsonism dementia,[3] Alzheimer's,[4] Parkinson's and Huntington's disease.[5] Deposition of TDP-43 in many such neurological diseases highlights a highly susceptible nature of the protein to undergo misfolding and aggregation. Structurally, TDP-43 is a multi-domain protein containing four major domains- the N-terminal domain, two RNA Recognition motifs termed together as tethered RRM domain (TDP-43^{tRRM}), and a C-terminal domain.[6, 7] TDP-43 performs vital functions inside the cell such as mRNA splicing, mRNA binding including auto-regulation of its own mRNA.[8, 9 10] TDP-43^{tRRM} carries out the major function of binding to nucleic acid through its highly conserved residues.[11] It also regulates the localization of the protein inside the cell through its nuclear localization signal (NLS) and nuclear export signal (NES),[12] impairment of which has been shown to cause accumulation of the protein predominantly in the cytoplasm under the diseased condition.[13] It is therefore comprehensible that any change within this domain might affect the overall function and localization and thus affect the proteostatic equilibrium within the cell. Various environmental factors such as concentration, crowding agent, pH, temperature, mutation, etc., regulate the proteostatic equilibrium inside the cell.[14, 15] In particular, many disease-associated mutations have been shown to predominantly shift the proteostatic balance, often resulting in aggregation and plaque formation within the cell.[16]

The TDP-43^{tRRM} domain, in particular, harbors important disease-associated point mutations- D169G and P112H.[17] Interestingly, both these mutations are electrostatic in nature. The D169G mutation has been observed in ALS patients[17], while the P112H mutation has been found in FTD cases.[18] The D169G mutation has been assessed for its role in stress granule formation.[19] Although the crystal structure of D169G-RRM1 does not show a significant change compared to TDP-43^{tRRM} protein[20], a recent study using solution NMR pointed out a substantial change in the dynamics of the residues clustered around the mutation site.[21] The mutation has also been shown to affect the ATP-binding capacity of the protein.[21] Additionally, it has been shown to slightly increase the thermal stability and has been proven to have a higher propensity to undergo proteolytic cleavage by caspase.[20] Such an increase in thermal stability has also been observed for P112H mutation, which has been shown to induce local conformational change at the site of mutation.[22] P112H mutation has been shown to affect the stacking interactions of W113 with nucleic acids, thus impairing the RNA-binding activity of

the protein.[22] Under neutral pH condition, both these mutations have been shown to decrease the aggregation reaction.[22, 23] These studies suggest several factors that may play a key role in accelerating the disease progression, such as post-translational modifications,[23] resistance to degradation due to increased stability,[23] or the impaired functional binding capacity of the protein.[21, 22] However it is essential to consider that these mutations are electrostatic in nature and therefore may contribute to specific electrostatic interactions that may govern the aggregation kinetics. A comprehensive study exploring the electrostatic effect of these mutations on the protein is missing. Therefore, in our study, we have modulated the electrostatic environment of the protein by changing the pH and the ionic strength in order to explore the role of these electrostatic mutations on the structure and aggregation reaction. In our previous study,[24], we showed how protonation-deprotonation and ionic strength regulate the assembly-disassembly of TDP-43[tRRM] protein molecules, suggesting an essential role of electrostatics in governing the conformation and the intermolecular interactions (including amyloid-like assemblies) of the protein molecules. It will thus be interesting to understand the magnitude of the electrostatic contribution of these mutations on the protein structure and the aggregation kinetics.

In our study, we found that the mutants D169G and P112H do not cause any global change in the conformation of the native structure of the protein. However, both the mutants show higher chemical stability compared to TDP-43[tRRM]. Interestingly, we observed that the mutant P112H shows a higher propensity to undergo aggregation at low pH, suggesting that the histidine may govern the protonation-deprotonation equilibrium which is linked to the assembly-disassembly equilibrium of the protein molecules. In contrast, the mutant D169G shows a comparable kinetic dependency as that of the TDP-43[tRRM] protein. We also observed that the mutant P112H shows a very weak dependency on the ionic strength, as compared to D169G and TDP-43[tRRM] protein molecule.

5.2 Materials and Methods

Protein purification

TDP-43[tRRM] protein has been purified using the protocol described in a previous study.[25] The purity of the protein was confirmed by SDS-PAGE, and the mass was determined to be 19,429 Da using electrospray ionization-mass spectrometry.

Site-directed mutagenesis

Point mutations TDP-43^{tRRM} P112H and TDP-43^{tRRM} D169G were generated using QuikChange site-directed mutagenesis kit (Qiagen) using wild-type plasmid as the template. The expression and purification for both the mutants were performed using the same protocol followed for TDP-43^{tRRM} protein.[25] Once purified, the mutants were run in SDS-PAGE to confirm the purity of the sample. A mass of 19,371 Da for D169G and 19,468 Da for P112H was determined using electrospray ionization-mass spectrometry.

Aggregation condition

Aggregation conditions used were similar to those described in a previous study (Chapter 4). In brief, the protein was concentrated between 600-800 μM and then transferred to 20 mM Gly-HCl pH 3 buffer containing 150 mM KCl. A final protein concentration of 10 μM was maintained for all the aggregation reactions.

pH-induced equilibrium experiments

Different buffers were used in different pH ranges as described in previous work (Chapter 2). All the samples were equilibrated for 24 h to achieve equilibrium before taking the spectroscopic measurements.

Circular Dichroism

Far-UV CD measurements were performed in a Jasco J-815 spectropolarimeter. Far-UV CD scan was collected between 200-250 nm using a 0.1 cm cuvette. For the kinetic experiment, the absorbance was measured at 216 nm at regular time intervals. The deadtime for the kinetic experiment was calculated by measuring the time difference between the addition of protein to the aggregation buffer and the very first signal measurement. Following settings were used for far-UV CD measurements- bandwidth- 2nm, DIT – 4s, and scan speed – 20 nm/s.

Thioflavin T (ThT) assay

The ThT assay was performed by adding 1 μM protein to the ThT assay buffer containing 50 mM Tris-HCl and 40 μM ThT dye. For the ThT scans, the sample was excited at 440 nm, and the emission was collected from 452 nm to 535 nm. For the kinetic experiment, the sample was excited at 440 nm, and the emission was collected at 482 nm for 20 s. The data

were then averaged and plotted as a function of time. The background ThT fluorescence contributed by the buffer was subtracted from all the data.

Trp Fluorescence and anisotropy

Protein samples (~ 10 µM) were incubated in pH 7.5 buffer, and the steady-state fluorescence was measured by exciting the Trp residues at 295 nm and measuring the fluorescence from 310 to 400 nm. For steady-state anisotropy measurements, the samples were excited at 295 nm and the anisotropy was measured at 340 nm for 300s, using an anti-photobleaching setting. The excitation and emission slit width was 2 nm and 5 nm, respectively.

Guanidinium Chloride (GdmCl) -induced equilibrium unfolding transition monitored by Trp fluorescence and far-UV CD

For the GdmCl-induced unfolding transition, protein samples (4-5 µM) were incubated in 20 mM MOPS pH 7.5 buffer containing 150 mM KCl. The samples were incubated overnight to achieve equilibrium. The Trp fluorescence was monitored by exciting at 295 nm and measuring the fluorescence at 340 nm. For far-UV CD measurement, the change in the absorbance was monitored at 222 nm. The concentration of the denaturant was measured using a refractometer (Abbe type refractometer from Rajdhani Scientific Instrument Co.).

Size Exclusion Chromatography

Size exclusion chromatography was performed for the native form of TDP-43^{tRRM}, D169G and P112H protein using GE Superdex 75 10/300 GL high-performance gel filtration column. The column was pre-equilibrated with 20 mM MOPS pH 7.5 buffer containing 150 mM KCl. Protein samples (~ 25 µM) were injected, and the sample was run at 0.8 mL/min flow rate.

Transmission electron microscopy

The amyloid-like assemblies of the mutants were placed on 300-mesh Formvar carbon-coated grid (Electron Microscopy Science) for 5 min. The excess sample was then removed, and the grid was then negatively stained using 2% uranyl acetate solution for 1.5 min. After the staining, the grid was rinsed with Milli-Q for a min and then air-dried before examining using transmission electron microscopy (Technai-T20) at an accelerating voltage of 200 kV.

5.3 Results

5.3.1 The native fold of the protein is maintained in D169G and P112H mutants

Site-directed mutagenesis was used to create single point mutations (D169G and P112H) within the TDP-43^{tRRM} protein. Both the mutations are located within the RRM1 domain of TDP-43^{tRRM} protein (Figure 5.1).
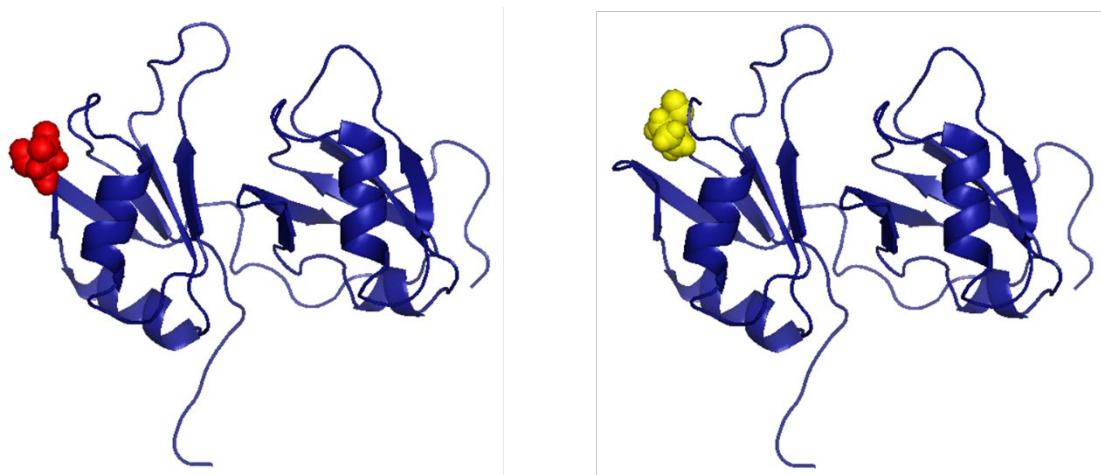
Figure 5.1: *PDB structure of the TDP-43^{tRRM} protein (4BS2). The mutant D169G is indicated by the red sphere, while the mutant P112H is indicated by the yellow sphere.*

To understand the effect of these mutations on the native state of the protein, we first compared the structure and the stability of the mutants to TDP-43^{tRRM} using local and global spectroscopic probes (Figure 5.2). The native state of TDP-43^{tRRM} has been structurally characterized previously and is used as a reference for comparison with the mutants' here.[25] Figure 5.2A shows the normalized fluorescence spectrum of the Trp residues (W113 and W172) in TDP-43^{tRRM} compared to D169G and P112H at a similar protein concentration. We observe that the local side-chain packing indicated by the Trp fluorescence remains similar across the three variants. In order to further investigate the solvation environment around the Trp residue across the three variants, we compared the maximum fluorescence emission, λ_{max}, of the fluorophore, which is an excellent measurement of the polarity of the surrounding medium. We observed that the λ_{max} for both the mutants, D169G and P112H, was 350 nm which is highly comparable to that of TDP-43^{tRRM} protein (349 ± 0.5 nm), suggesting that the Trp residues have a similar solvation environment in the mutants as compared to TDP-43^{tRRM} protein. We then compared the anisotropy of the Trp residues across the three variants to understand the local dynamics of the fluorophore within the protein molecules (Figure 5.2B).

The anisotropy of the Trp residues in D169G is 0.0814, and P112H is 0.0828, which appears to be marginally higher than TDP-43^{tRRM} (0.0750). This nominal increase in the anisotropy of the Trp might indicate a small increase in the local rigidity around the fluorophore that does not affect the solvation around the protein core. Thus, our data indicate that the overall side-chain packing remains intact, and the solvation of the fluorophore appears similar in the mutants, D169G and P112H and highly comparable to that of TDP-43^{tRRM} protein.
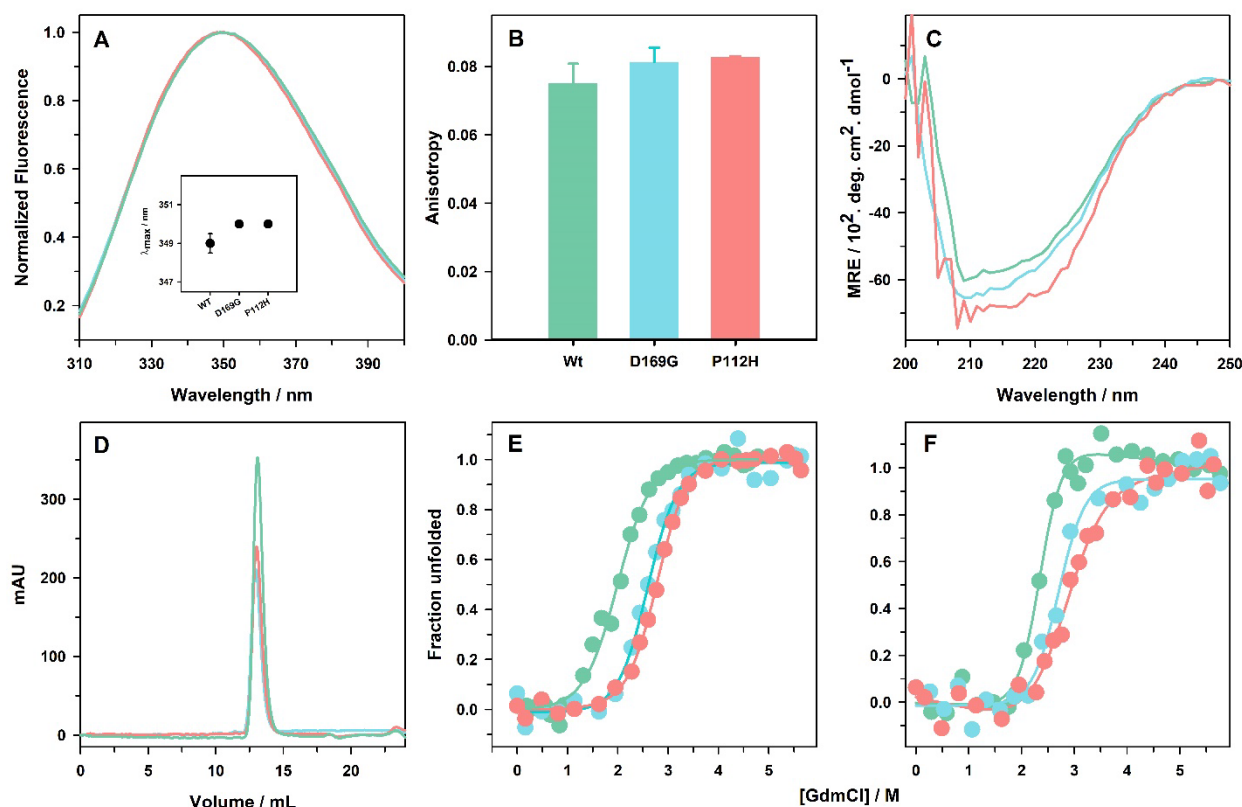
Figure 5.2: *Structure and stability comparison of the mutants, D169G and P112H, with TDP-43^{tRRM} protein. The native structure, size and stability of the mutants D169G (blue color) and P112H (red color) are compared with TDP-43^{tRRM} protein (cyan color) at pH 7.5. (A) Normalized fluorescence spectrum of all the mutants is compared with TDP-43^{tRRM} protein. The inset of panel A shows the maximum emission of the Trp residues across the three proteins. (B) Anisotropy of the Trp fluorophore of the mutants in comparison with TDP-43^{tRRM} protein. (C) Comparison of the far-UV CD spectrum of the mutants with TDP-43^{tRRM} protein. (D) Size exclusion chromatogram of the mutants compared with TDP-43^{tRRM} protein. Guanidinium Chloride (GdmCl)-induced equilibrium unfolding transition of the mutants is compared with TDP-43^{tRRM} by monitoring the change in the Trp fluorescence at 340 nm (Panel E) and far-UV CD at 222 nm (Panel F). The fraction unfolded is plotted as a function of denaturant concentration in panels E and F. The solid lines in panels E and F are to guide the eye.*

We next compared the global secondary structure of the mutants using the far-UV CD spectrum (Figure 5.2C). The mean residue ellipticity (MRE) at 222 nm for D169G was measured to be -5337 mdeg.cm^2.dmol^{-1} while for P112H was -6347 mdeg.cm^2.dmol^{-1}. This value is highly comparable to that of TDP-43^{tRRM} protein, which shows a similar α and β sheet content as that of the mutants. Upon comparing the size of the mutants using size exclusion chromatography, we observed that the mutants elute at the same retention volume as that of TDP-43^{tRRM} protein, whose size corresponds to ~19 kDa (Figure 5.2D). Overall, our data indicate that the mutants D169G and P112H do not significantly alter the local tertiary and the global secondary structure of the protein molecule.

5.3.2 The mutants show increased stability as compared to TDP-43^{tRRM} protein

The native state of the mutants appears similar to TDP-43^{tRRM} protein, but the chemical stability of the mutants appears to be different compared to TDP-43^{tRRM} protein (Figures 5.2E and 5.2F). Denaturant-induced unfolding transition of the mutants and TDP-43^{tRRM} were monitored by Trp fluorescence at 340 nm (Figure 5.2E) and far-UV CD at 222 nm (Figure 5.2F). We used the midpoint of the unfolding transition as measured by fluorescence and far-UV CD as a proxy for chemical stability. We observed that the midpoint of unfolding transition, when measured by Trp fluorescence for TDP-43^{tRRM} is 1.9M GdmCl while that for D169G and P112H are 2.6M GdmCl and 2.8M GdmCl, respectively. The midpoint of unfolding transition of secondary structure for TDP-43^{tRRM} is 2.3M GdmCl, while for D169G and P112H are 2.7M GdmCl and 3M GdmCl, respectively. Our data indicates that both secondary and tertiary structure stability is high for the mutants as compared to TDP-43^{tRRM}. A similar result has been observed for both the mutants using thermal stability, which concluded higher stability for the mutants as compared to TDP-43^{tRRM} protein.[22, 23]
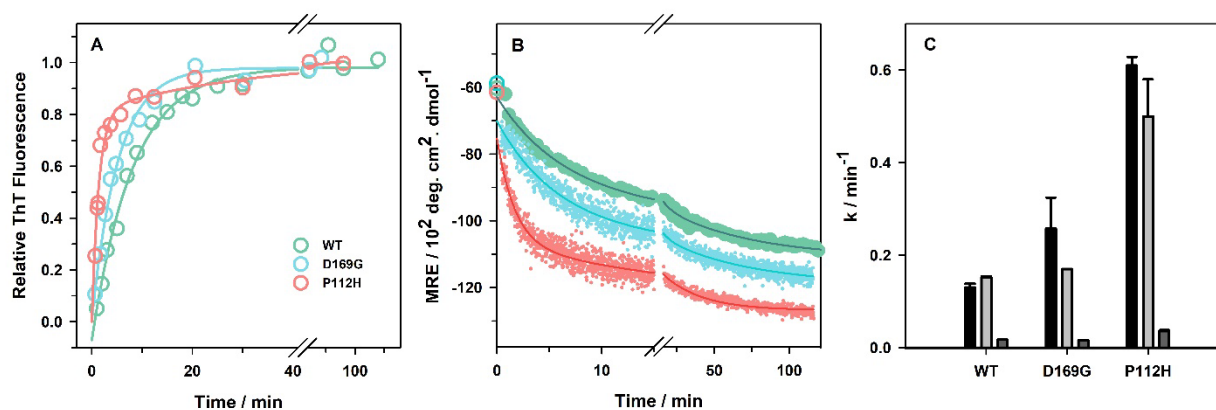
Figure 5.3: *Kinetics of the aggregation of the mutants compared to TDP-43^{tRRM}. (A) Elongation process is monitored by the change in the ThT fluorescence at 482 nm as a function of time. (B) Conformational change was monitored using far-UV CD at 216 nm as a function of time. The solid lines to panel A are fit to a single exponential equation, while the solid lines in panel B are fit to a double exponential equation. (C) Comparison of the rate of the elongation process, and the fast and the slow rate of the conformational change process for all the three protein variants (black bar- elongation, light gray- fast phase of CD and dark gray- slow phase of CD).*

5.3.3 P112H mutant shows increased rate of misfolding as compared to TDP-43^{tRRM}

We had observed in our previous study that TDP-43^{tRRM} undergoes a conformational change at low pH and physiological ionic strength to form ThT-positive β sheet-rich amyloid-like assemblies in multiple steps.[24, 25] We determined experimentally that TDP-43^{tRRM} undergoes a monophasic elongation reaction and a biphasic conformational change consisting of a fast and a slow phase (Chapter 4). The transition from the native state to the amyloid-like assemblies thus occurs in multiple steps (Chapter 4) with electrostatic interactions playing a major role in the formation of amyloid-like assembly. To understand how the electrostatic point mutations affect the kinetics of the aggregation reaction and to identify which steps in the reaction pathway are sensitive to the electrostatic effects, we compared the aggregation reaction of the mutants with TDP-43^{tRRM} under low pH and physiological ionic strength conditions.

Upon transferring the protein to the aggregation condition, we observed an increase in the ThT fluorescence at 482 nm. Kinetic analysis of the ThT fluorescence revealed an exponential change in the elongation process (Figure 5.3A). Both the mutants, D169G and P112H, showed an exponential dependence of the elongation process with the rates corresponding to 0.25 ± 0.06 min^{-1} and 0.61 ±0.018 min^{-1}, respectively (Figure 5.3C). The rate of elongation for D169G compares well with the rate observed for TDP-43^{tRRM} (0.13 ± 0.007 min^{-1}) in a previous study (Chapter 4). However, the P112H mutant shows an almost 4-fold

increase in the rate of elongation, suggesting a role for the charge mutation in accelerating the process of elongation.

The process of elongation is also accompanied by a change in the conformation of the protein molecules. We observed that under aggregation conditions, the protein molecules undergo conformational change resulting in an increased ellipticity as a function of time (Figure 5.3B). Kinetic data reveals a bi-exponential change in the conformation of the mutants as has been observed for TDP-43^{tRRM} previously (Chapter 4).[24] The rate of the fast phase for D169G is 0.17 min^{-1}, while that for P112H is 0.5 $\pm$ 0.08 min^{-1} (Figure 5.3C). The rate of the slow phase for D169G is 0.016 min^{-1}, while that for P112H is 0.036 $\pm$ 0.002 min^{-1} (Figure 5.3C). The bi-exponential rates for the mutant D169G compare well with TDP-43^{tRRM} (fast phase- 0.15 $\pm$ 0.0058 min^{-1} and slow phase- 0.017 $\pm$ 0.0012) rate (Chapter 4). However, the mutant P112H shows an almost 3-fold higher rate of the fast phase while a 2-fold higher rate for the slow phase of the CD compared to TDP-43^{tRRM}. All these data indicate that the removal of proline 112 and the introduction of the histidine side chain is substantially affecting the overall rate of the aggregation process.
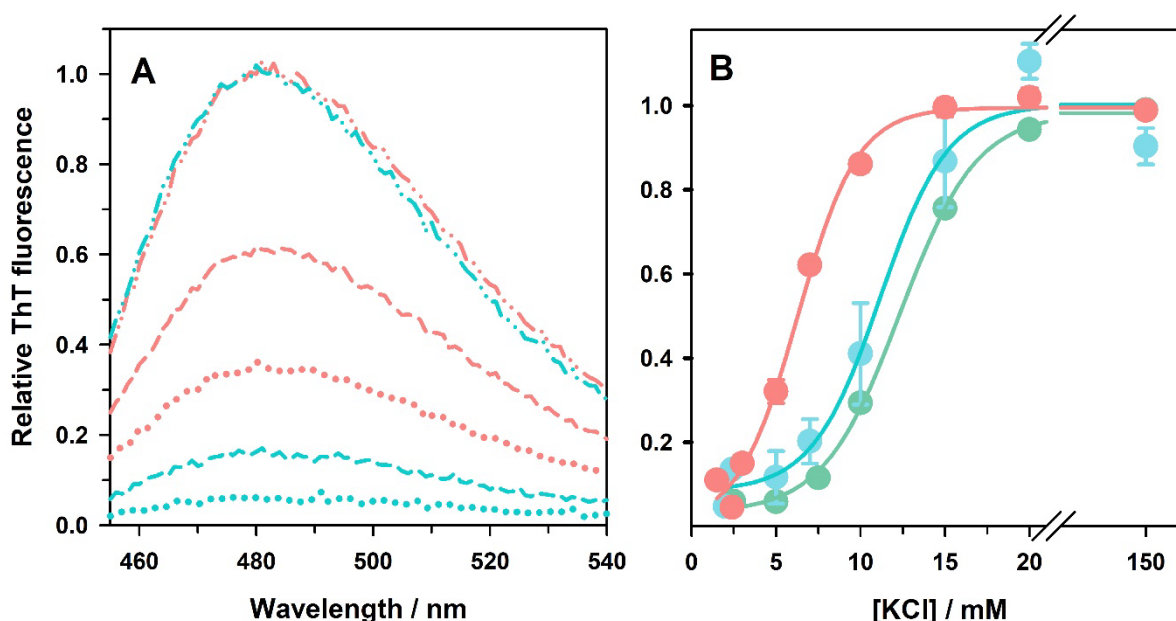
Figure 5.4: *Ionic strength dependency of the aggregation process of the mutants compared to TDP-43^{tRRM} protein. (A) ThT fluorescence scan of D169G (blue color) and P112H (red color) at different ionic strength – 5 mM (dotted), 7 mM (dashed lines) and 150 mM (dot-dot-dash). (B) Comparison of the ThT fluorescence at 482 nm at different ionic strengths under equilibrium conditions for TDP-43^{tRRM} (cyan color), D169G (blue color) and P112H (red color). The solid lines through the data are to guide the eyes.*

We understand that the electrostatic mutants have a different effect on the rate of the aggregation process. Thus in order to further investigate how the electrostatics modulate the intermolecular interactions, we exposed the mutants to a solution containing different ionic strengths at low pH and monitored the ThT fluorescence (Figure 5.4). We found that the native to amyloid-like transition is highly dependent on the ionic strength of the solution. The ThT fluorescence scan at 5 mM, 7 mM and 150 mM illustrate the difference in the ThT binding under equilibrium conditions for D169G and P112H mutant (Figure 5.4A). When monitored using a gradient of salt, we found that the mutant P112H shows a very weak dependency on the ionic strength compared to D169G (Figure 5.4B). The ionic strength dependency of D169G corresponds well with that of TDP-43^{tRRM} data (Figure 5.4B). It appears that at low ionic strength (< 5mM), the protein molecules do not bind to ThT dye, suggesting that the secondary structure remains native-like. Besides, the global secondary structure of TDP-43^{tRRM} in a previous study was shown to remain native-like at low pH and low ionic strength conditions (<5 mM). Our data suggests that the ionic strength at low pH plays a crucial role in modulating the electrostatic interactions within the protein molecules that might prime the protein to undergo aggregation.
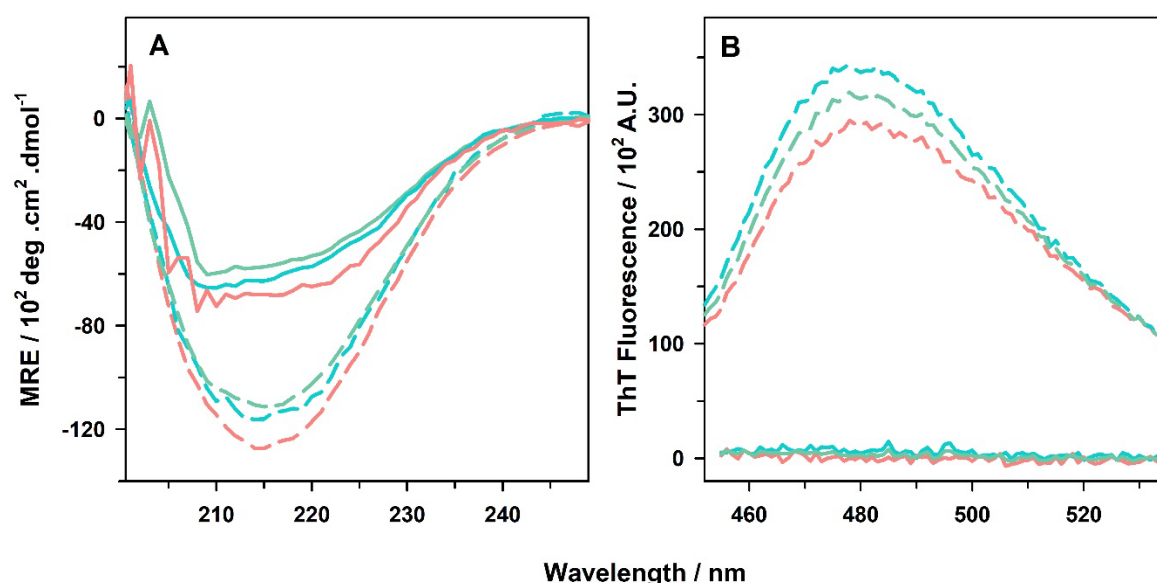
Figure 5.5: *Comparison of the amyloid-like assemblies (dashed lines) of D169G (blue), P112H (red) and TDP-43^{tRRM} (cyan) with their corresponding native structure (solid lines). (A) Comparison of the global secondary structure of the native form and the amyloid-like assemblies of D169G, P112H and TDP-43^{tRRM} using the far-UV CD. (B) ThT fluorescence scan of the native structure of D169G, P112H and TDP-43^{tRRM} compared with their corresponding amyloid-like assemblies.*

The above data suggest that P112H shows a higher rate of aggregation and weak ionic strength dependency as compared to TDP-43^{tRRM} and D169G mutant. However, we observe that the final amyloid-like assemblies formed across the mutants and TDP-43^{tRRM} are very similar. We characterized the structural properties of the amyloid-like assemblies by far-UV CD and ThT fluorescence (Figure 5.5). Figure 5.5A shows the global secondary structure of the amyloid-like assemblies of the mutants and TDP-43^{tRRM}. A representative far-UV CD spectrum of the protein molecule in the native state is shown. The far-UV CD spectrum of the amyloid-like assemblies shows increased ellipticity and a significant shifting of the overall peak towards 216 nm, implying the formation of an ordered β-sheet rich structure. The global secondary structure is almost similar across all the three variants pointing out the similarity in the overall arrangement of the cross-β sheet structure within the assemblies. Figure 5.5B shows the ThT fluorescence of the native states of the mutant and the TDP-43^{tRRM} compared with the ThT fluorescence of their corresponding amyloid-like assemblies under aggregation conditions. The ThT fluorescence of the native forms of all the variants is negligible, suggesting that the native structure does not bind to the ThT dye and does not have any amyloid-like characteristics. In contrast, the ThT fluorescence of the amyloid-like assemblies of the mutants and the TDP-43^{tRRM} is high, suggesting that the dye is able to stack within the cross-β sheet structure of the amyloid-like assemblies. Interestingly, the ThT fluorescence is highly comparable between amyloid-like assemblies of TDP-43^{tRRM} and the mutants, suggesting a similar internal structure for the assemblies. Although the reaction rate for the P112H mutant varies considerably compared to TDP-43^{tRRM} and the D169G mutant, the final amyloid-like assemblies formed at the end of the aggregation reaction seem to achieve a similar ThT and CD signal.
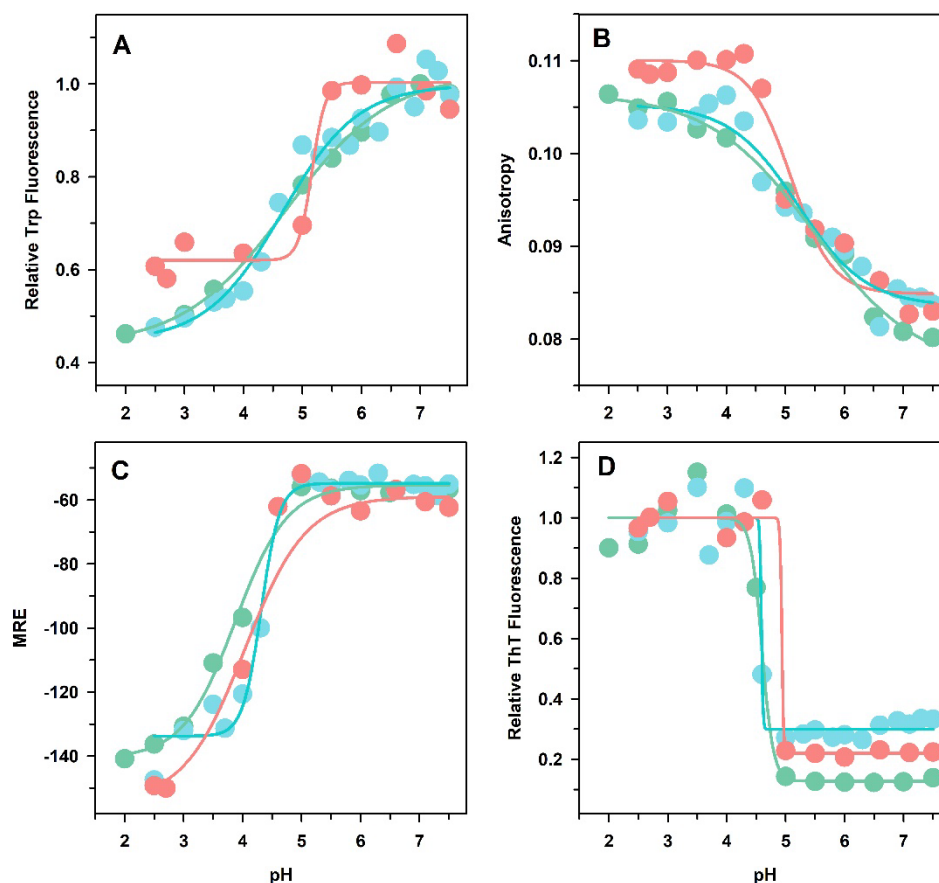
Figure 5.6: *pH dependence of the mutants, D169G and P112H under equilibrium conditions. Protein (10µM) was incubated in different pH buffers containing 150 mM KCl and monitored using Trp fluorescence (panel A), Trp anisotropy (panel B), far-UV CD (panel C) and ThT fluorescence (panel D). The cyan dots indicate TDP-43^{tRRM} protein; the blue dots indicate D169G, while the red dots correspond to P112H protein. The solid lines in all the panels are to guide the eye. The data for the TDP-43^{tRRM} has been taken from a previous study.*

5.3.4 pH- induced equilibrium transition from native to β sheet structure at 150 mM KCl

The mutants have been shown to undergo structural and conformational changes under low pH and physiological ionic strength. Thus we identify that pH and salt play a crucial role in governing the intermolecular interactions that regulate the structure of the mutants. In order to further investigate the effect of these electrostatic mutations on the pH-dependent misfolding transition, we monitored the misfolding transition under equilibrium conditions using different probes such as Trp fluorescence, Trp anisotropy, CD and ThT fluorescence (Figure 5.6). A previous study[26] from our lab suggested that for TDP-43^{tRRM}, the midpoint of misfolding transition is around pH 4, which can also be seen in the pH-induced misfolding of TDP-43^{tRRM} in figure 5.6. We then compared the pH dependency of the mutants under constant

physiological ionic strength. We observed a sigmoidal change across the three different protein variants for all the probes. Between pH 7.5 and pH 5, TDP-43[tRRM] and the mutant D169G show a gradual decrease in the Trp fluorescence (Figure 5.6A) with a corresponding increase in the anisotropy (Figure 5.6B). The CD signal for TDP-43[tRRM], D169G and P112H remains native-like (Figure 5.6C) with a negligible increase in ThT fluorescence (Figure 5.6D). In contrast, the Trp fluorescence of P112H remains similar between pH 5 and pH 7.5, while the anisotropy increases in accordance with the TDP-43[tRRM] and the D169G mutant. The sigmoidal transition observed for all the probes saturates at around pH 3.5. Between pH 3.5 and pH 2, the protein molecules show decreased Trp fluorescence (Figure 5.6A), increased anisotropy (Figure 5.6B), and higher ellipticity value with the ability to bind to ThT dye. This shows that the protein molecules have now transformed into cross-β sheet rich structure. Our previous study has shown that TDP-43[tRRM] protein undergoes a transformation into cross-β sheet amyloid-like structure at low pH.[24] This transition from the native state to β sheet rich amyloid-like structure is also seen for the mutants, D169G and P112H. If this transition occurred due to the protonation of D169, then replacement with glycine (an amino acid with a neutral side-chain) would eliminate the transition completely. However, the native to amyloid-like transition remains similar for the D169G mutant suggesting that the protonation of Asp plays no role in this misfolding transition. In contrast, for the P112H mutant, we see that the transition has become very sharp and distinct, suggesting a very critical dependence on the protonation of histidine in this transition.

5.4 Discussion

Mutations have been studied to show varied effects on the structure, stability and function of the protein molecules. Some studies exploring the C-terminal domain mutations within the TDP-43 protein have shown to aggravate the aggregation kinetics of the protein, thus causing a deleterious effect.[27] While other mutations have been shown to have nearly no effect to even causing a reduction in the aggregation propensity.[22, 28] The effect of mutations on the protein molecule is thus dependent on many factors such as the charge of the mutated residue, size of the residue,[29] sequence in which the mutation occurs which in turn might bring about various changes in the form of altered dynamics, stabilization or destabilization effect,[30, 31] change in the hydrophobicity or by simply affecting the transition state of the reaction.[32]

Here, in our study, we observed that the pathological mutations D169G and P112H do not significantly affect the native fold of the protein, as has been noted in a previous study.[22,

[23] We also observe increased chemical stability which is in line with the increased thermal stability observed for these mutants in the literature.[22, 23] A general accepted hypothesis is that destabilization of the native state often increases the formation of the partially folded molecules which are linked to an increased rate of misfolding and aggregation.[33] Yet, this is not adequate to explain the effect of all mutations as they might also exert an effect on the properties of partially misfolded, oligomeric and aggregation-prone species.[30, 34] Therefore, it is essential to clarify the role of mutations in exerting their effect on the distinct steps in the process of misfolding and aggregation.

Prior study has shown that TDP-43[tRRM] undergoes the formation of amyloid-like assemblies under low pH and physiological ionic strength in a multi-step pathway (Chapter 4). Upon examining the effect of these charged mutations on the distinct steps in the pathway, we observed an exponential dependence for the elongation and conformational change processes (Figure 5.3). The P112H mutation shows an almost 4-fold increase in the rate of elongation and a 3-fold increase in the rate of conformational change (Figure 5.3C), suggesting that the change in the electrostatics owing to the presence of protonated His brings about critical change within the electrostatic network of the protein molecules. The synchronized effect of protonation and salt plays a key role in lowering the activation barrier associated with elongation and conformational change processes, implying how important the electrostatics are in determining the transition state of the reaction. For the D169G mutant, the rate of these processes is highly comparable to that of TDP-43[tRRM] (Figure 5.3). The glycine present at 169 position with a neutral side chain does not undergo any protonation reaction and hence behave similar to TDP-43[tRRM] protein. Additionally, we can also point out that the removal of Asp at 169 position doesn't abate the native to amyloid-like transition suggesting that the role of Asp in controlling the electrostatics of the amyloid-like interactions is inconsequential.

In our previous work we showed how the protonation of key residues within the TDP-43[tRRM] protein causes destabilization of the protein molecules resulting in loosening of the side-chain packing and formation of partially unfolded molecules.[24] At low pH, TDP-43[tRRM] shows a sigmoidal dependency on the ionic strength (Figure 5.4).[24] Under the low salt condition, we initially observed the formation of molten globule-like (MG-like) species. With time, these MG-like species associate through weak interactions to form native-like oligomers. These weak interactions govern the metastable nature of the native-like oligomers. Presence of high salt (>15 mM KCl) however disrupts these weak interactions and favors the formation of highly structured cross-β sheet rich amyloid-like assemblies. The ionic strength dependency at low

pH thus regulates the size, conformation and the amyloidogenicity of the protein assemblies. Remarkably, we noted a similar dependency on the ionic strength for the D169G mutant as that of TDP-43^{tRRM} (Figure 5.4). The glycine present at 169 position with a neutral side chain does not bring about any significant change in the electrostatic network within the protein molecule and hence shows a similar ionic strength dependency. In contrast, the P112H mutant shows a very weak dependency on the ionic strength (Figure 5.4). This can be linked to the presence of protonated His at the 112 position whose electrostatic interactions with other charged residues such as D169, R171, E117, Q118 and D119 in the spatial proximity might favor the formation of amyloid-like interactions in the protein molecules. Additionally, the protonation of His might also weaken the unfavorable interactions that govern the initial structure and conformation of the protein, resulting in less amount of salt required to break those interactions and allowing the protein molecules to come together to form amyloid-like assemblies. This suggests that ionic strength plays a critical role in governing the intermolecular interactions between the protein molecules, thereby dictating the structure and the conformation of the protein molecules. The altered electrostatic effect manifested through an increased rate of aggregation at low pH suggests that the protonation of His at 112 position coupled with salt effect plays a critical role in regulating the amyloid-like interactions around this residue. Our data thus allows us to imply that the critical interactions that initiate the amyloid-like assembly in TDP-43^{tRRM} lie in the vicinity of P112 residue and are strongly influenced by the electrostatic environment around the residue.

Broadly, many pathogenic mutations in aggregation-prone proteins such as prion,[35, 36] β2 microglobulin, α-synuclein[37] and Aβ42[29] involve charge reversal or substitution of charge residue with uncharged residues, implying the significance of the electrostatic effect in the aggregation process. In this study, we tried to understand the consequence of charge mutations on the structure and aggregation kinetics of the protein molecules. By identifying the critical dependency of the charge and protonation state of the mutations we can fine tune the overall electrostatic interactions to modulate the aggregation timescale.

5.5 References

(1) 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.

(2) 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.

(3) Hasegawa, M., Arai, T., Akiyama, H., Nonaka, T., Mori, H., Hashimoto, T., Yamazaki, M., and Oyanagi, K. (2007) TDP-43 is deposited in the guam parkinsonism-dementia complex brains. *Brain* 130, 1386-94.

(4) Higashi, S., Iseki, E., Yamamoto, R., Minegishi, M., Hino, H., Fujisawa, K., Togo, T., Katsuse, O., Uchikado, H., Furukawa, Y., Kosaka, K., and Arai, H. (2007) Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of Alzheimer's disease and dementia with lewy bodies. *Brain Res.* 1184, 284-94.

(5) Chen-Plotkin, A. S., Lee, V. M., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease. *Nat. Rev. Neurol.* 6, 211-20.

(6) 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, Article 25.

(7) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.

(8) Ayala, Y. M., De Conti, L., Avendano-Vazquez, S. E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E., and Baralle, F. E. (2011) TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* 30, 277-88.

(9) Cohen, T. J., Lee, V. M., and Trojanowski, J. Q. (2011) TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends Mol. Med.* 17, 659-67.

(10) Birsa, N., Bentham, M. P., and Fratta, P. (2020) Cytoplasmic functions of TDP-43 and FUS and their role in ALS. *Semin. Cell Dev. Biol.* 99, 193-201.

(11) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* 37, 1799-1808.

(12) 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.

(13) Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., and Lee, V. M. (2008) Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *J. Biol. Chem.* 283, 13302-9.

(14) Kikis, E. A. (2019) The intrinsic and extrinsic factors that contribute to proteostasis decline and pathological protein misfolding. *Adv. Protein Chem. Struct. Biol.* 118, 145-161.

(15) Vakkayil, K. L., and Hoppe, T. (2022) Temperature-dependent regulation of proteostasis and longevity. *Front. Aging* 3.

(16) Yerbury, J. J., Ooi, L., Dillin, A., Saunders, D. N., Hatters, D. M., Beart, P. M., Cashman, N. R., Wilson, M. R., and Ecroyd, H. (2016) Walking the tightrope: Proteostasis and neurodegenerative disease. *J. Neurochem.* 137, 489-505.

(17) Kabashi, E., Valdmanis, P. N., Dion, P., Spiegelman, D., McConkey, B. J., Velde, C. V., Bouchard, J. P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P. F., Camu, W., Meininger, V., Dupre, N., and Rouleau, G. A. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat. Genet.* 40, 572-574.

(18) Moreno, F., Rabinovici, G. D., Karydas, A., Miller, Z., Hsu, S. C., Legati, A., Fong, J., Schonhaut, D., Esselmann, H., Watson, C., Stephens, M. L., Kramer, J., Wiltfang, J., Seeley, W. W., Miller, B. L., Coppola, G., and Grinberg, L. T. (2015) A novel mutation P112H in the TARDBP gene associated with frontotemporal lobar degeneration without motor neuron disease and abundant neuritic amyloid plaques. *Acta Neuropathol. Commun.* 3, Article 19.

(19) Bentmann, E., Neumann, M., Tahirovic, S., Rodde, R., Dormann, D., and Haass, C. (2012) Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* 287, 23079-94.

(20) Chiang, C. H., Grauffel, C., Wu, L. S., Kuo, P. H., Doudeva, L. G., Lim, C., Shen, C. K., and Yuan, H. S. (2016) Structural analysis of disease-related TDP-43 D169G mutation: Linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Sci. Rep.* 6, 21581.

(21) Dang, M., and Song, J. (2020) ALS-causing d169g mutation disrupts the ATP-binding capacity of TDP-43 RRM1 domain. *Biochem. Biophys. Res. Commun.* 524, 459-464.

(22) Agrawal, S., Jain, M., Yang, W. Z., and Yuan, H. S. (2021) Frontotemporal dementia-linked P112H mutation of TDP-43 induces protein structural change and impairs its RNA binding function. *Protein Sci.* 30, 350-365.

(23) Austin, J. A., Wright, G. S. A., Watanabe, S., Grossmann, J. G., Antonyuk, S. V., Yamanaka, K., and Hasnain, S. S. (2014) Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4309-4314.

(24) 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.

(25) 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.

(26) Patni, D., and Jha, S. K. (2021) Protonation-deprotonation switch controls the amyloid-like misfolding of nucleic-acid-binding domains of TDP-43. *J. Phys. Chem. B* 125, 8383-8394.

(27) Guo, W., Chen, Y., Zhou, X., Kar, A., Ray, P., Chen, X., Rao, E. J., Yang, M., Ye, H., Zhu, L., Liu, J., Xu, M., Yang, Y., Wang, C., Zhang, D., Bigio, E. H., Mesulam, M., Shen, Y., Xu, Q., Fushimi, K., and Wu, J. Y. (2011) An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struct. Mol. Biol.* 18, 822-30.

(28) Jiang, L. L., Zhao, J., Yin, X. F., He, W. T., Yang, H., Che, M. X., and Hu, H. Y. (2016) Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Sci. Rep.* 6, 23928.

(29) Yang, X., Meisl, G., Frohm, B., Thulin, E., Knowles, T. P. J., and Linse, S. (2018) On the role of sidechain size and charge in the aggregation of abeta42 with familial mutations. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5849-E5858.

(30) Liemann, S., and Glockshuber, R. (1999) Influence of amino acid substitutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry* 38, 3258-67.

(31) Singh, J., and Udgaonkar, J. B. (2015) Structural effects of multiple pathogenic mutations suggest a model for the initiation of misfolding of the prion protein. *Angew Chem.* 54, 7529-33.

(32) Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C. M. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805-8.

(33) Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* 8, 101-6.

(34) Hammarstrom, P., Wiseman, R. L., Powers, E. T., and Kelly, J. W. (2003) Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science* 299, 713-6.

(35) Singh, J., and Udgaonkar, J. B. (2016) Unraveling the molecular mechanism of pH-induced misfolding and oligomerization of the prion protein. *J. Mol. Biol.* 428, 1345-1355.

(36) Singh, J., and Udgaonkar, J. B. (2015) Molecular mechanism of the misfolding and oligomerization of the prion protein: Current understanding and its implications. *Biochemistry* 54, 4431-42.

(37) Flagmeier, P., Meisl, G., Vendruscolo, M., Knowles, T. P., Dobson, C. M., Buell, A. K., and Galvagnion, C. (2016) Mutations associated with familial parkinson's disease alter the initiation and amplification steps of α-synuclein aggregation. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10328-33.

Chapter 6

Electrostatic modulation of the intramolecular and the intermolecular interactions during the formation of an amyloid-like assembly

6.1 Introduction

Aggregation of proteins to form amyloid fibrils and amorphous aggregates is a common characteristic of all neurodegenerative diseases.[1] Many debilitating diseases such as amyotrophic lateral sclerosis, frontotemporal lobar degeneration, Alzheimer's and Parkinson's show deposition of the amyloid fibrils.[2-4] Specifically, the deposition of TDP-43 protein has been observed in a multitude of neurodegenerative diseases.[5-9] TDP-43 is a multi-domain protein consisting of an N-terminal domain, two RNA Recognition Motif domains tethered by a linker (termed here as TDP-43^{tRRM}) and a C-terminal domain.[10] Functionally, the nucleic acid-binding domain of TDP-43 (TDP-43^{tRRM}) performs vital functions inside the cell, such as mRNA splicing, mRNA transcription and translation and autoregulation of its own mRNA.[11, 12] Apart from performing critical functions, TDP-43 and its variant, have been shown to form different assemblies under destabilizing and misfolding conditions.[13-16] While previous studies have shown that protonation of the buried residue coupled with the salt plays a key role in regulating different kinds of assemblies[17, 18] formed for TDP-43^{tRRM}, we do not yet clearly understand how the electrostatic modulates the conformational transitions between the protein molecules.

The aggregation proclivity of a particular protein is influenced by two major determinants– its primary sequence which includes the distribution of residues that are hydrophobic, charged, polar or aromatic and their propensity to form β sheet content.[19] The other factor that significantly influences the aggregation propensity is the various physiological stresses that the cell experiences in its lifetime. These can be thermal fluctuations, starvation stress, and oxidative stress, which functions by bringing about change in the solvation condition of the cellular milieu, often initiating misfolding of proteins. Both intrinsic and extrinsic factors critically regulate the thermodynamics and the kinetics of the aggregation process. In particular, pH and salt are well-known modulators of the electrostatic interactions and have been shown to affect the protein molecules by bringing about changes in structure, conformation, hydrophobicity and the oligomeric nature of the protein molecules. The effect of electrostatics on the protein, in general, involves competition between many phenomena such as repulsion or attraction between charged residues, local dipolar interactions and desolvation penalties of ionizable residues.[19] Moreover, the contribution and balance of both electrostatic and hydrophobic interactions determine the rate and the equilibrium properties of protein species populated. In the context of age-associated diseases involving neurodegeneration, the accumulation of highly charged proteins in the cell is held to be a

significant source of protein destabilization.[20] This is majorly due to the change in the net charge brought about by oxidative damage to the cell under the aging condition.[20-22] In this context, it is vital to understand how the changes in the solvation condition modulate the electrostatic network in TDP-43 proteins. In particular, TDP-43[tRRM] protein has been shown to undergo complex multi-step aggregation reactions in the presence of physiological ionic strength at low pH (Chapter 4). Depending on the ionic strength of the solution, the protein populates into dry molten globule (no salt), native-like oligomers (< 5 mM KCl) and ThT-positive amyloid-like assemblies (>15 mM KCl).[18] However, it is not clear how the changes in the electrostatic network affects the structural conformation of the protein molecules and populate different types of species. Therefore in this work, we wanted to systematically dissect the effect of pH and ionic strength on the intra- and the intermolecular interactions governing the protein molecules. To understand the magnitude of this dependency and characterize the electrostatic effect on the protein, we studied both thermodynamic and kinetic dependency of the observed rate constants by varying the pH and salt concentration. We observed that the enthalpic destabilization caused by the protonation of buried ionizable residue is coupled to the structural opening of the protein molecules to form partially unfolded molecules. The pH-induced structural changes create an intra-molecular destabilizing effect which acts as the key structural requirement for initiating the aggregation process. In the presence of salt, the protein molecules undergo significant conformational changes to re-arrange their backbone-backbone interactions to form ordered amyloid-like assemblies. Salt, however, affects the aggregation process in two steps represented by the two transition states of the aggregation reaction. The salt majorly functions by binding specifically to the charged side-chains of the protein molecules to exert its effect, thus following electroselectivity series of anion binding.

6.2 Materials and methods

Protein purification

The purification of TDP-43[tRRM] protein was performed using a protocol followed in a previous study (Chapter 2). The purity of the protein was confirmed by both SDS-PAGE and electrospray ionization mass spectrometry.

Buffers, solutions and experimental conditions

All the chemicals used are of the highest purity grade from Sigma and Sisco research laboratories (SRL). Buffers at different pH were as reported previously (Chapter 2). For the

pH-induced experiment, the protein was incubated in different pH buffers, one set containing < 5 mM KCl and the other set of the sample containing 150 mM KCl in the solution. Once the sample achieved equilibrium, the far-UV CD was measured at 216 nm. The protein concentration used in all the experiments is kept constant at 10 μM.

Circular Dichroism

Jasco-J815 spectropolarimeter was used to measure the far-UV CD spectrum using the same parameters as described previously (Chapter 2). For equilibrium spectrum measurement, the absorbance spectra was collected between 250-200 nm using the following parameters- 2 nm, bandwidth; 1s D.I.T; 20 nm/min, scan speed. The pH-induced equilibrium data was plotted at 216 nm as a function of different pH.

Thioflavin T assay

Thioflavin T assay was performed by transferring 1 μM of protein sample in ThT assay buffer containing 40 μM ThT dye. The dye was excited at 440 nm, and the emission spectra were collected between 452 to 535 nm using excitation and emission slit widths of 2.5 nm and 2.5 nm, respectively. For kinetic measurements, the emission was collected at 482 nm for 20 s and the average of all the measurements was plotted as a function of time. Buffer spectrum was collected using the same parameter and subtracted from the sample data.

6.3 Results

6.3.1 Concentration of salt determines the fate of the protein at low pH

The effect of electrostatics on TDP-43^{tRRM} was analyzed by monitoring the secondary structure of the protein across the different pH ranges and under different salt conditions. Figure 6.1A shows the far-UV CD spectrum of the protein molecules at two extreme pH and ionic strength. The protein molecules at pH 7.5 remain native-like in the presence of low (< 5 mM KCl) or physiological salt (150 mM KCl) concentration. The protein shows a secondary structure rich in α and β sheet as described in a previous study.[17] When the protein is transferred to pH 3 buffer, in the presence of low salt (< 5 mM KCl), the molecules maintain an intact native-like secondary structure; however, in the presence of 150 mM KCl, the molecules undergo a disorder-to-order transformation seen by increased ellipticity. Additionally, the spectrum shows a maxima around 216 nm, characteristic of a β sheet rich structure, called here as amyloid-like assemblies.
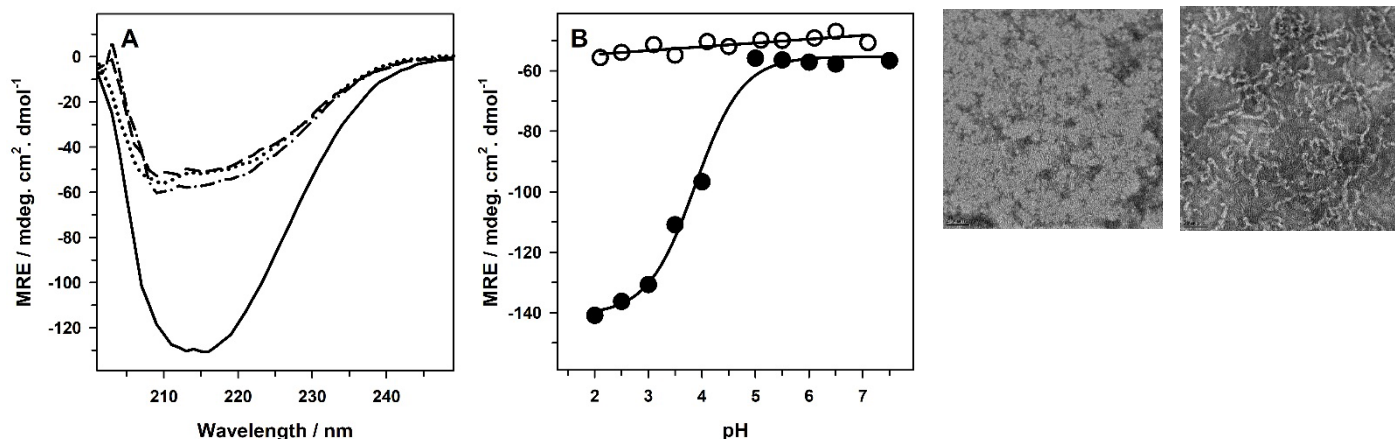
Figure 6.1: *Effect of pH and ionic strength on the equilibrium transition from native to the amyloid-like state of TDP-43^tRRM protein. (A) Far-UV CD spectrum of the protein incubated at pH 7.5 and pH 3 under two different salt conditions (< 5 mM and 150 mM KCl). (B) Mean residue ellipticity of the protein molecules when incubated under different pH solutions containing < 5 mM KCl (unfilled circles) and 150 mM KCl (filled circles) at 216 nm. (C) Transmission electron microscopy images of the protein molecules at pH 3 in the presence of < 5 mM and 150 mM KCl are shown.*

In order to understand the pH and ionic strength dependency on the structural conformation of the protein, we transferred the protein molecules to different pH buffers, one set of which contained low salt concentration (< 5 mM KCl) and the other contained physiological salt concentration (150 mM KCl) (Figure 6.1B). We next measured the far-UV CD absorbance of these samples and plotted the absorbance at 216 nm as a function of pH. We observed that in the presence of low salt concentration, the mean residue ellipticity (MRE) remains native-like across the pH range. However, in the presence of physiological salt concentration, we observed a sigmoidal dependency on pH, saturating around pH 3. The MRE increases from – 6000 mdeg. cm^2. dmol^{-1} under native-like conditions (between pH 7.5 and pH 5) to around – 14000 mdeg. cm^2. dmol^{-1} under low pH (less than pH 3). Previous studies have characterized these species formed under low pH, and high salt concentration and were found to be ThT-positive amyloid-like assemblies.[18] Our data thus suggests that under low pH conditions, salts play a significant role in bringing about structural transformation from a native-like secondary structure to a structure rich in cross-β sheets.

The TEM images in Figure 6.1C depict the appearance of the protein molecules at pH 3 under low salt and physiological salt concentration. The protein molecules appear as small diffused structures in the presence of low salt concentration; however, in the presence of high salt, the protein molecules appear as small curly worm-like assemblies.

6.3.2 Dependence of the kinetics of the formation of amyloid-like assembly on pH

In presence of 150 mM KCl, we studied the kinetics of the native to amyloid-like assembly transition as a function of pH using ThT fluorescence (Figure 6.2). ThT dye is a widely used probe to study the evolution of the aggregation reaction, as it is known to stack within the protein molecules during elongation and fluoresce.[23, 24]. We observed that at pH 7.5 and pH 5, the protein molecules do not bind to ThT and hence do not show any change in the ThT fluorescence. At pH 4.5, the protein molecules show some change in the ThT fluorescence, but the rate of the increase in the ThT fluorescence is relatively slow. Between pH 4 and pH 2.5, we see an exponential increase in the ThT fluorescence, suggesting that the protein molecules assemble together rapidly to form amyloid-like assemblies.
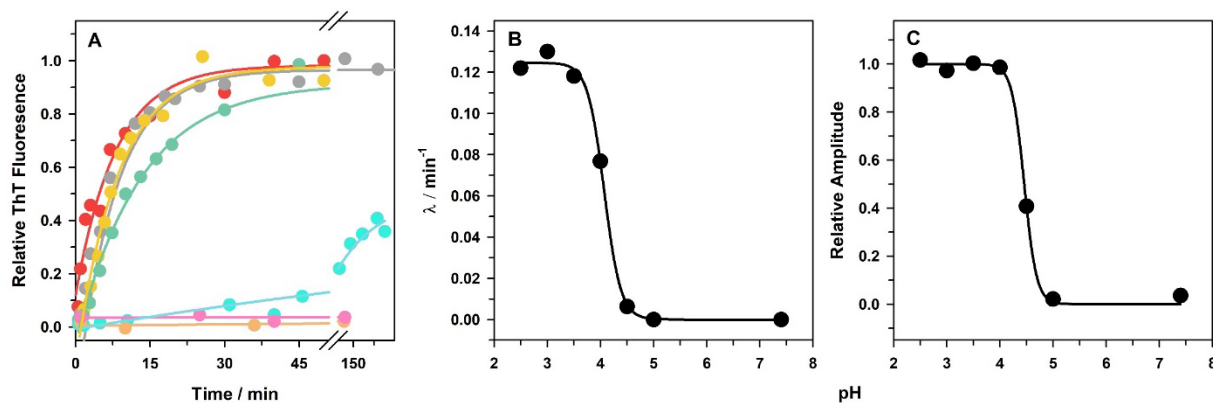
Figure 6.2: *pH-dependent kinetics of the amyloid-like assemblies at physiological salt concentration. (A) Aggregation kinetics monitored by ThT fluorescence at 482 nm in different pH buffers containing 150 mM KCl - pH 7.5 (orange), pH 5 (pink), pH 4.5 (blue), pH 4 (cyan), pH 3.5 (yellow), pH 3 (gray) and pH 2.5 (red). All the data except pH 7.5 and pH 5 in panel A fits best to a single exponential equation. For pH 7.5 and pH 5, a linear fit was used. (B) The apparent rate constant is plotted as a function of pH. (C) The relative amplitude of the aggregates is plotted as a function of pH. The maximum ThT fluorescence at low pH was used to calculate the relative amplitude across all the data. The solid lines in panels B and C are to guide the eyes.*

The apparent rate constant derived by fitting the data to a single exponential equation from figure 3A displays a sigmoidal dependency on the pH (Figure 6.2B). Since there is no increase in the ThT fluorescence at pH 7.5 and pH 5, and there is no formation of any amyloid-like assemblies, we assumed a nominal rate constant value of 0. The sigmoidal transition, which reflects the transformation from native to amyloid-like assemblies, has a midpoint of around pH 4.2. The apparent rate constant at pH 4 appears on the transition zone, while the rate

constant appears to have saturated beyond pH 3.5. Between pH 3.5 and pH 2.5, the apparent rate constant has become highly comparable.

Figure 6.2C shows the relative amplitude of the aggregation reaction plotted as a function of pH. The change in the final ThT fluorescence signal shows a sigmoidal transition with a change in the pH. The ThT fluorescence at pH 7.5 and pH 5 did not change, and hence we have assumed a negligible value of 0 as the amplitude. The ThT fluorescence at pH 4.5 has only achieved 40% of the total signal suggesting that around 40% of the population may be undergoing an amyloid-like conformational change reaction. The relative amplitude saturates beyond pH 4, and the value remains similar up to pH 2.5.

6.3.3 Ionic strength dependency of the aggregation reaction as a function of pH

We incubated the protein at different pH containing different concentrations of salt and measured its ThT fluorescence under equilibrium conditions (Figure 6.3). We monitored the ionic strength dependency by a 100-fold change in the salt concentration across the different pH. The protein molecules show varied dependency on the ionic strength at different pH (Figure 6.3A). The protein molecules at pH 5 show no change in the ThT fluorescence upon changing the salt concentration from 1.5 mM to 150 mM. However, we observe a sigmoidal dependency on the ionic strength between pH 4 to pH 2.5. The slope of this sigmoidal dependency is, however, shallow at pH 4 and pH 3.5, while it is quite steep at pH 3 and pH 2.5. The extent of the change in the ThT fluorescence was monitored by calculating the salt concentration at which 50% of the ThT signal is achieved as compared to the maximum fluorescence (Figure 6.3B). Our data imply that at low pH, where the tertiary structure is disrupted completely, the effect of salt is more pronounced (Figure 6.3B).
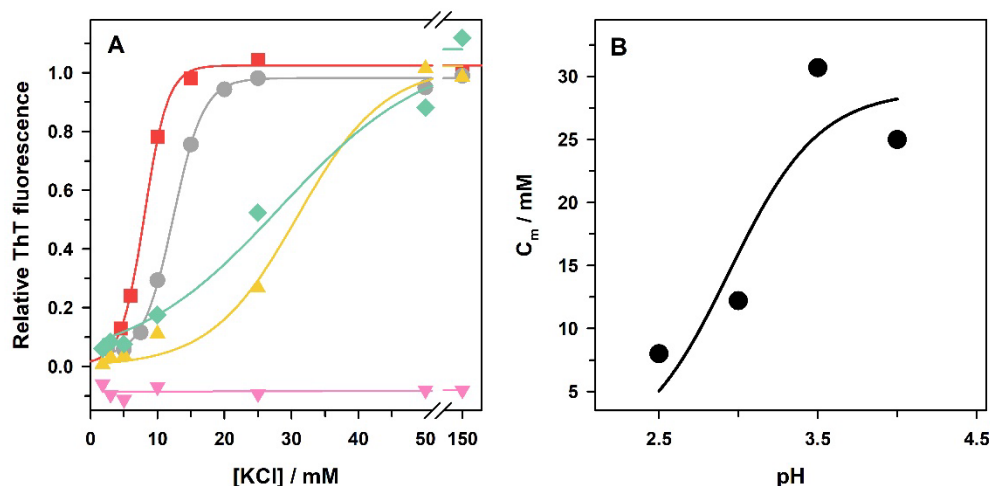
Figure 6.3: *Influence of the ionic strength on amyloid-like conformational transition of the protein molecules as a function of pH. The protein was transferred into different pH buffers containing different KCl concentrations and incubated overnight- pH 5 (pink), pH 4 (cyan), pH 3.5 (yellow), pH 3 (gray) and pH 2.5 (red). The ThT fluorescence of the equilibrated sample was measured at 482 nm. (A) Relative ThT fluorescence of the protein at different pH & ionic strength solution is plotted as a function of KCl concentration. All the data in panel A except pH 5 data fits best to the sigmoidal equation. The pH 5 data is fit to a linear equation. (B) The midpoint of the sigmoidal transition obtained from panel A is plotted as a function of pH.*

Around 25-30 mM of salt concentration is required to achieve 50% of the change in the ThT fluorescence at pH 4 and pH 3.5. In contrast, only around 12 mM and 7 mM of salt concentration are required to achieve a similar change at pH 3 and pH 2.5, respectively. It appears that beyond 50 mM of salt concentration, the ThT fluorescence saturates across the different pHs except for pH 5. Even at 150 mM, we do not observe any change in the ThT fluorescence at pH 5. Our data suggest that the weakening of the native-like interactions brought about by protonation is coupled with the conformational change in the protein molecules induced by the salt.
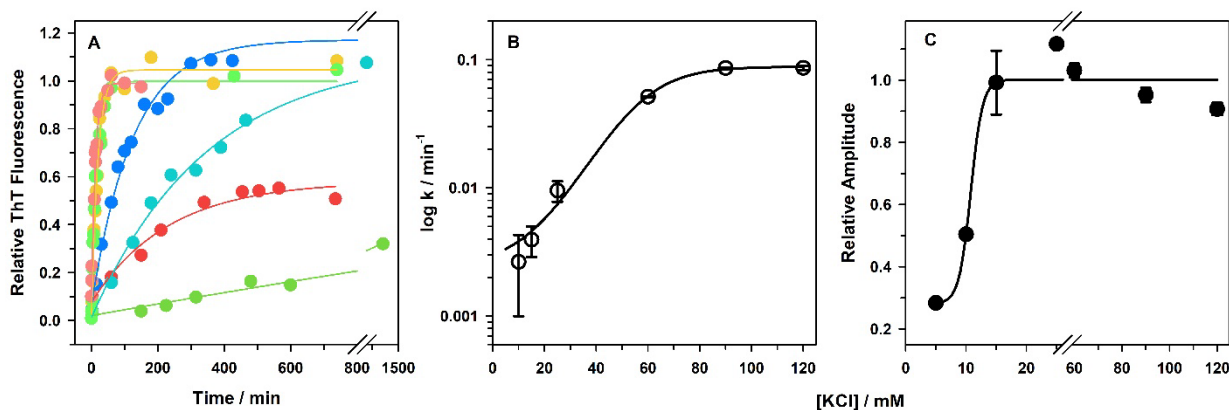
Figure 6.4: *Ionic strength dependency of the kinetics of aggregation. (A) Aggregation kinetics monitored by ThT fluorescence of the protein molecules as it undergoes aggregation in the presence of different ionic strengths of KCl under fixed pH (pH 3) –5 mM (green), 10 mM (red), 15 mM (blue), 25 mM (dark blue), 60 mM (yellow), 90 mM (light green) and 120 mM (pink). (B) The apparent rate constant obtained from panel A is plotted on a logarithmic scale as a function of the ionic strength. (C) The relative change in the signal of the ThT fluorescence calculated by dividing the final ThT fluorescence by the saturating signal of the ThT fluorescence at higher ionic strength (90- 120 mM). All the data in panel A is a fit to a single exponential equation. The data in panel B is a fit to the sigmoidal equation. The fit in panel C is a guide to the eye. The error bar represents the spread in the data from two independent experiments.*

6.3.4 Dependence of the kinetics of the amyloid-like aggregation on ionic strength

We observed that ionic strength plays a crucial role in dictating the structural conformation of the protein molecules. In order to quantitatively dissect the contribution of the ionic strength on the transition state of the aggregation reaction, we studied the kinetics by systematically varying the concentration of the ionic strength at fixed pH (pH 3). We changed the ionic strength from 5 mM to 150 mM and monitored the ThT fluorescence with time (Figure 6.4A). At the low salt concentration of 5 mM, the ThT fluorescence increases very slowly. With a further increase in the ionic strength of the solution, we noted an increase in not just the rate of ThT fluorescence but also the amplitude of the ThT signal. The increase in the ThT fluorescence follows an exponential dependence, and we do not observe any lag phase even at the lowest salt concentration. At the lowest salt concentration, although the rate and the amplitude are very low, we still observe a marginal increase in the ThT signal. The apparent rate constant of the aggregation kinetics across the different ionic strengths was obtained by fitting the data in figure 6.4A to a single exponential equation except for 5 mM data, whose signal increases at a very nominal rate. A semi-log plot of the apparent rate constant on the

ionic strength shows two different slopes, indicating two transition states associated with the aggregation reaction. Between 10 mM to 50 mM, the apparent rate constant is highly dependent on the ionic strength, as seen by the steep slope. However, beyond 60 mM, the dependency of the apparent rate on the ionic strength seems to have diminished.

The amplitude of the aggregation reaction also shows a sigmoidal dependence on the ionic strength of the solution at low pH (Figure 6.4C). At 5 mM salt concentration, only 25-30% of the ThT signal is achieved relative to the ThT signal at the highest salt concentration. With a further increase in the salt concentration to 10 mM, we observe that 50% of the total ThT signal has been achieved. We observe that beyond 15 mM of salt concentration, the ThT signal saturates, suggesting that all of the protein molecules have undergone an amyloid-like transition and no further increase in the ThT fluorescence occurs. It is interesting to observe that although the rate of the aggregation kinetics only saturates at 80 mM of salt concentration, the amplitude of the ThT signal saturates well before 20 mM of salt concentration.

6.3.5 An electroselectivity series governs the intermolecular interactions
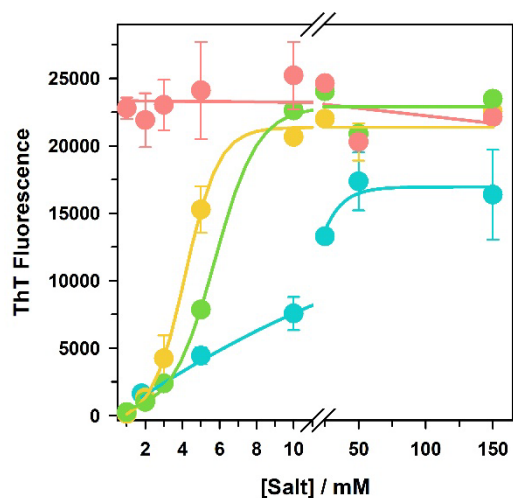
Figure 6.5: *Modulation of the TDP-43^{tRRM} aggregation at a fixed pH by different anions. The formation of amyloid-like fibrils was monitored by ThT fluorescence at 482 nm as a function of the ionic strength of different salts- KCl (blue), KNO$_3$ (green), KI (yellow) and K$_2$SO$_4$ (red). The fit to all the data is a guide to the eye. The error bar represents the spread in the data from two independent experiments.*

The aggregation of TDP-43^{tRRM} was carried out in the presence of different salt at a fixed pH condition (pH 3) - KCl, KNO$_3$, KI and K$_2$SO$_4$. The extent of aggregate formation was monitored by measuring the change in the ThT fluorescence under each aggregation condition.

The amplitude of the fluorescence signal appears to be dependent upon both the nature and the concentration of the anion (Figure 6.5). The degree of the effect of the salt on the TDP-43^{tRRM} aggregation was determined by salt concentration at which half the ThT signal (50%) is achieved as compared to the maximum ThT signal. The efficacy of the salt on TDP-43^{tRRM} aggregation was found to be: $SO_4^{-2} > I^- > NO^{3-} > Cl^-$. The minimum salt concentration at which maximum amyloid fibril formation obtained are 6 mM, 10 mM and 25 mM for KI, KNO$_3$ and KCl, respectively. For K$_2$SO$_4$, we observed that even at the lowest salt concentration of 1.5 mM maximum ThT fluorescence signal was reached. Therefore we could not estimate the lowest K$_2$SO$_4$ concentration required to bring about the amyloid formation.

6.4 Discussion

The structural opening reaction is the critical event in initiating the aggregation reaction

TDP-43^{tRRM} exists in multiple conformations under different solvation conditions. In neutral pH (in presence of low or physiological salt concentration) the protein molecules remain as a monomeric native conformer. At low pH in the presence of low salt, the protein molecules show an intact secondary structure (Figure 6.1). However, it has been shown in a previous study that the intramolecular tertiary packing is highly disrupted under low pH conditions. The opening/closing reaction that the protein molecules experience in the form of conformational breathing might allow the buried ionizable residue[25] in the protein molecule to undergo protonation under low pH conditions. This pH dependent transition observed for TDP-43^{tRRM} (Figure 6.1B) is due to the differential pK$_a$ of the buried ionizable side-chain in the folded and the unfolded states. The protonation of the buried ionizable residue is coupled to the structural opening of the protein molecules that exposes the hydrophobic patches which was earlier buried inside the core, thereby, populating partially unfolded molecules (PUF).[18] In order to compensate for this enthalpic destabilization, the partially unfolded molecules come together to form an oligomer with an intact native-like secondary structure at low pH and low salt concentration (Figure 6.1A and 6.1B). In the presence of physiological salt concentration, the protein molecules undergo dramatic change in the conformation, resulting in both intra- and the intermolecular structural re-arrangement to form amyloid-like assemblies (Figure 6.1 and 6.2). This initial pH-induced structural opening occurring early in the reaction is thus an important event that marks the beginning of the aggregation reaction. This observation is also in line with the conformational change hypothesis that strongly suggests that changes in the

tertiary structure of the protein molecules under partially denaturing conditions promote the formation of alternative conformations that self-assemble to form amyloid fibrils.[26]

At physiological salt concentration, we observed that the kinetics of the formation of the amyloid-like assemblies is also dependent on pH (Figure 6.2A). This dependence of the kinetics of aggregation on pH suggests that large scale structural opening occurs in the transition state of the reaction. Interestingly, the midpoint of the transition of the apparent rate constant, amplitude of the pH-dependent kinetics (Figure 6.2B and 6.2C) and the equilibrium of the conformational transition seen by CD (Figure 6.1B) is around pH 4.0. These data suggest that both the kinetics and the equilibrium transition depends on the protonation of similar buried ionizable residue which drives the misfolding reaction. Between pH 5 to pH 7.5, we do not observe ionization of any buried residue, and hence no opening reaction occurs. Thus, we do not observe formation of any amyloid-like assemblies up to pH 5 (Figures 6.1 and 6.2). Below pH 5, the protonation results in structural opening that might expose the "aggregation-prone" regions in the protein. The salt present in the solution might modulate the electrostatic interactions within this region to bring about crucial structural rearrangement to form amyloid-like assemblies. The amount of salt required to bring about the amyloid-like interactions in the protein molecules is highly dependent on the extent of destabilization caused by pH-induced structural changes (Figure 6.3). Thus our data unambiguously suggest that electrostatics play a critical role in dictating the different structural conformers of TDP-43[tRRM] protein.

Salt modulates the activation barrier

The presence of salt plays a significant role in modulating the conformation of the protein molecules. Salt affects both the rate and the amplitude of the aggregation reaction (Figure 6.4). However, we observed that salt affects the aggregation reaction in two steps characterized by the two transition states (Figure 6.4B). The aggregation reaction at lower salt concentration is sharply dependent on the ionic strength indicated by the steep slope that represents the dependency of the aggregation rate on the salt concentration in the solution. Our data suggest that salt modulates key interactions within the transition state to promote favorable amyloid-like interactions. In contrast, the narrow slope observed at higher salt concentrations (beyond 60 mM) does not show a substantial dependence of the rate on the ionic strength. This indicates that there is second transition state occurring at higher salt concentrations whose barrier height is weakly affected by the electrostatic interactions. Based on these data, we can conclude that the charge interactions brought about by salt play a crucial role in determining

the first free energy barrier while the second barrier seems to be limited by the structural opening reaction brought about by pH. Nonetheless, our data suggest once the protein is destabilized due to pH-induced structural changes, salt functions to bring about conformational changes in the protein molecules such that the main chain –main chain hydrogen bonding is promoted to create amyloid-like interactions within the protein molecules. Additionally, the presence of salt may modulate the salt bridge networks within the protein molecules or give rise to cation-pi interactions. Overall, the cooperative nature of the electrostatic interactions gives rise to amyloid-like assemblies.

Specific ion binding drives the protein molecules towards aggregation reaction

Our data suggest that salt plays a significant role in governing the intermolecular interactions of the protein molecules at low pH. However, we do not understand how salt brings about these changes. There are several ways in which salt can modulate the interactions of the protein molecules. The simplest is the Debye-Huckel screening which suggests that salt can exert its effect by screening the repulsive charges on the protein molecules, thereby allowing them to interact to form assemblies. There are other ways in which salt can exert its effect-either by directly binding to the protein molecules (Electroselectivity effect) [27-29] or by affecting the structure of the water molecule surrounding the protein molecules (Hofmeister effect).[30-32] If salts had been acting only by screening the charges, then the extent of fibril formation for 5 mM KI and 5 mM KCl would have been the same. However, we observe from our data that the extent of fibril formation is relatively less in KCl as compared to KI. This suggests that ionic strength alone cannot be responsible for the aggregation of TDP-43^{tRRM} under low pH. Moreover, screening of charges requires high salt concentration, and we observe that amyloid-like assembly formation for TDP-43^{tRRM} occurs at a minimal salt concentration. If the salts had been perturbing the water structure around the protein molecules, then it would have followed the Hofmeister series, in which case the dependency of ions should be: $SO_4^{-2} > Cl^- > NO^{3-} > I^-$. However, in our study, ions that promote the fibril formation do not follow the above series, and hence we can rule out the effect of salt in modulating the water structure around the protein molecules and causing aggregation. We observed the following dependency in our study: $SO_4^{-2} > I^- > NO^{3-} > Cl^-$ (Figure 6), which is well in line with the electroselectivity series of ions. The electroselectivity series implies the role of anions in specifically binding to the positive charges on the protein and modulating the amyloid-like fibril formation. A minimal amount of anions (~15 mM KCl) is required to bring about the amyloid-fibril formation, suggesting that specific

binding of the anions to the positively charged side-chains at low pH is essential to bring about structural re-arrangement and promote extensive hydrogen-bonding to form amyloid-like assemblies.

Many aggregation-prone proteins such as prion, α-synuclein,[33] β2-microglobulin and Aβ (1-40) have been shown to be influenced by the anions.[27, 34] Anions act in a diverse way to affect the aggregation reaction; e.g., in the case of the mouse prion protein, the anions function by binding to the protein specifically, following electroselectivity series of anion binding.[35] In contrast, the α-synuclein and the yeast prion protein Sup35 show formation of amyloid-fibril through the Hofmeister effect of anion binding.[33, 36]

6.5 References

(1) Koo, E. H., Lansbury, P. T., Jr., and Kelly, J. W. (1999) Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9989-90.

(2) Dugger, B. N., and Dickson, D. W. (2017) Pathology of neurodegenerative diseases. *Cold Spring Harb. Perspect. Biol.* 9.

(3) Jellinger, K. A. (2001) The pathology of parkinson's disease. *Adv. Neurol.* 86, 55-72.

(4) Rowland, L. P., and Shneider, N. A. (2001) Amyotrophic lateral sclerosis. *N. Engl. J. Med.* 344, 1688-1700.

(5) 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.

(6) 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.

(7) Hasegawa, M., Arai, T., Akiyama, H., Nonaka, T., Mori, H., Hashimoto, T., Yamazaki, M., and Oyanagi, K. (2007) TDP-43 is deposited in the guam parkinsonism-dementia complex brains. *Brain* 130, 1386-94.

(8) Higashi, S., Iseki, E., Yamamoto, R., Minegishi, M., Hino, H., Fujisawa, K., Togo, T., Katsuse, O., Uchikado, H., Furukawa, Y., Kosaka, K., and Arai, H. (2007) Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of alzheimer's disease and dementia with lewy bodies. *Brain Res.* 1184, 284-94.

(9) Chen-Plotkin, A. S., Lee, V. M., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease. *Nat. Rev. Neurol.* 6, 211-20.

(10) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.

(11) Cohen, T. J., Lee, V. M., and Trojanowski, J. Q. (2011) TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends Mol. Med.* 17, 659-67.

(12) Ayala, Y. M., De Conti, L., Avendano-Vazquez, S. E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E., and Baralle, F. E. (2011) TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* 30, 277-88.

(13) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem.* 289, 8264-8276.

(14) Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takahashi, R., Kitahara, R., Ito, H., Fujiwara, N., and Urushitani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* 288, 14886-905.

(15) Wang, Y. T., Kuo, P. H., Chiang, C. H., Liang, J. R., Chen, Y. R., Wang, S. Y., Shen, J. C. K., and Yuan, H. N. S. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* 288, 9049-9057.

(16) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* 132, 1186-1187.

(17) 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.

(18) 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.

(19) Vascon, F., Gasparotto, M., Giacomello, M., Cendron, L., Bergantino, E., Filippini, F., and Righetto, I. (2020) Protein electrostatics: From computational and structural analysis to discovery of functional fingerprints and biotechnological design. *Comput. Struct. Biotechnol. J.* 18, 1774-1789.

(20) de Graff, A. M., Hazoglou, M. J., and Dill, K. A. (2016) Highly charged proteins: The achilles' heel of aging proteomes. *Structure* 24, 329-36.

(21) Stadtman, E. R. (2006) Protein oxidation and aging. *Free Radic. Res.* 40, 1250-8.

(22) Davies, K. J., Delsignore, M. E., and Lin, S. W. (1987) Protein damage and degradation by oxygen radicals. Ii. Modification of amino acids. *J. Biol. Chem.* 262, 9902-7.

(23) Biancalana, M., and Koide, S. (2010) Molecular mechanism of thioflavin-t binding to amyloid fibrils. *Biochim. Biophys. Acta* 1804, 1405-12.

(24) Sulatskaya, A. I., Kuznetsova, I. M., and Turoverov, K. K. (2012) Interaction of thioflavin t with amyloid fibrils: Fluorescence quantum yield of bound dye. *J. Phys. Chem. B* 116, 2538-44.

(25) Patni, D., and Jha, S. K. (2021) Protonation-deprotonation switch controls the amyloid-like misfolding of nucleic-acid-binding domains of TDP-43. *J. Phys. Chem. B* 125, 8383-8394.

(26) Kelly, J. W. (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* 8, 101-6.

(27) Raman, B., Chatani, E., Kihara, M., Ban, T., Sakai, M., Hasegawa, K., Naiki, H., Rao Ch, M., and Goto, Y. (2005) Critical balance of electrostatic and hydrophobic interactions is required for beta 2-microglobulin amyloid fibril growth and stability. *Biochemistry* 44, 1288-99.

(28) Goto, Y., Calciano, L. J., and Fink, A. L. (1990) Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. U. S. A.* 87, 573-7.

(29) Goto, Y., Takahashi, N., and Fink, A. L. (1990) Mechanism of acid-induced folding of proteins. *Biochemistry* 29, 3480-8.

(30) Collins, K. D., and Washabaugh, M. W. (1985) The Hofmeister effect and the behaviour of water at interfaces. *Q. Rev. Biophys.* 18, 323-422.

(31) Baldwin, R. L. (1996) How Hofmeister ion interactions affect protein stability. *Biophys. J.* 71, 2056-63.

(32) Zhang, Y., and Cremer, P. S. (2006) Interactions between macromolecules and ions: The Hofmeister series. *Curr. Opin. Chem. Biol.* 10, 658-63.

(33) Munishkina, L. A., Henriques, J., Uversky, V. N., and Fink, A. L. (2004) Role of protein-water interactions and electrostatics in alpha-synuclein fibril formation. *Biochemistry* 43, 3289-300.

(34) Klement, K., Wieligmann, K., Meinhardt, J., Hortschansky, P., Richter, W., and Fandrich, M. (2007) Effect of different salt ions on the propensity of aggregation and on the structure of alzheimer's abeta(1-40) amyloid fibrils. *J. Mol. Biol.* 373, 1321-33.

(35) Jain, S., and Udgaonkar, J. B. (2010) Salt-induced modulation of the pathway of amyloid fibril formation by the mouse prion protein. *Biochemistry* 49, 7615-24.

(36) Yeh, V., Broering, J. M., Romanyuk, A., Chen, B., Chernoff, Y. O., and Bommarius, A. S. (2010) The Hofmeister effect on amyloid formation using yeast prion protein. *Protein Sci.* 19, 47-56.

Chapter 7

Site-specific mapping of the core of the amyloid-like aggregates of the nucleic acid binding domain of TDP-43 using hydrogen-deuterium exchange coupled to mass spectrometry

7.1 Introduction

Deposition of fibrils is the universal feature among the spectrum of neurodegenerative diseases such as Alzheimer's,[1] Parkinson's,[2] Amyotrophic lateral sclerosis (ALS),[3, 4] Fronto-temporal lobar degeneration (FTLD),[5] Huntington's[6] and Guam-Parkinsonism.[7] In the case of ALS & FTLD, proteinaceous inclusions consisting of TDP-43 protein are the major pathological hallmark of the diseased condition.[3, 4] TDP-43 consists of the N-terminal domain, tandem RNA Recognition Motif domains and the C-terminal domain.[8, 9] Distinct pathological oligomers and aggregates have been observed in the tissues of the diseased patient and under invitro condition, although it still remains to be debated if these aggregates resemble amyloid fibrils or not.[10-15] Such distinct aggregates may contribute towards heterogeneity and result in phenotypic diversity in the symptoms of the disease.[16, 17] Consistent with these findings, many invitro studies have tried to characterize the aggregates formed for different regions of TDP-43 protein.[18-24] High-level conformational information on the structures of the amyloid fibrils for a number of different proteins is available in the literature.[25-29] However, similar studies in TDP-43 protein is only limited to the low complexity domain (LCD) domain of the protein and therefore does not completely shed light on the role of other domains in the fibril formation.[30-33] Multiple studies have shown RRM domains of TDP-43 (TDP-43^{tRRM}) form aggregates or fibrils under stress-like condition.[22, 34-38] But we have very few high-level structural information on the type of fibrils or aggregates adopted by this domain.[39] In order to bridge this critical gap, we have adopted hydrogen-deuterium exchange (HDX) coupled to mass spectrometry to decode the core and the dynamic behavior of the secondary structural segments within the assemblies.

In our previous work, we had gained mechanistic evidence on the multistep pathway of aggregation of TDP-43^{tRRM} under destabilizing conditions. Our data suggested that multiple transient oligomeric species are populated along the aggregation pathway (Chapter 4). The amyloid-like assemblies formed at the end of the aggregation reaction seemed to be around 20-30 nm in its hydrodynamic radius, suggesting a certain level of heterogeneity in the overall size of the amyloid-like assemblies.[37] It thus becomes interesting to characterize and understand the dynamicity and understand how this gives rise to the heterogeneity observed within the aggregates. Hydrogen-deuterium exchange (HDX) provides substantial information on the structure and the dynamics of the amyloid fibrils. HDX measures the exchange of the amide hydrogen and thus can distinguish between the solvent accessibility of the exposed amide hydrogen and the ones involved in the core of the structure.[40, 41] Significant structural re-

arrangement occurs when the native protein undergoes a conformational transition to form ordered amyloid-like assembly. The core of the ordered amyloid-like assembly involves extensive network of hydrogen bonds; therefore, using HDX becomes an effective tool for monitoring the core of the aggregates. In the current study, we attempt to probe the site-specific behavior of individual structural segments within the protein assemblies.

In this study, we mapped the structural core and the dynamics of the different secondary structural segment within the amyloid-like aggregate of TDP-43^{tRRM}. Upon analyzing the hydrogen exchange of the backbone amide of the aggregate, we found that the sequences 124-131 and 248-255 do not exchange with the deuterium in the solvent, suggesting that these regions are buried deep inside the aggregates and forms the core of the supramolecular assembly. The boundaries of the core offer a certain level of protection to the nearby secondary structure segments, such that almost 48% of the segment 132-148 remain unexchanged. Interestingly, we observed bimodal isotope distribution patterns for specific segments (107-122, 156-200, 220-231, 235-243), suggesting the existence of heterogeneity within the assemblies. Mapping the core and the surface secondary segments gives a snapshot of the critical residues involved in the aggregate conformation and, therefore, can be used to develop biomarkers or therapeutic molecules.

7.2 Materials and methods

Protein purification

TDP-43^{tRRM} was expressed and purified following the protocol described in a previous study (Chapter 2). The purified protein was stored in storage buffer – 20 mM Sodium phosphate buffer containing 150 mM KCl (pH 7.2) with 1 mM DTT. The purity of the protein was confirmed using SDS-PAGE and ESI-MS analysis.

Buffers, solutions and experimental conditions

All the buffers used for the mass spectrometry experiment are of the highest purity grade from Sigma unless otherwise mentioned. The buffer for amyloid formation (aggregation buffer) consisted of 20 mM Glycine-HCl with 150 mM KCl (pH 3). The amyloid-like assemblies were prepared by jumping the required volume of the protein from the storage solution to the aggregation buffer such that the final concentration of the protein is around 100 µM. The protein was incubated in aggregation buffer for 2 h before performing any experiment. For the hydrogen-deuterium exchange, buffers were prepared in deuterium oxide (D$_2$O)

solution, and the pH of the solution was adjusted using deuterium chloride (DCl) and sodium deuteroxide (NaOD). The pH of the buffer was corrected for the isotopic effect by adding a factor of 0.4 to the pH meter reading to get the final pH of the deuterated buffer. For solutions containing guanidinium hydrochloride (GdmCl), the concentration of GdmCl was determined using the refractive index. The buffer solutions were filtered using 0.2 μm filters before use.

Peptide Mapping

To generate a peptide map for TDP-43^{tRRM}, we followed a previously established protocol (Acharya, N; Thesis, 2020). In brief, TDP-43^{tRRM} was first desalted in Gly-HCl pH 3.0 buffer, and then an online pepsin digestion was carried out on an immobilized pepsin column (Applied Biosystems) by passing 0.05% formic acid at a flow rate of 50 μL/min on a nanoAquity UPLC system (Waters). The digested peptides were collected on a trap column (C18 reversed-phase chromatography column, Waters), where excess salts were removed and finally eluted on an analytical C18 reversed-phase chromatography column (Waters). The elution of the peptides was done using 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 μL/min for efficient elution. All the columns were placed in an HDX module with temperature maintained at 4˚C. The eluted peptides were detected using a Synapt G2 HD mass spectrometer (Waters). The following parameters were used for mass detection: desolvation gas 600 L/hr, source temperature 50°C, desolvation temperature 100°C and capillary voltage 3 kV. The peptides were sequenced using the MS^E method and then analyzed using ProteinLynx Global Server (Waters) and manual inspection.

Table 7.1: Peptide sequence and their corresponding secondary structure element. (Loops/Disordered regions - L, β-sheets - β and α-helix - α)

S. No.	Peptide sequence	Secondary structural content
1	97-106	L_1
2	107-122	$\beta_1 L_2 \alpha_1$
3	124-131	L_3
4	132-148	$\beta_2 L_4 \beta_3$
5	156-174	$\alpha_2 L_6 \beta_4 L_7 \beta_5$
6	175-200	$L_8 \beta_6 L_9$
7	211-219	$\alpha_3 L_{10} \beta_7$
8	220-231	L_{11}
9	235-243	$L_{12} \alpha_4$
10	248-255	$\beta_9 L_{14} \beta_{10}$

Hydrogen-deuterium exchange experiment

Amyloid assemblies were first formed by transferring the required volume of protein into aggregation buffer from a concentrated stock solution such that the final concentration of protein was around 250-300 μM and incubated for 2 h. The exchange was initiated by diluting the protein 20-fold into the exchange buffer at pH 3 at 25˚C. At different time points of exchange, an aliquot of the sample (25μL) was withdrawn, and the exchange reaction was quenched by mixing with 175 μL of ice-cold 7.5 M GdmCl under quenching conditions (100 mM glycine-HCl, pH 2.5) for 1 min. The salt was then desalted using a G-25 spin trap desalting column in an ice-cold 100 mM Gly-HCl buffer (pH 2.5). The samples were injected at a flow rate of 50 μL/min (0.05% formic acid in water) into the HDX module coupled with UPLC for online pepsin digestion. The peptides eluted from the C18 trap column were washed to remove any salt. The peptides were then passed through C18 RPC analytical column and eluted using acetonitrile gradients (3-20% in 2 min, 20-65% in 4 min, 65-95% in 2 min, 95-3% in 1min) with 0.1% formic acid. A flow rate of 40 μL/min was used. In order to account for any back exchange and forward exchange, control experiments were carried out using protein unfolded in D$_2$O containing GdmCl. Using same experimental setup, the unfolded protein sample was processed to obtain the peptide mass of 95% labeled sample for each peptide.

At each time point of exchange, the peptide masses were calculated from the centroid of the isotopic envelope using MassLynx software. The extent of deuterium incorporation was calculated using the following equation-

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

where, m(t) and m(0%) denotes the centroid mass of the peptide at different time points and the undeuterated sample, respectively, while m(95%) is the centroid mass of the fully deuterated peptide (in 95% D$_2$O).

7.3 Results

7.3.1 TDP-43^{tRRM} forms ordered β-sheet rich amyloid-like assembly

The native form of TDP-43^{tRRM} protein is a monomeric molecule with an intact secondary structure that transforms in a multiphasic manner to form amyloid-like assemblies (Figure 7.1A). The far-UV CD scan of the native protein shows a structure rich in α-helix and β-sheet (Figure 7.1B). However, in the presence of the stress-like condition, the protein molecules undergo a drastic conformational change associated with global secondary structure

change that results in increased ellipticity and shift in the signal maxima to 216 nm, indicative of a structure rich in β-sheets (Figure 7.1B). A global structural analysis of the protein thus suggests that TDP-43^{tRRM} might form extensive H-bonding networks to give rise to an ordered β sheet-rich structure, as seen in the far-UV CD spectra. With the help of transmission electron microscopy (TEM) images, we can clearly observe small worm-like assemblies (Figure 7.1C) formed under the aggregation condition. The above data provide global-level information on the structural changes in the protein molecules as they form aggregate. In order to acquire more sequence-specific information, we used HDX-MS.
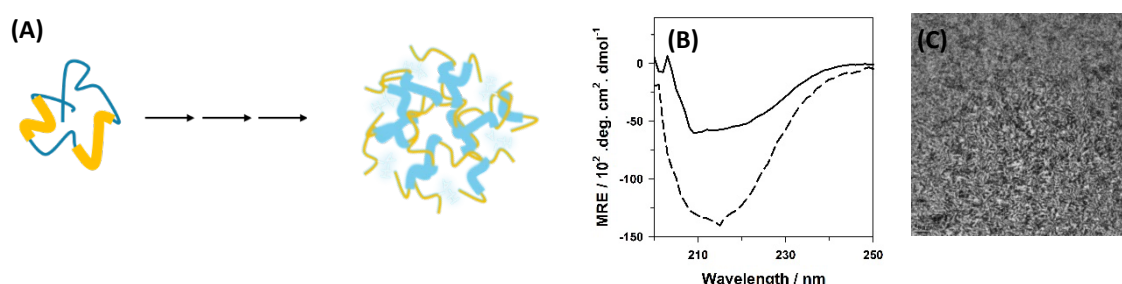
Figure 7.1: *Structural transformation of the protein molecule to an ordered worm-like fibril. (A) Schematic representation of the transformation of the native monomeric protein to an ordered assembly. (B) Far-UV CD spectra of the native form (solid line) of the protein and the amyloid-like assembly (dashed line). (C) Transmission Electron Microscope (TEM) image of the worm-like amyloid fibrils.*

7.3.2 Peptide mapping of TDP-43^{tRRM}

In HDX-MS, we obtain information on the solvent accessibility of the main-chain amide hydrogen sites of the different segments of the protein by proteolytically cleaving them at low pH. We first generated the peptide map of TDP-43^{tRRM} by controlled proteolysis using pepsin as described in the 'Material and Methods' section. A total of 10 segments were obtained upon cleavage, which is diagrammatically represented below, along with the missing gaps in the peptide.
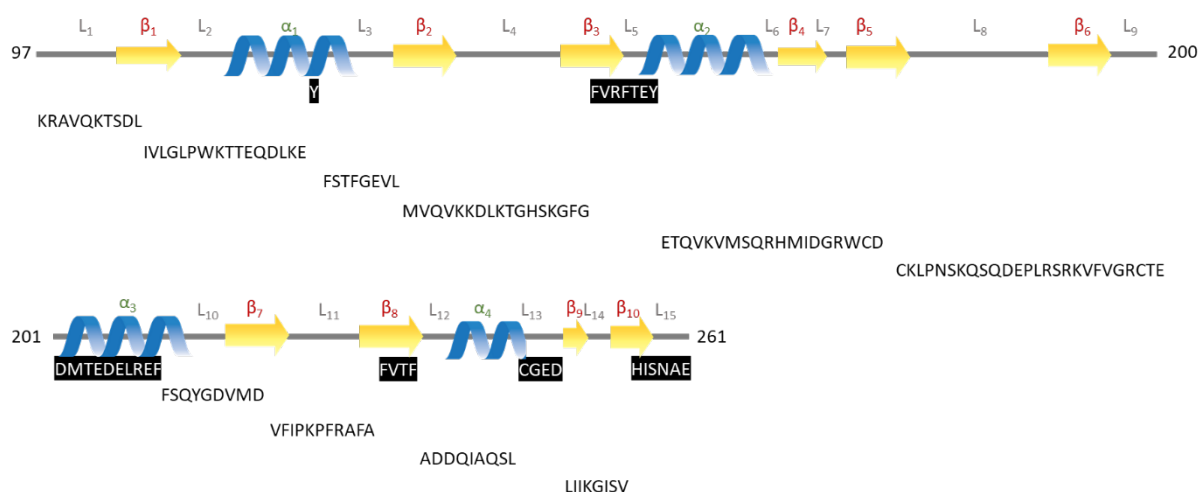
Figure 7.2: *Peptide map of TDP-43^(tRRM).Schematic representation of α-helix (blue ribbon), β-sheet (yellow arrow) and the loop regions (grey line) of TDP-43^(tRRM). The amino acid sequence corresponding to the different secondary structure elements is shown. The 'missing gap' of the peptide is shown in the black box.*

7.3.3 HDX-MS characterization of the amyloid-like assemblies
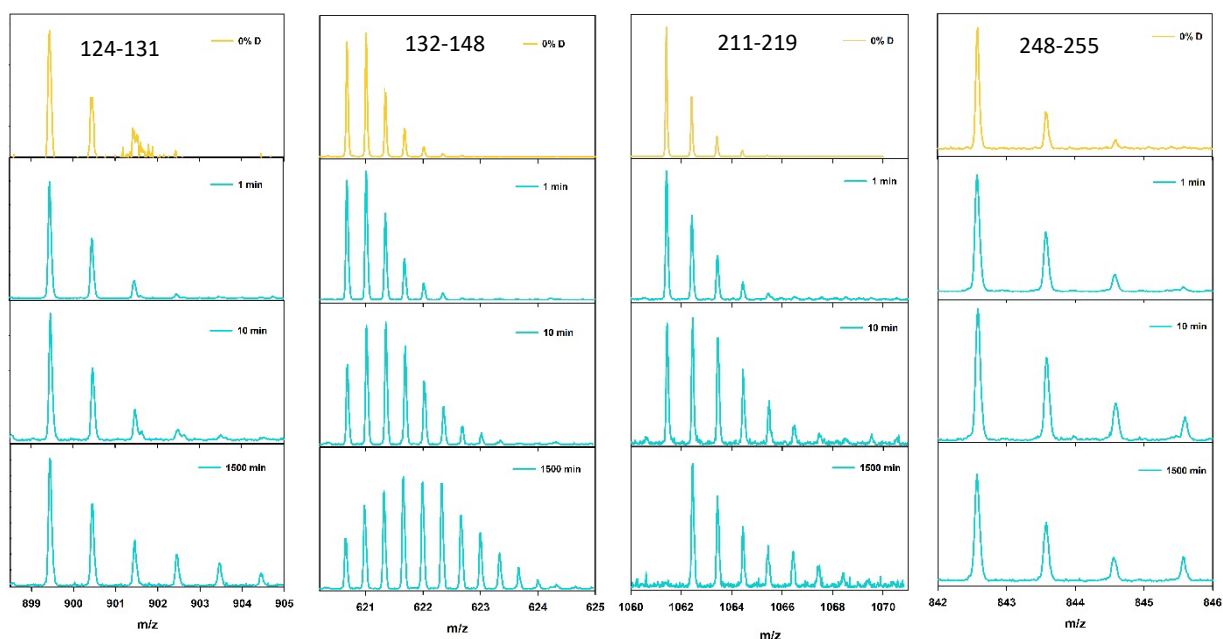
Figure 7.3: *Mass spectra of selected fragments of the aggregates at different time points of exchange reaction are compared with the undeuterated (0% D) control.*

The change in the mass for a few sequence segments of the aggregates at different times of HDX is summarized in Figure 7.3. Mass spectra for four secondary structural segments and their respective controls (0% D) are shown. The mass spectra show unimodal isotope distribution patterns for the segments- 124-131, 132-148, 211-219 and 248-255, implying that these fragments exist primarily in a single conformation. Each of these fragments shows a different extent of deuterium incorporation, suggesting that they differ in the protection they afford against the exchange. The secondary structure segments (132-148 and 211-219) show an increase in the mass with time. In contrast, for the segments- 124-131 and 248-255, we do not observe any change in the mass even at the highest time point. Our data suggest that the regions 124-131 and 248-255 show the least deuterium incorporation and thus correspond to the most protected segments of the aggregate.
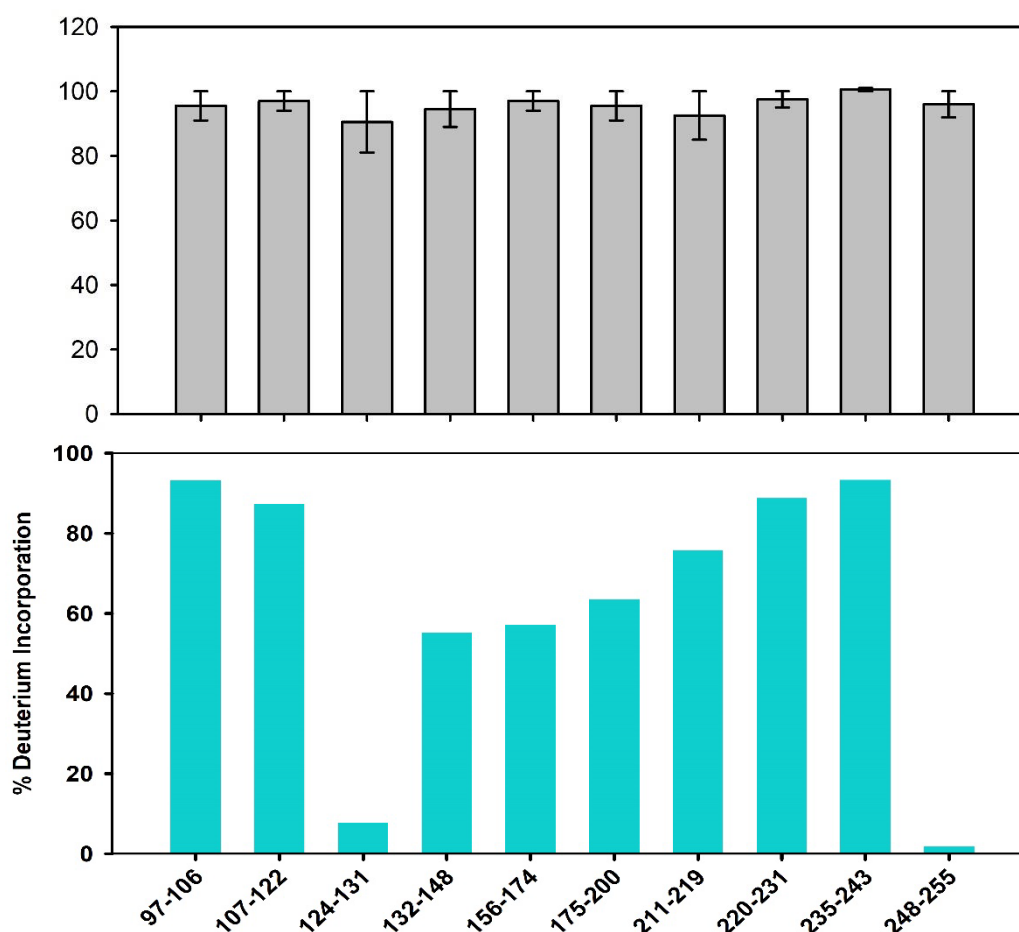
Figure 7.4: *The extent of deuterium incorporation in different segments of TDP-43^{tRRM}. The grey bar graph represents the deuterium incorporation by the different segments when the protein molecule is in its native state. This image has been adopted from the thesis of Acharya, N.; 2020 (unpublished data). The blue bar graph represents the deuterium incorporation by different secondary structure segments of the amyloid-like aggregates.*

The extent of deuterium incorporation in the monomer native form and the aggregates across the different segments of the protein are shown in Figure 7.4. All the segments under the native condition show maximum deuterium incorporation (Figure 7.4). For the amyloid-like aggregates, we observe a varied extent of deuterium uptake for different fragments. For the four fragments (97-106, 107-122, 220-231 and 235-243), we observe nearly 85-95% deuterium incorporation, suggesting that these segments are part of the aggregates with similar structural protection as observed in the native state. In contrast, the segments (124-131 and 248-255) show the least deuterium uptake in the aggregates, but the same segments in the native form show full deuterium incorporation. Our data imply that this secondary structure segment is protected in the aggregates but remains unprotected in the native state (Figure 7.4). For the segments (132-148, 156-174, 175-200 and 211-219), deuterium incorporation varied between 55-75%, suggesting that these regions form part of the aggregates which remains partially accessible to the solvent and the exchange reaction (Figure 7.4).
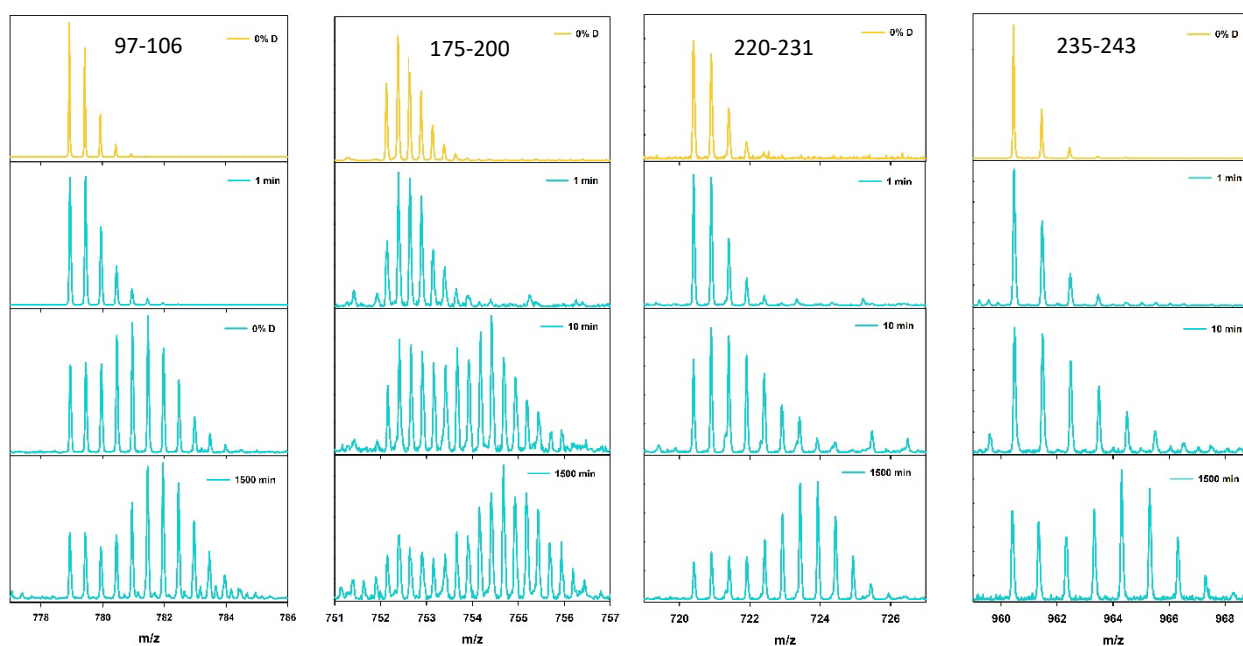
Figure 7.5: *Conformational heterogeneity in the amyloid form. Mass spectra of selected fragments are shown at a different time of exchange reaction and compared with 0% deuterated control.*

7.3.4 Conformational heterogeneity observed in the aggregates

We observed bimodal isotope distribution patterns for six segments of the aggregates-97-106, 107-122, 156-174, 175-200, 220-231 and 235-243 (Figure 7.5). A bimodal distribution implies that there exist at least two different populations that are significantly different in their protection against the exchange. Out of the two conformations, one remains highly protected and does not show any exchange reaction, while the other is weakly protected and undergoes an exchange reaction. Our data suggest that these structural segments may form the core in atleast some population of the aggregates. Such a multimodal distribution pattern has also been observed in oligomers and amyloid fibrils of other proteins.[29, 42, 43]
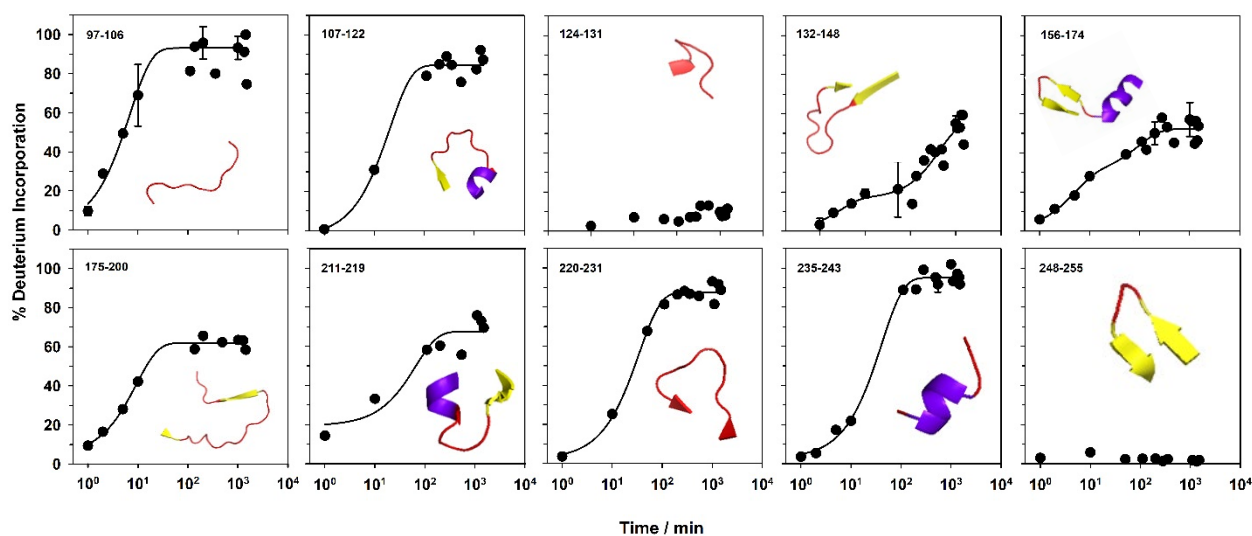
Figure 7.6: *Kinetics of hydrogen-deuterium exchange. Percent deuterium incorporation in different segments of the aggregates at different time periods of exchange at pH 3 and 25˚C is shown. The continuous lines through all the kinetic profiles fit either mono- or bi-exponential equation except for segments 124-131 and 248-255, where no change is observed. In the inset of each panel, cartoon representations of the secondary structural segments are shown in different colors- yellow (β-sheet), blue (α-helix) and red (loop). The error bar represents data from two independent experiments.*

7.3.5 Structural dynamics of the aggregates of TDP-43[tRRM]

The kinetics for the deuterium uptake in different secondary structural segments of the aggregates were monitored at pH 3 (Figure 7.6). At pH 3, the intrinsic HDX rate is at its minimum, which allows the exchange of all the amide protons (spread across the secondary structure region) to be monitored. The percent deuterium incorporation profiles fit either to mono- or biexponential equation. The segments 124-131 and 248-255 do not show any deuterium incorporation suggesting that the secondary structural segment is highly buried in the inner core of the aggregates. Except for segments 132-148 and 156-174, all other segments

show a monoexponential change in deuterium incorporation. Additionally, we observed 20% missing burst phase for the segment 211-219, indicating very fast deuterium incorporation for around 20% of the population (Table 7.2). Previous data has shown that all the segments undergo an EX2 exchange reaction (Acharya, N.; 2020, Thesis). Hence, using the observed rate constant from the fit, we calculated the protection factor for each of the segments (Table 7.2). The protection factor (P_f) was calculated by first determining the intrinsic exchange rate constant for each of the fragments and then dividing this by the observed rate constant obtained by the fits of the deuterium incorporation.

Sequence	k_{in}/min^{-1}	k_{obs}/min^{-1} (%amp)	P_f	Extent of Protection
97-106	0.17	ND (8%) 0.14 (92%)	Burst phase 1.22	Unprotected Unprotected
107-122	0.15	0.05 (87%) ND (13%)	3 Very slow	Unprotected Protected
124-131	1.3	ND (100%)	>10^5	Protected
132-148	0.27	0.36 (16%) 0.0022 (39%) ND (45%)	0.75 125 Very slow	Unprotected Moderately protected Protected
156-174	0.56	0.211 (28%) 0.0107 (24%) ND (48%)	2.7 52.6 Very slow	Unprotected Weakly protected Protected
175-200	0.67	ND (5%) 0.1076 (57%) ND (38%)	Burst phase 6.25 Very slow	Unprotected Unprotected Protected
211-219	0.98	ND (19%) 0.0153 (48%) ND (33%)	Burst phase 64 Very slow	Unprotected Weakly protected Protected
220-231	0.23	0.0290 (85%) ND (15%)	7.93 Very slow	Unprotected Protected
235-243	2.2	0.02 (92%) ND (8%)	111.11 Very slow	Moderately protected Protected
248-255	0.1	ND (100%)	>10^5	Protected

ND: Not determined

Table 7.2: *H/D 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.*
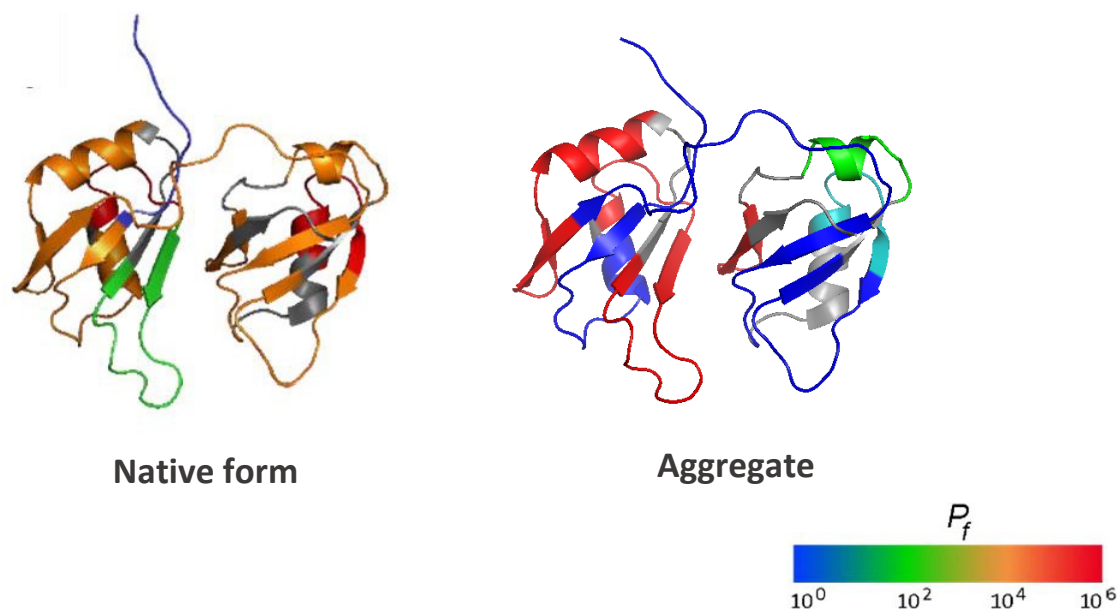
Figure 7.7: *Comparison of the heat map of the aggregate with the native form. Heat-map of the aggregate is compared with that of the native form (adapted from Acharya, N.; 2020, Thesis). Heat-map is designed using the protection factor values derived from the HDX kinetics of the aggregates. P_f value of 10^0-10^1 is shown by blue color, 10^1-10^2 is indicated by teal color, 10^2-10^3 is shown by green color and values >10^5 is displayed by red color. Secondary structural segments with amplitude nearly 50% or greater (with P_f > 10^5) are shown in red color. The missing fragments have been colored in gray.*

A comparison of the protection factors across the different segments of the aggregates implies significant differences in the kinetics of the deuterium uptake. The values of the P_f were divided into four different categories- highly protected (P_f>10^5), moderately protected (P_f= 10^2- 10^5), weakly protected (P_f> 10^1 - 10^2), and unprotected (P_f< 10^1) (Table 7.2). The highest protection factor is shown by two segments 124-131 and 248-255, which correspond to the L$_3$ region and β$_9$L$_{14}$β$_{10}$ region, respectively. This highest protection level observed corresponds to the highly protected core region of the aggregates. The loop region thus gets protected in the aggregates and may contribute to the disorder-to-order transition observed as the protein molecules transform from the native to the aggregate state. The segments 132-148 and 156-174 corresponding to β$_2$L$_4$β$_3$ and α$_2$L$_6$β$_4$L$_7$β$_5$ shows almost 45-48% of the secondary structural

region remaining unexchanged, indicating a high level of protection. Interestingly, the segment 132-148 is near the highly protected region 124-131 and hence may form the boundary of the structural core region. While the structural segments 97-106, 107-122, 220-231 corresponding to L_1, $\beta_1 L_2 \alpha_1$ and L_{11} remain unprotected. Nearly 16% and 28% of the population of the aggregate conformers belonging to the segments 132-148 and 156-174, respectively, remain unprotected. A heat map generated using protection factor values helps to visually dissect the solvent accessibility and the dynamics of different segments of protein molecules within the aggregates.

7.4 Discussion

In this study, we have performed a high-level structural characterization of the amyloid-like aggregates formed by TDP-43^{tRRM}. Our data suggest that the core of the aggregates encompasses the regions 124-131 and 248-255 (Figure 7.3). Interestingly, the core of the aggregates is divided into discontinuous stretches, as has been observed for the prion protein.[29] While the boundary of the core region comprising the segment 132-148 shows some degree of protection (48% protection) while no such protection is observed in the segment 235-243, which is also present near the core region 248-255 of the protein (Figure 7.6, Table 7.2). The protected regions of the structure segments, which form the core of the aggregates, contain a significant portion of the loop region (Table 7.2). It is observed in the global secondary structure that as the protein molecules transform from a native state to an aggregate state, the loop regions become more structured, consequently giving rise to a disorder-to-order transition (Figure 7.1B). So, therefore, we can speculate that the protected segments that form the core may be responsible for the disorder-to-order transition.

Previous data on the HDX of the native form indicated that the secondary structural segments of the native form show nearly 95% deuterium uptake after 24 h but for the same segments we do not observe complete protection in the amyloid-like aggregates suggesting that the aggregate structure is relatively protected as compared to the native form (Figure 7.4). Under the aggregation condition, we examined two segments 124-131 and 248-255 showing complete protection while other segments (132-148 and 156-174) showing nearly 50% deuterium incorporation, suggesting that the amyloid-like aggregates are much more ordered. Our data also indicates that the secondary structural segments belonging to the RRM1 domain (aa 102-177) are relatively more protected than those belonging to the RRM2 domain (aa 192-261), although the segment 248-255 in RRM2 remains protected. We can propose that the

RRM1 domain is relatively more ordered as compared to RRM2 domain within the arrangement of the aggregate structure. Overall, the HDX of the aggregates show a remarkable difference in the protection that each segment affords as a result of its structure and conformation within the aggregates (Figure 7.4).

It has been known that along the energy landscape, the formation of amyloid-like assemblies are thermodynamically most stable state,[44] but there exists heterogeneity and plasticity within the structural conformations of the aggregates. This might be due to the lack of any evolutionary pressure in defining the structure of the amyloid fibril.[45] Although the external morphology of the aggregates, as observed by the TEM image, does not seem to be polymorphic (Figure 7.1C), we observe bimodal isotopic distribution for certain secondary structural segments that may point to the presence of heterogeneity. Upon analysis of the mass spectra, we noticed that six structural elements showed a bimodal exchange pattern in the deuterium uptake (Figure 7.5) which can be co-related to the difference in the solvent-exposed area of the sequence segment. Therefore, we can speculate the co-existence of at least two different aggregate conformers under the equilibrium condition. While measuring the kinetics of the deuterium uptake (Figure 7.6), we observed the presence of both mono- and biexponential kinetics for the segments undergoing exchange at pH 3 under the EX2 regime. In particular, the biexponential kinetics seen for the structural segments– 132-148 and 156-174 may be due to the differences in the secondary structure elements in the protein that affords a different level of protection against HDX. However, there can also be a possibility that the same segment may have different levels of protection due to the co-existence of two different types of aggregates as noted under equilibrium condition. Subtle changes in the conformations without much change in the size or the morphological appearance of the amyloid-like assemblies in TEM (Figure 7.1C) demonstrate the importance of HDX as a powerful tool to map the polymorphism that exists within the amyloid structure. The structural plasticity observed within the amyloid-like assemblies thus contributes to the growing concept of polymorphism that exists in the field. Many structural studies have identified the presence of polymorphic structures among the aggregates formed using the low complexity domain of TDP-43.[30, 31] Consistent with the above findings, our study is an attempt to understand the structural dynamics and polymorphism of the assemblies formed by TDP-43[tRRM] under stress-like condition.

7.5 Reference

(1) Selkoe, D. J. (1991) The molecular pathology of alzheimer's disease. *Neuron* 6, 487-98.

(2) Poewe, W., Seppi, K., Tanner, C. M., Halliday, G. M., Brundin, P., Volkmann, J., Schrag, A. E., and Lang, A. E. (2017) Parkinson disease. *Nat. Rev. Dis. Primers* 3, 17013.

(3) 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.

(4) 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.

(5) Seltman, R. E., and Matthews, B. R. (2012) Frontotemporal lobar degeneration: Epidemiology, pathology, diagnosis and management. *CNS Drugs* 26, 841-70.

(6) Walker, F. O. (2007) Huntington's disease. *Lancet* 369, 218-28.

(7) Steele, J. C. (2005) Parkinsonism-dementia complex of guam. *Mov. Disord.* 20 Suppl 12, S99-S107.

(8) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry* 56, 809-823.

(9) 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, Article 25.

(10) Fang, Y. S., Tsai, K. J., Chang, Y. J., Kao, P., Woods, R., Kuo, P. H., Wu, C. C., Liao, J. Y., Chou, S. C., Lin, V., Jin, L. W., Yuan, H. S., Cheng, I. H., Tu, P. H., and Chen, Y. R. (2014) Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun.* 5, Article 4824.

(11) Kao, P. F., Chen, Y. R., Liu, X. B., DeCarli, C., Seeley, W. W., and Jin, L. W. (2015) Detection of TDP-43 oligomers in frontotemporal lobar degeneration-TDP. *Ann. Neurol.* 78, 211-221.

(12) French, R. L., Grese, Z. R., Aligireddy, H., Dhavale, D. D., Reeb, A. N., Kedia, N., Kotzbauer, P. T., Bieschke, J., and Ayala, Y. M. (2019) Detection of TAR DNA-binding protein 43 (TDP-43) oligomers as initial intermediate species during aggregate formation. *J. Biol. Chem.* 294, 6696-6709.

(13) Laferriere, F., Maniecka, Z., Perez-Berlanga, M., Hruska-Plochan, M., Gilhespy, L., Hock, E. M., Wagner, U., Afroz, T., Boersema, P. J., Barmettler, G., Foti, S. C., Asi, Y. T., Isaacs, A. M., Al-Amoudi, A., Lewis, A., Stahlberg, H., Ravits, J., De Giorgi, F., Ichas, F., Bezard, E., Picotti, P., Lashley, T., and Polymenidou, M. (2019) TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci.* 22, 65-77.

(14) Capitini, C., Conti, S., Perni, M., Guidi, F., Cascella, R., De Poli, A., Penco, A., Relini, A., Cecchi, C., and Chiti, F. (2014) TDP-43 inclusion bodies formed in bacteria are structurally amorphous, non-amyloid and inherently toxic to neuroblastoma cells. *Plos One* 9, e86720.

(15) Capitini, C., Fani, G., Vivoli Vega, M., Penco, A., Canale, C., Cabrita, L. D., Calamai, M., Christodoulou, J., Relini, A., and Chiti, F. (2021) Full-length TDP-43 and its C-terminal domain form filaments in vitro having non-amyloid properties. *Amyloid* 28, 56-65.

(16) Millecamps, S., Salachas, F., Cazeneuve, C., Gordon, P., Bricka, B., Camuzat, A., Guillot-Noel, L., Russaouen, O., Bruneteau, G., Pradat, P. F., Le Forestier, N., Vandenberghe, N., Danel-Brunaud, V., Guy, N., Thauvin-Robinet, C., Lacomblez, L., Couratier, P., Hannequin, D., Seilhean, D., Le Ber, I., Corcia, P., Camu, W., Brice, A., Rouleau, G., LeGuern, E., and Meininger, V. (2010) SOD1, ang, vapb, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: Genotype-phenotype correlations. *J. Med. Genet.* 47, 554-60.

(17) Kawakami, I., Arai, T., and Hasegawa, M. (2019) The basis of clinicopathological heterogeneity in TDP-43 proteinopathy. *Acta Neuropathol.* 138, 751-770.

(18) Berning, B. A., and Walker, A. K. (2019) The pathobiology of TDP-43 C-terminal fragments in ALS and ftld. *Front. Neurosci.* 13, 335.

(19) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc.* 132, 1186-1187.

(20) Mompean, M., Buratti, E., Guarnaccia, C., Brito, R. M., Chakrabartty, A., Baralle, F. E., and Laurents, D. V. (2014) "Structural characterization of the minimal segment of TDP-43 competent for aggregation". *Arch. Biochem. Biophys.* 545, 53-62.

(21) Chang, C. K., Chiang, M. H., Toh, E. K., Chang, C. F., and Huang, T. H. (2013) Molecular mechanism of oxidation-induced TDP-43 RRM1 aggregation and loss of function. *FEBS Lett.* 587, 575-582.

(22) Zacco, E., Martina, S. R., Thorogate, R., and Pastore, A. (2018) The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.* 11, Article 372.

(23) Tsoi, P. S., Choi, K. J., Leonard, P. G., Sizovs, A., Moosa, M. M., MacKenzie, K. R., Ferreon, J. C., and Ferreon, A. C. M. (2017) The N-terminal domain of ALS-linked TDP-43 assembles without misfolding. *Angew. Chem.* 56, 12590-12593.

(24) Wang, A., Conicella, A. E., Schmidt, H. B., Martin, E. W., Rhoads, S. N., Reeb, A. N., Nourse, A., Ramirez Montero, D., Ryan, V. H., Rohatgi, R., Shewmaker, F., Naik, M. T., Mittag, T., Ayala, Y. M., and Fawzi, N. L. (2018) A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J.* 37.

(25) Gremer, L., Scholzel, D., Schenk, C., Reinartz, E., Labahn, J., Ravelli, R. B. G., Tusche, M., Lopez-Iglesias, C., Hoyer, W., Heise, H., Willbold, D., and Schroder, G. F. (2017) Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science* 358, 116-119.

(26) Carulla, N., Zhou, M., Giralt, E., Robinson, C. V., and Dobson, C. M. (2010) Structure and intermolecular dynamics of aggregates populated during amyloid fibril formation studied by hydrogen/deuterium exchange. *Acc. Chem. Res.* 43, 1072-1079.

(27) Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H., and Goto, Y. (2002) Mapping the core of the beta(2)-microglobulin amyloid fibril by h/d exchange. *Nat. Struct. Biol.* 9, 332-336.

(28) Kheterpal, I., Zhou, S., Cook, K. D., and Wetzel, R. (2000) Abeta amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13597-13601.

(29) 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, 217-31.

(30) 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.

(31) Li, Q., Babinchak, W. M., and Surewicz, W. K. (2021) Cryo-em structure of amyloid fibrils formed by the entire low complexity domain of TDP-43. *Nat. Commun.* 12, 1620.

(32) Zhuo, X. F., Wang, J., Zhang, J., Jiang, L. L., Hu, H. Y., and Lu, J. X. (2020) Solid-state NMR reveals the structural transformation of the TDP-43 amyloidogenic region upon fibrillation. *J. Am. Chem. Soc.* 142, 3412-3421.

(33) Mompean, M., Hervas, R., Xu, Y., Tran, T. H., Guarnaccia, C., Buratti, E., Baralle, F., Tong, L., Carrion-Vazquez, M., McDermott, A. E., and Laurents, D. V. (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett.* 6, 2608-2615.

(34) Wang, Y. T., Kuo, P. H., Chiang, C. H., Liang, J. R., Chen, Y. R., Wang, S. Y., Shen, J. C. K., and Yuan, H. N. S. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem.* 288, 9049-9057.

(35) Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takahashi, R., Kitahara, R., Ito, H., Fujiwara, N., and Urushitani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* 288, 14886-905.

(36) 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.

(37) 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.

(38) 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.

(39) Shenoy, J., El Mammeri, N., Dutour, A., Berbon, M., Saad, A., Lends, A., Morvan, E., Grelard, A., Lecomte, S., Kauffmann, B., Theillet, F. X., Habenstein, B., and Loquet, A. (2020) Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. *FEBS J.* 287, 2449-2467.

(40) Engen, J. R. (2009) Analysis of protein conformation and dynamics by hydrogen/deuterium exchange ms. *Anal. Chem.* 81, 7870-7875.

(41) Wales, T. E., and Engen, J. R. (2006) Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* 25, 158-70.

(42) Mysling, S., Betzer, C., Jensen, P. H., and Jorgensen, T. J. (2013) Characterizing the dynamics of alpha-synuclein oligomers using hydrogen/deuterium exchange monitored by mass spectrometry. *Biochemistry* 52, 9097-103.

(43) Paslawski, W., Mysling, S., Thomsen, K., Jorgensen, T. J., and Otzen, D. E. (2014) Co-existence of two different alpha-synuclein oligomers with different core structures determined by hydrogen/deuterium exchange mass spectrometry. *Angew Chem.* 53, 7560-3.

(44) Baldwin, A. J., Knowles, T. P., Tartaglia, G. G., Fitzpatrick, A. W., Devlin, G. L., Shammas, S. L., Waudby, C. A., Mossuto, M. F., Meehan, S., Gras, S. L., Christodoulou, J., Anthony-Cahill, S. J., Barker, P. D., Vendruscolo, M., and Dobson, C. M. (2011) Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* 133, 14160-14163.

(45) Lansbury, P. T., Jr. (1999) Evolution of amyloid: What normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3342-3344.

Chapter 8

Conclusion and Future directions

This chapter reviews the most important findings of the thesis and discuss some future directions of the current work.

8.1 Summary

A detailed understanding on stress-detection by proteins and the subsequent molecular and structural level changes remains an evolving area of research. The central focus of research in this study have been to decode the structural and conformational changes to the protein molecules as it detects changes in the solvation environment. We briefly discuss the specific aims and conclusion of the study:

a) ***Chapter 2 & 3:*** In these two chapters, we explored the early molecular events that occurs as TDP-43^{tRRM} detects change in the solvation environment. Under the stress-like condition, the protonation-deprotonation equilibrium is linked to the assembly-disassembly equilibrium. Upon low pH stress, we observe that protonation of buried ionisable residue is coupled to the conformational opening reaction, resulting in the formation of partially unfolded molecules. These molecules have exposed hydrophobic patches which allows the protein molecules to assemble through weak intermolecular interactions forming metastable assemblies. The type of assemblies formed depends upon the type of solvation environment around the protein molecule. We characterized two different assemblies A form and L form at pH 3 and pH 2.5, respectively. The structural organization of the A form is different as compared to the L form, the tryptophan residue have restricted mobility and relatively buried in the A form as compared to the L form. With persistent stress, these metastable assemblies form amyloid-like aggregates. Thus, we conclude that TDP-43^{tRRM} can act as stress-sensor and respond to stress by undergoing reversible structural changes.

b) ***Chapter 4:*** Under prolong stress-like condition, we observed that the monomeric native form of TDP-43^{tRRM} undergoes dramatic change in its conformation to form β-sheet rich amyloid-like assemblies. In this study, we used a range of spectroscopic probe to target different structural events during the protein aggregation process. We kinetically mapped the process to establish four distinct steps in the aggregation reaction – very fast, fast, slow and very slow. The very fast phase corresponds to the formation of 'partially unfolded molecules' that undergo elongation and conformational conversion to form 'higher order intermediates' in fast step. The

temporal order of elongation and conformation conversion process depends upon the protein concentration. We observed that the HOI further undergo structural re-arrangement to form protofibrils in the slow step of the reaction. Further, the assembly of protofibrils results in the formation of amyloid-like assembly. We do not observe lag phase in any of the probes across all the protein concentration range studied, suggesting that the aggregation reaction follows an isodesmic polymerization mechanism. Our results conveyed that the process of TDP-43^{tRRM} aggregation follows a multi-step reaction pathway, forming different types of intermediates with distinct spectroscopic properties.

c) ***Chapter 5:*** The TDP-43^{tRRM} consist of two disease associated mutations, P112H and D169G, which are electrostatic in nature. We wanted to address the role of these electrostatic point mutations on the aggregation kinetics of TDP-43^{tRRM}. We observed that under low pH, both the mutants undergo aggregation. However, the P112H mutation showed substantial increase rate of aggregation as compared to D169G which was highly comparable to the TDP-43^{tRRM}. Additionally, we also observed a weak ionic strength dependency of P112H as compared to D169G and TDP-43^{tRRM}. Overall our data proves that the mutant P112H contributes significantly to the aggregation process. We propose in additional to other protonated residues, the histidine at 112 position might also undergo protonation. The synchronized effect of this protonation and salt might lower the activation barrier associated with the aggregation reaction. Our data thus suggest an important role of electrostatic interactions in determining the transition state of the reaction.

d) ***Chapter 6:*** We understand that electrostatic mutations play a critical role in governing the interactions during the aggregation formation. In order to further explore the role of electrostatics, we systematically investigated the thermodynamic and kinetic dependency of the protein aggregation by varying pH and salt, which are important modulators of electrostatics. The pH-induced structural changes results in intra-molecular destabilization effect which is a critical requirement for initiating the protein aggregation process. Salt affects the aggregation kinetics in two steps through two transition state. Overall our data suggests that electrostatics play a co-operative

role in promoting favourable intra- and intermolecular interactions to gives rise to amyloid-like aggregates.

e) ***Chapter 7:*** In the above chapters, we showed that under low pH and physiological ionic strength, we observe a dramatic change in the backbone conformation from the native structure consisting of α-helix, β-sheet and loop regions to a structure rich in cross-β sheet in the amyloid-like aggregate. In the final chapter, we have deduced the structural core, dynamics and the heterogeneity among the secondary structural segments of the aggregate at the site-specific level using hydrogen-deuterium exchange coupled to mass spectrometry. We observed a significant level of structural protection in the aggregates as compared to the native form. The two secondary structural elements 124-131 and 248-255 forms the core of the aggregates. While monitoring the deuterium uptake, we observed that the RRM1 region remains significantly more protected relative to the RRM2 region.

8.2 Contributions to the field:

Our study has made following important contributions to the field:

a) We have shown how TDP-43^{tRRM} can act as biosensor and detect environmental stress by undergoing metastable assembly formation with weak intermolecular interactions. Our data supports the evolving hypothesis suggesting some proteins can act as stress-sensor and respond to stress quickly by metastable assembly formation.

b) Along the aggregation energy landscape of TDP-43^{tRRM}, we observe multiple species which are linked kinetically through isodesmic polymerization mechanism. The kinetic pathway mapping the aggregation reaction showed multiple species with distinct spectroscopic characteristics.

c) Disease-associated electrostatic point mutation P112H shows increased rate of aggregation. The kinetic barriers of the aggregation reaction is highly sensitive to the electrostatic interactions.

d) The core of the amyloid-like aggregate of TDP-43^{tRRM} consists of secondary structural segments spanning 124-131 and 248-255, which are discontinuous regions. We observed heterogeneity within the amyloid-like aggregates supporting the current notion that there exists polymorphism within the amyloid-like conformations.

8.3 Future Directions:

a) We observe disorder-to-order transition during the native to amyloid-like transition. We know through experimental data that the loop regions in the protein turn into β-sheet rich structure, but we do not clearly understand the role of α-helix in this transition. In order to address this, we can use site-specific FRET pairs within the α-helix and monitor the change in the distance as the helix unfolds and re-arranges to form sheet like structure.

b) Our current study deduces the internal core of the amyloid-like aggregate, but we do not yet understand how during the kinetic conversion from the native to amyloid-like aggregate, the changes in the secondary structural elements occur as a function of kinetic reaction. In order to address this, we can use pulse HDX coupled to mass spectrometry. We can use deuterium pulse to label the protein molecule at different time point of the aggregation reaction. We have experimental rate constant of the aggregation reaction, using the rate constant we can decide the pulse label time and understand the structural changes during the aggregation reaction.

c) The single point mutation P112H shows a drastic increase in the chemical stability of the protein molecule. Using HDX coupled to mass spectrometry, we can point out the secondary structural elements that contribute towards increased stability by calculating their opening and closing rates. This information can be further used to calculate the stability of the individual segments which can further be co-related to the global stability of the protein.

ABSTRACT

Name of the Student: Meenakshi Pillai   **Registration No. : 10BB16A26037**
Faculty of Study: Biochemical Science   **Year of Submission: 2022**
CSIR Lab: NCL, Pune   **Name of the Supervisor: Dr. Santosh K. Jha**

 Title of the thesis: Mapping the energy landscape of amyloid-like aggregation of the nucleic acid binding domain of TDP-43

Stress-induced misfolding and aggregation have been the underlying cause of many neurodegenerative diseases. Multiple studies have shown the effect of stresses on TDP-43 protein, implicated in ALS and FTLD but what remains to be understood is how the functional form of the protein can detect stress and undergo subsequent molecular level changes upon stress detection. In this thesis, we have probed the early structural changes that occur due to stress detection in the functional RRM domain of TDP-43 (TDP-43^{tRRM}). We observed that during low pH stress, protonation of buried ionizable residue is linked to a conformational opening reaction to form partially unfolded molecules which further assemble in a time-dependent manner to form metastable assemblies. These assemblies were highly sensitive to the solvation condition and could further undergo a conformational transition to form an ordered β sheet-rich structure. Interestingly, different stresses result in the formation of different kinds of metastable assemblies. We further kinetically mapped the structural events to show that the transition from native to ordered β sheet-rich amyloid-like aggregates occurs in four distinct steps following an isodesmic polymerization mechanism. The temporal sequence of elongation and conformational conversion were dependent on the protein concentration, which suggested the presence of two pathways under low and high protein concentrations. We further dissected the role of disease-relevant electrostatic mutations, P112H and D169G, on the distinct steps of the aggregation pathway. P112H mutation showed a substantial increase in the aggregation rate, while the mutant D169G behaved similarly to TDP-43^{tRRM} protein. Moreover, P112H also showed a weak ionic strength dependency compared to D169G and TDP-43^{tRRM}, helping us examine the role of targeted electrostatic mutations on the misfolding and aggregation of the protein. In order to further explore the role of electrostatics, we systematically investigated the effect of pH and salt on the thermodynamic and kinetic dependency of aggregation. We observed that the kinetics is modulated in two steps through two transition states during the aggregate formation. Finally, we deduced the internal core, dynamics and structural heterogeneity within the amyloid-like aggregates using hydrogen-deuterium exchange coupled to mass spectrometry. Our study provides strong evidence to support the hypothesis that proteins can act as biosensors and undergo multiple structural changes to form metastable assemblies that can act as precursor for amyloid-like aggregates.

List of Publications

1. Pillai, M. and Jha, S.K. The folding and aggregation energy landscapes of tethered RRM domains of human TDP-43 are coupled via a metastable molten globule-like oligomer. Biochemistry, 2019, 58, 608–620.

2. Pillai, M. and Jha, S.K. Early metastable assembly during the stress-induced formation of worm-like amyloid fibrils of nucleic acid binding domains of TDP-43. Biochemistry, 2020, 59, 315–328.

List of Papers at national/international conferences/seminars

1. Presented poster at "Annual Meeting of the Indian Biophysical Society" at IISER, Pune, India. March, 2018.

An amyloidogenic native-like molten globule intermediate during the pH induced transition of TDP-43

Amyotrophic lateral sclerosis is a type of neurodegenerative disease characterized by accumulation of intracellular inclusion of TDP-43 in the neurons of central nervous system. TDP-43 is primarily a nuclear protein involved in many normal physiological functions that renders this protein as absolutely essential for normal growth and development. RNA Recognition Motif (RRM) forms the functional domain of TDP-43 (TDP-43^{tRRM}). Under pH and salt stress, we observed that TDP-43^{tRRM} undergoes a transition to form an intermediate that has substantial secondary structure and a disordered tertiary structure, called the A form. Here we describe the characteristics of the A form as monitored by fluorescence spectroscopy, size exclusion chromatography, circular dichroism spectroscopy, anisotropy, Thioflavin T and 8-anilino-1-napthalenesulfonate (ANS). The A form with exposed hydrophobic patch undergoes amyloid-like conformational change in presence of salt, denaturant or thermal fluctuation as measured by increase in ThT binding and a change in CD signal at 216 nm. The A form is reversible when the pH stress is removed suggesting that the metastable intermediate acts as a link between folding and aggregation energy landscape. The results suggest that accessibility of amyloidogenic state can occur from a local, partially unfolded state.

2. Presented poster at "Biophysics Paschim" at IIT, Bombay, India. August 2019.

Early metastable intermediate during the stress-induced transition of nucleic acid binding domain of TDP-43

Stress-induced misfolding and intraneuronal aggregation of the highly conserved nucleic acid binding protein TDP-43 and its fragments have been implicated in amyotrophic lateral sclerosis and multiple other neurodegenerative diseases. However, the physico-chemical mechanism of its misfolding from the functional folded state is poorly understood. We show that under low pH conditions, TDP-43^{tRRM} undergoes a conformational opening reaction that is coupled to the protonation of buried charged residues which further assembles to form a metastable structure termed as the L form. The structure of the L form is held by weak interactions and therefore it shows a steep dependence on the ionic strength of the solution. Under persistent stress, the L form undergoes a conformational change to form β-sheet rich 'β form' in at least two distinct step. Moreover, the structure of the L form and the previously studied A form shows considerable difference suggesting that different stresses can result in different kinds of assembly formation. Overall our data indicates that TDP-43^{tRRM} can act as stress-sensor and undergo transition to form metastable assemblies under stress-like condition.

3. Presented poster on National Science Day at CSIR-NCL, India. February 2021.

A step-wise mechanism of the misfolding and the aggregation of the nucleic acid binding domain of TDP-43

Stress-induced aggregation of TDP-43 has been implicated in the neurodegeneration of ALS and various other TDP-43 proteinopathies. However we do not understand the temporal order of events upon stress detection. In this study we delineate the step-wise mechanism of the aggregation of TDP-43^{tRRM}. The formation of the β form occurs in multiple steps and has been kinetically mapped using a set of probes such a fluorescence, Thioflavin T, CD and DLS targeting different steps during the aggregation process. The very first and the fast step is the formation of MG-like species monitored through a decrease in the fluorescence. At low protein concentration, we observed that the process of conformational conversion precedes the elongation and oligomerization process, resulting in formation of conformationally converted molecules. However at high protein concentration, the process of elongation occurs faster, resulting in formation of elongated molecules. This is followed by conformational conversion and increase in the hydrodynamic radius, leading to formation of higher order species of size around 10 nm. Interestingly, we observed that there is further association of these higher order species to form β form. This association of higher order species to β form results in both change in the conformation and the hydrodynamic radius of the protein molecules. For all the probes, we observed that there is no lag phase across different protein concentration, indicating that the process follows isodesmic polymerization mechanism. Overall, our study indicates that the process of formation of β form occurs in multiple step and the pathway is dictated by the protein concentration.
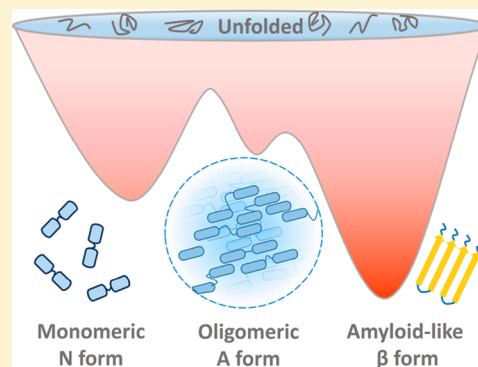
Publications

Biochemistry

The Folding and Aggregation Energy Landscapes of Tethered RRM Domains of Human TDP-43 Are Coupled via a Metastable Molten Globule-like Oligomer

Meenakshi Pillai and Santosh Kumar Jha*

Physical and Materials Chemistry Division, Academy of Scientific and Innovative Research (AcSIR), CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune 411008, Maharashtra, India

S *Supporting Information*

ABSTRACT: Stress-induced misfolding and intraneuronal aggregation of the highly conserved nucleic acid binding protein TDP-43 (transactive response DNA binding protein 43 kDa) and its fragments have been implicated in amyotrophic lateral sclerosis and several other neurodegenerative diseases. However, the physicochemical mechanism of its misfolding from the functional folded state is poorly understood. TDP-43 is a four-domain protein and performs the essential nucleic acid binding function with the help of its two tandem RNA recognition motif domains naturally tethered by a linker (called here the tethered RRM domain of TDP-43 or TDP-43^{tRRM}). Here, we show that the monomeric native form of TDP-43^{tRRM} remains in a pH-dependent and reversible thermodynamic equilibrium with a protonated, nanosized, 40-meric form (the A form). Under the stress-like low-pH condition, the A form becomes predominantly populated. In the A form, protein molecules have restricted dynamics of surface side-chain residues but native-like secondary structure. This self-assembled form possesses a loosely packed core in which the intrinsically disordered and aggregation-prone regions are in the proximity. The A form is metastable and swiftly aggregates into a highly stable amyloid-like protofibrillar form (β form) mediated by the disorder-to-order transition of intrinsically disordered regions upon small environmental perturbations. Interestingly, the A form and the β form are not formed when TDP-43^{tRRM} is bound to DNA, indicating that the nucleic acid binding regions of the protein participate in their formation. Our results reveal how the energy landscapes of folding and aggregation of TDP-43^{tRRM} are coupled by a metastable molten-globule like oligomeric form and modulated by stress-like conditions.

TDP-43 (transactive response DNA binding protein 43 kDa) is a highly conserved and ubiquitously expressed nucleic acid binding protein that participates in vital cellular processes in the nucleus and the cytoplasm, including gene transcription, translation, and mRNA processing.[1,2] Misfolding and prion-like aggregation of TDP-43 and its fragments in neurons into a toxic conformation, primarily induced by chronic stress, have been implicated in the pathology of a wide spectrum of fatal neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD), Alzheimer's disease, Parkinson's disease, Huntington's disease, and Guam-Parkinsonism dementia, collectively known as TDP-43 proteinopathies.[1,3−5] Most importantly, in the case of the incurable motor neuron disease ALS, ∼97% of the patients show deposition of TDP-43 and its fragments in motor neurons in the brain and spinal cord.[6] Despite its central role in multiple neurodegenerative diseases, the molecular mechanism of misfolding and aggregation of TDP-43 from its functional native form is very poorly understood. In particular, we do not understand how the protein begins to misfold. It is important to understand the nature and structure of early intermediates during the misfolding process because for many

neurodegenerative diseases the oligomeric intermediates en route to aggregation have been shown to be much more toxic than the final fibrillar state, and hence, early oligomeric intermediates are important targets for therapeutic intervention.[7,8]

TDP-43 is a 414-amino acid residue protein consisting of four different domains (Figure 1A). The two tandem RNA recognition motifs, RRM1 and RRM2, of the protein are naturally tethered by a 15-residue linker [called hereafter the tethered RRM domain of TDP-43 or TDP-43^{tRRM} (Figure 1A)] and are flanked by an N-terminal domain and a C-terminal domain (UniProtKB/Swiss-Prot entry Q13148). The N-terminal domain has a structural fold similar to that of the axin 1 protein,[9] and the C-terminal domain is largely disordered.[10] The aggregation-prone regions of TDP-43 are primarily located in the TDP-43^{tRRM} and the C-terminal domain.[11−15] However, the mechanistic role and contribution of different regions of the protein to the aggregation process
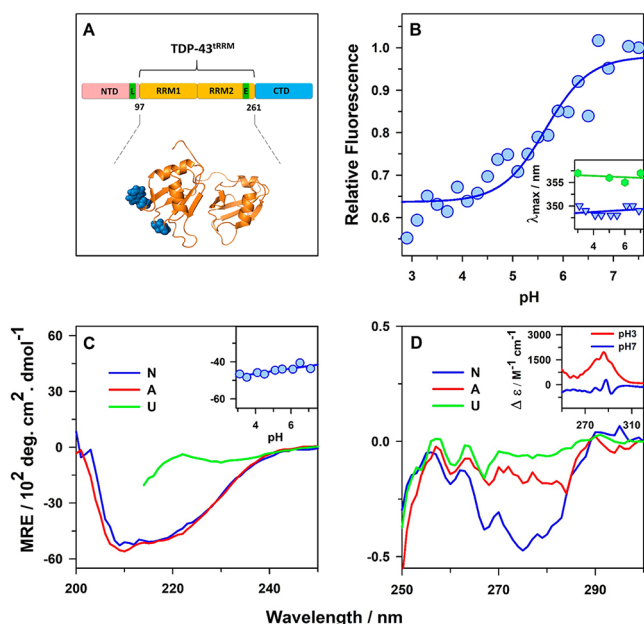
Figure 1. pH-induced formation of the A form of TDP-43[tRRM] that has N-like hydration and secondary structure but a disrupted tertiary structure. (A) Schematic representation of different domains of human TDP-43 (top). Abbreviations: NTD, N-terminal domain; RRM1 and RRM2, two tandem RNA recognition motifs (TDP-43[tRRM]); CTD, C-terminal domain. The nuclear localization sequence (L) in NTD and the nuclear export sequence (E) in the RRM2 domain are also shown. TDP-43[tRRM] (amino acid residues 97−261) is studied in this work, and its solution structure (Protein Data Bank entry 4BS2) is shown at the bottom. The side chains of the two tryptophan residues, W113 and W172, of TDP-43[tRRM] that are located in the RRM1 domain are shown as blue spheres. (B) Change in the fluorescence emission of the protein at 340 nm plotted as a function of pH. The solid blue line through the data is a nonlinear least-squares fit to eq 3. The inset of panel B shows the λ_{max} of the tryptophan residues as a function of pH, when incubated in the absence (blue inverted triangles) and presence of 6 M GdmCl (green hexagons). (C) Far-UV CD and (D) near-UV CD spectra of TDP-43[tRRM] for the N form, the A form, and the U form at pH 7. The inset of panel C shows the change in mean residue ellipticity (MRE) at 222 nm as a function of pH. The inset of panel D shows the protein was incubated at pH 7 and 3 in the absence and presence of 6 M GdmCl, and the difference in the UV absorbance spectrum between the two conditions at pH 7 and 3 is shown.

have remained elusive. TDP-43[tRRM] is the principal functional region of TDP-43 that performs the essential nucleic acid binding functions of the protein.[15−18] The binding of nucleic acids to the TDP-43[tRRM] region increases the solubility and thermodynamic stability of the protein and reduces the level of protein aggregation.[19,20] TDP-43[tRRM] also contains two disease-associated mutations,[21] P112H and D169G, and multiple disease-relevant caspase and calpain cleavage sites.[22] The highly conserved mutation in ALS, D169G, has been shown to increase the thermodynamic stability of both apo and DNA-bound TDP-43[tRRM].[14] Surprisingly, very little is known about the initial structural events and the nature of aggregation-primed early intermediates that couple the folding and aggregation energy landscape of TDP-43[tRRM]. For example, we do not understand whether the aggregation of TDP-43[tRRM] begins with the formation of protein droplet-like intermediates as has been observed for the N-terminal domain and the C-terminal domain of TDP-43.[10,23] It is also not

understood when and how the conformation of the protein molecule changes from the $\alpha+\beta$-containing structure to the predominantly β-sheet rich toxic assembly and whether the aggregated state resembles amyloids.[17,24,25]

One of the prime reasons for pathological assembly of TDP-43 is environmental stress.[17,26] In one of the models of its pathological aggregation, during stress, native TDP-43 reversibly associates to form stress granules primarily via the TDP-43[tRRM] region, which could undergo irreversible aggregation upon prolonged stress.[26] It has been shown recently that some stress-like conditions like starvation acidify the cytoplasm,[27−29] which could perturb the electrostatic interactions and promote disease-prone self-assembly of proteins.[30,31] In this study, using an array of spectroscopic methods and tools of thermodynamics and kinetics, we show for the first time that TDP-43[tRRM] remains in a pH-dependent reversible equilibrium with an oligomeric state in which protein molecules have native-like secondary structure but a disrupted tertiary structure that is fully populated at pH 3 (termed the acidic or A form hereafter). The A form is highly metastable and readily undergoes an amyloid-like misfolding to a predominantly β-sheet rich aggregated state upon small environmental perturbations. Our results shed important light on the earliest step of misfolding of TDP-43[tRRM] from its folded state.

■ MATERIALS AND METHODS

Protein Expression and Purification. The pGEX-6p1 plasmid containing the gene for the human TDP-43[tRRM] (amino acid residues 97−261 from TDP-43, which is UniProtKB/Swiss-Prot entry Q13148) was a generous gift from J. A. Zitzewitz.[15] The details of the plasmid are described elsewhere.[15] The gene contains a His$_6$ tag and a PreScission protease cleavage site for the removal of the His$_6$ tag. The plasmid was transformed into BL21(DE3) *Escherichia coli* cells. The cells were grown to an OD$_{600}$ of 0.8 in Luria-Bertani (LB) broth containing 100 μg/mL ampicillin. Then, the cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 20 °C for 24 h to overexpress the protein. The cells were lysed by sonication in lysis buffer [20 mM NaP$_i$ (pH 7.4), 300 mM NaCl, 30 mM imidazole, and DNaseI]. After pelleting the cell debris, we passed the supernatant through Ni Sepharose 6-Fast Flow beads (GE Healthcare) pre-equilibrated with lysis buffer. The beads were washed with lysis buffer, and the protein was eluted with elution buffer containing 300 mM imidazole. The eluted protein was buffer exchanged in protease cleavage buffer [50 mM Tris buffer (pH 8.0) and 1 mM dithiothreitol (DTT)] using a HiPrep 26/10 desalting column (GE Healthcare) followed by overnight cleavage of the His$_6$ tag at 15 °C using PreScission protease. Anion exchange chromatography using a HiPrep Q FF 16/10 column (GE Healthcare) was performed to obtain pure His$_6$ tag-cleaved protein. The pure protein was buffer exchanged in storage buffer [10 mM KP$_i$ (pH 7.2), 150 mM KCl, and 1 mM DTT]. The protein was >98% pure as determined by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) (Figure S1A). Electrospray ionization mass spectrometry showed the presence of a single protein species with a molecular weight of 19429 Da corresponding to the mass of the His$_6$ tag-cleaved TDP-43[tRRM] protein (Figure S1B).

Buffer, Solutions, and Experimental Conditions. All of the chemicals used are of the highest purity grade obtained from Sigma and Sisco Research Laboratories (SRL). The

following buffers were used at different pHs: 20 mM glycine for pH 2.9−3.5, 20 mM sodium acetate for pH 3.7−5.3, 20 mM 2-(N-morpholino)ethanesulfonic acid (MES) for pH 5.5−6.7, and 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS) for pH 6.9−7.5. All of the buffers contained 1 mM DTT. All of the buffers and solutions were filtered with a 0.2 μm filter before use. The protein concentration was determined by measuring the absorbance at 280 nm using a molar extinction coefficient of 15470 M^{-1} cm^{-1}.

The buffer for the N form consisted of 20 mM MOPS and 1 mM DTT (pH 7). The buffer for the A form consisted of 20 mM glycine and 1 mM DTT (pH 3). The N form and the A form were equilibrated for 3 h at room temperature before any spectroscopic measurement was performed. The unfolding buffer at pH 7 (pH 7U) consisted of 20 mM MOPS, 6 M guanidinium chloride (GdmCl), and 1 mM DTT (pH 7). The unfolding buffer at pH 3 (pH 3U) consisted of 20 mM glycine, 6 M GdmCl, and 1 mM DTT (pH 3). The concentration of GdmCl was determined by measuring the refractive index on an Abbe type refractometer from Rajdhani Scientific Instruments Co. (model RSR-2).

pH-Induced Equilibrium Structural Transition Monitored by Fluorescence. All of the fluorescence measurements were taken on a FluoroMax-4 spectrofluorometer (HORIBA Scientific). The protein samples (8 μM) at different pHs were equilibrated for 4−5 h at room temperature before the fluorescence measurements were taken. The fluorescence spectrum of each sample was acquired by exciting the tryptophan residues at 295 nm and collecting the emission from 310 to 400 nm. The excitation and emission slit widths were set to 1 and 8 nm, respectively. Data were averaged over three acquisitions. The fluorescence spectra of the native and GdmCl-unfolded protein were also acquired with the excitation and emission slit widths set to 1.5 and 3 nm, respectively. The wavelength of maximum fluorescence emission (λ_{max}) in the native (350 nm) and unfolded (357 nm) forms remains invariant with respect to the slit width settings.

Circular Dichroism. Circular dichroism (CD) measurements were taken on a Jasco J-815 spectropolarimeter. The spectra were collected using a data pitch of 1 nm, a data integration time of 4 s, a bandwidth of 2 nm, and a scan speed of 20 nm/min. Each CD spectrum was the average of three scans. For the far-ultraviolet (far-UV) CD measurements, the spectra were collected in the wavelength range of 200−250 nm using a quartz cuvette with a path length of 0.1 cm. For the near-UV CD measurements, the range was set to 250−300 nm and the measurements were taken using 1 cm path length cuvette. The concentration of the protein used for the far-UV CD measurements was 25 μM and for the near-UV CD measurements was 60 μM. The CD spectrum of the buffer was subtracted from the CD spectrum of each sample.

Difference UV Absorbance Spectroscopy. A double-beam UV−visible spectrophotometer (UV 3200) from LABINDIA analytical was used to measure the absorbance of all of the protein samples. Initially, a stock solution of protein (600 μM) was prepared in pH 7 buffer. From this stock, an equal amount of protein sample was added to the following four buffers [pH 3, pH 3U, pH 7, and pH 7U (see above)], so that the final concentration of the protein remains the same in all of the samples (15 μM). The absorbance of all of the samples was measured over a wavelength range of 250−320 nm in a 1 cm quartz cuvette. Each sample was blanked with its

respective buffer. The absorbance values were divided by the protein concentration to obtain a spectrum of the molar extinction coefficient for each sample. The difference absorbance spectrum at pH 3 was obtained by subtracting the extinction coefficient spectrum of pH 3U from that of the pH 3 protein sample. A similar spectrum at pH 7 was obtained by subtracting the extinction coefficient spectrum of pH 7U from that of the pH 7 protein sample.

Steady-State Fluorescence Anisotropy Measurements. Steady-state fluorescence anisotropy was measured on a FluoroMax-4 spectrofluorometer (HORIBA Scientific) that is coupled to an FM4-Pol polarization accessory in the L-format configuration. Protein samples (10 μM) at different pH values varying from pH 3 to 7 were equilibrated at room temperature for 4−5 h. The anisotropy data were collected by exciting the tryptophan residues at 280 nm and collecting emission at 340 nm for 300 s, using the anti-photobleaching setting. The excitation and emission slit width were set to 2 and 10 nm, respectively, while an integration time of 1 s was used. The steady-state fluorescence anisotropy (r) is related to emission intensity I as follows:

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G} \tag{1}$$

where the subscripts V (vertical) and H (horizontal) refer to the position of polarizers in the excitation beam (first subscript) and the emission beam (second subscript) for every measurement. The instrumental correction factor, G, is equal to the I_{HV}/I_{HH} ratio.

Dynamic Light Scattering (DLS). The DLS measurements were taken in a DynaPro 99 unit (Wyatt). The buffers used for DLS were filtered thrice with a 0.2 μm filter. All of the tips and centrifuge tubes were washed with 0.2 μm filtered MilliQ water. The N form (25 μM) and the A form (25 μM) were centrifuged at 14000 rpm for 15 min just before the DLS measurement. The protein sample was placed in a 60 μL cuvette maintained at 25 °C inside the sample chamber of the instrument. The sample was illuminated with a laser with a wavelength of 829.4 nm, and the scattering intensity at 90° and the autocorrelation function were measured simultaneously. The following settings were used: acquisition time, 3 s; signal-to-noise threshold, 2.5; and sensitivity, 90%. The mean autocorrelation function was obtained by averaging over 70 acquisitions and used to obtain the distribution of hydrodynamic radii (R_H) using DynaLS software (Wyatt).

Size Measurement Using a High-Molecular Weight Cutoff Centrifugal Concentrator. The N form (5 μM) and the A form (5 μM) were spun in a Vivaspin 20 (100 kDa molecular weight cutoff) centrifugal concentrator from GE Healthcare. The protein sample was centrifuged at 1000 rpm for 5 min, and the absorbance values of the inside and outside fractions were measured over a wavelength range of 250−300 nm.

8-Anilino-1-naphthalenesulfonic Acid (ANS) Fluorescence Assay. A stock solution of 10 mM ANS was prepared in dimethyl sulfoxide (DMSO). The concentration of ANS was determined by measuring the absorbance at 350 nm and using an extinction coefficient of 5000 M^{-1} cm^{-1}.[32] The N form (2 μM) and the A form (2 μM) were incubated with varying concentrations of ANS, ranging from 0 to 100 μM, for 15 min. After incubation, the fluorescence spectrum of each sample was measured by excitation at 380 nm and monitoring the emission

from 400 to 600 nm. The excitation and emission slit widths were 1 and 5 nm, respectively. For fluorescence resonance energy transfer (FRET) measurements, the protein was excited at 295 nm and the emission was monitored from 320 to 580 nm.

Reversibility of the N ⇌ A Transition. The A form (40 μM) was prepared from the N form by a pH jump from 7 to 3. Following this, the A form was diluted 10-fold into pH 7 buffer and pH 3 buffer (A form control) by manual mixing (dead time of ~15 s), and the temporal change in fluorescence was monitored for each sample immediately after mixing. A sample of 40 μM N form was also diluted 10-fold into pH 3 buffer and pH 7 buffer (N form control) by manual mixing, and the temporal change in fluorescence was monitored for each sample immediately after mixing. The intrinsic tryptophan fluorescence of each sample was recorded by exciting at 295 nm and monitoring the emission from 310 to 400 nm. The excitation and emission slit widths were set to 1.1 and 10 nm, respectively. The final protein concentration in all of the samples was 4 μM. Each sample was also equilibrated at the room temperature for 2 h, and the fluorescence signal, ANS binding, and steady-state fluorescence anisotropy were measured.

Characterization of the A ⇌ β Transition. The A form (25 μM) transforms into a β-sheet rich oligomeric form (β form) upon being heated in pH 3 buffer for 2 h at 40 °C or by being incubated in pH 3 buffer containing KCl or GdmCl. The A ⇌ β transition was characterized spectroscopically using far-UV CD and thioflavin T (ThT) fluorescence. The far-UV CD spectrum of the β form was measured as described above. To monitor the kinetics of formation of the β form, the equilibrated A form was heated in a 0.1 cm cuvette in a PTC-424S/15 Peltier system attached to the CD instrument. The change in ellipticity at 216 nm was measured at different times during the structural transition. For the kinetic experiment, a bandwidth of 2 nm and a data integration time of 4 s were used.

ThT Fluorescence Assay. The ThT fluorescence assay was performed in 50 mM Tris buffer (pH 8). In the assay solution, the protein concentration was 2 μM and the ThT concentration was 40 μM. The required amount of protein was withdrawn from the heated protein sample (25 μM) and cooled to 25 °C before being added to assay buffer. The ThT fluorescence was measured within 1 min of protein addition. The following experimental settings were used: excitation wavelength, 440 nm; emission wavelength, 482 nm; excitation slit, 1 nm; and emission slit, 10 nm. An identical setting was used to measure the fluorescence of the buffer containing the free dye. All of the data have been adjusted for the background fluorescence contributed by the free dye.

Size Exclusion Chromatography. The hydrodynamic properties of the N form and the salt-induced β form were studied using size exclusion chromatography (SEC) on an AKTA Pure M FPLC system (GE Healthcare). SEC was performed on a HiLoad 16/600 Superdex 200 pg column having a fractionation range of 10−600 kDa and a void volume of 39.8 mL. SEC was performed by loading a 25 μM protein sample onto the column, pre-equilibrated at the respective pH. All of the buffers used for SEC contained 150 mM KCl. A flow rate of 0.8 mL/min was used. SEC was also performed for the N form on a HiLoad 16/600 Superdex 75 pg column (Figure S1C).

For the determination of the apparent molecular weight (M_w^{app}) of the N form and the salt-induced β form, we first created a calibration curve between the partition coefficient (K_{av}) of four standard biomolecules (myoglobin, ovalbumin, albumin, and IgG) and their respective molecular weights, as described in the GE Healthcare manual. The value of K_{av} was calculated from the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \tag{2}$$

where V_e is the elution volume of the protein, V_0 is the void volume of the column, and V_t is the bed volume. The values of V_0 and V_t were taken from the column specification sheet, while the values for V_e for all of the known standard biomolecules were taken from the manufacturer-provided manual of the HiLoad Superdex 16/600 200 pg gel filtration column. We used the experimentally measured elution volumes of the N form and the salt-induced β form to calculate their partition coefficient and in turn determined their apparent molecular weight from the calibration curve.

GdmCl-Induced Equilibrium Unfolding Transitions. For the equilibrium unfolding transition, protein samples (4 μM) were incubated with varying concentrations of GdmCl at room temperature for 3 h. The unfolding transition was measured by exciting the tryptophan residues at 295 nm and recording the fluorescence spectrum from 310 to 400 nm. Buffer-corrected data were plotted at 340 nm as a function of GdmCl concentration. It is important to note that the A form transforms itself into a β form even with a minute amount of GdmCl. Hence, for pH 3, the equilibrium unfolding transition represents the β ⇌ U transition.

Experiments with DNA-Bound TDP-43^{tRRM}. TDP-43^{tRRM} was incubated in the presence of $(TG)_6$ single-stranded DNA for 2 h at pH 7.5 to allow the formation of the protein−DNA complex. To check the effect of DNA binding on the formation of the A form, pH 3 buffer was added to the protein−DNA complex. Additionally, 150 mM KCl was added in the pH 3 sample to examine the formation of the β form. The samples were analyzed using far-UV CD and the ANS binding assay as described above. The final concentrations of protein and DNA in the far-UV CD experiments were 15 and 18 μM and in the ANS binding assay were 2 and 5 μM, respectively.

■ RESULTS AND DISCUSSION

pH-Induced Structural Transition of TDP-43^{tRRM}. We recombinantly expressed and purified human TDP-43^{tRRM} from *E. coli* (Materials and Methods). Under physiological conditions at pH 7, TDP-43^{tRRM} is monomeric as determined by size exclusion chromatography (Figure S1C) and is hereafter termed the native (N) form. To understand the coupling of the thermodynamic energy landscape of folding, misfolding, and aggregation, we perturbed the electrostatic interactions of TDP-43^{tRRM} by modulating the pH of the protein environment. TDP-43^{tRRM} contains two tryptophan residues, W113 and W172, which are located in the RRM1 domain (Figure 1A). We observed that the intrinsic tryptophan fluorescence of TDP-43^{tRRM} decreases in a sigmoidal fashion when the pH of the environment is decreased from 7.5 to 2.9 (Figure S2A and Figure 1B). We fitted the pH-dependent fluorescence data to a model in which the structural transition from the N form to the A form is coupled to the protonation of

a single ionizable group, given by a transformed Henderson−Hasselbalch equation[33]

$$Y_{obs} = \frac{Y_A + Y_N \times 10^{pH-pH_m}}{1 + 10^{pH-pH_m}} \qquad (3)$$

where Y_{obs} denotes the observed fluorescence signal for a given pH value, Y_A and Y_N are the signals of the protonated A form and the deprotonated N form, respectively, and pH_m is the midpoint of the transition. The value of pH_m was determined to be 5.6. The observation that a structural change is required for protonation suggests that the ionizable group is buried inside the protein structure.

Tryptophan Residues in the A Form Have N-like Solvation. We observed that the solvation environment of the two tryptophan residues remains N-like during the N ⇌ A transition, as measured by the fluorescence Stokes shift (Figure S2B and Figure 1B, inset). The λ_{max} in the N form is 350 nm, which shifts to 357 nm upon unfolding of the protein in 6 M guanidinium chloride (GdmCl) at pH 7 (U form) (Figure S2B and Figure 1B, inset). We observed that there is no pH-dependent change in the λ_{max} of the tryptophan residues between pH 7 and 3, in the absence or presence of 6 M GdmCl. The λ_{max} in the A form is 350 nm and in the GdmCl-unfolded U form at pH 3 is 357 nm (Figure 1B, inset). In addition to the similar λ_{max}, the unfolded forms at pH 7 and 3 also have nearly identical far-UV CD spectra (Figure S2C). Together, these results suggest that tryptophan residues have similar solvation in the N form and the A form and that the unfolded forms at pH 7 and 3 are similar.

The Protein Molecules in the A Form Resemble a Molten Globule. We measured the global secondary and tertiary structure of the N form and the A form using far-UV and near-UV CD spectroscopy (Figure 1C,D). The mean residue ellipticity (MRE) at 222 nm, which is a measure of the total content of the secondary structure (α-helix + β-sheet) of proteins, in the N form is -4559 ± 403 deg cm^2 dmol^{-1} and in the U form is -548 ± 239 deg cm^2 dmol^{-1}. In comparison, the MRE at 222 nm in the A form is -5063 ± 437 deg cm^2 dmol^{-1}, which is very similar to the value in the N form. There is almost no pH-dependent change in the secondary structure during the N ⇌ A transition (Figure 1C, inset). In contrast, we observed that the A form has a highly disrupted tertiary structure compared to that of the N form. The near-UV CD spectrum of TDP-43^{tRRM} is dominated by the chirality of its thirteen phenylalanine (wiggles in the spectrum in the 255−270 nm region), three tyrosine, and two tryptophan residues (giving rise to a broad band around 278 nm). The MRE at 278 nm in the N form is -40 deg cm^2 dmol^{-1} and in the U form is -5 deg cm^2 dmol^{-1}. The MRE at 278 nm in the A form is -16 deg cm^2 dmol^{-1}, which is very similar to the MRE in the U form. These results indicate that in the A form, the asymmetric packing of the hydrophobic residues in the core is disrupted but the secondary structure remains N-like. Hence, the structure of the protein molecules in the A form is similar to a molten globule (MG).

Differential Molecular Arrangement of the Aromatic Amino Acid Residues in the N Form and the A Form. The change in the UV absorbance spectrum of buried aromatic residues upon unfolding is a sensitive indicator of their molecular environment. The exposure of aromatic residues to the polar aqueous solvent upon partial or complete unfolding leads to a blue shift in the absorbance wavelength and overall

changes in the intensity.[34] The spectral shift manifests itself in the difference spectrum between the folded and unfolded forms as peaks in 285−288 nm region for tyrosine residues and around 291−294 nm for tryptophan residues.[34] We measured the difference in the UV absorbance spectrum of the N form and the U form at pH 7 and the A form and the U form at pH 3 (Figure 1D, inset). The difference spectrum at pH 7 has very small peaks at 279, 286, and 292 nm, while in contrast, intense peaks are observed for the same wavelengths at pH 3. These results indicate that there is a large change in the extinction coefficient of the tyrosine and tryptophan residues upon unfolding of the A form, compared to that upon unfolding of the N form. We observed that this difference in the extinction coefficient is primarily due to the different molecular arrangement of aromatic amino acids in the A form compared to the N form and not due to the change in the polarity of the solvent. The difference in the UV absorbance spectrum of the unfolded forms at pH 3 and 7 is a good measure of the change in the absorbance spectrum due to the change in the polarity of the solvent. We observed that the difference in the absorbance spectrum of the unfolded forms at pH 3 and 7 has small and broad peaks (Figure S2D). In contrast, the difference in the absorbance spectrum of the N form and the A form has sharp and highly intense peaks at 284 nm (with a shoulder at 276 nm) and 292 nm (Figure S2D). The highly intense peaks compared to the small and broad peaks of the difference spectrum of the unfolded forms indicate that the difference in the molar extinction coefficient during the unfolding of the N form and the A form is not due to the changes in the polarity of the solvent but mostly due to the different molecular environments of the tyrosine and tryptophan residues in the N form and the A form.

Tryptophan Residues Have Restricted Mobility in the A Form. Surprisingly, we observed that the local dynamics of the two tryptophan residues is highly restricted in the A form compared to the N form (Figure 2A). We measured the pH-dependent changes in the steady-state fluorescence anisotropy of the tryptophan residues to understand the changes in their local dynamics during the N ⇌ A transition. The mean value of the steady-state anisotropy at 340 nm in the N form is 0.079 $\pm$ 0.002, which increases in an apparently sigmoidal fashion as a function of pH, to the value of 0.095 $\pm$ 0.002 in the A form. To rule out the possibility that the increase in anisotropy in the A form is not due to the decrease in the quantum yield of the tryptophan residues (Figure 1B and Figure S2A), we measured their steady-state anisotropy in the U form. Although the quantum yield of the U form is smaller than that of the N form, the steady-state anisotropy remains considerably smaller (Figure 2A, inset) because of the higher degree of flexibility accessible to the tryptophan side chain in the U form. This result indicates that the anisotropy experiment is primarily reporting on the flexibility of the fluorophore. Hence, the increase in anisotropy during the N ⇌ A transition is due to the restricted dynamics of the tryptophan residues in the A form compared to that of those in the N form.

The A Form Is Oligomeric. The near-UV CD and absorbance experiments suggest that the packing in the core of the A form is disrupted and hence the side-chain residues are expected to be more flexible and dynamic in the A form, but the steady-state anisotropy experiments suggest that the side chains of the tryptophan residues are dynamically restricted. The two tryptophan residues of TDP-43^{tRRM} are located very close to the surface of the protein (Figure 1A). We explored
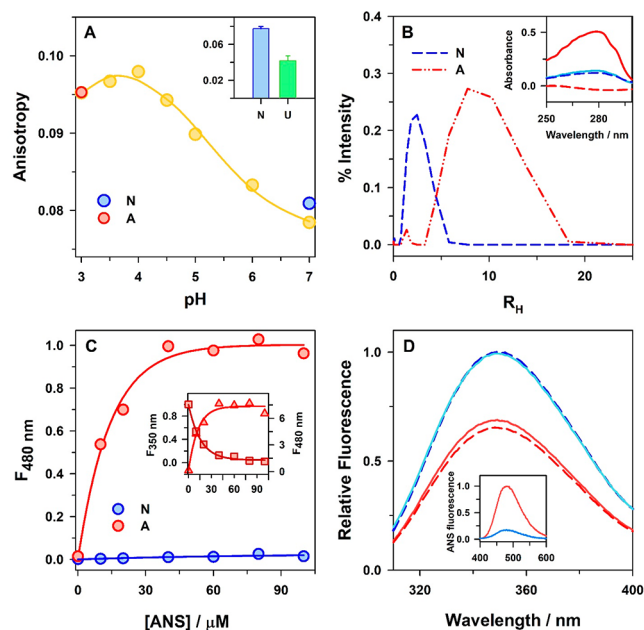
Figure 2. The A form has restricted motion of tryptophan side chains, is larger, has a single ANS binding site, and forms reversibly from the N form. (A) Change in fluorescence anisotropy of tryptophan residues at 340 nm as a function of pH (yellow circles). The solid yellow line through the data is drawn to guide the eyes. The red and blue circles represent the signal of the A form before and after, respectively, its transfer to the native condition at pH 7. The inset of panel A shows a comparison of the fluorescence anisotropy of tryptophan residues in the N form and U form. The error bar represents the spread in the data from three independent experiments. (B) Distributions of R_H as monitored by DLS. The inset of panel B shows the N form (blue line) and the A form (red line) when spun in a Vivaspin 20 centrifugal filter (molecular weight cutoff of 100 kDa). The UV absorbance spectrum of the solution inside (solid line) and outside (dashed line) the centrifugal filter is shown when the volume of the solution remaining inside the filter was one-fifth of the original volume. (C) Comparison of the binding of the ANS dye with the N form and the A form. The relative change in the intensity of the ANS fluorescence at 480 nm upon excitation at 380 nm is plotted vs varying concentrations of ANS. The inset of panel C shows the FRET between the tryptophan residues of TDP-43^{tRRM} and the protein-bound ANS dye in the A form. The changes in the fluorescence intensity of tryptophan residues at 350 nm (red square) and the ANS dye at 480 nm (red triangle) upon excitation at 295 nm are plotted vs the concentration of ANS, according to the left and right y-axes, respectively. The solid red lines through the data in panel C and its inset are nonlinear least-squares fits to eq 4, and the solid blue line through the N form data in panel C is drawn to guide the eyes. (D) Reversibility of the N ⇌ A transition. Fluorescence spectrum of the N form (blue solid line), formed after a jump from pH 3 to 7. The red solid line indicates the fluorescence spectrum of the A form formed after a jump from pH 7 to 3. As a control, the blue and red dashed lines represent the fluorescence spectra of the N form and the A form, respectively. In the inset of panel D, the red and blue lines show the ANS binding of the A form before and after, respectively, its transfer to the native condition at pH 7. The fluorescence emission spectra of ANS are shown after excitation at 380 nm.

whether the restriction in the dynamics of the surface residues in the A form is due to the self-assembly of the protein into an oligomeric structure. We measured the R_H of the N form and the A form using DLS (Figure 2B). We observed that the mean R_H of the A form is 8.5 nm, which is significantly larger than the mean R_H of the N form (2.3 nm). The width of the

distribution of R_H is larger in the A form than in the N form, indicating that the formation of the A form is accompanied by an increase in heterogeneity in the size of the population of protein molecules. These results indicate that the A form is oligomeric in nature.

In previous studies, data on the correlation among the molecular weight (M_w), the radius of gyration, and the R_H have been tabulated on a very large number of globular proteins.[35,36] We constructed a calibration curve using these previously reported data to estimate the M_w of the N form and the A form from their mean R_H values. The estimated M_w of the N form was around 20 kDa, which is very similar to the mass measured using mass spectrometry (Figure S1B). The M_w of the A form was estimated to be around 800 kDa. These results indicate that the A form is roughly a 40-mer.

We also confirmed the larger size of the A form compared to the N form by passing them through a 100 kDa molecular weight cutoff centrifugal filter. The inset of Figure 2B shows the absorbance of the protein retained and passed by the filter in the N form and the A form. We observed that the N form passed through the filter almost completely. In contrast, no protein molecule passed through the filter in the A form and almost all of the protein molecule is retained by the filter. These results indicate that almost the entire population of the protein molecules has formed the oligomeric A form at pH 3.

The A Form Has a Loosely Packed Hydrophobic Core. We observed that the A form binds to ANS, a dye that is known to bind the loosely packed hydrophobic core of MG.[37] The N form does not bind to ANS, but the fluorescence intensity of ANS increases dramatically in the presence of the A form (Figure S3A). We measured the binding of ANS with the A form and the N form as a function of ANS concentration. The N form does not bind to ANS even when the ANS concentration is as high as 100 μM (Figure 2C). In contrast, the binding of the A form increases hyperbolically with an increase in the ANS concentration (Figure 2C). To further characterize the binding of ANS to the A form, we performed FRET experiments (Figure 2C, inset). The absorbance spectrum of ANS overlaps with the fluorescence emission spectrum of tryptophan, and they form a FRET pair.[37] We observed that when the tryptophan residues of TDP-43^{tRRM} in the A form are excited at 295 nm, their fluorescence emission intensity decreases with an increase in the ANS concentration, which is accompanied by a concomitant increase in ANS fluorescence (Figure 2C, inset). These results suggest that ANS is binding to the A form near the tryptophan residues. We analyzed the binding of ANS with the A form using the following ligand binding equation:[37]

$$\Delta F(L) = n\Delta F_{max}\frac{[L]}{[L] + K_D} \qquad (4)$$

where $\Delta F(L)$ is the change in fluorescence upon addition of ANS at a concentration $[L]$, n is the number of identical, non-interacting binding sites for ANS, ΔF_{max} is the change in fluorescence when all of the molecules of the A form are bound to ANS, and K_D is the dissociation constant for the binding of ANS to the A form. All of the data in Figure 2C and its inset fitted well to eq 4. The mean values of n and K_D obtained from fits are 1.1 and 11 μM, respectively. These results indicate that the A form has a loosely packed hydrophobic core where ANS binds with a dissociation constant of around 11 μM. The observation that the 40-meric A form has only one binding site

for ANS indicates that the A form has a symmetrical and spherical micelle-like structure with a centrally located hydrophobic core, as has been observed previously for the low-pH intermediate of a small protein barstar.[38]

The N ⇌ A Transition Is Reversible. Although the A form is significantly distinct from the N form with respect to its side-chain packing, local dynamics, and size, it is interesting to note that the N ⇌ A transition is fully reversible. The fluorescence intensity of the tryptophan residues of TDP-43[tRRM] is significantly quenched upon formation of the A form from the N form. We observed that the fluorescence intensity of tryptophan residues is fully recovered and becomes similar in value to that of the N form when the A form is transferred to the native conditions at pH 7 (Figure 2D). The reversible change in fluorescence during the N ⇌ A transition (during both the forward and the reverse reactions) occurs within the dead time (~15 s) of the manual-mixing experiment (Figure S4). Upon being transferred to the native condition at pH 7, the A form also loses its ability to bind ANS (Figure 2D, inset) and regain the flexible local dynamics of the tryptophan residues (Figure 2A), like the N form. These results suggest that the N ⇌ A transition is fast and completely reversible in nature.

The A Form Is an Aggregation-Primed Intermediate and Is Highly Amyloidogenic. Intriguingly, we observed that the A form is amyloidogenic in nature. The far-UV CD spectrum of the A form is almost identical to that of the N form (Figure 1C) and shows minima at 209 and 222 nm, a characteristic signature of $\alpha+\beta$ protein structure (Figure 1A). We observed that upon heating the A form at 40 °C for 2 h, it undergoes a structural change to an amyloid-like conformation that contains predominantly β-sheet rich structure (β form), which is evident from the huge dip in the far-UV CD spectrum at 216 nm (Figure 3A). The MRE at 216 nm increases from -5645 ± 611 to -10365 ± 815 deg cm^2 dmol^{-1} during this structural change. Interestingly, the N form does not transform to the β form upon heating for 2 h at 40 °C (Figure 3A, inset). There is only a marginal change in the far-UV CD signal at 210 nm upon heating, which might be due to the partial unfolding of the protein as 40 °C being very close to the beginning of its thermal unfolding transition.[14] Figure 3B compares the kinetics of formation of the β form from the A form and the N form at 40 °C as determined by the measurement of the change in MRE at 216 nm. The transformation in the secondary structure during the A ⇌ β transition occurred in a single-exponential phase with a mean time constant of 26 min (Figure 3B). There is no observable lag phase. In contrast, for the N form, the MRE at 216 nm remains almost constant as a function of heating time.

The β form binds to the amyloid staining dye ThT[39] but not the A form or the N form (Figure 3C). The λ_{max} of the free ThT dye (in the absence of any protein) is 440 nm, which shifts to 482 nm in the presence of the β form (Figure 3C). The shift in λ_{max} was also accompanied by a dramatic increase in the fluorescence emission intensity. These results indicate that the β form strongly binds to ThT. In contrast, the A form and the N form do not show any binding to ThT, indicating the absence of any amyloid-like ordered β-sheet structure. In combination with the results of the far-UV CD experiments (Figure 3A,B), these findings strongly indicate that ordered β-sheet structure, akin to what has been observed in amyloids or soluble precursor protofibrils, is present in the β form but not in the A form or the N form.
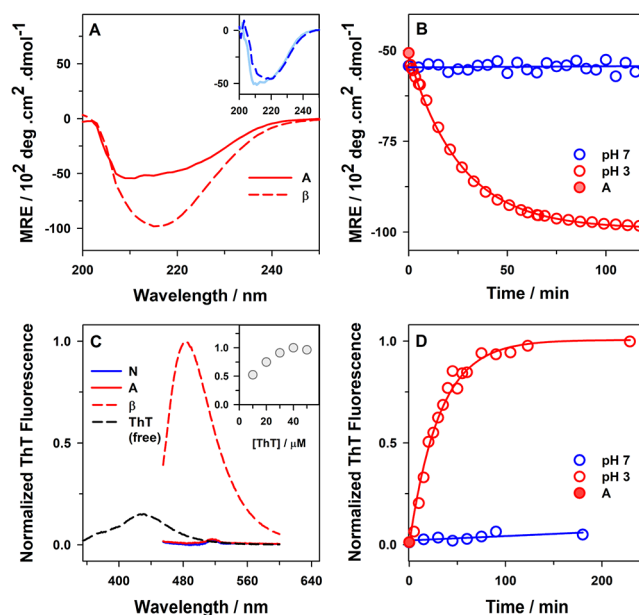

Figure 3. The A form is metastable and amyloidogenic and undergoes an amyloid-like misfolding to a β-sheet rich β form upon being heated. (A) Far-UV CD spectra of 25 μM A form before and after heating (β form) at 40 °C for 2 h. The inset shows the far-UV CD spectra of 25 μM N form before (solid blue line) and after heating (dashed blue line) at 40 °C for 2 h. (B) Comparison of the kinetics of the conformational change of 25 μM TDP-43[tRRM] at pH 7 and 3 at 40 °C, as monitored by the change in mean residue ellipticity at 216 nm. (C) Changes in the fluorescence emission spectra of the ThT dye upon binding to the N form, the A form, and the β form. The excitation wavelength in the case of the free dye was 340 nm, and in all other cases, it was 440 nm. The inset shows the change in the fluorescence intensity of ThT measured at 482 nm in the presence of the β form (2 μM) as a function of ThT concentration. (D) Comparison of the kinetics of the conformational change of 25 μM TDP-43[tRRM] at pH 7 and 3 at 40 °C, as monitored by the change in ThT fluorescence at 482 nm. In panels C and D, all of the data have been normalized to unity to the value of the ThT signal of the β form at 482 nm. In all of the panels, the β form is formed after heating 25 μM A form at 40 °C for 2 h. In panels B and D, the solid red lines through the data are fits to a single-exponential equation and the solid blue lines are drawn to guide the eye.

We observed that the fluorescence intensity of ThT depends upon the [dye]:[protein] ratio in the assay solution (Figure 3C, inset). The inset in Figure 3C shows the variation of ThT emission intensity as a function of ThT concentration when the concentration of the β form was kept fixed at 2 μM. When the dye concentration is 10 μM (5:1 [dye]:[protein]), the ThT signal is only 50% of the highest saturating signal, indicating that only half of the available binding sites are occupied by ThT. The ThT emission intensity did not increase and remained constant when the ThT concentration was >30 μM (Figure 3C, inset), indicating that all of the available binding sites are occupied by ThT when the [dye]:[protein] ratio is >15:1. In all of the ThT experiments in this study, the concentration of ThT (40 μM) was 20-fold greater than the protein concentration (2 μM) in the assay solution.

We monitored the kinetics of formation of the β form from the A form at 40 °C using the change in ThT fluorescence intensity (Figure 3D). ThT fluorescence measures the elongation of the ordered β-sheet structure during the process of amyloid-like assembly.[39] We observed that the increase in
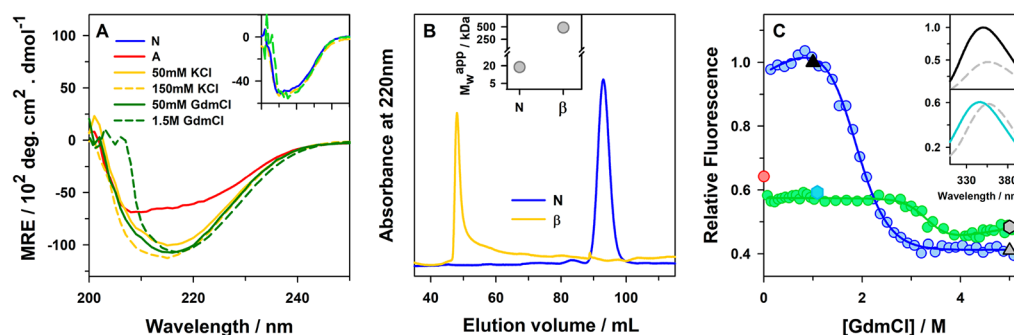
Figure 4. An increase in the ionic strength or partial destabilization of the A form leads to the formation of the large and highly stable misfolded β form. (A) The main panel and its inset show the far-UV CD spectra of TDP-43^{tRRM} at pH 3 and 7 in the absence and the presence of KCl and GdmCl. (B) Size exclusion chromatography of the N form and the β form. The inset shows a comparison of the M_w^{app} values of the N form and the β form. (C) GdmCl-induced equilibrium unfolding transitions of the N form (blue circles) and the β form (green circles) monitored by the change in tryptophan fluorescence at 340 nm. The filled red circle represents the signal of the A form. The solid lines through the data in panel C are drawn to guide the eye. The gray and black triangles represent the signal of the U form [at 5 M GdmCl (pH 7)] before and after, respectively, its transfer to a buffer containing 1.1 M GdmCl (pH 7). The gray and cyan hexagons denote the signal of the U form [at 5 M GdmCl (pH 3)] before and after, respectively, its transfer to a buffer containing 1.1 M GdmCl (pH 3). The insets shows the N $\rightleftharpoons$ U and $\beta \rightleftharpoons$ U transitions are reversible. The top inset shows the fluorescence spectrum of the protein in the U form [at 5 M GdmCl (pH 7)] before (gray dashed line) and after (black solid line) its transfer to a buffer containing 1.1 M GdmCl (pH 7), where the N form is predominantly populated. The bottom inset shows the fluorescence spectrum of the protein in the U form [at 5 M GdmCl (pH 3)] before (gray dashed line) and after (cyan solid line) its transfer to a buffer containing 1.1 M GdmCl (pH 3), where the β form is predominantly populated.

ThT fluorescence occurred in a single-exponential phase with a mean time constant of 31 min, without any lag phase (Figure 3D). The kinetics and the rate constants of formation of the β form are similar when monitored by ThT fluorescence and far-UV CD (panels D and B of Figure 3, respectively). The absence of the initial lag phase during the formation of the β form suggests that the A form is an aggregation-primed intermediate. No change in ThT fluorescence signal was observed when the N form was heated at 40 °C for 2 h, indicating the absence of formation of the ordered β-sheet-containing structure.

The A Form Is Metastable. Apart from the temperature perturbation, the β form is also formed from the A form upon partial destabilization in the presence of a small amount of denaturant or an increase in the ionic strength of the solution. Figure 4A shows the far-UV CD spectra of the A form in the presence of two different concentrations of the denaturant, GdmCl (50 mM and 1.5 M), and two different concentrations of the salt, KCl (50 and 150 mM). The N-like far-UV CD spectrum of the A form changes dramatically in the presence of GdmCl or KCl, with the disappearance of two minima at 209 and 222 nm and the appearance of a huge minimum at 216 nm, indicating the formation of the β-sheet rich structure. The far-UV CD spectra under all four conditions appear to be very similar to each other and to those of the heat-induced β form (Figure 3A). The MRE at 216 nm, under all four conditions, is around −10000 deg cm^2 dmol^{-1}, which is similar to the value of the heat-induced β form. These results suggest that the β form formed in the presence of GdmCl or KCl is similar in structure to the heat-induced β form. In contrast, no changes in the far-UV CD spectrum of the N form were observed in the presence of 1.5 M GdmCl or 150 mM KCl (Figure 4A, inset). The result that the A form is transformed to the β form in the presence of a minute amount of salt or destabilizing agent suggests that the A form is highly metastable.

The Misfolded β Form Is Large and More Resistant to Chemical Denaturation. We compared the M_w^{app} of the β form to that of the N form using size exclusion chromatography on a HiLoad 16/600 Superdex 200 pg gel filtration

column (Figure 4B and Materials and Methods). The column has a fractionation range of 10−600 kDa and a void volume of 39.8 mL. We observed that the elution volume for the N form is 92.7 mL and for the β form is 48 mL (Figure 4B). Using the data on elution volumes, we determined the M_w^{app} (Figure 4B, inset) of the N form and the β form. We first constructed a calibration curve between the known elution volumes and M_w^{app} values of a few standard biomolecules (Figure S5 and Materials and Methods). Using the calibration curve (Figure S5), we determined the M_w^{app} of the N form to be 18.9 kDa and for the β form to be 485 kDa. The M_w^{app} of the N form as determined from size exclusion chromatography is similar to the mass measured using mass spectrometry (Figure S1B). However, the β form elutes at the low-resolution regime of the column (Figure S5), and the calculated mass is only a lower limit of its size. Nevertheless, these results indicate that the β form is a β-sheet-containing soluble protein aggregate that is much larger than the N form.

The A form is highly metastable and transforms to the β form upon either heating or addition of minute amounts of a chemical denaturant (Figures 3A and 4A). The A $\rightleftharpoons \beta$ transition occurs at a GdmCl concentration as low as 2 mM, and the midpoint of the transition is 3.5 mM (Figure S6). The steep dependence at such a low concentration of GdmCl reflects the weak and transient nature of interactions that govern the self-assembled A form. In contrast, we observed that the β form is highly stable. Figure 4C compares the chemical stability of the β form and the N form. We observed that the fluorescence intensity of the tryptophan residues in the N form and the β form changes in a sigmoidal fashion, in an apparently two-state manner, as a function of GdmCl concentration. It is unlikely that as the β form is oligomeric in nature, all of the elements of its structure will unfold in a single step. Similarly, the N form of TDP-43^{tRRM} has also been proposed to unfold via two intermediate states in a previous study.[15] Because of the potential multistep nature of both the N $\rightleftharpoons$ U and the $\beta \rightleftharpoons$ U transitions, we used the midpoint of the unfolding transition as a proxy for chemical stability. The midpoint of the unfolding of the N form is 1.9 M GdmCl,

which is very similar to that observed previously.[15] The midpoint of the unfolding of the β form is 3.3 M, indicating that it is highly resistant to chemical denaturation. Such a high resistance to unfolding reflects the strong nature of interactions in the β form.

It is important to note that both the N ⇌ U and the β ⇌ U transitions are reversible in nature (Figure 4C and its insets and Figure S7). Upon dilution of the denaturant at pH 7, the U form fully regains the fluorescence spectrum of the N form (Figure 4C, top inset). Similarly, upon dilution of the denaturant at pH 3, the U form completely regains the fluorescence spectrum (Figure 4C, bottom inset) and the far-UV CD spectrum of the β form (Figure S7). These results indicate that the β form is highly thermodynamically stable and that the structure in the β form is held by much stronger interactions compared to the A form and the N form.

The Disordered Regions Acquire Ordered β-Sheet Structure during the A ⇌ β Transition. All of the variants of the β form appear to have similar β-sheet contents, which is evident from their similar far-UV CD spectra (Figures 3A and 4A). We observed that the MREs at 216 nm in the β forms are very high compared to those of the A form and the N form. This result is highly surprising. The MRE at 216 nm for the β-sheet is usually lower than that of the α-helix and lowest for the random coil or the disordered segment.[40] Accordingly, if only the α-helical segments of the A form transform to β-sheet during the A ⇌ β transition, the β form would have a lower MRE at 216 nm. However, this is not observed to be the case. It is important to note that a substantial part of the three-dimensional structure of TDP-43^tRRM (Figure 1A) is disordered.[15] The partially disordered nature of the structure of TDP-43^tRRM in both the N form and the A form is also evident in the smaller MREs in their near-UV and far-UV CD spectra (Figure 1C,D) compared to those of the other structured α+β proteins.[37,40] TDP-43^tRRM has a polyampholyte-like primary sequence that is rich in both positively and negatively charged amino acid residues. It contains 26 basic residues (14 lysines, 9 arginines, and 3 histidines) and 25 acidic residues (14 aspartic acids and 13 glutamic acids) that accounts for around one-third of the total number of amino acids in its sequence. Such polyampholyte-like sequences have been theoretically predicted[41] and experimentally observed[42] to behave like intrinsically disordered regions (IDRs). If the disordered regions of TDP-43^tRRM convert to β-sheet during the A ⇌ β transition, as during the aggregation of intrinsically disordered proteins like α-synuclein and amyloid-β,[43,44] that will dramatically increase the MRE at 216 nm. Hence, our results indicate that the IDRs of TDP-43^tRRM have converted to β-sheet in the β forms, in addition to the α-helical segments, if any.

The β Form Resembles an Amyloid-like Protofibril. The β form of TDP-43^tRRM strongly binds to ThT, a dye that is known to bind amyloid fibril-like structures (Figure 3C,D). It is important to note that the ability of the oligomers of different variants and fragments of TDP-43 to bind ThT and their resemblance to amyloids are strongly debated.[17,24,25,45] Aggregates formed by some short peptide fragments of TDP-43 have been shown to bind ThT and resemble amyloids, but many other TDP-43 aggregates do not appear to bind to ThT.[11,24,45,46] In a few cases, for example for full length TDP-43, the aggregates that do not bind ThT do not appear to contain β-sheet rich structure, which is evident from the far-UV CD spectrum.[47] In contrast, in a recent study,[48] some

fragments of TDP-43^tRRM have been shown to form cross-β structure similar to that of amyloids. We observed that the β forms of TDP-43^tRRM strongly bind to ThT (Figure 3C). The ability of the β forms to bind ThT is strongly dependent on the [dye]:[protein] ratio, and an at least 15-fold excess concentration of the dye is required for complete binding (Figure 3C, inset). The β forms have very high β-sheet contents (Figures 3A and 4A). They are oligomeric and much larger than the N form (Figure 4B) but are soluble. Soluble oligomers with β-sheet rich structure often termed protofibrils are observed frequently as the precursor to amyloid fibrils during their assembly.[49,50] It is possible that the β forms observed in this study are similar to amyloid protofibrils. This is an important result because recent studies have shown that soluble oligomers or protofibrils, rather than insoluble fibrils, may be the toxic species in protein aggregation-related neurodegenerative diseases.[51,52] We observed that the kinetics of the A ⇌ β transition studied by far-UV CD and ThT fluorescence (Figure 3B,D) yielded similar rate constants, indicating that the change in the secondary structure (conformational conversion) and elongation of the protofibrillar structure occur simultaneously. However, it is likely that the A ⇌ β interconversion occurs in more than one step. A detailed investigation of the stepwise mechanism of the process is currently underway in our laboratory.

The A Form and the β Form Are Not Formed for DNA-Bound TDP-43^tRRM. We examined the effect of DNA binding on the formation of the A form and the β form (Figure 5). The
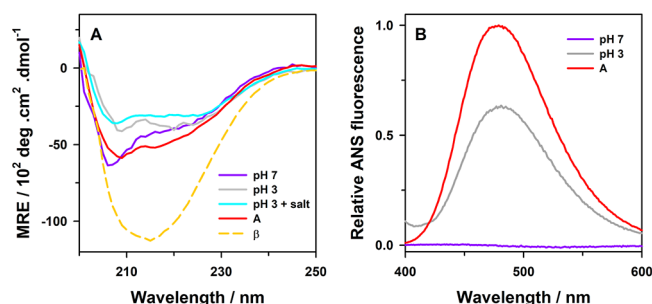
Figure 5. Effect of DNA binding on the formation of the A form and the β form. (A) Far-UV CD spectra of TDP-43^tRRM bound to single-stranded (TG)_6 DNA at pH 7, pH 3, and pH 3 with 150 mM KCl. For comparison, far-UV CD spectra of the A form and the salt-induced β form (150 mM KCl) at pH 3 formed in the absence of DNA are shown. (B) Comparison of the extent of ANS binding of the protein−DNA complex at pH 7 and 3 with the extent of ANS binding of the A form formed in the absence of DNA.

protein−DNA complex at pH 7 shows a far-UV CD spectrum similar to that of the unbound protein (Figures 1C and 5A). This result indicates that the global secondary structure of the N form remains similar upon DNA binding. The far-UV CD spectrum of the protein−DNA complex at pH 3, however, showed a smaller MRE compared to that of the A form (formed in the absence of the DNA), indicating a partially disrupted secondary structure similar to an MG. Interestingly, the far-UV CD spectrum of the protein−DNA complex at pH 3 does not change upon addition of as high as 150 mM KCl. These results imply that the protein−DNA complex does not undergo a disorder-to-order transition as observed for the A ⇌ β conversion for the free protein (Figure 5). Hence, the protein−DNA complex is not metastable at pH 3 and the β form is not formed. We observed that the protein−DNA

complex at pH 7 does not bind ANS (Figure 5B), similar to the free protein (Figure S3A and Figure 2C). In contrast, the ANS binding capacity of the protein−DNA complex at pH 3 is substantially decreased compared to that of the A form (Figure 5B). The weakened ANS binding of the protein−DNA complex at pH 3 shows that the A form is not completely formed but the protein molecules retain the MG character. Combined, these results indicate that the A form and the β form are not formed for DNA-bound TDP-43$^{\text{tRRM}}$.

The results presented above suggest that the DNA binding regions of TDP-43$^{\text{tRRM}}$ are involved in the self-assembly of the A form and that the formation of the A form is an important step toward the formation of the amyloid-like β form. It is possible that RRM1 domain plays a major role in the formation of the A form because nucleic acids bind strongly to the RRM1 domain compared to the RRM2 domain.[16,18] The RRM1 domain is also less stable than the RRM2 domain to chemical and thermal denaturation,[15,18] which might make it more prone to participate in the oligomerization process. The high efficiency of FRET between tryptophan residues located in the disordered regions of the RRM1 domain and ANS also suggests that the RRM1 domain is involved in the formation of the self-assembled A form.

The Metastable A Form Couples the Folding and Misfolding Energy Landscapes. One of the most important results of this study is that the A form is the common link among the energy landscapes of folding, misfolding, and aggregation of TDP-43$^{\text{tRRM}}$ (Figure 6). The A form is formed by the self-assembly of the monomeric protein without misfolding into an oligomeric structure consisting of ~40 molecules of TDP-43$^{\text{tRRM}}$. The protein molecules in the A form have N-like secondary structure but a disrupted tertiary structure (Figures 1 and 2). The A form readily misfolds and aggregates into the β-sheet rich β form (Figures 3 and 4), and the formation of the A form is crucial for the formation of the β form (Figure 5). Hence, the A form couples the folding and the aggregation energy landscapes of TDP-43$^{\text{tRRM}}$.

The N ⇌ A Transition Has Some Similarities to the Monomer ⇌ Droplet Transition of Proteins. It has recently been reported that proteins containing low-complexity IDRs, including the C-terminal domain and the N-terminal domain of TDP-43,[10,23] begin to misfold by condensation of protein molecules into nanometer- to micrometer-sized protein droplets, akin to the vapor-to-liquid phase transition of matter.[10,17,53−56] The 40-meric A form is only 8.5 nm in size, soluble, and not a protein droplet. However, the N ⇌ A transition appears to have some similarities to the monomer ⇌ droplet transition. For example, like the monomer ⇌ droplet transition,[55] a thermodynamic equilibrium exists between the monomeric N form and the self-assembled A form that is highly reversible and dynamic (Figure 2 and Figure S4). The A form is metastable and primed to undergo misfolding and aggregation reaction to the amyloid-like β form upon small environmental perturbations (Figures 3 and 4), akin to the liquid-to-solid phase transition of protein droplets.[55,57,58] The mechanisms and the reasons behind the formation of protein droplets are poorly understood, but one important determinant is the multivalency of the protein molecules, in terms of the presence of the positive, negative, and hydrophobic centers, like Janus particles,[55] for self-assembly.[59,60] The protein molecules in the A form are protonated and have a molten globular structure that can provide multiple charged and hydrophobic centers required for assembly. This is in line
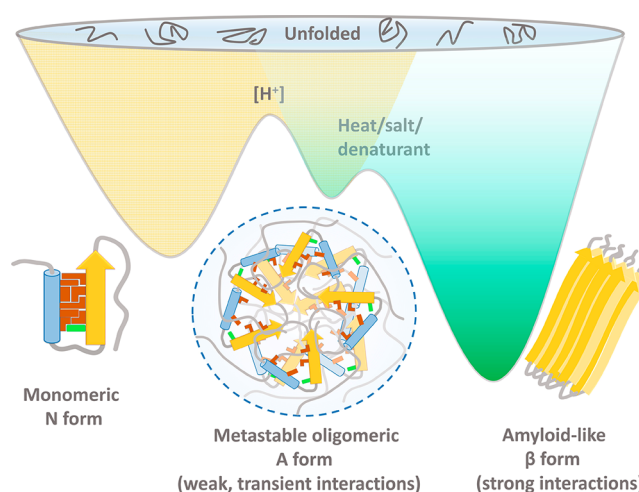

Figure 6. Model energy diagram for the coupling of the folding and misfolding energy landscapes of TDP-43$^{\text{tRRM}}$. The schematic also shows how the changes in environmental conditions like pH, co-solutes, and temperature modulate the energy landscape. Under native-like conditions at pH 7, the yellow patterned portion of the landscape is operational where the monomeric N form is the most stable state. It remains in a pH-dependent thermodynamic equilibrium with a high-energy nanosized 40-meric A form. The protein molecules in the A form retain the secondary structure of the N form but have loose side-chain packing (brown bars). The N ⇌ A transition is coupled to the protonation of at least one buried ionizable side-chain residue (green bar) of the protein. The overall structure of the A form is like a micelle with a loosely packed central hydrophobic core in which the intrinsically disordered and aggregation-prone regions of the protein are in the proximity. The loosely packed structure of the A form is held by weak and transient interactions. At pH 3, the protonated A form is highly stabilized, the N form is destabilized, and the N ⇌ A equilibrium shifts to predominantly populate the A form. Hence, the green portion of the energy landscape becomes operational at pH 3. The metastable A form readily undergoes an amyloid-like conformational conversion and amyloidogenic assembly to a β-sheet rich β form upon partial destabilization (by addition of GdmCl or heating) or an increase in the ionic strength of the medium. The IDRs of the protein acquire ordered β-sheet structure during the A ⇌ β transition. The structure in the β form is amyloid-like and appears to be held by strong interactions. The A form thus couples the folding and aggregation energy landscapes of TDP-43$^{\text{tRRM}}$.

with the recent proposal that globules (like proteins in a θ solvent) can self-assemble to form large structures easily compared to the fully structured native proteins.[55,57,60] The self-assembled A form binds to a single molecule of ANS near the tryptophan residues (Figure 2C and its inset) and has a centrally located loosely packed hydrophobic core similar to a micelle. As tryptophan residues are located in the IDRs of the RRM1 (Figure 1A), the loosely packed hydrophobic core of the A form must involve IDRs of the RRM1, in addition to the possible involvement of the IDRs of RRM2, if any. The IDRs of the A form undergo a disorder-to-order transition during the A ⇌ β conversion in the presence of co-solutes like KCl and GdmCl (Figure 4), similar to the liquid-to-solid transition of the droplets of other proteins.[57,58,61] It appears that the high local concentration of the protein molecules in the A form brings the aggregation-prone and IDRs of TDP-43$^{\text{tRRM}}$ into the proximity of each other (Figure 6) and makes it prime to undergo misfolding and aggregation reaction upon small environmental perturbations.

pH Modulates the N ⇌ A Transition. The partitioning of TDP-43[tRRM] into the monomeric N form versus the oligomeric A form is highly dependent on the pH of the environment (Figures 1 and 2). The population of the A form increases as the pH decreases from 7 to 3 (Figures 1 and 2). The midpoint of the N ⇌ A transition is at pH 5.6 (Figure 1B). The formation of the A form is coupled to the protonation of at least one buried charged residue (Figure 1B). Among the amino acid residues that could titrate near pH 5.6, one histidine (H166) and three aspartic acids (D105, D138, and D247) are almost fully buried inside the protein structure. It is possible that one of these residues might have a slightly modified pK_a due to the burial, and its protonation is coupled to the N ⇌ A transition. It is important to note that many disease-related mutations of TDP-43,[21] like in the case of many variants of prions,[62] perturb the electrostatic interactions of the protein.

Crossing of the Major Unfolding Barrier Is Not Required for Misfolding. Although the structure of the A form resembles an MG, we do not observe the population of a monomeric MG during the N ⇌ A transition. Nevertheless, the MG-like structure of the protein molecules in the highly amyloidogenic A form suggests that the misfolding and aggregation of TDP-43[tRRM] could begin from the partially unfolded states having N-like secondary structure and that the complete crossover of the unfolding barrier is not required for misfolding, as has been proposed for a few proteins, including acylphosphatase, human β2 microglobulin, and human trans-thyretin.[63]

Modulation of the Folding and Misfolding Energy Landscapes by Stress-like Conditions. The results of this study report on only the RRM domains of TDP-43 and do not reveal the structural changes that occur in the N-terminal and C-terminal domains of the protein. Previous studies have shown that both the N-terminal and the disordered C-terminal domains play an important role in the aggregation of TDP-43 and assemble through liquid−liquid phase separation-driven droplet formation.[10,17] A majority of the disease-associated mutations have also been observed in the C-terminal domain of the protein.[21] It is important to note, however, that the genetic mutations contribute in only ∼5% of the ALS cases (including familial as well as sporadic), and in the remaining ∼95% of the cases, the disease is caused by environmental stress.[6,17] An attractive alternative model of the pathological aggregation of TDP-43 is that environmental stress induces the reversible association of native TDP-43 into granules primarily via its TDP-43[tRRM] region.[26] Upon chronic exposure to stress, TDP-43 associated with the stress granules undergo irreversible aggregation into disease-prone assemblies.[26,64,65] It is important to note that some stress-like conditions like starvation and energy depletion decrease the pH of the cellular environment.[27−29] It has been reported recently that acidification of a cellular environment by stress leads to the perturbation of electrostatic interactions and formation of disease-prone large assemblies of proteins like yeast prion Sup35[31] and poly(A) binding protein.[30] In the case of stress granule-associated poly(A) binding protein, higher-order metastable assemblies form via its RRM domain.[30] It has also been suggested that at the molecular level, proteins might sense stress using the protonation−deprotonation of the protein side chains.[29,31] In the current study, we observed that the monomeric N form of TDP-43[tRRM] exists in a pH-dependent reversible thermodynamic equilibrium with a metastable molten globule-like oligomeric A form. The A form is fully populated only below pH 4.5 (Figure 1B), in response to the low-pH stress-like conditions, and transforms to the β form with a further increase in the level of stress. Our results shed light on how stress-like conditions might modulate the energy landscape of TDP-43[tRRM] inside the cell. Our data suggest (see the model in Figure 6) that under physiological conditions at neutral pH, the fully folded form of the protein is populated predominantly. Upon environmental stress that perturbs the electrostatic interactions of the protein, the protein self-assembles into the aggregation-primed intermediate A form. Upon prolonged heat or salt stress or in the presence of destabilizing conditions, the equilibrium shifts to populate the disease-relevant β form. Hence, the results of this study reveal a physicochemical and thermodynamic mechanism of coupling and modulation of the folding and the aggregation energy landscapes of TDP-43[tRRM].

■ ASSOCIATED CONTENT

Ⓢ **Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b01013.

> SDS−PAGE and mass spectrometry data on protein purity (Figure S1A,B); size exclusion chromatography of the N form (Figure S1C); changes in the fluorescence spectrum during the pH-induced N ⇌ A transition (Figure S2A); local solvation environment around tryptophan residues in the N form, the A form, and the U form (Figure S2B); far-UV CD spectra of the unfolded protein at pH 7 and 3 (Figure S2C); dissection of the changes in the difference UV absorbance spectrum due to the changes in the solvation environment and molecular structure (Figure S2D); ANS binding of the N form and the A form monitored by changes in ANS fluorescence and FRET between bound ANS and tryptophan residues (Figure S3); kinetics of the N ⇌ A transition (Figure S4); calibration curve for molecular weight determination using size exclusion chromatography (Figure S5); sensitivity of the A ⇌ β transition to GdmCl concentration (Figure S6); and reversibility of the β ⇌ U transition monitored by far-UV CD (Figure S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sk.jha@ncl.res.in. Telephone: 91-20-25902588. Fax: 91-20-25902615.

ORCID Ⓘ

Santosh Kumar Jha: 0000-0003-1339-7409

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Lee, E. B., Lee, V. M. Y., and Trojanowski, J. Q. (2012) Gains or losses: Molecular mechanisms of TDP43-mediated neurodegeneration. *Nat. Rev. Neurosci. 13*, 38−50.

(2) 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.

(3) 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. Y. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science 314*, 130−133.

(4) 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.

(5) Chen-Plotkin, A. S., Lee, V. M. Y., and Trojanowski, J. Q. (2010) TAR DNA-binding protein 43 in neurodegenerative disease. *Nat. Rev. Neurol. 6*, 211−220.

(6) Ling, S. C., Polymenidou, M., and Cleveland, D. W. (2013) Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron 79*, 416−438.

(7) Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science 300*, 486−489.

(8) Caughey, B., and Lansbury, P. T. (2003) Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci. 26*, 267−298.

(9) Mompean, M., Romano, V., Pantoja-Uceda, D., Stuani, C., Baralle, F. E., Buratti, E., and Laurents, D. V. (2016) The TDP-43 N-terminal domain structure at high resolution. *FEBS J. 283*, 1242−1260.

(10) Conicella, A. E., Zerze, G. H., Mittal, J., and Fawzi, N. L. (2016) ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure 24*, 1537−1549.

(11) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc. 132*, 1186−1187.

(12) Shodai, A., Morimura, T., Ido, A., Uchida, T., Ayaki, T., Takahashi, R., Kitazawa, S., Suzuki, S., Shirouzu, M., Kigawa, T., Muto, Y., Yokoyama, S., Takahashi, R., Kitahara, R., Ito, H., Fujiwara, N., and Urushitani, M. (2013) Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kda TDP-43. *J. Biol. Chem. 288*, 14886−14905.

(13) Wang, Y. T., Kuo, P. H., Chiang, C. H., Liang, J. R., Chen, Y. R., Wang, S., Shen, J. C., and Yuan, H. S. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *J. Biol. Chem. 288*, 9049−9057.

(14) Austin, J. A., Wright, G. S. A., Watanabe, S., Grossmann, J. G., Antonyuk, S. V., Yamanaka, K., and Hasnain, S. S. (2014) Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc. Natl. Acad. Sci. U. S. A. 111*, 4309−4314.

(15) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem. 289*, 8264−8276.

(16) Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ule, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol. 20*, 1443−1449.

(17) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry 56*, 809−823.

(18) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res. 37*, 1799−1808.

(19) Huang, Y. C., Lin, K. F., He, R. Y., Tu, P. H., Koubek, J., Hsu, Y. C., and Huang, J. J. T. (2013) Inhibition of TDP-43 aggregation by nucleic acid binding. *PLoS One 8*, No. e64002.

(20) Sun, Y. L., Arslan, P. E., Won, A., Yip, C. M., and Chakrabartty, A. (2014) Binding of TDP-43 to the 3′ UTR of its cognate mRNA enhances its solubility. *Biochemistry 53*, 5885−5894.

(21) Buratti, E. (2015) Functional significance of TDP-43 mutations in disease. *Adv. Genet. 91*, 1−53.

(22) Wobst, H. J., Delsing, L., Brandon, N. J., and Moss, S. J. (2017) Truncation of the TAR DNA-binding protein 43 is not a prerequisite for cytoplasmic relocalization, and is suppressed by caspase inhibition and by introduction of the A90V sequence variant. *PLoS One 12*, No. e0177181.

(23) Wang, L., Kang, J., Lim, L. Z., Wei, Y. Y., and Song, J. X. (2018) TDP-43 NTD can be induced while CTD is significantly enhanced by ssDNA to undergo liquid-liquid phase separation. *Biochem. Biophys. Res. Commun. 499*, 189−195.

(24) Mompean, M., Hervas, R., Xu, Y. Y., Tran, T. H., Guarnaccia, C., Buratti, E., Baralle, F., Tong, L., Carrion-Vazquez, M., McDermott, A. E., and Laurents, D. V. (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett. 6*, 2608−2615.

(25) Mompean, M., Baralle, M., Buratti, E., and Laurents, D. V. (2016) An amyloid-like pathological conformation of TDP-43 is stabilized by hypercooperative hydrogen bonds. *Front. Mol. Neurosci. 9*, 125.

(26) 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.

(27) Petrovska, I., Nuske, E., Munder, M. C., Kulasegaran, G., Malinovska, L., Kroschwald, S., Richter, D., Fahmy, K., Gibson, K., Verbavatz, J. M., and Alberti, S. (2014) Filament formation by metabolic enzymes is a specific adaptation to an advanced state of cellular starvation. *eLife 3*, No. e02409.

(28) Munder, M. C., Midtvedt, D., Franzmann, T., Nuske, E., Otto, O., Herbig, M., Ulbricht, E., Muller, P., Taubenberger, A., Maharana, S., Malinovska, L., Richter, D., Guck, J., Zaburdaev, V., and Alberti, S. (2016) A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife 5*, No. eaa09347.

(29) Kroschwald, S., and Alberti, S. (2017) Gel or die: Phase separation as a survival strategy. *Cell 168*, 947−948.

(30) Riback, J. A., Katanski, C. D., Kear-Scott, J. L., Pilipenko, E. V., Rojek, A. E., Sosnick, T. R., and Drummond, D. A. (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell 168*, 1028−1040.

(31) Franzmann, T. M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A. S., Nuske, E., Richter, D., Baumeister, W., Grill, S. W., Pappu, R. V., Hyman, A. A., and Alberti, S. (2018) Phase separation of a yeast prion protein promotes cellular fitness. *Science 359*, No. eaao5654.

(32) Stryer, L. (1965) The interaction of a naphthalene dye with apomyoglobin and apohemoglbin: a fluorescent probe of non-polar binding sites. *J. Mol. Biol. 13*, 482−495.

(33) 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.

(34) Creighton, T. E. (1997) *Protein structure: A practical approach*, IRL Press at Oxford University Press, Oxford, U.K.

(35) Tyn, M. T., and Gusek, T. W. (1990) Prediction of diffusion coefficients of proteins. *Biotechnol. Bioeng. 35*, 327−338.

(36) Smilgies, D. M., and Folta-Stogniew, E. (2015) Molecular weight-gyration radius relation of globular proteins: A comparison of light scattering, small-angle X-ray scattering and structure-based data. *J. Appl. Crystallogr. 48*, 1604−1606.

(37) Khurana, R., and Udgaonkar, J. B. (1994) Equilibrium unfolding studies of barstar - evidence for an alternative conformation which resembles a molten globule. *Biochemistry 33*, 106−115.

(38) Juneja, J., Bhavesh, N. S., Udgaonkar, J. B., and Hosur, R. V. (2002) NMR identification and characterization of the flexible regions in the 160 kda molten globule-like aggregate of barstar at low pH. *Biochemistry 41*, 9885−9899.

(39) Naiki, H., and Gejyo, F. (1999) Kinetic analysis of amyloid fibril formation. *Methods Enzymol. 309*, 305−318.

(40) Greenfield, N. J. (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc. 1*, 2876−2890.

(41) Das, R. K., and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A. 110*, 13392−13397.

(42) Muller-Spath, S., Soranno, A., Hirschfeld, V., Hofmann, H., Ruegger, S., Reymond, L., Nettels, D., and Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A. 107*, 14609−14614.

(43) Nath, S., Meuvis, J., Hendrix, J., Carl, S. A., and Engelborghs, Y. (2010) Early aggregation steps in α-synuclein as measured by FCS and FRET: Evidence for a contagious conformational change. *Biophys. J. 98*, 1302−1311.

(44) Jha, A., Kumar, M. G., Gopi, H. N., and Paknikar, K. M. (2018) Inhibition of β-amyloid aggregation through a designed β-hairpin peptide. *Langmuir 34*, 1591−1600.

(45) Neumann, M., Kwong, L. K., Sampathu, D. M., Trojanowski, J. Q., and Lee, V. M. (2007) TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: Protein misfolding diseases without amyloidosis. *Arch. Neurol. 64*, 1388−1394.

(46) Saini, A., and Chauhan, V. S. (2014) Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir 30*, 3845−3856.

(47) Fang, Y. S., Tsai, K. J., Chang, Y. J., Kao, P., Woods, R., Kuo, P. H., Wu, C. C., Liao, J. Y., Chou, S. C., Lin, V., Jin, L. W., Yuan, H. S., Cheng, I. H., Tu, P. H., and Chen, Y. R. (2014) Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat. Commun. 5*, 4824.

(48) Guenther, E. L., Ge, P., Trinh, H., Sawaya, M. R., Cascio, D., Boyer, D. R., Gonen, T., Zhou, Z. H., and Eisenberg, D. S. (2018) Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat. Struct. Mol. Biol. 25*, 311−319.

(49) Ionescu-Zanetti, C., Khurana, R., Gillespie, J. R., Petrick, J. S., Trabachino, L. C., Minert, L. J., Carter, S. A., and Fink, A. L. (1999) Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy. *Proc. Natl. Acad. Sci. U. S. A. 96*, 13175−13179.

(50) Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B., and Teplow, D. B. (2003) Amyloid β -protein (Aβ) assembly: Aβ 40 and Aβ 42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U. S. A. 100*, 330−335.

(51) Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campioni, S., Tzitzilonis, C., Soragni, A., Jessberger, S., Mira, H., Consiglio, A., Pham, E., Masliah, E., Gage, F. H., and Riek, R. (2011) In vivo demonstration that α-synuclein oligomers are toxic. *Proc. Natl. Acad. Sci. U. S. A. 108*, 4194−4199.

(52) He, Y., Zheng, M. M., Ma, Y., Han, X. J., Ma, X. Q., Qu, C. Q., and Du, Y. F. (2012) Soluble oligomers and fibrillar species of amyloid β-peptide differentially affect cognitive functions and hippocampal inflammatory response. *Biochem. Biophys. Res. Commun. 429*, 125−130.

(53) Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell 163*, 123−133.

(54) Mitrea, D. M., and Kriwacki, R. W. (2016) Phase separation in biology; functional organization of a higher order. *Cell Commun. Signaling 14*, 1.

(55) Bergeron-Sandoval, L. P., Safaee, N., and Michnick, S. W. (2016) Mechanisms and consequences of macromolecular phase separation. *Cell 165*, 1067−1079.

(56) Li, H. R., Chiang, W. C., Chou, P. C., Wang, W. J., and Huang, J. R. (2018) TAR DNA-binding protein 43 (TDP-43) liquid-liquid phase separation is mediated by just a few aromatic residues. *J. Biol. Chem. 293*, 6090−6098.

(57) Pappu, R. V., Wang, X., Vitalis, A., and Crick, S. L. (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys. 469*, 132−141.

(58) Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell 162*, 1066−1077.

(59) Li, P. L., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., Banani, S. F., Russo, P. S., Jiang, Q. X., Nixon, B. T., and Rosen, M. K. (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature 483*, 336−340.

(60) Brangwynne, C. P., Tompa, P., and Pappu, R. V. (2015) Polymer physics of intracellular phase transitions. *Nat. Phys. 11*, 899−904.

(61) Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E., and Zweckstetter, M. (2017) Liquid-liquid phase separation of the microtubule-binding repeats of the alzheimer-related protein tau. *Nat. Commun. 8*, 275.

(62) Singh, J., and Udgaonkar, J. B. (2016) Unraveling the molecular mechanism of pH-induced misfolding and oligomerization of the prion protein. *J. Mol. Biol. 428*, 1345−1355.

(63) Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol. 5*, 15−22.

(64) Parker, S. J., Meyerowitz, J., James, J. L., Liddell, J. R., Crouch, P. J., Kanninen, K. M., and White, A. R. (2012) Endogenous TDP-43 localized to stress granules can subsequently form protein aggregates. *Neurochem. Int. 60*, 415−424.

(65) Aulas, A., and Vande Velde, C. (2015) Alterations in stress granule dynamics driven by TDP-43 and FUS: A link to pathological inclusions in ALS? *Front. Cell. Neurosci. 9*, 423.

Early Metastable Assembly during the Stress-Induced Formation of Worm-like Amyloid Fibrils of Nucleic Acid Binding Domains of TDP-43

Meenakshi Pillai and Santosh Kumar Jha*

Cite This: https://dx.doi.org/10.1021/acs.biochem.9b00780

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: TDP-43 protein travels between the cytosol and the nucleus to perform its nucleic acid binding functions through its two tandem RNA recognition motif domains (TDP-43^{tRRM}). When exposed to various environmental stresses, it forms abnormal aggregates in the cytosol of neurons, which are the hallmarks of amyotrophic lateral sclerosis and other TDP-43 proteinopathies. However, the nature of early structural changes upon stress sensing and the consequent steps during the course of aggregation are not well understood. In this study, we show that under low-pH conditions, mimicking starvation stress, TDP-43^{tRRM} undergoes a conformational opening reaction linked to the protonation of buried ionizable residues and grows into a metastable oligomeric assembly (called the "low-pH form" or the "L form"). In the L form, the protein molecules have disrupted tertiary structure, solvent-exposed hydrophobic patches, and mobile side chains but the native-like secondary structure remains intact. The L form structure is held by weak interactions and has a steep dependence on ionic strength. In the presence of as little as 15 mM KCl, it fully misfolds and further oligomerizes to form a β-sheet rich "β form" in at least two distinct steps. The β form has an ordered, stable structure that resembles worm-like amyloid fibrils. The unstructured regions of the protein gain structure during L $\rightleftharpoons$ β conversion. Our results suggest that TDP-43^{tRRM} could function as a stress sensor and support a recent model in which stress sensing during neurodegeneration occurs by assembly of proteins into metastable assemblies that are precursors to the solid aggregates.
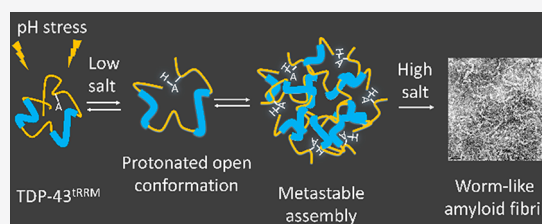
Amyotrophic lateral sclerosis (ALS) is a lethal disease that leads to progressive impairment of motor neurons resulting in weakening of muscles, paralysis, and eventually death.[1] A defining feature of ALS, in almost all of the patients examined to date, is the presence of abnormal aggregates of the TDP-43 (transactive response element DNA binding protein of 43 kDa) protein in the cytosol of motor neurons in the central motor cortex and glial cells.[2−5] The intraneuronal deposits of TDP-43 are also observed in multiple other TDP-43 proteinopathies.[6,7] TDP-43 is a highly conserved nucleic acid binding protein that has an important role in many vital cellular processes like the synthesis of RNA and proteins and the processing of mRNA.[8] It is known that sporadic factors, like chronic environmental stress,[5,6] are responsible for its aggregation in a majority (90−95%) of disease conditions, but the underlying molecular mechanism of the self-assembly process is poorly understood. It has been postulated that under stress-like conditions TDP-43 participates in the formation of large protein assemblies, such as stress granules.[9,10] These assemblies are highly dynamic and metastable and form irreversible aggregates under chronic stress.[9,11,12] However, very little is known about the nature of the initial conformational changes that begin the stress-induced self-assembly and misfolding of TDP-43.

The formation of stress granules, which are large assemblies consisting primarily of proteins and translationally stalled mRNA, has been reported under a variety of mortal environmental stress conditions, such as nutrient starvation[13] and osmotic,[14] thermal,[15] and oxidative stress[16] as a mean of stress adaptation. However, the mechanism by which cells detect stress at the molecular level is not well understood. In one of the models, it has been hypothesized that some proteins in the cell could function as biosensors and sense stress by site-specific protein modifications.[13,17,18] For example, when the cells are exposed to oxidative insults, the cysteine and methionine residues of proteins become oxidized.[19] A change in cytosolic pH due to starvation stress is detected by protonation and/or deprotonation of charged residues.[13,15,17,18,20] Thermal stress is detected by the partial unfolding of proteins.[13,15,18] These site-specific molecular changes trigger the self-assembly of proteins into reversible oligomeric structures with largely native-like (N-like) secon-

dary structure.[13,17,18] It has been postulated that coupling of assembly and disassembly of these oligomers with sensitization and desensitization of environmental stress could be a rapid and energy-efficient method of regulating stress.[13] The initial oligomeric assemblies grow larger[21] and eventually demix and phase separate into a variety of structures with distinct morphologies and material properties (liquids, gels, filaments, colloids, dynamic fibers, etc.), depending upon the stress.[15,17,18] Upon persistent stress, these assemblies irreversibly change their structure to form abnormal solid aggregates,[11,12] which have been implicated in various neurodegenerative diseases like prion disease,[20] ALS,[22,23] and FTLD.[23] To understand the early stages of these neurodegenerative diseases, including ALS, it is critical to understand the nature of the initial conformational changes that allow stress sensing and formation of N-like oligomers. It is also essential to understand how the characteristics of the initial assemblies formed under various stress conditions differ and how they might lead to polymorphism of the diseased condition.

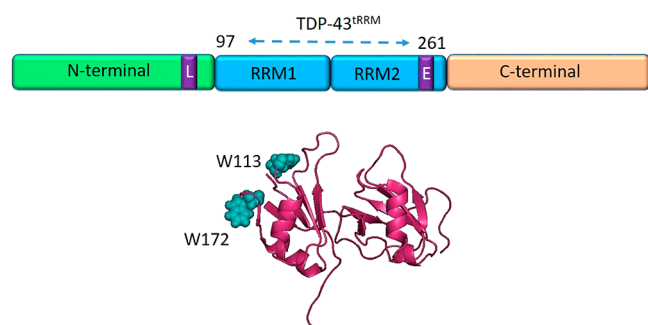Structurally, TDP-43 protein has four distinct domains (Figure 1). The N-terminal domain is involved in dimeriza-

Figure 1. Diagrammatic representation of TDP-43 domains. There are four distinct domains comprising human TDP-43: the N-terminal domain, two RNA recognition motifs (RRM1 and RRM2) together termed TDP-43^{tRRM}, and the C-terminal domain. The letters L and E represent the nuclear localization signal sequence and nuclear export signal sequence, respectively. The bottom part of the figure shows the solution structure of TDP-43^{tRRM} (amino acid residues 97−261) taken from Protein Data Bank entry 4BS2. The tryptophan residues, W113 and W172, situated in the RRM1 domain are represented by blue spheres.

tion,[24] has a ubiquitin-like fold,[25] and has been recently reported to be undergoing aggregation without becoming misfolded.[26] The C-terminal domain is mostly involved in protein−protein interactions[27] and carries a large number of mutations, which have been implicated in the diseased condition.[28] However, mutations are responsible for only 5− 10% of the diseased cases[5] and thus are marginal contributors to the diseased state. It is possible that a large number of mutations occur in the C-terminus because it is intrinsically disordered and hence mutations could be better tolerated than in the highly conserved N-terminus or the nucleic acid binding domains. It is important to note that vastly diverse C-terminal sequences are present across the different species in the otherwise highly conserved TDP-43 protein.[28,29] This also points toward the possibility that mutations may simply be disease modifiers but not the major causal factor as the same mutation has been shown to have different pharmacological/ pathological effects in different patients.[28] TDP-43 also

contains two RNA recognition motif domains (RRM1 and RRM2) in tandem that are linked by a 15-residue natural linker (called together TDP-43^{tRRM}, hereafter). The TDP-43^{tRRM} region is highly conserved and functionally important and binds to nucleic acids in a sequence-specific manner to perform a gamut of nucleic acid binding functions in the cell.[6,30,31] TDP-43^{tRRM} forms abnormal aggregates in various stress-like conditions,[32−34] which is modulated by the presence of nucleic acids.[33−36] TDP-43 is co-deposited in the stress granules with other RNA binding proteins (e.g., PABP, TIA, and eIF3)[37,38] possibly via its TDP-43^{tRRM} region.[35,39] In this work, we show that under low-pH conditions mimicking starvation stress TDP-43^{tRRM} undergoes a partial unfolding reaction that is linked to the protonation of buried ionizable residues and grows into an oligomeric assembly (called hereafter the "low-pH form" or the "L form") in which the protein molecules have an unfolded-like (U-like) tertiary structure and solvation environment but an N-like secondary structure. The L form is extremely metastable and swiftly misfolds into a β-sheet rich amyloid-like aggregated state (the β form) upon a minute change in the ionic strength. The flexible loops of TDP-43^{tRRM} become more ordered during the L to β transition. The results presented here give a fresh perspective on the initial intermediates formed during the aggregation of TDP-43^{tRRM}.

■ MATERIALS AND METHODS

Protein Expression and Purification. The expression and the purification protocol of TDP-43^{tRRM} (UniProtKB/ Swiss-Prot entry Q13148) have been described previously.[34] The pure protein was stored in a buffer containing 10 mM potassium phosphate, 150 mM KCl, and 1 mM dithiothreitol (DTT) (pH 7.2). Sodium dodecyl sulfate−polyacrylamide gel electrophoresis revealed that the protein was highly pure. The molecular weight of the His$_6$ tag-cleaved TDP-43^{tRRM} protein as determined by electrospray ionization mass spectrometry was 19429 Da.

Buffer, Solutions, and Experimental Conditions. All of the reagents are of the highest purity grade procured from Sisco Research Laboratories (SRL) and Sigma. The buffers used at different pH values have been reported previously.[34] A molar extinction coefficient of 15470 M^{-1} cm^{-1} at 280 nm was used to determine the protein concentration.

The N form and the L form were prepared by transferring the required volume of protein from a concentrated stock solution of around 600−800 μM protein, which is in the storage buffer. The buffer used for the N form contained 20 mM MOPS and 1 mM DTT (pH 7.5). The buffer used for the L form contained 20 mM glycine and 1 mM DTT (pH 2.5). The unfolding buffer contained 20 mM MOPS, 6 M guanidinium chloride (GdmCl), and 1 mM DTT (pH 7.5). The concentration of GdmCl was measured as described previously.[34,40] Before their use, all of the buffers were filtered with a 0.2 μm filter. The final concentration of residual KCl in all of the experiments was around 3−5 mM unless otherwise mentioned.

pH-Induced Equilibrium Structural Transition Monitored by Fluorescence. For the pH-induced fluorescence measurements, the protein samples (8 μM) were equilibrated at different pH values at room temperature before the fluorescence spectrum was recorded on a FluoroMax-4 spectrofluorometer (HORIBA Scientific). The excitation wavelength was set to 295 nm, and the emission was collected from 310 to 400 nm. The excitation slit width was 1 nm, while

the emission slit width was 8 nm. Each fluorescence spectrum was averaged over three scans. The fluorescence spectra recorded after incubation for 5 and 24 h remained identical.

The data on the pH dependence of the change in fluorescence of the protein are fitted to a model in which the structural change from pH 7 to 2.5 occurs in two steps (N $\rightleftharpoons$ A $\rightleftharpoons$ L) corresponding to the protonation of at least two ionizable groups, given by a modified Henderson−Hasselbalch equation:

$$Y_{obs} = \frac{Y_N + Y_A \times 10^{pH_{m1}-pH}}{1 + 10^{pH_{m1}-pH}} + \frac{Y_L + Y_A \times 10^{pH-pH_{m2}}}{1 + 10^{pH-pH_{m2}}}$$

$$(1)$$

where Y_{obs} denotes the observed fluorescence signal for a particular pH value, Y_N, Y_A, and Y_L are the signals of the N form, the A form (acidic form), and the L form (low pH form), respectively, and pH_{m1} and pH_{m2} are the midpoints of the N to A and A to L transitions, respectively.

Dynamic Light Scattering (DLS). A DynaPro 99 unit (Wyatt) was used for the DLS measurements as reported previously.[34] In brief, all of the buffers used for the experiment were filtered three times with a 0.2 μm filter to remove dust particles. The centrifuge tubes and the tips were rinsed with 0.2 μm filtered Milli-Q water just before use. The scattering intensity at 90° and the autocorrelation function were acquired at the same time by illuminating the samples with a 829.4 nm laser. The experimental settings were as reported previously.[34] DynaLS software (Wyatt) was used to obtain the mean autocorrelation function by taking an average of 50 acquisitions and resolving them into Gaussian distributions of hydrodynamic radii (R_H). The mean of all of the acquisitions from the cumulant analysis was used to calculate the total light scattering intensity.

Steady-State Fluorescence Anisotropy Measurements. Protein samples (10 μM) were equilibrated in buffers of different pHs from 7 to 2, and steady-state fluorescence anisotropy was measured as reported previously.[34] The tryptophan residues were excited at 280 nm, and the emission was collected for 300 s at 340 nm, using an antiphotobleaching setting. The excitation and emission slit widths were set to 2 and 10 nm, respectively. The integration time was set to 1 s. The steady-state fluorescence anisotropy (r) is related to emission intensity I as follows:

$$r = \frac{I_{VV} - I_{VH}G}{I_{VV} + 2I_{VH}G}$$

$$(2)$$

where subscript V (vertical) refers to the position of polarizers in the excitation beam and the subscript H (horizontal) to the emission beam. G refers to the instrumental correction factor and is equal to the ratio of I_{HV} to I_{HH}. The measured values of anisotropy remained identical after incubation for 5 and 24 h.

Circular Dichroism (CD). A Jasco J-815 spectropolarimeter was used for the CD measurements, and experiments were performed using the same parameters as reported previously.[34] For far-ultraviolet (far-UV) CD and near-UV CD measurements, the spectra were collected in the wavelength ranges of 200−250 and 250−300 nm using quartz cuvettes with path lengths of 0.1 and 1 cm, respectively. The near-UV CD measurements were carried out at a protein concentration of 60 μM (1.1 mg/mL), while 25 μM protein (0.48 mg/mL) was used for far-UV CD measurements.

8-Anilino-1-naphthalenesulfonic Acid (ANS) Fluorescence Assay. For the ANS assay, the N form (2 μM) and the L form (2 μM) were incubated for 15 min in the dark with different ANS concentrations varying from 0 to 100 μM. ANS binding was assessed by excitation of the dye at 380 nm and acquisition of the emission spectrum from 400 to 600 nm. For FRET measurements, the tryptophan residues were excited at 295 nm and the emission was monitored from 320 to 580 nm. The excitation slit was set to 1 nm, and the emission slit was set to 5 nm. The binding affinity of the ANS dye for the L form is determined by fitting the data in Figure 4D and its inset to eq 3:

$$\Delta F(ANS) = n\Delta F_{max} \frac{[ANS]}{[ANS] + K_D}$$

$$(3)$$

where $\Delta F(ANS)$ is the change in the fluorescence signal in the presence of varying concentrations of ANS and ΔF_{max} is the maximum fluorescence when all of the ANS is bound to the L form. In the equation, K_D is the dissociation constant and n is the number of binding sites.

The kinetics of formation of the L form was monitored using ANS binding on a PerkinElmer fluorescence spectrometer (LS 55). After the protein had been transferred at pH 2.5, at different times of the formation of the L form, the binding of the ANS to the protein was measured. The excitation wavelength was set to 295 nm, and the emission was measured from 310 to 570 nm. The excitation slit width was 7 nm, and the emission slit width was 6.5 nm.

Salt-Induced Formation of the β Form. For the salt-dependent assay, the protein (10 μM) was incubated with varying salt concentrations ranging from 2.5 to 200 mM, for 24 h. Following the incubation, far-UV CD spectra of all of the samples were collected. All of the samples were also analyzed by the thioflavin T (ThT) fluorescence assay as described below.

For the measurement of the kinetics of formation of the β form, 10 μM protein was transferred in pH 2.5 buffer containing 150 mM KCl at 25 °C. The kinetics of the conformational change was monitored by measuring the change in ellipticity at 216 nm by CD. The bandwidth was set to 2 nm, and the data integration time was set to 4 s for all of the kinetic experiments. The elongation kinetics was measured by monitoring the change in ThT fluorescence as described below.

ThT Fluorescence Assay. To perform the ThT assay, the protein sample (1 μM) was added to ThT buffer [50 mM Tris (pH 8) and 20 μM ThT] and the ThT fluorescence was acquired within 1 min. The ThT dye was excited at 440 nm, and the emission was collected from 455 to 535 nm. The excitation and emission slit widths were set to 1 and 10 nm, respectively. The emission spectra of the buffer having the dye were recorded under similar settings. For the L $\rightleftharpoons \beta$ transition kinetics, the emission was collected at 482 nm for 20 s. The background fluorescence contributed by the free dye was subtracted from all of the data.

Size Exclusion Chromatography. Size exclusion chromatography (SEC) was performed for the N form and the β form using a HiLoad 16/600 Superdex 200 pg column on an AKTA Pure M FPLC system (GE Healthcare). The fractionation range of this column is 10−600 kDa, and it has a void volume of 39.8 mL. The column was first equilibrated with buffer at the respective pH before 25 μM protein was loaded onto the column. The flow rate was set to 0.8 mL/min.
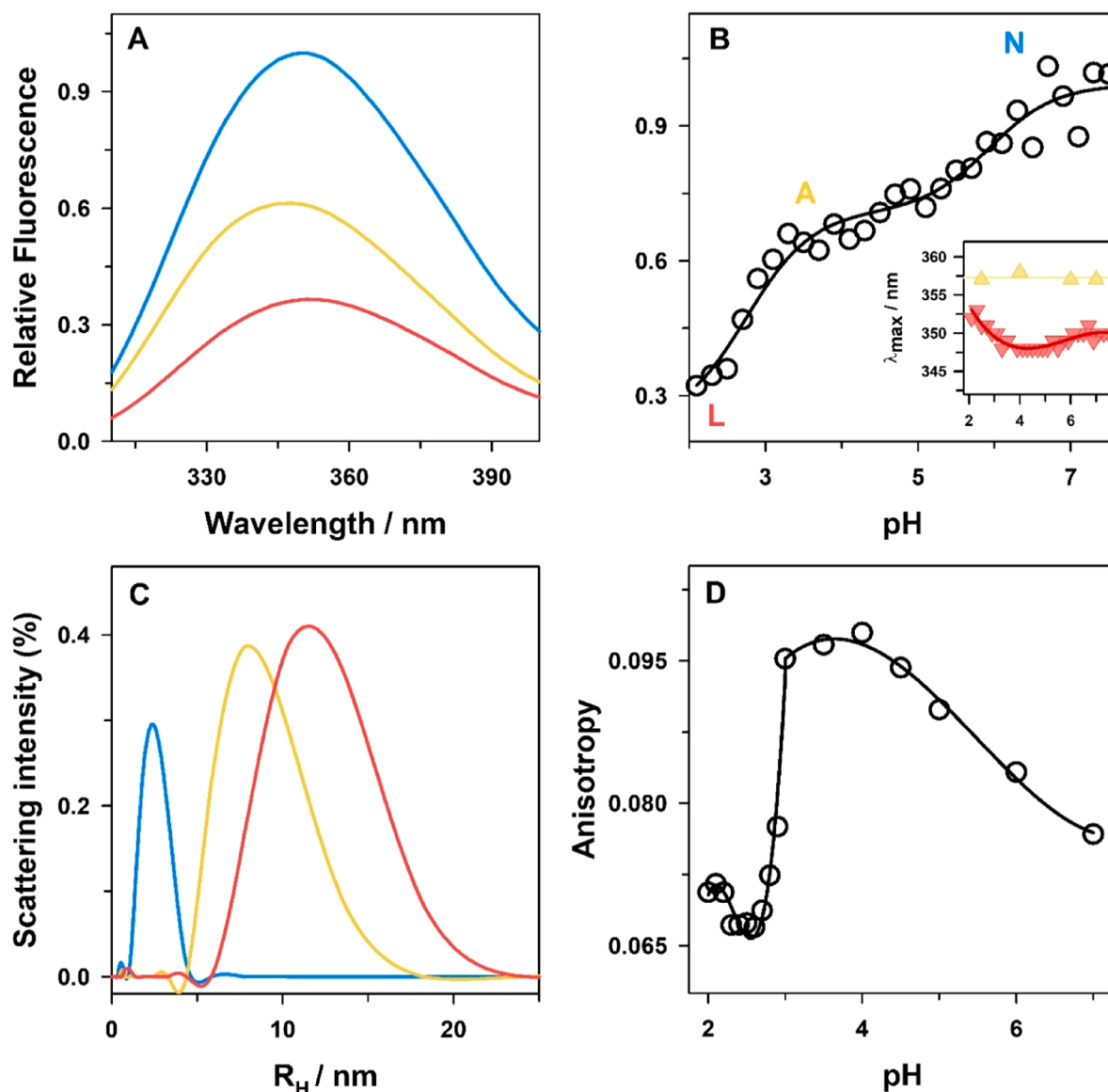
Figure 2. pH-induced monomer to oligomer transition of TDP-43^{tRRM}. (A) Relative fluorescence spectrum plotted at pH 7 (blue), pH 4.1 (yellow), and pH 2.5 (red). (B) Fluorescence intensity of TDP-43^{tRRM} (O) at different pH values plotted at 340 nm. The black line through the data is a nonlinear least-squares fit to eq 1. The relative fluorescence values of the N, A, and L forms are indicated. The inset of panel B displays the λ_{max}^{em} of the tryptophan residues at different pH values in the presence (yellow triangles) and absence (red inverted triangles) of 6 M GdmCl. The red and yellow lines across the data are drawn to guide the eye. (C) R_H distributions of the N form (blue), the A form (yellow), and the L form (red) as monitored by DLS. (D) Fluorescence anisotropy (tryptophan residues) at 340 nm plotted at different pH values (O). The black line across the data is shown to guide the eye.

The buffers contained 150 mM KCl to prevent the protein from sticking to the column. The apparent molecular weights of the N form and the β form were determined from the known molecular weights of standard biomolecules as described previously.[34]

Transmission Electron Microscopy (TEM). The L form and the β form (10 μL of a 25 μM solution) were placed on a 300-mesh Formvar carbon-coated copper grid (Electron Microscopy Science) for 5 min. After the excess solution had been removed, a 2% uranyl acetate solution was used to negatively stain the grid for 1.5 min. The grid was then washed with filtered Milli-Q water for a min and allowed to air-dry. The samples were examined via transmission electron microscopy (Technai-T20) at an accelerating voltage of 200 kV.

GdmCl-Induced Equilibrium Unfolding Transitions. The protein samples (4 μM) were first equilibrated at room temperature for 3 h with different concentrations of GdmCl. The fluorescence spectra of tryptophan residues were measured as described previously,[34] with the excitation slit set to 1.2 nm and the emission slit set to 10 nm. The fluorescence spectrum of the buffer was collected under identical settings. Buffer-subtracted data are plotted at 340 nm as a function of GdmCl concentration.

■ RESULTS

pH-Dependent Biphasic Structural Change Monitored by Fluorescence. Figure 2A displays the variation in the fluorescence spectrum of the protein (Figure 1) as a function of pH. The fluorescence intensity of the protein
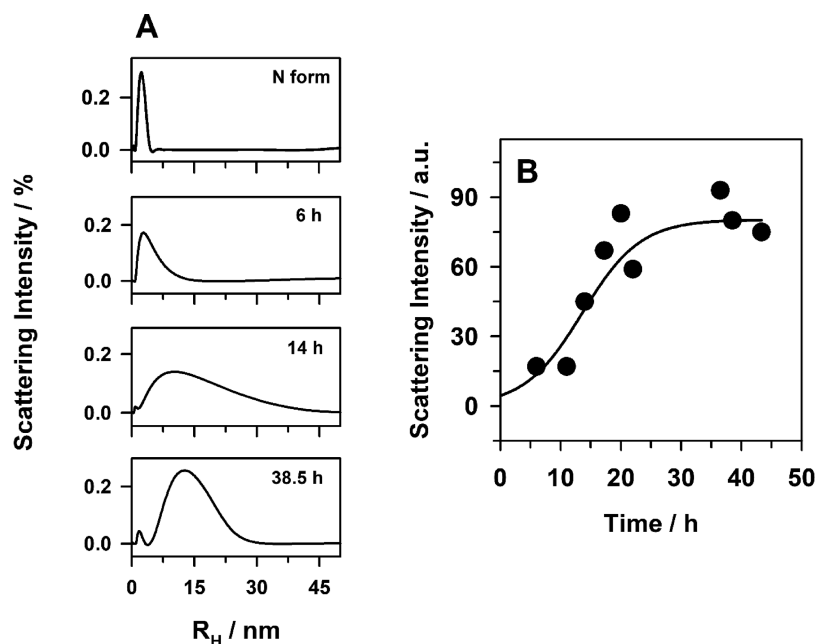
Figure 3. Kinetics of formation of the L form as measured by DLS. (A) DLS distributions of the N form and at three different times after the protein had been transferred at pH 2.5 and room temperature. (B) Change in the intensity of the scattered light (●) plotted with respect to time during the formation of the L form. The black solid line across the data is shown to guide the eye. In both panels, the concentrations of the protein and KCl were 25 μM and 3 mM, respectively.

decreases in two discrete sigmoidal steps with a decrease in pH from 7.5 to 2 (Figure 2B). This result implies that the local tertiary structure around the tryptophan side chains is disrupted in two steps and the protein undergoes two structural transitions with a decrease in pH, linked to the protonation of at least two buried ionizable groups (Materials and Methods). The first transition has a midpoint around pH 5.6 and plateaus between pH 4.5 and 3. The product of the first transition is termed the "acidic form" or the "A form" of the protein (Figure 2B). The midpoint of the second transition is assessed to be at pH 2.75, and it plateaus around pH 2.5. The product of the second transition is termed the "low-pH form" or the "L form" of the protein (Figure 2B). The structural properties of the A form have been described previously.[34] In this study, we dissect the structural character-istics of the L form and report its amyloidogenic nature and metastable behavior.

U-like Solvation of the Tryptophan in the L Form. In addition to the change in the fluorescence intensity, the wavelength of the maximum fluorescence emission (λ_{max}^{em}) of the tryptophan residues also changes in two steps during the N to L transition (Figure 2B, inset). The change in the fluorescence Stokes shift reports on the extent of the polar environment around the fluorophore. We observed that the λ_{max}^{em} in the N form is 348 nm that shifts slightly to the blue to 346 nm in the A form,[34] indicating that the solvation environment of the tryptophan residues in the A form is either similar or slightly hydrophobic in comparison with that of the N form. With a further decrease in pH, the λ_{max}^{em} changes to 351 nm in the L form, reflecting a highly solvated polar environment around the tryptophan residues in the L form. In contrast, the GdmCl-induced U form has a λ_{max}^{em} of 356 nm across a range of different pH values. These results indicate that the L form is not completely unfolded but the local tryptophan environment remains U-like.

The N ⇌ L Transition Is a Monomer ⇌ Oligomer Transition. The disrupted packing and the U-like solvation environment around the tryptophan residues suggested a partially folded structure for the L form. However, the hydrodynamic measurements using DLS revealed that the L form is an oligomer in addition to being partially folded. The N form is monomeric in nature[34] and has an R_H of ∼2.3 nm (Figure 2C). The R_H of the L form was observed to be ∼12 nm. Intriguingly, the N to L transition is accompanied by an increase in the width of the distribution that suggests an increased heterogeneity of the population during the transition. The L form with a mean R_H of 12 nm is a larger species compared to the A form, whose R_H is 8.5 nm.[34] These results indicate that the transition from the N form to the L form occurs along with an increased level of self-association of partially folded molecules, resulting in the formation of an intermediate with a larger hydrodynamic radius.

Change in the Local Dynamics of Tryptophan Side Chains during the N ⇌ L Transition. Interestingly, the local dynamics of the tryptophan residues changes in a two-step manner during the monomer ⇌ oligomer transition. We observed that the mean value of the steady-state fluorescence anisotropy increased from 0.0767 to 0.0952 as we moved from the N form to the A form, indicating an increase in the rigidity of the tryptophan residues in the A form.[34] However, the A ⇌ L transition shows a very steep decrease in the anisotropy value, which changes to 0.0674 in the L form. A low value of anisotropy reflects an unfolded-like flexible environment around the tryptophan residues in the L form,[34] which is in line with the U-like solvation for the tryptophan residues as seen by the λ_{max}^{em} (Figure 2B, inset). Increased dynamics might also afford multiple relative orientations to the tryptophan residues, resulting in a lower anisotropy value. These results indicate that TDP-43[tRRM] has a completely different structural organization in the L form, the A form, and the N form.
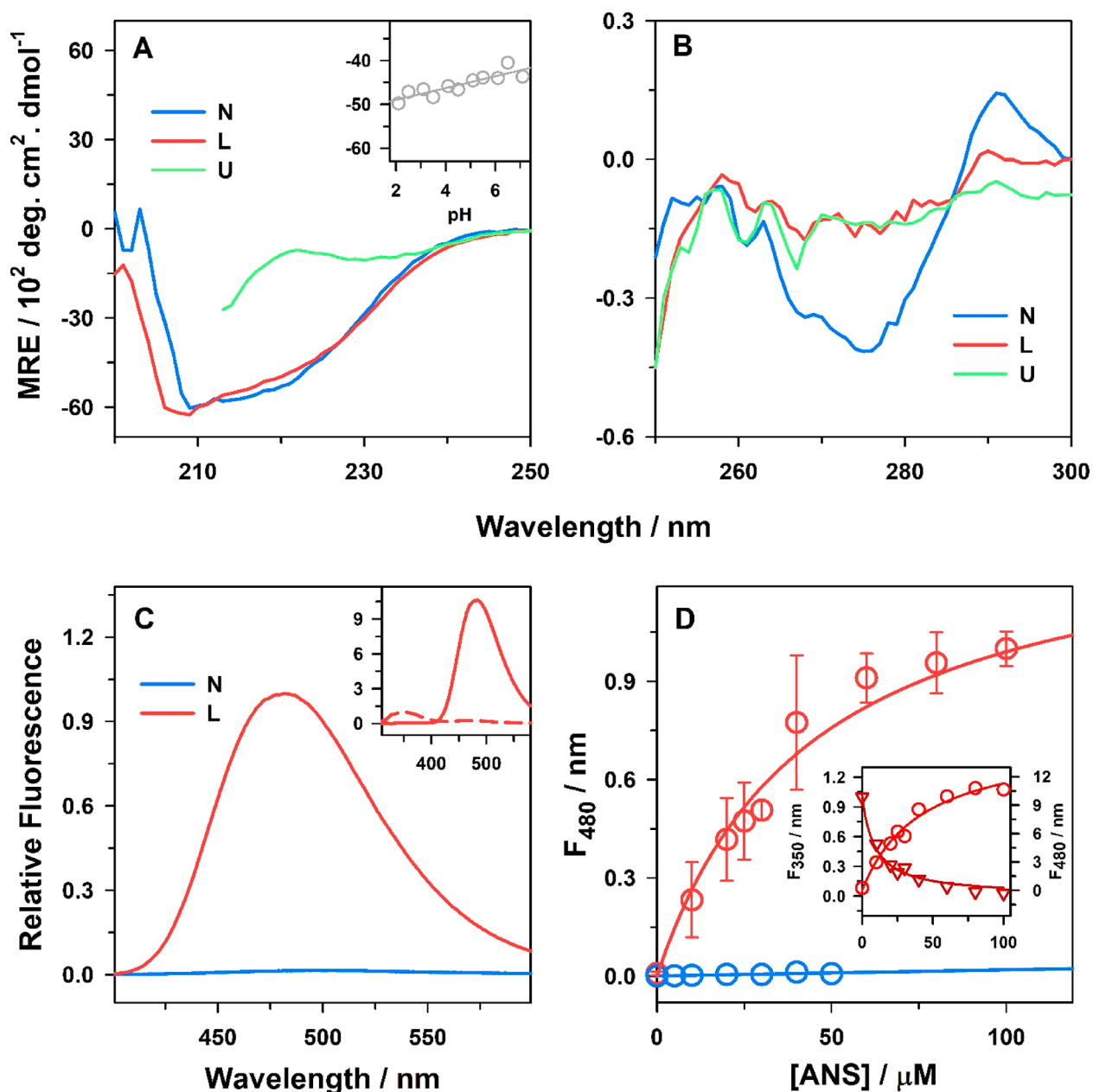
Figure 4. L form that consists of protein molecules that are molten globular in nature with exposed hydrophobic patches. (A) Far-UV CD and (B) near-UV CD spectra of the N form, the L form, and the U form. The MRE at 222 nm at different pH values is plotted in the inset of panel A. (C) Fluorescence spectrum of the ANS dye in the presence of the N form and the L form. The tryptophan residues in the L form exhibit FRET with protein-bound ANS. The inset in panel C displays the fluorescence spectrum of the L form in the absence and presence of ANS after excitation at 295 nm. The fluorescence of tryptophan is quenched in the presence of ANS, but there is a huge increase in the fluorescence of ANS indicating energy transfer. (D) Measurement of the binding of ANS to the N form (blue circles) and the L form (red circles). Upon excitation of the ANS dye at 380 nm, the relative change in the fluorescence intensity is plotted at 480 nm as a function of ANS concentration. FRET between tryptophan residues and the bound ANS dye in the L form is shown in the inset of panel D. The left y-axis shows the tryptophan fluorescence at 350 nm (inverted red triangles), while the right y-axis shows the ANS fluorescence at 480 nm (red circles) upon excitation at 295 nm in the presence of increasing concentrations of ANS. The solid red lines across the data in panel D and its inset are nonlinear least-squares fits to eq 3, and the solid blue line across the N form data is shown to guide the eye.

DLS-Monitored Kinetics of Formation of the Oligomeric L Form. The L form is oligomeric in nature, and its R_H is larger than that of the monomeric N form (Figure 2C). Figure 3A shows the temporal evolution of the distribution of R_H of 25 μM protein at room temperature as obtained from DLS experiments during the formation of the L form. The mean and the width of the distribution of R_H increase progressively with time during the N $\rightleftharpoons$ L transition. This result indicates that both the size and heterogeneity of the protein molecules increase during the formation of the L form. Figure 3B shows the increase in total light scattering intensity with time during the formation of the L form. The increase in scattering intensity occurs in two steps. The increase is slow at the beginning with a prominent rise around 5 h. The scattering intensity achieves a plateau after 15 h. The solution of the L form remains considerably clear during the entire transition,
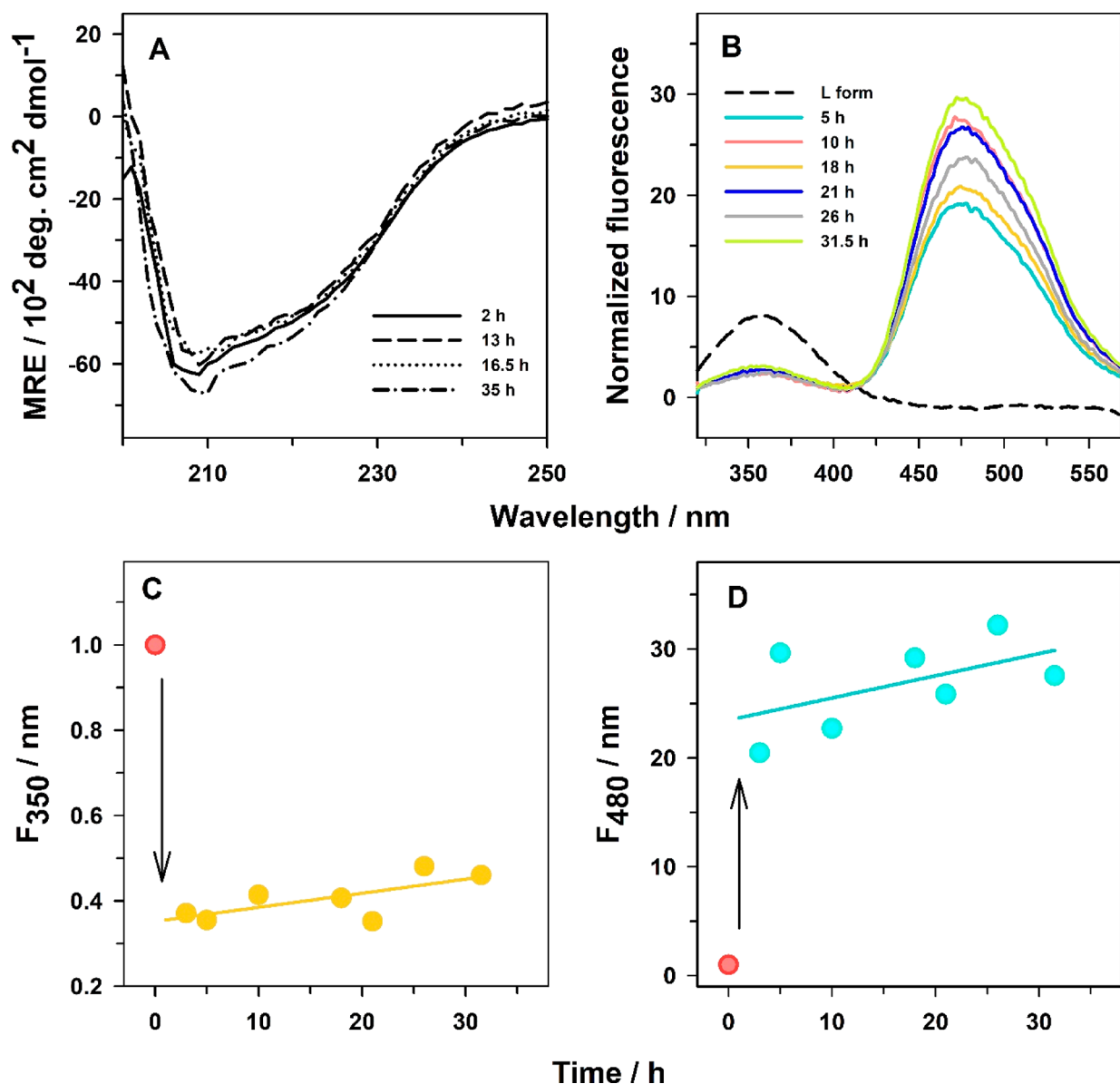
Figure 5. Kinetics of the formation of the L form as measured by far-UV CD and ANS binding. The formation of the L form is initiated by transferring the protein at pH 2.5. (A) Far-UV CD spectrum of the protein at different times during the formation of the L form. (B) Binding of ANS to the protein monitored at different times during the formation of the L form. Panel B shows the fluorescence spectrum upon excitation at 295 nm. As a control, the fluorescence spectrum of the L form in the absence of ANS is also plotted (black dashed line). (C) Relative change in tryptophan fluorescence intensity (yellow circles) at 350 nm upon excitation at 295 nm in the presence of ANS plotted at different times during the formation of the L form. (D) Relative change in ANS fluorescence intensity (blue circles) at 480 nm upon excitation at 295 nm plotted at different times during the formation of the L form. In panels C and D, the red circle indicates the fluorescence emission of tryptophan residues at 350 nm and ANS emission at 480 nm upon excitation at 295 nm in the absence of ANS, respectively, at pH 2.5. The yellow and blue lines in panels C and D are shown to guide the eye.

and we do not see species higher than ∼30 nm at any time during the formation of the L form (Figure S1).

Molten Globular Nature of Protein Molecules in the L Form. We next used far-UV and near-UV CD spectroscopy to analyze the global secondary and tertiary structure of the protein, respectively. Figure 4A compares the far-UV CD spectra of the L form, the N form, and the U form of the protein. The far-UV CD spectrum of the L form is very similar to that of the N form, reflecting an intact secondary structure of the protein molecules in the L form. The mean residue ellipticity (MRE) at 222 nm shows only a small change across the pH range, suggesting very little change in the secondary structure of the protein during the N ⇌ L conversion (Figure

4A, inset). The MRE of the L form at 222 nm is −4961 ± 124, which is comparable to the MRE of the N form (−4664 ± 156). Figure 4B compares the near-UV CD spectrum of the L form, the N form, and the U form. The N form shows weak but distinct absorbance in the range of 270−285 nm due to the asymmetric packing of its three tyrosine (Y123, Y155, and Y214) and two tryptophan residues in the protein structure. In the U form, the protein shows near-zero absorbance in the near-UV CD spectrum due to the loss of asymmetric packing of aromatic residues. The near-UV CD spectrum of the L form is very similar to that of the U form. These results suggest that the global tertiary structure of the protein molecules in the L form is dramatically disrupted (Figure 4B), but the secondary
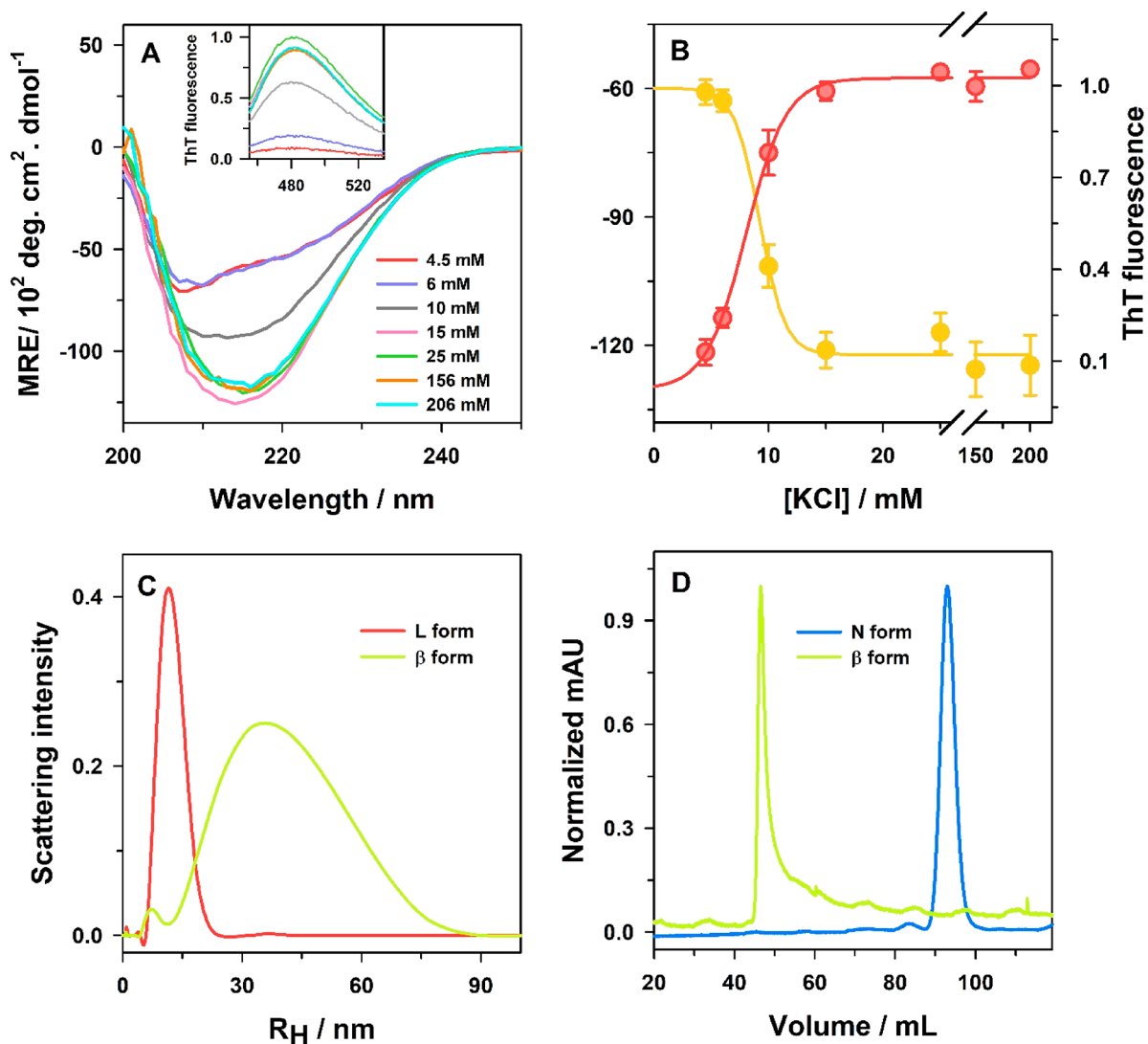
Figure 6. Extremely metastable L form that undergoes conformational conversion coupled with oligomerization in the presence of salt to form the amyloid-like misfolded β form. (A) Far-UV CD spectra of the protein at pH 2.5 in the presence of an increasing concentration of KCl. The inset of panel A shows ThT fluorescence spectra of corresponding samples upon excitation at 440 nm. (B) The yellow circles show the change in MRE as a function of salt concentration (left-hand y-axis) at 216 nm, while the red circles show the change in the fluorescence of ThT at 482 nm (right-hand y-axis). The solid lines across the data in panel B are shown to guide the eye. (C) DLS distributions of the L form and the β form. (D) Size of the N form and the β form as monitored by SEC. The data for the N form were taken from a previous study.[34]

structure is intact (Figure 4A). Thus, the protein molecules that make up the L form are molten globular in structure.

The L Form Has Exposed Hydrophobic Patches. The hydrophobic dye ANS shows a huge increase in fluorescence upon binding to the L form, indicating that the L form contains a loosely packed solvent-exposed hydrophobic patch. In contrast, the ANS dye does not show any fluorescence in the presence of the N form, indicating that the N form fails to bind to ANS (Figure 4C). We observed that there is a transfer of energy from tryptophan residues of the protein to ANS upon photoexcitation of tryptophan. The inset of Figure 4C shows that when the tryptophan is excited at 295 nm in the presence of ANS, there is a decrease in tryptophan fluorescence emission with a simultaneous increase in ANS fluorescence. The result that the tryptophan and ANS act as a FRET pair indicates that the tryptophan residues are located close to the hydrophobic patch.

Figure 4D shows the changes in the fluorescence of ANS when the L form and the N form are incubated with increasing

concentrations of ANS. The inset of Figure 4D shows the decrease in the fluorescence emission of tryptophan residues and the increase in the emission of the ANS dye upon excitation at 295 nm in the L form with increasing concentrations of ANS. We estimated the binding affinity of the ANS dye for the L form by fitting the Figure 4D data and its inset to a ligand binding model (Materials and Methods). The K_D value for the binding of ANS to the L form is 40 μM. The binding of ANS dye to the L form is weak relative to the binding of ANS to the A form (K_D = 11 μM).[34] The weak binding capacity of the hydrophobic dye hints at the possibility of solvation of the hydrophobic patches in the L form.

A Conformational Opening Reaction Precedes Oligomerization. The results of the DLS data show that the oligomerization of the protein during the formation of the L form occurs in two steps (Figure 3). To understand the nature of the conformational changes during the formation of the L form, we monitored the L form kinetics using far-UV CD and ANS binding (Figure 5). We observed that the secondary

structure remains native-like during the formation of the L form (Figure 5A). We also examined the kinetics of formation of the L form using ANS binding and observed that the change during the transition occurs in two steps (Figure 5B−D). The major decrease in the tryptophan emission and the consequent increase in the ANS fluorescence occur very fast during the first step of the formation of the L form (Figure 5B−D). In addition, there is a very small change in the fluorescence during the second step of the formation of the L form on a slower time scale (Figure 5B−D). These experiments suggest that when the N form is subjected to low pH, it first undergoes a very fast transition into a partially unfolded conformation in which the hydrophobic patches are exposed but the secondary structure remains N-like. Combined with the DLS data (Figure 3), these results indicate that the partially unfolded conformation oligomerizes in the second slower step. Hence, a partial unfolding reaction and exposure of hydrophobic patches precede the process of oligomerization during the formation of the L form. The small change in the ANS fluorescence in the second slower step might be due to the creation of a small number of hydrophobic patches during the process of oligomerization.

The Structural Organization of the L Form Is Highly Sensitive to the Ionic Strength of the Solution. Interestingly, we observed that the structure of the L form is very sensitive to the concentration of salt in the solution. The L form is an oligomer (Figure 2C) that consists of protein molecules that have a broken tertiary structure but an intact N-like secondary structure (Figure 4). Panels A and B of Figure 6 show that the L form retains its N-like secondary structure only when the concentration of KCl in the solution is below 5−6 mM. The L form also does not bind to the amyloid-staining dye ThT (Figure 6A, inset), indicating it does not contain an amyloid-like structure. With an increase in the concentration of salt, there is an increase in the value of MRE in the far-UV CD spectrum accompanied by a switch in the signal maximum to 216 nm (Figure 6A). This result indicates that at high salt concentrations the L form is transformed into a state that is highly rich in β-sheet structure (β form). The structural transformation of the L form to the β form is complete by 15 mM KCl, above which the β form is predominantly populated (Figure 6A,B). The β form also binds strongly to the ThT dye (Figure 6A, inset, and Figure 6B), revealing the presence of cross β-sheet structure similar to that of amyloids. The structural transition of the L form to the β form is quite steep, and the midpoint of the transition, monitored by both far-UV CD and ThT fluorescence, occurs at ∼10 mM KCl (Figure 6B). The structural transformation of the L form in the presence of a very low concentration of salt indicates the weak nature of the force that governs the structure of the L form. Interestingly, the conformational conversion of the L form to the β form is accompanied by an increased size, as monitored by DLS (Figure 6C). The mean values of R_H for the L form and the β form are 12 and 36 nm, respectively. The width of the DLS distribution is larger for the β form than for the L form, indicating that the β form is more heterogeneous than the L form. These results suggest that the oligomerization status and the structural organization of the L form are highly sensitive to the ionic strength of the solution.

The β Form Is a Larger Species. We estimated the molecular weight of the β form using SEC. A size exclusion chromatogram of the L form could not be obtained as the protein sticks to the column at low salt concentrations (<5

mM). The N form and the β form were passed through a HiLoad 200 pg Superdex 16/600 column. The N form eluted at 92.8 mL, while the elution of the β form was found to occur at 46.7 mL (Figure 6D). A small trail is seen for the β form, which implies the heterogeneous nature of the soluble oligomers. A decrease in the elution volume for the β form reflects a larger hydrodynamic radius and a higher apparent molecular weight compared to those of the N form. The apparent molecular weight of the N form was calculated to be 18.8 kDa[34] (consistent with a monomer), which increased considerably for the β form and was found to be ∼530 kDa. The elution volume of the β form (46.7 mL), however, is near the void volume (39.8 mL) of the column. Hence, the determined molecular weight is only a lower limit, and the molecular weight of the β form is expected to be >530 kDa, signifying that the β form is at least a 25-mer.

The β Form Has a Morphology Similar to That of Worm-like Amyloid Fibrils. We examined the external morphology of the L form and the β form using TEM (Figure 7). In the TEM micrograph, the L form appears to be ∼10 nm
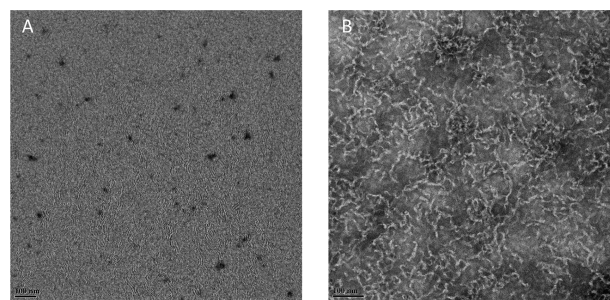

Figure 7. External morphology of the L form and the β form. Transmission electron microscopy (TEM) images of (A) the L form and (B) the β form (150 mM KCl). The scale bars in the bottom left corners are 100 nm.

oligomers (Figure 7A). In contrast, upon prolonged incubation, the β form showed a twisted worm-like morphology that was a few hundred nanometers in length (Figure 7B). The worm-like appearance of the β form is similar to the worm-like amyloid fibrils seen for multiple proteins, including amyloid-β,[41] β_2-microglobulin,[42] and mouse prion protein.[43]

Kinetics of the L $\rightleftharpoons$ β Transition. We monitored the kinetics of the L $\rightleftharpoons$ β transition by measuring the change in the ThT fluorescence and far-UV CD signal at 216 nm (Figure 8A). ThT is an amyloid-staining dye that has been used to measure the formation and elongation of cross β-sheet rich structure.[44] The L form does not bind to ThT, but the β form effectively binds to ThT (Figure 6A, inset). In the presence of 150 mM KCl where the β form is fully populated (Figure 6B), we observed that the kinetics of elongation of cross β-sheet structure during the formation of the β form as measured by ThT fluorescence is characterized by a single-exponential change with a rate constant of 0.12 min^{-1} (Figure 8A). No lag phase is observed in the ThT-monitored kinetics.

Interestingly, the kinetics of the L $\rightleftharpoons$ β transition as measured by far-UV CD occurs in two exponential phases (Figure 8A). The rate constant of the fast phase was 0.20 min^{-1}, while the slow phase had a rate constant of 0.032 min^{-1} (in the presence of 150 mM KCl). Fifty percent of the total change in the far-UV CD signal occurred in the fast phase, and
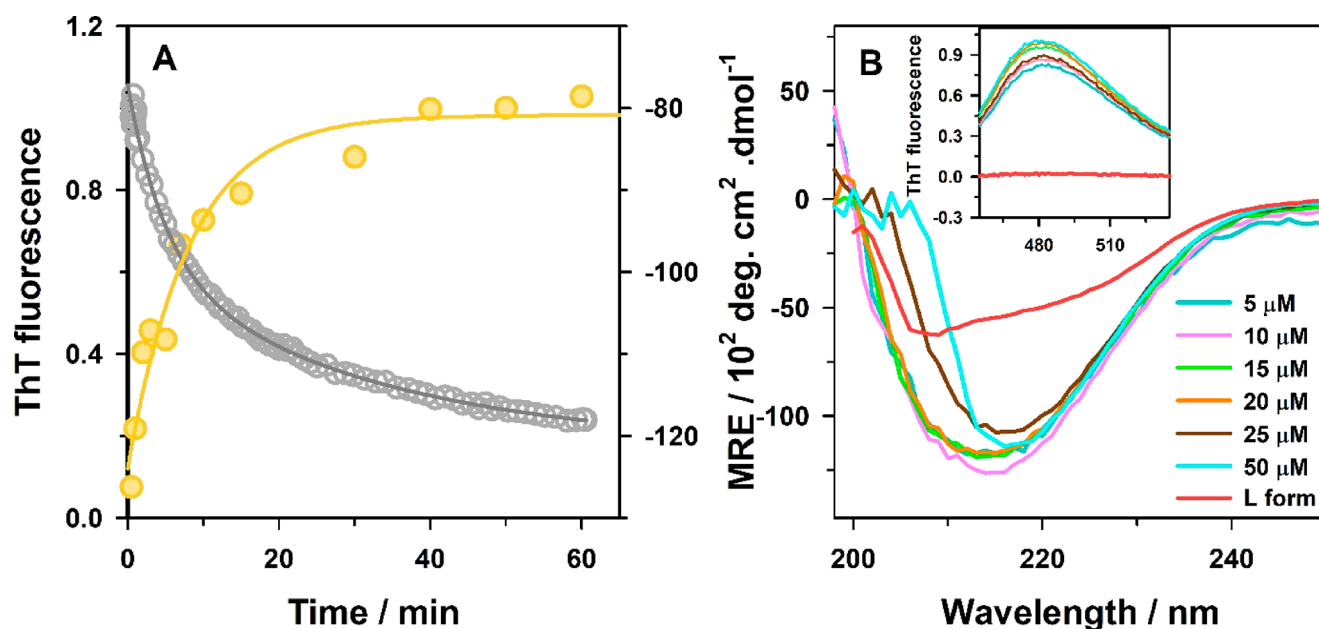
Figure 8. Kinetics of the L ⇌ β transition. (A) Change in kinetics monitored through ThT fluorescence (yellow circles, left-hand y-axis) and far-UV CD (gray circles, right-hand y-axis) when 10 μM protein was transferred to pH 2.5 buffer containing 150 mM KCl. The yellow line through the ThT data is a fit to a single-exponential equation, while the black line through the far-UV CD data is a fit to a double-exponential equation. (B) Far-UV CD spectra of different concentrations of the protein upon incubation at pH 2.5 with 150 mM KCl. The red line indicates the far-UV CD spectrum of the L form. The inset shows ThT fluorescence spectra of the corresponding protein samples.
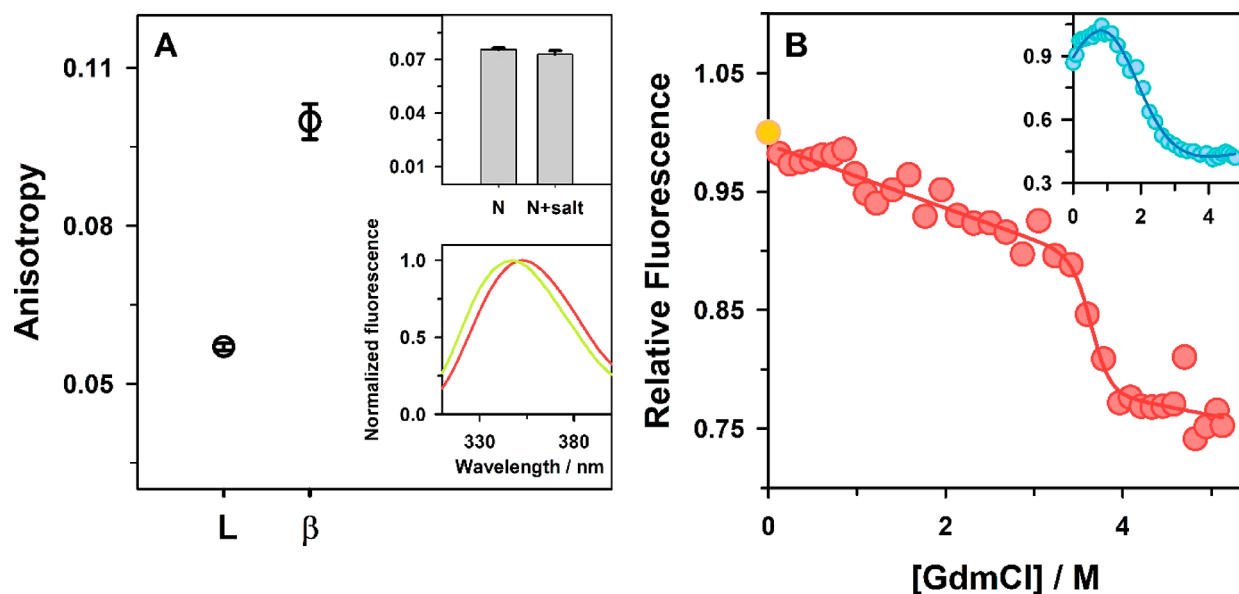

Figure 9. β form that is structurally ordered and stable. (A) Steady-state fluorescence anisotropy (tryptophan residues) in the L form and the β form (150 mM KCl). The top inset shows the steady-state fluorescence anisotropy (tryptophan residues) in the N form in the absence and presence of 150 mM KCl. The bottom inset compares the normalized fluorescence spectrum of the L form (red line) and the β form (yellow line). (B) Change in the fluorescence of the protein at pH 2.5 (main panel, red circles) and pH 7.5 (inset, blue circles) with varying GdmCl concentrations. The L form signal is indicated by a filled yellow circle. The solid lines in the main panel and its inset are shown to guide the eye.

the remaining 50% occurred in the slow phase of the kinetics. These results imply that the conformational conversion of the N-like secondary structure of the protein molecules in the L form to the β-sheet rich β form occurs in multiple steps. No lag phase was observed in far-UV CD-monitored kinetics, similar to the ThT-monitored kinetics (Figure 8A). The absence of a lag phase in the process of conformational conversion and elongation during the L ⇌ β transition implies that the L form acts as a nucleus in the formation of the β form.

The kinetics of the L ⇌ β transition in Figure 8A has been monitored at a monomeric protein concentration of 10 μM. However, we observed that the formation of the β form is independent of the protein concentration in the concentration range of 5−50 μM, in the presence of 150 mM KCl (Figure 8B). The β forms formed at different protein concentrations have similar far-UV CD spectra (Figure 8B), and they bind to the ThT dye to similar extents (Figure 8B, inset). These results

indicate that the β forms formed at different protein concentrations have a similar structure.

The β Form Is Structurally Rigid and Highly Stable in Comparison to the L Form and the N Form. The increase in the MRE value in the β form in comparison to the L form (Figure 6A) may be due to the conversion of the disordered structure of the protein (Figure 1) into a more ordered structure in the β form. This random coil to β-sheet transition, however, does not occur in the N form whose CD structure remains largely unchanged with the increase in the ionic strength of the solution.[34] We observed that the L $\rightleftharpoons$ β transition is accompanied by an increased rigidity of protein side chains (Figure 9A). The tryptophan residues in the L form are located in a region that has a U-like solvation environment (Figure 2B and its inset) and is highly dynamic as indicated by the low steady-state fluorescence anisotropy (Figure 2D). The mean value of the steady-state fluorescence anisotropy of tryptophan residues in the L form is 0.057, which increases to 0.099 in the β form (Figure 9A). The higher anisotropy in the β form indicates a relatively ordered structure around the tryptophan residues in the β form. In contrast, the fluorescence anisotropy of tryptophan residues in the N form does not change in the absence or presence of 150 mM KCl (Figure 9A, top inset).

The L $\rightleftharpoons$ β conversion is also marked by a blue shift in the λ_{max}^{em} of tryptophan residues. The mean value of λ_{max}^{em} in the L form is found to be 351 $\pm$ 1 nm, which shows a considerable blue shift to 345 $\pm$ 1 nm in the β form (Figure 9A, bottom inset). These results indicate that tryptophan residues are in a more hydrophobic environment in the β form than in the L form due to the exclusion of water molecules and the formation of ordered protein structure.

We examined the structural stability of the L form, the β form, and the N form using a denaturant (GdmCl)-induced unfolding transition (Figure 9B). We observed that the L form is extremely metastable and converts to the β form upon addition of as little as 120 mM GdmCl (Figure S2), as inferred from changes in the far-UV CD spectrum (Figure S2A), an increase in the fluorescence anisotropy of tryptophan residues (Figure S2B), a blue shift of ~6 nm in their λ_{max}^{em} values (Figure S2C), and the elution volume determined via size exclusion chromatography (Figure S2D). The β form, however, showed an apparently sigmoidal unfolding transition with the midpoint at 3.6 M GdmCl (Figure 9B). In comparison, the monomeric N form also showed an apparently sigmoidal transition but with the midpoint at 1.9 M GdmCl (Figure 9B, inset). The observation that the β form has a relatively high resistance to chemical denaturation can be attributed to the fact that the protein molecules in the β form are held by strong interactions.

■ DISCUSSION

The aggregation of TDP-43 in the cytosol of neurons is the hallmark of ALS and multiple other TDP-43 proteinopathies.[6] A majority of the cases of ALS (90–95%) are sporadic in nature,[5] and chronic environmental stress in the form of osmotic stress, oxidative insults, and changes in pH or temperature has been suggested to cause the disease.[9] TDP-43 has been reported to be recruited in the stress granules in response to heat stress, oxidative stress, and chemical stressors and to co-deposit with the other stress granule proteins (e.g., PABP, TIA, and eIF3) in the diseased brain.[16,38,39] Stress granules are dynamic biomolecular condensates consisting of proteins and nucleic acids that largely retain the native

structure of proteins and protect the cell from environmental stress, but under prolonged stress, they transform into abnormal aggregates.[9] Starvation stress decreases the pH of the cytosol from 7.4 to 5.5, resulting in an ~100-fold change in proton ion concentration.[13,15] It has been proposed that some proteins in the cell could function as biosensors and sense stress by site-specific protein modifications.[13] For example, such a drastic change in the concentration of proton ions is efficiently sequestered by biosensor proteins by protonation of their amino acid residues.[13,15] The process of protonation triggers the formation of metastable assemblies, and the coupling of protonation–deprotonation equilibria to assembly and disassembly of proteins has been proposed to be a fast and energy-efficient method for mitigating a decrease in the pH of the cytosol due to starvation stress.[13] We, however, very poorly understand how the changes in proton ion concentration affect the assembly–disassembly equilibria and in turn modulate the nature of the initial assemblies formed in the case of TDP-43 protein during its aggregation. In our quest to understand the steps involved in the formation of different initial assemblies of TDP-43 that could form in response to a change in proton ion concentration, we employ a range of pH values in this study and deduce the conformational change that accompanies metastable assembly formation.

In this study, we observed that under low-pH conditions, which mimic starvation stress, TDP-43^{tRRM} undergoes a conformational opening reaction (Figure 5) to form a molten globule-like structure that allows the protonation of the side chains of buried ionizable groups (Figure 2B). There are three aspartate residues (D105, D138, and D247) and one histidine residue (H166) that are buried in the hydrophobic core of TDP-43^{tRRM}, and it is possible that the structural change protonates any of these critical residues. These events trigger the assembly of protein molecules to form an oligomeric L form (Figure 3). In the L form, the protein molecules possess secondary structure similar to that of the native protein but the tertiary structure is disrupted (Figure 4A,B) and the hydrophobic core is wet (Figure 4C,D). The L form does not show any binding to ThT, indicating that amyloidogenic structure is absent in the L form. At different times of its formation, the protein molecules maintain the native-like secondary structure (Figure 5A). There is no larger amorphous aggregate[45] at any time during the oligomeric transition (Figure S1).

The L form is highly metastable, and upon small changes in the ionic strength of the solution, it further aggregates to a β-sheet rich form (the β form) (Figure 6). As little as 15 mM KCl is enough to fully convert the L form into the β form, indicating that the L form is held together by very weak interactions (Figure 6B). In contrast, the β form is highly resistant to chemical unfolding (Figure 9B), indicating that the β form is structurally held by very strong interactions. The TEM experiments revealed that the external morphology of the β form resembles worm-like amyloid fibrils (Figure 7). The β form also binds to the ThT dye. In the literature, there has been a debate about whether the aggregates of TDP-43 and its fragments resemble amyloids.[6,46–49] Our results show that TDP-43^{tRRM} forms worm-like amyloid fibrils under starvation stress-like conditions.

The L to β transition measured by far-UV CD (Figure 6A) is accompanied by an increase in the value of MRE, in addition to the shift in maxima to 216 nm. This result suggests that the unstructured and disordered regions of the L form assemble into ordered β-sheet during this conversion. The flexible side

chains of W113 and W172 located in the RRM1 domain become rigid during the L to β transition, as indicated by steady-state fluorescence anisotropy experiments (Figure 9A), indicating that the RRM1 domain gains structure in the β form. Previous studies have also suggested that the RRM1 domain is less stable than the RRM2 domain, which might render it more prone to aggregation.[31,50] The β form forms in two-exponential phases (Figure 8A), suggesting that the ordering of structure occurs in at least two distinct steps. In totality, our results indicate that the L to β conversion is a disorder-to-order transition, akin to the "liquid-to-solid" transition of stress granules to amyloids.

We observed that under mild stress conditions, the oligomeric A form is formed (Figure 2), which has been characterized in detail previously.[34] The A form has a structural organization that is different from that of the L form. Contrary to the L form, in the A form tryptophan residues have restricted mobility[34] (Figure 2D) and are relatively buried in the hydrophobic protein environment (Figure 2B, inset). It is also smaller than the L form (Figure 2C). These results indicate that different metastable assemblies can form under the different stress-like conditions as has been proposed recently for Pub1 protein.[15] This result is important because TDP-43 forms stable assemblies of discrete densities and morphologies under different disease subtypes.[28,51] It is possible that the alternate forms of disease occur due to the alternate pathological conformations of initial oligomeric assemblies of TDP-43.[51]

One very important suggestion from our results is that the TDP-43[tRRM] region can function as a stress sensor as reported for Sup35,[20] Pab1,[18] and Pub1,[15] in addition to performing its nucleic acid binding functions. The sensing of pH stress begins via an initial conformational change and protonation that allows the protein to oligomerize into the L form. The L form is structurally held by weak interactions and is metastable. With small changes in the ionic strength of the solution, it transforms into the amyloid-like β form and disordered regions of TDP-43[tRRM] gain structure during this conversion. In the oligomeric L form, the local concentration of the protein will be high, which can bring the disordered region of TDP-43[tRRM] into the proximity, making it highly prone to undergoing higher-order assembly with a small fluctuation in the solution conditions. This behavior of the L form is similar to the behavior of large stress granule-like assemblies that are held by weak interactions, are metastable, and form abnormal, stable, solid-like aggregates under persistent stress.[9] Stress granules have recently been reported to form in multiple steps with the initial formation of the small soluble oligomeric core,[21] and it is possible that the L form resembles the initial stages of stress granule-like assemblies. In this way, the results of this study shine new light on the initial stages of the stress-induced aggregation of TDP-43 and the pathogenesis of ALS.

■ ASSOCIATED CONTENT

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biochem.9b00780.

DLS data of the L form transition until 1000 nm (Figure S1) and conversion of the L form to the β form upon addition of a very small amount of GdmCl (Figure S2) (PDF)

Accession Codes

TAR DNA binding protein 43 (TDP-43), UniProtKB/Swiss-Prot entry Q13148 (TADBP_HUMAN).

■ AUTHOR INFORMATION

Corresponding Author

Santosh Kumar Jha − *CSIR-National Chemical Laboratory, Pune, India, and Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India;* ⓞ orcid.org/0000-0003-1339-7409; Phone: 91-20-25902588; Email: sk.jha@ncl.res.in; Fax: 91-20-25902615

Other Author

Meenakshi Pillai − *CSIR-National Chemical Laboratory, Pune, India, and Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.biochem.9b00780

■ REFERENCES

(1) Rowland, L. P., and Shneider, N. A. (2001) Amyotrophic lateral sclerosis. *N. Engl. J. Med. 344*, 1688−1700.

(2) 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.

(3) 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. Y. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science 314*, 130−133.

(4) 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.

(5) Ling, S. C., Polymenidou, M., and Cleveland, D. W. (2013) Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron 79*, 416−438.

(6) Sun, Y. L., and Chakrabartty, A. (2017) Phase to phase with TDP-43. *Biochemistry 56*, 809−823.

(7) Gao, J., Wang, L., Huntley, M. L., Perry, G., and Wang, X. (2018) Pathomechanisms of TDP-43 in neurodegeneration. *J. Neurochem. 146*, 7−20.

(8) 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.

(9) 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.

(10) McGurk, L., Gomes, E., Guo, L., Mojsilovic-Petrovic, J., Tran, V., Kalb, R. G., Shorter, J., and Bonini, N. M. (2018) Poly(ADP-Ribose) prevents pathological phase separation of TDP-43 by promoting liquid demixing and stress granule localization. *Mol. Cell 71*, 703−717.

(11) Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T., and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell 163*, 123−133.

(12) Lin, Y., Protter, D. S., Rosen, M. K., and Parker, R. (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell 60*, 208−219.

(13) Rabouille, C., and Alberti, S. (2017) Cell adaptation upon stress: The emerging role of membrane-less compartments. *Curr. Opin. Cell Biol. 47*, 34−42.

(14) Dewey, C. M., Cenik, B., Sephton, C. F., Dries, D. R., Mayer, P., Good, S. K., Johnson, B. A., Herz, J., and Yu, G. (2011) TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Mol. Cell. Biol. 31*, 1098−1108.

(15) Kroschwald, S., Munder, M. C., Maharana, S., Franzmann, T. M., Richter, D., Ruer, M., Hyman, A. A., and Alberti, S. (2018) Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery. *Cell Rep. 23*, 3327−3339.

(16) McDonald, K. K., Aulas, A., Destroismaisons, L., Pickles, S., Beleac, E., Camu, W., Rouleau, G. A., and Vande Velde, C. (2011) TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum. Mol. Genet. 20*, 1400−1410.

(17) Kroschwald, S., and Alberti, S. (2017) Gel or Die: Phase separation as a survival strategy. *Cell 168*, 947−948.

(18) Riback, J. A., Katanski, C. D., Kear-Scott, J. L., Pilipenko, E. V., Rojek, A. E., Sosnick, T. R., and Drummond, D. A. (2017) Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell 168*, 1028−1040.

(19) Cohen, T. J., Hwang, A. W., Unger, T., Trojanowski, J. Q., and Lee, V. M. (2012) Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. *EMBO J. 31*, 1241−1252.

(20) Franzmann, T. M., Jahnel, M., Pozniakovsky, A., Mahamid, J., Holehouse, A. S., Nuske, E., Richter, D., Baumeister, W., Grill, S. W., Pappu, R. V., Hyman, A. A., and Alberti, S. (2018) Phase separation of a yeast prion protein promotes cellular fitness. *Science 359*, eaao5654.

(21) Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R., and Parker, R. (2016) Distinct stages in stress granule assembly and disassembly. *eLife 5*, e18413.

(22) Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell 162*, 1066−1077.

(23) Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S. K., Rees, E., Miyashita, A., Costa, A. R., Dodd, R. B., Chan, F. T. S., Michel, C. H., Kronenberg-Versteeg, D., Li, Y., Yang, S. P., Wakutani, Y., Meadows, W., Ferry, R. R., Dong, L., Tartaglia, G. G., Favrin, G., Lin, W. L., Dickson, D. W., Zhen, M., Ron, D., Schmitt-Ulms, G., Fraser, P. E., Shneider, N. A., Holt, C., Vendruscolo, M., Kaminski, C. F., and St George-Hyslop, P. (2015) ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron 88*, 678−690.

(24) Shiina, Y., Arima, K., Tabunoki, H., and Satoh, J. (2010) TDP-43 dimerizes in human cells in culture. *Cell. Mol. Neurobiol. 30*, 641−652.

(25) Qin, H. N., Lim, L. Z., Wei, Y. Y., and Song, J. X. (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.

(26) Tsoi, P. S., Choi, K. J., Leonard, P. G., Sizovs, A., Moosa, M. M., MacKenzie, K. R., Ferreon, J. C., and Ferreon, A. C. M. (2017) The N-terminal domain of ALS-linked TDP-43 assembles without misfolding. *Angew. Chem., Int. Ed. 56*, 12590−12593.

(27) 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.

(28) Buratti, E. (2015) Functional significance of TDP-43 mutations in disease. *Adv. Genet. 91*, 1−53.

(29) Ayala, Y. M., Pantano, S., D'Ambrogio, A., Buratti, E., Brindisi, A., Marchetti, C., Romano, M., and Baralle, F. E. (2005) Human, Drosophila, and C.elegans TDP43: Nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol. 348*, 575−588.

(30) Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ule, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., and Allain, F. H. T. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol. 20*, 1443−1449.

(31) Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J., and Yuan, H. S. (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res. 37*, 1799−1808.

(32) 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.

(33) Zacco, E., Grana-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.

(34) 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.

(35) Mann, J. R., Gleixner, A. M., Mauna, J. C., Gomes, E., DeChellis-Marks, M. R., Needham, P. G., Copley, K. E., Hurtle, B., Portz, B., Pyles, N. J., Guo, L., Calder, C. B., Wills, Z. P., Pandey, U. B., Kofler, J. K., Brodsky, J. L., Thathiah, A., Shorter, J., and Donnelly, C. J. (2019) RNA binding antagonizes neurotoxic phase transitions of TDP-43. *Neuron 102*, 321−338.

(36) Huang, Y. C., Lin, K. F., He, R. Y., Tu, P. H., Koubek, J., Hsu, Y. C., and Huang, J. J. (2013) Inhibition of TDP-43 aggregation by nucleic acid binding. *PLoS One 8*, e64002.

(37) Volkening, K., Leystra-Lantz, C., Yang, W. C., Jaffee, H., and Strong, M. J. (2009) Tar DNA binding protein of 43 kDa (TDP-43), 14−3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). *Brain Res. 1305*, 168−182.

(38) Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y. J., Vanderwyde, T., Citro, A., Mehta, T., Zaarur, N., McKee, A., Bowser, R., Sherman, M., Petrucelli, L., and Wolozin, B. (2010) Tar DNA binding protein-43 (TDP-43) associates with stress granules: Analysis of cultured cells and pathological brain tissue. *PLoS One 5*, e13250.

(39) Colombrita, C., Zennaro, E., Fallini, C., Weber, M., Sommacal, A., Buratti, E., Silani, V., and Ratti, A. (2009) TDP-43 is recruited to stress granules in conditions of oxidative insult. *J. Neurochem. 111*, 1051−1061.

(40) Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266−280.

(41) Walsh, D. M., Hartley, D. M., Kusumoto, Y., Fezoui, Y., Condron, M. M., Lomakin, A., Benedek, G. B., Selkoe, D. J., and Teplow, D. B. (1999) Aβ-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. *J. Biol. Chem. 274*, 25945−25952.

(42) Gosal, W. S., Morten, I. J., Hewitt, E. W., Smith, D. A., Thomson, N. H., and Radford, S. E. (2005) Competing pathways

determine fibril morphology in the self-assembly of β2-microglobulin into amyloid. *J. Mol. Biol. 351*, 850−864.

(43) Jain, S., and Udgaonkar, J. B. (2008) Evidence for stepwise formation of amyloid fibrils by the mouse prion protein. *J. Mol. Biol. 382*, 1228−1241.

(44) Naiki, H., and Gejyo, F. (1999) Kinetic analysis of amyloid fibril formation. *Methods Enzymol. 309*, 305−318.

(45) Adachi, M., Noji, M., So, M., Sasahara, K., Kardos, J., Naiki, H., and Goto, Y. (2018) Aggregation-phase diagrams of β2-microglobulin reveal temperature and salt effects on competitive formation of amyloids versus amorphous aggregates. *J. Biol. Chem. 293*, 14775−14785.

(46) Chen, A. K., Lin, R. Y., Hsieh, E. Z., Tu, P. H., Chen, R. P., Liao, T. Y., Chen, W., Wang, C. H., and Huang, J. J. (2010) Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *J. Am. Chem. Soc. 132*, 1186−1187.

(47) Neumann, M., Kwong, L. K., Sampathu, D. M., Trojanowski, J. Q., and Lee, V. M. (2007) TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: Protein misfolding diseases without amyloidosis. *Arch. Neurol. 64*, 1388−1394.

(48) Mompean, M., Hervas, R., Xu, Y., Tran, T. H., Guarnaccia, C., Buratti, E., Baralle, F., Tong, L., Carrion-Vazquez, M., McDermott, A. E., and Laurents, D. V. (2015) Structural evidence of amyloid fibril formation in the putative aggregation domain of TDP-43. *J. Phys. Chem. Lett. 6*, 2608−2615.

(49) Saini, A., and Chauhan, V. S. (2014) Self-assembling properties of peptides derived from TDP-43 C-terminal fragment. *Langmuir 30*, 3845−3856.

(50) Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R., and Zitzewitz, J. A. (2014) Folding of the RNA recognition motif (RRM) domains of the amyotrophic lateral sclerosis (ALS)-linked protein TDP-43 reveals an intermediate state. *J. Biol. Chem. 289*, 8264−8276.

(51) Laferriere, F., Maniecka, Z., Perez-Berlanga, M., Hruska-Plochan, M., Gilhespy, L., Hock, E. M., Wagner, U., Afroz, T., Boersema, P. J., Barmettler, G., Foti, S. C., Asi, Y. T., Isaacs, A. M., Al-Amoudi, A., Lewis, A., Stahlberg, H., Ravits, J., De Giorgi, F., Ichas, F., Bezard, E., Picotti, P., Lashley, T., and Polymenidou, M. (2019) TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci. 22*, 65−77.