# Structural characterization of amyloid fibrils by solid-state NMR

Jayakrishna Shenoy Krishnashenoy Padmabai

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

# DOCTEUR DE

# L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

SPÉCIALITÉ : CHIMIE PHYSIQUE

Par Jayakrishna Shenoy K.P.

# STRUCTURAL CHARACTERIZATION OF AMYLOID FIBRILS BY SOLID-STATE NMR

Sous la direction de : Antoine LOQUET

Soutenue le 11 Décembre 2020

Membres du jury :

| | | |
|---|---|---|
| Erick DUFOURC | Directeur de Recherche, CNRS, Bordeaux | Président |
| Salvador VENTURA | Professeur, Autonomous University of Barcelona | Rapporteur |
| Vinesh VIJAYAN | Professeur Associé, IISER-TVM, Thiruvanathapuram | Rapporteur |
| Alexander K. BÜLL | Professeur, Technical University of Denmark, Kgs. Lyngby | Examinateur |
| Philippe DERREUMAUX | Professeur, Université Paris Diderot, Paris | Examinateur |
| Antoine LOQUET | Directeur de Recherche, CNRS, Bordeaux | Directeur |

**Acknowledgment**

Completing this would not have been possible without many people's immense help, including my research supervisor, colleagues, teachers, friends, and Family. It gives me great pleasure to express my gratitude to all those who have supported me in completing this thesis.

This thesis is carried out as a collaboration between the Loquet group of the European Institute of Chemistry and Biology (IECB) and the Institute of Chemistry & Biology of Membranes & Nano-objects (CBMN), both located in Bordeaux. I want to thank the directors of both institutes, **Remi Fronzes** and **Sophie Lecompte,** for hosting me in their laboratories.

Foremost, I owe my deepest gratitude to my supervisor **Antoine Loquet** for his continuous support for my Ph.D. study. I would like to thank him for his patience, valuable guidance, motivation, and immense knowledge. It's a great privilege to work with him. Special thanks to **Erick Dufourc** for being cool with the endless University administration procedures and validations and support.

I would like to convey my sincere gratitude to **Salvator Ventura** and **Vinesh Vijayan** for their interest in my work and for being the referees of my thesis. I would also like to thank **Erick Dufourc**, **Alexander K. Büll,** and **Philippe Derreumaux** for being part of the jury committee.

I would like to thank **Sophie Lecompte, Christopher Cullin**, **Alexander KAI BüLL, Benjamin Bardiaux, and Sven J. Saupe** for their immense support in collaboration studies and for offering precious inputs.

I would like to express my immense gratitude to **Melanie, Antoine, Mathilde, and Alix** for their patience and constant support in the bio-lab and sample preparations. I want to thank **Axelle Grelard**, **Estelle, Morvan, Brice Kauffmann, and Marion Decossas** for providing continuous support in using NMR, XRD, and EM Platforms. I also would like to thank **Myriam Mederic Dargere, Thierry Dakhil, Jean-Michel Blanc** for their technical support in using the biochemistry platform at the IECB.

My special thanks to **Alons Lends, Asen Dasklov, Ahmad Saad, Sujeesh, Arpita Tawani, and Denis Martinez.** I consider it a privilege to work with these incredible researchers. I thank my fellow lab mates in the Loquet lab for sharing these years with me and for the stimulating discussions, and for all the fun we have had in the last few years.

Special thanks to **Alons and Asen** for their valuable suggestions, mentoring throughout the thesis. I also thank them for their friendship and cheerful footballing analysis. Thanks to **Mathias Ferber** for the molecular simulation studies and the critical reading of the thesis. Special gratitude to **Nadia, Anthony, and Gaelle** for their immense personal and academic support throughout the thesis and their company.

Many thanks to all the staff at IECB, UB, and CBMN for helping me with the administrative procedures.

I am indebted to **CNRS** (Centre national de la recherche Scientifique) and **ERC** (European research council) for funding my thesis.

Thanks to my fellow lab mates: **Tanguy**, **Coralie**, **Hicham, Issa, Valentine, Martin, Axelle, Mathilde, Bamba, and Rio** for sharing these years with me and their support all the fun we have had in the last few years. Special thanks to my colleagues in IECB: **Antoine Amestoy, Antoine Scalable, Bala, Shaheen**, **Marti, Leonardo, Salgado, and Eduardo** for all their help and their company.

Thanks to all the friends in Bordeaux: Ani, Reenu, Maneesha, Sujeesh, Karthi, Pramod, Arjun, and Sreedevi, making life in Bordeaux pleasant and memorable. Thanks to all my IISERTVM-friends, FC Aishwarya, Team Shaji, and Counterstrike: GO team for all the fun. Thanks to Deschamps, Zizou, Klopp, and their team, the GOAT(S), Jordan Peterson, for the inspiration. Thanks to JRE podcasts, Rick and Morty, Dave Chappelle, Bill Burr, South park, Peaky blinders, for entertainment. Thanks to my beloved Family.

My sincere gratitude to mother nature for all the blessings!

**-Jaya**

“We are going to die, and that makes us the lucky ones. Most people are never going to die because they are never going to be born. The potential people who could have been here in my place but who will in fact never see the light of day outnumber the sand grains of Arabia. Certainly those unborn ghosts include greater poets than Keats, scientists greater than Newton. We know this because the set of possible people allowed by our DNA so massively exceeds the set of actual people. In the teeth of these stupefying odds it is you and I, in our ordinariness, that are here. We privileged few, who won the lottery of birth against all odds, how dare we whine at our inevitable return to that prior state from which the vast majority have never stirred?”

**― Richard Dawkins, Unweaving the Rainbow: Science, Delusion and the Appetite for Wonder**

# TABLE OF CONTENTS

# List of figures and tables

# List of tables

# List of suplimentary figures

# Abstract

Amyloids are an important class of protein supramolecular self-assemblies in the biological world. These fibrous protein aggregates are rich in β-sheet structure and display a characteristic cross-beta pattern observed by X-ray diffraction (XRD). Sometime, the self-templating ability helps them to propagate as prions (infectious proteins). Some amyloids perform vital cellular functions like chemical storage, structure, signaling, etc. and are called functional amyloids. At the same time, some are associated with lethal neurodegenerative diseases such as Alzheimer's, Parkinson's disease, etc. and are called pathological amyloids.

In my thesis presented here, I use a combination of electron microscopy, X-ray fiber diffraction, Fourier-transform infrared spectroscopy, and solid-state nuclear magnetic resonance spectroscopy (NMR) techniques for the structural studies of (1) pathological amyloids formed by TDP-43 concerned with histopathology of amyotrophic lateral sclerosis, (2) the molecular mechanism of signal transduction via functional amyloid proteins NWD2/HET-s involved in programmed cell death in fungi, and (3) the molecular basis of binding modes of Thioflavin T (ThT) on amyloid fibrils.

**Project 1- Structural dissection of amyloid fibrils of different TDP-43 constructs by solid-state NMR**

The TAR DNA binding protein with a molecular size of 43 kDa (TDP-43) is observed as the main component in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar dementia (FTLD) cytoplasmic inclusions. TDP-43 consists of a well-folded N-terminal domain (NTD), two RNA recognition motif domains (RRM1 and RRM2), and an intrinsically disordered C-terminal domain (CTD). The prion-like C-terminal domain possesses most of the pathologically relevant mutations and plays a critical role in the spontaneous aggregation of TDP-43 and associated proteinopathy. We managed to obtain a qualitatively and quantitatively superior amount of TDP-43 amyloid fibrils for biophysical analysis. We analyzed different fibril-forming TDP constructs using electron microscopy, X-ray diffraction, and solid-state NMR, namely TDP-43(1-414), TDP-35(90-414), TDP-16(267-414), TDP-43(1-366), TDP-13(300-414), TDP-11(300-399), TDP-10(314-414) and two TDP splice forms TDP43-1 and sTDP-43-2 to gain insights into the molecular organization of TDP-43 in the context of amyloid aggregation.

Although the different TDP constructs share similar amyloid properties of fibrillar morphology and a typical cross-β signature by X-ray diffraction, solid-state NMR indicates that TDP-43 and TDP-35 share the same polymorphic molecular structure, while other C-terminal fragments encompass a well-ordered amyloid core. We identified several residues in the so-called C-terminal GaroS2 (368–414) domain that participates in the rigid core of TDP-16 fibrils, underlining its importance during the aggregation process. Our findings demonstrate that C-terminal fragments can adopt a different molecular conformation in isolation or in the context of the full-length assembly, suggesting that the N-terminal domain and RRM domains play an important role in the TDP-43 amyloid transition.

Later, we identified several β–breaker segments located at the hydrophobic, Q/N rich, and GaroS2 subdomains in the amyloid fold of a CTF, named TDP-13 (300-414). Several aromatic residues, including W334 and F367 paired with glycines, are identified as crucial in the amyloid core. We observed that the different CTFs (TDP-16, TDP-13, TDP-11, TDP-10) adopt different polymorphs, maintaining a common cross-β fibrillar architecture. Analogous to other reported pathologic amyloid fibrils, the TDP CTFs can display structurally distinct and self-propagating amyloid fibrils depending on the aggregation conditions and sequence length. Interestingly, TDP-13 itself readily engages alternative packing arrangements in response to different aggregation conditions. Our results complement the recent cryo-EM study pointing out that two peptide segments from CTD termed SegA (residues 311-360) and SegB (residues 286-331) form multiple structurally different fibrillar polymorphs highlighting the role of the hydrophobic region. Even though the CTD (TDP-16) includes the amyloid core and forms amyloid filaments, the NTD dimerization alongside putative interactions from RRM domains may play a decisive role in governing CTDs aggregation interface, therefore in constructing the structural architecture of the final amyloid assembly.

Further, we observed that the TDP-43 splice isoforms (sTDP43-1, sTDP43-2) could aggregate in the absence of CTD and display amyloid-like characteristics like ThT fluorescence and XRD cross-b diffraction pattern. Our SSNMR data reports that the TDP43 isoforms adopt an amyloid fold similar to full-length TDP-43, suggesting a misfolding mechanism of NTD and RRM domains responsible for the amyloid transition in TDP43 isoforms.

Our work paves the way for understanding the role of the N-terminal and RRM domains in the TDP-43 amyloid transition.

**Project 2- Molecular mechanism of heterotypic amyloid signaling proteins NWD2/HET-s involved in Programmed cell death in fungi**

The [Het-s] prion activates HET-S pore-forming protein and triggers programmed cell death (PCD) in the fungus *Podospora anserina.* The HET-s β-solenoid fold acts as a template for the amyloid transition of the HET-S prion-forming domain (PFD). A Nod-like receptor (NLR) protein called NWD2, encoded by the gene next to *het-S,* possesses an N-terminal motif (R0) homologous to the elementary repeat unit of HET-s β–solenoid fold (R1, R2). The NLRs are associated with host defense and programmed cell death in plants, animals, and fungi. It is proposed that upon activation, the N-terminal motif (R0) can adopt the β-solenoid fold, template the amyloid transition of the HET-S prion-forming domain, and trigger the downstream cell death pathway.

In this work, we deduced the 3D model of the β–solenoid fold adopted by the functional amyloid NWD2 (R0) based solid-state NMR restraints. Similar to the HET-s PFD model, the R0-motifs in the NWD2 NLR signalosome comprises a central triangular hydrophobic core designed by three β–sheets (β1, β2 and β3) with a fourth β–sheet pointing away from the central amyloid core. Furthermore, we successfully reported the 3D structural model for the heterotypic-amyloid interface (R0/R2), i.e., amyloid interface made of two different protein segments, here R2 motif in HET-s and its templating precursor amyloid (NWD2), involved in the fungal signal transduction to programmed cell death. Although the triangular hydrophobic core is virtually conserved in R0-R2 hetero-amyloids, our SSNMR data suggests that the 4$^{th}$ and eighth β–strands are dynamically flexible and display a high degree of structural heterogeneity. This crucial structural difference in the R0/R2 interface could exfoliate prionized HET-S from the NWD2 signalosome. This mode of amyloid nucleation will allow repeated use of a signalosome, enabling signal amplification. Here, we report a molecular mechanism of a potential advantage of achieving signal amplification in two-gene architecture in contrary to the more straightforward all-in-one structure in fugal NLR signaling via the heteroamyloid interface potentially designed with lower stability. The R0/R2 interface may have to perform a dual role, i.e., act as a templating surface as well as a fragmentation interface.

It thus appears that in addition to the ability to form the β-solenoid fold efficiently, sequence variation of elementary R-motifs is finely tuned at the sequence level, with (i) an R1

tailored to a lower the amyloid-forming ability potentially to avoid self-activation of HET-S and (ii) R0 reduced in charge content to allow for homotypic polymerization and (iii) capable of forming different heterotypic interfaces with either R1 or R0 with the R0/R2 interface potentially conductive to slightly lower stability to allow for the recycling of the NWD2 signalosome.

**Project 3- Binding modes of Thioflavin T (ThT) on β–solenoid amyloids**

Fluorescent dyes like ThioflavinT (ThT) and congo red are widely used for accurate and conclusive diagnosis of various amyloid-related diseases. ThT exhibits enhanced fluorescence emission intensity upon amyloid binding (absorption at ~440 nm and emission at ~480 nm for ThT). Various uncharged Thioflavin-derivatives with good brain entry and clearance are used for imaging and quantifying amyloid plaques associated with various neurodegenerative diseases. However, many amyloids with cross–β architecture, including the model fungal prion HET-s (218-289), do not affect ThT fluorescence raising the prospect of false-negative results. Remarkably, HET-s-like β-solenoid architecture as the fundamental element is identified in human prion proteins like PrP and tau in its amyloid fold. Bioinformatic studies point out the commonality of β-solenoids and β-arches in amyloids as structural elements. So, we set out to determine the structural and molecular basis of tertiary-fold independent ThT fluorescence in amyloids adopting a β-solenoid prion fold.

We examined the deterministic factors of fluorescence property of ThT molecules upon binding different amyloid systems, namely HET-s(218-289), R0-R0, R0-R1, R0-R2, and R0-R1-R2. Interestingly, only R0-R0, R0-R1, and R0-R1-R2 fibrils were found to be ThT positive, while HET-s(218-289) and R0-R2 fibrils exhibited very low or no ThT fluorescence near physiological pH (pH 7.5). To investigate the molecular mechanism of ThT binding mode on β–solenoid amyloid fibrils, we employed solid-state NMR experiments.

We observed the deterministic interaction of ThT on β–solenoid adopting amyloids facilitating enhanced fluorescence and hence an accurate measurement of aggregation kinetic parameters. The solid-state NMR results were in excellent agreement with the ThT binding site on grooves designed by residues 9H-10A-11E in R0-R0 amyloids. THT's protonated amino group forms a salt bridge with E11 and 9H forming π–π-stacking interactions with the aromatic rings in ThT. Amyloid surfaces with aromatic residues, hydrophobicity, and acidic residues improve accessibility for ThT binding, where basic residues hinder ThT binding. The

potential salt bridge formation between the glutamate carboxyl groups within the fibrils of R0-R2/HET-s(218-289) may compete with the stable formation of a salt bridge with ThT. Also, the lack of aromatic residues in R0-R2 and HET-s(218-289) at the grooves make them ThT negative. This inference could also explain the differential magnitude in ThT intensity on other β–solenoid amyloids. R0-R1 shows a very high order of magnitude in ThT fluorescence intensity in similar conditions as they lack salt bridges between the amyloid β–turns, resulting in free salt bridges with glutamate (E11) and aspartate (D230). The ThT quantum yield differs with binding mode, binding free energy, ThT conformation on amyloid fibrils, and stoichiometry. The ThT preferably binds on solvent-exposed fibril grooves composed of acidic residues compatible for the salt-bridge formation or aromatic residues compatible with a π–π stacking.

As Thioflavin-derived compounds are used for qualitative and quantitative analysis of amyloid plaques, our results would pave the way for systematic research to discover novel tracer molecules to diagnose amyloid diseases accurately and conclusively. This study may potentially help for a rational design of specific protein aggregation inhibitors in a therapeutical perspective.

**Keywords:** SSNMR, TDP-43, amyloids, β–solenoid, HET-s, NWD2, Thioflavin-T

# Résumé

Les amyloïdes sont une classe importante d'auto-assemblages supramoléculaires de protéines dans le monde biologique. Ces agrégats protéiques fibreux sont riches en β-feuillets et présentent un signal caractéristique observé par diffraction des rayons X (XRD). Leur capacité d'auto-assemblage les aide parfois à se propager sous forme de prions (protéines infectieuses). Un groupe d'amyloïdes remplit des fonctions cellulaires vitales comme le stockage de molécules, la participation à des échafaudages structuraux, la signalisation, etc. et sont appelés amyloïdes fonctionnelles. En même temps, certains amyloïdes sont associées à des maladies neurodégénératives mortelles telles que la maladie d'Alzheimer, la maladie de Parkinson, etc. et sont appelés amyloïdes pathologiques.

Dans ma thèse présentée ici, j'utilise une combinaison de techniques regroupant la microscopie électronique, la diffraction des rayons X, l'analyse par spectroscopie infrarouge à transformée de Fourier, et des techniques de pointe en spectroscopie de résonance magnétique nucléaire du solide (SSNMR) afin d'étudier (1) l'amyloïde pathologique formé par la protéine TDP-43 qui est liée à l'histopathologie de la sclérose latérale amyotrophique, et (2) le mécanisme moléculaire de transduction du signal par les protéines amyloïdes fonctionnelles NWD2/HET-s impliquées dans la mort cellulaire programmée chez les champignons. Nous décrivons également (3) la base moléculaire du colorant de liaison amyloïde Thioflavin T(ThT) sur les fibrilles amyloïdes en nous basant sur d'expériences de NMR.

**Projet 1- Dissection structurelle de fibrilles amyloïdes de différentes constructions de TDP-43 par RMN du solide**

La protéine de liaison à l'ADN TAR d'une taille moléculaire de 43 kDa (TDP-43) est observée comme le principal composant des inclusions cytoplasmiques de la sclérose latérale amyotrophique (SLA) et de la démence lobaire frontotemporale (DLF). La TDP-43 se compose d'un domaine N-terminal bien replié (NTD), de deux domaines de motifs de reconnaissance d'ARN (RRM1 et RRM2) et d'un domaine C-terminal intrinsèquement désordonné (CTD). Le domaine C-terminal de type prion possède la plupart des mutations pathologiques pertinentes et joue un rôle essentiel dans l'agrégation spontanée de TDP-43 et de la protéinopathie associée. Nous avons réussi à obtenir une quantité qualitativement et quantitativement supérieure de fibrilles amyloïdes TDP-43 pour l'analyse biophysique. Nous

avons analysé différentes constructions de TDP formant des fibrilles en utilisant la microscopie électronique, la diffraction des rayons X et la spectroscopie RMN du solide, TDP-43(1-414), TDP-35(90-414), TDP-16(267-414), TDP-43(1-366), TDP-13(300-414), TDP-11(300-399), TDP-10(314-414) et deux formes d'épissage TDP TDP43-1 et sTDP-43-2 pour mieux comprendre l'organisation moléculaire de TDP-43 dans le contexte de l'agrégation amyloïde.

Bien que les différentes constructions TDP partagent des propriétés amyloïdes similaires en terme de morphologie fibrillaire et une signature typique par diffraction des rayons X, le RMN du solide indique que TDP-43 et TDP-35 partagent la même structure moléculaire polymorphe, tandis que d'autres fragments C-terminaux englobent un noyau amyloïde bien ordonné. Nous avons identifié plusieurs résidus dans le domaine dit C-terminal GaroS2 (368-414) qui participe au coeur rigide des fibrilles TDP-16, ce qui souligne son importance lors du processus d'agrégation. Nos résultats démontrent que les fragments C-terminaux peuvent adopter une conformation moléculaire différente en isolation ou dans le contexte de l'assemblage complet, ce qui suggère que le domaine N-terminal et les domaines RRM jouent un rôle important dans la transition amyloïde TDP-43.

Plus tard, nous avons identifié plusieurs segments de rupture situés dans les sous-domaines hydrophobe, riche en Q/N et GaroS2 dans le repliement amyloïde d'un CTF, à savoir TDP-13 (300-414). Plusieurs résidus aromatiques, dont W334 et F367 associés à des glycines, sont identifiés comme étant essentiels dans le coeur amyloïde. Nous avons observé que les différents CTF (TDP-16, TDP-13, TDP-11, TDP-10) adoptent des structures polymorphiques différentes, tout en ayant une architecture fibrillaire commune. Les CTF TDP peuvent présenter des fibrilles amyloïdes structurellement distinctes et auto-propagatives en fonction des conditions d'agrégation et de la longueur de la séquence. Il est intéressant de noter que le TDP-13 peut adopter différentes conformations en réponse à différentes conditions d'agrégation. Nos résultats complètent une récente étude en microscopie électronique sur TDP-43 et mettent en avant que deux segments peptidiques du CTD, appelés SegA (résidus 311-360) et SegB (résidus 286-331), forment de multiples polymorphes fibrillaires structurellement différents, ce qui met en évidence le rôle de la région hydrophobe. Bien que le CTD (TDP-16) comprenne le coeur amyloïde et forme des filaments amyloïdes, la dimérisation du NTD ainsi que les interactions supposées des domaines RRM peuvent jouer un rôle décisif dans la gouvernance de l'interface d'agrégation des CTD, donc dans la construction de l'architecture structurelle de l'assemblage amyloïde final.

En outre, nous avons observé que les isoformes d'épissage TDP-43 (sTDP43-1, sTDP43-2) pourraient s'agréger en l'absence de CTD et présenter des caractéristiques de type amyloïde comme la fluorescence ThT et un signal typique en diffraction des rayons X. Nos données de RMN du solide indiquent que les isoformes TDP43 adoptent un coeur amyloïde similaire à la protéine entière, suggérant un mécanisme de mauvais repliement des domaines NTD et RRM responsable de la transition amyloïde dans les isoformes TDP-43.

Notre travail ouvre la voie à la compréhension du rôle des domaines N-terminal et RRM dans la transition amyloïde TDP-43.

**Projet 2- Mécanisme moléculaire des protéines de signalisation amyloïdes hétérotypiques NWD2/HET-s impliquées dans la mort cellulaire programmée chez les champignons**

Le prion [Het-s] active la protéine HET-S formant des pores et déclenche la mort cellulaire programmée (PCD) dans le champignon Podospora anserina. Le coeur solénoïde de HET-s est un template pour la transition amyloïde du domaine de formation des prions (PFD) HET-S. Une protéine du type Nod-like receptor (NLR) appelée NWD2, codée par le gène voisin de het-S, possède un motif N-terminal (R0) homologue à l'unité répétitive élémentaire du coeur amyloïde de HET-s (R1, R2). Les NLR sont associés à la défense de l'hôte et à la mort cellulaire programmée chez les plantes, les animaux et les champignons. Il est proposé qu'après activation, le motif N-terminal (R0) puisse adopter le repliement β-solénoïde, servant de modèle à la transition amyloïde du domaine HET-S formant le prion, et déclenche la voie de mort cellulaire en aval.

Dans ce travail, nous avons déterminé la structure 3D du du solénoïde adopté par le motif amyloïde fonctionnel NWD2 (R0) à l'aide de contraintes structurales de RMN du solide. Comme dans le modèle HET-s PFD, les motifs R0 du signalosome NLR NWD2 comprennent un noyau central triangulaire hydrophobe conçu par trois feuillets avec un quatrième feuillet pointant vers l'extérieur du coeur amyloïde central. De plus, nous avons réussi à déterminer un modèle 3D de l'interface hétérotypique entre les amyloïdes R0 et R2, c'est-à-dire l'interface amyloïde constituée de deux segments de protéines différents, ici le motif R2 dans HET-s et son précurseur amyloïde (NWD2), impliqué dans la transduction du signal fongique jusqu'à la mort cellulaire programmée. Bien que le noyau hydrophobe triangulaire soit virtuellement conservé dans les hétéroamyloïdes R0-R2, nos données de RMN du solide suggèrent que les 4ème et 8ème β-brins sont dynamiquement flexibles et présentent un haut degré d'hétérogénéité structurelle. Cette différence structurelle cruciale dans l'interface

R0/R2 pourrait exfolier le HET-S sous forme prion du signalosome NWD2. Ce mode de nucléation amyloïde permettrait l'utilisation répétée d'un signalosome, permettant l'amplification du signal. Ici, nous rapportons un mécanisme moléculaire d'un avantage potentiel de réaliser l'amplification du signal dans une architecture à deux gènes, contrairement à la structure tout-en-un plus simple de la signalisation NLR fugitive via l'interface hétéroamyloïde potentiellement conçue avec une stabilité moindre. Par conséquent, l'interface R0/R2 pourrait jouer un double rôle, c'est-à-dire agir comme une surface de référence ainsi qu'une interface de fragmentation du prion.

Il apparaît donc qu'en plus de la capacité à former efficacement le coeur solénoïde amyloïde, la variation de séquence des motifs élémentaires amyloïdes (R0, R1 et R2) est finement réglée, avec (i) un R1 adapté à une capacité de formation d'amyloïde plus faible potentiellement pour éviter l'auto-activation du HET-S et (ii) R0 réduit en contenu de charge pour permettre une polymérisation homotypique et (iii) capable de former différentes interfaces hétérotypiques avec R1 ou R0. L'interface R0/R2 étant potentiellement d'une stabilité légèrement plus faible pour permettre le recyclage du signalosome NWD2.

**Projet 3 - Modes de liaison de la ThioflavinT (ThT) sur les amyloïdes solénoïdes**

Les colorants fluorescents comme le ThioflavinT (ThT) et le rouge congo sont largement utilisés pour un diagnostic de diverses maladies liées aux amyloïdes. Le ThT présente une intensité d'émission de fluorescence accrue lors de la liaison avec l'amyloïde (absorption à ~440 nm et émission à ~480 nm pour le ThT). Divers dérivés de la ThioflavinT non chargés sont utilisés pour l'imagerie et la quantification des plaques amyloïdes associées à diverses maladies neurodégénératives. Cependant, de nombreuses amyloïdes, dont le prion fongique modèle HET-s (218-289), n'affectent pas la fluorescence du ThT, ce qui laisse entrevoir des résultats faussement négatifs. Des études bio-informatiques ont mis en évidence le caractère commun des $\beta$-solénoïdes et des $\beta$-arches dans les amyloïdes en tant qu'éléments structurels. Nous avons donc entrepris de déterminer la base structurelle et moléculaire de la fluorescence ThT indépendante du repli tertiaire dans les amyloïdes adoptant une structure $\beta$-solénoïde.

Nous avons examiné les facteurs déterministes de la propriété de fluorescence des molécules ThT lors de la liaison de différents systèmes amyloïdes, à savoir HET-s(218-289), R0-R0, R0-R1, R0-R2 et R0-R1-R2. Il est intéressant de noter que seules les fibrilles R0-R0, R0-R1 et R0-R1-R2 se sont révélées positives au ThT, tandis que les fibrilles HET-s(218-289) et R0-

R2 ont présenté une fluorescence ThT très faible ou nulle à proximité du pH physiologique (pH 7,5). Pour étudier le mécanisme moléculaire du mode de liaison du ThT sur les fibrilles amyloïdes solénoïdes, nous avons utilisé d'expériences de RMN.

Nous avons observé l'interaction du ThT sur des solénoïdes d'amyloïdes, ce qui a permis d'augmenter la fluorescence et donc de mesurer précisément les paramètres cinétiques d'agrégation. Les résultats d'expériences de RMN étaient en accord avec le site de liaison du ThT sur les résidus 9H-10A-11E. Le groupe amino protoné du ThT forme un pont salin avec E11 et 9H formant ainsi des interactions avec les cycles aromatiques du ThT. Les surfaces amyloïdes étudiées comportent des résidus aromatiques, l'hydrophobicité et les résidus acides améliorent l'accessibilité pour la liaison du ThT et le rendement quantique là où les résidus basiques empêchent la liaison du ThT. La formation potentielle d'un pont salin entre les groupes carboxyle du glutamate dans les fibrilles de R0-R2/HET-s(218-289) peut concurrencer la formation stable d'un pont salin avec le ThT. De plus, l'absence de résidus aromatiques dans R0-R2 et HET-s(218-289) au niveau des sillons les rend ThT négatifs. Cette inférence pourrait également expliquer la différence d'intensité du ThT sur d'autres amyloïdes solénoïdes. R0-R1 montre un ordre de grandeur très élevé dans l'intensité de la fluorescence ThT dans des conditions similaires, car ils manquent de ponts salins entre les spires amyloïdes, ce qui donne des ponts salins libres glutamate (E11) et aspartate (D230). Le rendement quantique du ThT diffère en fonction du mode de liaison, de l'énergie libre de liaison, de la conformation du ThT sur les fibrilles amyloïdes et de la stœchiométrie. Le ThT se lie de préférence sur des sillons de fibrilles exposés aux solvants, composés de résidus acides compatibles pour la formation de ponts de sel ou de résidus aromatiques compatibles pour l'empilement.

Comme les composés dérivés de la thioflavin T sont utilisés pour l'analyse qualitative et quantitative des plaques amyloïdes, nos résultats ouvriraient la voie à une recherche systématique pour découvrir de nouvelles molécules traceuses permettant de diagnostiquer les maladies amyloïdes de manière précise et concluante. Cette étude peut potentiellement aider à une conception rationnelle des inhibiteurs spécifiques de l'agrégation des protéines dans une perspective thérapeutique.

Mots-clés : SSNMR, TDP-43, amyloïdes, β–solénoïde, HET-s, NWD2, Thioflavin-T

# Unité de recherche

[CBMN UMR 5248]

# Chapter 1 INTRODUCTION

Proteins are biomolecular machines that perform vital functions in cells. The newly synthesized mRNA undergoes splicing and translates to the corresponding protein with the help of ribosomes. These are essential steps in the central dogma. After translation, the protein exists as an unfolded polypeptide or random coil, which lacks long-term stability. The protein should be folded into a well-defined three-dimensional structure to perform its functions precisely with long-term stability.[1] The protein folding process begins even before the translation is complete. The folding process is complex and carried out either independently or by the enzymatic assistance of molecular chaperones.

The mutual interactions of the amino acids depending on their nature (charge, hydrophobicity, size, etc.) lead to a consistent, efficient, and functionally active, well-defined 3D structure called the "native state" of the protein. This folding process depends not only on the primary sequence of the protein but also on various molecular crowding influences from the cellular environment.[2] A free-energy landscape based on a statistical approach is generally used to describe an unfolded polypeptide pathway to the native state.[1,3] A newly synthesized unfolded polypeptide undergoes a cascade of conformational adaptation through intermolecular interactions and population intermediates. These intermediate changes are conformationally heterogeneous and determined by both the kinetic and thermodynamic properties of the polypeptide and may not have a single route towards the native state.[4] And these partially folded population states progressively reorganize into the formation of the *de novo* native state. By the strong influence of evolution, the proteins have a rugged funnel-like shaped landscape biased towards the native structure. The surface of the folding funnel trailed by each polypeptide is unique at specific conditions depending on its kinetic and thermodynamic parameters[5]. Failure to restructure into the native state, the polypeptide chain remains inactive. In some cases, the misfolded proteins lead to toxic oligomers and further assembles to form amorphous or thermodynamically more favored amyloid fibrillar aggregates.[6]

However, there exists a class of proteins or regions of proteins called intrinsically disordered proteins (IDP), or intrinsically disordered protein regions (IDPRs) that show significant conformational flexibility and plasticity and low sequence complexity but plays a vital role including protein-protein interaction, signal transduction, gene expression, etc.[7–10] The remarkable ability to form liquid-liquid phase separation via the multitude of transient interactions is fascinating by IDPRs.[11] It's functional and pathological aspects led to extensive research on IDPRs in science today.[11–13] Under pathological conditions, these IDPs can also undergo a structural transition from a disordered state to beta-sheet rich aggregates called amyloids.[14] The deposition of amyloid aggregates or plaques is considered a pathological hallmark of various neurodegenerative diseases such as Alzheimer's, Parkinson's, Huntington's, and Lou Gehrig's diseases.[1516] Therefore, it is essential to study the mechanism

of protein misfolding, aggregation, and amyloid formation to develop an efficient therapeutic strategy.

### 1.1 Amyloids

Amyloids are a unique class of protein self-assembly structures in the biological world that are linked to many pathological conditions and functionally relevant. The term "amyloid" was coined initially by Schleiden and followed by Virchow as it has been found in iodine stained deposits in the liver mistaken as starch-like materials (derived from the Latin word "amylum" for starch).[17] Later, these deposits were identified as proteinaceous aggregates owing to the high nitrogen content.[17] A more generally accepted definition of amyloids can be stated as aggregates of proteins, rich in β-sheets with a fibrillar morphology of about 7-13 nm in diameter as observed by electron microscopy (EM) or Atomic force microscopy (AFM) and a few microns in length, which binds organic dyes, such as Thioflavin-T (ThT), Congo red (CR) or its derivatives and displays a characteristic cross-beta diffraction pattern by X-ray diffraction (XRD).[14,18–21] The cross-beta architecture involves the arrangement of β-strands perpendicular to the fibril axis and assembled into β-sheets that run along the fibrillar axis.[2223]

### 1.2 Cross-β structure, and stability

Interestingly, many amyloid proteins, regardless of the difference in their primary sequence, share a common structural characteristic called cross-b structure, characterized by characteristic X-ray diffraction (XRD) pattern, first observed in silk.[24] The particular core cross-β structure of the amyloid fibrils consists of an orderly arrangement of β-sheets. The hydrogen bonding direction runs parallel to the fibril axis and the β-strands in orthogonal, as in the rungs of a ladder.[24] This continuous packing of cross β-sheets generates a unique X-ray diffraction pattern with a meridional reflection at ~ 4.7 Å and an equatorial diffraction signal at ~ 6-12 Å corresponding to inter-strand and inter-sheet spacing, respectively (**Figure 1.1).** [24–26] Several high-resolution structures of 3D microcrystals of short peptide segments derived from pathogenic amyloid fibrils (*e.g.,* Aβ, tau, α-synuclein, TDP-43, and sup35) have been reported.[27–29] These studies have formulated a distinct class of amyloid structures that could serve as the cross-β spine termed as steric zippers. Here, the adjacent β–sheets are tightly held together with facing side chains interdigitated via an interface devoid of water (dry interface) between them. There is a substantial gain in entropy for forming a "dry face" via the complete exclusion of water molecules between the β-sheets and forming a steric zipper.[27,30] The crystallographic structures of microcrystals formed by many pathologically relevant amyloidogenic short peptides have revealed atomistic details of the cross-β architecture.[27,30] Although cross-β patterns are a diagnostic hallmark of amyloid, any structural information leading to high-resolution structures is limited due to the difficulty in growing highly ordered crystals from amyloid fibrils due to polymorphic nature and failure to provide any of the tertiary and quaternary arrangement that can be very different among amyloid aggregates.

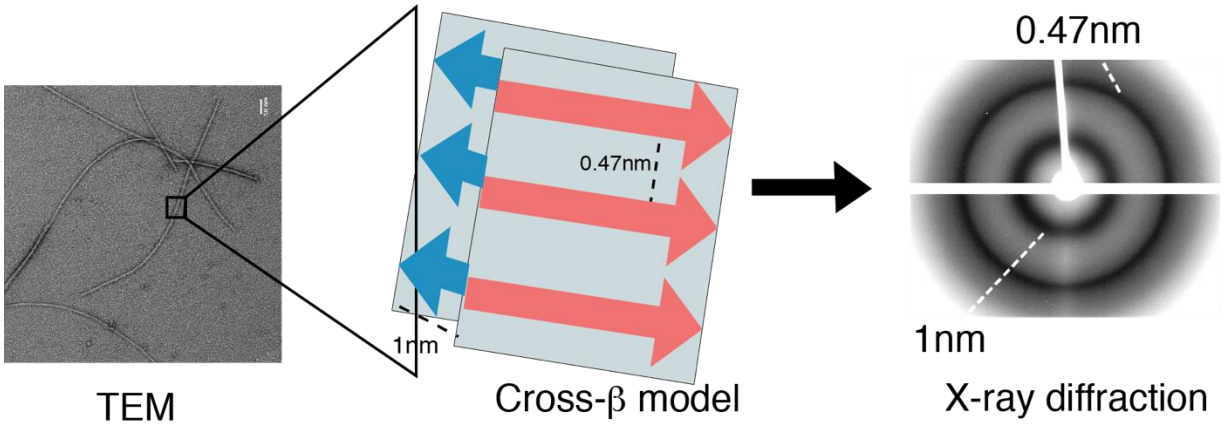



*Figure 1.1 Schematic representation of the cross-β x-ray diffraction pattern displayed by amyloid fibrils. The amyloid fibrils generate a unique X-ray diffraction pattern with a meridional reflection at ~4.7 Å and an equatorial diffraction signal at ~6-12 Å corresponding to inter-strand and inter-sheet spacing, respectively.*[31]

The one-dimensional nature of its order hinders the formation of three-dimensional crystals of amyloid fibrils. Only smaller amyloidogenic peptides are short enough to be able to grow in a three-dimensional lattice. As a result, the characteristic cross-β structure is one dimensional with a single rotational and translational symmetry element. The structural stability of amyloid fibrils is contributed by several factors, including the hydrogen bond network that runs parallel to the fibrillar axis formed between the backbone amide groups and carbonyls. Additionally, the side-chain carbonyls and amides also contribute to the hydrogen bonds. Amyloid fibrils formed by proteins rich in Q/N residues in their amino acid sequence can be stabilized by the hyper-cooperative hydrogen bonds unique to Gln/Asn side chains.[32,33]Another significant interaction includes the aromatic stacking via π–π interaction between side-chains of aromatic residues. Also, side-chains of Asn, Gln, Thr, Ser can form extensive hydrogen bonds that run up and down the fibrils. These structural elements are called ladders and can also provide structural stability.[2730] The formation of salt bridges between side chains of complimentary charged residues can significantly contribute to fibril stability.[34] Nevertheless, van der Waals interactions between closely spaced β-sheets also strengthen the fibril structure. Although the amyloid fibrils formed by full-length proteins have complex structures than the ones created by short peptides, the underlying stabilizing interactions are likely to be the same.

### 1.3 Mechanism of amyloid transition and prion property

Structural biologists have always been fascinated by amyloids and prions for a long due to their conversion from a soluble state to an insoluble form and their unique self-templating property, resulting in robust supramolecular self-assemblies with high stability. More than 40 proteins or protein segments are documented to form pathological amyloid fibrils and almost the same number of functionally relevant amyloid in the ever-expanding field of amyloid fibrils self-assemblies. The underlying process in the transition of soluble state

to insoluble aggregates is similar to molecular crystallization involving nucleation and elongation.[35–37]

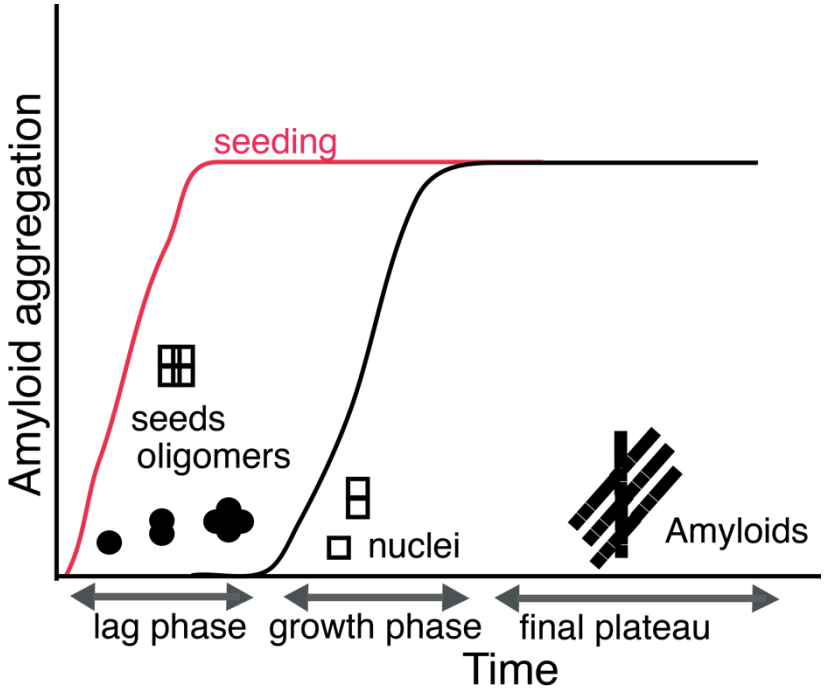



*Figure 1.2 Aggregation kinetics of amyloids displaying the lag phase, growth phase, and saturation phase and alternate path via seeding.*

The aggregation kinetics of the amyloid transition and aggregation can be followed by monitoring turbidity or ThT fluorescence displaying a sigmoidal curve with three distinct phases: a lag phase, an exponential phase, or growth phase or elongation phase, and a saturation phase or plateau (**Figure 1.2**). The main event occurring in the lag phase is the formation of critical nuclei, i.e., the smallest aggregating species stable enough to pursue the elongation phase rather than getting dissolved by various interactions of monomers. The nucleation phase is inherently a slow process determined by a high free-energy barrier.[38] Once the critical nuclei are formed, the additional growth occurs at a relatively higher rate with a lower free energy barrier. In a classic nucleated polymerization mechanism, the monomers get partially or entirely misfolded and undergo a set of the thermodynamically unfavorable process generating the nuclei.[14,38] This model can be used to explain a variety of aggregation for the proteins such as Aβ, α-synuclein, insulin, etc.[39–41] In other cases, the monomeric proteins rapidly get converted to misfolded unstable aggregates which cannot grow in to mature fibrils then undergo structural reorganization to generate competent nuclei, acting as a template for later nucleation events.[42] This nucleated conformational conversion of the pre-formed oligomers to critical nuclei has been observed in many systems, including Aβ, α-synuclein, IAPP, etc.[41,43,44] The natively folded globular proteins with buried amyloidogenic core must undergo partial misfolding resulting in solvent-exposed prion domains and trigger aggregation.[45] A combination of nucleated conformational conversion or nucleated polymerization mechanism could contribute towards the formation of native-like

aggregates. Cu-Zn superoxide dismutase, β2-microglobulin, lysozyme, etc., have been reported to show a native-like aggregation mechanism.[46–48]

The elongation phase describes a steep transition where the rate of amyloid transition is greatest. The elongation phase rate depends on two processes; the first one involves a monomer independent pathway such as fragmentation fibrils, where the rate relies on the concentration of existing fibrils. The second pathway describes a monomer-independent process such as secondary nucleation, where the fibrils surface act as a catalyst for the nucleation of monomers to generate new aggregates (surface catalysis).[37,49,50]The process of formation of critical nuclei can be surpassed by the addition of competent nuclei obtained from pre-formed fibrils, thereby reducing the lag phase. This addition of external competent nuclei is called seeding, critical in prion-like propagation of many neurodegenerative diseases.[51,52] However, the processes in elongation phases are fundamentally the same as in the lag phase. The lag phase is simply when the aggregate concentration is below the detection limit even though fibrils exist in this step.[37]

The self-templating ability of amyloid proteins helps them to propagate as prions. Once a natively folded protein/IDP acquires the amyloid fold, essentially acting as a catalyst in templating the soluble form of the protein to adopt the same fold and aggregate. The amyloid aggregates shown to have very high thermal stability and protease resistance. It was hypothesized that the infectious agent responsible for transmissible spongiform encephalopathies (TSEs), including scrapie in sheep, chronic wasting disease (CWD) in deer, mad cow disease in cattle, and Creutzfeldt–Jakob disease (CJD) in humans, is exclusively a variant form of the protein expressed in the organism itself rather than infection from pathogens.[53,54] The term "prions" was first coined in 1982 by Stanley B. Prusiner in short for "proteinaceous infectious particle" about the self-perpetuating and infectious nature of the prion variant of the prion protein (PrP). The disease propagation mechanism involves converting the native, cellular protein PrP into $PrP^{Sc}$ (scrapie PrP = PrPSc), acquiring a structure with high β-sheet content through post-translational modification. The PrP protein prion variant could act as the template of the trans conformation of nascent $PrP^{Sc}$ protein to adopt the same prion fold.[54] Compelling evidence from genetic, biochemical, and neuropathological studies proved the prion hypothesis's authenticity, and Prusiner was awarded the Nobel prize in Physiology or Medicine in 1997 for the discovery of prions. Later, a variety of other mammalian proteins with prion-like domains (PrLD) were identified in the etiology of neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration with ubiquitin-positive inclusions (FTLD-U), and Huntington's disease, etc.[55,56] A handful of fungal prions were discovered which behave similarly to PrP, but, in general, they are non-toxic to their hosts and gave strong support to the protein-only concept and demonstrated the self-templating and propagating nature of the prion protein.[57–59] However, fungal prions are shown to be evolved to perform specific functions advantageous to the microorganism and enhance their ability to adapt and survive in diverse environments.[60]

### 1.4 Structural polymorphism in amyloids

Another peculiar property of amyloid fibrils includes polymorphism, a phenomenon in which a polypeptide sequence under the same environmental conditions adopts multiple structurally distinct molecular conformation in amyloid form. Amyloids, in general, can form different polymorphs depending on the aggregation conditions like protein concentration, type of buffer, incubation temperature, pH (affected by the isoelectric point of the protein), protein concentration, presences of external compounds such as metal ions, or chelating agents, chaotropic or redox reagents, etc. Changing the aggregation conditions *in-vitro* can affect the distribution of polymorphs in the sample and the magnitude of the different polymorphic populations.[61,62]

However, functional amyloids typically form isomorphic structures compared to pathological amyloids that adopt a diverse set of polymorphic entities. The presence of severe polymorphism in disease-related amyloids could be attributed to the uncontrolled misfolding of the proteins. The misfolding protein can attain multiple local minima in the aggregation energy landscape with relatively small energy differences energy barriers but larger than thermal energy. Depending on the changes in the environmental conditions, they can lead to distinct polymorphs corresponding to different local minima.[63] Misfolded proteins inherently comprise a high degree of disorder. Their structural plasticity plays a critical role in distributing a wide range of conformational polymorphs when such proteins form self-assembled macrostructures. In functional amyloids, the amino acid sequences are evolutionarily modified to attain a specific conformation with a steep local minimum in the aggregation energy landscape.[64] Unlike functional amyloids, protein sequences of pathological amyloid fibrils are not evolutionarily optimized.

EM or AFM can be successfully employed to differentiate polymorphic fibril morphology based on the number of protofilaments, degree of fibril twisting, cross-sectional area, or mass-per-length of the fibrils.[65] Fibril polymorphs with a difference in these parameters can co-exist in the same sample. However, identical spectral characteristics via EM or AFM measurements do not always assure an isomorphic sample. The polymorphism in the fibrils could occur due to the different binding modes between the protofilaments at the mesoscopic level. Also, the observation of spectroscopic heterogeneity may not indicate different polymorphs; instead, the occurrence of multiple conformers of polypeptides (monomers) contributing to the fibril morphology is equally possible, as observed in several A$\beta$ (1-40) fibril samples.[66,67] The presence of intra-sample polymorphism is spotted in *in-vitro* samples, which are chemically synthesized or recombinantly expressed proteins, and in patient-derived amyloid fibril samples.[68]

In general, the polymorphism in cross–$\beta$ fibrils occurs primarily from the difference in the number of protofilaments contributing to the mature fibril, variations in the relative orientation of the protofilaments concerning each other differences in monomeric subunits in the protofilaments. However, these characteristics are intertwined so that variations in the protofilaments can lead to differences in orientation in the nearby protofilament subunits due to altered interaction surface for the possibility of secondary nucleation.[37,69] The

difference in monomeric subunits can be further characterized as packing polymorphism where the cross-β subunits are arranged differently (like parallel versus anti-parallel strands), and the segmental polymorphism attributes to the residues involved in the amyloid core.

Further, there also exists side-chain polymorphism reflecting the difference in conformation of side-chains in subunits in protofilaments. These structural heterogeneities can be identified by solid-state NMR (SSNMR) at an atomic level, where distinct amyloid polymorph provides distinct SSNMR fingerprints detecting residue-specific conformation details in the cross-β structure. Different polymorphs in the sample are indicated by more than one set of resonances in the spectrum. Additionally, more than one NMR peak for a unique set of residues or side-chains suggests the presence of polymorphism at the atomic level for the corresponding residue. The variations in side-chain and backbone conformations can also lead to numerous and undistinguishable chemical shifts or broad line-width.[70,71] However, the presence of repetitive sequence motifs, specific dynamic properties of amino acids in the NMR time scale (static or dynamic disorder) can also lead to broad SSNMR signals.[72] Star-like and bundle-like arrangements or amorphous fibrils meshwork formed by the different fibril networks is also observed.[73–77] All kinds of polymorphism can be jointly called aggregate morphotypes displaying fibril morphology variations. Aggregate morphotypes can be defined as the distinguishable subtype of amyloid aggregates formed by a specific protein from its other aggregates based on morphology or staining features.[65]

Polymorphism of amyloid fibrils is of great interest in research as their structural variation is accounts for clinical subtypes of distinct neurodegenerative diseases. Different polymorphs display differences in properties like nucleation (fragmentation or secondary nucleation via acting as a surface catalyst), the potential to propagate as prion (propagation rate), and cytotoxicity. [37,51,78–82]

In the later part of the chapter, we will address TDP-43, which forms pathological amyloid aggregate responsible for ALS and FTLD, and a HET-s/NWD2 paradigm, which plays an essential role in the fungal cell death signaling pathway. Chapter 2 will address the research progress in the structural characterization of TDP amyloid fibrils, and Chapter 3 on the molecular mechanism of prison-based signal transduction in the fungal cell death signaling pathway addressing the HET-s/NWD2 paradigm.

### 1.5 Pathological amyloid - TDP-43

This section will be reviewing the available bibliography on TDP-43 structure, its role in cellular functions, prion behavior, chemical properties, and pathological aggregates.

Protein misfolding and aggregation are associated with numerous fatal neurodegenerative diseases.[83] The formation of senile plaques and neurofibrillary tangles (NFTs) composed of paired helical filaments of tau (Tau PHF) and amyloid-β (Aβ) is considered to be the biological hallmark of Alzheimer's disease.[84] $A\beta_{1-40}$ and $A\beta_{1-42}$ formed by the proteolytic fragmentation of the amyloid precursor protein (APP) are the most prominent forms of amyloid fibrils found in amyloid plaques of Alzheimer's patients.[85] The aggregation of the microtubule-associated protein, tau, is associated with a broad range of brain diseases

generally termed as tauopathies, which includes Alzheimer's disease (AD), FTLD with tau inclusion (FTLD-tau) with progressive supranuclear palsy, argyrophilic grain disease, corticobasal degeneration, and amyotrophic lateral sclerosis (ALS).[86,87] Similarly, the deposition of insoluble aggregates of intrinsically disordered protein α-synuclein in Lewy bodies and Lewy neurites is involved in etiologies Lewy body-related disorders including Parkinson's disease, dementia with Lewy bodies (DLB) with other neurodegenerative conditions, characterized as synucleinopathies.[88–90]

The TAR DNA binding protein of 43 kDa (TDP-43) is observed as the main component in the neuronal cytoplasmic inclusion bodies of patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar dementia (FTLD).[91,92] ALS is a lethal neurogenerative disease characterized by the gradual degeneration of both upper and lower motor neurons that affects ~5 out of 100,000 individuals each year worldwide.[93,94] The ALS patients suffer from muscle spasticity and hyper-excitability, and gradual paralysis, eventually followed by respiratory failure and pneumonia and death within about 3–5 years of exhibiting the disease symptoms.[9596] FTLD is associated with progressive neurodegeneration of frontal and temporal lobes leading to dementia, affecting ∼15–22 per 100,000 individuals each year worldwide.[97–99] Surprisingly, ∼90% of the ALS cases and∼50% of all FTLD cases are associated with TDP-43's aggregation.[100,101]

### 1.5.1 Structure of TDP-43

TDP-43 was initially discovered as an HIV-transcription repressor protein, critical for the viral gene expression in 1995, later reported to be involved in RNA splicing of cystic fibrosis transmembrane conductance regulator (CFTR) exons.[102] TDP-43 is composed of 414 amino acids with a theoretical molecular weight of 44740 Daltons. It belongs to the large family of nuclear factors known as heterogeneous nuclear ribonucleoproteins (hnRNP), which binds to specific RNA sequences via highly conserved RNA recognition motifs (RRMs). TDP-43 is composed of a well-folded N-terminal domain (NTD, aa 1-81), a nuclear localization signal (NLS, aa 82–106), two highly conserved RNA recognition motifs RRM1 (aa 106-175) and RRM2 (aa 192-262), a nuclear export signal (NES, aa 239-274) and an intrinsically disordered C-terminal domain (aa 275-414). The C-terminal domain (CTD) comprises of consisting of the hydrophobic segment (318–340), the Q/N rich segment (341–367), and two subdomains rich in glycines, aromatic residues, and serines, called GaroS1 and GaroS2 (**Figure 1.3**).[103–107]

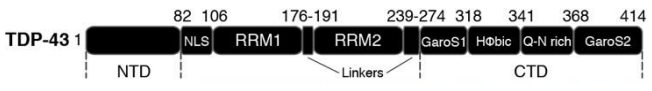



*Figure 1.3 The primary sequence of TDP-43 (1–414) boxes are used to represent different functional domains.*

The TDP-43 N-terminal domain can form reversible homodimers and higher-order oligomers in-vitro and in-vivo, essential to perform physiological functions like mRNA splicing. Arguably, the formation of TDP-43 dimers is one of the reasons for its pathological

aggregation.[108–110]About three monomeric and two dimeric high-resolution TDP-43 NTD structures have been reported to date (Table 1).[107,111,112] The monomeric NTD adopts a ubiquitin-like β-grasp fold consisting of one α-helix and six β-strands with a β1–β2–α1–β3–β4–β5–β6 arrangement, a structure homologous to the DIX domain of Axin 1.[107] The dimeric structures display a head-to-tail arrangement of the NTD monomeric subunits essential for splicing activity and shown to spatially separate CTDs and prevent pathological aggregation.[110,113] Additionally, a phosphomimic substitution at the conserved residue Ser 48 discourages liquid-liquid phase separation (LLPS) *in vitro* and disrupts RNA splicing activity.[113] However, a recent study using single-molecule fluorescent techniques shows that the NTD can undergo reversible oligomerization and enhance the intrinsically disordered C-terminal domains' aggregation propensity full-length TDP-43.[114] The nuclear localization signal (NLS) connects the NTD and RRM domain, essential for the transport of TDP-43 into the nucleus from cytoplasm mediated by Importin-α that binds the NLS of TDP-43.[115,116] Mutations or deletions on NLS leads to cytoplasmic mislocalization of TDP-43 and aggregation. Caspase cleavages at positions at Asp89, Met-85, and A90V mutation enhances the accumulation of TDP-43 in the cytoplasm and formation of insoluble aggregates.[117–121]
The two TDP-43 RNA Recognition Motifs (RRMs), RRM1, and RRM2 are connected by a 15 amino acid linker and composed of five β-strands and two α-helices arranged in a β1–α1–β2–β3–α2–β4–β5 manner with the conserved RNA-binding aromatic and hydrophobic residues present in the β-strands. The TDP-43 tandem RRMs uniquely binds the single-stranded RNA or DNA bases and sugar rings in the 5’ to 3’ direction, unlike typical tandem RRMs.[104,105] Accumulating studies point out that TDP-43 preferentially binds to (TG)/(UG) sequences.[104,122–125] Like other hnRNP proteins with multiple RRMs, TDP-43 acquires nucleic acid binding specificity by an interaction between the two RRMS, where RRM2 acts as a switch to regulate the sequence specificity via Arg151.[126] Interestingly, the RRMs are shown to possess amyloidogenic regions and misfold and initiate or accelerate TDP-43 aggregation.[127,128] The RRM2 has been shown to form cystine linked homodimers that could be converted into amyloid aggregates.[129] A synthetic peptide derived from the RRM-2 domain (246-255) is highly aggregation-prone.[130] Recently, the atomic resolution structure of RRM2 $^{247}$DLIIKGISVHI$^{257}$ has been documented, forming multiple amyloid polymorphs characterized by the difference in backbone conformation and symmetry classes of steric zippers.[131] Interestingly, RRMs are associated with several post-translational modifications and mutations ( D169, E24,6, and D247 to glycines) relevant in the pathological aggregation of TDP-43. Additionally, several mutations on RRM domains adversely affect the RNA binding capacity.[132127133104] The nuclear export signal (NES) regulates the transport of TDP-43 between the nucleus and cytoplasm, and the deletion of this segment results in the accumulation of cytoplasmic TDP-43 and aggregation.[120]

Like other reported prion domains, the TDP-43 C-terminal domain (CTD) is intrinsically disordered and highly aggregation-prone, playing a critical role in TDP-43 aggregation.[134] Interestingly, 90% of disease-related mutations and many post-translation modification sites are located at the CTD.[135] The TDP-43 CTD is essential for RNA splicing activity and interaction

with several partner proteins.[136–140] Ubiquitinated and phosphorylated aggregates of truncated TDP-43 C-terminal fragments of size ~25–35 kDa formed by the aberrant cleavage by the caspase are shown to be highly cyto-toxic and is considered as the biological hallmark of ALS.[141–143] Strikingly, ~30% of human proteins possessing a prion-like domain belong to the family of proteins, which are of RNA/DNA-binding proteins and several of which, including TDP-43, FUS, hnRNPs, are associated with the etiology of numerous neurodegenerative diseases.[144]

The CTD can be divided into distinct sub-regions based on the amino acid sequence, namely the hydrophobic segment (318–340), the Q/N rich segment (341–367), and two segments rich in glycines, aromatic residues, and serines, called GaroS1 and GaroS2.[33] Multiple studies indicate that the C-terminal domain is necessary and sufficient for amyloid aggregates.[145–148] The hydrophobic segment can form amyloid filaments and adopt a helical structure in solution and displays membrane interaction via the Met311-Gln343 subdomain adopting a well-folded Ω-loop-helix structure. ALS mutations like G335D and Q343R disrupts the helix-turn-helix system. [147–149] The Q/N rich segment is similar to prions like domains in fungal proteins Sup35, Ure2p that can form *in vitro* aggregates which are heterogeneous, and rich in β-sheet structures.[55,148,150,151] Eisenberg and coworkers have shown that the six peptides derived from CTD can form different classes of steric zippers significant in pathological TDP-43 aggregation. Also, the segment 312-317 with ALS mutations A315T and A315E can form irreversible aggregates from labile Low complexity Aromatic-Rick Kinked Structures (LARKS) on phosphorylation via interaction with close LARK segments facilitated by the formation of SGs (stress granules) also able to form phase-separated droplets and hydrogels unlike other peptides from the CTD.[152] A wealth of studies suggests the Q/N region adopts an extended β-hairpin structure stabilized by hyper-cooperative hydrogen bonds.[148,153–155] Interestingly, the recent cryo-EM research of *Cao et al*. describes the ability of two peptide segments from CTD termed SegA (residues 311-360) that can form three structurally distinct fibrillar polymorphs, all sharing a common dagger-shaped fold and SegB (residues 286-331) containing an ALS mutation A315E to create an R-shaped fold which drives to irreversible aggregates aggravating the pathology.[155] A recent solid-state NMR study suggests a helix to the β-sheet helix-to-sheet structural transformation of TDP-43 (residues 311–360) adopting parallel in-register cross-β structure in the fibrillar form composed of five β-strands.[156] Recently, a crystal structure of TDP-43 fragment containing amino acids 1-270 comprised of NTD and RRMs has been reported, revealing only the NTD with the molecules arranged as parallel spirals with adjacent molecules positioned in a head to tail manner.[157] A table containing all the reported structures of TDP-43 and peptides is shown below **(Table 1**).

*Table 1.1 reported structures of TDP-43 domains and their amyloid structures*

| Technique | Sequence | PDB code | Reference |
|---|---|---|---|
| **N-terminal domain** | | | |
| NMR | 1-77 | 2N4P | *Mompeán et al., 2016* |
| XRD | 2-79 | 5MDI | *Afroz et al., 2017* |
| NMR | 1-102 | 5MRG | *Mompeán et al., 2017* |
| NMR | 1-77 | 5X4F | *Jiang et al., 2017* |
| NMR | 1-80 | 6B1G | *Wang A. et al., 2018* |
| XRD | 1-80 | 6T4B | *Wright et al., 2020* |
| **RRM domains** | | | |
| NMR | 102-269 | 4BS2 | *Lukavsky et al., 2013* |
| NMR | 96-185 | 2CQG | |
| XRD | 103-179 | 4IUF | *Kuo et al., 2014* |
| XRD | 103-168 | 4Y00 | *Chiang et al., 2016* |
| XRD | 103-180 | 4Y0F | *Chiang et al., 2016* |
| NMR | 193-267 | 1WF0 | – |
| XRD | 192-261 | 3D2W | *Kuo et al., 2009* |
| **RRM amyloid structures** | | | |
| XRD | 248–253 | 5W50 | *Guenther et al., 2018* |
| Electron microscopy | 247–257 | 5W52 | *Guenther et al., 2018* |
| Electron microscopy | 247–257 | 5W7V | *Guenther et al., 2018* |
| **C-terminal domains** | | | |
| NMR | 307–348 | 2N2C | – |
| NMR | 311–360 | 2N3X | *Jiang et al., 2016* |
| NMR | 311–360 G335D | 2N4G | *Jiang et al., 2016* |
| NMR | 311–360 Q343R | 2N4H | *Jiang et al., 2016* |
| **C-terminal amyloid structures** | | | |
| XRD | 312–317 | 5WHN | *Guenther et al., 2018* |
| XRD | 312–317 A315T | 5WHP | *Guenther et al., 2018* |
| XRD | 370–375 | 5WIA | *Guenther et al., 2018* |
| XRD | 396–402 | 5WIQ | *Guenther et al., 2018* |
| Electron microscopy | 312–317 | 5WKB | *Guenther et al., 2018* |
| XRD | 300–306 A315E | 5WKD | *Guenther et al., 2018* |
| X-ray diffraction and micro-electron diffraction | 328–333 | 6CB9 | *Guenther et al., 2018* |
| XRD | 321–326 | 6CEW | *Guenther et al., 2018* |
| Electron microscopy | 312–317 Phosphorylated A315T | 6CF4 | *Guenther et al., 2018* |

| | | | |
|---|---|---|---|
| Electron microscopy | 333–343 | 6CFH | *Guenther et al., 2018* |
| Electron microscopy | 311–360 | 6N3A | *Cao et al., 2019* |
| Electron microscopy | 312–352 | 6N3B | *Cao et al., 2019* |
| Electron microscopy | 288–314 A315E | 6N3C | *Cao et al., 2019* |
| Electron microscopy | 312–347 | 6N37 | *Cao et al., 2019* |

Interestingly, the TDP-43 can exist in dynamic protein droplets or LLPS via its intrinsically disordered C-terminal domain. The hydrophobic subregion undergoes transient helix fold facilitating the TDP-43 phase separation, and ALS associated mutations in this region disrupt the helical stability and inhibit phase separation.[138] In conclusion, the prion-like C-terminal domain could transiently adopt several higher order self-assemblies in LLPS. A detailed analysis of TDP-43 associated with LLPs is discussed further.

### 1.5.2 TDP-43 physiological functions

TDP-43 is predominantly a nuclear protein that shuttles between the nucleus and cytoplasm, engaging in diverse functions in mRNA metabolism, which involves mRNA transcription, splicing, maturation, stability, transport, and translation (**Figure 1.4**).[100,158] TDP-43 is associated with about 30% of mRNA targets of the entire transcriptome and predominantly binds the 3' UTRs and plays a critical role in mRNA stability, maturation, and transport.[124]

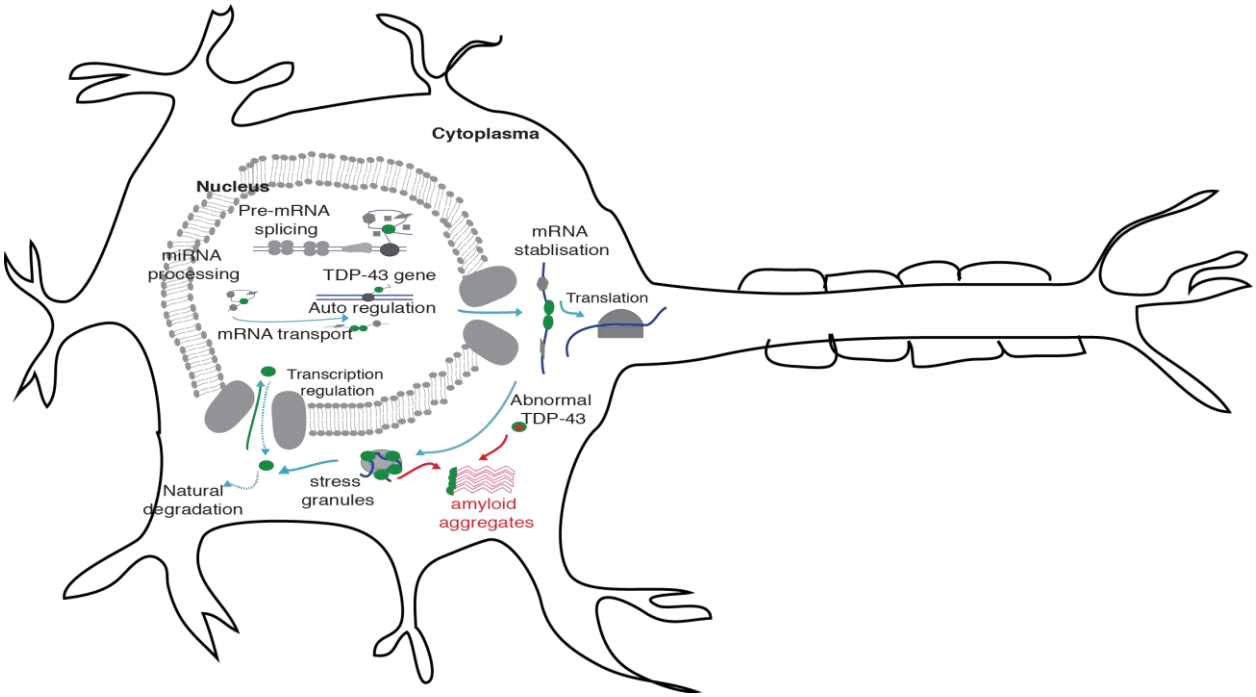



*Figure 1.4 Normal neuronal function and pathological aggregation of TDP-43. TDP-43 (green ellipses) is involved in many vital cellular activities such as RNA splicing, transport, translation regulation, and autoregulation of its expression before undergoing physiological degradation or transferring back to the nucleus. However, mutation facilitated protein misfolding (red*

*asterisk), aberrant cleavage, cytoplasmic mislocalization, aging, or increased local concentration in stress granules may lead to the formation of amyloid-like fibrillar aggregates.*

The aberrant aggregation of TDP-43 directly affects mRNA splicing of various essential genes like Cystic fibrosis transmembrane conductance regulator (CFTR), TARDBP (TDP-43 and others), FUS, SNCA (α-synuclein), HTT (Huntingtin), and APP (Amyloid precursor protein, amyloid-β), etc.[126,159–161] TDP-43 binding the 3' UTRs of mRNAs to have diverse roles like its autoregulatory activity, human low molecular weight (hNFL) mRNA stability, and destabilizing activity on Vascular endothelial growth factors (VEGFs) and granulin (GRN) mRNAs that affects progranulin expression.[123,162,163] RNA molecules are transported in the cell with the assistance of microtubules via ribonucleoprotein (RNP) granules formed by the interaction of TDP-43 with RNA molecules.[164] TDP-43 also controls the translation of futsch mRNA and its expression at the neuromuscular junction in Drosophila and suggests a microtubule-dependent mechanism in the context of aggregation of TDP-43 and ALS.[165] TDP-43 is also involved in the translational machinery, acting as a translational repressor. Its interaction with polyribosome mediated by the interaction with the specific ribosomal protein, RACK1, enhances aggregation of TDP-43 in the cytoplasm under ALS conditions.[166] TDP-43 can interact with the nuclear Drosha complex and bind directly to the relevant primary miRNAs (pre-miRNAs) and enables the production of precursor miRNAs. Also, TDP-43 can interact with the Dicer complex promoting the pr-mi-RNA processing by binding to their terminal loops.[167]

### 1.5.3 Role of post-translational modifications in TDP-43 aggregation and Toxicity

TDP-43 aggregates with several post-translational modifications like ubiquitination, hyper-phosphorylation, and aberrant cleavage are observed in the cytoplasm inclusion bodies.[168] Several other post-translational modifications such as acetylation, poly ADP-ribosylation, and cysteine oxidation has been identified linked to ALS.[169] Though the ubiquitination sites in TDP-43 remain mostly vague, numerous phosphorylation sites at serine, threonine, or tyrosine residues (41 serine, 15 threonine, and eight tyrosine residues in fl-TDP-43) have been identified mainly located on the C-terminal domain.[170] Mutations affecting ubiquitination are found to decrease the accumulation of TDP-43, suggesting a modulating role of ubiquitination in TDP-43 aggregation.[171] However, ubiquitination plays a differential role in the regulation of soluble and aggregated TDP-43species. The K-48 linked polyubiquitin chains direct proteasomal-mediated degradation while K-63 linked polyubiquitin chains refer to autophagic removal of the TDP-43.[172] The TDP-43's phosphorylation accelerates cytoplasmic mislocalization and aggregation in the neuronal cells.[121,173–175] Interestingly, the brain cortex predominantly shows deposition of phosphorylated C-terminal fragments and the spinal cords with phosphorylated full-length TDP-43 in ALS and FTLD patients.[176] Antibodies designed to bind phosphorylated TDP-43 are clinically used to detect pathological TDP-43 inclusions. Out of 20 lysine residues in TDP,43 only a few are reported to be prone to acetylation, such as the K-145 and K-192, which impairs RNA binding, mitochondrial dysfunction, and enhances aggregation.[177,178]

Poly ADP-ribosylation (PARylation) is another post-translational modification associated with DNA damage repair relevant in cell cycle regulation, chromatin reorganization, and cancer.[179] This involves adding ADP-ribose units to the carboxyl group of glutamate or aspartate on the target proteins forming an ester bond catalyzed by Poly (ADP-ribose) polymerase (PARP) enzymes.[180] Polymeric PAR chains can be created by ribose bonds between the subunits. The negative charge on PAR can alter the protein-DNA/RNA and the protein-protein interactions by changing the structure of the target proteins. A PARylation enzyme, tankyrase, reduces TDP-43's aggregation by promoting phase separation and accumulation of TDP-43 in the stress granules.[181] Another post-translational modification of TDP-43 is cysteine oxidation, which causes the re-localization of TDP-43 into cytoplasm and form stress granules.[133,182] The Cys39 and Cys50 at the N-terminal domain form disulfide bridges forming homodimers at first, then tetramers, and inhibit TDP-43 aggregation.[112] The cleavage of TDP-43 into CTD fragments of size 20–25 kDa or 35 kDa has been observed in patients with ALS or FTLD mediated by several proteases like calpain, caspase 3, and caspase 7.[92,143,183–186]

### 1.5.4 Prion-like behavior of TDP-43

TDP-43 exhibits prion-like characteristics in cell culture studies.[100,187,188] This prion-like behavior of TDP-43 essentially categorizes ALS and FTLD as a prion-like disease. The amyloid cross-β forming capacity of the TDP-43 C-terminal domain and its ability to seed native TDP-43 to an amyloid state plays the underlying principle behind its prion-like behavior.[145,147,189] The aggregated post-translationally modified full length or C-terminal fragments of TDP-43 recovered in the detergent-insoluble fraction derived from ALS brain extracts are shown to induce amyloid transition in nuclear, soluble, recombinant wild-type TDP-43 in neuronal and glial cultures.[190–192] Interestingly, TDP-43 can independently seed the aggregation irrespective of superoxide dismutase-1 (SOD1) fibrils or α-synuclein fibers.[193,194] Shreds of evidence suggest that TDP-43 can propagate to neighboring cells, and elevated levels of free and exosomal TDP-43 are identified in the cerebrospinal fluid (CSF) of ALS patients.[195–197] The mechanism of cell-to-cell propagation is still elusive. Still, some hypothesis proposes transfer via exosomes, endocytosis, and even passive diffusion.[191] Nevertheless, the absence of extensive spread of TDP-43 aggregates in the somata of degenerated neurons on post-mortem analysis of patient samples suggests that the TDP-43 aggregates do not spread to adjacent brain cells.[198] However, the spread of ALS is observed all over the neuraxis in both upper and lower motor neurons, possibly by the active transport of TDP-43 in the motor neuron axons via zones of axonal contacts.[198–200]

Accumulating data suggests that the prion-like behavior of aberrant aggregation of TDP-43 contributes to TDP-43 pathology. Additional studies are essential to the mechanism of cell-to-cell transfer of TDP-43 in ALS and FTLD to propose a novel therapeutic strategy.

### 1.5.5 Phase separation of TDP-43

The formation of membrane-less condensed liquid-like droplets of proteins, RNA, and other biomolecules called a liquid-liquid-phase separation (LLPS) is associated with several neurodegenerative diseases. TDP-43 and other RNA binding proteins like FUS and hnRNP A1/2 undergo phase separation via a transient-multivalent interaction of intrinsically disordered regions. The low complexity domains (LCD) of proteins generally consist of amino acid likes glycine, serine, glutamine, proline, glutamic acid, lysine, and arginine.[201,202] The sequences found in liquid droplets also consists of aromatic residues, particularly tyrosine and phenylalanine, where the residues are arranged in such a fashion that the polypeptide backbone can enable charge-charge, charge-π, and π-π stacking interactions.[203] Atomic simulation studies on low complexity domains (LCD) suggest that the degree of charge distribution of polar and charged residues leads to a conflicting interaction between intrachain electrostatic interactions and chain-solvent interaction.[204] LLPS is involved in many biological functions like RNA stability, regulating heterogeneous ribonucleoprotein assembly, and disruption of phase separation pathway is linked with several diseases.[11,205,206] The phase separation property of IDPs is associated with the formation of stress granules. Also, LCD mediated LLPS contributes to the amyloid transition RNA binding proteins and develop as amyloid-like aggregates.

Fawzi and coworkers have demonstrated that the hydrophobic subregion cooperatively and transiently adopts a helix conformation and mediates TDP-43 phase separation. At very low salt concentration, the LCD of TDP-43 remains monomeric but instantly undergoes phase separation at physiological salt concentrations. The addition of a meager amount of RNA results in phase-separated TDP-43 in the absence of RRM domains stabilized by the saturation of protein-RNA interactions suggesting the RRM domain-mediated RNA interactions in TDP-43 and other RNA binding proteins may play a critical role in RNA-mediated phase separation.

Interestingly, the ALS mutations like A321G, Q331K, and M337V and helix breaking A326P substitution disrupt the phase separation suggesting the importance of the helix forming hydrophobic subregion in LLPS.[138] Mutation studies suggest that the tryptophan may serve as a critical residue in the TDP-43 LLPS pathway, and a mutant, W334G, can disrupt LLPS without affecting the α-helical propensity.[207,208] The N-terminal dependent dimerization of TDP-43 could physically bring together the C-terminal domains close and increase the effective local concentration and thereby the possibility of transient interaction between the C-terminal domains. The ALS mutation on a uniquely well-conserved hydrophobic segment of the C-terminal domain could alter the inherent TDP-43 self-interactions and direct the TDP-43 amyloid-like aggregation.[138207,208] Also, a phosphomimetic substitution at S48 is shown to disrupt liquid-liquid phase separation LLPS of TDP -43 polymeric assembly *in vitro* and disrupts RNA splicing activity.[113]

The downregulation of tankyrase, a poly (ADP-ribose) (PAR) polymerase, is documented to disrupt the accumulation of TDP-43 in the cytoplasm. TDP-43 is modified by

the non-covalent binding of PAR to the PAR-binding motifs present in TDP-43 NLS and promotes LLPS of TDP-43 in-vitro. In mammalian cells and neurons, the PAR binding facilitates TDP-43 accumulation in stress granules. The storage of TDP-43 in stress granules is likely to protect TDP-43 initially from disease-associated phosphorylation and aggregation but on persisting stress condition leads to phosphorylation and aggregation of TDP-43 in inclusions.[181] WT RNP granules demonstrate distinct biophysical properties depending on axon location, and ALS-linked mutations like M337V and G298S interrupt TDP-43 RNP granule motility and enhances the granule viscosity.[209] Intriguingly, the phase separation of RNA binding protein depends on RNA/protein ratios, where low ratios promote phase separation, and high ratios prevent droplet formation *in vitro*. Also, low levels of nuclear RNA levels and mutations on RNA binding proteins promote phase separation and the accumulation of proteins in cytotoxic inclusion bodies, thereby suggesting an RNA binding dependent aberrant phase transitions in TDP-43.[209] Strikingly, pathological mutations on the low complexity domains on other RNA binding proteins like FUS, hnRNP A1, and hnRNP A2 are associated with phase-separated granules and ALS inclusions. [14,206,209]

In summary, the disease mutation disrupts the phase transition forming functional granules, leading to the misregulation of RNP granules related to RNA metabolism and associated toxicity. A deep understanding of the TDP-43 phase transition and formation of RNP granules could enable therapeutic intervention in amyloidosis by targeting phase transition behavior.

### 1.5.6 TDP-43 polymorphs, isomers, and associated disease subtypes

Ubiquitinated, phosphorylated, and proteolytically cleaved TDP-43 C-terminal fragment characterizes most instances of ALS and FTLD.[91,210] Interestingly, a distinct polymorph of the amyloid proteins is responsible for particular pathology subtypes and displays variation on clinical and neuropathological features.[211] Accumulating data on a-$\beta$, tau, and $\alpha$-synuclein protein aggregates displaying differential polymorphs associated with specific disease subtypes reinforces the notion of prion strains.[212–215]

TDP-43 shows a variation in histological features corresponding to specific clinical observation of primary nonfluent aphasia, behavior variant frontotemporal, and semantic dementia. Based on the morphology and distribution of TDP-43 amyloid aggregates in the brain area, FTLD-TDP is classified into three categories, namely FTLD-TDP type A, B, and C. Subtype A is featured by the occurrence of TDP-43 aggregates in neuronal cytoplasmic inclusion bodies, dystrophic neurites, and intranuclear inclusions in upper cortical layers but not contributing motor neuron disease (MND). These subtypes can be caused by sporadic or familial mutations on granulin (GRN) and chromosome 9 open reading frame 72 (C9orf72) genes. FTLD-TDP type B is characterized by the occurrence of TDP-43 aggregates in neuronal cytoplasmic inclusion bodies distributed in all cortical layers and displays a combination of FTLD and MND. Subtype B is also associated with mutations in the C9orf72 gene. Lastly, FTLD-TDP type C shows long and dense dystrophic neurites and no neuronal cytoplasmic inclusion bodies, predominantly observed in superficial cortical layers. These disease subtypes are

sporadic and involved in semantic dementia exclusive of MND. However, recent studies have identified unprecedented clinical and pathological features of TDP-43 polymorphs and distribution and are not categorized. [216–219]

By employing a combination of sarkospin, biochemical assay, and imaging methods, it has been demonstrated that TDP-43 fibrils extracted from different disease subtypes display distinct morphological and biochemical properties. Specific FTLD types can be consistently distinguished by the intrinsic density of the isolated TDP-43 aggregates. Also, subtype C displays very high-density protease-resistant TDP-43 totals with low-level ubiquitination exclusively associated with sporadic type dementia. Strikingly, the subtype C and A diseases, TDP-43 show variation in seeding and neurotoxicity. The subtype A TDP-43 aggregates are fast-progressing and toxic, where subtype C TDP-43 aggregates are slowly progressing and relatively non-toxic under similar conditions with equivalent phosphorylation.[219] Similarly, four major ALS subtypes are associated with accumulating proteinaceous aggregates, including TDP-43 and several other proteins like superoxide dismutase 1 (SOD1) and other RNA/DNA binding proteins.[220]

The TARDBP gene encoding TDP-43 consists of six exons, and evidences suggest that there exist 19 different alternative transcripts formed by various combinations of splicing events. Interestingly, cytoplasmic TDP-isoforms (sTDP-43) interferers RNA splicing and causes upregulation of endogenous *TARDBP* suggesting loss of autoregulatory capacity of TDP-43 may contribute to TDP-43 Proteinopathy.[221] Alternatively, spliced TDP-43 isoforms sTDP43-1 and sTDP43-2forms inclusion bodies that sequester with WT TDP-43 and display neuronal toxicity when overexpressed in mammalian neurons. The occurrence of TDP-43 isoform aggregates in vulnerable motor neurons, glia of ALS patients implicates a crucial role of sTDP43 in ALS. [222]

A deep insight among the various genetic, clinical and pathological subtypes of FTD and ALS associated with different TDP-43 and its isoforms is essential to mechanistically link other neurodegenerative disorders and crucial for developing targeted therapeutic strategies for these conditions.

**1.6 Functional amyloids-NWD2**

This section will be reviewing the available bibliography on the HET-s/NWD2 paradigm based on available homology modeling data, structural and biological studies.

Although amyloids are associated with the etiology of neurogenerative diseases (Parkinson's, Alzheimer's), the self-templating ability of the amyloids leads to prion-like behavior.[223–225] This unique prion-like action of amyloids is advantageous in some biological systems and are called functional amyloids.[226] The amyloids involved in biological functions include structure, storage, information, loss of function, and gain of function.[227] Some amyloids are involved in structural operations like Pmel17 that templates small chemicals, and bacterial curli that form biofilms.[228,229] Some proteins/ peptides are stored as amyloid fibrils to be inert from their surroundings or conversely, e.g., some hormones store mammalian secretory granules.[230] The HET-s prion in fungus *Podospora anserina* is involved

in its immune system as a signaling mechanism.[231,232] The yeast enzyme Cdc19 and the termination factor sup35 loses its activity in its amyloid form.[233] The RIP1/RIP3 amyloid system forms hetero amyloids and involves in the cell signaling pathway.[234] Functional amyloids are widespread in diverse organisms, including bacteria.[235]

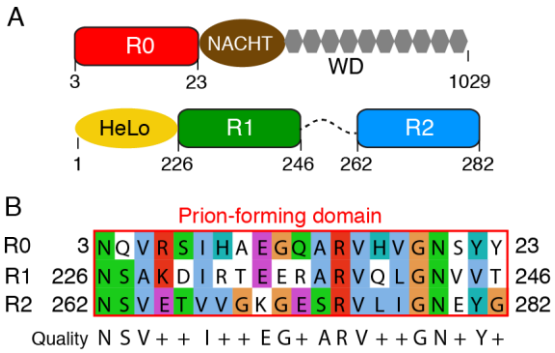



*Figure 1.5 Sequences and homology between NWD2 (3-23), called here as R0, and HET-s (226-287), called here as R1-R2. A) Domain organization of NWD2 and HET-s. B) Sequence alignment of R0, R1, and R2. The prion-forming domains are highlighted with a red box.*

The fungal prion protein HET-s is a well-studied prion system where the atomic resolution structure of prion propagating β-solenoid fold is available.[34,236] The HET-s prion ([Het-s]) is involved in the heterokaryon incompatibility in the fungus *Podospora anserina*. Incompatibility is a cell death mechanism that avoids the fusion of unlike strains.[237,238] This mechanism limits the fusion with genetically unidentical partners to prevent the passage of harmful plasmids, viral particles, and conspecific parasitism.[237,238] The *het-s* locus exists as two agonistic alleles, het-s, and het-S, which encodes the proteins HET-s and HET-S, and these proteins are 95% identical.[239] The Het-s/HET-S proteins share a two-domain structure containing a C-terminal prion forming domain (PFD, residue 218 to 289) necessary and sufficient for prion propagation, the N-terminal pore-forming toxin domain (1-227) termed HeLo domain.[240,241] Due to specific mutations on the HeLo domain, HET-s lacks the pore-forming activity, which helps them to propagate as a prion.[240,242]

The PFD of HET-S/s are functionally interchangeable and contains two pseudo-repeats of 21 amino acids (called repeats R1 and R2), which adopts a left-handed β-solenoid fold in the prion form. Each pseudo-repeat corresponds to one rung of a two-layer β-solenoid structure connected by a flexible loop of 15 amino acids. Each repeat comprises of four β-strands where the first three β-strands form the triangular hydrophobic core, and the aromatic loop at the C-terminal end of PFD, along with the third and fourth β-strands, forms a semi-hydrophobic pocket. The HET-s prion fold is stabilized by two asparagine ladders (N226/N262 and N243/N279), three salt bridges (K229/E265, E234/K270, and R236/E272), and polar interaction within the hydrophobic core (S273-T233). The cell death is induced when the [Het-s] prion templates the amyloid transition of HET-S PFD into the β-solenoid fold.[34,236] This switch of HET-S PFD from a soluble state to an amyloid state, in turn, refolds

the HeLo domain and triggers the pore-forming activity by exposing an N-terminal transmembrane helix and relocates it to the cell membrane and induces toxicity.[243]

In *P. anserina*, the gene immediately adjacent to *het-S* encodes a protein called NWD2, a Nod-like receptor (NLR), which possesses an N-terminal motif (NWD2 (3-23)), also called here as repeat R0. This R0 repeat motif is homologous to the elementary repeat unit of the β-solenoid motif of HET-s PFD (R1, R2) (**Figure 1.5**). The NLRs are signal-transducing ATPases which belong to the family of proteins called STAND (Signal Transduction ATPase with Numerous Domains). STAND proteins regulate diverse biological activities in bacteria and eukaryotes, involving in gene expression, innate immune responses, and cell death. STAND proteins generally consist of a C-terminal ligand-binding domain, a central NOD domain, and an N-terminal effector domain.[244–247] Upon ligand binding, it undergoes oligomerization and downstream signaling pathways via the N-terminal domain.[248]

Similarly, the NWD2 is composed of a C-terminal WD repeat domain, a central NACHT domain, and an N-terminal motif (R0) homologous to HET-S/s PFD. The R0-motif is necessary and sufficient for prion-inducing activity *in vivo* and the shortest sequence to exhibit prion activity. Homology modeling and SSNMR analysis of R0 peptide show that this motif can adopt a HET-S/s like fold.[249–251] NWD2 is one of many NLRs in the fungal genome, activated by ligand-induced oligomerization. It was hypothesized that based on homology and other biological analysis, that the binding of a putative ligand on the WD repeats and oligomerization of NWD2 through its NACHT domain leads to the spatial clustering of the N-terminal motif (R0), adopting a β-solenoid fold, which in turn templates HET-S to transform to the prion state that triggers HeLo domain pore-forming activity.[249,250,252] Recently, *Cai et al*. have shown that the R0 motif of NWD2 and HET-s PFD can functionally interchange with Pyrin signaling domains in human NLR (NLRP3) and its cognate effector protein (ASC), providing direct evidence of NWD2 mediated prion transition of HET-S/s PFD.[251]

### 1.7 Aim of the study

Aggregation of amyloid proteins are involved in many neurodegenerative diseases and also in biological functional aspects. It is essential to obtain high-resolution amyloid structures, to gain insights into the mechanism of protein aggregation, and to develop therapeutic strategies. Unfortunately, conventional structure determination tools like X-ray crystallography and solution NMR cannot be used for amyloids as they form non-crystalline, insoluble aggregates. The SSNMR spectroscopy with magic angle spinning (MAS) can successfully be employed to obtain amyloid structures in atomic resolution at near-physiological conditions.[34,253–255] My thesis is constructed of three parts. The first part is dedicated to the structural studies of pathological amyloid formed by TDP-43 concerned with the histopathology of ALS. The second part is dedicated to the molecular mechanism of signal transduction via functional amyloid proteins NWD2/HET-s involved in programmed cell death mechanisms in fungi using SSNMR and other biophysical techniques.

Although the role of TDP-43 in ALS and associated diseases have been studied for many years, the mechanism of aggregation remains elusive in the absence of a high-resolution structure of TDP-43 in its fibrillar state. To date, only a handful of amyloid

structures of TDP-43 peptides are reported, and the structure of the full-length protein in the amyloid state remains unknown. Furthermore, an understanding of the amyloid fold adopted by TDP43 isoforms and TDP-43 polymorphs is essential to characterize ALS subtypes. To achieve these goals, I employed a "divide and conquer" approach where I created several TDP-43 constructs, including the full-length protein and its C-terminal fragments, namely, TDP-16, TDP-13, TDP-11, and TDP-10 from the intrinsically disordered region to locate the amyloid core and obtain an insight about the different polymorphs formed by the C-terminal fragments. Also, I used two TDP-43 isoforms, namely sTDP43-1 and sTDP43-2, to examine the amyloid fold they adopt in their fibrillar state. A combination of biophysical techniques like electron microscopy, X-ray fiber diffraction, FT-IR analysis, and SSNMR spectroscopy to examine the molecular organization of different TDP constructs. A structural basis of TDP-43 polymorphs will shed new insights into the molecular mechanism of TDP-43 mediated neurodegeneration and design novel methods to increase amyloid therapeutic efficiency, especially in the treatment of neurodegenerative diseases such as ALS and FTLD. These studies will be described in **chapter 3**.

Secondly, I examined the role of β-solenoid fold in fungal Nod-like receptor (NLR) function and signal transduction via a heterotypic amyloid interface in immunity-related cell fate pathways. The NLRs are associated with host defense and programmed cell death (PCD) in plants, animals, and fungi. Upon activation, the N-terminal motif (R0) can adopt the β-solenoid fold, template the amyloid transition of the HET-S prion-forming domain, and trigger the downstream cell death pathway. To understand the underlying mechanism of amyloid based signaling in fungal programmed cell death, we used SSNMR to obtain the atomic resolution structure of the β-solenoid fold adopted by the functional amyloid NWD2 (R0) and the heterotypic amyloid interface formed between R0 and HET-S. This study could explain a universal mechanism of amyloid based NLR signal transduction. These studies will be described in **chapter 4**.

Thirdly, I examined the amyloid-specific dye Thioflavin T (ThT) binding mode on amyloids, which is used for an accurate and conclusive diagnosis of amyloids on model amyloid system HET-s and R0-R0. To identify the most probable binding epitopes on amyloids in atomic resolution, we use chemical shift perturbations as detected by SSNMR. A deep insight into ThT binding modes on amyloids will provide a basis for the efficient search and design of novel molecules crucial for the conclusive diagnosis of amyloid-related diseases or to develop novel therapeutical advances. These studies will be described in **chapter 5**.

# Chapter 2 SSNMR FOR BIOLOGICAL RESEARCH

This chapter attempts to explain the basic SSNMR concepts and the protocol to elucidate structural models of amyloid fibrils from SSNMR data combined with other biophysical techniques.

## 2.1 Interactions in SSNMR

The magnetic and electric properties of nuclear spins are utilized in NMR spectroscopy. Nuclear spins can be considered as tiny magnets with an intrinsic magnetic moment (I) originating from the currents of charged particles and the quark magnetic moments inside the nucleus. In general, a spin 1/2 nucleus in a magnetic field ($B_0$) experiences four significant interactions expressed as a separate Hamiltonian ($\hat{H}$). The sum of all Hamiltonians constitutes the spin interactions during the NMR experiments.

$$\hat{H} = \hat{H}_Z + \hat{H}_{CSA} + \hat{H}_J + \hat{H}_D,$$

Where $\hat{H}_Z$ is the Zeeman Hamiltonian, which describes the interaction between the magnetic moment of spins(I) and the external magnetic field($B_0$) and directly proportional to the gyromagnetic ratio ($\gamma$).

$\hat{H}_{CSA}$ is the full chemical-shift Hamiltonian, representing the interaction between the nucleus and the external magnetic field, as mediated by the environment of the nucleus. This quantity is observed as a shift in the resonance frequency, considering isotropic and anisotropic contributions. The chemical shielding directly provides information about the local chemical environment of nuclear spin, local structure, and molecular coordination. In the case of polypeptides, the chemical shifts provide the secondary structure information.

$\hat{H}_J$ is the spin-spin interaction between the same (homonuclear) or different (heteronuclear) nuclei mediated by bonding electrons called J-coupling or scalar coupling. This coupling interaction is isotropic and reports about connectivities and dihedral angles.

$\hat{H}_D$ is the spin-spin interaction (homo- or hetero nuclei) through space and is anisotropic. Dipolar coupling directly reports distance information between interacting spins. The J coupling and dipolar coupling are exploited in polarization transfers between nuclei to obtain structural information about the bonding(J) and proximity(dipolar) of the nuclei.

The chemical shift anisotropy ($\hat{H}_{CS}$) and dipolar coupling ($\hat{H}_D$) Hamiltonians are anisotropic interactions, meaning they depend on the spatial orientation of the spins to the applied magnetic field. Theoretically, this relation can be expressed as $3\cos^2\theta - 1$, where the $\theta$ is the angle between the Z principal axis of the interaction and the external applied magnetic field. These anisotropic interactions lead to broad signals in SSNMR called the ''powder pattern'' due to the random orientation of spins with the external magnetic field. This line broadening is not observed in solution-state NMR due to the rapid tumbling of molecules, which averages any anisotropic interactions. However, these anisotropic interactions can be

averaged out in SSNMR by spinning the sample at an angle of $\theta^{=}$ 54.7° with respect to the magnetic field. This technique historically is called magic-angle spinning (MAS).[256,257] The schematic representation of the rotor alignment is shown in **Figure 2.1A.**

## 2.2 SSNMR for biological research

In order to get a response from magnetically active nuclei (see 2.3. discussions on which biological nuclei are active for NMR), certain timely placed radio frequency (RF) pulses are being applied. The response signal is then acquired, stored, and converted from the time domain to the frequency domain using the Fourier-transformation technique.

Advances in solid-state NMR (SSNMR) spectroscopy have been employed successfully in determining structural and dynamic information of various supramolecular assemblies. SSNMR with MAS (MAS-SSNMR) has been used as a powerful technique to obtain the atomic-resolution structures, study interactions, and dynamics of various biomolecular assemblies like amyloids, microcrystalline proteins, viral capsids, bacterial pilus, membrane protein, large molecular complexes, etc.[34,157,258–276] It is often impossible to grow macromolecular protein assemblies like amyloids into single crystals for X-ray diffraction studies or carry out structural studies analysis by solution-state NMR due to their large molecular size and insolubility. SSNMR techniques do not possess any of these limitations and can be used as a powerful tool to obtain atomic-level information of amyloids. Moreover, recent technical advances in sample preparation, access to high magnetic fields (up to 1.2 GHz in 2020), probe technology, and methodology developments have made MAS-SSNMR a robust technique along with cryo-EM to explore the atomistic details of amyloids in biologically relevant conditions.

Interestingly, 3D structural models of over 100 proteins and polypeptides solved by SSNMR have been deposited in the protein data bank (PDB). Several amyloid fibril structures at atomic resolution have been solved based on SSNMR data, including HET-s(218-289), α-synuclein, and Aβ(1-42), and β-endorphin.[34,236,254,276–281] Compared to other structure determination methods, SSNMR also allows us to obtain the site-specific protein dynamics on the various time scales.[282]

## 2.3 Biological sample preparation for SSNMR studies

Due to the low natural abundance of $^{13}C$ (1.1%) and $^{15}N$ (0.37%) nuclei ($^{12}C$ is not magnetically active and $^{14}N$ is quadrupole nuclei) in biomolecules, the SSNMR studies of organic biomolecules usually require enrichment of $^{13}C$ and $^{15}N$ isotopes. The samples for biomolecular SSNMR are often produced using recombinant expression techniques, and cell-free peptide synthesis can be applied. [283–285] In this thesis, only the recombinant expression will be reviewed. This protein preparation technique is based on the expression in minimal media, which is supplied with isotopically labeled nutrients. Additional labeling schemes based on heterologous expression are also employed for structural analysis. Sparse labeling schemes with selectively $^{13}C$-labeled samples can be produced by supplying selectively $^{13}C$-labeled sources such as (1,3-$^{13}C$)- and (2-$^{13}C$)-glycerol/glucose to achieve spin dilution and spectral resolution. Mixed labeled (each monomer exclusively of either $^{15}N$ or $^{13}C$

labeled)samples are used to distinguish inter and intramolecular distance restraints[286–291]. The amyloid proteins are generally observed in bacterial inclusion bodies, later on, extracted in buffers containing chaotropic agents like urea, guanidium hydrochloride, or guanidinium thiocyanate. The protein can be purified using various chromatographic techniques based on the protein construct design.

**2.4 Assembling conditions**

The success in 3D structure determination of amyloids by SSNMR largely depends on the availability of homogeneous samples. Typically, the high sample homogeneity can be achieved by removing the denaturing agents from the sample via dialysis or size exclusion columns. Different aggregation conditions are chosen for different amyloid fibrils by providing buffer conditions with the right molecular environments (pH, salts, metal ions, redox agents, etc.) are essential to meet the requirement of homogenous fibrils. After fibrillization, care should be taken to tightly pack the amyloid samples in the SSNMR rotors considering the maximum filling efficiency and sample hydration for long-term stability. As the SSNMR is very sensitive to atomic level structural order, even slight differences in local conformation can be essential, and closely associated chemical shifts result in broad signals, adversely affects the SSNMR data analysis.[292–294] It is not uncommon to observe different polymorphs with broad SSNMR signals of pathological amyloids owing to their structural plasticity, repetitive motifs with similar chemical shifts, or their structural dynamics. Attaining consistent batches of homogenous amyloid fibrils is challenging but is necessary for the more feasible data analysis. It is also very relevant from a pathological perspective as differential amyloid polymorphs display differential pathological phenotypes.[78,295–300] However, as the amyloids follow a nucleation polymerization mechanism of aggregation, a consistent set of homogenous amyloid samples can sometimes be achieved by repeated steps of "seeding" (supplying competent nuclei able to polymerize native proteins to amyloid fold) and sonication of the polymerization mixture.[280,301–305]

**2.5 SSNMR experimental concepts**

In this session, the main SSNMR experimental approaches conducted in this thesis will be reviewed in short.

The main bottleneck of SSNMR spectroscopy is its relatively low sensitivity, especially for $^{13}C$ and $^{15}N$ nuclei. This could be improved by applying different magnetization transfer techniques from the $^{1}H$ nucleus which is more sensitive. The most common magnetization transfer technique is the cross-polarization (CP) method. This approach is based on polarization transfer from a nucleus of high $\gamma$ value (usually $^{1}H$) to a nucleus to a lower $\gamma$ nucleus X (X= $^{13}C$ or $^{15}N$) via heteronuclear dipolar coupling interactions by applying simultaneous spin locks at both channels. The two contact pulses amplitudes in the CP experiment are carefully set to achieve the Hartmann–Hahn matching condition where the energy difference is equal between the respective rotating frame spin states of $^{1}H$ and X spins. However, at MAS conditions, achieving cross-polarization is tricky as the dipolar coupling has

no isotropic (constant) term, unlike CSA. So, the cross-polarization is achieved by adjusting the amplitudes of the $^{1}H$ ($\omega_H$) and X ($\omega_x$) contact pulse frequencies considering the MAS frequency ($\omega_r$) to meet the condition as $\omega_H$-$\omega_x$=$\pm\omega_r$ or $\pm 2\omega_r$. The CP experiment typically allows us to detect the rigid (dynamics around ms scale) parts of the molecules. For a theoretical explanation of CP, the reader is kindly directed to chapter 2 of "Solid-State NMR Spectroscopy Principles and Applications" by Melinda J. Duer.[306] A simple CP based pulse sequence with $^{13}C$ detection is shown in **Figure 2.1B.** An opening 90° pulse creates $^{1}H$ magnetization along the XY plane, followed by an on-resonance contact pulse is applied. Based on the choice system, the contact time of the pulse ($\tau_{CP}$) can be varied from µs to several milliseconds. In this step, the polarization is transferred from $^{1}H$ to X (X= $^{13}C$ or $^{15}N$), and the efficiency of this transfer depends on the magnitude of the dipolar interaction between them. The fast-longitudinal relaxation ($T_1$) of $^{1}H$ helps to minimize the duration of the pulse sequences. By that, the experimental time can be significantly reduced. The CP transfer step is followed by acquisition on the X channel, applying a robust heteronuclear decoupling scheme (TPPM, SPINAL, XiX, etc.) to improve the line-widths and sensitivity further.[307–309]

In order to initiate the homonuclear $^{13}C$-$^{13}C$ and $^{15}N$-$^{15}N$ magnetization transfer, several recoupling schemes such are RFDR, DREAM, PDSD, DARR, and PAR are often used in biomolecular SSNMR. [290,291,310–314] Here in this thesis, the PDSD based recoupling experiment with various mixed times are used to probe the structural information.

Suppose the dipolar couplings are averaged out, which is typical for species with high flexibility. In that case, J coupling based transfer experiments can be applied to increase the sensitivity for the heteronuclei. In this technique, the polarization is transferred from the nucleus with a high $\gamma$ value spin (usually $^{1}H$) to a spin with lower $\gamma$ value X (X= $^{13}C$ or $^{15}N$). This transfer relies on the $^{1}H$-X scalar couplings ($J_{HX}$ coupling. The method historically is called Insensitive Nuclei Enhanced by Polarization Transfer or INEPT.[315,316] The INEPT scheme in SSNMR is similar to the one used in solution-NMR. The scalar coupling is independent of directionality and molecular dynamics. However, the efficiency of polarization transfer is lost due to the rapid transverse relaxation ($T_2$) of rigid parts of the macromolecular assemblies during the long spin-echo sequence and hence not detected in SSNMR INEPT experiments **(Figure 2.1C)**. The INEPT based experiments selectively detect the flexible part of the macromolecular assemblies having correlation time in the range of ns to µs. The CP based and INEPT based experiments can be used as a complementary strategy for determining rigid and flexible part of the amyloids selectively. An example of such spectra with structural information on NWD2-R2 amyloid fibrils is shown in **Figure 2.1D**.

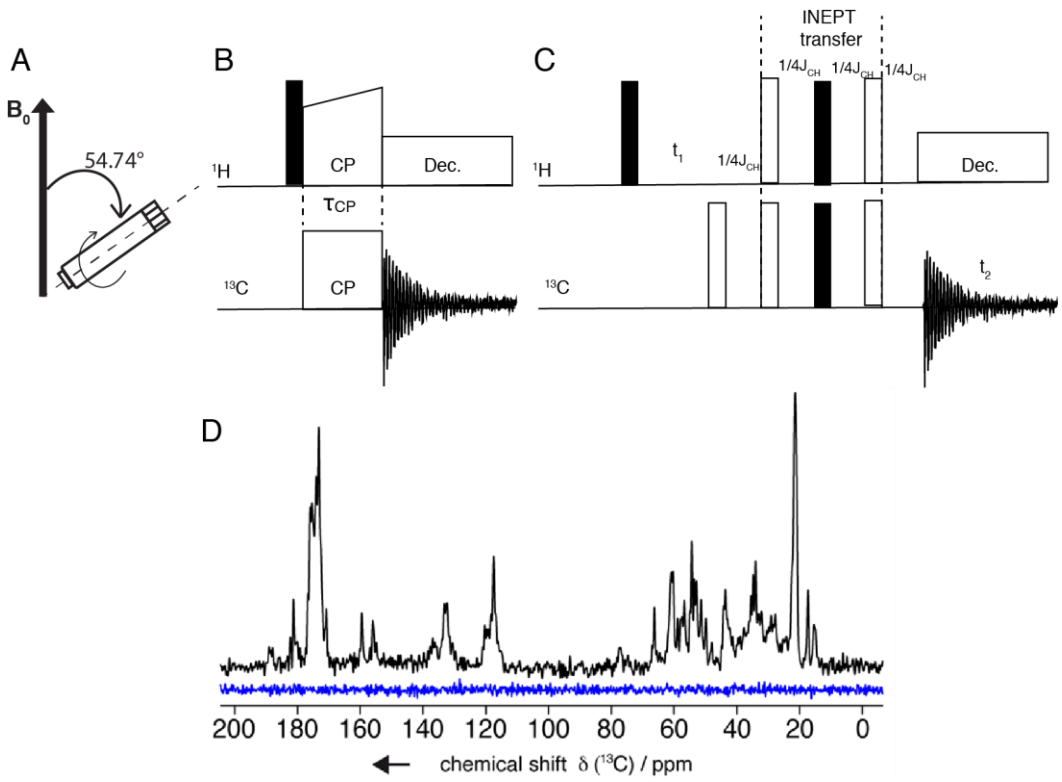



*Figure 2.1 A) Magic angle spinning (MAS) of the sample. B) The 1D $^{13}$C-detected CP pulse sequence. C) The 2D $^{1}$H-$^{13}$C INEPT pulse sequence. The black rectangle represents the 180° pulse, and an open rectangle represents the 90° pulse. D) Solid-state NMR 1D $^{13}$C- spectra on natural abundance NWD2(3–23), recorded using a CP (black) and INEPT (blue) polarization transfers. (Figure adapted from Daskalov et al. [317])*

In order to resolve peaks in large molecular systems, multi-dimensional NMR experiments are required. In this thesis, we mostly employed a 2D $^{13}$C-$^{13}$C correlation with the PDSD scheme for the structural characterization of amyloid fibrils. The experiment starts with using a CP transfer from $^{1}$H to $^{13}$C nucleus, followed by an indirect evolution (t1) of the resonance frequency of the $^{13}$C spins. This evolution is recorded for several consecutive experiments by incrementing the points in the indirect dimension. Subsequently, PDSD recoupling scheme is applied to transfer magnetization between $^{13}$C nuclei, followed by a t2 acquisition period. $^{1}$H decoupling schemes are applied during t1 and t2 periods to average out more heteronuclear dipolar coupling **(Figure 2.2A)**. With the additional transfers, this overall scheme can be extended to any number of dimensions and nuclei. The resulting 2D spectra of the off-diagonal peaks correspond to two nuclei frequencies, close in space (~7Å) **(Figure 2.2B)**.

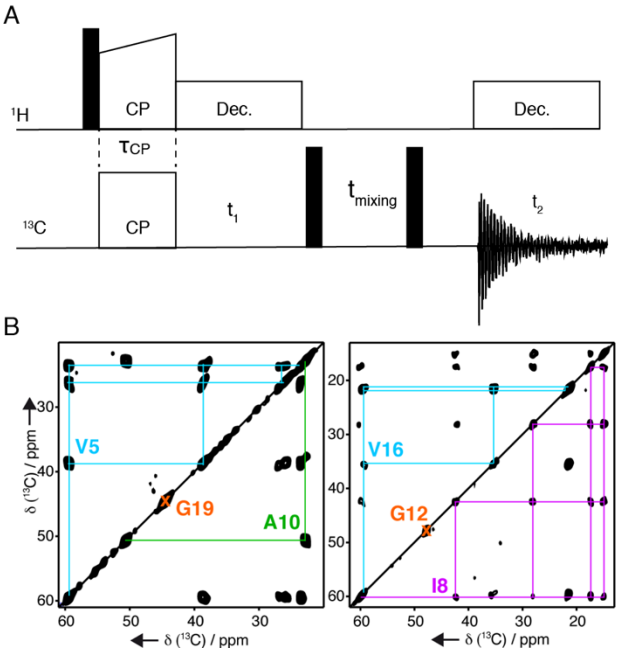



*Figure 2.2 A) The 2D $^{13}C$-$^{13}C$ PDSD pulse sequence with the mixing time $t_{mix}$ used for the $^{13}C$-$^{13}C$ transfer, here we used the PDSD scheme. The black rectangle represents the 90° pulse, and an open rectangle represents the 180° pulse. B) Two-dimensional solid-state NMR $^{13}C$-$^{13}C$ correlation spectra of NWD2(3–23) fibrils, isotopically $^{13}C$ labeled on Val5, Ala10, Gly19 (left) and Ile8, Gly12, Val16 (right), intra-residue correlations are shown. (Figure adapted from Daskalov et al..[317])*

**2.6 SSNMR data analysis**

The following steps can be followed for the SSNMR data analysis of amyloid fibrils. A detailed setup for SSNMR experiments is reviewed in *Loquet et al*. with visual communications.[318]

1. A 1D $^{1}H$ spectrum is recorded under MAS conditions (generally 16 scans) to identify the internal reference compound signal (Sodium trimethylsilylpropanesulfonate, DSS) to calibrate the $^{1}H$ (set to 0 ppm). Based on the $^{1}H$ resonance, the $^{13}C$ and $^{15}N$ chemical shifts are calibrated following the IUPAC recommendations.[319] The relative shift of the bulk water $^{1}H$ resonance δ(H2O) to the DSS signal reveals the sample's temperature using the equation δ(H2O) = 7.83- T/96.9 with a precision of 1 - 2K.[320,321]
2. A 1D $^{1}H$-$^{13}C$ CP spectrum is recorded to estimate the rigid parts of the protein self-assembly.
3. A 1D $^{1}H$-$^{13}C$ INEPT spectrum is recorded to estimate the highly flexible parts of the protein self-assembly.
4. A two-dimensional (2D) $^{1}H$-$^{13}C$ INEPT experiment is carried out to probe highly mobile parts of the protein self-assembly. Generally, the spectra show peaks originating from 1 bond $J_{CH}$ couplings. Furthermore, a TOtal through-Bond correlation SpectroscopY (TOBSY) transfer scheme is used for longer polarization transfer based on the homonuclear J-couplings and suppressing the dipolar coupling.[322,323] The chemical shifts obtained from such spectra generally correspond to that of random coil

conformation, which can be readily checked on the Biological Magnetic Resonance Data Bank(BMRB) database(https://bmrb.io/ref_info/stats.php?restype=aa&set=full) or *Wang et al*. and identify the flexible segments located in the protein sequence.[324,325]

### 1.1.1. Conformational analysis and distance restraints collection

1. A 2D $^{13}C$-$^{13}C$ PDSD experiment with a short mixing time (~50 ms) on uniformly $^{13}C$ labeled samples can be used to identify the intra-residue $^{13}C$-$^{13}C$ correlations, including side chains. Either CcpNmr analysis or SPARKY software can be chosen for NMR data analysis.[326,327]
2. Further, A 2D $^{13}C$-$^{13}C$ PDSD experiment with intermediate mixing time (~150-200 ms) can be used to identify sequential $^{13}C$-$^{13}C$ correlations by overlaying on the short mixing time PDSD spectrum. The additional peaks are visible in the intermediate-mixing PDSD spectrum resulting from sequential (residue i - residue i±1) correlation. However, the PDSD experiment with intermediate mixing time occasionally delivers i±2 and i±3 or even i>4 residue correlations subject to the rigidity and proximity based on the structure.
3. Additional experiments like 2D $N_iCA_i$, $N_iCO_i$, $N_i$ $(CA_i)CB_i$, and $N(CA_i)CO_i$ spectra can be used to avoid ambiguity in $^{13}C$-$^{13}C$ PDSD assignments with the additional $^{15}N$ dimension in intra residue $^{15}N$-$^{13}C$-$^{13}C$ connectivity. Additionally $N_i(CO_{i-1})CA_{i-1}$ and $N_i(CO_{i-1})CX_{i-1}$ experiments can be used for (i) to residue (i-1) residue correlations.
4. Once the amino acid sequence is identified corresponding to the rigid part of the amyloid fibrils, then their secondary structure can be calculated by comparing with chemical shifts of random coil conformation using the equation, ΔδCα–ΔδCβ. Positive and negative values correspond to α-helical and β-strand conformation, respectively.[325]
5. Further, distance restraints can be obtained using uniformly labeled $^{13}C$, $^{15}N$ samples, and sparse labeling schemes. Samples produced with selectively $^{13}C$-labeled sources such as (1,3-$^{13}C$)- and (2-$^{13}C$)-glycerol/glucose are used to achieve spin dilution and spectral resolution owing to the average labeling of amino acids synthesized via the citric acid cycle.[286–289,328–331] To obtain long-range molecular distance restraints ($^{13}C$-$^{13}C$ correlations contain i to >i+4 contacts), PDSD spectra with long mixing time (400-500ms) are acquired from several samples, including $^{13}C$/$^{15}N$-uniformly enriched (U-[$^{13}C$-$^{15}N$]) or sparsely $^{13}C$ enriched (1,3-$^{13}C$- and 2-$^{13}C$- glycerol/glucose labeled) of amyloid fibril samples. Obtaining many such long-range distance restraints is essential during the structural calculation step.[286–289,328–331] These distance restraints are separated into unambiguous and ambiguous intra/inter-molecular distances**.**
6. Mixed labeled samples are used to distinguish inter and intramolecular distance restraints with an equimolar mixture of either selectively $^{15}N$- and $^{13}C$-labeled or (1,3-$^{13}C$)- and (2-$^{13}C$)-glycerol/glucose samples using PAIN or PAR recoupling

schemes using polarization transfer trough space between a $^{15}N$ and $^{13}C$ or between $^{13}C$ depending on the labeling schemes.[286–291]

### 2.6.1 Structural calculation from SSNMR distance restraints

The backbone chemical shifts are used to predict the dihedral angle restraints on ϕ/Ψ backbone angles using softwares like TALOS+.[332] The classified distance restraints such as unambiguous and ambiguous intra/inter-molecular distances along with the protein sequence, TALOS-based dihedral angle restraints, additional information from complementary structural biology techniques like cryo-EM, Mass-per-length measurement, XRD, etc and structural similarity with published structures of homologous amyloids can be combined using dedicated softwares like ARIA, CYANA, UNIO, Rosetta or XPLOR-NIH for robust structure calculation.[333–336] Amyloids have the peculiar properties of stacking subunits stacked on top of each other stabilized by hydrogen bonds running parallel to the fibril axis. These hydrogen bond restraints between paired β-strands can be maintained during structural calculation by imposing an upper limit distance of 3.0 Å for the donor (N) and acceptor (O) and 2.0Å for amide hydrogen (HN) and acceptor. However, in case of ambiguity in establishing hydrogen bond restraints, multiple structural calculations by testing various possibilities of hydrogen-bandings and eventually the lowest energy structure(with the lowest restraint violations) are selected for analysis.[276,337] Above mentioned molecular modeling softwares can perform these rigorous structural calculations. Additionally, the CYANA software can help achieve automated structure calculation by collecting possible assignments based on matching the cross-peaks positions concerning the assigned chemical shifts with a tolerance range defined by the user.

The fibrils structures in this study are determined with ARIA2.3 in 8 cycles of structural calculations and restraint analysis, followed by a cycle of refinement in water.[333] At each iteration, 100 structures were calculated, and the ten lowest-energy structures were passed to the next iteration.[338] Sequential short and medium-range assignments ($|i-j| \leq 4$) were set manually to intra-chain restraints. When $|i-j|>4$, and in the absence of evidence whether the assignments were intra- or inter-chain correlations, the ARIA protocol was determining the intra- or inter- nature of each restraint throughout the iterations, as exemplified by **Figure 2.3**. The fibrils structures were calculated as pentamers using simulated annealing performed with CNS 1.2.[339] The ladder topology was maintained during the calculation through distance restraints, ensuring that the distance between equivalent Cα atoms in neighboring monomers is constant throughout the pentamer, without fixing a particular distance value, i.e., dm/m+1 = dm+1/m+2 = dm+2/m+3 = dm+3/m+4 = dm+4/m+5.[236,340] Additionally, an NCS (Non-crystallographic Symmetry) restraint was added to minimize the RMSD between atomic coordinates of the monomers.[340] Finally, the conformational database potential from *Kuszewski et al*..[341] was applied to the whole system to maintain the dihedral geometry, especially in poorly determined regions. Finally, before depositing the amyloid structure in the Protein Data Bank (PDB), the structure's validity is examined with standard validation

tools like PROCHECK or MolProbity. [342–344342–344] A General procedure for 3D structure calculation of amyloid fibrils from SSNMR data is displayed in Figure **2.**3.[72]

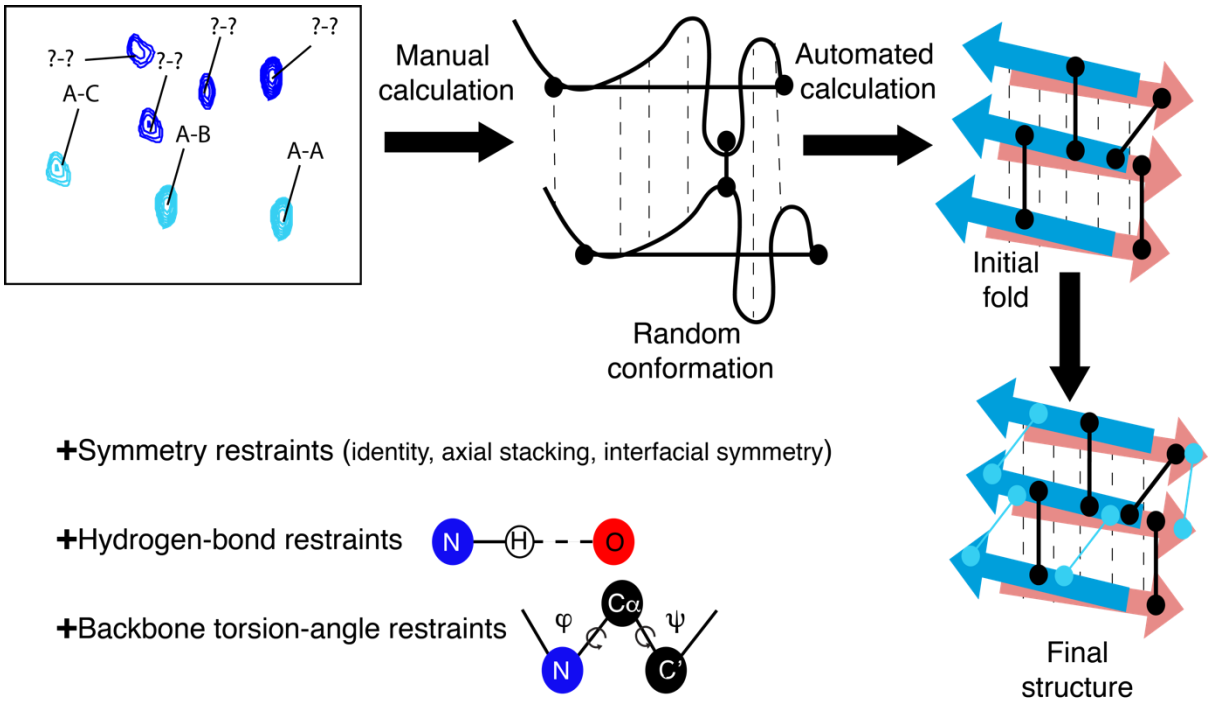



*Figure 2.3 General procedure for 3D structure calculation of amyloid fibrils from SSNMR data. First, an initial fold is calculated from the distance restraints obtained from unambiguously assigned peak lists (Manual calculation). A restraint set is then supplemented with all the identified cross-peaks obtained from different SSNMR spectra (high spectral ambiguity in general) that will automatically be assigned to yield the final structure as a high precision bundle of conformers (Automated calculation). The higher-order symmetry of amyloid fibrils arising from the symmetric arrangement of subunits is ensured by maintaining symmetry restraints during calculation (identity restraints, axial stacking, and interfacial symmetry, if suitable). Further, the hydrogen bond restraints, the calculated secondary structure from chemical shifts, the backbone dihedral angle restraints are applied to enforce an appropriate stacking and an ordered arrangement of β–strands. The figure is adapted from Loquet et al.* [72]

SSNMR has been considered an essential role in understanding the molecular mechanism underlying the folding/misfolding events, dynamics, and structural information of amyloid proteins.[345,346] The recent advances in isotopic labeling schemes, aggregation conditions for obtaining homogenous amyloid fibrils, SSNMR instrumentations (novel pulse sequences, higher magnetic field, ultra-fast MAS NMR probes, Dynamic nuclear polarization(DNP)), structure calculation softwares have made a substantial leap in establishing SSNMR as a powerful tool for the structure elucidation of macromolecular assemblies like amyloids. The unique advantage of SSNMR over other complementary techniques lies in determining the structural homogeneity or inhomogeneity (polymorphism)

at the atomic level.[70–72] Investigation of amyloid polymorphism is crucial in many proteins misfolding related neurodegenerative diseases. Moreover, the variations in polymorph are associated with the difference in disease phenotypes, including cytotoxicity.[78,295–300] Advances in applying SSNMR in early aggregation stages of pathological amyloids like oligomers or “protofibers” provides an unprecedented potential to aid the design of better therapeutic strategies. The DNP techniques for polarization enhancement and ultrafast MAS NMR probes capable of spinning >100 kHz for $^{1}$H-detected approaches are yet to unravel the unforeseen possibilities of SSNMR in amyloid structural and dynamic investigations.[347–355]

# Chapter 3 STRUCTURAL DISSECTION OF AMYLOID AGGREGATES OF TDP-43, ITS C-TERMINAL FRAGMENTS, AND SPLICE ISOFORMS

## 3.1 Introduction

Self-assembly and deposition of proteinaceous fibrillar aggregates are common features of many neurodegenerative diseases, ranging from Alzheimer's disease to prion-associated disorders. [223,356–358] These fibrillar aggregates are usually characterized by a typical X-ray diffraction pattern, encoding inter-strand and inter-sheet distances of a protein fold named cross-β structure.[22] Although the cross-β arrangement is a structural hallmark for amyloid fibrils, solid-state NMR (SSNMR) and more recently, cryo-electron microscopy (cryo-EM) has revealed that the tertiary fold (i.e., the relative position of the β-strands) and the intermolecular arrangement (i.e., subunit-subunit packing and inter-protofilament packing) can considerably differ between amyloid proteins. In their fibrillar state, amyloid proteins are insoluble, non-crystalline, and often exhibit structural polymorphism, complicating high-resolution structure determination. [34,71,253–255,302,359,360] Protein aggregation into amyloid deposits has been associated with the pathogenesis of several human diseases. The structural characterization of these proteins in their fibrillar state greatly influences disease propagation's molecular and structural basis.

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTLD) are fatal neurodegenerative diseases, sharing several clinical and pathological features.[361,362] FTLD patients present progressive atrophy in the brain's temporal and frontal lobes. At the same time, ALS is caused by a progressive deterioration of motor neurons. Patients with ALS and FTLD disorders reveal a common biochemical feature, namely cytoplasmic inclusions associated with the TAR (trans-activation response) DNA-binding protein (TDP-43) present in an aggregated state. [91,92] TDP-43 is a 414 amino acid DNA/RNA-binding nuclear protein involved in several processes; it was initially described as a cellular factor involved in HIV-1 gene expression.[102] TDP-43 is ubiquitously present in nuclei and has essential functions in transcription, pre-mRNA splicing, and translational regulation in the cytosol.[363,364] TDP-43 and other heterogeneous nuclear ribonuclear proteins (hnRNPs) are involved in liquid-liquid phase-separated assemblies forming membrane-less organelles called stress granules.[11,113,138,152,365–371] Stress granules composed of mRNAs, TDP-43, and other hnRNPs have been observed in neurons of patients with ALS, FTD, and inclusion body myopathy. Mutations in genes encoding hhRNPs have been linked to ALS and FTD disorders.[168,372,373] Although the role of TDP-43 in regulating the assembly process of ALS- and FTD-related stress granules is considered the critical factor, the structural basis of its aggregation remains obscure in the absence of a high-resolution structure of TDP-43 in its fibrillar state.

The protein consists of an N-terminal domain (NTD), two RNA Recognition Motif domains (RRM1 and RRM2), and a disordered C-terminal (CTD) region (**Figure 3.1**), the full-length protein being inherently prone to aggregate and to cross-seed Aβ oligomers.[146,374,375] The NTD is folded and essential to fulfilling native functional processes and necessary for

pathological aggregation.[106,107,109,114,120,376,377] RRM1 and RRM2 domains are also folded and bind UG-rich segments of RNA, and it has been proposed that RRM2 may contribute to the pathological misfolding.[104,136,378–380] The pathological form of TDP-43 in ALS and FTLD is hyperphosphorylated, ubiquitinated, and proteolytically cleaved into C-terminal fragments.[91,183] The C-terminal domain (CTD) is amyloidogenic and, from a structural point of view, a predominantly disordered protein segment, interacting with other members of the heterogeneous nuclear ribonuclear proteins (hnRNP) family.[138,139,149,381] Interestingly, most of the missense mutations associated with ALS in TDP-43 are located at the CTD. Some of these mutations lead to more toxic species or various levels of aggregation capability of the protein.[374,382–384] The CTD is an N/Q/S/G-rich protein segment and is seen as a low-complexity domain (LCD, i.e., protein segments that contain repeats of short amino acid motifs). Several short peptide segments have been investigated to derive structural models, e.g., the 341-357 peptide from the Q-N rich segment by various biophysical techniques, six peptide segments from the CTD using X-ray crystallography, and the RRM2 polypeptide segment 247-257 by micro-electron diffraction, cryo-EM, and X-ray diffraction.[131,152,153] Recently, cryo-EM structures of the amyloid fibrils formed from two segments of the C-terminal domain termed SegA (residues 311-360) and SegB (residues 286-331) were reported by Eisenberg and coworkers.[155] later, a solid-state NMR study suggests a helix to the b-sheet helix-to-sheet structural transformation of TDP-43 (residues 311–360) adopting parallel in-register cross-b structure in the fibrillar form composed of five β-strands.[156] Furthermore, the ability of TDP-43 to convert normal proteins into aggregated and pathological aggregates is the basis of its classification into a prion-like protein.[55,190] The presence of an LCD, rich in polar amino acids with putative prion-like properties, is reminiscent of amyloid proteins with prion behavior identified in yeast such as Sup35 and Ure2p, for which the LCDs can also form *in vitro* aggregates, heterogeneous and rich in β-sheet structures.[150,151,385] Proteolytic cleavage leads to C-terminal fragments (CTFs) of TDP-43 under pathological conditions such as oxidative stress or acetylation, and deposition of TDP-43 CTFs are also considered as a hallmark of ALS and FTLD.[177,386]

Recent documentation suggests 19 different alternative transcripts are formed by various combinations of splicing events of the TARDBP gene encoding TDP-43 named TDP splice isoforms(sTDP-43s). Interestingly, these cytoplasmic TDP-isoforms interfere with RNA splicing and causes upregulation of endogenous *TARDBP* and loss of autoregulatory capacity of TDP-43, contributing to TDP-43 proteinopathy.[118,221] The TDP-43 splice isoforms, namely sTDP43-1 and sTDP43-2, are formed by the alternative TARDBP gene, and they differ by only three amino acids and lack the entire glycine-rich domain (residues 281–414). Instead, it is substituted by a new exon encoding a unique 18–amino acid C-terminus not found in FL-TDP43. [387] The potential NES (nuclear exporting signal) identified in the unique 18 amino acid sequence drives the cytoplasmic sTDP43 localization. NES's location at the extreme C-terminal enhances the accessibility for nuclear exporters and dominates over NLS (nuclear localization signal) present within the N-terminus, accelerates the mislocalization of sTDP43 into the cytoplasm.[222] These TDP-43 splice isoforms are observed in inclusion bodies that sequester

with WT TDP-43 and display neuronal toxicity when overexpressed in mammalian neurons. The occurrence of TDP-43 isoform aggregates in vulnerable motor neurons, glia of ALS patients implicates a crucial role of sTDP43 in ALS. [222]

Due to the inherent propensity of many amyloid proteins to exhibit structural heterogeneity in their fibrillar form, structural investigations of amyloid proteins encompassing low-complexity domains such as TDP-43 and its CTFs constitute a very challenging task. Magic-angle spinning (MAS) SSNMR is a powerful approach to study heterogeneous non-crystalline, fibrous amyloid proteins and can be readily combined with complementary biophysical techniques to characterize the amyloid fibrils of different TDP fragments.[70,72,294,345,346,388] Here, we have embarked on the molecular characterization of TDP-43, and its spice isoforms sTDP43-1 and sTDP43-2 in its amyloid state, as well as the CTFs, called TDP-35 (90-414), TDP-16 (267-414), TDP-13, TDP-11, TDP-10, and a C-terminal truncated construct (TDP-43 ΔGaroS2), using a combination of SSNMR, electron microscopy, and X-ray diffraction.

## 3.2 Results

### 3.2.1 Aggregation, morphology, and X-ray diffraction of TDP-43 constructs

We initially designed three recombinant protein constructs to explore the molecular conformation of TDP-43 in its fibrillar form: the full-length protein (residues 1-414, TDP-43), and CTFs called TDP-35 (TDP fragment of size 35 kDa, residues 90-414), TDP-16 (TDP fragment of size 16 kDa, residues 267-414) (**Figure 3.1**). TDP-35 excludes the N-terminal domain (NTD) and contains the two RRM domains and the intrinsically disordered, prion-like region; TDP-16 includes only the latter, consisting of the hydrophobic segment (318–340), the Q/N rich segment (341–367), and two segments rich in glycines, aromatic residues, and serines, called GaroS1 and GaroS2 following the annotation of *Mompean et al., 2016* (GaroS for Gly, Aromatic, Ser-rich segment).[33] To improve SSNMR sensitivity and later perform multidimensional spectroscopy, we overexpressed the three TDP constructs in E. coli grown in isotopically labeled media to obtain uniformly $^{13}C/^{15}N$-labeled protein preparations. The three constructs were then purified and self-assembled at high concentrations (1mg/ml) in 20 mM MES buffer pH 7.5.

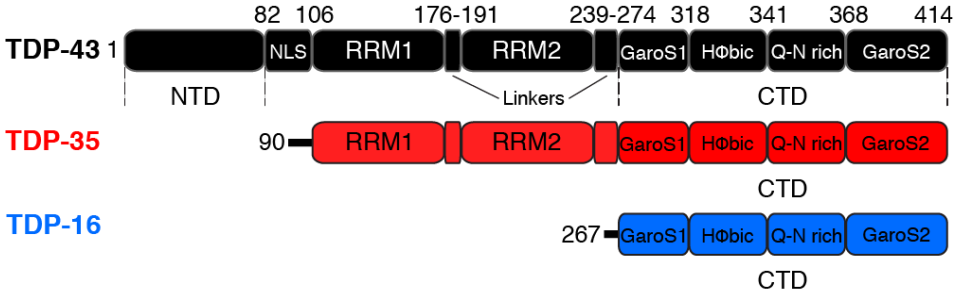



*Figure 3.1 The primary sequence of the three constructs TDP-43(1–414), TDP-35(90–414), and TDP-16(267–414): Boxes are used to represent different functional domains. All three proteins were produced using the recombinant expression in Escherichia coli.*

After one week of aggregation at room temperature under slow agitation, the different constructs reveal the formation of straight, unbranched filaments as observed by negative staining electron microscopy (**Figure 3.2A, B, and C**). Previous work using electron microscopy (EM) on full-length TDP-43 fibrils and short peptides showed fibrils around ~5-25 nm in diameter.[131,145,152,374] The CTFs (TDP-35,TDP-16) based on *in vitro* recombinant preparation lead to comparable filament diameter with a ~5-10 nm width for the protofilaments similar to TDP-43. All these TDP-43 constructs display homogeneous, thin, and featureless filaments at the macroscopic level. Although the EM resolution of the negatively stained samples is limited, the observed morphology for TDP-35 filaments does not reveal multiple protofilaments within the fibrils.

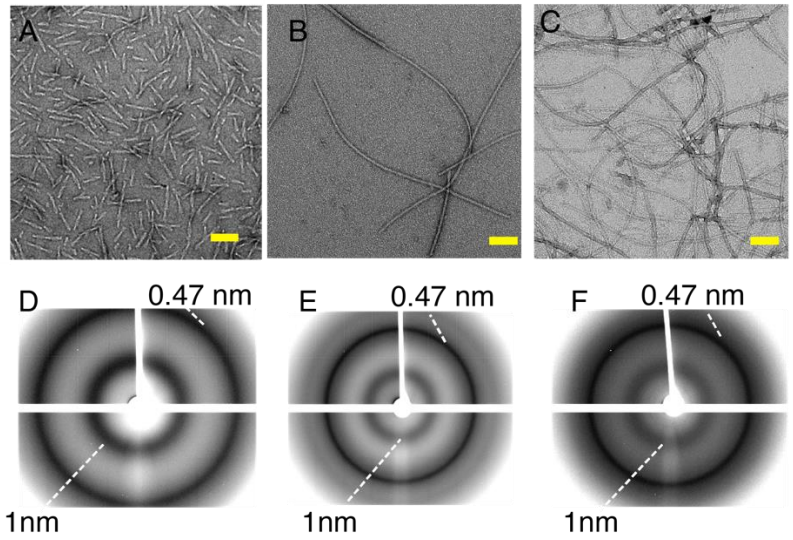



*Figure 3.2 Negatively stained electron micrographs of TDP-43 (A), TDP-35 (B), and TDP-16 (C) aggregates showing straight, unbranched filaments. Scale bars (in yellow) are 100 nm. X-ray diffraction pattern TDP-43 (D), TDP-35 (E), and TDP-16 (F) aggregate showing characteristic cross-β diffraction pattern in amyloid fibrils. Reflections at 4.7 and 10 Å are highlighted corresponding to the inter-strand and inter-sheet spacing, respectively. (G) X-ray scattering intensity for TDP-43 aggregates showing maxima at 4.7 and 10 Å.*

Next, we conducted X-ray fiber diffraction to gain insights into the TDP molecular arrangement in the filaments. The X-ray diffraction data for all these TDP constructs consist of two intense rings at 4.7 Å and 10 Å (**Figure 3.2D, E, and F**), a characteristic diffraction pattern observed cross-β structure in amyloid fibrils, corresponding to the inter-strand and inter-sheet spacing, respectively.[22] Our X-ray fiber diffraction results are comparable to most pathological amyloid fibrils displaying cross-β architecture.[389]

### 3.2.2 Solid-state NMR of TDP amyloid aggregates reveals a rigid core of various structural polymorphism

The molecular structures of the three TDP constructs in their fibrillar form were investigated using magic-angle spinning SSNMR. First, we probed the three-dimensional (3D)

fold of the rigid amyloid core using dipolar-based polarization transfer (cross-polarization). We employed two-dimensional $^{13}C$-$^{13}C$ spectroscopy with a short proton-driven spin-diffusion (PDSD) mixing time of 50 ms to assess intra-residue $^{13}C$-$^{13}C$ correlations. A global assessment of the $^{13}C$-$^{13}C$ experiments performed on TDP constructs reveals two types of SSNMR fingerprints. Amyloid aggregates of TDP-43 and TDP-35 lead to broad SSNMR signals (**Figure 3.3**), indicating a crucial structural heterogeneity at the local level. Full-width $^{13}C$-$^{13}C$ spectra are showed in **Figure 3.3**, and the representative region of alanine Cα-Cβ is shown in **Figure 3.4**. We estimate a $^{13}C$ linewidth of ~200-250 Hz (FWHH, ~1-1.25 ppm) for TDP-43 and TDP-35, similar to several pathological amyloid fibrils studied by SSNMR. The local polymorphism arises from multiple sidechain and backbone conformations leading to numerous and undistinguishable chemical shift variations resulting in broad lines. Such structural polymorphism has been observed for many pathological amyloid fibrils by SSNMR. It is generally accepted as an expression of the structural diversity of conformations that amyloid proteins can adopt during the misfolding process related to disease propagation.[70,71,356,390] The spectral region encoding for alanine Cα-Cβ correlations contains cross-peaks reflecting β–sheet, α–helical and random coil conformation. Overall, the chemical shift values indicate a largely β-sheet-rich conformation for the TDP-43 and TDP-35, a conformation expected for pathological amyloid fibrils.

Interestingly, SSNMR fingerprints of the TDP-35 and TDP-43 are superimposable, as illustrated in **Figure 3.4** for the alanine and proline signals, indicating that the molecular conformation of the amyloid core is virtually conserved between the TDP-43 and TDP35. The presence of a high degree of local structural heterogeneity observed by SSNMR contrasts with homogeneous filaments at the mesoscopic level (**Figure 3.2**). Interestingly, such results have been seen for several pathological amyloid fibrils.[391]

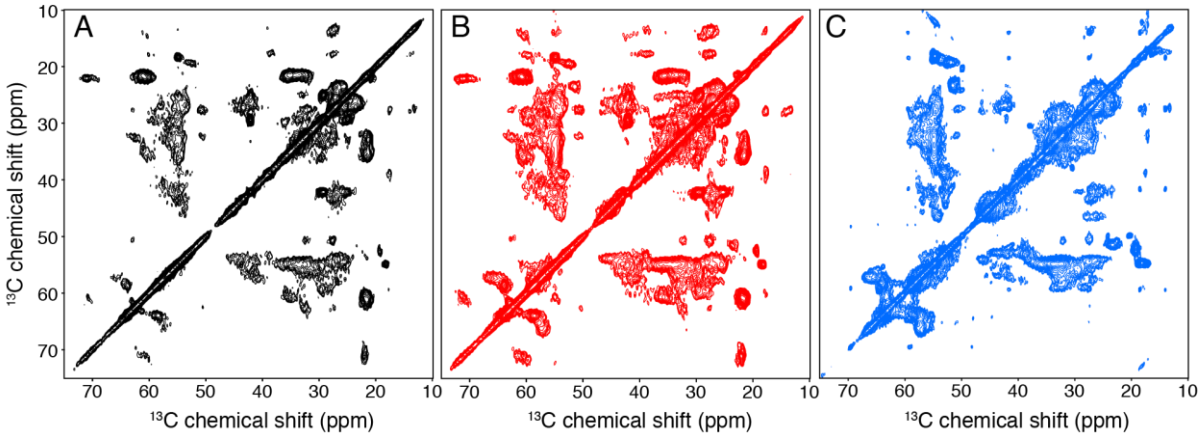



*Figure 3.3 Solid-state nuclear magnetic resonance $^{13}C$–$^{13}C$ correlation spectra of the different self-assembled [U–$^{13}C$, $^{15}N$]-labeled TDP fibrillar assemblies, the carbon-carbon magnetization transfer is achieved by a proton-driven $^{13}C$–$^{13}C$ spin diffusion. The aliphatic region of (A) full-length TDP-43 (black), (B) the TDP-35 construct (red), and (C) the TDP-16 construct (blue) recorded at 278 K at a $^{1}H$ Larmor frequency of 600 and 800 MHz, respectively, aan11 kHz MAS*

*frequency and a mixing time of 50 ms for obtaining intraresidue CC correlation. The numbers of scans are, respectively, 192, 192, and 600. The sample quantity per rotor is ~ 10 mg.*

### 3.2.3 Protein segments in TDP-43 and its C-terminal fragments undergo high mobility in the aggregated state

Next, we addressed the identification of highly mobile segments of the TDP-43 amyloid aggregates by employing dynamics-based spectral editing, here using J-based polarization transfer.[392] We use a $^{1}$H-$^{13}$C J-based INEPT correlation experiment to reveal the mobile residues of the fibrillar assembly. All three constructs show numerous signals in the INEPT spectra (**Figure 3.5 and Figure 3.6**), indicating highly mobile protein segments. This type of experiment reveals signals corresponding to residues in flexible regions, i.e., most often in random coil conformation, a site-specific resonance assignment is not possible, and only residue-type identification can be attempted. The INEPT signals observed in TDP-43 and TDP-35 are comparable, and a few new correlations are visible for the TDP-16 and shorter prion-domain fragments (Hα-Cα signals at ~55-60 ppm in **Figure 3.5**), indicating that the mobile segment(s) observed in the different protein constructs would be comparable. In a non-complete polymerization process, a part of the protein sample would stay in a non-assembled monomeric state.

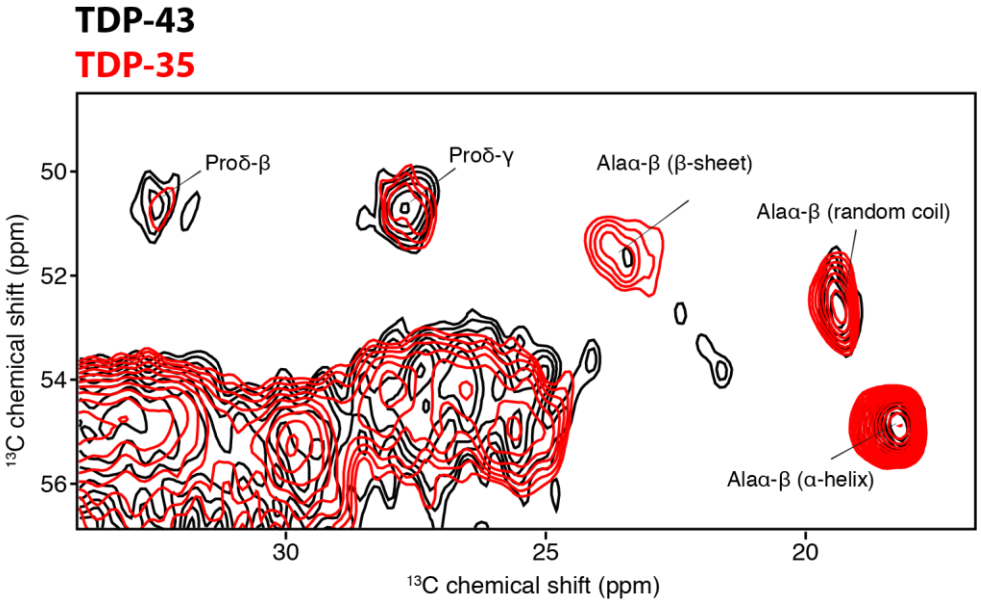



*Figure 3.4 An excerpt from the overlay of the $^{2}$D $^{13}$C–$^{13}$C PDSD correlation spectra obtained on (U–$^{13}$C,$^{15}$N)-labeled samples of full-length TDP-43 (black) and TDP-35 (red) recorded at 278 K and a $^{1}$H frequency of 600 MHz, 11 kHz MAS frequency, a mixing time of 50 ms and 600 scans. Labels indicate characteristic alanine Cα–Cβ and proline Cδ–Cβ correlations and chemical shift-dependent secondary structure in brackets.*

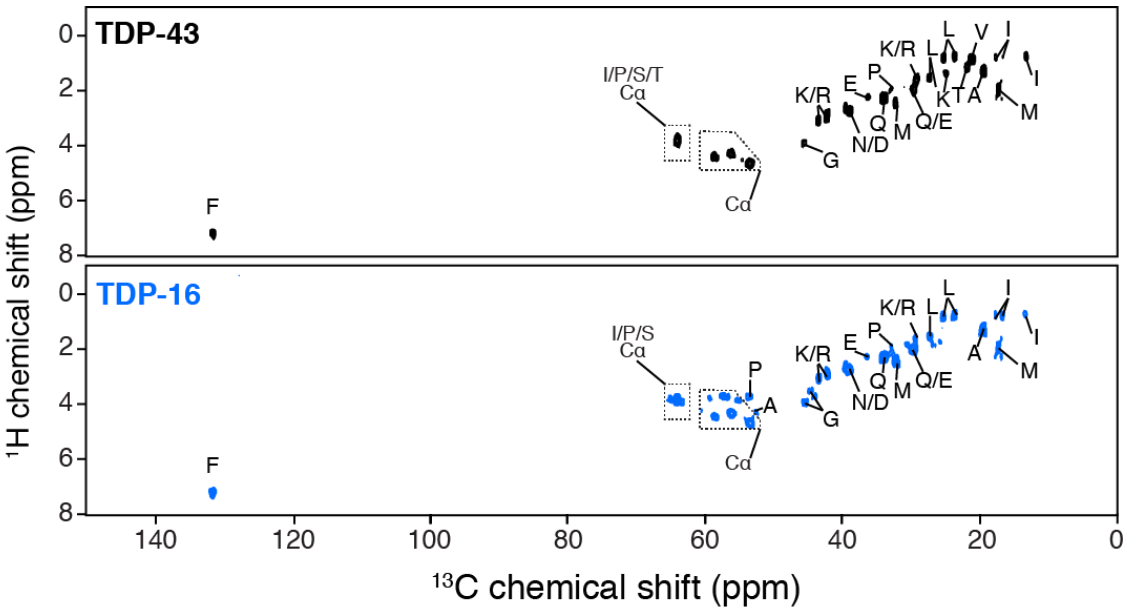



*Figure 3.5 SSNMR $^{1}H$–$^{13}C$ INEPT experiments of full-length TDP-43 (black) and TDP-16 (blue) amyloid aggregates recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency, and 128 scans at 278 K showing tentative assignments corresponding to residues in the flexible region of the fibrillar assembly.*

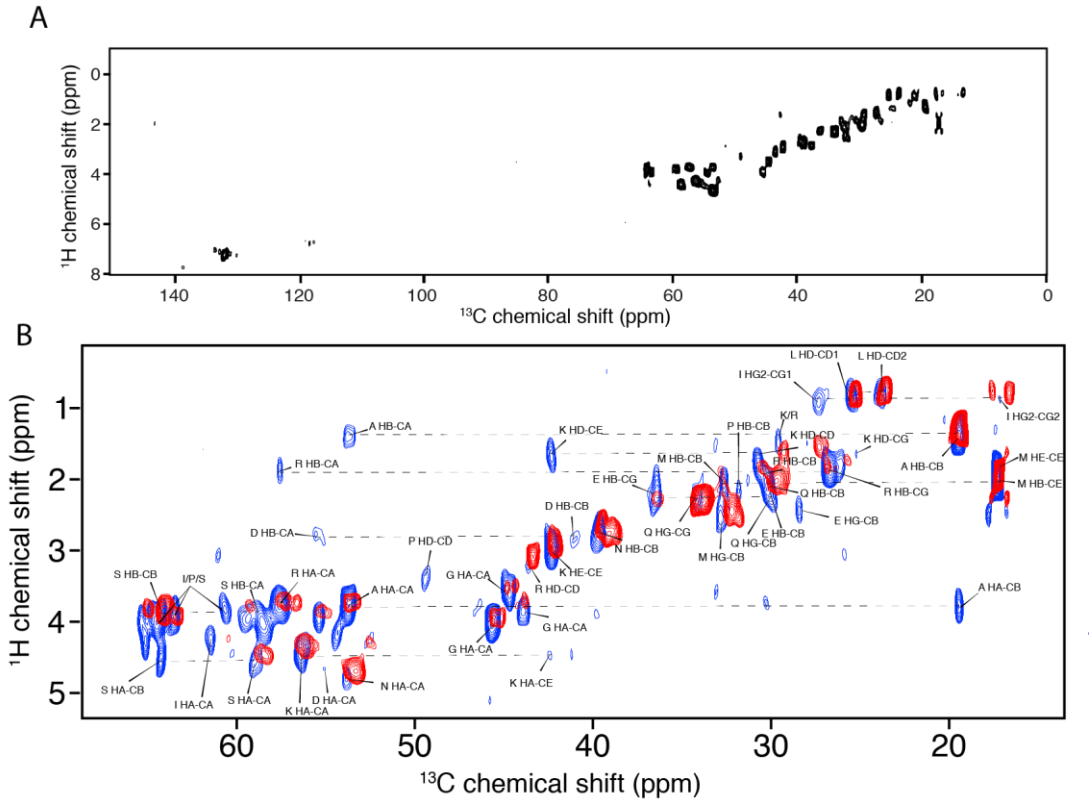



*Figure 3.6 (A) $^{1}H$–$^{13}C$ INEPT spectrum of TDP-35 recorded at a $^{1}H$ frequency of 600 MHz, at 278 K with 416 scans showing signals corresponding to residues in the flexible region of the fibrillar assembly. (B) SSNMR ($^{1}H$)–$^{1}H$–$^{13}C$ INEPT-TOBSY (in blue, 2048 scans) and $^{1}H$–$^{13}C$ INEPT (in red, 128 scans) spectra of TDP-16 fibrils recorded at a 1H frequency of 600 MHz, at 278 K showing*

*residue-type assignments of the mobile residues of the fibrillar assembly. Intra-residue CC correlations are indicated with dotted lines resulting from the TOBSY transfer step.*

To address the presence (or not) of monomeric TDP molecules in the SSNMR sample, we investigate the presence of residues that could be unambiguously identified in the INEPT experiment. More residue-type assignments were performed using a 2D ($^{1}$H)-$^{1}$H-$^{13}$C INEPT-TOBSY and shown in **Figure 3.6**. Aromatic residues have well defined and isolated chemical shifts, making them useful amino acids to perform an unambiguous assignment. We observed a single aromatic correlation (encoding for phenylalanine) in the INEPT experiments while other aromatic residues were not visible. TDP-43 contains 23 phenylalanines, 6 tryptophans, 11 histidines and 9 tyrosines (9, 3, 6 and 2 respectively in the TDP-16 construct). It indicates that the J-based SSNMR experiments' signals originate from very mobile protein segments within the fibrillar assemblies and not from monomeric species. It suggests that a part of the TDP-43 primary sequence is not involved in the rigid amyloid core. However, the presence of signals in the INEPT spectrum of fibrils of TDP-16 and shorter prion-domain fragments is more surprising, implying that the minimum sequence enabling aggregation is smaller than the corresponding prion-domain fragments. Indeed, it has been demonstrated that the shorter peptide fragments from CTD such as 341-357, 286-331, and 311-360 are sufficient to self - assemble into amyloid fibrils.[153,155,156]

### 3.2.4 The C-terminal fragment TDP-16 forms a well-ordered amyloid core within the fibrils.

In contrast to TDP-43 and TDP-35, amyloid fibrils formed by the TDP-16 fragments show a different SSNMR $^{13}$C-$^{13}$C fingerprint (**Figure 3.7**). We observed two types of $^{13}$C-$^{13}$C correlations in the TDP-16 spectrum: (i) broad lines comparable to those observed for TDP-43 and the TDP-35 fragments, and (ii) sharper lines indicating a more stringent molecular conformation, i.e., a much less pronounced structural polymorphism. Several sharp NMR signals are highlighted in Figure 7 and could be identified based on a residue-type assignment. We estimate a $^{13}$C linewidth of ~100 Hz (FWHH, 0.5 ppm) for these sharp NMR signals, comparable to linewidths observed for highly organized amyloid fibrils such as the prion domain of HET-s or $\alpha$–synuclein.[393,394] The broad signals detected in TDP-43 and TDP-35 samples are observed in the TDP-16 fragment, suggesting a common polymorphic β-sheet-rich part in the three constructs.

Interestingly, most of the sharp NMR signals observed in the TDP-16 fragment do not superimpose with signals from TDP-43 and TDP-35 aggregates. As illustrated with the alanine region in the excerpt presented in **Figure 3.7-B**, two C$\alpha$-C$\beta$ correlations are only present in the TDP-16 sample with unambiguous $\beta$–sheet conformation according to their chemical shifts. 1D $^{13}$C slices extracted from the 2D $^{13}$C-$^{13}$C spectra of all constructs investigated in our work reveal that these two signals are not detected for TDP-35, TDP-43, and TDP-43 ΔGaroS2 (**Figure 3.7C**). We observed that almost all sharp peaks identified in the TDP-16 amyloid fibrils display a $\beta$–sheet-rich conformation. It suggests that the well-ordered amyloid core in TDP-

16 displays a different molecular organization in the context of the isolated fragment and the full-length sequence assembly.

Although solid-state NMR chemical shifts offer a powerful tool to monitor the secondary structure, in the case of broad peaks, it remains challenging to distinguish the different contributions from α–helical, β–sheet, and random-coil Structure. Therefore, we carried out ATR FT-IR analysis to study the relative amount of secondary structure elements of TDP-43, TDP-35, and TDP-16 amyloid fibrils. Both TDP-43 and TDP-35 aggregates present a comparable FT-IR signal in the amide I and amide II range (**Figure 3.7-D**). A deconvolution analysis was used to extract the contribution of secondary structure elements, namely parallel and anti-parallel β–sheet (1634, 1620, and 1690 cm-1), α–helix (1658 cm-1), turns (1675 cm-1,) and random coil (1646 cm-1) according to values reported here.[395] TDP-43 and TDP-35 aggregates exhibit a comparable FT-IR signature, with a β–sheet secondary structure amount of ~50% (Table 1), while TDP-16 fibrils show a lower content of β–sheet (~40%). The β–sheet content in TDP-35 and TDP-16 is higher than what is observed in the previously solved Structure of TDP-43 domains (~20-25%), reinforcing the idea that the amyloid transition is associated with a misfolding of the NTD and RRM domains. SSNMR and FT-IR analysis are very complementary since SSNMR spectra specifically probe respectively rigid and highly mobile residues (CP- and J-coupling-based experiments). In contrast, FT-IR probes the overall protein content of the sample. Because only in TDP-16, a rigid and well-ordered amyloid core is observed, FT-IR data reveals that an extensive amount of non β–sheet conformation is present in the fibrillar architecture, which we interpret as an expression of the misfolding process and its resulting structural polymorphism. This non β–sheet contribution, which may surround the rigid amyloid core, is rigid enough to be observable in the SSNMR CP-based experiments.

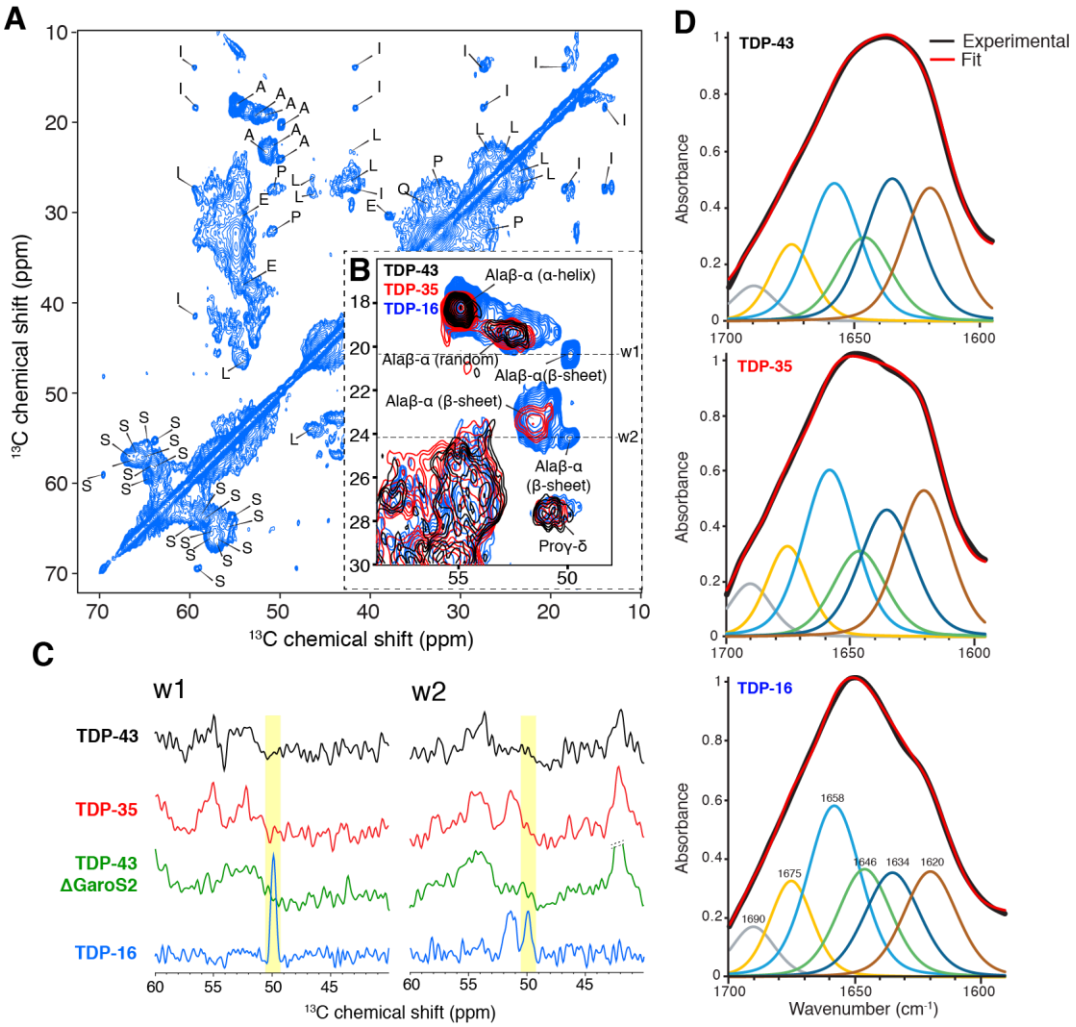



*Figure 3.7 (A) SSNMR $^{13}C$–$^{13}C$ correlation experiment of TDP-16 amyloid aggregates recorded at a $^{1}H$ frequency of 800 MHz, 11 kHz MAS frequency, and mixing time of 50 ms at 278 K, using 600 scans. (B) Excerpt of overlaid $^{13}C$–$^{13}C$ PDSD correlation spectra of TDP-16 (in blue), TDP-35 (in red), and TDP-43 (in black) for the alanine Cα–Cβ region with chemical shift-dependent secondary structure in the bracket and unique alanine signals in TDP-16 indicated by dotted lines. (C) 1D $^{13}C$ slices extracted from the 2D $^{13}C$–$^{13}C$ experiments showing the two alanine signals observed for TDP-16 (highlighted in yellow) are not detected for TDP-35, TDP-43, and TDP-43 ΔGaroS2. (D) FTIR spectra of TDP-16, TDP-35, and TDP-43 aggregates in the amide I and II range showing the experimental (black) and theoretically fitted curve (red).*

Next, we attempted the identification of protein segments in the rigid amyloid core of TDP-16 fibrils by conducting a $^{13}C$-$^{13}C$ correlation experiment with a proton-driven spin-diffusion long mixing time (200 ms) to detect sequentially (i.e., residue i to i±1) connectivity (**Figure 3.8**). A sequential correlation could be observed (A382-I383) since only one such combination in the sequence leads to identifying the G380-A-A-I383 segment. Furthermore, we assigned the stretch A391-G396 based on the AGS correlations. The poor spectral resolution limits further sequential assignments. To confirm the assignments, a point mutagenic strategy was employed in which the isoleucine at position 383 is mutated to valine,

i.e., I383V. This mutant formed amyloid fibrils having SSNMR fingerprints similar to other TDP-43 constructs. As illustrated with the alanine region in the excerpt presented in **Figure 3.8-B**, the resonances corresponding to I383 vanished along with the unique sharp alanine resonances present in the wild type construct.

*Table 3.1 Secondary structure content of TDP-43, TDP-35, and TDP-16 amyloid fibrils probed by ATR-FTIR analysis.*

| Secondary structure element | Wavenumbers ($cm^{-1}$) | Percentage of structural elements in percentage | | |
|---|---|---|---|---|
| | | TDP-16 | TDP-35 | TDP-43 |
| β-sheet (// and anti //) | 1620 1690 1634 | 41 | 50 | 52 |
| Random coil | 1646 | 18 | 13 | 15 |
| α-Helix | 1658 | 28 | 26 | 23 |
| Turn | 1675 | 13 | 11 | 10 |

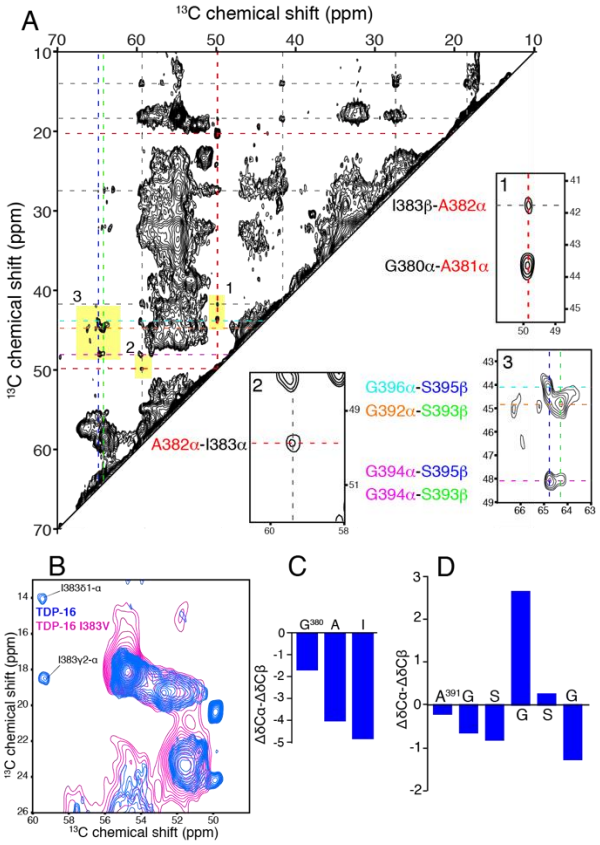



*Figure 3.8 Investigation of the amyloid core of TDP-16 amyloid aggregates. (A) SSNMR 2D $^{13}$C–$^{13}$C correlation experiment of TDP-16 amyloid aggregates recorded at 278 K and a $^{1}$H frequency of 800 MHz, 11 kHz MAS, and a mixing time of 200 ms with 792 scans for obtaining sequential (i.e., residue i to i ± 1) connectivity indicated by color-coded dotted lines. Example*

*of sequential assignment of the protein segment (highlighted in yellow) G380-I382 (1) and (2) and S393-G396 (3). (B) Excerpt of the superposition of TDP-16 (in blue), TDP-16 I383V (in pink), for the alanine Cα–Cβ region and isoleucine Cα–Cγ, Cδ region indicating the absence of I383 along with the unique sharp alanine resonances present in the wild-type construct. (C–D) The secondary chemical shift analysis of residues 380–382 and 391–396. Chemical shift-dependent secondary structure of amino acid sequence can be calculated by comparing experimentally obtained chemical shift with that of random-coil conformation using the equation ΔδCα–ΔδCβ. Positive or negative values correspond to α-helical or β-strand conformation, respectively.*

Interestingly, the signals corresponding to the sequential correlations to the amino acid stretch GAAI are also absent in the TDP-16 I383V spectra, which confirms these signals in TDP-16 are from the GaroS2 region (Fig. 8B). By comparing the experimentally obtained chemical shift values with random coil conformation, we calculated the secondary chemical shift for the GAAI segment (Fig. 8C) and the AGSGSG segment (Fig. 8D).[396] Positive and negative values correspond to α-helical and β-strand conformation, respectively. The secondary chemical shift analysis of the amino acid stretch G380-A-A-I383 in TDP-16 reveals a β-strand conformation. For the A391-G396 stretch, we observed β-strand conformation; however, a clear β-breaker was identified for residues G394 and S395 according to their chemical shifts. Sharp NMR linewidths are observed for these segments, located at the end of the CTD, and the loss of these signals for the I383V mutant indicates that the rigid and well-ordered amyloid core of TDP-16 is comprised in the GaroS2 segment (residues 368-414). A recent SSNMR study on inclusion body seeded LCD fibrils suggests that the GaroS2 segment comprises two b-sheets where the G380-A-A-I383 segment is accommodated.[156] However, the A391-G396 segment in this study adopts a helical conformation, possibly due to a different polymorph. Although the GaroS2 segment is not rich in N/Q or hydrophobic residues and therefore not intuitively associated with amyloid formation, our results pinpoint the importance of the last part of the CTD in forming a well-ordered amyloid core.

### 3.2.5 The GaroS2 domain is part of the rigid core of TDP-16 but is not necessary for the full-length TDP-43 aggregation

The two protein stretches identified in the TDP-16 fragment by solid-state NMR are comprised in the so-called GaroS2 domain (residues 368-41473). To evaluate the importance of this domain on the molecular conformation as observed by NMR, we designed a C-terminal TDP construct (**Figure 3.9A**) lacking the GaroS2 domain (TDP-43 ΔGaroS2, i.e., residues 1-366 of TDP-43).[33] TDP-43 ΔGaroS2 aggregates in the same buffer conditions. On comparing TDP-43 ΔGaroS2 with full-length TDP-43, we observed that comparable $^{13}C$-$^{13}C$ SSNMR fingerprints are exemplified for isolated cross-peaks of serine threonines (**Figure 3.9B**) and alanines (**Figure 3.9C**). This observation suggests that the amyloid core may accommodate the lack of the GaroS2 domain and still maintain a molecular conformation comparable to full-length TDP-43. It supports our previous results (**Figure 3.7**) as the GaroS2 domain in the TDP-16

fragment showed SSNMR signals not visible in the full-length construct. The 2D 1H-$^{13}$C INEPT spectrum of TDP-43 ΔGaroS2 (**Figure 3.10**) shows comparable signals compared to TDP-43; however, the phenylalanine contribution is absent. Because the GaroS2 contains three phenylalanine residues, it indicates in a mobile segment of TDP-43. Therefore, the GaroS2 segment is crucial for TDP-16 aggregation but not a limiting factor to assemble the full-length protein, in line with previous studies highlighting the importance of the hydrophobic and the Q/N rich regions for the aggregation.[91,131,152,153,155]

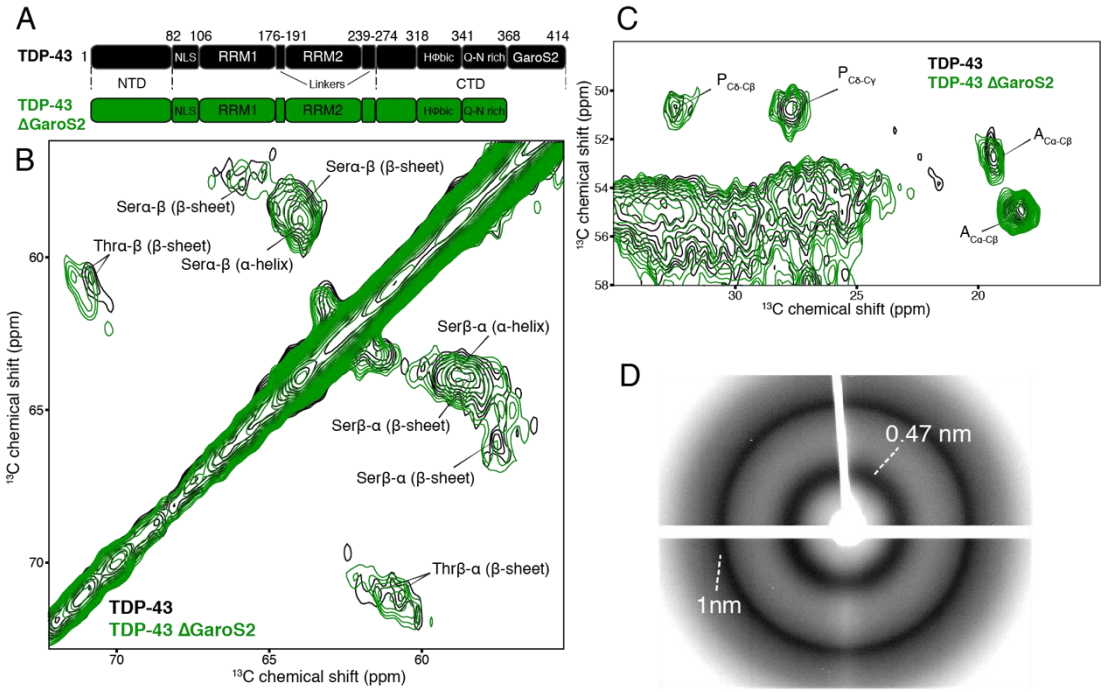



*Figure 3.9 (A) Sequence construction strategy of TDP-43 ΔGaroS2 that corresponds to TDP-431–366. (B-C) overlaid $^{13}$C–$^{13}$C PDSD correlation spectra of TDP-43 ΔGaroS2 (in green) and TDP-43 (in black) highlighted for the Ser/Thr Cα–Cβ region (B) and Ala Cα–Cβ (C) region recorded at a 1H frequency of 600 MHz, 11 kHz MAS frequency, and a mixing time of 50 ms at 278 K, with the calculated secondary structure of the residue in the bracket. (D) X-ray diffraction pattern of TDP-43 ΔGaroS2 filaments showing characteristic cross-β diffraction pattern in amyloid fibrils. Reflections at 4.7 Å and 10 Å are highlighted, corresponding to the inter-strand and inter-sheet spacing, respectively.*

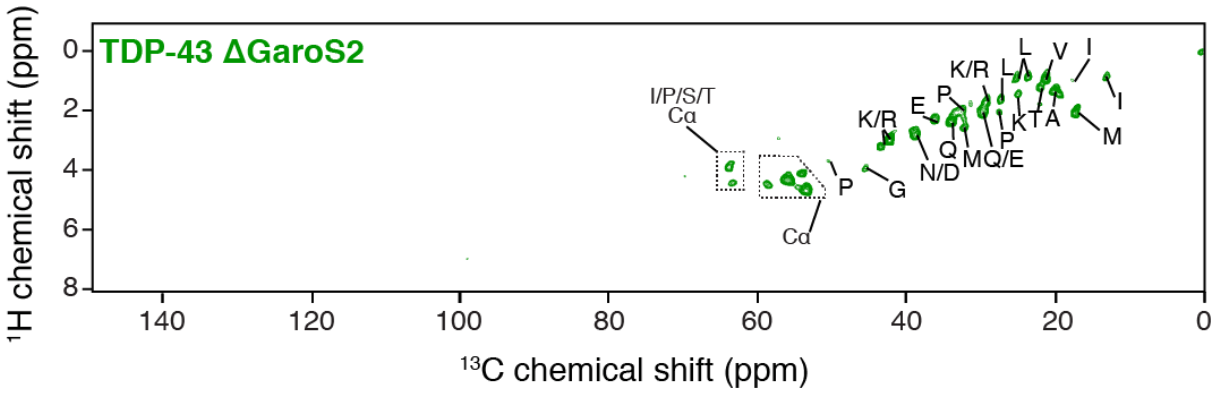



*Figure 3.10 SSNMR $^{1}H$–$^{13}C$ INEPT of TDP-43 deltaGaroS2 fibrils recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency at 278 K, with 192 scans showing tentative assignments corresponding to residues in the flexible region of the fibrillar assembly.*

### 3.2.6 Aggregation, morphology, and X-ray diffraction of TDP-CTF fragments

The high number of the INEPT spectrum peaks (shown in **Figure 3.5**) for the TDP-16 construct (LCD) indicates a significantly higher amount of flexible regions. To investigate the shortest part to encode the amyloid core in more detail, we employed a "divide and conquer" strategy by creating three shorter C-terminal fragments (CTFs): TDP-13 (300-414), TDP-11(300-399), and TDP-10 (314-414). These constructs consisted of the hydrophobic and Q/N rich regions but varied in sequences from GaroS1&2 subdomains (**Figure 3.11 A**). All these constructs were made based on recombinant bacterial expression and purification. The purified proteins were aggregated in comparable filaments with a diameter of ~5-10 nm (**Figure 3.11A**). The protofilaments were similar to those formed by full-length TDP-43, displaying homogeneous, thin, and featureless filaments at the macroscopic level (**Figure 3.11B**). Although the EM resolution of the negatively stained samples is limited, the observed morphology for TDP-CTF filaments does not reveal the presence of multiple protofilaments within the fibrils. The X-ray diffraction data for all these shorter CTFs comprises a characteristic diffraction pattern of two intense rings at 4.7 Å and 9 Å (**Figure 3.11E, F, G**). This pattern is typical for cross-β structural elements in amyloid fibrils. The data were consistent with other TDP-43 constructs (TDP43, TDP-35, and TDP-16).[22,31] The organic dye, thioflavin T is used as fluorescent binding probe for histology staining of amyloid fibrils. The ThT molecule display enhanced fluorescence on binding to amyloids.[397] The excitations at 440 nm resulted in an enhanced fluorescence signal at approximately 480 nm for shorter CTFs fibrils (**Figure 3.11D**). These observations indicated that the new CTF constructs (TDP-13, TDP-10, and TDP-11) form amyloid-type nself-assemblies. The different fluorescence intensities from ThT binding assays for CTF fibrils could be due to the difference in the fibril concentration used in the experiment (0.2-0.3 mg/ml) or different modes of ThT binding.[398]

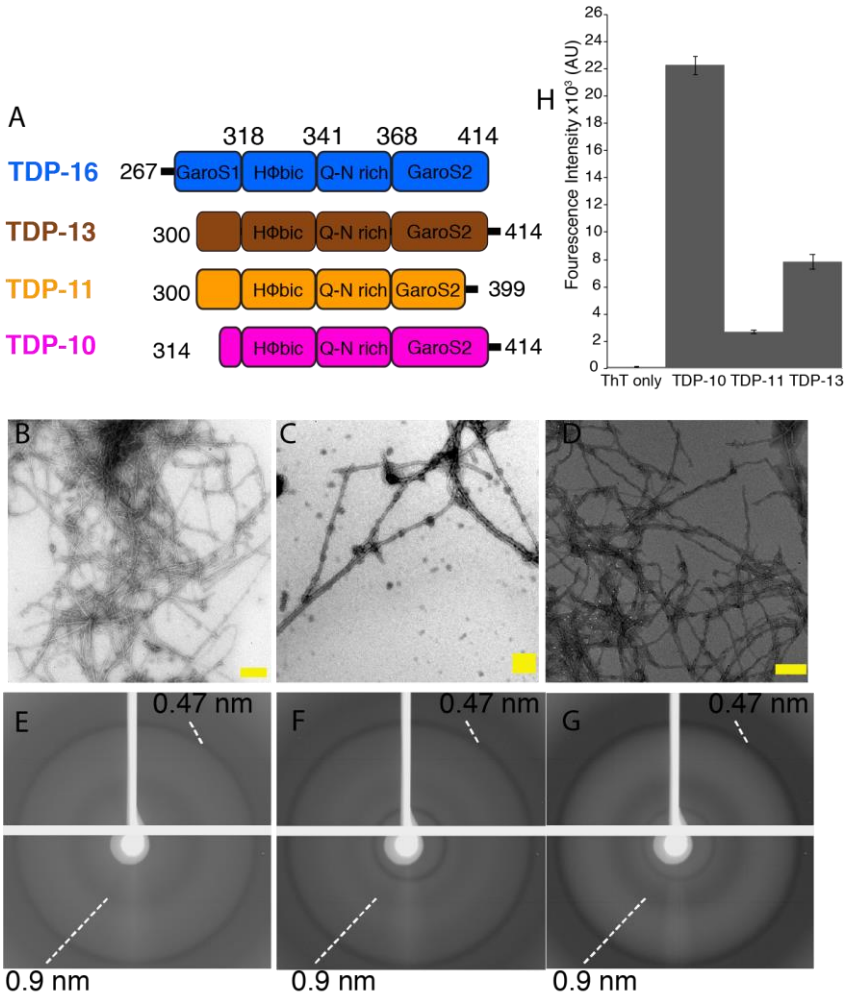



*Figure 3.11 A) Sequence construction strategy of TDP LCD fragments, TDP-$^{13}$(300-414), TDP-11(300-399), and TDP-10(314-414): Boxes are used to represent different functional domains. Negatively stained electron micrographs of B) TDP-13, C) TDP-11, and D) TDP-10 aggregates showing straight unbranched filaments. Scale bars (in yellow) are 200 nm. C) X-ray diffraction patterns of E) TDP-13, F) TDP-11, and G) TDP-10 aggregates showing characteristic cross-β diffraction pattern in amyloid fibrils. Reflections at 4.7 Å and 10 Å are highlighted. D)ThT fluorescence assay on TDP-13, TDP-11, and TDP-10 fibrils.*

### 3.2.7 SSNMR of TDP-CTF amyloid aggregates reveals the presence of flexible regions

Next, we addressed the identification of highly mobile segments of the TDP-CTF amyloid aggregates by employing the INEPT based $^{1}$H-$^{13}$C correlation experiment. Surprisingly, all three shorter CTFs show multiple signals in the INEPT spectra (**Figure 3.12**), indicating highly mobile protein segments. The INEPT signals observed in TDP-13 and TDP-10 were comparable, and a few new correlations are observed in the Cα region (~54-60 ppm) of the TDP-11 (Hα-Cα correlation peaks are indicated with the dotted squares in **Figure 3.12**). The similar chemical shift patterns from INEPT spectra (**Figure 3.12**) revealed that the mobile segment(s) observed in the different protein constructs were similar. The possibility of protein sample remaining in a non-assembled monomeric state can be excluded as we

observed only 1 or 2 aromatic residues (1 phenylalanine and 1 tyrosine), and no typical signals were observed of other aromatic residues such as histidine or tryptophan, which is present in the primary sequence. It suggests that a part of the shorter CTF's primary sequence is not involved in the rigid amyloid core. However, the presence of signals in the INEPT spectrum (**Figure 3.12**) of fibrils of shorter CTFs is more surprising, implying that the minimum sequence enabling aggregation is smaller than the corresponding prion-domain fragments. This inference is in line with the observation that the shorter peptide fragments from CTD such as 341-357, 286-331, and 311-360 are sufficient to self -assemble into amyloid fibrils. [153,155,156]

Interestingly, TDP-11 and TDP-13 comprised only a few resonances compared to TDP-10, which corresponded to random coil conformation. The results imply that a limited number of residues were in structurally flexible regions, consistent with the assumption that most residues in TDP CTFs, TDP-10, and TDP-13 are in the structurally rigid region. Unfortunately, these data were incapable of pointing out a specific flexible region, as the peaks in INEPT spectra (**Figure 3.12**) signals indicate serine, glycine, alanine, methionine, glutamine, and asparagine that are not localized in a specific region of LCD.

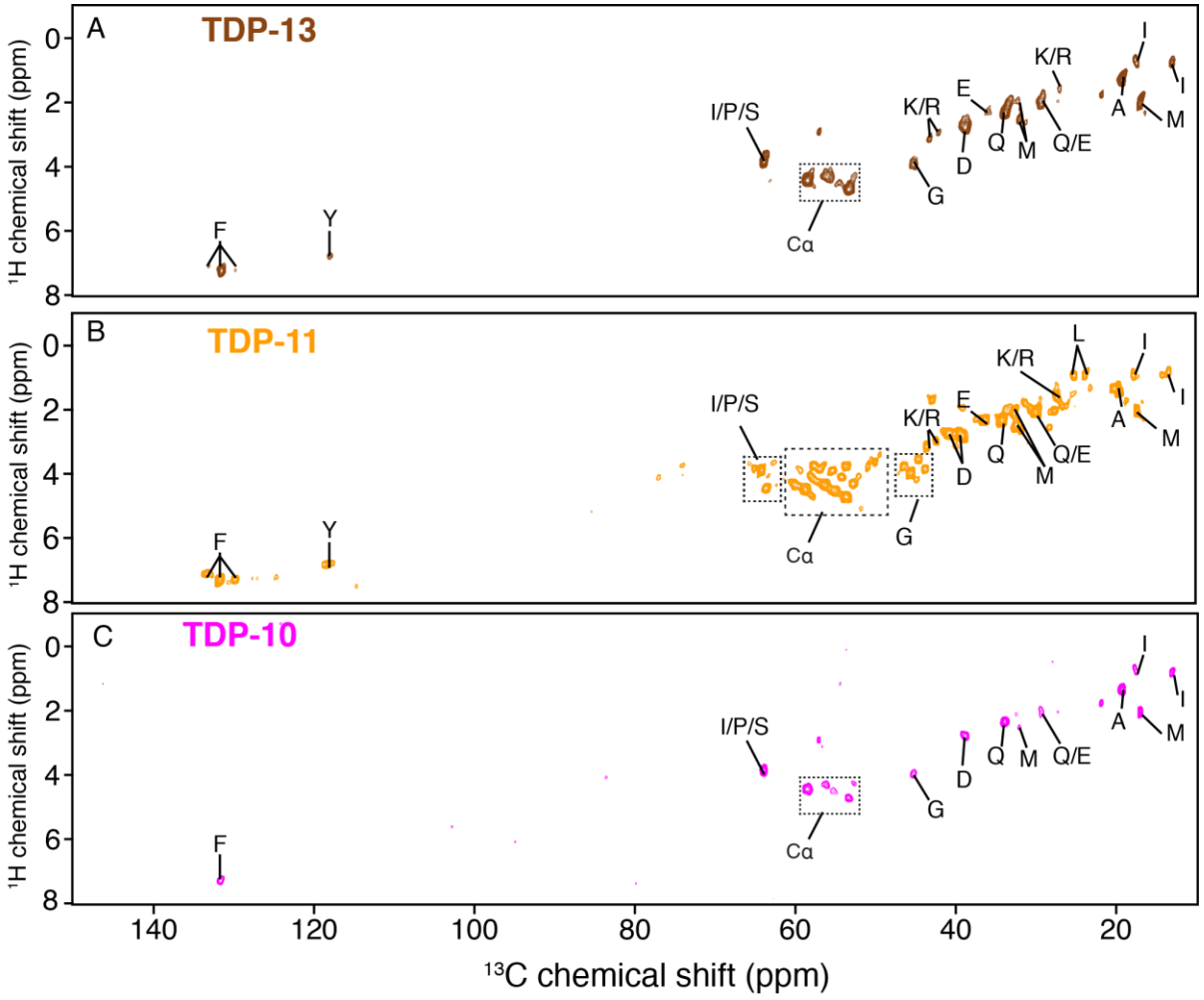



*Figure 3.12 SSNMR $^{1}H$–$^{13}C$ INEPT spectra of A)TDP-13(brown), B) TDP-11(orange), and C) TDP-10 (pink) amyloid aggregates showing tentative assignments corresponding to residues in the*

*flexible region of the fibrillar assembly. The data were recorded at $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency at 278 K.*

### 3.2.8 TDP-43 C-terminal short constructs formed different polymorphs with a well-ordered amyloid core

In contrast to TDP-43 and TDP-35, amyloid fibrils formed by the short TDP-CTFs showed a different SSNMR $^{13}C$-$^{13}C$ fingerprint, which was more similar to TDP-16 (**Figure 3.13**), indicating a more stringent molecular conformation with less pronounced structural polymorphism. Several sharp NMR signals could be identified with an estimated $^{13}C$ linewidth of ~100 Hz (measured at FWHH), comparable to linewidths observed for highly organized amyloid fibrils such as the prion domain of HET-s and α–synuclein.[393,394] The broad signals detected in the TDP-16 sample are also observed in the CTFs (**Figure 3.13, 3.14 A**), suggesting a common polymorphic β-sheet-rich part in these constructs.

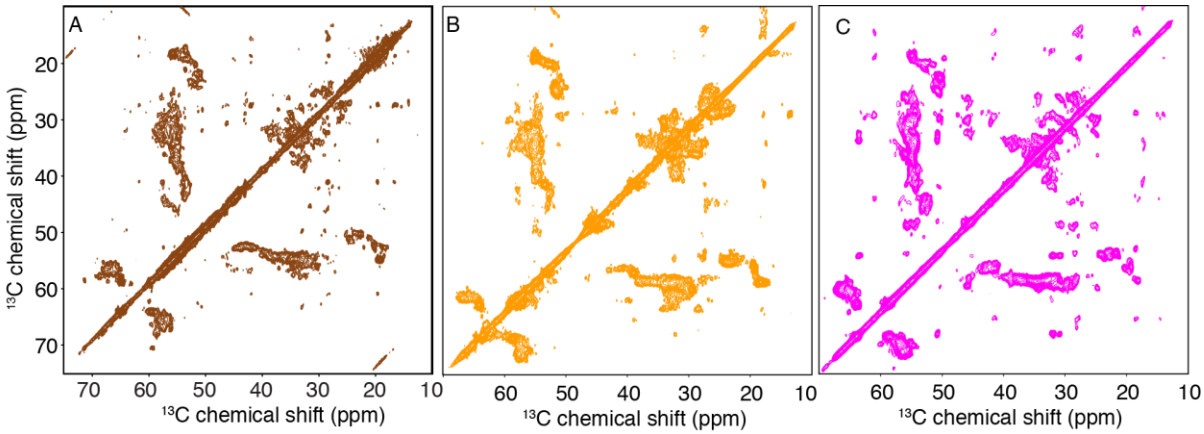



*Figure 3.13 SSNMR $^{13}C$–$^{13}C$ correlation spectra of the different self-assembled [U–$^{13}C$, $^{15}N$]-labeled TDP fibrillar assemblies, the carbon-carbon magnetization transfer is achieved by a proton-driven $^{13}C$–$^{13}C$ spin diffusion. The aliphatic region of (A) TDP-13(brown), (B)TDP-11(orange), and( C)TDP-10 (pink) aggregates recorded at 278 K at a $^{1}H$ Larmor frequency of 800 MHz, respectively, with an 11 kHz MAS frequency and a mixing time of 50 ms with the PDSD recoupling for obtaining intra-residue CC correlation.*

*Table 3.2 Secondary structure content of TDP-13, TDP-11, and TDP-10 fibrils probed by ATR-FTIR analysis.*

| Secondary structure element | Wavenumbers ($cm^{-1}$) | Percentage of structural elements (%) | | |
|---|---|---|---|---|
| | | TDP-10 | TDP-11 | TDP-13 |
| β–sheets // or anti// | 1620, 1635, 1690 | 37 | 44 | 46 |

| | | | | |
|---|---|---|---|---|
| Random coil | 1646 | 14 | 15 | 13 |
| α-Helices | 1658 | 34 | 29 | 28 |
| Turn | 1675 | 15 | 12 | 13 |

Interestingly, the different CTF fibrils adopted distinctive molecular architecture, as exemplified by comparing the alanine Cα-Cβ correlation region with TDP-16 (**Figure 3.14 A, B, C**), where each CTF fibrils display variation in the number of alanine signals observed or in its secondary structure or in $^{13}C$ linewidth representing the rigidity at the local level. However, the sharp signals observed in TDP-16 are not observed in any of the short CTFs. A significant amount of structural polymorphism is observed in TDP-13 and TDP-11 fibrils implied by duplicated resonances for the same residues (proline δ–γ, δ–β correlations) (**Figure 3.14 A, B, C**). However, these duplicated resonances may be encoded as broad signals in TDP-16 and TDP-10. The well-ordered amyloid core in each CTFs displays a different molecular organization depending on the amino acid sequence length representing the prion domain. This inference is similar to what's been observed by *Cao et al*. , where the CTD segment (311-360) can adopt three different polymorphs, maintaining a common "dagger-shaped" fold.[155]

Further, we carried out ATR FT-IR analysis to study the relative amount of secondary structure elements of shorter CTF amyloid fibrils. Both TDP-13 and TDP-11 aggregates present a comparable FT-IR signal in the amide I and amide II range (**Figure 3.14D, E, F**). TDP-13 and TDP-11 aggregates exhibit a comparable FT-IR signature, with a β–sheet secondary structure amount of ~45% (Table 2), while TDP-10 and TDP-16 fibrils show a lower content of β-sheet (~40%). SSNMR and FT-IR analysis are very complementary since SSNMR spectra specifically probe rigid or highly mobile residues, and FT-IR probes the sample's overall secondary structure content. Like TDP-16, the shorter CTFs constructs (TDP-13,11,10) display a significant amount of non-β–sheet conformation in its fibrillar architecture, which we interpret as flexible loops or helical turns that connect the β-strands. A recent SSNMR data acquired on fibrils seeded from aggregated LCD structures in inclusion bodies suggest that only ~50% of the C-terminal domain contributes to the amyloid core.[399] The non β-sheet contribution in CTFs can be attributed to semi-rigid segments surrounding the rigid amyloid core but rigid enough to be observable in the SSNMR CP-based experiments.

Next, we observed that TDP-13 displays a better spectral resolution with many isolated peaks. (**Figure 3.13A**) So, we attempted to identify protein segments in the rigid amyloid core of TDP-13 fibrils by conducting a $^{13}C$-$^{13}C$ correlation experiment with a 2D 150 ms long PDSD spectra to detect sequentially (i.e., residue i to i±1) connectivity (**Figure 3.15A**). A sequential correlation could be observed (I3I8-A322) since only one such combination in the sequence leads to identifying the I318-N-P-A322 segment. Furthermore, we assigned the stretch S333-W334-G335 and F367-S370 based on the single unique S-aromatic residue-G correlations. The poor spectral resolution limits further sequential assignments. The aromatic ring of W334 and F367 are very rigid, showing strong signals in the aromatic region (~110-140

ppm) of the 2D $^{13}C$-$^{13}C$ correlation spectrum (**Figure 3.15B**), suggesting the aromatic sidechain is involved in inter-residue interactions and participates in the amyloid core for these residues. Several unassigned correlations between glycines and aromatic residues (Try or Phe) can be observed in the aromatic region, suggesting they are part of the amyloid core (**Figure 3.15B**). These inter-residue correlations in the SSNMR spectra suggest that the TDP-13 fibrils have a single dominant fibril structure. However, the broad linewidth for cross-peaks reflects a conformational polymorphism at the local level.

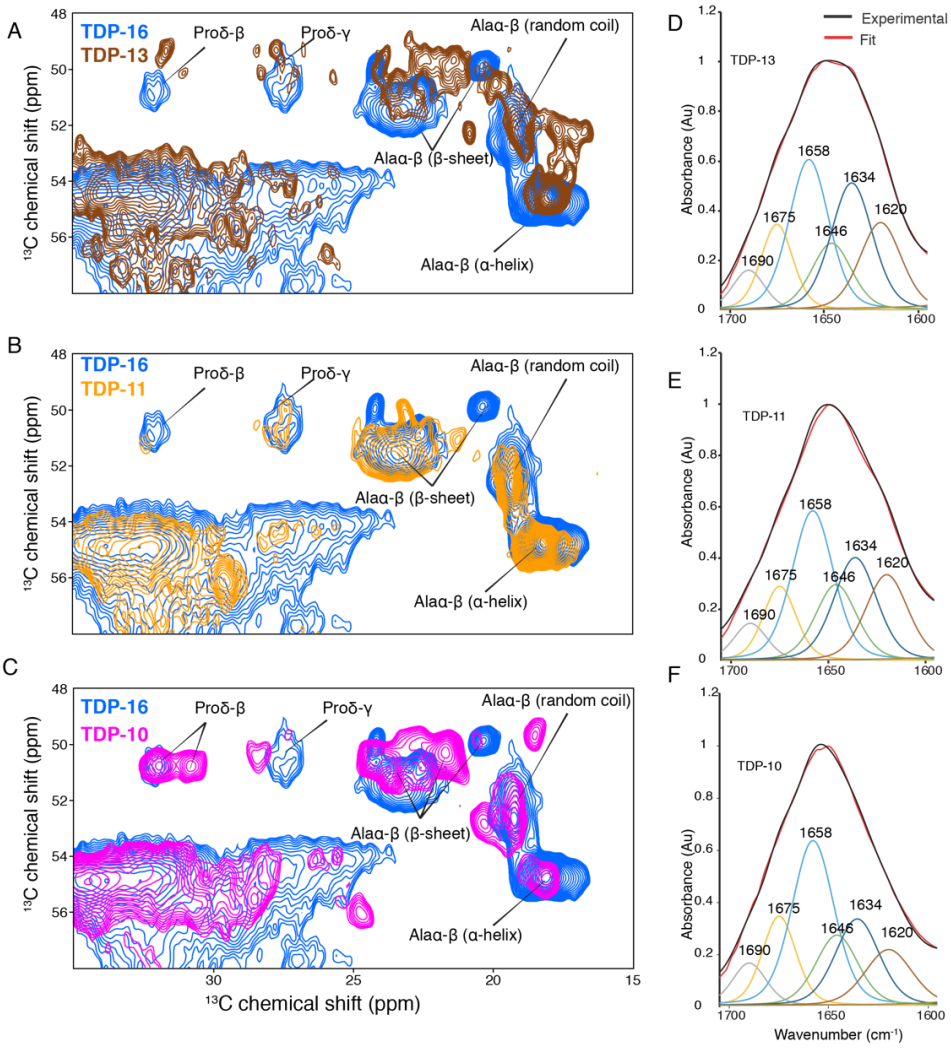



*Figure 3.14 An excerpt from the overlay of the 2D $^{13}C$–$^{13}C$ PDSD correlation spectra with a mixing time of 50 ms PDSD recoupling obtained on (U–$^{13}C$,$^{15}N$)-labeled samples of (A) TDP-13(brown), (B)TDP-11(orange), and( C)TDP-10 (pink) compared with TDP-16 (blue) recorded at 278 K and a $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency, a mixing time of 50 ms. Labels indicate characteristic alanine Cα–Cβ and proline Cδ–Cβ correlations and chemical shift-dependent secondary structure in brackets. FTIR spectra of (D) TDP-13, (E) TDP-11, and (F) TDP-10 aggregates in the amide I and II range showing the experimental (black) and theoretically fitted curve (red).*

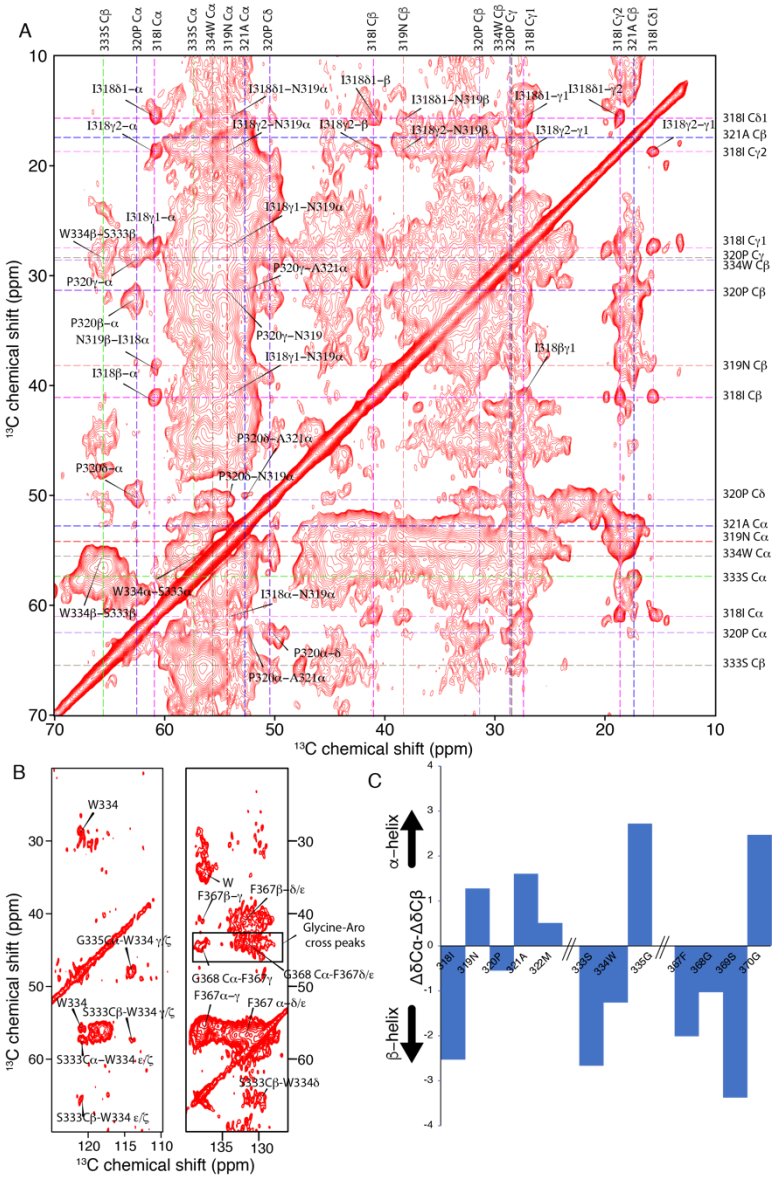



*Figure 3.15 Investigation of the amyloid core of TDP-13 amyloid aggregates. (A) Aliphatic and (B) aromatic region of SSNMR 2D $^{13}C$–$^{13}C$ correlation experiment of TDP-13 amyloid aggregates recorded at 278 K and a $^{1}H$ frequency of 800 MHz, 11 kHz MAS, and a mixing time of 150 ms for obtaining sequential (i.e., residue i to i ± 1) connectivity indicated by color-coded dotted lines. (C) The secondary chemical shift analysis of residues 318–322,333–334 and 367-370. Chemical shift-dependent secondary structure of amino acid sequence can be calculated by comparing experimentally obtained chemical shift with that of random-coil conformation[325]using the equation ΔδCα–ΔδCβ. Positive or negative values correspond to α-helical or β-strand conformation, respectively.*

Interestingly, the sequential correlations to the amino acid stretch I3I8-A322, S333-G335, and F367-S370. These amino acid stretches corresponded to the hydrophobic and the GaroS2 subdomain, respectively. The secondary chemical shift analysis comparing the experimentally obtained chemical shift values with random coil conformation of the amino acid stretch I3I8-A322, S333-G335, and F367-S370 in TDP-13 reveals a mixture of helical and

β-strand conformation (**Figure 3.15C**). Interestingly all the regions assigned here, I3I8-M322, S333-G335, and F367-S370 region, could contribute to a β–breaker according to their chemical shifts. This inference is in agreement with the SSNMR study on inclusion body seeded LCD fibrils and another one on the amyloidogenic region (311-360), indicating the I3I8-A322, S333-G335, and F367-S370 segments are located toward the end of β–strands and belong to the loop region connecting β–strands in the amyloid core.[156,399] Along with other published data, our results also convey the importance of the hydrophobic segment in the TDP-43 amyloid transition.[138,147,156,384,399] Although we mainly observe residues from the hydrophobic and GaroS2 regions, we cannot rule out the possibility of a Q/N rich region embedded in the broad signals as these residues share very close chemical shifts. We attempted to assign more residues and obtain structural information by proton-detected experiments on the perdeuterated samples (at 60 kHz MAS).[351,400] Unfortunately, the resolution of $^{1}$H detected spectra were also poor and insufficient (**Figure 3.16 A and B**) to conduct the assignments. The broad $^{1}$H peaks could arise from high polymorphism in samples and large residual similarity in the protein sequence (many Q and N residues or GaroS subdomains). Further on, the resolution in the $^{1}$H dimension will be improved by applying faster MAS frequencies in combination with larger magnetic fields.[355,401] To reduce the peak overlapping, we also tried to apply sparse labeling schemes[331]. The spectra obtained with 1,3 and 2-glycerol labeling schemes (**Figures 3.16C and D**) were insufficient to resolve sequential spin pairs for the assignments.

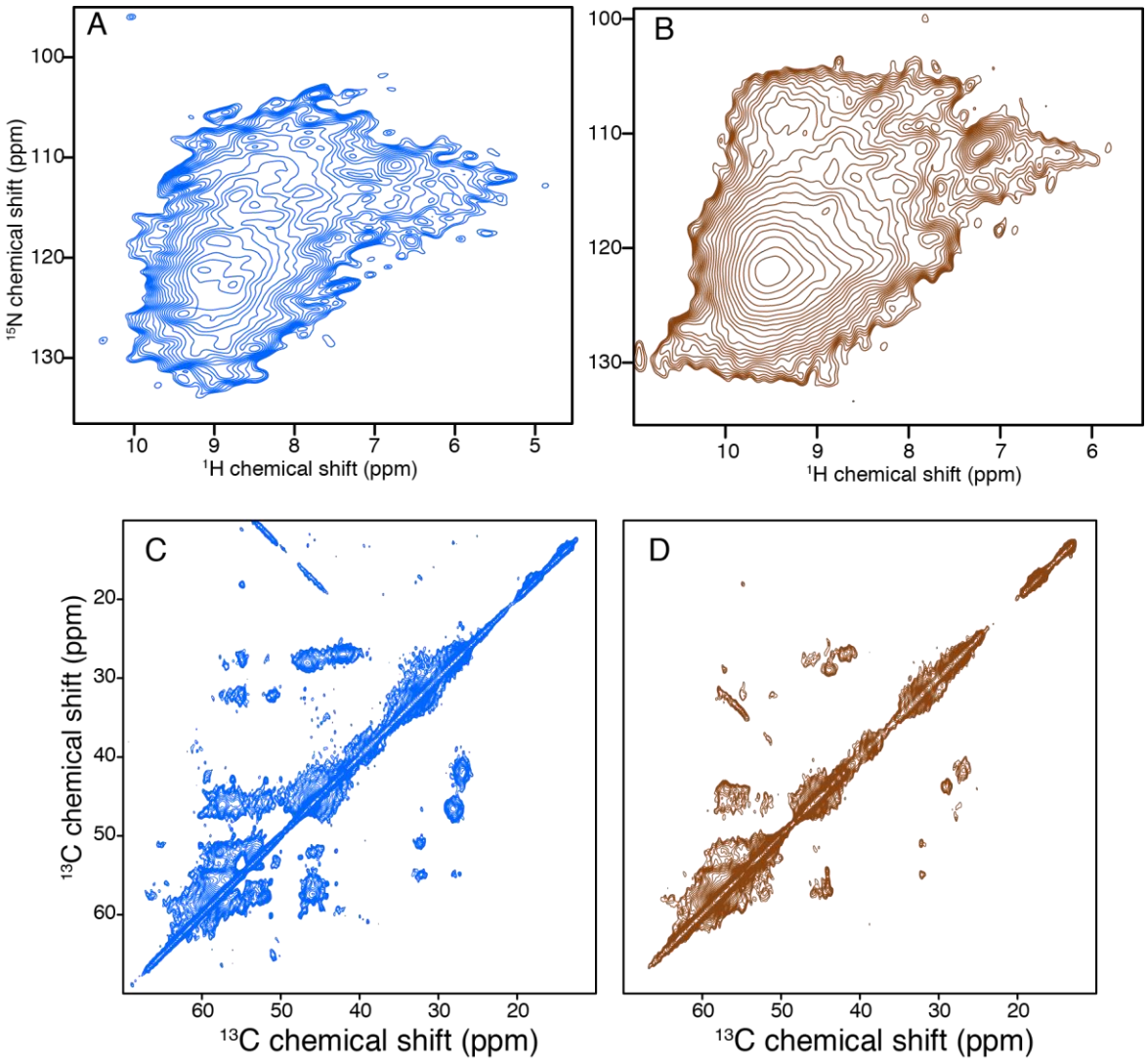



*Figure 3.16 2D hNH CP based spectra of (A) TDP-16 and (B) TDP-13 amyloid aggregates recorded at a $^{1}$H frequency of 600 MHz on the perdeuterated samples at 60 kHz MAS. SSNMR 2D $^{13}$C–$^{13}$C correlation experiment of 2-glycerol sample of (C) TDP-16 and (D) TDP-13 recorded at 278 K and a $^{1}$H frequency of 600 MHz, 11 kHz MAS, and a mixing time of 150 ms PDSD.*

### 3.2.9 Shortened C-terminal fragment, TDP-13 can adopt different polymorphs

As our investigation on shorter TDP-13 was going on, *Zhuo et al.* published an SSNMR based secondary structure on the amylogenic region 311-360. The protein was expressed and purified based on thioredoxin tag and assembled in 50 mM Tris-HCl, 300 mM NaCl, pH 8.0.[156] We tried to reproduce the same purification and aggregation protocols for our TDP-13 construct. The newly made NMR sample was investigated using the 2D $^{13}$C-$^{13}$C PDSD spectrum (shown in **Figure 3.17A**). The spectrum was overlaid with the reported chemical shifts on the TDP-13[156]. From these data, we concluded that TDP-13 adopts a different polymorph compared to the reported one by *Zhuo et al.* (**Figure 3.17A**). Interestingly, this new polymorph was also different from our previous one, as specified by the typical alanine Cα-Cβ correlation region (**Figure 17B**). Nonetheless, we cannot rule out the possibility of TDP-13 adopting an

identical amyloid core structure with minor conformational polymorphism as the chemical shifts and linewidths are very sensitive to the local chemical environment and dynamics. Strikingly, *Cao et al.* has reported that the CTD segment (311-360) can adopt three different polymorphs, conserving a common "dagger-shaped" fold. Our results indicate that the TDP CTFs can display structural plasticity, analogous to other reported pathologic amyloid fibrils of α–synuclein, Aβ, Tau, etc. depending on the aggregation conditions and sequence length.[71,78,402–404]

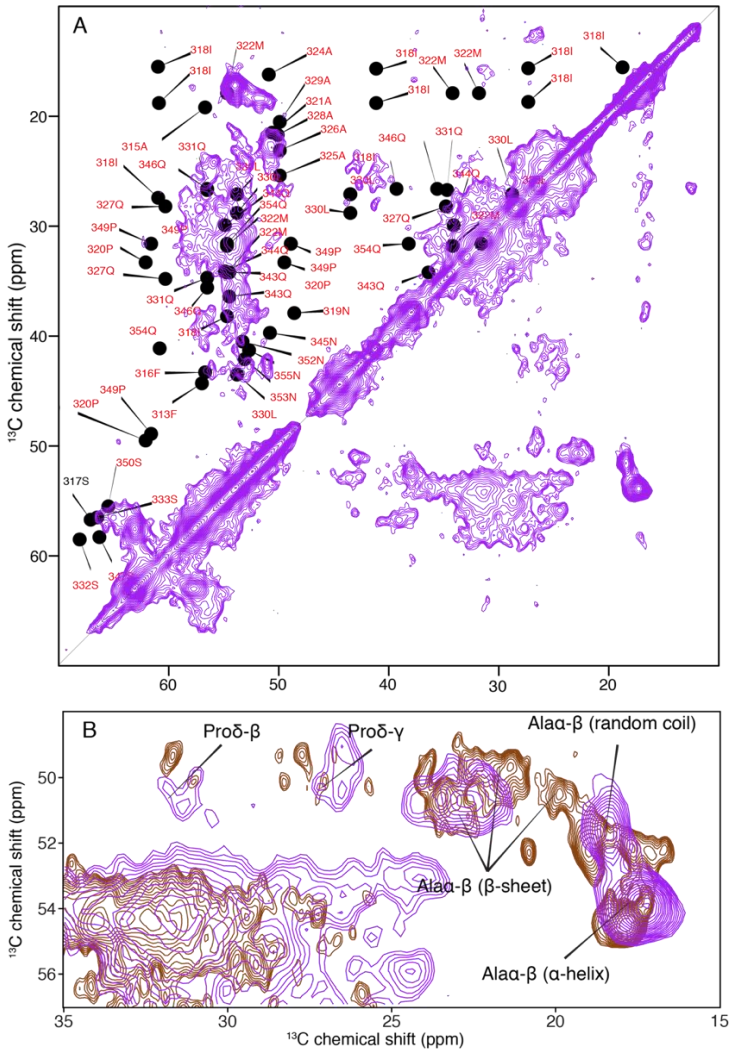



*Figure 3.17 Comparison of TDP-13 polymorphs. A) An overlay of chemical shifts obtained from Zhuo et al. .[156] (black dots) obtained for LCD segment (311-360) on to $^{13}C$-$^{13}C$ correlation spectrum of TDP-13 aggregated in identical conditions recorded at 278 K and a $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency, a mixing time of 50 ms. B) An overlay of $^{13}C$-$^{13}C$ correlation spectrum of TDP-$^{13}$ aggregated in identical conditions to Zhuo et al. [156] vs. 20 mM MES buffer pH 7.5.*

### 3.2.10 Shortened TDP-43 isoforms adopt a similar TDP-43 like amyloid fold

Next, we proceeded with further investigation of the recent developments in TDP-43 alternative splice isoforms in ALS pathology.[222] To evaluate the importance of TDP-43 splice isoforms, namely sTDP43-1 & -2 aggregation in ALS related pathology, we used a combination of various biophysical techniques, including SSNMR.[222] The two TDP-43 splice isoforms differed by only three amino acids and the entire glycine-rich domain (residues 281–414). The substituted a unique 18–amino acid C-terminus was not found in FL-TDP43 (**Figure 3.18A**).[387] Although the TDP splice isoforms do not form fibrillar aggregates like the full-length TDP-43, they show characteristic XRD patterns of amyloid fibrils displaying the characteristic diffraction pattern of two intense rings at 4.7 Å and 10 Å (**Figure 3.18B, C**). However, TDP-43 splice isoforms formed mesh-like aggregates. We assume that isoforms can form multi-valent interactions like LARKS with domains apart from the prion-like domain.[28,131,155] Such multi-valent interactions showed to form amorphous aggregates or fibril clumps.[406] Interestingly, these TDP splice isoforms exhibited the ThT fluorescence binding assays similar to other amyloid aggregates (**Figure 3.18D**).[375,407] These isoforms were intrinsically insoluble and readily aggregated even at low concentrations. The aggregates of TDP splice isoforms displayed amyloid-like properties. This finding was counter-intuitive as accumulating evidence suggested that the TDP amyloid core was present in the prion domain.[91,210] *In vivo*-studies suggests that overexpressed TDP-43 accumulates in the cytoplasm forming large inclusion bodies.[171,222,387]

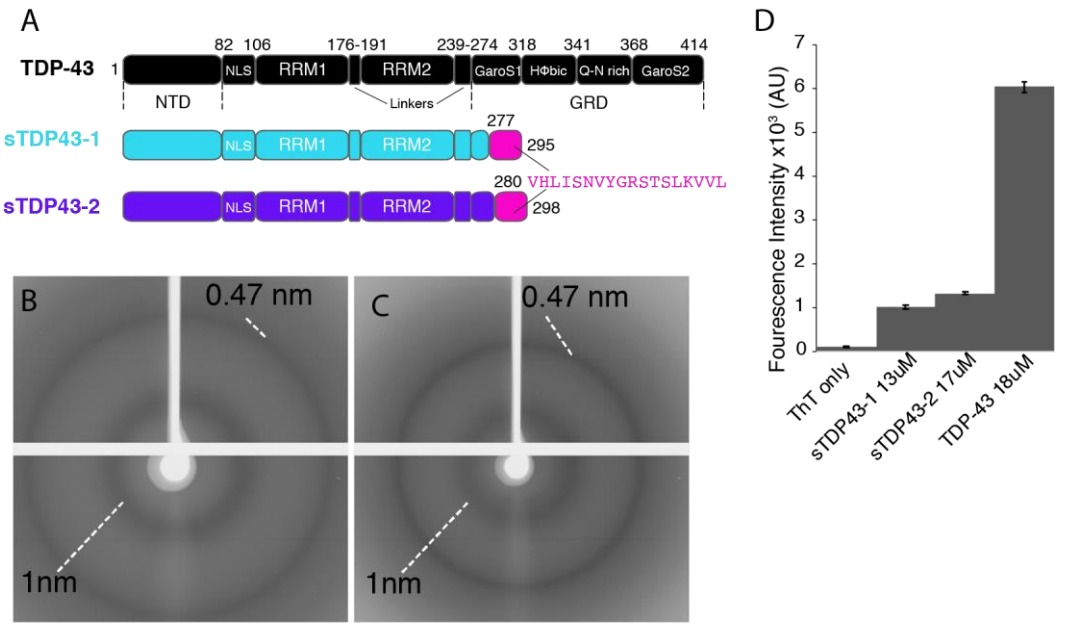



*Figure 3.18 A) Sequence construction strategy of TDP-43 splice isoforms sTDP43-1 and sTDP43-2. B) X-ray diffraction pattern of sTDP43 aggregates showing characteristic cross-β diffraction pattern in amyloid fibrils. Reflections at 4.7 Å and 10 Å are highlighted. C) ThT fluorescence assay on aggregates of TDP-43 and its splice isoforms.*

Further, we studied the TDP splice isoform's molecular architectures in their aggregated form using SSNMR. No significant Cα or Cβ resonances were detected in the sTDP43-1 INEPT spectra, implying only the sidechains are flexible, unlike sTDP43-2 or fl-TDP-43 (**Figure 3.19**). Also, no signals in the aromatic region were observed, implying the absence of monomers in the sample and a complete fibrillization. Broad SSNMR peaks of $^{13}C$ (linewidth of ~200-250 Hz, FWHH, ~1-1.25 ppm) indicated a central structural heterogeneity of TDP splice isoforms at the local level (**Figure 3.20A, B**). Surprisingly, a global assessment of aggregates of shortened TDP splice isoforms by SSNMR revealed that these aggregates adopt a TDP-43 like amyloid fold, represented by isolated Cα-Cβ cross-peaks of serines, threonines (**Figure 3.21A**), and alanines (**Figure 3.21B**). The spectral region encoding for alanine Cα-Cβ correlations contains mostly of β-sheet content as observed in TDP-43, TDP-35, and other pathological amyloid fibrils.[31,408] Moreover, the absence of the intense alanine signals in a helical conformation in the isoform aggregates indicates these signals originate from the TDP-43 LCD. Previously, alanines in the hydrophobic segment have reported being involved in transient helix essential in liquid-liquid phase separation of LCD.[113,138] Strikingly, no extra signals corresponding to the unique 18 amino acid in the isoforms were detected, given several valines, leucines, and serines are present in this sequence, which has isolated chemical shifts (**Figure 3.20A, B**). The absence of these residues implies their absence in the rigid amyloid core. However, the presence of these residues cannot be ruled out due to the possible broadening of the peaks below the detection limits.

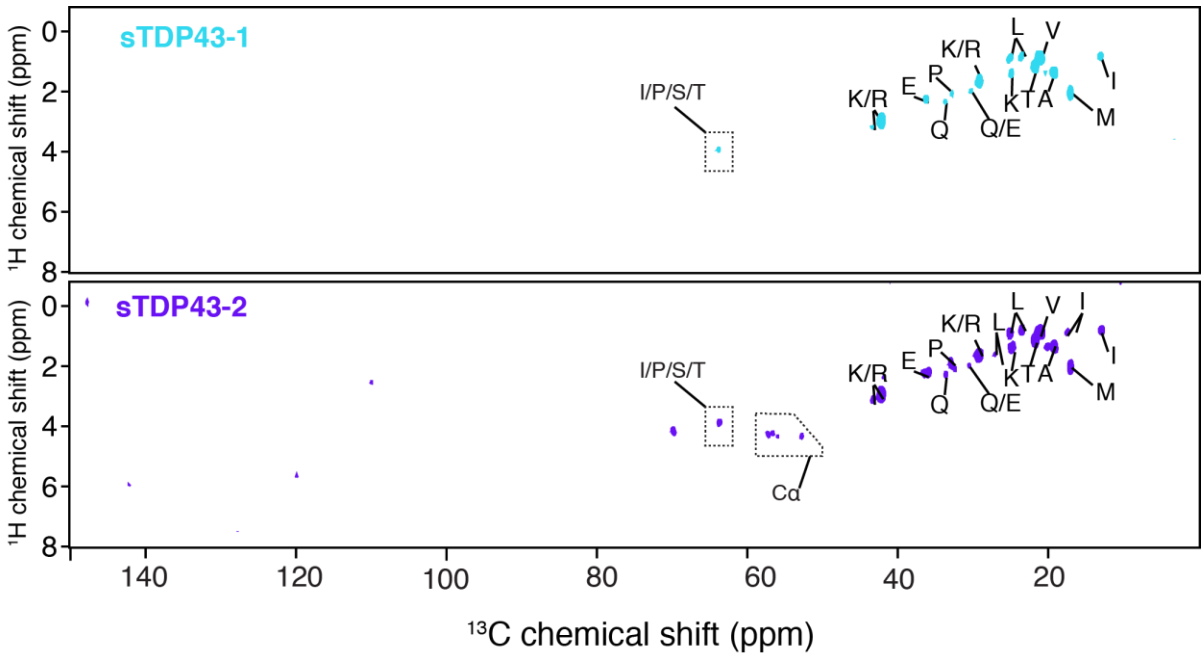



*Figure 3.19 SSNMR $^{1}H$–$^{13}C$ INEPT experiments of sTDP43-1 (cyan) and sTDP43-2 (violet) amyloid aggregates recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS frequency at 278 K showing tentative assignments corresponding to residues in the flexible region of the fibrillar assembly.*

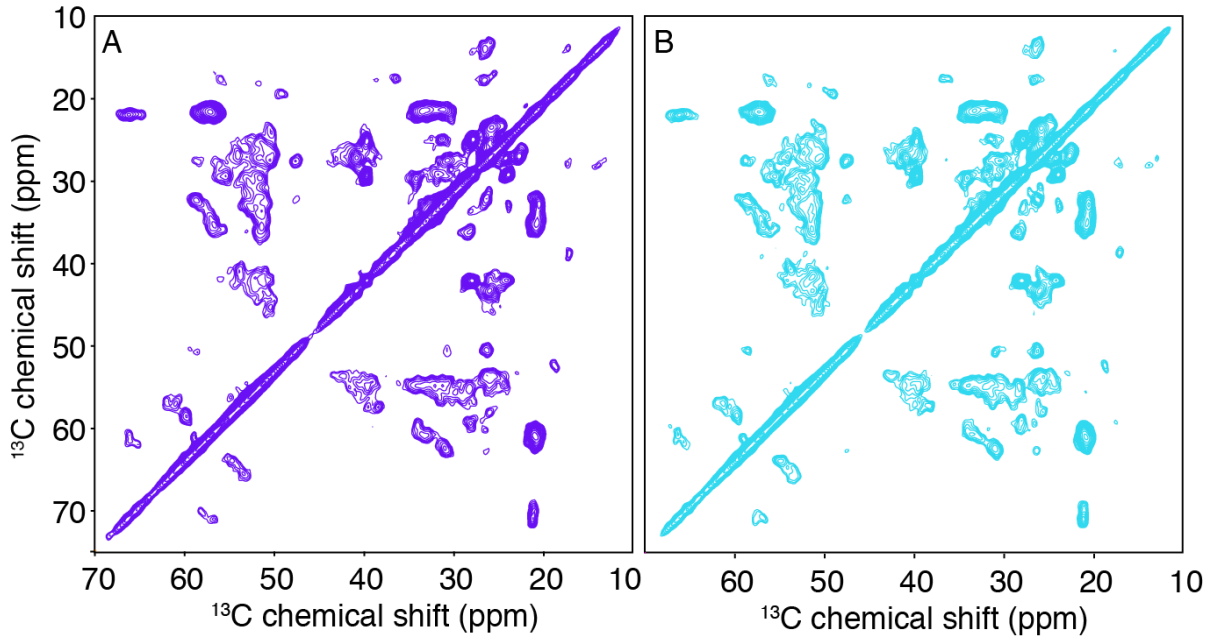



*Figure 3.20 Solid-state nuclear magnetic resonance $^{13}C$–$^{13}C$ correlation spectra of the different self-assembled [U–$^{13}C$, $^{15}N$]-labeled TDP fibrillar assemblies, the carbon-carbon magnetization transfer is achieved by a proton-driven $^{13}C$–$^{13}C$ spin diffusion. The aliphatic region of (A) sTDP43-1 (cyan) and (B) sTDP43-2 (violet) recorded at 278 K at a $^{1}H$ Larmor frequency of 600 MHz, respectively, an 11 kHz MAS frequency and a mixing time of 50 ms for obtaining intra-residue CC correlation.*

Further, ATR-FTIR analysis suggested that the TDP-43 splice isoforms have a similar signature (**Figure 3.21C, D**) but have a slightly higher amount of β-sheet content (~ 55%) than TDP-43 or TDP-35 (~ 50%) (Table 3). This β-sheet content is higher than what is observed in the previously solved structure of TDP-43 domains (~ 20–25%), reinforcing the idea that the amyloid transition is associated with misfolding NTD and RRM domains. These results suggest that TDP-43 native proteins rapidly may be converted to misfolded unstable aggregates that cannot grow into mature fibrils. Then TDP-42 can undergo structural reorganization to generate competent nuclei, acting as a template for later nucleation events.[14,42] This nucleated conformational conversion of the pre-formed oligomers to critical nuclei has been observed in many systems, including Aβ, α–synuclein, and IAPP.[41,43,44] A combination of nucleated conformational conversion or nucleated polymerization mechanism could contribute towards the formation of native-like aggregates as Cu-Zn superoxide dismutase, β2-microglobulin, lysozyme, etc. have been reported to show native-like aggregation mechanism.[14,46–48,409]

*Table 3.3 FTIR spectra of TDP-43 splice isoforms aggregates in the amide I and II range showing the experimental (black) and theoretically fitted curve (red).*

| Secondary structure element | Wavenumbers ($cm^{-1}$) | Percentage of structural elements (%) | |
|---|---|---|---|
| | | STDP43-1 | STDP43-2 |
| β-sheets // or anti// | 1620, 1635, 1690 | 55 | 54 |
| Random coil | 1646 | 10 | 10 |
| α–helices | 1658 | 24 | 25 |
| Turn | 1675 | 11 | 11 |

Altogether, our SSNMR data suggests that the TDP-43 can form amyloid fold irrespective of the LCD and still maintain a molecular conformation comparable to full-length TDP-43. It supports our previous results (**Figures 3.7 and 3.9**) as the TDP-43 ΔGaroS2 can form fibrils without the GaroS2 domain in the prion domain. This conclusion of amyloidogenic segments present in the NTD and RRMs agrees with previous studies, which displayed that peptides derived from RRMs can form aggregates, and NTD is playing a crucial role in aggregation.[131,171,186,210,377,410–412] Emerging studies suggest that the N-terminal TDP-43 fragments are found in cytoplasmic inclusions in ALS patients and contribute to ALS pathogenesis.[186,210] Generally, the RRMs are classified as crucial in misfolding and aggregation in the absence of LCD, regulating the solubility of RNA binding proteins. [171,377,410–412] Irrespective of RRMs, the N-terminus can enhance TDP-43 aggregation and toxicity.[171,377,412,413] In this analysis, we report for the first time that the aggregates of shortened TDP isoforms are amyloid in nature and can adopt a TDP-43 like amyloid fold even in the absence of the LCD.

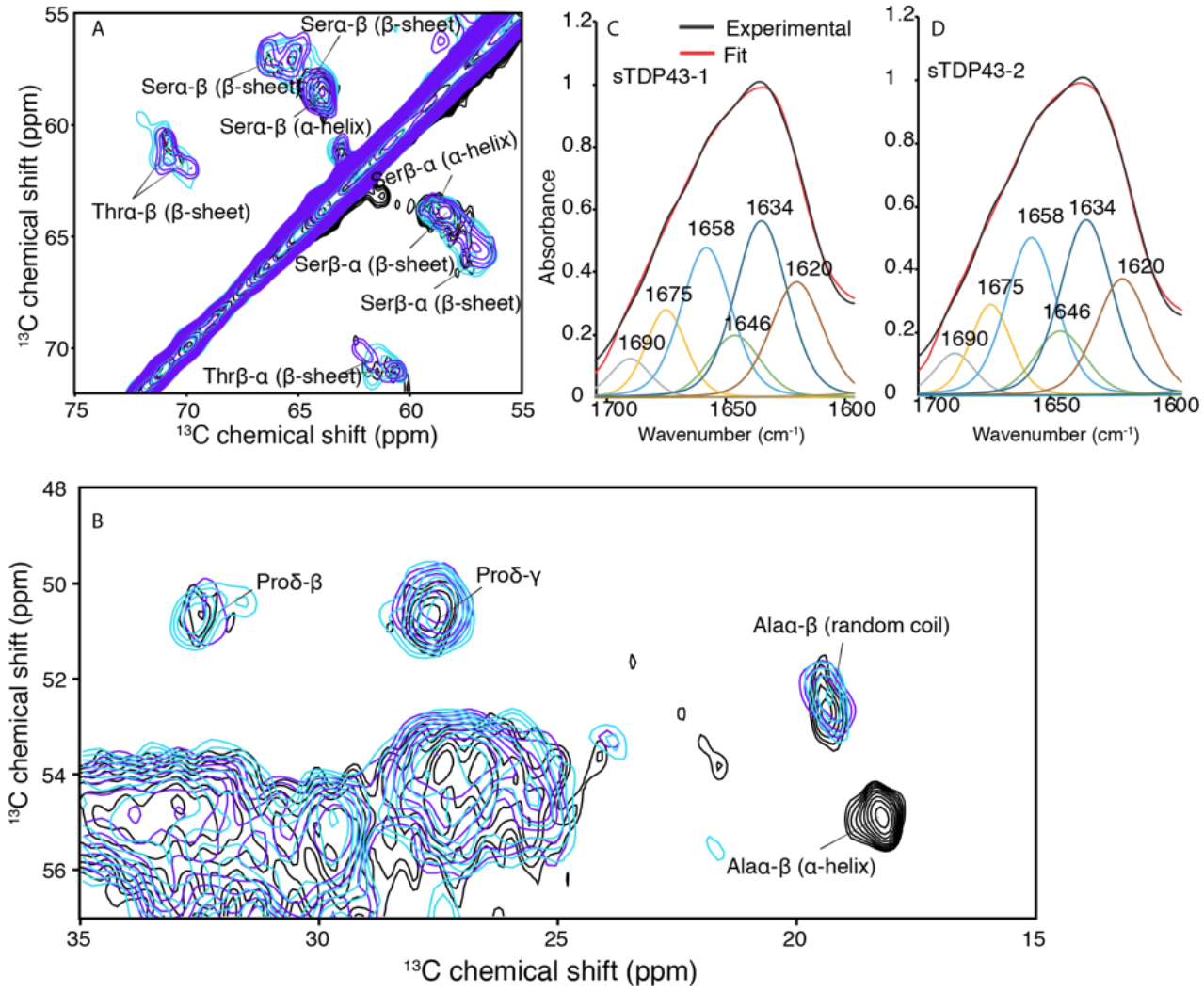



*Figure 3.21 Overlaid $^{13}C$–$^{13}C$ PDSD correlation spectra of sTDP43-1 (cyan), sTDP43-2 (violet) and TDP-43 (black) highlighted for the A) Ala Cα–Cβ B) Ser/Thr Cα–Cβ region. FTIR spectra of C) sTDP43-1 D) sTDP43-2 aggregates in the amide I and II range showing the experimental (black) and theoretically fitted curve (red).*

### 3.3 Discussion

In summary, we analyzed different fibril-forming TDP constructs using electron microscopy, X-ray diffraction, and solid-state NMR spectroscopy, TDP-43(1-414), TDP-35(90-414), TDP-16(267-414), TDP-43(1-366), TDP-13(300-414), TDP-11(300-399), TDP-10(314-414) and two TDP splice forms TDP43-1 and sTDP-43-2 to gain insights into the molecular organization of TDP-43 in the context of amyloid aggregation. All constructs except the two TDP splice forms form filaments but share features of amyloid aggregates observed from X-ray diffraction patterns and negative-staining EM analysis. *In vitro* preparation reported here for the TDP constructs; TDP-35, TDP-16, TDP-43, TDP-13, TDP-11, TDP-10 and two TDP splice forms TDP43-1 and sTDP43-2 are based on recombinant protein expression in E. coli. The EM images reveal morphologies comparable to those reported by *Johnson et al.* on TDP-43, using the recombinant heterologous expression by *Guo et al.* on synthetized TDP-43.[145,374] It has been reported that after a short period of aggregation, oligomeric species could be observed

by EM for TDP-43 and various mutants with a morphotype comparable to TDP-43 inclusions observed in neurons of ALS patients.[374,414,415] In the present study, unbranched fibrils observed after one week of aggregation indeed represent the final stage of *in vitro* TDP aggregation (i.e., mature fibrils).

X-ray diffraction analysis reveals rings at 4.7 Å and 10 Å, characteristic of pathological amyloid fibrils.[389] Similar patterns were observed for short peptides (6-11 amino acids) of TDP-43.[131,152,156] Crystal structures of these peptides have revealed steric zippers with various interfaces.[152] Although a structural comparison of the amyloid fold between short peptide segments and full length or truncated protein preparations should be considered with care, our data are in line with results reported by Eisenberg and coworkers, and the cross-β fold would dominate the amyloid architecture in TDP-43 filaments.[131,152]

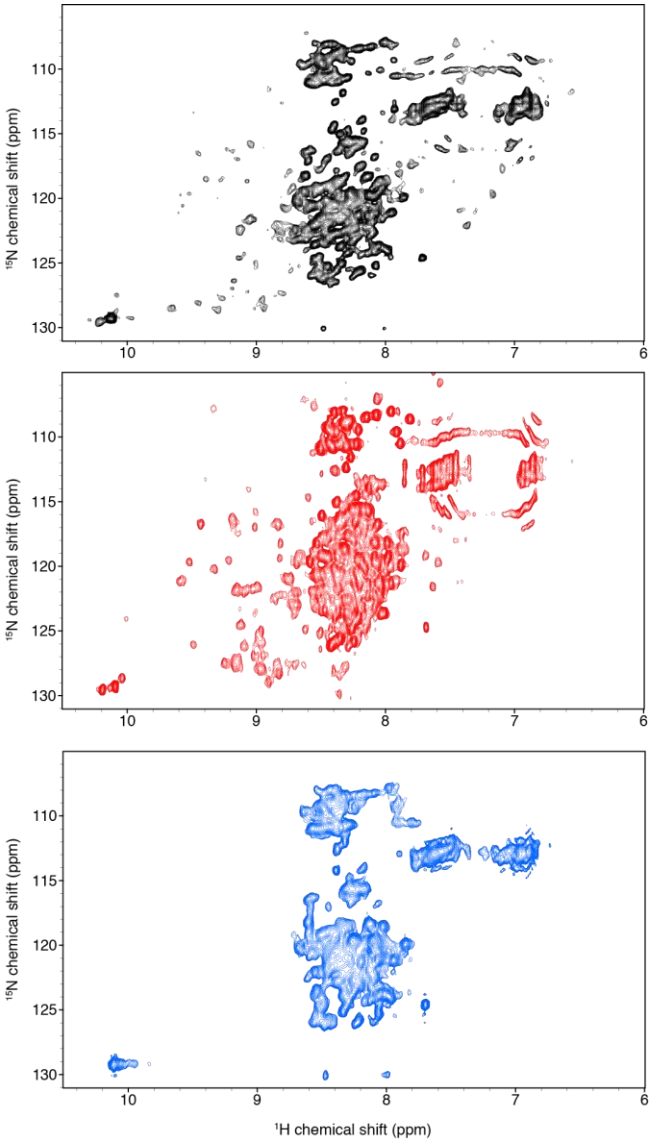



*Figure 3.22 Solution NMR $^{15}$N–$^{1}$H spectra show recombinant TDP-43 (black, 36 scans) and TDP-35 (blue, 128 scans) proteins exhibiting a wide range of $^{1}$H-$^{15}$N chemical shift dispersion while TDP-16 (blue, 512 scans) with a limited $^{1}$H-$^{15}$N chemical shift dispersion. Proteins were $^{15}$N labeled and prepared as described in the Materials and methods.*

Our SSNMR experiments highlight the main differences in the amyloid core conformation between the different TDP constructs. First, because the SSNMR fingerprints of TDP-43, TDP-35, and TDP splice isoforms sTDP43-1 and 2 superimpose well, it indicates that the rigid amyloid core is conserved between these protein sequences. The TDP-35 construct lacks the NTD, an 80-residue domain that adopts a globular fold when isolated.[107] And in TDP splice isoforms, the prion domain is substituted by a unique 18 amino acid sequence not found in WT-TDP-43.[222] Our SSNMR data on TDP-43 does not show the presence of sharp NMR signals and therefore exclude the presence of a rigid globular NTD decorating the amyloid core, as SSNMR demonstrated it for the CTD of Ure2p in full-length Ure2p fibrils.[416,417] Our data, however, cannot exclude the presence of the NTD in a molten globular state, as observed for the NTD of HET-s in full-length HET-s fibrils.[418] FT-IR analysis reveals an extensive amount of non β-sheet secondary structure in all constructs (~40-50%). Interestingly, the TDP-43 NTD is involved in the sequestration of full-length proteins into inclusions and is required for toxic aggregation *in vitro* and *in vivo*.[109,146,377] Fawzi and coworkers recently identified a key amino acid position (S48) disrupting the liquid-liquid phase separation, reinforcing the idea of an NTD strongly involved in the assembly process.[113] Strikingly, our SSNMR data indicate a similar amyloid core for the NTD-lacking TDP-35 construct, suggesting that the NTD is not involved in the rigid core of the mature fibrils. It has been demonstrated that the RRM2 domain also plays a role in the aggregation of TDP-43.[379,419] Solution NMR analysis of our TDP-43 constructs (**Figure 3.22**) shows that recombinant TDP-43 and TDP-35 proteins under non-denaturant conditions can refold. At the same time, TDP-16 exhibits a limited 1HN chemical shift dispersion, indicating an unstructured state. Because TDP-16 does not possess domains that can refold (i.e., the NTD and the RRM domains), it suggests a rather classical process going from an unstructured state to amyloid fibrils TDP-16. In contrast, aggregation of TDP-43 and TDP-35 would require the misfolding of the NTD and the RRM domains. These results can also be extrapolated to the aggregation mechanism of TDP splice isoforms.

Observations on TDP-16 and other shorter CTFs are quite compelling because SSNMR fingerprints reveal that the molecular architecture of these constructs might differ from TDP-35, TDP-43, and splice isoforms. Overall, our analysis reveals that the GaroS1, the hydrophobic, and the Q-N rich regions are involved in the common rigid core of TDP constructs except for the splice isoforms. The structural differences can be observed at different levels: (i) the local polymorphism reflected by the NMR linewidths is more critical in TDP-43, TDP-35, and TDP splice isoforms sTDP43-1 and 2 than the TDP CTD-fragments. Although several NMR signals superimpose between these protein constructs and share a high β-sheet content, a set of resonances associated with narrow and resolved NMR signals is only observed in the TDP CTD fragments. The excellent NMR spectral resolution of these resonances points to a high structural order for a part of the amyloid core of TDP CTD fragments, not present in the TDP-43, TDP-35, and TDP splice isoforms sTDP43-1 and 2 amyloid aggregates. (ii) The SSNMR resonance assignment indicates that the segment G380-I383 and A391-G396 are included in the highly ordered rigid core of TDP-16. These protein stretches belong to the so-called GaroS2 segment, rich in glycines, aromatic, and serine amino

acids. This observation underlines the importance of TDP-43 CTD in the formation of amyloid fibrils, which agrees with numerous studies implying the CTD role in the aggregation process (reviewed here[144]). Several ALS-associated mutations have been identified close to the 380-383 segment, e.g., S379, A382, I383, and G384. [183,420–422] Furthermore, the Huang group has recently identified W385 as crucial amino acid involved in the TDP-43 aggregation.[423] iii). The SSNMR resonance assignment of TDP-13 indicates that the segment I3I8-M322 and S333-G335 corresponding to the hydrophobic subdomain and the F267-S370 segment in the GaroS2 segment are included in the highly ordered rigid core. Interestingly, all these segments could contribute to a β–breaker according to their chemical shifts, indicated by the recent SSNMR studies on inclusion body seeded LCD fibrils, an amyloidogenic region (311-360), indicating the I3I8-M322, S333-G335, and F267-S370 segments are located towards the end of β–strands and may belong to the loop region connecting β–strands. [156,399] The aromatic ring of W334 and F367arisesy rigid, showing strong signals in the 2D $^{13}C$-$^{13}C$ correlation spectrum, suggesting that the aromatic side chains are involved in inter-residue interactions within the amyloid core. Also, several unassigned Tryp/Phe- glycine correlations can be observed in the aromatic region, suggesting they are part of the amyloid core. Interestingly, most of these aromatic-glycine interactions in TDP-13 belong to the GaroS2 subdomain (two phe-gly rich motifs at positions 367-368 and 396-403, and two Trp-Gly motifs at positions 384-386 and 411-413). This inference is in line with our previous findings that the GaroS2 domain is involved in the TDP amyloid core.[31] Moreover, at the time of writing this thesis, *Mompean et al*. have pointed out that the Trp-Gly and Phe-Gly motifs interactions as a driving force for the TDP-43 prion domain fibrilization by using solution-NMR and SSNMR analysis.[424] Interestingly, the aromatic residue W334 is shown to be a key residue that closely interacts with the sidechain carbonyl of Q343 and A325, A326, A329, and S332 and is involved in TDP-43 (311-360) liquid-liquid phase separation.[156] Also, the hydrophobic segment is shown to be involved in reversible liquid-liquid phase separation meditated by α–helical structure, and ALS-associated mutations in this region disrupt the phase separation by inhibiting interaction and helical stabilization.[138] Strikingly, a mutation of W334 with G residue disrupts phase separation without disturbing helical propensity.[207] The hydrophobic region is shown to be essential for TDP-43 aggregation, and helix-to-sheet structural transformation initiates TDP-43 aggregation.[147,156] Our results also convey the importance of the hydrophobic segment in the TDP-43 amyloid transition, harmonious with other published results.[138,147,156,384,399]

Our SSNMR analysis of the TDP amyloid core provides additional experimental evidence towards the importance of the C-terminal domain, which can adopt different amyloid conformation depending on its environment (in the full-length construct or isolated fragments). Also, different CTFs (TDP-16, TDP-13, TDP-11, TDP-10) adopt different polymorphs maintaining a common cross beta fibrillar architecture. Analogous to other reported pathologic amyloid fibrils, the TDP CTFs can display structurally distinct and self-propagating amyloid fibrils depending on the aggregation conditions and sequence length. Interestingly, TDP-13 itself readily engage alternative packing arrangements in response to different aggregation conditions.[71,78,402–404] Our results complement the recent cryo-EM study

of *Cao et al*., where they point out two peptide segments from CTD termed SegA (residues 311-360) and SegB (residues 286-331) to form multiple structurally different fibrillar polymorphs highlighting the role of the hydrophobic region.[155] Interestingly, SegA itself can form three different polymorphs (named sym, asym, and slow), and the hydrophobic region (321-340) adopts a "dagger-shaped" fold. Here, in our study, the sharp signals emerging from the GaroS2 region in TDP-16 and hydrophobic segment in TDP-13 indicate the structural homogeneity in this region and its importance in forming a well-ordered amyloid core of the mature fibrils.

The amyloid proteins are known to form different polymorphs in varying aggregation conditions. We compared the polymorphs reported by *Zhuo et al*. with our TDP-13 fibrils assembled in similar aggregation conditions.[156] Interestingly, our TDP-13 construct adopted a different polymorph based on the chemical shift analysis. Strikingly, the TDP-13 fibrils aggregated under these conditions displayed an extensive structural heterogeneity than the fibrils aggregated in our aggregation conditions (dialysis against 20mM MES buffer pH 7.5) based on the $^{13}C$ lineshapes comparison. However, the identified β-breaker regions (I3I8-M322, S333-G335, and F267-S370) of TDP-13 in our aggregation conditions are located at the exact positions reported by *Zhuo et al*., suggesting the TDP-13 fibrils could maintain a common amyloid core displaying a differential amount of polymorphism in different aggregation conditions. We also identified the key aromatic residue W334 participating in the amyloid core with strong signals originating from the aromatic rings suggesting its presence in the hydrophobic core via forming sidechain interactions supporting the inference of *Zhuo et al.* that the TDP amyloid core adopts a parallel in-register conformation.[156]

Further, we tested the amyloid nature of the TDP-43 splice isoforms displays in ALS pathology. The TDP-43 splice isoforms mesh-like or clumpy aggregates but displays amyloid features such as ThT fluorescence and a characteristic cross-β diffraction pattern observed by X-ray diffraction. The SSNMR analysis indicates that this splice form adopts a TDP-43 like amyloid fold, despite lacking the LCD. Nevertheless, the initiation and amplification of α–synuclein aggregation are shown to be inhibited by its homolog β–synuclein by competing for binding sites for secondary nucleation and acting as a natural suppressor for α–synuclein aggregation associated with Parkinson's diseases.[425] Similarly, the splice isoforms may alter the TDP-43 aggregation kinetics via secondary nucleation as they share identical amyloid fold. Interestingly, TDP-43 splice isoforms are observed in inclusion bodies that sequester with WT TDP-43 and display neuronal toxicity when overexpressed in mammalian neurons.[222] Also, the absence of LCD domains hinders the formation of reversible liquid-phase separation and accelerates sTDP43 aggregation, suggesting the protective nature of LCD against misfolding and aggregation at times of supersaturation. Hence, the intrinsically disordered region in RNA binding proteins may be evolutionarily designed to regulate protein solubility and phase behavior and protect against proteotoxic damage.[411] Indeed, some yeast RNA binding proteins like PUB1, PAB1, and Sup35, the LCD acts as environmental stress sensors and adapt the to the unstable environment via modifying protein phase behavior and display enhanced aggregation in the absence of LCD.[411,426–428] Extensive research is required for a better

understanding of splice isoform associated TDP-43 pathology apart from its loss of function associated with the lack of LCD.[170,429]

Also, following on the latest model for TDP self-assembly presented by the Fawzi group here[113] and considering recent structural studies of the isolated TDP NTD, our current working hypothesis consists of the NTD and the RRM domains mediating the assembly mechanism, notably through the NTD dimerization, which could bring CTDs close together to trigger protein aggregation, while not participating directly to the cross-β architecture of the mature amyloid core.[110–114] Because TDP-43, TDP-35, and sTDP43-1&2 amyloid aggregates exhibit the same SSNMR fingerprints, the truncation of the NTD domain or lack of LCD is not enough to significantly change the amyloid architecture. In contrast, the truncation of the NTD and the two RRM domains (i.e., the TDP-16 construct) profoundly impact the structure. Therefore, the difference in the SSNMR spectra of TDP-43/35 with LCD fragments could only be partially explained by the lack of NTD dimerization-mediated aggregation, suggesting a crucial role of the RRM domains. We observed that the TDP-16 and other shorter LCD fragments can still form amyloid fibrils without the NTD. Even though the CTD (TDP-16) includes the amyloid core and forms amyloid filaments, the NTD dimerization alongside putative interactions from RRM domains may play a decisive role in governing CTDs aggregation interface, therefore in constructing the structural architecture of the final amyloid assembly. Notably, the difference between the SSNMR fingerprints of TDP-16 with full-length TDP-43 stresses the importance of considering the presence (or not) of the NTD to question the amyloid aggregation from amyloid fibril structure. Because TDP-43 was very recently identified as functional amyloid involved in the regeneration of muscle through the formation of amyloid-like oligomeric assemblies called myo-granules, a careful structural examination of its different domains during the aggregation process, in the context of neurodegeneration or during the formation of myo-granules, is required.[430] Our work paves the way for understanding the role of the N-terminal and RRM domains in the TDP-43 amyloid transition, also raises the question about the protective nature of LCD in amyloid protein aggregation.

## 3.4 Materials and methods

### 3.4.1 Preparation of the pET24-TDP expression vectors

The three different TDP constructs encoding genes all contain a 6xHis-tag and TEV cleavage site in Nter and are flanked by NdeI and XhoI restriction sites. The constructs were ordered from Eurofins Genomics in a standard pEX-A vector. The TDP inserts and the expression vector pET24 (Novagen) were prepared for cloning by digestion with NdeI and XhoI and ligated using T4 DNA ligase (New England Biolabs).

### 3.4.2 Recombinant protein expression

E.coli strain BL21(DE3) pLysS was transformed with pET24-TDP vector and plated onto LB agar plates containing 30 µg/mL kanamycin initial culture; a 10 mL LB medium was

inoculated with a single clone and incubated at 37°C under shaking overnight. For expression and purification, 1L of LB medium was inoculated at 2% and incubated at 37°C while shaking at 220 rpm until the growth exponential log phase (OD600 0.8). Expression was induced with 0.75 mM IPTG at 30°C, 20h for TDP-43/sTDP43-1& -2 / TDP-35 and at 37°C, 3h30 for TDP-16 / TDP-43 ΔGaroS2/ TDP-16 I383V/TDP-11/TDP-10/TDP-13. For the production of $^{13}C$ /$^{15}N$ labeled proteins, the expression was carried out in an M9 minimal medium supplemented with 1g/L of $^{15}NH_4Cl$ and 2g/L $^{13}C_6$-D-glucose as nitrogen and carbon sources. Cells from the fresh initial 10 mL LB culture were harvested and resuspended in 100mL of labeled M9 medium and grown overnight at 30°C to accustom bacteria to the labeled nitrogen/carbon sources. 1 L of the minimal labeled medium was inoculated at OD600 0.2 and let grow at 37°C, 220 rpm until growth exponential log phase, and expression was induced with 0.75 mM IPTG. Cells were harvested by centrifugation (6000 g, 30 min, 4°C) and frozen (-80°C) until purification.

### 3.4.3 Protein purification

Frozen bacterial pellets were thawed and resuspended on ice in 20 mL of lysis buffer (Tris 50 mM, $^{15}$0 mM NaCl, pH 8), sonicated (30W, 3x45sec), and centrifuged at $^{15}$ 000 g for one hour at 4°C. The pellet containing the protein was then resuspended in 20 mL of lysis buffer supplemented with 2% Triton X-100 by vortexing vigorously, incubated at 37°C (20 min, 220 rpm), and centrifuged (50 000 g, 10 min, 4°C). The pellet was washed with lysis buffer (2 h, 37°C, 220 rpm), centrifuged (50 000 g, 10 min, 4°C). The pellet thus obtained was resuspended in 10 mL extraction buffer (50 mM Tris, 0.5 M NaCl, pH 8) and supplemented with dry guanidine hydrochloride until saturation. The suspension was incubated overnight at 60°C, sonicated on ice as above, and centrifuged (250 000 g, one h, 16°C). The supernatant was recovered and loaded on a HisTrap HP 5mL column (GE Healthcare) equilibrated in a binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Proteins were eluted from the column with a linear gradient of elution buffer (50 mM Tris, 0.5 M NaCl, 500mM imidazole, 8 M urea, pH 8). The fractions containing protein were monitored by UV absorbance at 280nm and analyzed on 12% tris-tricine SDS PAGE.

### 3.4.4 *In vitro* assembly of TDP fibrils

[U-$^{13}C$ /$^{15}N$]-labeled TDP samples were assembled *in vitro*. The pure protein recovered after affinity purification was dialyzed two times against 20 mM MES buffer pH 7.5 to remove urea and imidazole and was allowed to self-assemble for one week at room temperature. It was then extensively washed in water, centrifuged at 300 000 g, and transferred to the SSNMR rotor. The sample quantity per rotor is ~10 mg.

### 3.4.5 Solid-state NMR spectroscopy

SSNMR experiments were performed at a MAS frequency of 11 kHz on a triple resonance 4 mm MAS probe (Bruker Biospin) using 600 MHz and 800 MHz $^{1}$H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~278 K, and chemical shifts were calibrated using DSS as an internal reference. A cross-polarization contact time of 1 ms was used for a decoupling strength of 90 kHz using SPINAL-64. The rigid part of the fibrils was probed using two-dimensional $^{13}$C-$^{13}$C correlation spectra using a proton-driven spin-diffusion (PDSD) mixing scheme. Intra-residue correlations were probed using a PDSD mixing time of 50 ms for acquisition times of 20 ms (direct) and 7.5 ms (indirect), leading to a total experimental time of ~5 days. Sequential correlations were probed using a PDSD mixing time of 200 ms for acquisition times of 20 ms (direct) and 6.3 ms (indirect), leading to a total experimental time of ~7 days. Two-dimensional 1H-$^{13}$C INEPT were recorded using acquisition times of 20 ms and 6 ms in direct and indirect dimensions, respectively, for a total experimental time of ~18h. INEPT-TOBSY (INEPT-total through-bond correlation spectroscopy) experiment was recorded at a MAS rate of 8.33 kHz with TOBSY mixing time of 6 ms for acquisition times of 18.4 ms (direct) and 4 ms (indirect), leading to a total experimental time of ~3 days to establish homonuclear $^{13}$C–$^{13}$C connectivities. All two-dimensional spectra were processed using a sine-bell shift of pi/4 and plotted identically using the same contour level (50) and scaling factor (1.18) in SPARKY.

### 3.4.6 Solution NMR spectroscopy

The eluted samples were concentrated and diluted in PBS buffer, 20 mM DTT to reach 0.5 M urea, and concentrated again before loading on a gel filtration column Superdex 75 10/300GL or Superdex 200 10/300GL (GE Healthcare) for solution NMR experiments. The equilibration buffer used for size exclusion chromatography (SEC) were 20 mM sodium phosphate pH 7.5, 140 mM NaCl, 20 mM dithiothreitol for TDP-43, 20 mM sodium phosphate pH 7.5, 20 mM dithiothreitol for TDP-16 and 20 mM sodium phosphate pH 6.6, 300 mM NaCl, 20 mM dithiothreitol for TDP-35 in order to avoid rapid aggregation. The desalted samples were concentrated up to 40 µM for TDP-35 and TDP-16 and 10 µM for TDP-43 for the solution NMR experiments. All experiments were performed on uniformly $^{15}$N labeled protein samples in their SEC buffer supplied with 10% D20 (v/v) on Bruker 800 MHz or 600 MHz $^{1}$H Larmor frequency spectrometer equipped with a Triple resonance cryoprobe at 283K or 293K. The chemical shifts were calibrated using DSS or TSP. NMR data were processed with Bruker Topspin 4.0.6 and analyzed using SPARKY[327](T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

### 3.4.7 X-ray diffraction

Fiber diffraction patterns were measured at 4°C on a Rigaku FRX rotating anode X-Ray generator equipped with a Pilatus 200K hybrid pixel detector at the copper wavelength. The

concentrated hydrated samples were mounted in a MicroLoops™ from Mitegen on a goniometer head under the cold nitrogen flow. Each diffraction pattern corresponds to a 360° rotation along the phi axis with an exposure time of 720 sec after subtraction of a "blank" image of the same exposure time with only the loop on the goniometer head. The raw data are shown in **Figure 3.2G**.

### 3.4.8 Transmission Electron microscopy

One droplet of 2-fold serial dilutions of TDP constructs filaments in $^{15}$0mM acetic acid buffer was applied to glow-discharged 300 mesh carbon-coated copper grids for 1min, washed with water, stained with a 2% uranyl acetate (w/v) solution for 1min and dried under dark condition. Samples were observed in a FEI CM120 transmission electron microscope at an accelerating voltage of 120 kV under TEM low-dose mode at a magnification of 40,000X TDP-35 and 20,000X for TDP-16, TDP-13, TDP-11, TDP-10, TDP-43. TEM images were recorded using a Gatan USC1000 2k x 2k camera. For TDP-16 fibrils, a 100-fold dilution of the solid-state NMR sample was used.

### 3.4.9 Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR)

ATR-FTIR spectroscopy was performed on a Nicolet 6700 FT-IR spectrometer (Nicolet Instrument, Madison, WI) equipped with liquid nitrogen-cooled mercury–cadmium–telluride detector (ThermoFisher Scientific, San Jose, CA, USA), with a spectral resolution of 4 $cm^{-1}$ and a one-level zero filling. The hydrated TDP fibril samples were deposited (10 µL) on a germanium ATR crystal. The excess water was removed under nitrogen flow, and the spectrum was recorded with 200 scans. The spectra were analyzed to determine the secondary structure elements of each protein with an algorithm based on a second-derivative function and a self-deconvolution procedure (GRAMS and OMNIC softwares, ThermoFisher Scientific) to examine the number and wavenumber of individual bands within the spectral range of the amide I band (1700- 1600 $cm^{-1}$).

### 3.4.10 ThT fluorescence experiments

The monomeric TDP samples obtained from size exclusion chromatography (Superdex 10/300GL) were supplied with 15uM ThT. These samples were then transferred to 96 well corning plates (bottom transparent) with 100uL/well. The kinetics were initiated by placing the 96-well plate at 25°C and shaking at 100 rpm in a CLARIOstar®Plus plate reader and measured with different gains. The Th-T fluorescence measurements were made every 2 minutes (with excitation wavelength 449 nm and emission wavelength 482 nm) with shaking after each scan. An average of 10 values from the saturation phase is plotted as a histogram plot.

# Chapter 4 MOLECULAR MECHANISM OF HETEROTYPIC AMYLOID SIGNALING PROTEINS NWD2/HET-S INVOLVED IN PROGRAMMED CELL DEATH IN FUNGI

## 4.1 Introduction

Amyloids are proteinaceous aggregates rich in β–sheets and involved in many human cognitive diseases.[223,224] However, a class of amyloids called functional amyloids participate in various biological functions such as structure, storage, information, loss of function, and gain of function.[16,226,227] Amyloids display a unique self-templating ability, which allows them to propagate as prions in many transmissible diseases. [54,225,431,432] [Het-s] is one of the prions of the filamentous fungus *Podospora anserina* involved in the non-self-recognition process heterokaryon incompatibility, a form of programmed cell death (PCD).[231,433] The cell death is induced by a fusion of two agonistic allelic strains of the *het-s* gene termed *het-s(small s)* and *het-S* (capital S) by forming an abnormal contact line termed barrage.[433] The two agonistic alleles, het-s, and het-S encode the proteins HET-s and HET-S that are 95% identical.[239] The proteins share a two-domain structure containing a C-terminal prion forming domain (PFD, residue 218 to 289) necessary and sufficient for prion propagation and an N-terminal pore-forming toxin domain (1-227) termed HeLo domain (**Figure 4.1A**).[240,241] The PFD of HET-S is a cell death-inducing protein, activated by fusion with a prion-infected [Het-s] strain, regulated by amyloid templating. The Het-s/HET-S PFD is functionally interchangeable. However, the HET-s HeLo domain lacks the pore-forming activity due to specific mutations and propagate as a prion.[240,242] The regulatory PFD adopts a specific β-solenoid amyloid fold characterized by the stacking of two rungs of β–strands, each formed of 21 amino acid long pseudo-repeats, namely R1 and R2, connected by a flexible loop of 15 amino acids.[34,236,434] Once the PFD adopts the β-solenoid prion fold, the HeLo domain refolds, exposing an N-terminal transmembrane helix (TMH) and targets the cell membrane, where it exerts toxicity.[243,435] The fungal HELL (HeLo-Like) domain proteins are homologous to the N-terminal helical cell death execution domain of the Mixed Lineage Kinase domain-Like (MLKL) protein, which regulates mammalian necroptosis and the RPW8 and Rx-N CC-domains triggering cell death in plants suggests an evolutionary conservation.[436–438]

However, there exists another mode of activating HET-S, from which the [HET-S]/[Het-s] incompatibility system is evolutionarily derived from. And that is activation by a Nod-like receptor (NLR)is termed NWD2, encoded by the gene adjacent to het-S in the genome, associated with immune defense and programmed cell death pathways.[249,250] The NWD2/HET-S system belongs to a large group of affector-effector protein systems involved in cell death and host defense cascades based on prion-like or amyloid signaling mechanism.[251,439] Interestingly, the R0 motif of NWD2 and HET-s PFD can functionally interchange with PYRIN signaling domains based on prion-like polymerization in human NLR (NLRP3) and its cognate effector protein (ASC), providing direct evidence of NWD2 mediated prion transition of HET-S/s PFD.[251] Another example observed in mammals is where the RIP1 and RIP3 kinases form hetero amyloids via short amyloid motifs termed RIP homotypic interaction motifs (RHIMs) involved in programmed necroptosis.[234] Interestingly, the MLKL is a downstream effector of the RIP1/RIP3 complex in necroptosis.[440]

The NWD2 protein comprises a central NACHT nucleotide-binding and oligomerization domain, a C-terminal WD40 repeat domain, and an N-terminal effector motif (R0) (**Figure 4.1A**). R0 (NWD2$^{3-23}$) is homologous to the two elementary PFD pseudo-repeat of the HET-s prion (R1, R2) (**Figure 4.1B**). It is proposed that the ligand-induced oligomerization of NWD2 allows spatial clustering of R0 motifs adopting the HET-s like prion fold, which in turn templates the HET-S PFD and activates the HeLo domain.[249–251] Analogous to the NWD2/HET-S, several other gene clusters, including FNT1/HELLF with adjacent genes encoding an NLR and a putative cell death effector protein, with the N-terminal motif of the NLR protein and C-terminal end of the effector sharing homologous sequences have been identified in the fungal kingdom.[249,441] These inferences suggest that a signal transduction mechanism via prion-like polymerization is evolutionarily conserved across species.[251,436]

The R0 motif is necessary and sufficient for prion-inducing activity *in vivo* and is the shortest sequence to exhibit prion activity.[250] Homology modeling and SSNMR analysis of the R0 peptide suggests that this motif can adopt HET-S/s like fold. [249–251] Moreover, a distant homolog of HET-s from the fungus *Fusarium graminearum* called FgHET-s is shown to have a PFD, FgHET-s(218–289) that adopts a HET-s like prion fold and able to cross-seed HET-s(218–289) prion formation, suggesting the relevance of β–solenoid, prion-fold based cross seeding.[442] Interestingly, our group has recently pointed out, based on atomic-resolution structural studies of the HELLF prions that a distant homolog of HET-s showing only 17% of sequence identity with the PFD repeats of HET-S/s that shares a virtually identical amyloid backbone but not sufficient for cross-seeding. This suggests that the underlying templating mechanism could rely on the critical position of amino acid side-chains in the amyloid fold (unpublished results).[443] The amyloid heterotypic seeding has implications on neurodegenerative diseases.[444–446]

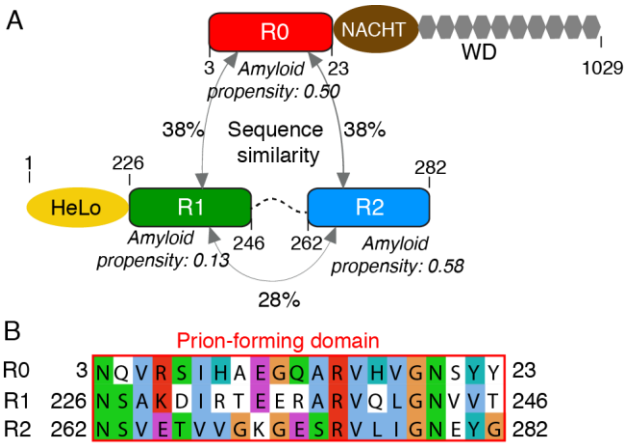



*Figure 4.1 Sequences and homology between NWD2 (3-23), called here R0, and HET-s (226-287), called here R1-R2. A) Domain organization of NWD2 and HET-s and compares the sequence similarity between R0, R1, and R2. Amyloid propensity calculated from Archcandy is given for each domain. B) Sequence alignment of R0, R1, and R2, prion-forming repeats are highlighted with a red box.*

Given the importance of the β–solenoid fold in fungal NLR functions, we sought to discover the structural basis of amyloid-based signal transduction by taking the NWD2/HET-S paradigm as a model system. While *in vivo* functional evidence, bioinformatics data, and existing structural evidence support the notion that R0 adopts a HET-s-like β-solenoid fold,

the molecular mechanism of the transmission of prion fold from NWD2 to HET-S remains elusive.[249–251] The consensus sequence of the NWD2 motif (R0) is very similar to the HET-s PFD pseudo repeats with conserved hydrophobic residues at the core positions (3, 6, 8, 12, 14, and 16), glycine residues at the β-arch positions (10 and 17), the asparagine ladders (1 and 18), and the positive charge in position 13 (**Figure 4.1B**). The sequence identity of R0 with individual repeats of HET-s PFD (R1 and R2) is respectively about 38% and 28% (**Figure 4.1A**). The R0 peptide has been shown to form amyloid fibrils irrespective of its ArchCandy score of 0.50, which is slightly lower than the fibril-forming threshold (0.58) predicted by ArchCandy.[250,447] This anomaly of ArchCandy probably because it is not explicitly designed to predict β-solenoidal structures and, by default, considers all predicted structures form in-register parallel β-arches.[447] However, ArchCandy successfully predicted the amyloid propensity of R2 (ArchCandy score 0.58), which readily forms amyloid fibrils, and R1 (ArchCandy score 0.12), which does not self-assemble invitro (**Figure 1A**).

## 4.2 Results

To determine the structural basis of the prion-based signal transduction in the NWD2/HET-s paradigm, we herein created several chimeric constructs, namely R0-R0, R0-R1, and R0-R2. We employed a combination of SSNMR, molecular modeling, and other biophysical techniques. Magic-angle spinning (MAS) SSNMR is shown to be a powerful technique for the structural studies of non-crystalline and fibrous amyloid aggregates or in combination with complementary biophysical techniques.[70,72,294,345,346,388]

### 4.2.1 Structural investigation and infectivity of the chimeric R0-R0 functional amyloids

The structural characterization of amyloid fibrils by SSNMR requires high quality isotopically labeled protein samples. These samples can be obtained from a recombinant bacterial expression system supplied with isotopically labeled nutrients (described in **Materials and Methods**). The R0 motif was too short and detrimental to bacteria in recombinant protein expression, and isotopic labeling of synthetic R0 peptide was costly. To overcome this issue and explore molecular architecture in the fibrillar form, we relied on a recombinant chimeric construct made of two R0 repeats connected by the flexible loop in HET-s PFD (called "loop" here) (**Figure 4.2A**). The R0-R0 chimeric construct spontaneously aggregated in vitro, similar to the conditions of the HET-s PFD.[243] The R0-R0 chimeric fibrils appeared as straight, unbranched fibrils of ~5 nm in diameter comparable with synthetic R0 peptide aggregates and HET-s (218-289) fibrils as observed by electron microscopy **(Figure 4.2B)**.[250,442] The R0 peptide is regarded as the shortest motif to exhibit [Het-s]-infectivity and functionally substitute HET-s PFD for prion formation.[250] Our previous SSNMR study using synthetic peptides had shown that the R0 motif adopts a rigid HET-S/s like fold.[250] However, the stabilizing effects that lead to homotypic R0 stacking in the presence of unfavorable side-chain electrostatic repulsion remain unclear in the absence of a high-resolution structural model.

Further, we analyzed the prion infectivity of these fibrils by barrage assay. The chimeric R0-R0 amyloids transfected into [Het-s*] strains resulted in prion conversion, indicating the presence of a functional HET-s-like prion fold (**Figure 4.2C**).[250] To obtain the R0-R0 functional amyloid structural model, we probed the residues contributing to the rigid amyloid core using dipolar-based polarization transfer (cross-polarization). We performed 2D $^{13}C$–$^{13}C$ spectroscopy with a short proton-driven spin diffusion (PDSD) mixing time of 50 ms

to correlate intra-residue carbon atoms **(Figure 4.2D).** About 23 spin systems (amino acid residues) were identified as contributing to the rigid amyloid core. We estimate an average $^{13}C$ linewidth of about ~0.5ppm, comparable to the linewidth of HET-s (218-289) amyloids, indicating a well-organized amyloid core.[393] A unique set of resonances in the PDSD spectrum indicates the lack of conformational polymorphism in R0-R0 amyloid fibrils. Based on the HET-s(218-289) amyloid model, the residues that stack on top of each other in the two rungs of the β–solenoid model in R0-R0 are the same and share the same chemical environment, results in a single set of resonances for each corresponding residue. Hence they display excellent spectral resolution.[34]

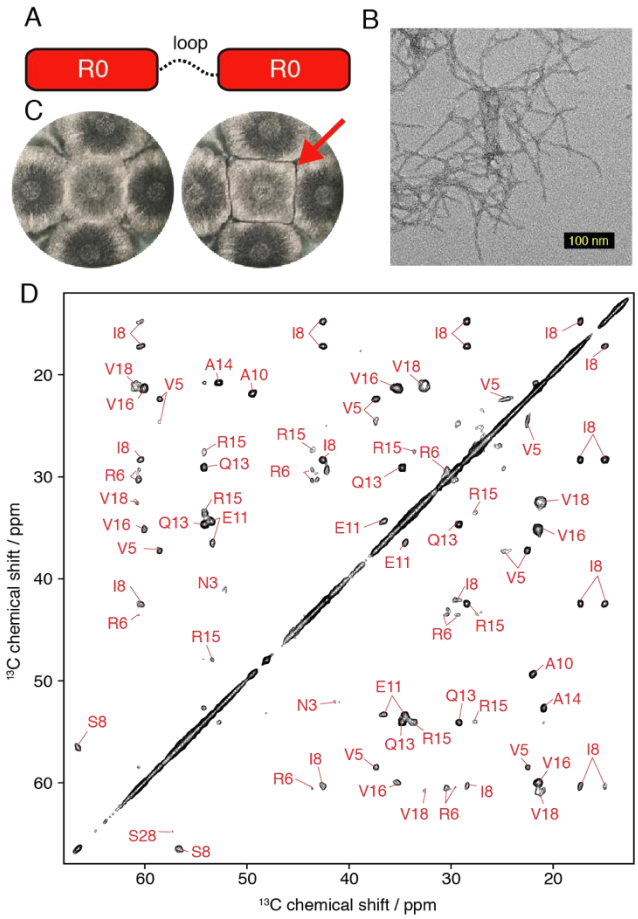



*Figure 4.2 Structural investigation and infectivity of the chimeric R0-R0 functional amyloids. A) Schematic representation of R0-R0 chimeric construct strategy. Each monomer is made of two R0 repeats connected by the R1-R2 loop. B) Negatively stained electron micrograph of R0-R0 functional amyloid fibrils, scale bar, is 100 nm. C) Transfection essays of [Het-s*] with R0-R0 functional amyloid fibrils. Four transfected strains are tested on each plate, and the tester strain in the center of the plate is [Het-S]. R0-R0 fibrils convert strains to the [Het-s] phenotype (detected by the formation of a barrage reaction to the central [Het-S] tester, highlighted by a red arrow). On left: Control test. D) Two-dimensional solid-state NMR $^{13}C$-$^{13}C$ correlation experiment of R0-R0 functional amyloid fibrils recorded at a $^{1}H$ frequency of 800 MHz, 11 kHz MAS frequency with a PDSD mixing time of 50 ms at 278 K. The residue assignment is shown in red.*

### 4.2.2 Structural analysis of the chimeric R0-R0 functional amyloids

Owing to the excellent spectral resolution of the R0-R0 functional amyloids, we further performed 2D SSNMR on selectively $^{13}C$ labeled samples (2-$^{13}C$-glycerol and 1,3-$^{13}C$ -glycerol samples) with a PDSD long mixing time (400 ms) to detect sequential (i.e., residue i to i±1) and long-distance connectivity, revealing the amyloid core-forming residues N3-Y23 (**Figure 4.3A, B**). By comparing experimentally obtained chemical shifts with random coil conformation, we probed the secondary structure conformation adopted by the amino acids using the so-called secondary chemical shift analysis, based on the value, ΔδCα–ΔδCβ. Positive and negative values correspond to α-helical and β-strand conformation, respectively.[325] The secondary structure analysis of R0-R0 amyloids displayed magnitude and patterns similar to HET-s (218-289) amyloids, indicating the presence of 8 β–strands (**Figure 4.3C).**[34] Apart from two alanine residues, the R1-R2 loop was not "SSNMR visible" due to high flexibility, suggesting that the R1-R2 loop does not contribute to the R0 amyloid rigid core.

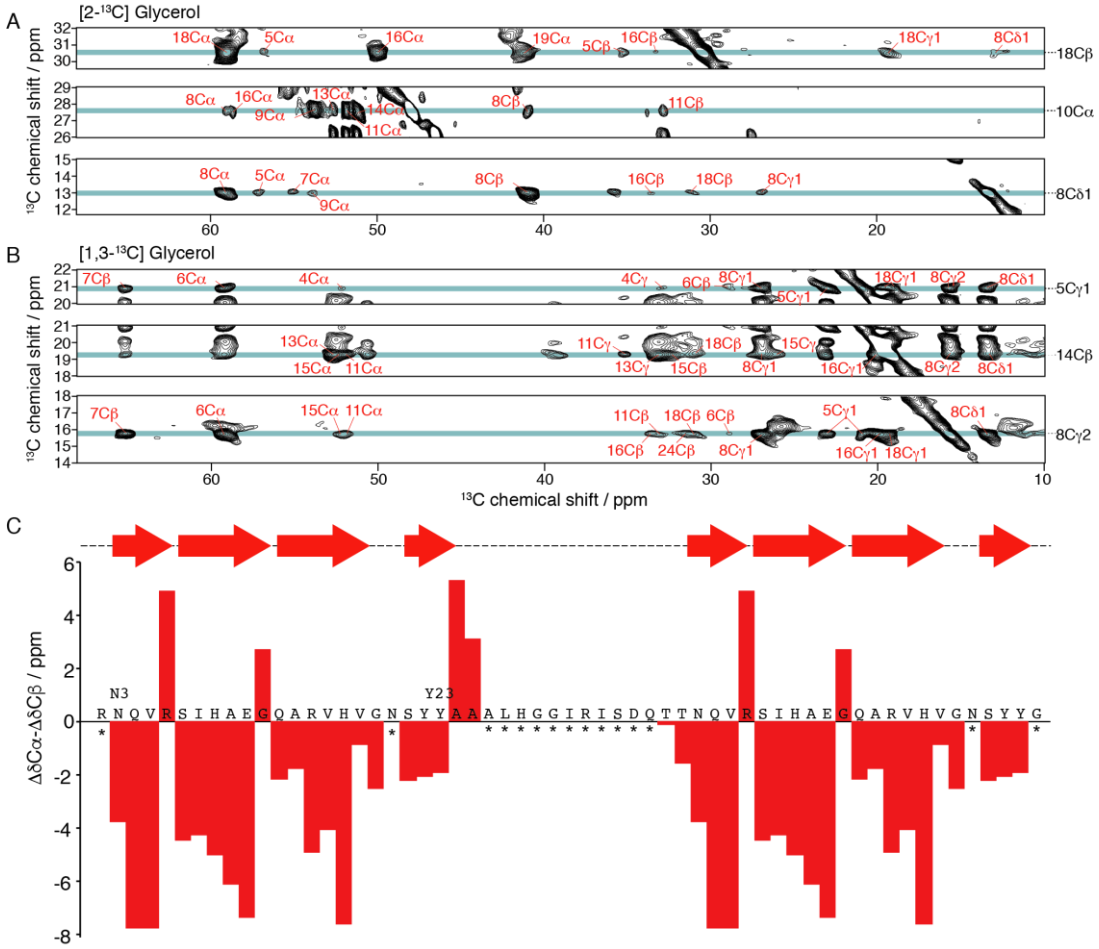



*Figure 4.3 Secondary structure of the chimeric R0-R0 functional amyloids. A) Two-dimensional solid-state NMR $^{13}C$ -$^{13}C$ correlation experiment of (2-$^{13}C$)-glycerol labeled R0-R0 functional amyloids, inter-residue correlations are highlighted for Ile8, Ala10 and Val18 . B) Two-dimensional solid-state NMR $^{13}C$ -$^{13}C$ correlation experiment of (1,3-$^{13}C$)-glycerol labeled R0-R0 functional amyloids, inter-residue correlations are highlighted for Val5, Ile8, Ala14. C) Solid-state NMR secondary chemical shift analysis of R0-R0 functional amyloids, negative or positive*

*values indicate β-strand or α-helix conformation, respectively. The eight identified β-strands are displayed.*

Further, the 2D NMR on sparsely labeled (2-$^{13}$C-glycerol and 1,3-$^{13}$C -glycerol samples) protein samples significantly increased spectral resolution and helped to avoid ambiguity in assignments, and provided many structurally relevant distance restraints.[331] The R0-R0 structure was successfully determined with ARIA {XX} on the basis of 453 assigned distance restraints, of which 100 were long-range (|i-j| > 4), 159 short-range (|i-j| ≤ 1) and 104 medium-range (1< |i-j| < 5) distance restraints (**Figure 4.4A).**

The average root means square deviations (RMSD) to the mean structure of the 10 lowest-energy conformers (200 calculated structures) for the backbone atoms, and heavy atoms are estimated to be 0.53Å and 1.08 Å, considering only the rigid amyloid core. The structure was derived based on the identified β-strands, and through similarity comparison with the published structures of HET-s(218-289) (PDB ID 2RNM, 2KJ3),[34,236] intra- and inter-monomer hydrogen bond restraints between β-strands (β1-β5, β2-β6, β3-β7, β4-β8) including in the calculation. Further structure calculation statistics are given in table 1. The R0-R0 functional amyloid adopts a β–solenoid molecular architecture composed of 8 β-strands (β1 to β8) separated by the disordered R1-R2 loop region. The first three β–strands per layer define the triangular hydrophobic core (6 β–strands per molecule β1–β3, and β5–β7), stabilized by the hydrophobic side-chains intertwined inside the amyloid core, forming a dry face (devoid of water). The non–β-sheet chemical shift of R6 and G12 can be explained by their positions at the β-arc. The two asparagine ladders at positions 3 and 20, the glutamine ladder at position 13, contribute to the fibril stability via side-chain hydrogen bonds. All these side-chain interactions, along with the backbone hydrogen bonds between continuous layers, contribute to the stability of the R0-R0 construct and its propagation as long, mature amyloid fibrils **(Figure 4.4B**).

To gain further structural insights into the β–solenoid fold designed by R0 repeat, we compared the R0 amyloid structure with the pseudo-repeats (R1, R2) of HET-s PFD in its fibrillar state (PDB 2RNM,2KJ3)[34,236] **(Figure 4.4C**). We observed that R0 forms a unique three residue β–arch between β2- and β3 stabilized by the glutamine ladder to accommodate the two alanines (A10 and 14) inside the hydrophobic core rather than a sharp turn in HET-s fold stabilized by the hydrogen bonds between T233 and S273. The homotypic parallel in-register stacking of the R0 motif was unlikely as per the prediction by ArchCandy calculations considering the electrostatic repulsion of three charged residues aligned on top of each other (2 arginines and one glutamate) (**Figure 4.4D**). However, the R0 repeat bears less charged residues than HET-s repeats; seven charged residues in R1 and four in R2 repeats, as depicted in **Figure 4.5D**. In HET-s (218-289), the charged residues in the pseudo repeats are located at conserved positions, arranged one on top of the other such that the charge complementarity could result in the formation of stabilizing salt-bridges (K229-E265, E234-K270, and R236-E272). Similar to HET-s (218-286) amyloids, the charged residues in R0 are located in the β–arch regions facing outside and largely solvent accessible, making the R0 self-assembly thermodynamically feasible.[34] Additionally, the reduced charge content in R0 compared to other repeat motifs is more compatible with homotypic polymerization. Interestingly, the R1 repeat with a considerably larger number of charged residues (7 residues in total) than R0 and R2 (3 and 5 respectively) did not form any self-assembly in-vitro (Data not shown). An R1 homotypic parallel stacking with charged residues arranged on top of each other could lead to severe electrostatic repulsion, making R1 self-assembly thermodynamically unfavorable.

This inference is in line with its low ArchCandy score of 0.12 of R1, given the fibril-forming threshold is 0.58.[250,447]

### 4.2.3 R2 peptide adopts a HET-S/s-Like β-solenoid Fold

Interestingly, the R2 peptide readily self-assembles in-vitro in similar conditions to the R0 peptide. R2 self-assembly wad was observed as long unbranched filaments of ~5nm diameter comparable with HET-s PFD fibrils by electron micrograph **(Figure 4.5A).**[250] We performed two-dimensional solid-state NMR on $^{13}$C-labeled V264/G269/I277 R2 amyloid fibrils to overview its molecular architecture. The V264/G269/I277 residues are located at the key position in the triangular hydrophobic core. A sharp $^{13}$C linewidth of ~0.5–0.7 ppm at 7 kHz magic-angle spinning (MAS) (on a 500 MHz proton frequency spectrometer) suggests a highly ordered 3D structure, comparable to HET-s(218–289) and R0 amyloids.[250,393] Further, we established homonuclear connectivity between V264/G269/I277 residues with a $^{13}$C-$^{13}$C correlation experiment with PDSD long mixing time (400 ms), inferring R2 adopts a rigid HET-S/s-like fold **(Figure 4.5B)**. The self-assembling property of R2 is in good agreement with ArchCandy predictions.[250,447]

*Table 4.1 R0 homo amyloid SSNMR structure calculation statistics.*

| **Number of restraints (per monomer)** | |
|---|---|
| Unambiguous NMR distance restraints | |
| ↳ Short-range (\|i-j\|≤1) | 159 |
| ↳ Medium-range (1<\|i-j\|<5) | 104 |
| ↳ Long-range (\|i-j\|≥5) | 100 |
| ↳ Ambiguous | 62 |
| ↳ Inter-monomers | 28 |
| ↳ Total | 453 |
| Hydrogen bond restraints (intra \| inter) | 28 \| 28 |
| Dihedral angle restraints (phi \| psi) | 30 \| 30 |
| **Restraints violation (per monomer)** | (mean ± SD) |
| NMR distance restraints | |
| ↳ RMS of violations (Å) | 0.01 ± 0.00 |
| ↳ Number of violations >0.3Å | 0.56 ± 0.65 |
| ↳ Number of violations >0.1Å | 13.52 ± 1.94 |

| | |
|---|---|
| Hydrogen bond restraints | |
| ↳ RMS of violations (Å) | 0.00 ± 0.00 |
| ↳ Number of violations >0.1Å | 0.00 ± 0.00 |
| Dihedral angle restraints | |
| ↳ RMS of angle violations (°) | 0.23 ± 0.052 |
| ↳ Number of violations >5° | 0.54 ± 0.858 |
| **Deviation from ideal geometry** | |
| RMS for bond lengths (Å) | 0.004 ± 0.0001 |
| RMS for bond angles (°) | 0.44 ± 0.015 |
| RMS for improper (°) | 1.44 ± 0.096 |
| **Ramachandran plot statistics (%)** | |
| most favored regions | 86.90 ± 3.06 |
| allowed regions | 12.80 ± 2.97 |
| generously allowed regions | 0.30 ± 0.63 |
| disallowed regions | 0.00 ± 0.00 |
| **RMSD from the average structure (Å)** | |
| All residues | |
| ↳ Backbone atoms | 5.05 ± 0.90 |
| ↳ Heavy atoms | 5.74 ± 0.91 |
| Prion forming domain (10-30,46-66) | |
| ↳ Backbone atoms | 0.53 ± 0.07 |
| ↳ Heavy atoms | 1.08 ± 0.18 |

a:-one hydrogen bond is encoded by two distance restraints (dH..O and dN..O), b:- from PROCHECK [343], c:- residues with S(phi) + S(psi) > 1.8. [448]

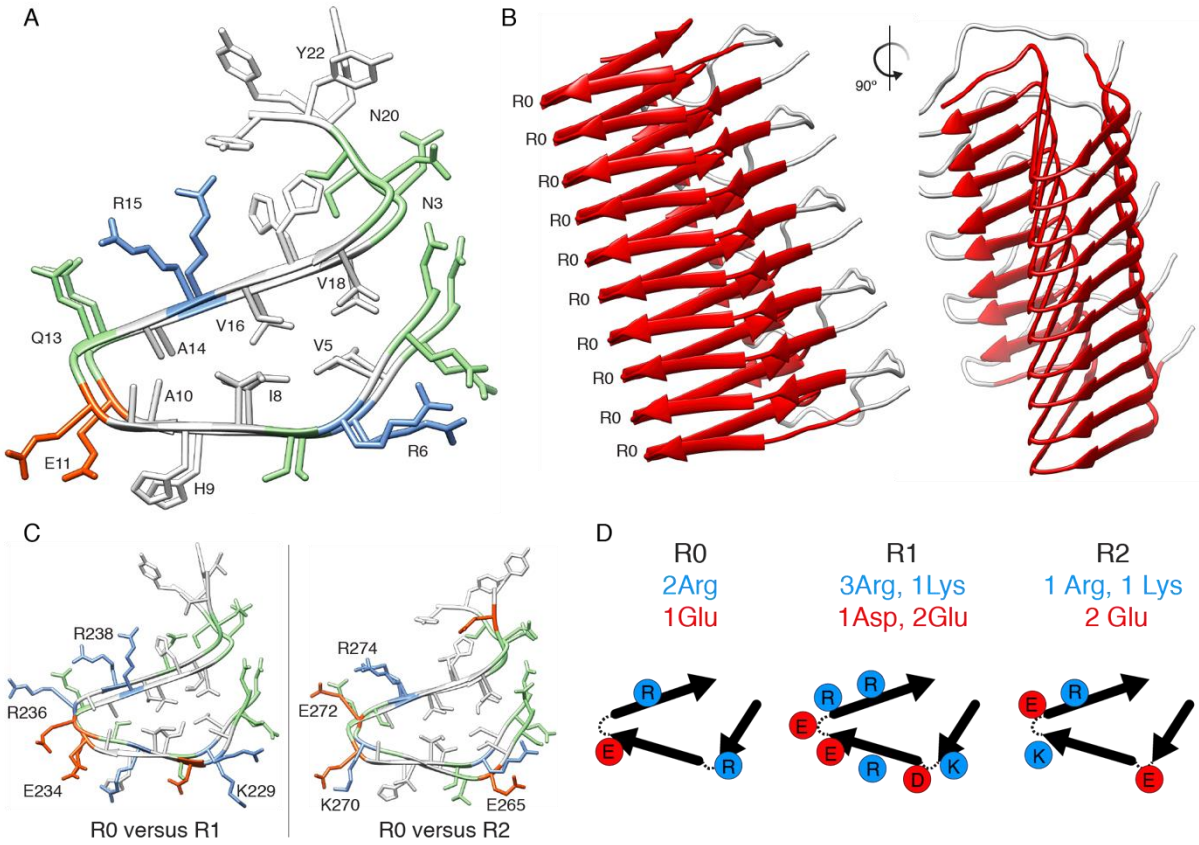



*Figure 4.4 Three-dimensional structure of R0-R0 functional amyloids. A) Atomic model of R0-R0 amyloid core. Hydrophobic residues are colored in white, positively charged in blue, negatively charged in orange. B) Side views of R0-R0 functional amyloids for a fibril composed of Five R0-R0 chimeric constructs. C) Structural comparison of R0-R0 and R1-R2 (PDB entry 2RNM)[34] amyloid structures, displayed for R0 and R1 (left), R0, and R2 (right). D) Schematic view of the position of charged residues in R0, R1, and R2.*

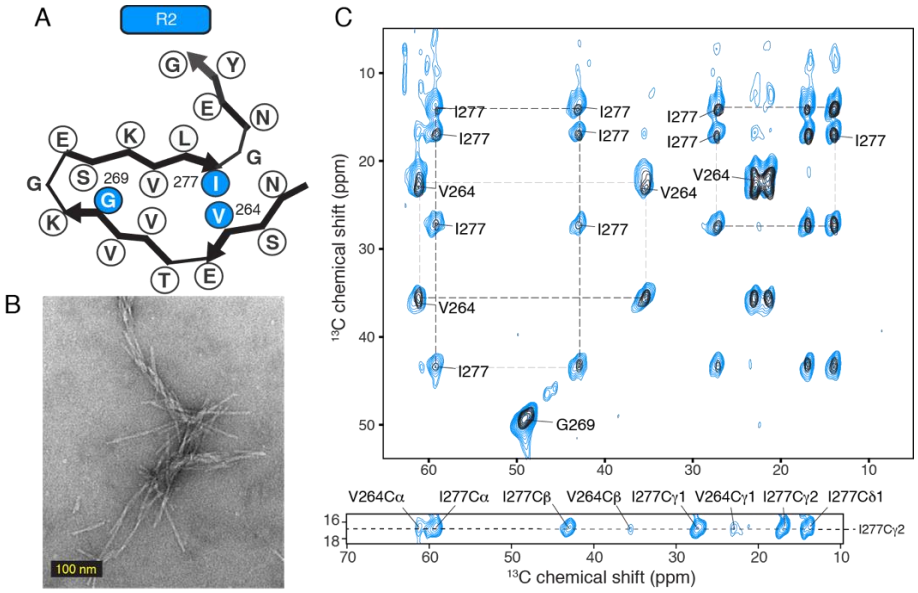



*Figure 4.5 R2 repeat forms amyloid fibrils and adopts b-solenoid fold. A) Schematic representation of R2 labeling strategy, made of three $^{13}$C-labeled amino acids (Val264, Gly269, and Ile277). B) Negatively stained electron micrographs of R2 fibrils, scale bar, is 100 nm. C) On top: two-dimensional solid-state*

*NMR $^{13}C$-$^{13}C$ correlation experiment of $^{13}C$-labeled V264/G269/I277 R2 amyloid fibrils, intra-residue correlations are shown. Bottom: Assignment of V264-I277 correlations.*

### 4.2.4 R0 forms a heterotypic amyloid interface with R2

Further**,** we checked the molecular mechanism of HET-S prionization, templated from the R0 prion-fold in the activated NWD2 complex. Our previous study proposed that the initial NWD2/PFD interaction occurs via a heterotypic amyloid interface designed by R0/R2. The conserved salt-bridges at positions 4 and 9 of R0 repeat, and a possible charge complementary with R2 repeat of HET-S could stabilize the heterotypic amyloid interface.[317] For a detailed structural analysis of this heterotypic amyloid signaling interface formed between R0 and R2, we designed a chimeric protein construct named R0-R2 with the pseudo repeats R0 and R2 connected by the R1-R2 loop **(Figure 4.6A).** This chimeric construct aggregated *in vitro* forming long unbranched fibrils of ~5nm in diameter comparable with HET-s PFD **(Figure 4.6B).**[250] Moreover, these amyloid fibrils displayed high prion infectivity on transfecting into [Het-s*] strains on a barrage assay **(Figure 4.6C**), suggesting the R0R2 can functionally cross-seed [Het-s*] strains by adopting a HET-s/S like prion fold**.**

We applied dipolar-based SSNMR experiments to elucidate the rigid amyloid core structure designed by the R0/R2 heterodimeric interface. The 2D $^{13}C$–$^{13}C$ spectroscopy with a PDSD mixing time of 50 ms, displayed a sharp $^{13}C$ linewidth of ~0.6 ppm, indicating the presence of highly ordered amyloid core as depicted in C$\alpha$ (~50–60ppm) and serine C$\alpha$-C$\beta$ region **(Figure 4.6D).** The consistent presence of a single set of resonances corresponding to 36 residues implies the lack of polymorphism in R0-R2 functional amyloid fibrils and provides an excellent spectral resolution (**See SI Figure4.1 and 4.2**). Sequential connectivity (i.e., residue i to i±1) from N3-19G and T260 to G278 can be established based on the 2D $^{13}C$ correlation experiment with a PDSD experiment with a long mixing time (200 ms) (**Figure 4.6E**). Interestingly, resonances corresponding to 24H was not observed owing to its specific dynamics or spectral overlap. Secondary structure analysis based on chemical shift was used to locate the six $\beta$–strand segments, displaying patterns and magnitude similar to HET-s PFD fibrils (**Figure 4.6F)**.[34,325] The non–β-sheet chemical shift of R6/D265 and G12/G271 can be explained by their positions at the $\beta$-arch. Interestingly, the $\beta$4 and $\beta$8 strands are not observed in CPMAS-based spectra along with the R1-R2 loop residues due to their fast dynamics.

To obtain molecular distance restraints, we collected SSNMR data from several samples, including $^{13}C$/$^{15}N$-uniformly enriched (U-[$^{13}C$-$^{15}N$]) or sparsely $^{13}C$ enriched (1,3-$^{13}C$- and 2-$^{13}C$- glycerol labeled) of R0-R2 functional amyloid fibril samples (**Figure 4.7 A, B**). The residue-residue plot corresponding to the peak-lists connecting between i and j residues with $35 \leq |i - j| \leq 37$ implies a parallel beta-sheet between the two pseudo-repeats (res. 5-263 and res. 18-277), which are separated by 36 residues in the chimeric construct (**Figure 4.7C**). However, some peaks at the C$\alpha$ region (~50-60ppm) in 2-$^{13}C$- glycerol spectra with a PDSD mixing time of 400 ms had no possible resonance assignment at all, which we assume either as the spectral artifacts or belongs to an unassigned spin system or originate between correlations between the protofilaments which are in close contact.

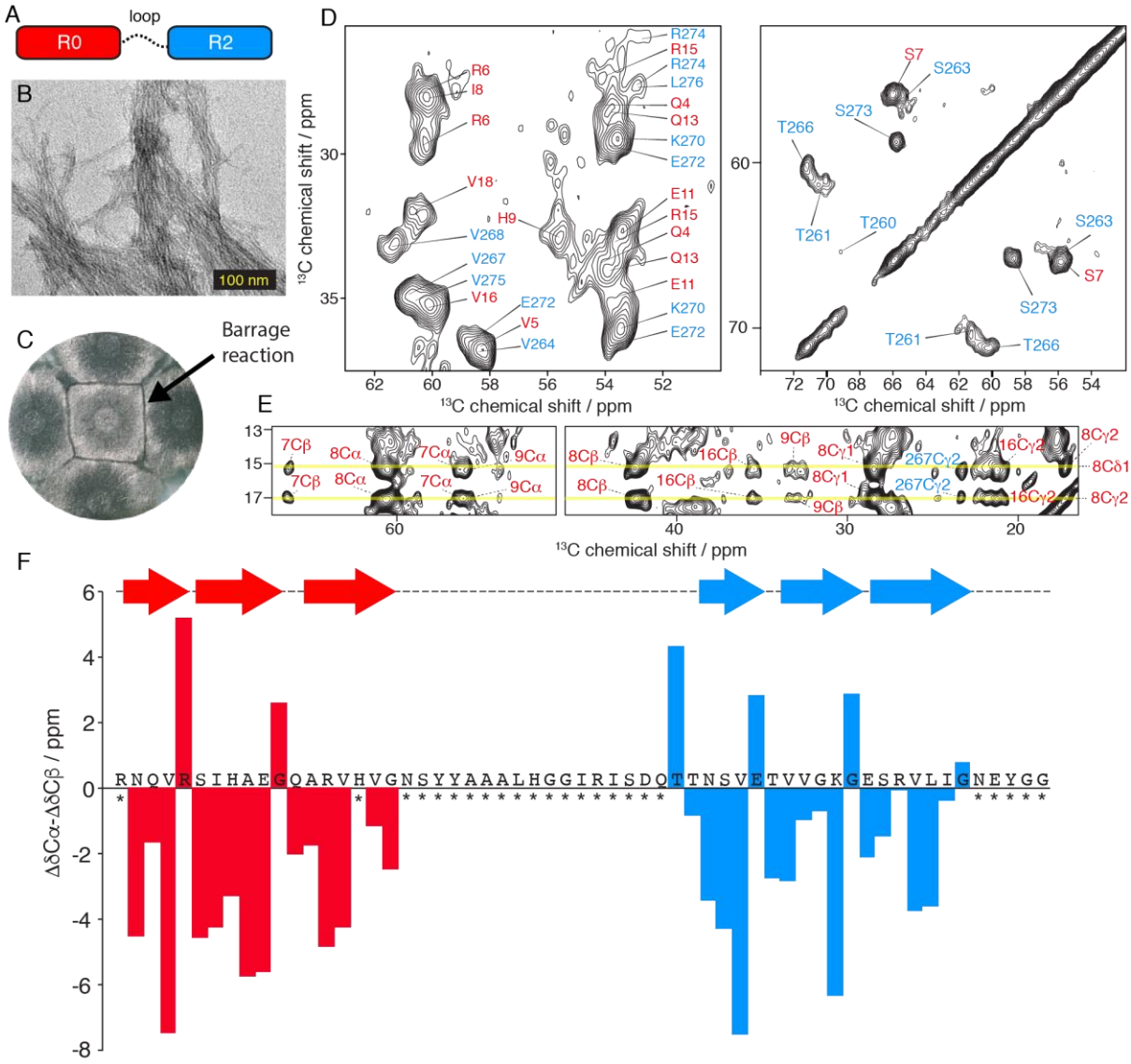



*Figure 4.6 Structural investigation of the chimeric R0-R2 functional amyloids. A) Schematic representation of R0-R2 chimeric construct strategy. Each monomer is made of the R0 and R2 repeats connected by the R1-R2 loop. B) Negatively stained electron micrographs of R0-R2 functional amyloid fibrils, scale bar, is 100 nm. C) Transfection essays of [Het-s*] with R0-R2 functional amyloid fibrils. Four transfected strains are tested on each plate, and the tester strain in the center of the plate is [Het-S]. R0-R2 fibrils convert strains to the [Het-s] phenotype (detected by the formation of a barrage reaction to the central [Het-S] tester, highlighted by an arrow). D) Two-dimensional solid-state NMR $^{13}C$-$^{13}C$ correlation experiment of R0-R2 functional amyloid fibrils. Residue assignment is shown in red for R0 and in blue for R2. E) Intra- and inter-residue correlations are highlighted for Ile8-C$\delta$1 and Ile8-C$\gamma$2. F) Solid-state NMR secondary chemical shift analysis of R0-R2 functional amyloids, negative or positive values indicate $\beta$-a strand or α-helix conformation, respectively. The six identified ß-strands are displayed.*

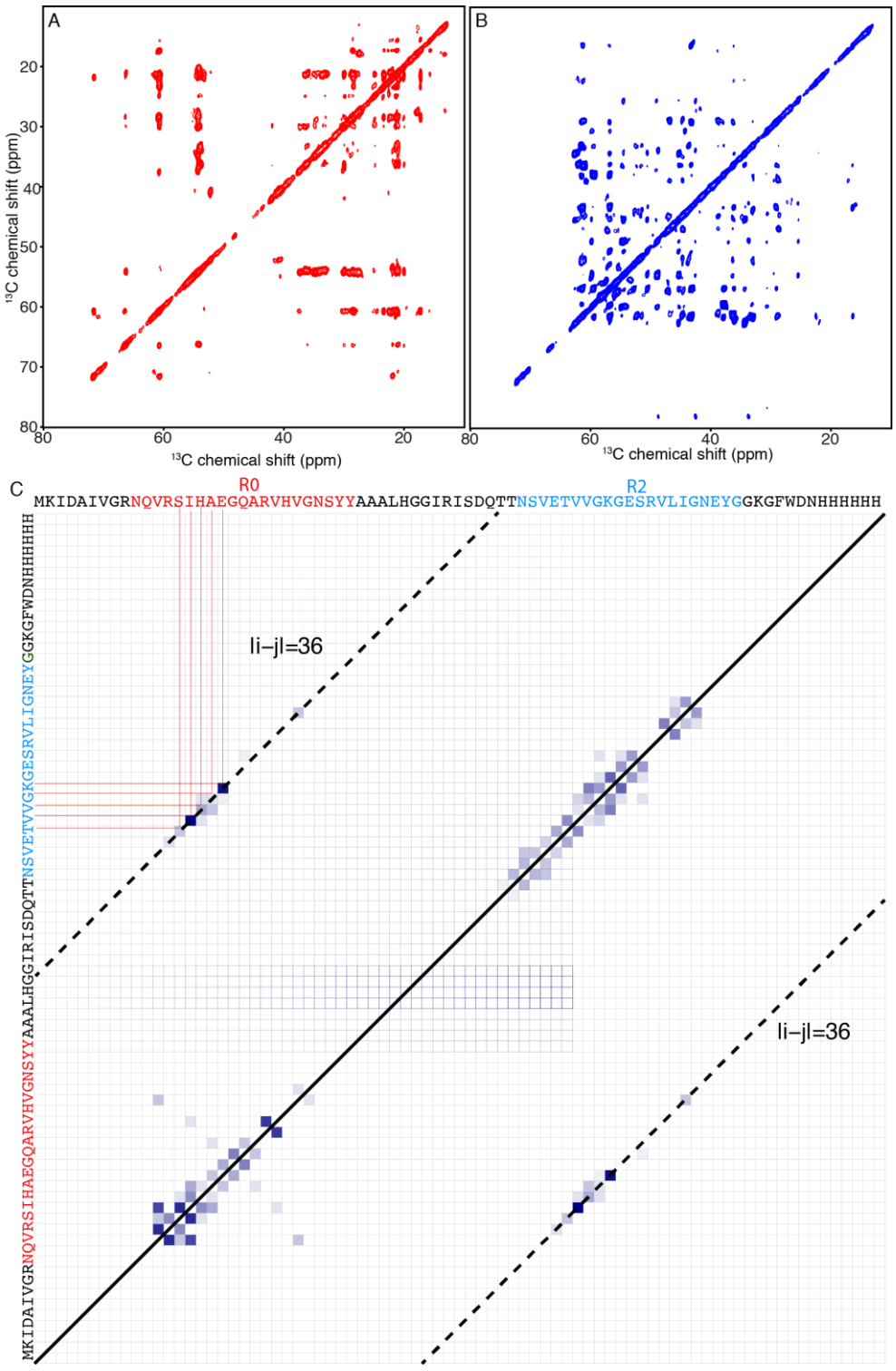



*Figure 4.7 Two-dimensional solid-state NMR $^{13}C$-$^{13}C$ correlation experiment of (A) 1,3-$^{13}C$- and (B) 2-$^{13}C$-glycerol labeled R0-R2 functional amyloid fibril samples with a PDSD mixing time of 400ms for the collection of distance restraints. C) Residue-residue plot of contacts observed in the SSNMR spectra displaying short-range distance ($|i-j| \leq 1$), long-range ($|i-j| > 4$) and medium-range, ($1< |i-j| < 5$) connectivity. The dotted lines indicate positions where peaks are*

*supporting in-register parallel $\beta$– sheets (|i-j|=36) as obtained in the final structure are expected. The darker the squares higher the number of restraint connectivity observed.*

*Table 4.2 R0-R2 hetero-amyloid SSNMR structure calculation statistics*

| **Number of restraints (per monomer)** | |
|---|---|
| Unambiguous NMR distance restraints | |
| ↳ Short-range (\|i-j\|≤1) | 226 |
| ↳ Medium-range (1<\|i-j\|<5) | 71 |
| ↳ Long-range (\|i-j\|≥5) | 138 |
| ↳ Ambiguous | 38 |
| ↳ Inter-monomers | 10 |
| ↳ Total | 483 |
| Hydrogen bond restraints (intra \| inter) | 22 \| 22 |
| Dihedral angle restraints (phi \| psi) | 24 \| 24 |
| **Restraints violation (per monomer)** | (mean ± SD) |
| NMR distance restraints | |
| ↳ RMS of violations (Å) | 0.00 ± 0.00 |
| ↳ Number of violations >0.3Å | 0.10 ± 0.30 |
| ↳ Number of violations >0.1Å | 8.82 ± 1.79 |
| Hydrogen bond restraints | |
| ↳ RMS of violations (Å) | 0.00 ± 0.00 |
| ↳ Number of violations >0.1Å | 0.00 ± 0.00 |
| Dihedral angle restraints | |
| ↳ RMS of angle violations (°) | 0.20 ± 0.061 |
| ↳ Number of violations >5° | 0.50 ± 0.500 |
| **Deviation from ideal geometry** | |
| RMS for bond lengths (Å) | 0.004 ± 0.0001 |

| | |
|---|---|
| RMS for bond angles (°) | 0.43 ± 0.018 |
| RMS for improper (°) | 1.35 ± 0.100 |
| **Ramachandran plot statistics (%)** | |
| most favored regions | 84.70 ± 2.54 |
| allowed regions | 15.30 ± 2.54 |
| generously allowed regions | 0.00 ± 0.00 |
| disallowed regions | 0.00 ± 0.00 |
| **RMSD from the average structure (Å)** | |
| All residues | |
| ↳ Backbone atoms | 9.10 ± 1.65 |
| ↳ Heavy atoms | 9.91 ± 1.76 |
| Prion forming domain (10-30,46-66) | |
| ↳ Backbone atoms | 2.44 ± 0.37 |
| ↳ Heavy atoms | 3.11 ± 0.39 |

a:-one hydrogen bond is encoded by two distance restraints (dH..O and dN..O), b:- from PROCHECK (Laskowski et al., 1993)[343], c:- residues with S(phi) + S(psi) > 1.8 (Bhattacharya et al., 2007)[448]

Further, these NMR restraints were converted to yield a 3D structure bearing a heterodimeric amyloid R0/R2 interface. Aria successfully retained 483 distance restraints in total, of which 138 were long-range (|i-j| > 4), 71 medium range, (1< |i-j| < 5) and 226 short-range distance (|i-j| ≤ 1) restrains. The ten resulting lowest-energy conformers bundled with an RMSD from the average structure of 2.44 Å and 3.11 Å for backbone atoms and heavy atoms, respectively, considering only the rigid amyloid core (**Figure 4.8A**). A detailed structural calculation statistic is given in **table 2**. Unlike HET-s PFD and R0 fibrils, the R0-R2 amyloid core comprises only the six β–strands forming the triangular hydrophobic core. Our SSNMR data do not show any strong NMR signals originating from the 4$^{th}$ and 8$^{th}$ β–strands or the R1-R2 loop and are therefore excluded from the rigid amyloid core. Moreover, the residues towards the end of the sixth β–strand L276-G279 in R2, possibly due to higher flexibility but are visible in higher PDSD mixing time (200/400ms). As a result, the orientation of the 4$^{th}$ and 8$^{th}$ β–strands could not be determined unambiguously. Interestingly, a predicted asparagine ladder between N20-N279 is not observed in R0-R2 stacking and may reduce the R0-R2 fibril stability. However, signals corresponding to aromatic residues from 4$^{th}$ and 8$^{th}$ β–strands (comprises of three tyrosines, 2 in R0 and one in R2 repeats) were absent on the J-based INEPT-HETCOR spectra, as shown in **SI Figure 4.3**, suggesting their possible existence in a molten state similar to the N-terminal domain in full-length HET-s fibrils.[449] Interestingly, the absence of 4$^{th}$ and 8$^{th}$ β–strands did not affect the formation of the central

hydrophobic core, suggesting the lesser importance of these β–strands for the amyloid stability and propagation. This inference is intuitive as the 4$^{th}$ and 8$^{th}$ β–strands are not directly involved in the central hydrophobic core. Also, no significant restraints corresponding to the C-terminus, residues beyond Y281 were observed as is the case in the newest structure of HET-s(218-289)(PDB 2KJ3), suggesting that the domain is more flexible in R0-R2.[236] Interestingly, mutational studies on HET-s PFD (218-289) suggests that the second asparagine ladder (N243/N279) at position 18 is not as strongly conserved as the first one and displays WT-like prion activity.[450] However, the lack of N-ladders significantly reduced fibril stability and displayed longer lag time.[450,451]

Our SSNMR data-based R0-R2 functional amyloid model suggests that the specific hydrophobic side-chains from the six β–strands (β1–β3 and β5–β7) interdigitate to form a dry face and stabilizes the triangular hydrophobic core, and contributes to the fibril stability. Analogous to HET-s PFD fold, the solvent-exposed charged residues are arranged on top of each other to compensate the charges by forming both intra and intermolecular salt bridges (6R-265E, 11E-270K), stabilizing the R0-R2 heterotypic amyloid stacking. Furthermore, the side-chain H-bonding between S7 and T266 and amide groups of asparagines forming the "asparagine ladders" (N3-N262) fortifies the R0-R2 stacking. The tightly packed hydrophobic side-chains in the amyloid core interior, several solvent-exposed stacking interactions outside the amyloid core, and the extended hydrogen bond network between adjacent β–strands stabilizes the R0-R2 β–solenoid structure to develop into mature amyloid fibrils (**Figure 4.8B**).

In summary, our SSNMR data suggests that the R0-R2 functional amyloid is capable of adopting a β–solenoid fold with a triangular hydrophobic core similar to HET-s PFD, independent of fourth and eighth β–strands. The residues from the fourth and eighth β–strands not being "SSNMR visible" on CP-based spectra imply a higher conformational disorder, correlating with lower fibril stability. The lack of interface stability can be attributed to a lack of formation of asparagine ladder N20-N279 or aromatic stacking Y23-Y281 that supports the amyloid interface. However, these inferences are in line with our working hypothesis that the ligand-induced oligomerization of NOD domains in NWD2, leading to a cooperative folding of R0 motifs adopting a central β–solenoid structure, which then serves as a structural template to nucleate HET-S PFD, facilitated via R0/R2 heterotypic amyloid interface (**Figure 4.8C**). The absence of fourth and eight β–strands could be crucial to smoothly detach prionized HET-S from the NWD2 signaling complex allowing repeated use of signalosome for many HET-S molecules enabling a signal amplification. Therefore, the formation of a stable NWD2/HET-S complex could be energetically highly demanding for a signaling complex.[439] However, an activated NLR complex directly involved in the pore-forming activity is observed in plants.[452,453]

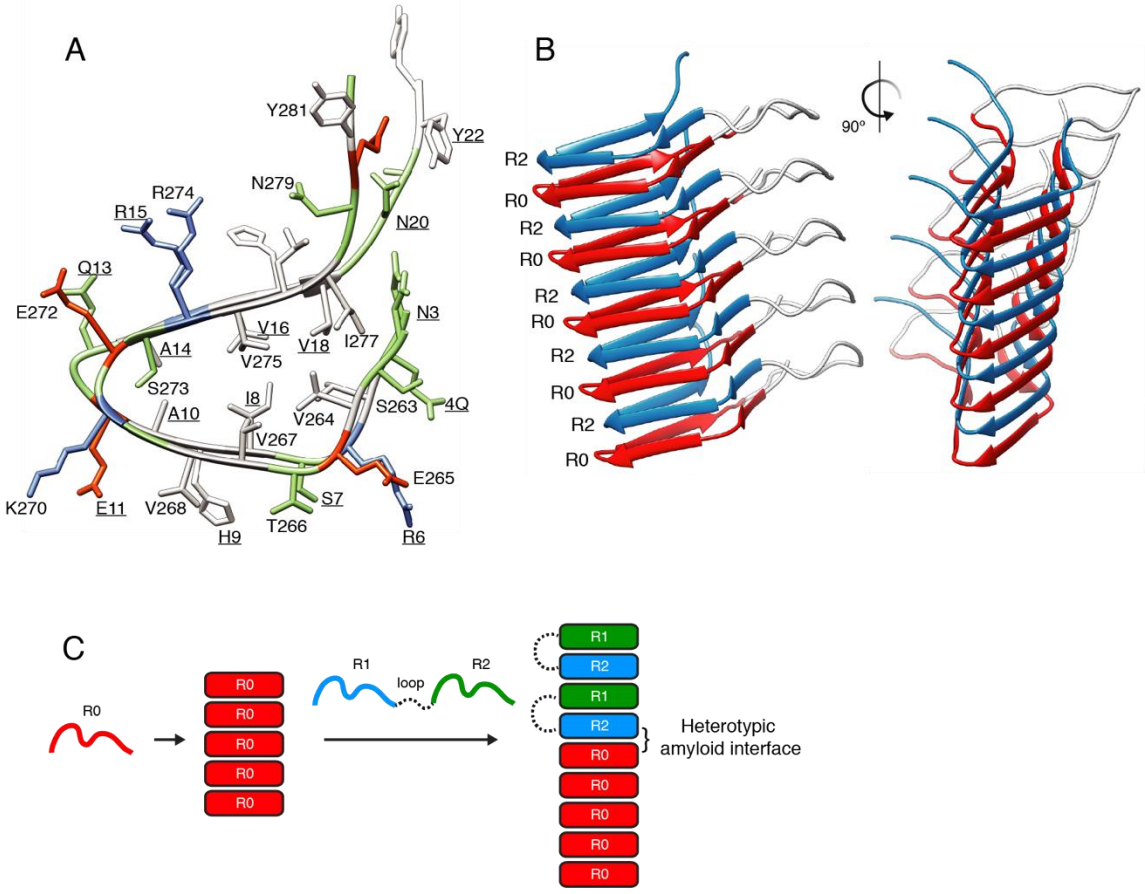



*Figure 4.8 A) Atomic model of R0-R2 amyloid core. Hydrophobic residues are colored in white, positively charged in blue, negatively charged in orange. Residues from R0 and repeats are shown with and without underline, respectively. B) Side views of R0-R0 functional amyloids for a fibril composed of five R0-R2 chimeric constructs. C) After activation of its NOD domains, R0 is converted into amyloid assemblies. R0 amyloid fold can serve as a structural template to nucleate R1-R2 in amyloid fibrils through a heterotypic interface between R0 and R2.*

### 4.2.5 R0 might form a heterotypic-amyloid interface preferably with R2 over R1 in the activation of HET-S PFD

We checked the determining factors for a successful hetero-amyloid interface by checking the possibility of an amyloid interface formed between R0 and R1 by SSNMR. For that, we employed a similar chimeric construction strategy where we connected R0 and R1 repeats by the R1-R2 loop **(Figure 4.9A**). The R0-R1 chimeric construct formed fibrillar aggregates of diameter ~5nm, as revealed by electron microscopy (**Figure 4.9B).** Surprisingly, these amyloid aggregates displayed high prion infectivity on transfecting into [Het-s*] strains on a barrage assay **(Figure 4.9C)**, suggesting their functional capacity to cross-seed [Het-s*] strains by virtually adopting a HET-s/S like prion fold**.** Nevertheless, the R0-R1 amyloid fibrils exhibited pronounced structural polymorphism as detected by broad peaks in the SSNMR spectrum arising from undistinguishable chemical shift variation due to multiple conformations, which are close in the energy landscape **(SI Figure 4.4**). Unlike R0-R2 hetero-amyloids, R0-R1 display the virtual presence of 8 β–strands identified from chemical shift

analysis with patterns and magnitude similar to HET-s PFD, suggesting the presence of a β–solenoid fold (**SI Figure 4.5**). A qualitative analysis of spectral resolution and variability in $^{13}C$ linewidth can be evaluated from the isoleucine and Cγ2/ Cδ1-Cα spectral regions to compare the amyloid core's rigidity in different amyloid fibrils **(Figure 4.8D)**. A narrower $^{13}C$ linewidth indicates a more stringent order in the amyloid assembly. The linewidth analysis reveals that the $^{13}C$ linewidth of various chimeric constructs follows a trend with R0-R0<R1-R2≤R0-R2<R0-R1, which means the R0-R1 amyloid core is liable to structural heterogeneity.
The increased disorder or polymorphic behavior of R0-R1 amyloid fibrils can be related to possible lower fibril stability or lack of stabilizing interaction that must be encountered by the structure to overcome energy barriers. These energetic backlashes could be due to severe charge repulsion and absence of salt-bridges when the pseudo repeats, R0 and R1, are arranged on top of each other (R6-K229, E11-E234, R15-R238). The presence of salt bridges shown to have a cumulative effect on kinetics and stability of HET-s amyloid fibrils and prion propagation rate.[64,451]

In HET-s PFD, two adjacent polar residues in the amyloid core stabilize together the fibril through a network of hydrogen bonds along the fibril axis (S273 and T233). Here, in the context of R0-R1, the side chain of T233 faces a less favorable core made of alanines side chains in close proximity. The relatively larger size of the threonine at that position (Alanine in R0, Glycine in R2) and the absence of potential polar interactions within the core are less favorable parameters for the stability and propagation of the fibril. Here in the R0-R1 spectra, we observed that the T233 Cα-Cβ peaks are week and the Cγ correlations are absent (**SI Figure 4.4**) in contrast to HET-s(218-289) fibrils, suggesting this residue is comparatively flexible and not able to form any polar interaction inside the R0-R1 hydrophobic core. Moreover, the residues T233 to A237 and A10 to A14 display broad signals compared to HET-s(218-289) and R0-R0 fibril spectra, suggesting a structural heterogeneity in this region. Buried polar interactions are shown to be favorable for early folding stages in HET-s(218-389).[451] Unfortunately, the particular location of T273 on the β2-strand blocks it from formulating any such interactions. The mutation studies on HET-sPFD by William et al. demonstrate that the T233 is stabilized by the adjacent S273 in WT but interacts unfavorably with the hydrophobic core in HET-s(218-289)S273A mutant.[451] The lack of polar compensation for T233 in the HET-s(218-289)S273A mutant leads to a longer lag time in aggregation, suggesting the destabilizing effect of T233 in the amyloid core. Also, they indicate that the S273 possibly could interact with the adjacent turn in the central hydrophobic core and display WT like fibrillation kinetics independent of stabilization by T233.[451] Similarly, FgHET-s(218-289), a distant HET-s homolog, accommodates a buried polar residue S273 at the identical position in its triangular hydrophobic core.[442] Interestingly, the clustering of HET-s-related amyloid motifs (HRAM) in fungi show that the serine at position 14 (especially in the HRAM4 family) is more conserved than the threonine at position 10 and do not display any sequence-covariance between them, suggesting the significance of a smaller polar residue at position 14 independent of stabilizing interaction from residues at position 10.[441]
In summary, our SSNMR data suggests that the β–solenoid fold is virtually conserved in R0-R1 hetero-amyloids but displays a higher degree of structural heterogeneity, as exemplified by the broader lineshapes. The presence of a kinetically unfavorable buried polar residue (T233), absence of buried polar interactions, or stabilizing salt-bridges could detrimentally affect the early folding stages of the R0/R1 interface formation. Additionally, the R0 clustered in the fungal NLR signalosome choosing R1 for a hetero-amyloid interface over relatively

easily accessible C-terminal R2 motif is highly unlikely due to the HeLo domain steric hindrance.

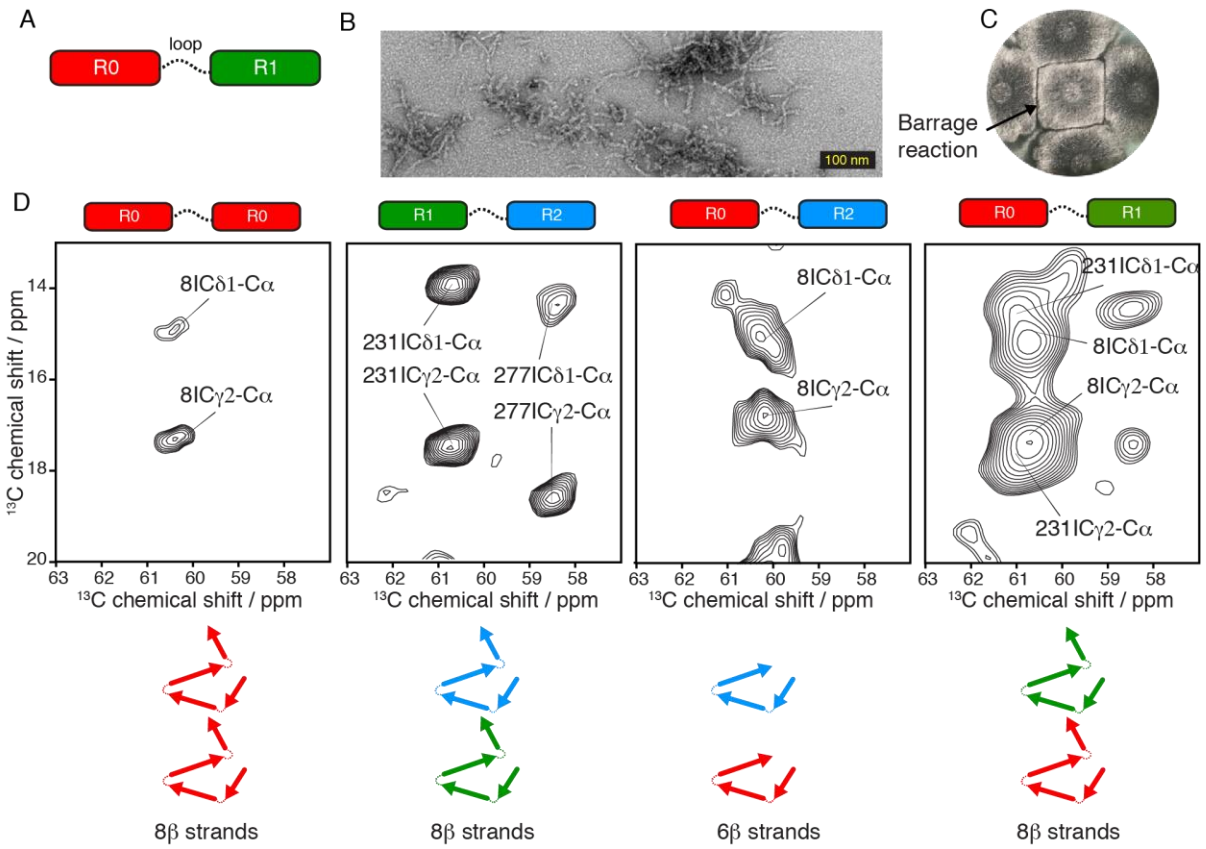



*Figure 4.9 Structural investigation of the chimeric R0-R1 functional amyloids. A) Schematic representation of R0-R1 chimeric construct strategy. Each monomer is made of the R0 and R1 repeats connected by the R1-R2 loop. B) Negatively stained electron micrographs of R0-R1 functional amyloid fibrils, scale bar, is 100 nm. C) Transfection essays of [Het-s*] with R0-R1 functional amyloid fibrils. Four transfected strains are tested on each plate, and the tester strain in the center of the plate is [Het-S]. R0-R2 fibrils convert strains to the [Het-s] phenotype (detected by the formation of a barrage reaction to the central [Het-S] tester, highlighted by an arrow). D) Comparison of R0-R0, R1-R2, R0-R2, and R0-R1-R2 $^{13}C$-$^{13}C$ correlation experiments for the isoleucine C$\delta$1-C$\alpha$ and C$\gamma$2-C$\alpha$ spectral region, illustrating a variability of $^{13}C$ linewidth. A schematic representation of the $\beta$–sheet arrangement observed in each construct is displayed at the bottom.*

### 4.3 Discussions

In this work, we analyzed different functional amyloids formed by chimeric constructs of pseudo repeats involved in the NWD2/HET-S paradigm, namely R0-R0, R0-R1, and R0-R2. We used a combination of SSNMR, molecular modeling, and other biophysical techniques to gain insights into the structural basis of amyloid based signal transduction by fungal NLRs in cell fate pathways. All the chimeric constructs form amyloid fibrils with a virtually conserved β–solenoid fold and functionally display high prion infectivity. It has been demonstrated that the R0 motif can adopt a HET-s/S like fold and exhibit prion activity.[250] Analogous to the HET-S/NWD2, several other gene clusters display an affector-effector correlation based on a prion-

fold transmission from a NOD-like STAND receptor protein to an effector protein involved in cell fate signaling pathways. Also, such gene clusters are widespread in the fungal kingdom, suggesting its evolutionary success.[249,441] However, the mechanistic details of the amyloid fold adopted by the R0 amyloid motif of fungal NLR protein, NWD2, and its putative amyloid interface with HET-SPFD is crucial to reveal the molecular mechanism underlying prion amyloid based signal transduction in fungi.

Our SSNMR based molecular model highlights the main differences in the amyloid core conformation between the different chimeric constructs. Similar to the HET-s PFD model, the R0-motifs in the NWD2 NLR signalosome comprises a central triangular hydrophobic core designed by three β–sheets (β1, β2 and β3) with a fourth β–sheet pointing away from the central amyloid core.[34,236] The hydrophobic core is designed by interdigitating hydrophobic side-chains. The charged and polar residues point towards the exterior. The two asparagine ladders at positions 3 and 20 and the glutamine ladder at positions 4, 13 contribute to the fibril stability on R0 parallel stacking.

Furthermore, we successfully reported the 3D structural model for the heterotypic-amyloid interface(R0/R2), i.e., amyloid interface made of two different protein segments, here R2 motif in HET-s and its templating precursor amyloid (NWD2), involved in the fungal signal transduction to programmed cell death. Hetero-amyloids are coupled with a variety of biological responses, including necroptosis and neurodegenerative diseases.[444–446] Although the triangular hydrophobic core is virtually conserved in R0-R2 hetero-amyloids, our SSNMR data suggests that the 4$^{th}$ and 8$^{th}$ β–strands are dynamically flexible and display a high degree of structural heterogeneity.

To gain further insights into the backbone-conformational changes associated with the formation R0/R2 hetero amyloid interface, we made a chemical shift comparison of Cα residues in R0 and R2 repeats in R0-R2 hetero-amyloids with NWD2 homo-amyloids and HET-s(218-289) fibrils (**Figure 4.10A, B**).[34] We observed R0 and R2 pseudo repeats on the hetero-amyloid interface displays a significant amount of chemical shift perturbation compared to their parent amyloids, suggesting a discreet structural transition involved in the hetero-amyloid interface in the proposed templating mechanism. Moreover, residues toward the end of the sixth β–strand L276-G279 in R2 is absent in a lower mixing time (50ms) but are visible in higher PDSD mixing time (200/400ms) marked in sky blue in **Figure 4.10A**, implying their existence in a semi-rigid conformation in the heterotypic-amyloid interface. The semi-rigidity in this region and the absence of the 4$^{th}$,8$^{th}$ β–strands can be attributed to the lack of formation of asparagine ladders (N20-N279) and associated chemical shift perturbations. The asparagine ladders are shown to seal the amyloid structures in many amyloids, including HET-s (218-289)-like and the Q/N rich amyloid proteins.[64,441,442,451,454–456] The stabilizing interactions via the hydrogen bond networks from side-chain amides, Q/N ladders act as a replication interface to promote prion conversion.[456]

The absence of a conserved asparagine ladder, an absent fourth β–sheet in the R0/R2 interface, may reduce the hetero-amyloid interface stability. However, a virtually conserved triangular hydrophobic core could be sufficient to nucleate HET-S PFD amyloid transition from the activated NWD2 complex. This crucial structural difference in the R0/R2 interface could exfoliate prionized HET-S from the NWD2 signalosome. This mode of amyloid nucleation will allow repeated use of a signalosome, enabling signal amplification. A stable NWD2/HET-S complex could be energetically highly demanding for a signaling complex.

The possibility of NWD2 signalosome interacting with R2 as a templating interface can be proposed by the following factors i) a certain number of HET-S/s homologs have altered

R1 motifs (with small deletions, for example) and also that a part of the R1 motif is involved in the HeLo domain fold[64,240,241] ii) The R2 repeat possesses conserved salt-bridges at positions 4 and 9, resulting in charge complementary between both R1 and R0 motifs, suggesting R2 could be the connecting link for the amyloid templating from an activated NWD2 complex to HET-S to overcome the nucleation barrier (**Figure 4.10C**). A charge complementarity with the R0 motif could be favorable for an initial NWD2/PFD interaction via an R0/R2 interface. Conversely, a plausible electrostatic charge repulsion occurring in an R0/R1 interface avoids such an intermediate step. iii) In the proposed model of the NWD2 signalosome, the central R0 β–solenoid fold, surrounded by NACHT domains in the fungal NLR signalosome, is more likely to interact with the readily accessible C-terminal R2 motif than the R1 motif attached with a sterically crowded HeLo domain. iv) Our SSNMR data suggests that the β–solenoid fold is virtually conserved in R0-R1 hetero-amyloids but displays a higher degree of structural heterogeneity, as exemplified by the broader lineshapes. The presence of kinetically unfavorable buried polar residue (T233) at the β–arc position, absence of the buried polar interactions, and charge complementarity could detrimentally affect the early folding stages of the R0/R1 interface formation.

Higher-order assemblies are often found useful in receptor signaling complexes in innate immunity such as the apoptosome, the myddosome, the resistosome, the RIP1/RIP3 necrosome, the ASC-dependent inflammasomes, and the MAVS CARD filaments.[234,251,439,452,457–460] We here propose a working hypothesis based on our SSNMR data and the ASC mediated inflammasome signaling;[251] 1) Putative ligand(s) bound on WD-domain induces oligomerization of NWD2 proteins via the NACHT domains, allowing a spatial clustering of R0 motifs and cooperative folding into the β–solenoid fold. 2) This NWD2 cluster with a central β–solenoid prion fold nucleates prionization of HET-S proteins via an R0/R2 heterotypic amyloid interface, which otherwise prevented from occurring due to insufficient expression level for spontaneous prionization of HET-S and high nucleation energy barrier. 3) The growth of HET-S prions on NWD2 signalosome central base (**Figure 4.10D**) followed by the fragmentation/exfoliation from the signalosome owing to lower fragmentation energy via R0/R2 hetero-amyloid interface. 4) HET-s prions then rapidly template other HET-S molecules / the free NLR signalosome prionize other HET-S molecules resulting in signal amplification. 5) HET-S prions with an activated HeLo domain relocates to the cell membrane and perform a pore-forming activity.

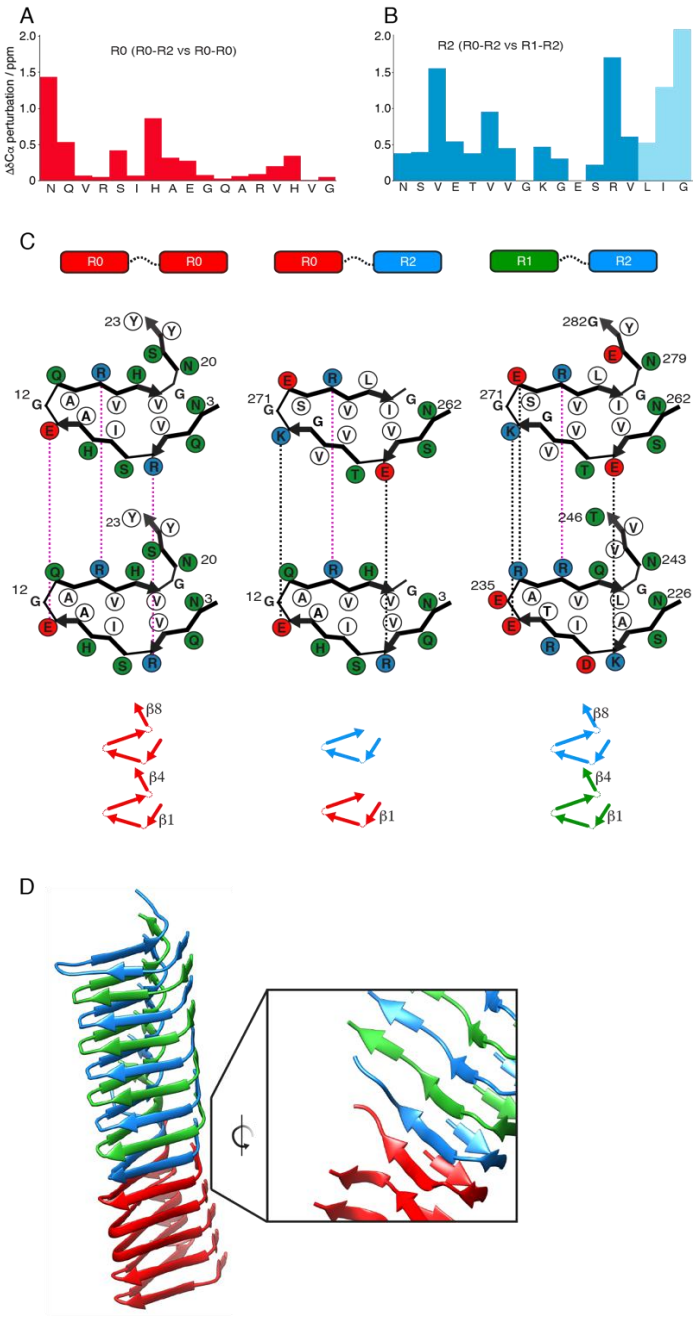



**FIGURE 4.10**

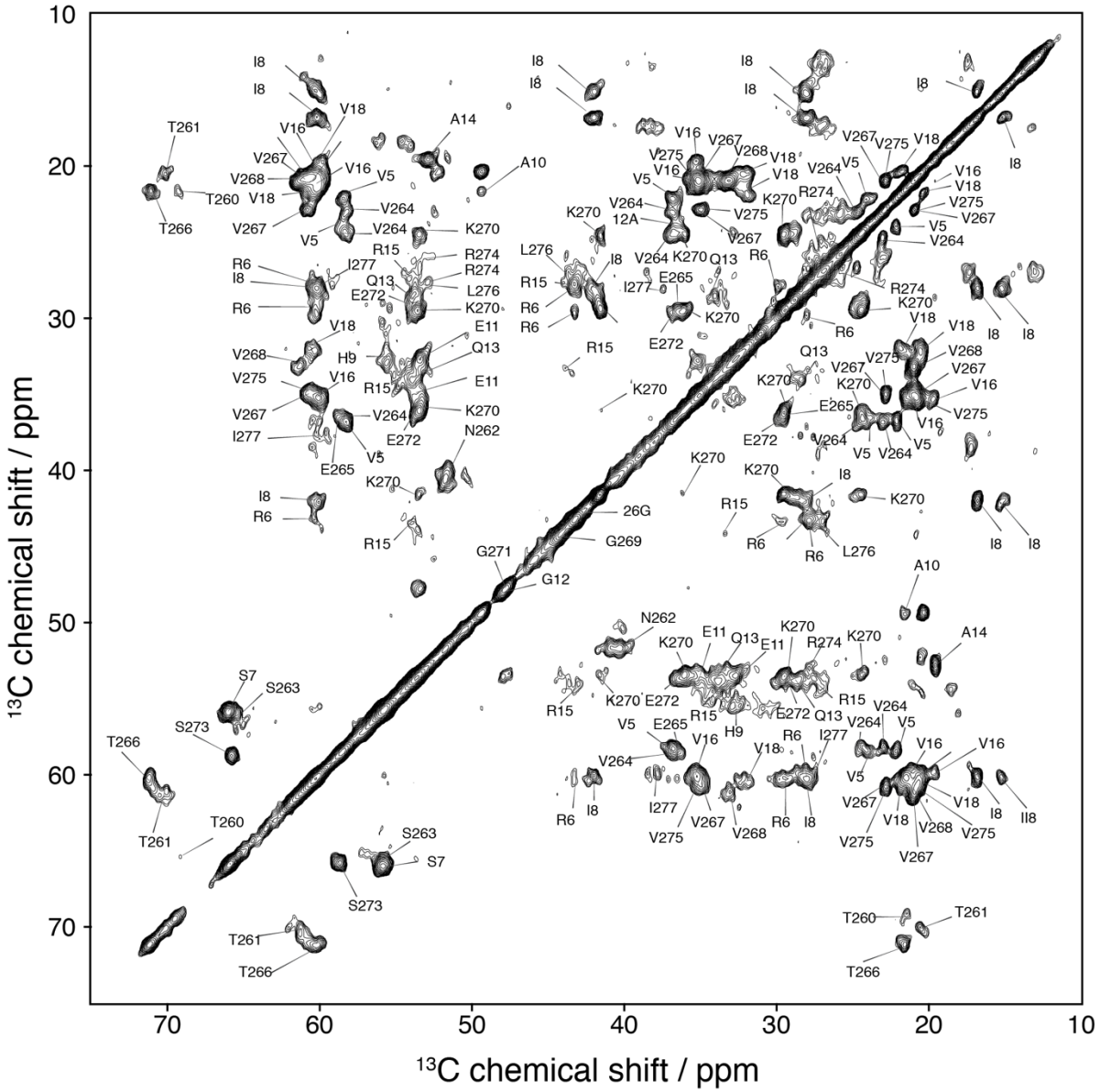
*Figure 4.11 Heterotypic amyloid signaling mechanism. A)$^{13}C\alpha$ chemical shift perturbation measured as the difference of R0 chemical shifts in R0-R2 and R0-R0 (in red) B) $^{13}C\alpha$ chemical shift perturbation measured as the difference of R2 chemical shifts in R0-R2 and R1-R2 (in blue), semi-rigid amino acids are displayed in sky blue. C) A schematic representation of the arrangement of b-sheets in R0-R0, R0-R2, and HET-s functional amyloids showing favorable interactions (In black) and unfavorable interactions (in pink) D) Proposed model of the R0/R1-R2 amyloid fibrils after R1-R2 elongation. The zoomed image shows the heterotypic interface between R0 and R2.*

The advantage of such an assembly is that a limited number of NWD2 (pentamer or heptamer) would be sufficient to activate the transconformation and activation of many HET-

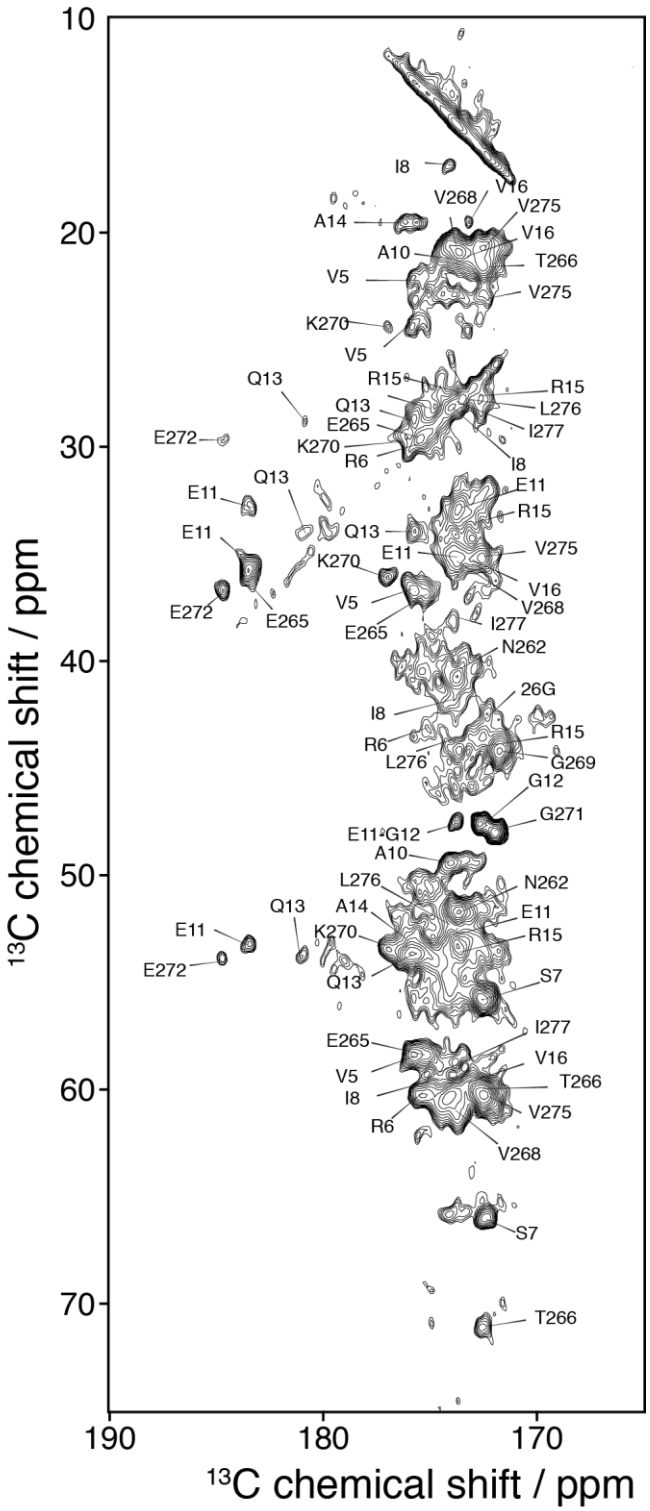
S molecules for membrane disruption.[439] Hence, it could be crucial to exfoliate HET-S prion from the activated NWD2 complex achieved via fragmentation of the R0/R2 interface and repeat the process allowing a signal amplification. Signal amplification can also be achieved by seeding the amyloid transition of native HET-S by prion form of HET-S. Such signal amplification associated with gene expression programs has been described in the temporal activation of transcription factor nuclear factor (NF)-κB by the signaling molecule tumour-necrosis factor (TNF)-α in a dose-dependent manner.[461] The higher-order assemblies, in general, formed via prion-like polymerization in fungal NLR signaling machines, facilitates essential features of signal transduction cascades such as signal amplification, threshold response, and noise reduction.[439] A signal transmission via a prion-fold is widespread across species, suggesting its evolutionary success due to strict cooperativity in amyloid folding, robustness in its templating and propagation, and extreme compactness.[462] Here, we report a molecular mechanism of a potential advantage of achieving signal amplification in two-gene architecture in contrary to the more straightforward all-in-one structure in fugal NLR signaling via the heteroamyloid interface potentially designed with lower stability. Here, the R0/R2 interface may perform a dual role, i.e., act as a templating surface and a fragmentation interface.

It thus appears that in addition to the ability to form the β-solenoid fold efficiently, sequence variation of elementary R-motifs is finely tuned at the sequence level, with (i) an R1 tailored to a lower the amyloid-forming ability potentially to avoid self-activation of HET-S and (ii) R0 reduced in charge content to allow for homotypic polymerization and (iii) capable of forming different heterotypic interfaces with either R1 or R0 with the R0/R2 interface potentially conductive to slightly lower stability to allow for recycling of the NWD2 signalosome.

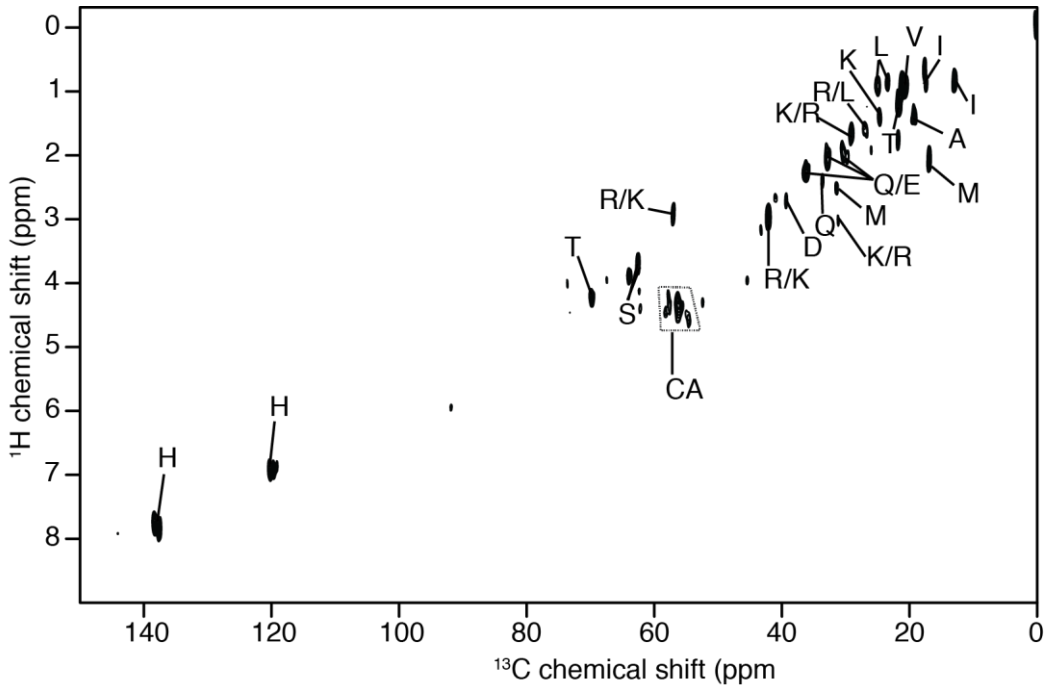
## 4.4 Supplementary figures



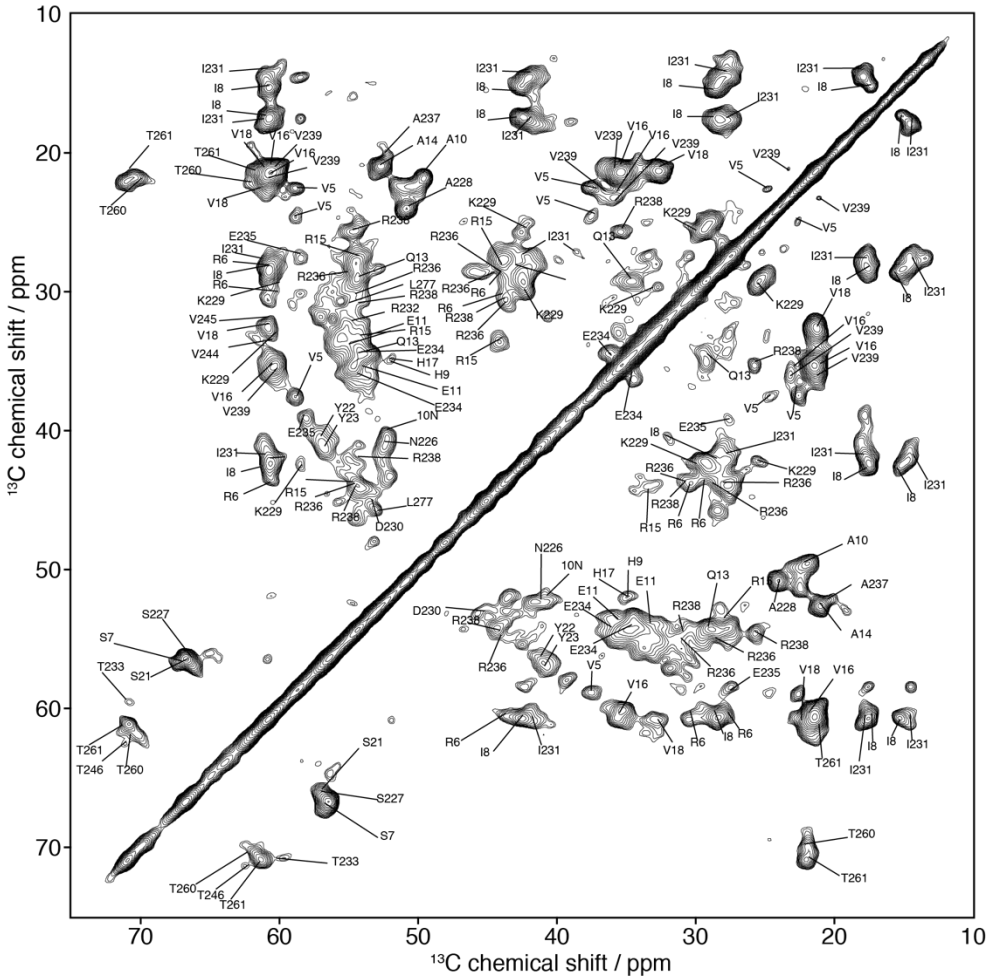

*SI Figure 4.1 Solid-state NMR $^{13}C$ -$^{13}C$ correlation experiment of the aliphatic region of R0-R2 fibrils recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS, and with a PDSD mixing time of 50 ms at 278K.*



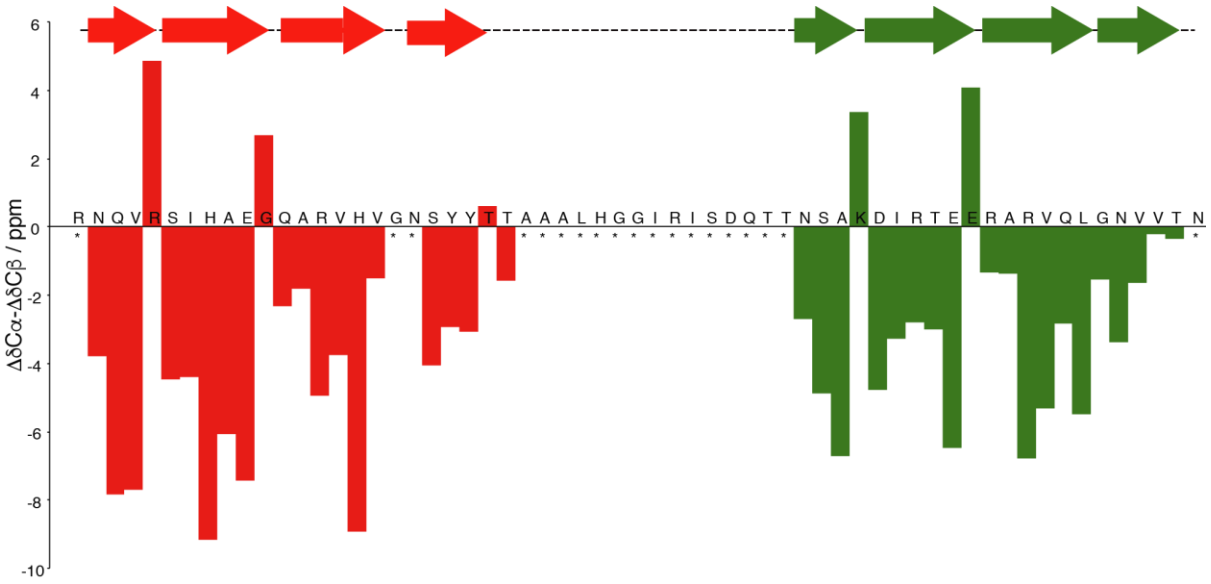

*SI Figure 4.2 Solid-state NMR $^{13}C$ -$^{13}C$ correlation experiment of carbonyl region of R0-R2 fibrils recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS, with a PDSD mixing time of 50 ms at 278K.*




*SI Figure 4.3 SSNMR $^{1}$H–$^{13}$C INEPT of R0-R2 amyloid fibrils recorded at a $^{1}$H frequency of 600 MHz, 11 kHz MAS frequency at 278 K, showing tentative assignments corresponding to residues in the flexible region of the fibrillar assembly.*




*SI Figure 4.4 Solid-state NMR $^{13}C$ -$^{13}C$ correlation experiment of the aliphatic region of R0-R1 fibrils recorded at a $^{1}H$ frequency of 600 MHz, 11 kHz MAS, with a PDSD mixing time of 50 ms at 278K.*




*SI Figure 4.5 Solid-state NMR secondary chemical shift analysis of R0-R1 functional amyloids, negative or positive values indicate β-strand or α-helix conformation, respectively. The eight identified β-strands are displayed.*

## 4.5 Materials and methods

### 4.5.1 Preparation of the pET24-NWD2 expression vectors

The different NWD2/HET-s constructs encoding genes are designed to contain a C-terminal 6xHis-tag and flanked by NdeI and XhoI restriction sites. The constructs were ordered from Eurofins genomics in a standard pEX-A vector. The gene inserts and the expression vector pET24 (Novagen) were prepared for cloning by digestion with NdeI and XhoI and ligated using T4 DNA ligase (New England Biolabs).

### 4.5.2 Recombinant protein expression

E.coli strain BL21 (DE3) pLysS was transformed with pET24-X vector (X= NWD2-NWD2 /5XNWD2/HET-s /R0-R2 /R0-R1 /R0-R1-R2) and plated onto LB agar plates containing 30 µg/mL kanamycin. As an initial culture, 10 mL LB medium was inoculated with a single clone and incubated in 37°C under shaking at 220 rpm until the exponential phase (OD600 0.8). For the production of $^{13}C/^{15}N$ labeled proteins, the expression was carried out in an M9 medium with 15NH4Cl and 13C6-D-glucose as sole nitrogen and carbon sources, which was inoculated with 2% (v/v) of the initial fresh culture. The use of M9 medium supplemented with $^{15}NH_4Cl$ and either 1,3-$^{13}$C-labelled glycerol or 2-$^{13}$C-labelled glycerol resulted in selectively producing $^{13}$C-labeled proteins to facilitate specific NMR analysis. The bacteria were grown at 37°C while shaking at 220 rpm, and expression was induced with 0.75 mM IPTG when the OD600 reached 0.7-0.8 for 20h at 30°C. Then, the Cells were harvested by centrifugation (6000 rpm, 30 min, 4°C) and frozen (-80°C) until purification.

### 4.5.3 Recombinant protein purification

Frozen cell pellets were thawed and resuspended on ice in 20 mL lysis buffer (Tris 50 mM, 150 mM NaCl, pH 8). Cells were lysed by sonication on ice with 30 W power (45-sec pulses, 45-sec intervals). The cell suspension was centrifuged at 15 000 g for 1h at 4°C to collect inclusion bodies. The pellet containing the protein was then resuspended in 20 mL lysis buffer supplemented with 2% Triton X-100 by vortexing vigorously, incubated at 37°C (40 min, 220 rpm), and centrifuged (50 000 g, 10 min, 4°C). The pellet was washed with lysis buffer and incubated at 37°C (1h, 220 rpm) to remove residual Triton X-100. After centrifugation (50 000 g, 10 min, 4°C), the pellet was resuspended in 10 mL extraction buffer (50 mM Tris, 0.5 M NaCl, 6M Guanidinium thiocyanate, pH 8). The suspension was incubated overnight at 60°C, sonicated on ice as above, and centrifuged (250 000 g, one h, 22°C). The supernatant was recovered and loaded on a HisTrap HP 5mL column (GE Healthcare) equilibrated in a binding buffer (50 mM Tris, 0.5 M NaCl, 20 mM imidazole, 8 M urea, pH 8). Proteins were eluted from the column with a linear gradient of elution buffer (50 mM Tris, 0.5 M NaCl, 500mM imidazole, 8 M urea, pH 8). 6M Guanidinium hydrochloride was used as a denaturing agent instead of 8M urea in binding and elution buffer for NWD2-NWD2 and R0-R1 to avoid any pre-formed aggregates. The fractions containing pure protein were monitored by UV absorbance at 280nm and analyzed on 12% tris-tricine SDS PAGE. The protein solution was further desalted with a HiPrep 26/10 desalting column (GE Healthcare) connected to an Äkta Pure 25 HPLC system in a continuous 7 mL/min flow of solvent A (1% acetic acid). Fractions containing protein were monitored by UV absorbance at 280 nm.

### 4.5.4 *In vitro* assembly of fibrils

The in vitro self-assembly of NWD2/HET-s samples in 1% acetic acid was initiated by rapidly changing the pH to 7.5 by adding 3M Tris buffer, pH 8. These samples were allowed to self-assemble for one week at room temperature. It was then extensively washed in double distilled water, centrifuged at 300 000 g, and transferred to the SSNMR rotor. R2 peptide was purchased from GENOSPHERE Biotechnologies (France) at purity >95%, dissolved in 1% acetic acid at 1mM, and aggregated at room temperature by adding 15% (v/v) of 3M Tris buffer, pH 8 in order to turn the pH to 7.5.

### 4.5.5 Prion propagation and incompatibility assays

The protein transfection assays with amyloid fibrils of recombinant R0-R1, R0-R2, R0-R1 were carried out following the protocol described by Benkemoun et al. and Dasklov et al. with minor modifications.[250,463]

### 4.5.6 Solid-state NMR spectroscopy

SSNMR experiments were performed at a MAS frequency of 11 kHz on a triple resonance 4 mm or 3.2 mm MAS probe (Bruker Biospin) using 500 MHz, 600 MHz, and 800 MHz 1H Larmor frequency spectrometers (Bruker Biospin). The sample temperature was set to ~278 K, and chemical shifts were calibrated using DSS as an internal reference. A cross-polarization contact time of 1 ms was used for $^{1}H$-$^{13}C$ polarization transfer. Proton decoupling with a strength of 90 kHz using the SPINAL-64 decoupling sequence was applied during acquisition times. The two-dimensional $^{13}C$-$^{13}C$ correlation spectra using a proton-driven spin-

diffusion (PDSD) mixing of 50 ms was used to correlate intra residue carbon atoms with acquisition times of 20 ms (direct) and 10 ms (indirect), leading to a total experimental time of ~1 day. Sequential correlations were probed using a PDSD mixing time of 200 ms for acquisition times of 20 ms (direct) and 9.5 ms (indirect). Long-range $^{13}C$-$^{13}C$ contacts were examined by using a PDSD mixing time of 400 ms for acquisition times of 20 ms (direct) and 7.5 ms (indirect), leading to a total experimental time of ~7 days on selectively $^{13}C$ labeled samples (2-glycerol and 1,3-glycerol samples). Long-range $^{13}C$-$^{13}C$ contacts were obtained using a PDSD mixing time of 400 ms using selectively $^{13}C$-labeled samples (2-glycerol and 1,3-glycerol). Mobile residues were probed using a two-dimensional $^{1}H$-$^{13}C$ INEPT experiment with acquisition times of 18.4 ms and 7.1 ms in direct and indirect dimensions, respectively, for a total experimental time of ~10h. All two-dimensional spectra were processed using a sine-bell shift of pi/4 and plotted identically setting the contour level (50) and scaling factor (1.18) in SPARKY/ CCPNMR analysis software.

### 4.5.7 NMR Structure calculation

The fibrils structures of R0R0 and R0R2 were determined with ARIA2.3 in 8 cycles of structural calculations and restraint analysis followed by a cycle of refinement in water.[333] At each iteration, 100 structures were calculated, and the 10 lowest-energy structures were passed to the next iteration.[338] Cross-peak assignments for $^{13}C$-$^{13}C$ correlations were converted into distance restraints with an upper-bound of 8.5 Å. A total of 679 and 506 restraints were given to ARIA as input restraints for R0R0 and R0R2, respectively. In both cases, sequential short and medium-range assignments ($|i-j| \leq 4$) were set manually to intra-chain restraints. When $|i-j|>4$, and in the absence of evidence whether the assignments were intra- or inter-chain correlations, the ARIA protocol was determining itself throughout the iterations, the intra- or inter- nature of each restraint.

Backbone dihedrals angles were predicted with TALOS+ from $^{13}C$ and $^{15}N$ chemical shifts.[332] Secondary chemical shifts were used to identify secondary structure elements. Later, the dihedral angles of predicted β-strands ($\delta_{C\alpha}$-$\delta_{C\beta}<0$) were converted into dihedral angle restraints with an error range corresponding to ± 1.5 times the TALOS error with a minimum of ± 15°. Based on the identified β-strands, and through similarity with the published structures of HET-s(218-289) (PDB ID 2RNM, 2KJ3), intra- and inter-monomer hydrogen bond restraints between β-strands (β1-β5, β2-β6, β3-β7, β4-β8) were included in the calculation. Final restraints and structure statistics are given in Tables 1 and 2.

The fibrils structures were calculated as pentamers using simulated annealing performed with CNS 1.2.[339] The ladder topology was maintained during the calculation through distance restraints, ensuring that the distance between equivalent Cα atoms in neighboring monomers is constant throughout the pentamer, without fixing a particular distance value, i.e., dm/m+1 = dm+1/m+2 = dm+2/m+3 = dm+3/m+4 = dm+4/m+5 .[236,340] Additionally, an NCS (Non-crystallographic Symmetry) restraint was added to minimize the RMSD between atomic coordinates of the monomers.[340] Finally, the conformational database potential from Kuszewski et al. [341] was applied to the whole system to maintain the dihedral geometry, especially in poorly determined regions.

### 4.5.8 Molecular Modelling

The structure of the R0/R1-R2 amyloid elongation fibril was created by homology with Modeller 9.25.[464] First, a manual structure was created with Chimera y juxtaposing 4.8Å-distant PFD in the order (R1-R2)x5-(R0)x5.[465] .The whole structure was then calculated based on multiple alignments that included the manual structure and the appropriate templates bearing the various interfaces (R0/R0 from the R0R0 structure, R0/R2 from the R0-R2 structure, and R1/R2 from the HET-s(218-289) structure.[34] Ten structures were generated, among which the one bearing the lowest Dope Score as selected as the final model.[466]
Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

### 4.5.9 Transmission Electron microscopy

One droplet of 2-fold serial dilutions of NWD2 constructs filaments in 150 mM acetic acid was applied to glow-discharged 300 mesh carbon-coated copper grids for 1min, washed with water, stained with a 2% uranyl acetate (w/v) solution for 1min and dried under dark condition. Samples were observed in an FEI CM120 transmission electron microscope at an accelerating voltage of 120kV under TEM low-dose mode at a magnification of 30 000. TEM images were recorded using a Gatan USC1000 2k x 2k camera.

# Chapter 5 BINDING MODES OF THIOFLAVIN T (THT) ON β–SOLENOID AMYLOIDS

## 5.1 Introduction

Amyloids are β–sheet rich protein aggregates and a pathological hallmark in several fatal human diseases.[223] Their unique self-templating ability to convert soluble native protein to insoluble aggregates helps them propagate as prions with relevance in many neurodegenerative diseases and essential biological activities.[53,467] About 40 pathologically relevant amyloids have been documented to date. One of the diagnostic strategies of progressive neurodegenerative diseases is centered on detecting biomarkers like amyloid aggregates. It is essential to develop conclusive tests to detect early-onset amyloid pathology.[468,469] Fluorescent dyes like ThioflavinT (ThT) and congo red are widely used for such accurate and conclusive diagnosis of various amyloid-related diseases.[470–475] Congo red displays apple-green birefringence on binding amyloids under polarized light, whereas ThT exhibits characteristic redshift with enhanced fluorescence emission intensity upon amyloid binding (absorption at ~440 nm and emission at ~480 nm for ThT ).[471,476,477] Unlike Congo red, a kinetic inhibitor molecule, ThT is a very useful tool for probing amyloid aggregation kinetics.[478] ThT is a benzothiazole salt composed of a pair of benzothiazole and benzaminic rings freely rotating around a shared C–C bond (**Figure 5.1A**). The rotation around this bond results in fluorescence quenching of the ThT excited state. However, ThT binding on the amyloid cross–β architecture blocks the rotation around the shared C–C bond in the dye, resulting in enhanced fluorescence, and is considered the 'gold standard' for amyloid staining.[397,453,471,473,479–485] Various uncharged Thioflavin-derivatives with good brain entry and clearance are used for imaging and quantifying amyloid plaques associated with various neurodegenerative diseases.[486–489] However, many amyloids with cross–β architecture, including the model fungal prion HET-s (218-289), do not affect ThT fluorescence raising the prospect of false-negative results.[474,482,490–492]

Previous studies on the ThT binding modes on various amyloid fibrils have failed to explain why the ThT negative results occurs on amyloids like HET-s (218-289).[398,491–493] Sabate and co-workers have reported that unlike the fibrils assembled at pH 2, HET-s (218-289) fibrils formed at higher pH do not exhibit fluorescence, even though ThT do bind the fibrils as established by ThT anisotropy studies.[492] Based on Krebs's ThT binding model, it is postulated that the long axis of ThT orients parallel to the fibril axis in a planar conformation, embedded in the hydrophobic channels or grooves establishing interactions with solvent-exposed side-chains predominantly via hydrophobic interactions or π-stacking interactions with aromatic residues.[493–495] However, in the absence of the ThT binding mode in atomic resolution, the underlying mechanism of ThT fluorescence specificity on binding amyloids remains elusive.

Interestingly, unlike HET-s(218-289) fibrils, its homolog Fg-HET-s(218-289) is ThT positive, meaning ThT fluoresce on binding this amyloid architecture. It has been puzzling that the Fg-HET-s(218-289) fibrils are ThT positive, adopting a similar HET-s-like prion fold, suggesting ThT fluorescence is independent of the amyloid fold.[442] Remarkably, HET-s-like β-solenoid architecture is identified as the fundamental element in human prion proteins like PrP and tau in its amyloid fold.[80,405,496,497] Bioinformatic studies have pointed out the commonality of β-solenoids and β-arches in amyloids as structural elements.[498–501] A systematic research is essential to discover novel tracer molecules to accurately and

conclusively diagnose amyloids or for the rational design of specific protein aggregation inhibitors in a therapeutical perspective. So, we set out to determine the structural and molecular basis of tertiary-fold independent ThT fluorescence in amyloids adopting a β-solenoid prion fold.

Here, we use our success in determining the structural model for NWD2 functional amyloids and its heterotypic amyloid interface R0/R2 with HET-sPFD involved in the signal transduction in the filamentous fungus, *Podospora anserina*.[250] We examined the deterministic factors of fluorescence property of ThT molecules upon binding different β–solenoid amyloid systems, namely HET-s(218-289), R0-R0, R0-R1, R0-R2, and R0-R1-R2 (**Figure 5.1B**). These are chimeric constructs of amyloid motifs involved in prion-based signal transduction in the NWD2/HET-s paradigm.[249,441,443,502] Recent *in vivo* functional evidence, bioinformatics data, and existing structural evidence demonstrate that all these functional amyloids adopt a left-handed β–solenoid fold invitro (also explained in **chapter 3**).[34,236,249,250,441,443] The R0-R0, R0-R1, R0-R2, and R0-R1-R2 are composed of the elementary repeats of HET-s PFD(R1, R2) and NWD2 (R0), with each repeat unit separated by the 15 amino acid long flexible loop in HET-s PFD.[34] Although these constructs were designed to understand the NWD2/HET-s paradigm, they came in handy to determine the molecular mechanism of ThT binding on amyloids. Each repeat in these chimeric constructs can adopt a triangular β–solenoid turn and stack on top of each other to form mature protofibrils (**Figure 5.1C**).[249,441]

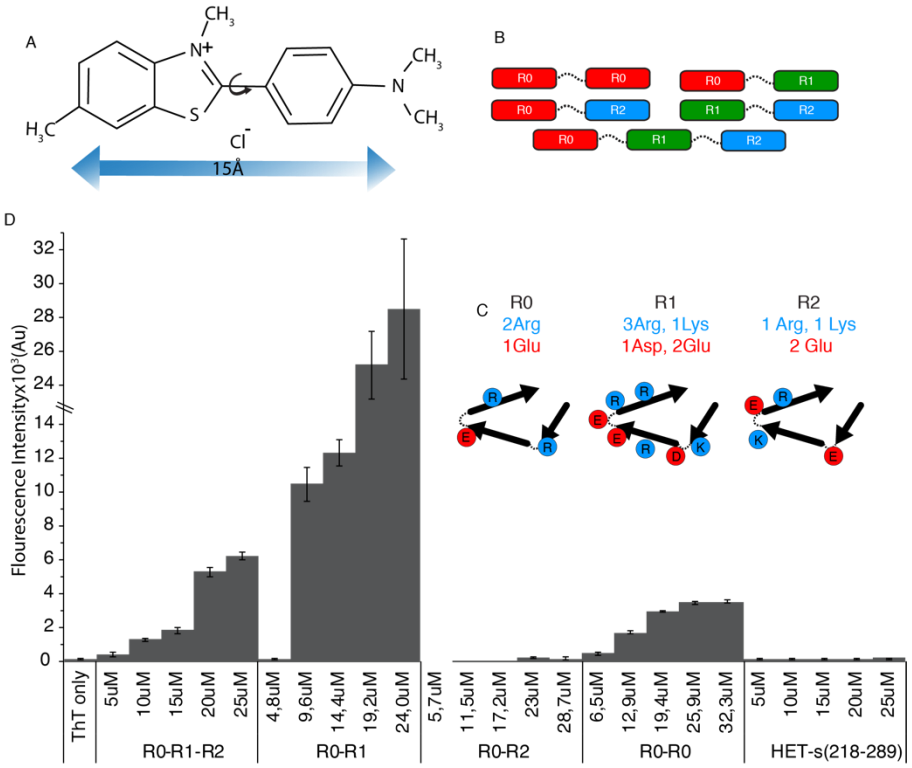



*Figure.5.1 A) chemical structure of Thiolfavin-T molecule and a model showing its head to the tail distance of 15Å. B) Sequence construction strategy of functional amyloids, namely R0-R0, R0-R1, R0-R2, and R0-R1-R2. C) Cartoon representation of charged residue in each amyloid motif.D) Concentration-dependent ThT-fluorescence assay on R0-R0, R0-R1, R0-R2, and R0-R1-R2 amyloid fibrils.*

**5.2 Results**

We conducted the ThT fluorescence experiment on each construct as described in the materials and methods. Interestingly, only R0-R0, R0-R1, and R0-R1-R2 fibrils were found to be ThT positive (i.e., ThT display fluorescence on binding these amyloids). In contrast, HET-s(218-289) and R0-R2 fibrils exhibited very low or no ThT fluorescence(**Figure 5.1D**) at near physiological pH (pH 7.5). Strikingly, ThT displays variations in quantum yield on binding different amyloid fibrils, exemplified by a greater fluorescence intensity on binding R0-R1 fibrils than other chimeric constructs.However, The molecular basis of the variation in ThT fluorescence intensity is poorly understood. It is proposed that the fluorescence quantum yield depends on many variables, such as the difference in binding mode, the number of binding sites available (stochiometry), binding affinity, and the ThT conformation.[472,503,504] The ThT conformation in the excited-state dimer (excimer) formed via binding in cavities is shown to exhibit enhanced fluorescence. [490,503,504]

Although all these NWD2/HET-s chimeric amyloids adopt a HET-s-like prion fold, the determining factors affecting the discrepancy in ThT fluorescence quantum yield is baffling. Based on Krebs's model of ThT binding on amyloids and the β–solenoid structural model at pH 7.5, nine ThT accessible channels or grooves can be formulated, located between the side-chains of i and i+2 residues pointing towards the solvent-exposed surface.[34,492,494] A cationic ThT fluorophore would preferably avoid fibril channels that possess positively charged residues like arginine or lysine due to electrostatic repulsion.[492] A difference in charge distribution on the fibril grooves in each chimeric constructs could play a critical role in recruiting ThT molecules and a planar conformation change, exhibiting variation in ThT fluorescence intensity (**Figure 5.1C**). The ThT anisotropy measurements on HET-s(218-289) fibrils have demonstrated that ThT binding depends on pH and ionic strength. A poor ThT-binding has been observed on HET-s(218-289) fibrils at a highly acidic pH due to the charge repulsion, which effectively gets shielded at high salt concentrations, enhancing the ThT binding affinity.[491,492] An alternate binding mode in HET-s(218-289) fibrils that do not arrest the rotation of aryl rings in ThT could also explain its ThT negative nature can also be proposed.[26]

Additionally, a comparison of XRD data on ThT bound and free R0-R0 and HET-s(218-289) amyloid fibrils display similar XRD patterns corresponding to two-rung β–solenoid structures with a reflection at 4.7 Å corresponding to intra-strand spacing and a second reflection near 9.4 Å revealing that the repeating unit of these fibrils is a two-strand repeat (Figure 5.2).[451,505] However, several diffraction patterns arising in ThT bound R0-R0 amyloid fibrils at ~5-6 Å suggests the possibility of intra-strand spacing alteration caused by ThT integration or changes in protofilament structure.[398,472] These extra signal intensities can originate from ThT itself, as the ThT powder display a series of reflections at 4.4 Å, 4.9 Å, 6.5 Å, 6.6 Å, 8.7 Å, 8.8 Å, 9.4 Å, and 9.5 Å.[398] Nevertheless, no significant change is observed for intensities at 10Å corresponding to inter-sheet spacing, suggesting the ThT has not aligned perpendicular to the fibril axis. Surprisingly, no considerable reflections are detected at 15 Å corresponding to amyloid incorporated ThT or at ~10-12 Å corresponding to the altered two β-strand's repeat spacing. These results is challenging to obtain a conclusive statement regarding ThT influence on the intra-strand spacing of β–solenoid amyloid fibrils. Nevertheless, at some level, structural changes are observed in the R0-R0 and HET-s (218-289) fibrils while accommodating ThT. As the XRD data is based on the ensemble average, a low amount of fibrils with a regular arrangement of ThT could lead to the low intensity of these patterns. After the fibrillization step, the ultracentrifugation spin-down assay suggests

that only a few ThT molecules are bound to the β–solenoid fibrils with a protein: ThT ratio of ~12:1. Also, the lack of instrumental sophistication to obtain better resolution hinders further analysis. Nevertheless, a detailed study on ThT bound amyloid fibrils is in progress by varying the hydration of amyloid fibrils and aligning the fibrils on to grids to obtain better resolution.

To identify the protein surface that interacts with ThT in β–solenoid forming amyloids, we used SSNMR spectroscopy data-based molecular modeling. We selected NWD2 and HET-s(218-289) amyloids to be the candidates to determine the difference in ThT binding modes by which NWD2 fibrils are ThT positive, whereas HET-s(218-289) fibrils are ThT negative. HET-s and NWD2 amyloid fibrils display a single set of resonances of spin systems (amino acid residues) consistently, implying the lack of structural polymorphism, providing excellent spectral resolution.[34,236,506] Moreover, the availability of a 3D structural model makes them the perfect amyloid system to detect a characteristic chemical-shift perturbation (CSP) upon ThT binding by comparing the SSNMR fingerprints of bound and unbound [$^{13}$C,$^{15}$N]-labeled fibrils.[34,236] The CSP analysis of two-dimensional $^{13}$C–$^{13}$C correlation spectra (proton-driven spin diffusion, PDSD) and $^{15}$N-$^{13}$C correlation spectra (NCA) can be used to distinguish perturbation in the local chemical environment of side-chain atoms and protein backbones upon ligand binding. [507–509] However, CSP data should be considered carefully, as they could originate either due to direct ligand interaction or by allosteric effects.

To find the possible ThT binding site(s) on HET-s(218-289) fibril, we first compared the 2D $^{13}$C–$^{13}$C correlation spectra (proton-driven spin diffusion, PDSD) for ThT bound (green) and free (black) HET-s(218-289) fibrils(**Figure 5.3**). A non-specific perturbation due to temperature and pH effects can be ruled out since there are no significant CSPs for Ser and Thr residues in ThT bound and free HET-s(218-289) fibril spectra. Remarkably, no significant CSPs are observed for the residues inside the triangular hydrophobic core, suggesting ThT biding is outside the central hydrophobic core and does not change the amyloid core structure. This inference is anticipated, as the hydrophobic core is tightly packed(dry face) and not spacious enough to accommodate 15 Å long ThT molecules. A cut-off of 0.2ppm was set to characterize extensive CSPs, suggesting a structural change. On comparing the 2D $^{13}$C-$^{13}$C spectra of free and ligand-bound HET-s (218-289) fibrils, we observed strong CSPs on the residues 246V, 263 S, 264V, 281Y, 286F, and 287W (a representative excerpt of these regions of 2D $^{13}$C–$^{13}$C correlation spectra are shown in **Figure 5.4**).

Similarly, residues 281Y, 282G, 286F, and 287W displayed significant CSP in NCA spectra suggesting local chemical environment changes in the amide backbone in HET-s(218-289) fibrils on ThT binding (**Figure 5.5**). Interestingly, these CSPs in NCA spectra are localized for the aromatic residues (281Y, 286F, and 287W ), which belong to the semi-rigid segment, designing the semi-hydrophobic pocket not directly participating in HET-s(218-289) rigid amyloid core. However, this segment is identified as essential for HET-s prion propagation.[64,236] It is plausible that ThT molecules interact with these aromatic residues in the C-terminal semi-flexible loop via aromatic π–π stacking interaction and re-orients their peptide backbone resulting in localized CSPs in the NCA spectrum. Interestingly, major CSPs are observed in the $^{13}$C–$^{13}$C correlation spectra for this region upon ThT binding. However, the free rotation about the ThT aryl rings may not be hindered, responsible for its fluorescence by the ThT interaction with the semi-flexible C-terminal aromatic segment.

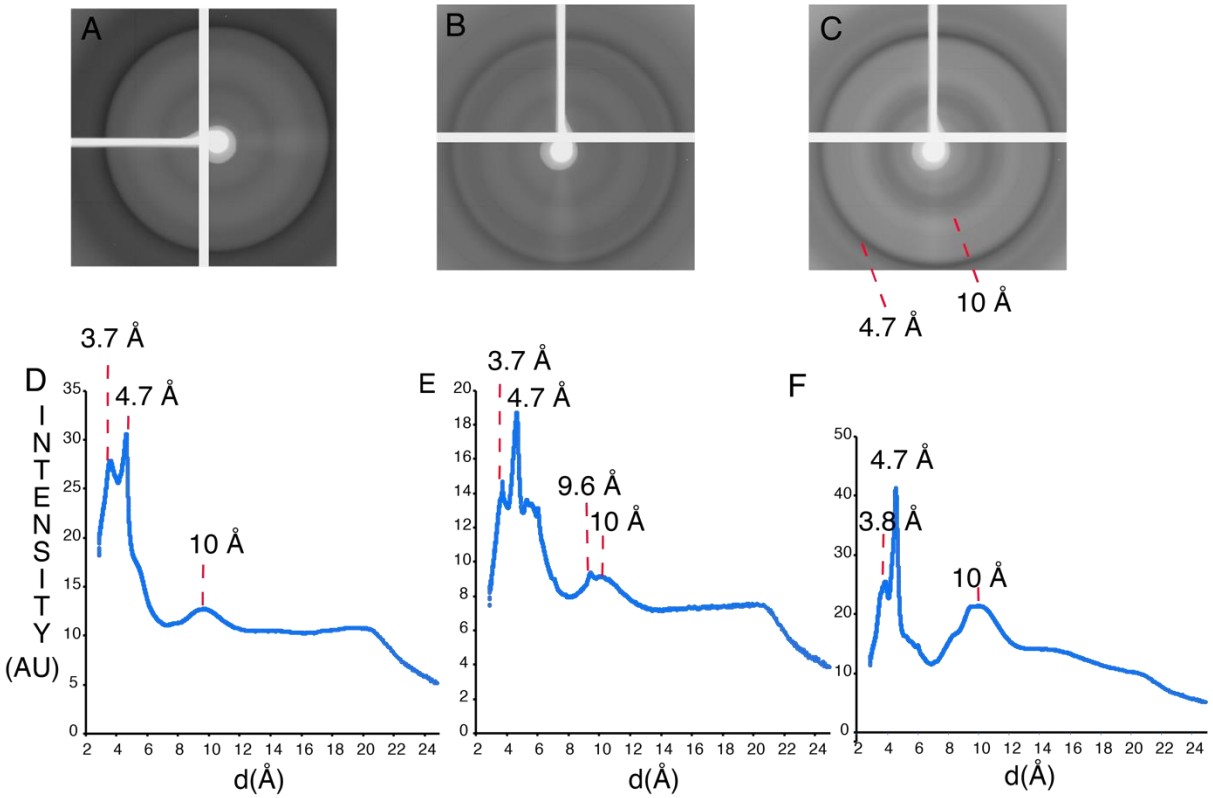



*Figure 5.2 XRD diffraction pattern of ThT free and bound amyloid fibrils. XRD diffraction pattern (A) R0-R0 amyloid fibrils, (B) ThT bound R0-R0 fibrils (C) ThT bound HET-s(218-289) amyloid fibrils. Reflections at 4.7 and 10 Å are highlighted corresponding to the interstrand and intersheet spacing, respectively. X-ray scattering intensity for (D) R0-R0 amyloid fibrils, (E) ThT bound R0-R0 fibrils, (F) ThT bound HET-s(218-289) amyloid fibrils are showing maxima at ~3.7, 4.7, and 10 Å.*

Further, we analyzed the CSPs in ThT bound R0-R0 amyloid fibrils. A cut off of 0.2ppm was set to characterize extensive CSPs for 2D $^{13}C$–$^{13}C$ correlation spectra and 0.3ppm for the NCA spectra to imply a structural change. As expected, the residues at the central hydrophobic core showed no significant CSPs, eliminating the possibility of ThT placed inside the core. The SSNMR CSP analysis of R0-R0 amyloid fibrils showed extensive CSPs for residues 9H,17H, 18V, 21S, 11E, and 19G in the 2D $^{13}C$–$^{13}C$ correlation spectra (**Figure 5.6**). Moreover, residues 21S, 9H, 17H, and 7S displayed CSP significant CSPs in the NCA spectra (**Figure 5.7**). Interestingly, these residues are surface exposed. Based on the SSNMR CSP analysis, we proposed two binding pockets for ThT in R0R0 fibrils; i) on the aperture formed by the 3$^{rd}$ and fourth β–sheets(β3-β4), ii) on the grooves formed by 9H -10A -11E on parallel stacking. The β–solenoid turns in R0-R0 amyloid fibrils display an average spacing of 4.8Å, and hence the ThT molecules could regularly align parallel to the fibril axis spanning four R0 turns(two R0-R0 monomers) (~14.5 Å in total), close to the head-to-tail length of a ThT molecule (~15Å), in line with the Krebs's model of ThT binding.[510] A similar binding mode of amyloid binding dyes (ThT and congo red) along the fibril axis has been documented for Aβ(1–42) fibrils and HET-s(218-289) fibrils.[493,507,511]

To investigate the molecular mechanism of ThT binding mode on β–solenoid amyloid fibrils, we are currently collaborating with B. Bardiaux and M. Ferber from Pasteur Institute to set up a molecular docking on ThT molecules on the SSNMR structure of HET-s. Chemical

shift perturbations are converted into residue-residue proximities and used as structural input in the modeling process.

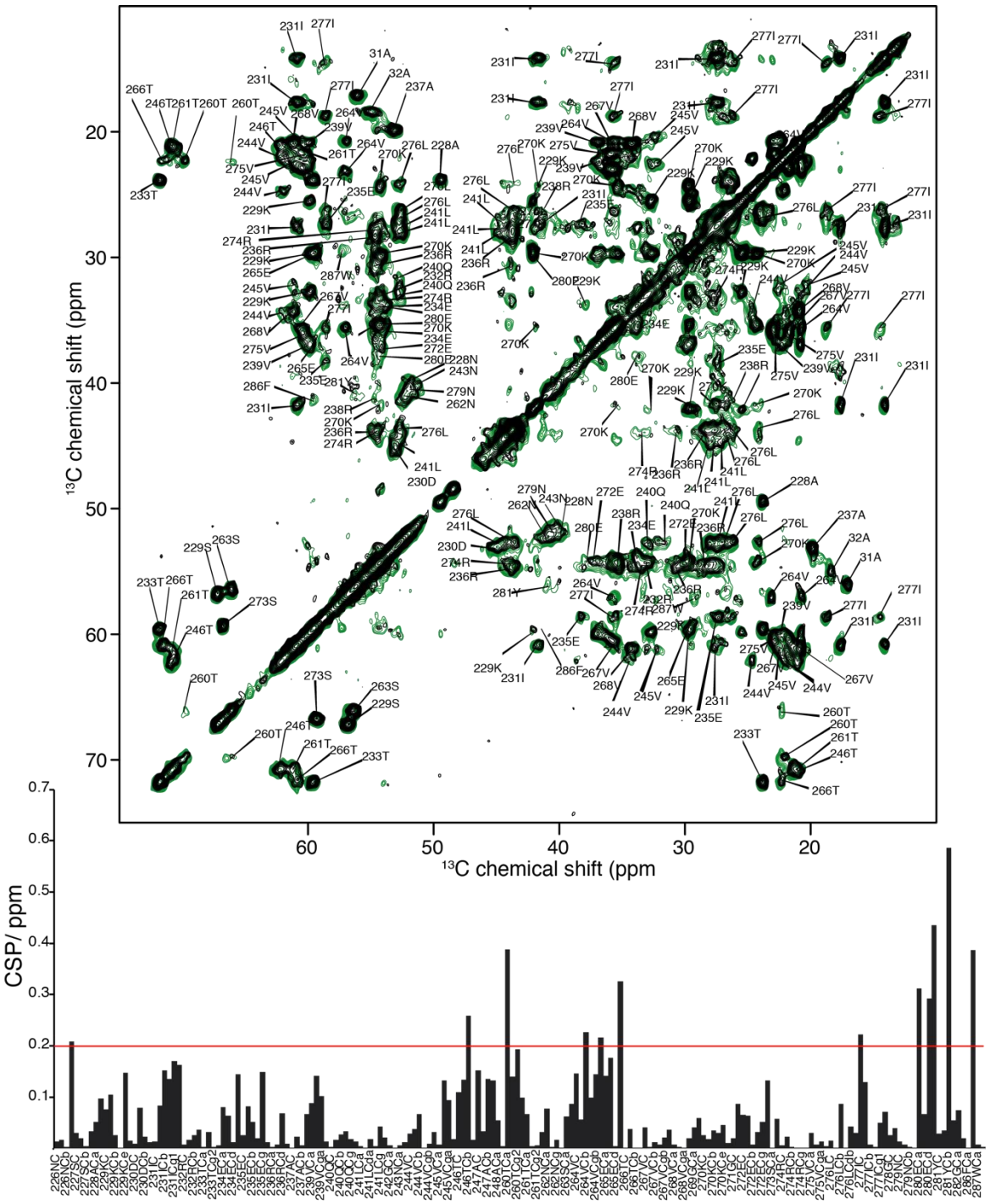



*Figure 5.3 Structural investigation of the ThT binding mode on HET-s(218-289). Top: Comparison of 2D $^{13}$C-$^{13}$C correlation experiment with PDSD mixing time of 50ms of bound (green) and free (black)HET-s(218-289) fibrils. Bottom: Chemical shift perturbation is measured as the difference of ThT bound and free HET-s(218-289) fibrils. A cut-off of 0.2 ppm is set for extensive CSP.*

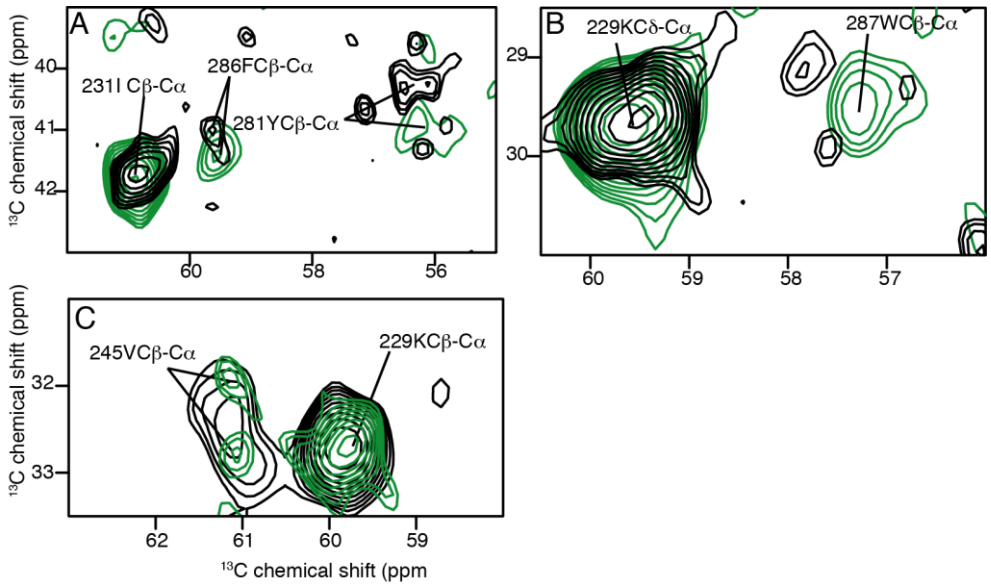



*Figure 5.4 Representative excerpt of regions of 2D-$^{13}$C–$^{13}$C correlation spectra with PDSD mixing time of 50ms, indicating CSP for ThT bound (green) and free (black) HET-s(218-289) fibrils.*

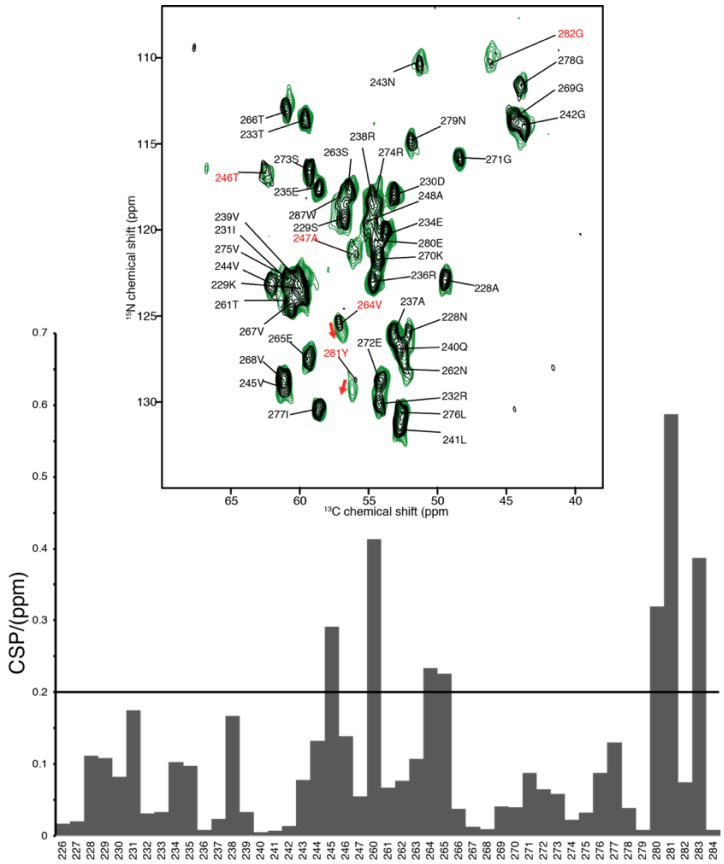



*Figure 5.5 Structural investigation of the ThT binding mode on HET-s(218-289) amyloids. Top: Comparison of two-dimensional SSNMR 2D NCA spectra of ThT bound(green) and free(black) HET-s(218-289) fibrils. Bottom: Chemical shift perturbation is measured as the difference of ThT bound and free HET-s(218-289) fibrils. A cut-off of 0.2 ppm is set for extensive CSP.*

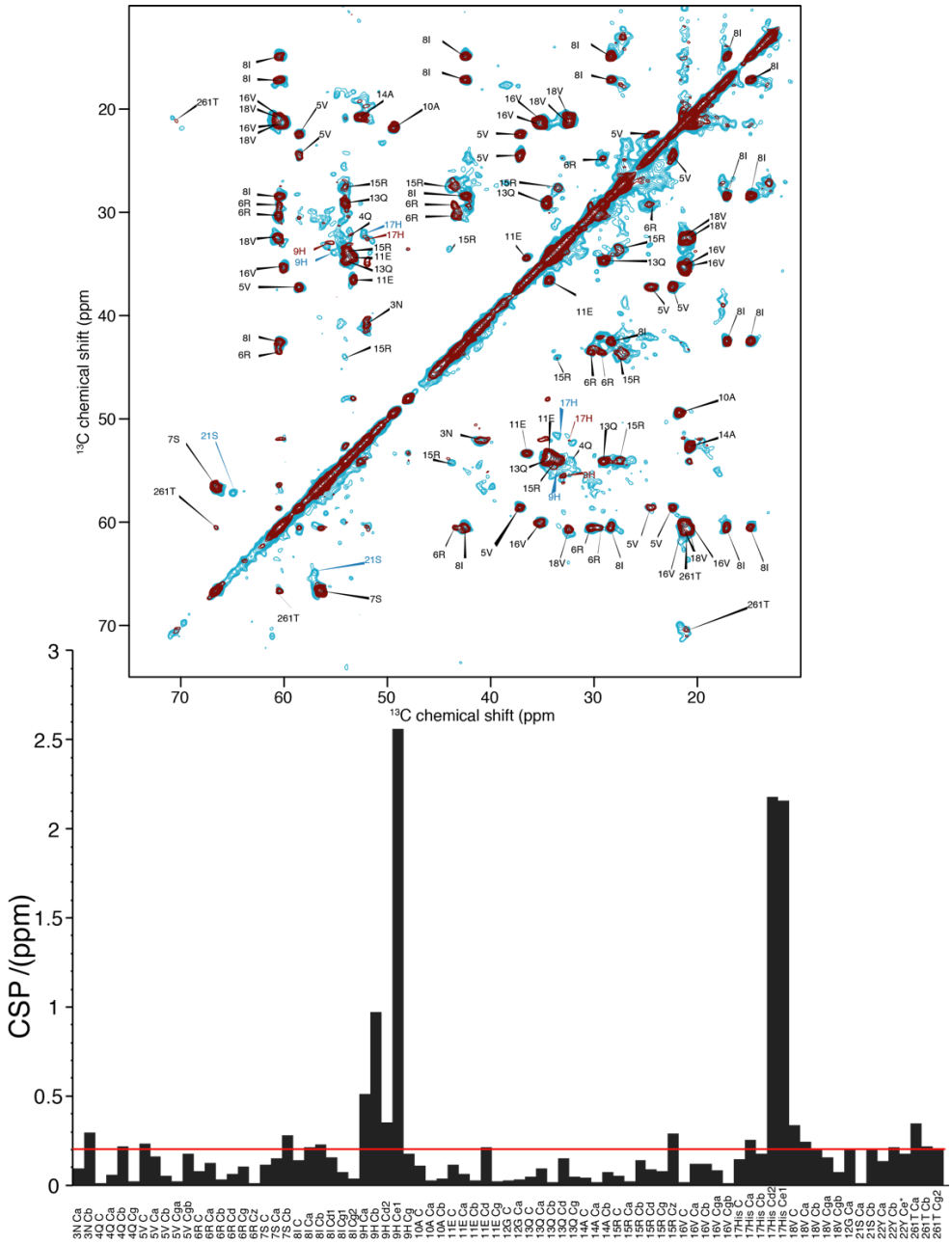



*Figure 5.6 Structural investigation of the ThT binding mode on R0-R0 functional amyloids. Top: Comparison of two-dimensional SSNMR 2D-$^{13}$C–$^{13}$C correlation spectra (proton-driven spin diffusion, PDSD) of ThT bound(brown) and free(cyan) R0-R0 fibrils. The residues with clear CSP or linewidth change are marked. Bottom: Chemical shift perturbation is measured as the difference of ThT bound and free R0-R0 fibrils. A cut-off of 0.2 ppm is set for extensive CSP.*

The ThT binding mode on grooves in R0-R0 amyloid fibrils could explain why ThT does not show fluorescence on binding HET-s(218-289), R0-R2 fibrils. Only a few hydrophobic channels are available in R0-R2 and HET-s(218-289) fibrils accessible for ThT due to unfavorable electrostatic repulsion from basic residues (lysines and arginines)at physiological pH. Moreover, the potential salt bridge formation between the glutamate carboxyl groups within the fibrils of R0-R2/HET-s(218-289) may compete with the stable formation of a salt bridge with ThT. Once the carboxyl group of Glutamate/Aspartate forms salt bridges within

the fibril, it leaves no binding option for the protonated head group of ThT. Hence, the absence of free acidic or aromatic residues and relatively higher presence of basic residues at solvent-exposed surfaces of R0-R2 and HET-sPFD fibrils hinders ThT binding. This inference regarding the basic residues (lysine or arginine) at the ThT accessible fibril grooves is in line with the observation of Sabate and co-workers, showing that ThT binding in HET-s fibrils formed at pH2 is dramatically affected by pH and ionic strength.[492] A similar argument can also be proposed for the ThT binding model on Fg-HET-s(218-289) fibrils owing to the presence of ThT accessible hydrophobic channels devoid of basic residues.[442]

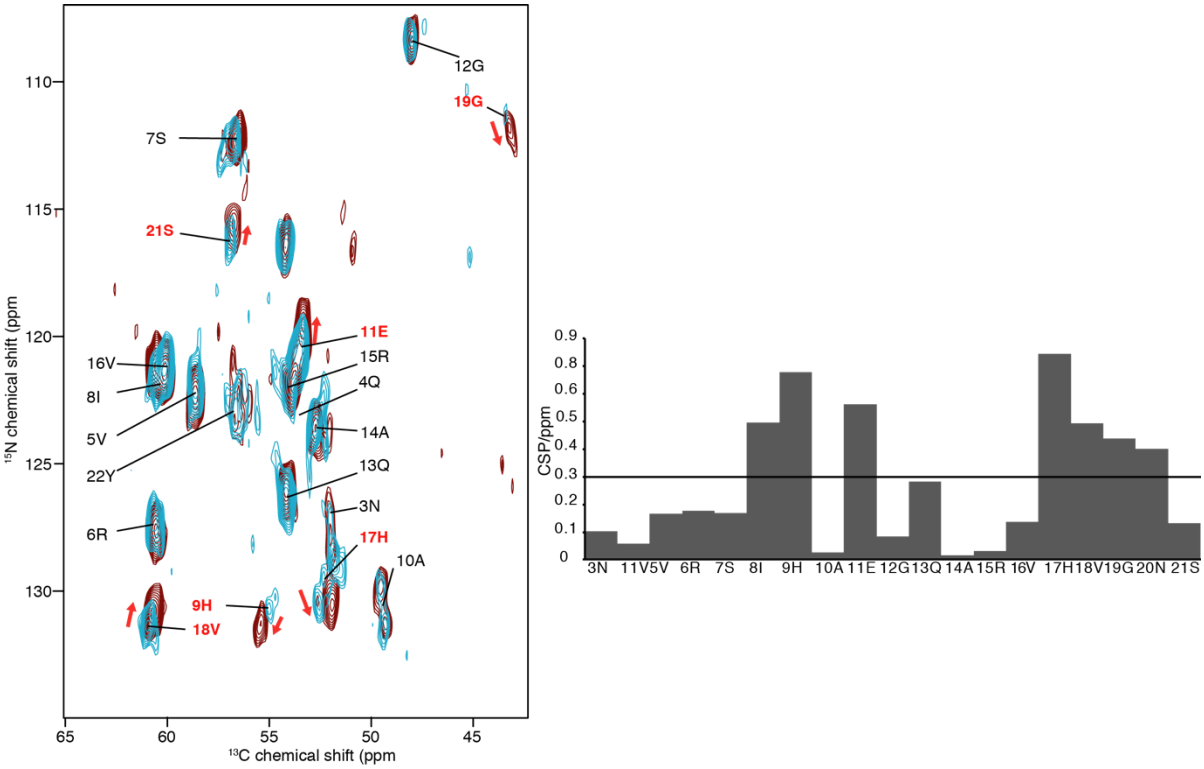



*Figure 5.7 Structural investigation of the ThT binding mode on R0-R0 functional amyloids. Left: Comparison of two-dimensional SSNMR 2D NCA spectra of ThT bound (brown) and free (cyan) R0-R0 fibrils. The residues with greater CSPs are indicated by red arrows. Right: Chemical shift perturbation is measured as the difference of ThT bound and free R0-R0 amyloid fibrils. A cut-off of 0.3 ppm is set for extensive CSP.*

On the contrary, the observation that the R0-R1 harboring six basic residues (5 arginines and one lysine) and being ThT positive and R0-R2 with only four basic residues (3 arginines and one lysine) being ThT negative is baffling. This effect can only be explained by the presence of free acidic residues at the external axial channels of R0-R1 fibrils(**Figure 5.1C**). In R0-R2 fibrils, all the acidic residues present in solvent-accessible grooves form salt bridges within the fibril (E11-K270 and E265-R6) or are surrounded with basic residues (E272 with R15 and R274). As a result, no ThT accessible acidic residues would be available in R0-R2 fibril to form a stabilizing electrostatic interaction. The acidic residues (3 glutamates and one aspartate) at the external axial channels of R0-R1 fibrils may help recruit the free ThT molecules from the solvent to the hydrophobic sites by electrostatic interaction and display enhanced fluorescence.

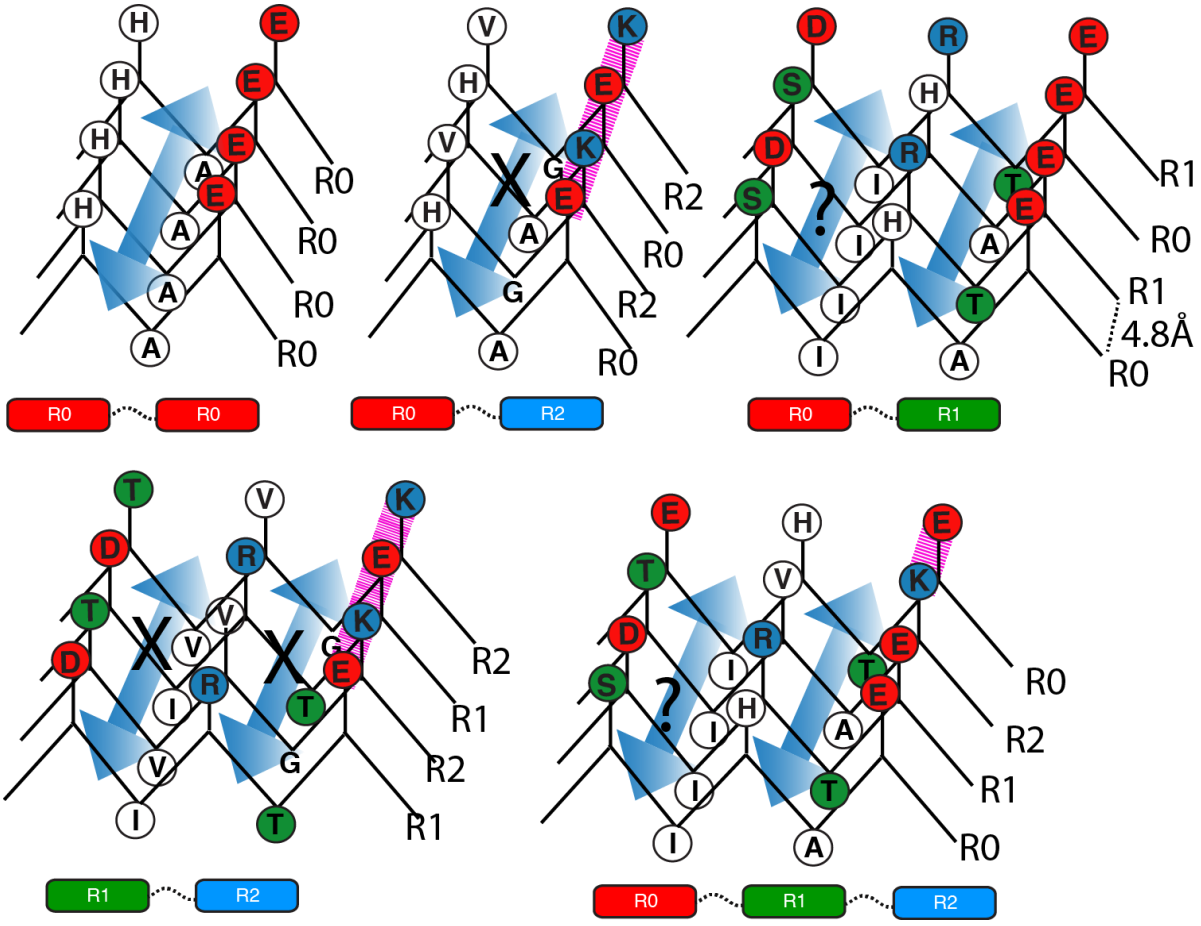



FIGURE 5.8

*Figure 5.9. A schematic model of ThT binding mode on different β–solenoid amyloid fibrils. The proposed binding site of ThT on the channels/grooves formed by side-chains of i and i + 2 residues of each β–stand in the cross–β amyloid structure fibrils running along the long axis. The salt bridges within the fibrils are marked by pink dashed lines.*

Interestingly, ThT shows a lower fluorescence intensity on binding R0-R1-R2 fibrils than R0-R1 but higher than R0-R0 fibrils. This phenomenon can partially be explained by the alternate presence of salt-bridges and free acidic residues in the ThT accessible fibril grooves. The reduced aromaticity (H to V), presence of basic residues compared to R0-R1 could lead to the lower ThT fluorescence of R0-R1-R2 fibrils. However, our study is limited by one binding site model as R0-R1 and R0-R1-R2 may have multiple ThT binding sites as there are other hydrophobic channels (Figure 5.8). Also, The ThT fluorescence quantum yield cannot be explained by only looking into binding sites (as we do here), as this phenomenon also depends on other factors such as the difference in binding mode, the number of binding sites available (stochiometry), and the ThT conformation.[472,503] The ThT conformation can play a critical role in fluorescence quantum yield, but that information can only be obtained by quantum-chemical calculations, which is beyond the scope of this thesis.[479,480,512–514]

In summary, we observed the deterministic interaction of ThT on β–solenoid adopting amyloids facilitating enhanced fluorescence and hence an accurate measurement of aggregation kinetic parameters. Amyloid surfaces with aromatic residues, hydrophobicity, and acidic residues improve accessibility for ThT binding, where basic residues hinder ThT binding. The lack of aromatic residues and charge complementarity in R0-R2 and HET-s(218-

289) at the grooves make them ThT negative. This inference could also explain the differential magnitude in ThT intensity on β–solenoid amyloids. Interestingly, ThT shows a very high order of magnitude in fluorescence intensity upon binding R0-R1 amyloids in similar conditions. This effect can be explained by the lack of salt bridges between the amyloid β-turns, resulting in hydrophobic channels with free acidic residues like glutamate (E11) and aspartate (D230) compatible for an ion pair interaction with positively charged ThT. However, the ThT quantum yield differs with binding mode, the number of binding sites available (stochiometry), binding affinity, and the ThT conformation.[472,503] Our approach based on SSNMR experiments compares only a single mode of binding in different β–solenoid amyloids and the relative binding affinity. As a result, our current model could not explain the variation in fluorescence intensity on binding different amyloid fibrils. However, our ThT binding model can explain the deterministic interaction in ThT binding on β–solenoid fibrils and predict the possible binding for ThT. The presence of acidic residues, glutamate (E11), and aspartate (D230) could facilitate an additional groove compatible for ThT binding via salt-bridge formation with ThT and π-π stacking with H9. Expected results from molecular modeling will help us to provide an atomistic view of the ThT-amyloid binding site.

### 5.3 Materials and methods

Protein constructs ($^{13}C/^{15}N$ or unlabeled) HET-s(218-289), R0-R0, R0-R1, R0-R2, and R0-R1-R2 are expressed and purified, as mentioned in chapter 3. Thioflavin T was purchased from Sigma Aldrich.

#### 5.3.1 ThT fluorescence experiments

Each protein was desalted to 1% acetic acid and made a stock solution of 34uM. A series dilution of protein was created and supplied with 15uM ThT and 23%(v/V) 3M tris pH 8 to change the pH and trigger aggregation. All these sample manipulations were done at four °C or in ice to avoid any pre-aggregates. The samples were then transferred to 96 well corning plates (bottom transparent) with 100uL/well. The kinetics were initiated by placing the 96-well plate at 25°C and shaking at 100 rpm in a CLARIOstar®Plus plate reader and measured with different gains. The Th-T fluorescence measurements were made every 5 minutes (with excitation wavelength 449 nm and emission wavelength 482 nm) with shaking after each scan. The ANS fluorescence was measured similarly with an excitation wavelength of 355 nm and an emission wavelength of 520 nm. Fluorescence intensity at the end of fibrillization marked by plateau has plotted against the concentration of each construct.

#### 5.3.2 Spin-down Assay

Excess of (~50uM) of protein samples are incubated with 15uM ThT at 25°C overnight. The samples were ultracentrifuged at 60 000 rpm, and the amount of ThT in the supernatant was measured by absorbance at 412nm ($\epsilon_{412}$ = 31600 $M^{-1}cm^{-1}$). Later, the amount of ThT bound to the fibrils is calculated from this.

#### 5.3.3 Invitro fibrillization of functional amyloids with ThT for SSNMR

The in vitro self-assembly of NWD2/HET-s samples supplied with ThT in 1:10 molar ratio (Protein: ThT) in 1% acetic acid was initiated by rapidly changing the pH to 7.5 by adding

3M Tris buffer, pH8. These samples were allowed to self-assemble for one week at room temperature, then extensively washed in double distilled water, centrifuged at 13500 g, and transferred to the SSNMR rotor. The control sample was prepared and treated identically through all steps without adding ThT, as described in chapter 3.

### 5.3.4 X-ray diffraction

XRD analysis was carried out as described in Chapter 3.

### 5.3.5 NMR measurements

All SSNMR spectra were recorded on a Bruker Avance 900 or 600 spectrometers with a Bruker 3.2 mm triple-resonance probe spinning at 11 kHz and at sample temperature ~278 K. The two-dimensional $^{13}C$-$^{13}C$ correlation spectra using a proton-driven spin-diffusion (PDSD) mixing were recorded similar to the description in chapter 3. The two-dimensional NCA spectra were recorded with a CP scheme with 4 ms $^{15}N$-$^{13}C$ contact time.

# Conclusions

We successfully employed SSNMR techniques for the structural characterization of pathological and functional amyloids, where the conventional structure determination tools like X-ray crystallography and solution-NMR fail to achieve. Here, we have presented a combination of electron microscopy, X-ray fiber diffraction, Fourier-transform infrared spectroscopy analysis, and state of the art SSNMR techniques for the structural studies of pathological amyloid formed by TDP-43 associated with ALS and the molecular mechanism of signal transduction via functional amyloid proteins NWD2/HET-s involved in Programmed cell death in fungi. We also describe the molecular basis of amyloid binding dye Thioflavin T (ThT) on amyloid fibrils based on SSNMR experiments.

In Chapter 3, to investigate the molecular organization and obtain an insight into the different polymorphs of different TDP constructs, we used a combination of SSNMR and other biophysical techniques. Our SSNMR analysis of the different TDP constructs provides additional experimental evidence towards the importance of the C-terminal domain in amyloidosis, which can assume different amyloid conformation depending on its environment (in the full-length construct or isolated fragments). However, the different C-terminal fragments (CTFs) adopt different polymorphs maintaining a typical cross-β fibrillar architecture. We identified several residues in the so-called C-terminal GaroS2 (368–414) domain that participates in the rigid core of TDP-16 fibrils, underlining its importance during the aggregation process. We also identified several β–breaker segments located at the hydrophobic, Q/N rich, and GaroS2 subdomains in the amyloid fold of a CTF, namely TDP-13 (300-414). Several aromatic residues, including W334 and F367 paired with glycines, are crucial in the amyloid core. Also, the TDP-13 construct is shown to engage in alternative packing arrangements in response to variations in aggregation conditions. Our SSNMR studies of TDP-43 isoforms (sTDP43-1, sTDP43-2) responsible for ALS demonstrate that these isoforms can form amyloids without CTD and adopt an amyloid fold similar to full-length TDP-43, suggesting a misfolding mechanism of NTD and RRM domains responsible for the amyloid transition in TDP43 isoforms. Our work paves the way for understanding the role of the N-terminal and RRM domains in the TDP-43 amyloid transition. It also raises the question about the protective nature of low complexity domains (LCDs) amyloid protein aggregation.

In Chapter 4, we investigated the role of β–solenoid fold in fungal NLR function and signal transduction via a heterotypic amyloid interface in immunity-related cell fate pathways. In this work, based on SSNMR restraints, we deduced the 3D model of the β–solenoid fold adopted by the functional amyloid NWD2 (R0) and the R0/R2 heterotypic amyloid interface involved in the prion-based templating mechanism in the filamentous fungi *Podospora anserina*. Our results suggest the ability to form the β-solenoid fold efficiently, sequence variation of elementary R-motifs is finely tuned at the sequence level, with (i) an R1 tailored to a lower the amyloid-forming ability potentially to avoid self-activation of HET-S and (ii) R0 reduced in charge content to allow for homotypic polymerization and (iii) capable of forming different heterotypic interfaces with either R1 or R0 with the R0/R2 interface potentially conductive to slightly lower stability to allow for recycling of the NWD2 signalosome, enabling signal amplification.

In Chapter 5, we examined the deterministic factors of fluorescence property of ThT molecules upon binding on β–solenoid forming amyloid systems. The ThT preferably binds on solvent-exposed fibril grooves composed of acidic residues compatible with the salt-bridge formation or aromatic residues for π–π stacking. As Thioflavin-derived compounds are used for qualitative and quantitative analysis of amyloid plaques, our results would pave the way for systematic research to discover novel tracer molecules to diagnose amyloid diseases accurately and conclusively. This study would also help for a rational design of specific protein aggregation inhibitors from a therapeutical perspective.

In summary, my thesis work enables a better understanding of amyloid folds and polymorphism in functional and pathological amyloid proteins. The results are in line with the general concept of functional amyloids being finely tuned at the sequence level that direct to a single unique polymorph with a well-executed amyloid transition. In contrast, the pathological amyloid exhibits structural plasticity due to spontaneous misfolding and aggregation of proteins caused by disease mutations or an inhospitable cellular environment. The variation in the polymorphism of pathological amyloids correlates to differential phenotypic features of the disease condition.

# REFERENCES


1. Dobson, C. M. Protein folding and misfolding. *Nature* (2003). doi:10.1038/nature02261

2. Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nature Structural and Molecular Biology* (2009). doi:10.1038/nsmb.1591

3. Mallamace, F. *et al.* Energy landscape in protein folding and unfolding. *Proc. Natl. Acad. Sci. U. S. A.* (2016). doi:10.1073/pnas.1524864113

4. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct. Funct. Bioinforma.* (1995). doi:10.1002/prot.340210302

5. Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. *Current Opinion in Structural Biology* (2004). doi:10.1016/j.sbi.2004.01.009

6. Selkoe, D. J. Folding proteins in fatal ways. *Nature* **426**, 900–904 (2003).

7. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* (2015). doi:10.1038/nrm3920

8. Uversky, V. N. Intrinsically Disordered Proteins. in *Brenner's Encyclopedia of Genetics: Second Edition* (2013). doi:10.1016/B978-0-12-374984-0.01698-3

9. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology* (2005). doi:10.1038/nrm1589

10. Van Der Lee, R. *et al.* Classification of intrinsically disordered regions and proteins. *Chemical Reviews* (2014). doi:10.1021/cr400525m

11. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).

12. Tompa, P. Intrinsically disordered proteins: A 10-year recap. *Trends in Biochemical Sciences* (2012). doi:10.1016/j.tibs.2012.08.004

13. Majumdar, A., Dogra, P., Maity, S. & Mukhopadhyay, S. Liquid-Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* (2019). doi:10.1021/acs.jpclett.9b01731

14. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* **86**, 27–68 (2017).

15. Dobson, C. M. The Amyloid Phenomenon and Its Links with Human Disease. *Cold Spring Harb Perspect Biol* (2017). doi:10.1101/cshperspect.a023648

16. Greenwald, J. & Riek, R. Biology of amyloid: Structure, function, and regulation. *Structure* (2010). doi:10.1016/j.str.2010.08.009

17. Friedreich, N. & Kekulé, A. Zur Amyloidfrage. *Arch. für Pathol. Anat. und Physiol. und für Klin. Med.* (1859). doi:10.1007/BF01945246

18. Toyama, B. H. & Weissman, J. S. Amyloid Structure: Conformational Diversity and Consequences. *Annu. Rev. Biochem.* (2011). doi:10.1146/annurev-biochem-090908-120656

19. Nilsson, M. R. Techniques to study amyloid fibril formation in vitro. *Methods* (2004). doi:10.1016/j.ymeth.2004.03.012

20. SUNDE, M. & BLAKE, C. The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. in

*Advances in Protein Chemistry Volume 50* (2008). doi:10.1016/s0065-3233(08)60320-4

21. Nesterov, E. E. *et al.* In vivo optical imaging of amyloid aggregates in brain: Design of fluorescent markers. *Angew. Chemie - Int. Ed.* (2005). doi:10.1002/anie.200500845

22. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273**, 729–739 (1997).

23. Astbury, W. T., Dickinson, S. & Bailey, K. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* (1935). doi:10.1042/bj0292351

24. Geddes, A. J., Parker, K. D., Atkins, E. D. & Beighton, E. "Cross-beta" conformation in proteins. *J. Mol. Biol.* (1968).

25. Eanes, E. D. & Glenner, G. G. X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* (1968). doi:10.1177/16.11.673

26. Makin, O. S., Atkins, E., Sikorski, P., Johansson, J. & Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U S A* **102**, 315–320 (2005).

27. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* (2005). doi:10.1038/nature03680

28. Guenther, E. L. *et al.* Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat Struct Mol Biol* (2018). doi:10.1038/s41594-018-0064-2

29. Wiltzius, J. J. W. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nat. Struct. Mol. Biol.* (2009). doi:10.1038/nsmb.1643

30. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* (2007). doi:10.1038/nature05695

31. Shenoy, J. *et al.* Structural dissection of amyloid aggregates of TDP-43 and its C-terminal fragments TDP-35 and TDP-16. *FEBS J.* **287**, (2020).

32. Tsemekhman, K., Goldschmidt, L., Eisenberg, D. & Baker, D. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.* (2007). doi:10.1110/ps.062609607

33. Mompean, M., Baralle, M., Buratti, E. & Laurents, D. V. An Amyloid-Like Pathological Conformation of TDP-43 Is Stabilized by Hypercooperative Hydrogen Bonds. *Front Mol Neurosci* **9**, 125 (2016).

34. Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science (80-. ).* **319**, 1523–1526 (2008).

35. Ferrone, F. Analysis of protein aggregation kinetics. *Methods Enzymol.* (1999). doi:10.1016/S0076-6879(99)09019-9

36. Morris, A. M., Watzky, M. A. & Finke, R. G. Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature. *Biochimica et Biophysica Acta - Proteins and Proteomics* (2009). doi:10.1016/j.bbapap.2008.10.016

37. Arosio, P., Knowles, T. P. J. & Linse, S. On the lag phase in amyloid fibril formation. *Phys. Chem. Chem. Phys.* (2015). doi:10.1039/c4cp05563b

38. Garcia, G. A., Cohen, S. I. A., Dobson, C. M. & Knowles, T. P. J. Nucleation-conversion-polymerization reactions of biological macromolecules with prenucleation clusters. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* (2014). doi:10.1103/PhysRevE.89.032712

39. Jarrett, J. T., Berger, E. P. & Lansbury, P. T. The Carboxy Terminus of the β Amyloid Protein Is Critical for the Seeding of Amyloid Formation: Implications for the Pathogenesis of Alzheimer's Disease.

*Biochemistry* (1993). doi:10.1021/bi00069a001

40. Uversky, V. N., Li, J. & Fink, A. L. Evidence for a Partially Folded Intermediate in α-Synuclein Fibril Formation. *J. Biol. Chem.* (2001). doi:10.1074/jbc.M010907200

41. Pease, L. F. *et al.* Probing the nucleus model for oligomer formation during insulin amyloid fibrillogenesis. *Biophys. J.* (2010). doi:10.1016/j.bpj.2010.10.010

42. Serio, T. R. *et al.* Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science (80-. ).* (2000). doi:10.1126/science.289.5483.1317

43. Lee, J., Culyba, E. K., Powers, E. T. & Kelly, J. W. Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. *Nat. Chem. Biol.* (2011). doi:10.1038/nchembio.624

44. Cremades, N. *et al.* Direct observation of the interconversion of normal and toxic forms of α-synuclein. *Cell* (2012). doi:10.1016/j.cell.2012.03.037

45. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nature Chemical Biology* (2009). doi:10.1038/nchembio.131

46. Banci, L. *et al.* Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M506637200

47. Eakin, C. M., Berman, A. J. & Miranker, A. D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat. Struct. Mol. Biol.* (2006). doi:10.1038/nsmb1068

48. Canet, D. *et al.* Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Biol.* (2002). doi:10.1038/nsb768

49. Törnquist, M. *et al.* Secondary nucleation in amyloid formation. *Chem. Commun.* (2018). doi:10.1039/c8cc02204f

50. Knowles, T. P. J. *et al.* An analytical solution to the kinetics of breakable filament assembly. *Science (80-. ).* (2009). doi:10.1126/science.1178250

51. Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* (2013). doi:10.1038/nature12481

52. Polymenidou, M. & Cleveland, D. W. The seeds of neurodegeneration: Prion-like spreading in ALS. *Cell* (2011). doi:10.1016/j.cell.2011.10.011

53. Prusiner, S. B. Molecular biology of prion diseases. *Science (80-. ).* (1991). doi:10.1126/science.1675487

54. Prusiner, S. B. Prions. in *Proceedings of the National Academy of Sciences of the United States of America* (1998). doi:10.1073/pnas.95.23.13363

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

56. Goedert, M. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Aβ, tau, and α-synuclein. *Science* (2015). doi:10.1126/science.1255555

57. Lindquist, S. *et al.* Investigating protein conformation-based inheritance and disease in yeast. in *Philosophical Transactions of the Royal Society B: Biological Sciences* (2001). doi:10.1098/rstb.2000.0762

58. Aguzzi, A. Unraveling prion strains with cell biology and organic chemistry. *Proceedings of the National Academy of Sciences of the United States of America* (2008). doi:10.1073/pnas.0710824105

59. Dong, J. *et al.* Probing the role of PrP repeats in conformational conversion and amyloid assembly of chimeric yeast prions. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M704952200

60. Newby, G. A. & Lindquist, S. Blessings in disguise: Biological benefits of prion-like mechanisms. *Trends in Cell Biology* (2013). doi:10.1016/j.tcb.2013.01.007

61. Kodali, R., Williams, A. D., Chemuru, S. & Wetzel, R. Aβ(1-40) forms five distinct amyloid structures whose β-sheet contents and fibril stabilities are correlated. *J. Mol. Biol.* (2010). doi:10.1016/j.jmb.2010.06.023

62. Klement, K. *et al.* 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**, 1321–1333 (2007).

63. Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D. & Riek, R. Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew. Chemie - Int. Ed.* (2010). doi:10.1002/anie.201000068

64. Daskalov, A. *et al.* Contribution of Specific Residues of the β-Solenoid Fold to HET-s Prion Function, Amyloid Structure and Stability. *PLoS Pathog.* (2014). doi:10.1371/journal.ppat.1004158

65. Fändrich, M. *et al.* Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. in *Journal of Internal Medicine* (2018). doi:10.1111/joim.12732

66. Lopez Del Amo, J. M. *et al.* An asymmetric dimer as the basic subunit in Alzheimer's disease amyloid β fibrils. *Angew. Chemie - Int. Ed.* (2012). doi:10.1002/anie.201200965

67. Sachse, C., Fandrich, M. & Grigorieff, N. Paired beta-sheet structure of an Abeta(1-40) amyloid fibril revealed by electron microscopy. *Proc Natl Acad Sci U S A* **105**, 7462–7466 (2008).

68. Annamalai, K. *et al.* Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo. *Angew. Chemie - Int. Ed.* (2017). doi:10.1002/anie.201701761

69. Jeong, J. S., Ansaloni, A., Mezzenga, R., Lashuel, H. A. & Dietler, G. Novel mechanistic insight into the molecular basis of amyloid polymorphism and secondary nucleation during amyloid formation. *J. Mol. Biol.* (2013). doi:10.1016/j.jmb.2013.02.005

70. Tycko, R. & Wickner, R. B. Molecular Structures of Amyloid and Prion Fibrils: Consensus versus Controversy. *Acc Chem Res* (2013). doi:10.1021/ar300282r

71. Tycko, R. Physical and structural basis for polymorphism in amyloid fibrils. *Protein Sci* **23**, 1528–1539 (2014).

72. Loquet, A. *et al.* 3D Structure Determination of Amyloid Fibrils using Solid-State NMR Spectroscopy. *Methods* (2018). doi:10.1016/j.ymeth.2018.03.014

73. Han, S. *et al.* Amyloid plaque structure and cell surface interactions of β-amyloid fibrils revealed by electron tomography. *Sci. Rep.* (2017). doi:10.1038/srep43577

74. Kollmer, M. *et al.* Electron tomography reveals the fibril structure and lipid interactions in amyloid deposits. *Proc. Natl. Acad. Sci. U. S. A.* (2016). doi:10.1073/pnas.1523496113

75. Bergström, J. *et al.* Amyloid deposits in transthyretin-derived amyloidosis: Cleaved transthyretin is associated with distinct amyloid morphology. *J. Pathol.* (2005). doi:10.1002/path.1759

76. Gondo, T. *et al.* Localized amyloidosis in squamous cell carcinoma of uterine cervix: Electron microscopic features of nodular and star-like amyloid deposits. *Virchows Arch. A Pathol. Anat. Histopathol.* (1993). doi:10.1007/BF01621806

77. Wisniewski, H. M., Wegiel, J., Wang, K. C. & Lach, B. Ultrastructural studies of the cells forming amyloid in the cortical vessel wall in Alzheimer's disease. *Acta Neuropathol.* (1992). doi:10.1007/BF00311383

78. Qiang, W., Yau, W. M., Lu, J. X., Collinge, J. & Tycko, R. Structural variation in amyloid-beta fibrils from Alzheimer's disease clinical subtypes. *Nature* **541**, 217–221 (2017).

79. Fändrich, M., Meinhardt, J. & Grigorieff, N. Structural polymorphism of Alzheimer Aβ and other amyloid fibrils. *Prion* (2009). doi:10.4161/pri.3.2.8859

80. Falcon, B. *et al.* Structures of filaments from Pick's disease reveal a novel tau protein fold. *Nature* (2018). doi:10.1038/s41586-018-0454-y

81. Xue, W. F. *et al.* Fibril fragmentation enhances amyloid cytotoxicity. *J. Biol. Chem.* (2009). doi:10.1074/jbc.M109.049809

82. Cohen, S. I. A. *et al.* A molecular chaperone breaks the catalytic cycle that generates toxic Aβ oligomers. *Nat. Struct. Mol. Biol.* (2015). doi:10.1038/nsmb.2971

83. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **81**, 678–699 (2003).

84. Rasool, C. G. & Selkoe, D. J. Alzheimer's disease: exposure of neurofilament immunoreactivity in SDS-insoluble paired helical filaments. *Brain Res* **322**, 194–198 (1984).

85. Haass, C. & Selkoe, D. J. Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide. *Cell* **75**, 1039–1042 (1993).

86. AVILA, J. Role of Tau Protein in Both Physiological and Pathological Conditions. *Physiol. Rev.* (2004). doi:10.1152/physrev.00024.2003

87. Lee, V. M., Goedert, M. & Trojanowski, J. Q. Neurodegenerative tauopathies. *Annu. Rev. Neurosci.* **24**, 1121–1159 (2001).

88. Breydo, L., Wu, J. W. & Uversky, V. N. Alpha-synuclein misfolding and Parkinson's disease. *Biochim Biophys Acta* **1822**, 261–285 (2012).

89. Goedert, M. Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.* **2**, 492–501 (2001).

90. Lippa, C. F. *et al.* DLB and PDD boundary issues: Diagnosis, treatment, molecular pathology, and biomarkers. *Neurology* (2007). doi:10.1212/01.wnl.0000256715.13907.d3

91. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (80-. ).* **314**, 130–133 (2006).

92. Arai, T. *et al.* 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 (2006).

93. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. in *The Lancet* (2011). doi:10.1016/S0140-6736(10)61156-7

94. Rowland, L. P. & Shneider, N. A. Amyotrophic lateral sclerosis. *New England Journal of Medicine* (2001). doi:10.1056/NEJM200105313442207

95. Mitchell, J. & Borasio, G. Amyotrophic lateral sclerosis. *Lancet* (2007). doi:10.1016/S0140-6736(07)60944-1

96. van Es, M. A. *et al.* Amyotrophic lateral sclerosis. *The Lancet* (2017). doi:10.1016/S0140-6736(17)31287-4

97. Mackenzie, I. R. A. *et al.* Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann. Neurol.* (2007). doi:10.1002/ana.21147

98. Dugger, B. N. & Dickson, D. W. Pathology of neurodegenerative diseases. *Cold Spring Harbor Perspectives in Biology* (2017). doi:10.1101/cshperspect.a028035

99. Onyike, C. U. & Diehl-Schmid, J. The epidemiology of frontotemporal dementia. *Int. Rev. Psychiatry*

(2013). doi:10.3109/09540261.2013.776523

100. Ling, S. C., Polymenidou, M. & Cleveland, D. W. Converging mechanisms in als and FTD: Disrupted RNA and protein homeostasis. *Neuron* (2013). doi:10.1016/j.neuron.2013.07.033

101. Tan, R. H., Ke, Y. D., Ittner, L. M. & Halliday, G. M. ALS/FTLD: experimental models and reality. *Acta Neuropathologica* (2017). doi:10.1007/s00401-016-1666-6

102. Ou, S. H., Wu, F., Harrich, D., Garcia-Martinez, L. F. & Gaynor, R. B. 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 (1995).

103. Cohen, T. J., Lee, V. M. Y. & Trojanowski, J. Q. TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends Mol Med* **17**, 659–667 (2011).

104. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol* **20**, 1443–1449 (2013).

105. Kuo, P. H., Chiang, C. H., Wang, Y. T., Doudeva, L. G. & Yuan, H. S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gkt1407

106. Qin, H., Lim, L. Z., Wei, Y. & Song, J. 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 (2014).

107. Mompean, M. *et al.* The TDP-43 N-terminal domain structure at high resolution. *FEBS J* **283**, 1242–1260 (2016).

108. Shiina, Y., Arima, K., Tabunoki, H. & Satoh, J. I. TDP-43 dimerizes in human cells in culture. *Cell. Mol. Neurobiol.* (2010). doi:10.1007/s10571-009-9489-9

109. Zhang, Y. J. *et al.* 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 (2013).

110. Afroz, T. *et al.* Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat Commun* **8**, 45 (2017).

111. Mompean, M. *et al.* Point mutations in the N-terminal domain of transactive response DNA-binding protein 43 kDa (TDP-43) compromise its stability, dimerization, and functions. *J Biol Chem* **292**, 11992–12006 (2017).

112. Jiang, L. L. *et al.* The N-terminal dimerization is required for TDP-43 splicing activity. *Sci Rep* **7**, 6196 (2017).

113. Wang, A. *et al.* A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J* **37**, (2018).

114. Tsoi, P. S. *et al.* The N-Terminal Domain of ALS-Linked TDP-43 Assembles without Misfolding. *Angew Chem Int Ed Engl* **56**, 12590–12593 (2017).

115. Winton, M. J. J. *et al.* 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 (2008).

116. Nishimura, A. L. *et al.* Nuclear import impairment causes cytoplasmic trans-activation response DNA-binding protein accumulation and is associated with frontotemporal lobar degeneration. *Brain* (2010). doi:10.1093/brain/awq111

117. Suzuki, H., Lee, K. & Matsuoka, M. TDP-43-induced death is associated with altered regulation of BIM

and Bcl-xL and attenuated by caspase-mediated TDP-43 cleavage. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.197483

118. Nishimoto, Y. *et al.* Characterization of alternative isoforms and inclusion body of the TAR DNA-binding protein-43. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M109.022012

119. Xiao, S. *et al.* Low molecular weight species of TDP-43 generated by abnormal splicing form inclusions in amyotrophic lateral sclerosis and result in motor neuron death. *Acta Neuropathol.* (2015). doi:10.1007/s00401-015-1412-5

120. Winton, M. J. *et al.* 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 (2008).

121. Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J. Neurosci.* (2010). doi:10.1523/JNEUROSCI.4988-09.2010

122. Kuo, P. H., Doudeva, L. G., Wang, Y. T., Shen, C. K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp013

123. Colombrita, C. *et al.* TDP-43 and FUS RNA-binding proteins bind distinct sets of cytoplasmic messenger RNAs and differently regulate their post-transcriptional fate in motoneuron-like cells. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M111.333450

124. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat. Neurosci.* (2011). doi:10.1038/nn.2778

125. Furukawa, Y. *et al.* A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci. Rep.* (2016). doi:10.1038/srep20576

126. Sephton, C. F. *et al.* Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.190884

127. Shodai, A. *et al.* Conserved Acidic Amino Acid Residues in a Second RNA Recognition Motif Regulate Assembly and Function of TDP-43. *PLoS One* (2012). doi:10.1371/journal.pone.0052776

128. Shodai, A. *et al.* Aberrant assembly of RNA recognition motif 1 links to pathogenic conversion of TAR DNA-binding protein of 43 kDa (TDP-43). *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.451849

129. Prasad, A., Sivalingam, V., Bharathi, V., Girdhar, A. & Patel, B. K. The amyloidogenicity of a C-terminal region of TDP-43 implicated in Amyotrophic Lateral Sclerosis can be affected by anions, acetylation and homodimerization. *Biochimie* (2018). doi:10.1016/j.biochi.2018.05.003

130. Saini, A. & Chauhan, V. S. Delineation of the core aggregation sequences of TDP-43 C-terminal fragment. *ChemBioChem* (2011). doi:10.1002/cbic.201100427

131. Guenther, E. L. L. *et al.* Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat Struct Mol Biol* **25**, 311–319 (2018).

132. Chiang, C. H. *et al.* Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Sci Rep* **6**, 21581 (2016).

133. Buratti, E. TDP-43 post-translational modifications in health and disease. *Expert Opinion on Therapeutic Targets* (2018). doi:10.1080/14728222.2018.1439923

134. Santamaria, N., Alhothali, M., Alfonso, M. H., Breydo, L. & Uversky, V. N. Intrinsic disorder in proteins involved in amyotrophic lateral sclerosis. *Cell. Mol. Life Sci.* (2017). doi:10.1007/s00018-016-2416-6

135. Mackenzie, I. R. A. & Rademakers, R. The role of transactive response DNA-binding protein-43 in

amyotrophic lateral sclerosis and frontotemporal dementia. *Current Opinion in Neurology* (2008). doi:10.1097/WCO.0b013e3283168d1d

136. Ayala, Y. M. *et al.* Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J Mol Biol* **348**, 575–588 (2005).

137. Freibaum, B. D., Chitta, R. K., High, A. A. & Taylor, J. P. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *J. Proteome Res.* (2010). doi:10.1021/pr901076y

138. Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure* **24**, 1537–1549 (2016).

139. Buratti, E. *et al.* 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 (2005).

140. D'Ambrogio, A. *et al.* Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp342

141. Zhang, Y. J. Y.-J. *et al.* Progranulin Mediates Caspase-Dependent Cleavage of TAR DNA Binding Protein-43. *J. Neurosci.* (2007). doi:10.1523/jneurosci.3421-07.2007

142. Zhang, Y.-J. *et al.* Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0900688106

143. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (80-. ).* **314**, 130–133 (2006).

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

145. Guo, W. *et al.* An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat Struct Mol Biol* **18**, 822–830 (2011).

146. Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F. E. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Hum Mol Genet* **24**, 9–20 (2015).

147. Jiang, L. L. *et al.* Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.463828

148. Mompeán, M. *et al.* Structural characterization of the minimal segment of TDP-43 competent for aggregation. *Arch. Biochem. Biophys.* (2014). doi:10.1016/j.abb.2014.01.007

149. Lim, L., Wei, Y., Lu, Y. & Song, J. ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43. *PLoS Biol* **14**, e1002338 (2016).

150. Wickner, R. B. Yeast and Fungal Prions. *Cold Spring Harb Perspect Biol* **8**, (2016).

151. Shewmaker, F., Wickner, R. B. & Tycko, R. Amyloid of the prion domain of Sup35p has an in-register parallel beta-sheet structure. *Proc Natl Acad Sci U S A* **103**, 19754–19759 (2006).

152. Guenther, E. L. *et al.* Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. *Nat Struct Mol Biol* (2018). doi:10.1038/s41594-018-0064-2

153. Mompean, M. *et al.* Structural Evidence of Amyloid Fibril Formation in the Putative Aggregation Domain of TDP-43. *J Phys Chem Lett* **6**, 2608–2615 (2015).

154. Mompean, M., Chakrabartty, A., Buratti, E. & Laurents, D. V. Electrostatic Repulsion Governs TDP-43 C-terminal Domain Aggregation. *PLoS Biol* **14**, e1002447 (2016).

155. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structures of four polymorphic TDP-43 amyloid cores. *Nat. Struct. Mol. Biol.* (2019). doi:10.1038/s41594-019-0248-4

156. Zhuo, X. F. *et al.* Solid-State NMR Reveals the Structural Transformation of the TDP-43 Amyloidogenic Region upon Fibrillation. *J. Am. Chem. Soc.* (2020). doi:10.1021/jacs.9b10736

157. Wright, G. S. A. *et al.* Purification and Structural Characterization of Aggregation-Prone Human TDP-43 Involved in Neurodegenerative Diseases. *iScience* (2020). doi:10.1016/j.isci.2020.101159

158. Coyne, A. N., Zaepfel, B. L. & Zarnescu, D. C. Failure to deliver and translate—new insights into RNA dysregulation in ALS. *Frontiers in Cellular Neuroscience* (2017). doi:10.3389/fncel.2017.00243

159. Buratti, E. & Baralle, F. E. 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.* (2001). doi:10.1074/jbc.M104236200

160. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* (2011). doi:10.1038/nn.2779

161. Polymenidou, M. *et al.* Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Research* (2012). doi:10.1016/j.brainres.2012.02.059

162. Strong, M. J. *et al.* TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Mol Cell Neurosci* **35**, 320–327 (2007).

163. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J.* (2011). doi:10.1038/emboj.2010.310

164. Alami, N. H. *et al.* Axonal Transport of TDP-43 mRNA Granules Is Impaired by ALS-Causing Mutations. *Neuron* (2014). doi:10.1016/j.neuron.2013.12.018

165. Coyne, A. N. *et al.* FUTSCH/MAP1B mRNA is a translational target of TDP-43 and is neuroprotective in a Drosophila model of amyotrophic lateral sclerosis. *J. Neurosci.* **34**, 15962–15974 (2014).

166. Russo, A. *et al.* Increased cytoplasmic TDP-43 reduces global protein synthesis by interacting with Rack1 on polyribosomes. *Hum. Mol. Genet.* **26**, 1407–1418 (2017).

167. Kawahara, Y. & Mieda-Sato, A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proc. Natl. Acad. Sci. U. S. A.* (2012). doi:10.1073/pnas.1112427109

168. Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science (80-. ).* **319**, 1668–1672 (2008).

169. Kametani, F. *et al.* Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains. *Sci. Rep.* (2016). doi:10.1038/srep23281

170. Gendron, T. F., Josephs, K. A. & Petrucelli, L. TDP-43: Mechanisms of neurodegeneration. *Neuropathol Appl Neurobiol* (2010). doi:10.1111/j.1365-2990.2010.01060.x

171. Dammer, E. B. *et al.* Coaggregation of RNA-binding proteins in a model of TDP-43 proteinopathy with selective RGG motif methylation and a role for RRM1 ubiquitination. *PLoS One* (2012). doi:10.1371/journal.pone.0038658

172. Scotter, E. L. *et al.* Differential roles of the ubiquitin proteasome system and autophagy in the clearance of soluble and aggregated TDP-43 species. *J. Cell Sci.* (2014). doi:10.1242/jcs.140087

173. Nonaka, T. *et al.* Phosphorylated and ubiquitinated TDP-43 pathological inclusions in ALS and FTLD-U are

recapitulated in SH-SY5Y cells. *FEBS Lett* **583**, 394–400 (2009).

174. Liachko, N. F., Guthrie, C. R. & Kraemer, B. C. Phosphorylation promotes neurotoxicity in a Caenorhabditis elegans model of TDP-43 proteinopathy. *J. Neurosci.* (2010). doi:10.1523/JNEUROSCI.2911-10.2010

175. Choksi, D. K. *et al.* TDP-43 phosphorylation by casein kinase iε promotes oligomerization and enhances toxicity in vivo. *Hum. Mol. Genet.* (2014). doi:10.1093/hmg/ddt498

176. Neumann, M. *et al.* Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. *Acta Neuropathol.* (2009). doi:10.1007/s00401-008-0477-9

177. Cohen, T. J. *et al.* An acetylation switch controls TDP-43 function and aggregation propensity. *Nat. Commun.* (2015). doi:10.1038/ncomms6845

178. Wang, P., Wander, C. M., Yuan, C. X., Bereman, M. S. & Cohen, T. J. Acetylation-induced TDP-43 pathology is suppressed by an HSF1-dependent chaperone program. *Nat. Commun.* (2017). doi:10.1038/s41467-017-00088-4

179. Bai, P. Biology of Poly(ADP-Ribose) Polymerases: The Factotums of Cell Maintenance. *Molecular Cell* (2015). doi:10.1016/j.molcel.2015.01.034

180. Leung, A. K. L. Poly(ADP-ribose): An organizer of cellular architecture. *Journal of Cell Biology* (2014). doi:10.1083/jcb.201402114

181. McGurk, L. *et al.* Poly(ADP-Ribose) Prevents Pathological Phase Separation of TDP-43 by Promoting Liquid Demixing and Stress Granule Localization. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.07.002

182. Valle, C. & Carrì, M. T. Cysteine modifications in the pathogenesis of ALS. *Frontiers in Molecular Neuroscience* (2017). doi:10.3389/fnmol.2017.00005

183. Hasegawa, M. *et al.* Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol* **64**, 60–70 (2008).

184. Inukai, Y. *et al.* Abnormal phosphorylation of Ser409/410 of TDP-43 in FTLD-U and ALS. *FEBS Lett* **582**, 2899–2904 (2008).

185. Zhang, Y.-J. *et al.* Progranulin Mediates Caspase-Dependent Cleavage of TAR DNA Binding Protein-43. *J. Neurosci.* (2007). doi:10.1523/jneurosci.3421-07.2007

186. Yamashita, T. *et al.* A role for calpain-dependent cleavage of TDP-43 in amyotrophic lateral sclerosis pathology. *Nat. Commun.* (2012). doi:10.1038/ncomms2303

187. Bräuer, S., Zimyanin, V. & Hermann, A. Prion-like properties of disease-relevant proteins in amyotrophic lateral sclerosis. *Journal of Neural Transmission* (2018). doi:10.1007/s00702-018-1851-y

188. Nonaka, T. & Hasegawa, M. TDP-43 prions. *Cold Spring Harb. Perspect. Med.* (2018). doi:10.1101/cshperspect.a024463

189. Sun, Y. & Chakrabartty, A. Phase to Phase with TDP-43. *Biochemistry* (2017). doi:10.1021/acs.biochem.6b01088

190. Nonaka, T. *et al.* Prion-like properties of pathological TDP-43 aggregates from diseased brains. *Cell Rep* **4**, 124–134 (2013).

191. Smethurst, P. *et al.* In vitro prion-like behaviour of TDP-43 in ALS. *Neurobiol. Dis.* (2016). doi:10.1016/j.nbd.2016.08.007

192. Ishii, T., Kawakami, E., Endo, K., Misawa, H. & Watabe, K. Formation and spreading of TDP-43 aggregates in cultured neuronal and glial cells demonstrated by time-lapse imaging. *PLoS One* (2017).

doi:10.1371/journal.pone.0179375

193. Furukawa, Y., Kaneko, K., Watanabe, S., Yamanaka, K. & Nukina, N. A seeding reaction recapitulates intracellular formation of sarkosyl-insoluble transactivation response element (TAR) DNA-binding protein-43 inclusions. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.231209

194. Nonaka, T. *et al.* Prion-like properties of pathological TDP-43 aggregates from diseased brains. *Cell Rep* **4**, 124–134 (2013).

195. Ding, X. *et al.* Exposure to ALS-FTD-CSF generates TDP-43 aggregates in glioblastoma cells through exosomes and TNTs-like structure. *Oncotarget* (2015). doi:10.18632/oncotarget.4680

196. Kasai, T. *et al.* Increased TDP-43 protein in cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *Acta Neuropathol.* (2009). doi:10.1007/s00401-008-0456-1

197. Sproviero, D. *et al.* Pathological proteins are transported by extracellular vesicles of sporadic amyotrophic lateral sclerosis patients. *Front. Neurosci.* (2018). doi:10.3389/fnins.2018.00487

198. Brettschneider, J. *et al.* Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Ann. Neurol.* (2013). doi:10.1002/ana.23937

199. Beekes, M. & McBride, P. A. The spread of prions through the body in naturally acquired transmissible spongiform encephalopathies. *FEBS Journal* (2007). doi:10.1111/j.1742-4658.2007.05631.x

200. Ravits, J., Paul, P. & Jorg, C. Focality of upper and lower motor neuron degeneration at the clinical onset of ALS. *Neurology* (2007). doi:10.1212/01.wnl.0000260965.20021.47

201. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* (2001). doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

202. Vucetic, S., Brown, C. J., Dunker, A. K. & Obradovic, Z. Flavors of protein disorder. *Proteins Struct. Funct. Genet.* (2003). doi:10.1002/prot.10437

203. Brangwynne, C. P., Tompa, P. & Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* (2015). doi:10.1038/nphys3532

204. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1304749110

205. Jiang, H. *et al.* Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. *Cell* (2015). doi:10.1016/j.cell.2015.08.010

206. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* (2015). doi:10.1016/j.cell.2015.07.047

207. Li, H. R. *et al.* The physical forces mediating self-association and phase-separation in the C-terminal domain of TDP-43. *Biochim Biophys Acta* **1866**, 214–223 (2018).

208. Li, H. R., Chiang, W. C., Chou, P. C., Wang, W. J. & Huang, J. rong. 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 (2018).

209. Gopal, P. P., Nirschl, J. J., Klinman, E. & Holzbaurb, E. L. F. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proc. Natl. Acad. Sci. U. S. A.* (2017). doi:10.1073/pnas.1614462114

210. Igaz, L. M. *et al.* Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Am. J. Pathol.* (2008). doi:10.2353/ajpath.2008.080003

211. Aguzzi, A. & Falsig, J. Prion propagation, toxicity and degradation. *Nat Neurosci* **15**, 936–939 (2012).

212. Aguzzi, A., Heikenwalder, M. & Polymenidou, M. Insights into prion strains and neurotoxicity. *Nature Reviews Molecular Cell Biology* (2007). doi:10.1038/nrm2204

213. Sanders, D. W. *et al.* Distinct tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* (2014). doi:10.1016/j.neuron.2014.04.047

214. Peelaerts, W. *et al.* α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature* (2015). doi:10.1038/nature14547

215. Meyer-Luehmann, M. *et al.* Exogenous induction of cerebral β-amyloidogenesis is governed bf agent and host. *Science (80-. ).* (2006). doi:10.1126/science.1131864

216. Mackenzie, I. R. A. *et al.* A harmonized classification system for FTLD-TDP pathology. *Acta Neuropathologica* (2011). doi:10.1007/s00401-011-0845-8

217. Tan, R. H. *et al.* Distinct TDP-43 inclusion morphologies in frontotemporal lobar degeneration with and without amyotrophic lateral sclerosis. *Acta Neuropathol. Commun.* (2017). doi:10.1186/s40478-017-0480-2

218. Vernay, A., Sellal, F. & René, F. Evaluating behavior in mouse models of the behavioral variant of frontotemporal dementia: Which test for which symptom? *Neurodegenerative Diseases* (2016). doi:10.1159/000439253

219. Laferrière, F. *et al.* TDP-43 extracted from frontotemporal lobar degeneration subject brains displays distinct aggregate assemblies and neurotoxic effects reflecting disease progression rates. *Nat. Neurosci.* (2019). doi:10.1038/s41593-018-0294-y

220. Guerrero, E. N. *et al.* TDP-43/FUS in motor neuron disease: Complexity and challenges. *Progress in Neurobiology* (2016). doi:10.1016/j.pneurobio.2016.09.004

221. D'Alton, S., Altshuler, M. & Lewis, J. Studies of alternative isoforms provide insight into TDP-43 autoregulation and pathogenesis. *RNA* (2015). doi:10.1261/rna.047647.114

222. Weskamp, K. *et al.* Shortened TDP43 isoforms upregulated by neuronal hyperactivity drive TDP43 pathology in ALS. *J. Clin. Invest.* (2020). doi:10.1172/JCI130988

223. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333–366 (2006).

224. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203 (2012).

225. Lansbury, P. T. Mechanism of scrapie replication [3]. *Science* (1994). doi:10.1126/science.8079159

226. Fowler, D. M., Koulov, A. V, Balch, W. E. & Kelly, J. W. Functional amyloid--from bacteria to humans. *Trends Biochem Sci* **32**, 217–224 (2007).

227. Otzen, D. & Riek, R. Functional amyloids. *Cold Spring Harb. Perspect. Biol.* (2019). doi:10.1101/cshperspect.a033860

228. Fowler, D. M. *et al.* Functional amyloid formation within mammalian tissue. *PLoS Biol* **4**, e6 (2006).

229. Chapman, M. R. *et al.* Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science (80-. ).* **295**, 851–855 (2002).

230. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science (80-. ).* **325**, 328–332 (2009).

231. Coustou, V., Deleu, C., Saupe, S. & Begueret, J. The protein product of the het-s heterokaryon

incompatibility gene of the fungus Podospora anserina behaves as a prion analog. *Proc Natl Acad Sci U S A* **94**, 9773–9778 (1997).

232. Maddelein, M. L., Dos Reis, S., Duvezin-Caubet, S., Coulary-Salin, B. & Saupe, S. J. Amyloid aggregates of the HET-s prion protein are infectious. *Proc Natl Acad Sci U S A* **99**, 7402–7407 (2002).

233. Saad, S. *et al.* Reversible protein aggregation is a protective mechanism to ensure cell cycle restart after stress. *Nat. Cell Biol.* (2017). doi:10.1038/ncb3600

234. Li, J. *et al.* The RIP1/RIP3 necrosome forms a functional amyloid signaling complex required for programmed necrosis. *Cell* **150**, 339–350 (2012).

235. Zhou, Y., Blanco, L. P., Smith, D. R. & Chapman, M. R. Bacterial amyloids. *Methods Mol Biol* **849**, 303–320 (2012).

236. Meier, B. H. *et al.* Atomic-Resolution Three-Dimensional Structure of HET-s(218-289) Amyloid Fibrils by Solid-State NMR Spectroscopy. *J. Am. Chem. Soc.* **132**, 13765–13775 (2010).

237. Saupe, S. J. Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiol Mol Biol Rev* **64**, 489–502 (2000).

238. Glass, N. L. & Dementhon, K. Non-self recognition and programmed cell death in filamentous fungi. *Current Opinion in Microbiology* (2006). doi:10.1016/j.mib.2006.09.001

239. Turcq, B., Deleu, C., Denayrolles, M. & Bégueret, J. Two allelic genes responsible for vegetative incompatibility in the fungus Podospora anserina are not essential for cell viability. *MGG Mol. Gen. Genet.* (1991). doi:10.1007/BF00282475

240. Greenwald, J. *et al.* The mechanism of prion inhibition by HET-S. *Mol Cell* **38**, 889–899 (2010).

241. Balguerie, A. *et al.* Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina. *EMBO J* **22**, 2071–2081 (2003).

242. Seuring, C. *et al.* The mechanism of toxicity in HET-S/HET-s prion incompatibility. *PLoS Biol* **10**, e1001451 (2012).

243. Seuring, C. *et al.* The mechanism of toxicity in HET-S/HET-s prion incompatibility. *PLoS Biol* **10**, e1001451 (2012).

244. Leipe, D. D., Koonin, E. V. & Aravind, L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* (2004). doi:10.1016/j.jmb.2004.08.023

245. Danot, O., Marquenet, E., Vidal-Ingigliardi, D. & Richet, E. Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure* (2009). doi:10.1016/j.str.2009.01.001

246. Saleh, M. The machinery of Nod-like receptors: Refining the paths to immunity and cell death. *Immunol. Rev.* (2011). doi:10.1111/j.1600-065X.2011.01045.x

247. DeYoung, B. J. & Innes, R. W. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology* (2006). doi:10.1038/ni1410

248. Bonardi, V., Cherkis, K., Nishimura, M. T. & Dangl, J. L. A new eye on NLR proteins: Focused on clarity or diffused by complexity? *Current Opinion in Immunology* (2012). doi:10.1016/j.coi.2011.12.006

249. Daskalov, A., Paoletti, M., Ness, F. & Saupe, S. J. Genomic clustering and homology between HET-S and the NWD2 STAND protein in various fungal genomes. *PLoS One* **7**, e34854 (2012).

250. Daskalov, A. *et al.* Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLoS Biol* **13**, e1002059 (2015).

251. Cai, X. *et al.* Prion-like polymerization underlies signal transduction in antiviral immune defense and inflammasome activation. *Cell* (2014). doi:10.1016/j.cell.2014.01.063

252. Daskalov, A. On the evolutionary trajectories of signal-transducing amyloids in fungi and beyond. *Prion* **10**, 362–368 (2016).

253. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4976-84 (2016).

254. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic Abeta42 Amyloid Fibrils. *J Am Chem Soc* **138**, 9663–9674 (2016).

255. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627 e16 (2017).

256. Andrew, E. R., Bradbury, A. & Eades, R. G. Nuclear Magnetic Resonance Spectra from a Crystal Rotated at High Speed. *Nature* **4650**, 1659 (1958).

257. Maricq, M. M. & Waugh, J. S. NMR in rotating solids. *J. Chem. Phys.* (1979). doi:10.1063/1.437915

258. Habenstein, B. & Loquet, A. Solid-state NMR: An emerging technique in structural biology of self-assemblies. *Biophys. Chem.* (2016). doi:10.1016/j.bpc.2015.07.003

259. Mandala, V. S., Williams, J. K. & Hong, M. Structure and Dynamics of Membrane Proteins from Solid-State NMR. *Annu. Rev. Biophys.* **47**, 201–222 (2018).

260. Meier, B. H. & Bockmann, A. The structure of fibrils from “misfolded” proteins. *Curr Opin Struct Biol* **30**, 43–49 (2015).

261. Hong, M., Zhang, Y. & Hu, F. Membrane protein structure and dynamics from NMR spectroscopy. *Annual Review of Physical Chemistry* (2012). doi:10.1146/annurev-physchem-032511-143731

262. Korukottu, J. *et al.* High-resolution 3D structure determination of kaliotoxin by solid-state NMR spectroscopy. *PLoS One* (2008). doi:10.1371/journal.pone.0002359

263. Loquet, A., Habenstein, B. & Lange, A. Structural investigations of molecular machines by solid-state NMR. *Acc. Chem. Res.* (2013). doi:10.1021/ar300320p

264. Miao, Y. & Cross, T. A. Solid state NMR and protein-protein interactions in membranes. *Current Opinion in Structural Biology* (2013). doi:10.1016/j.sbi.2013.08.004

265. Tycko, R. & Wickner, R. B. Molecular structures of amyloid and prion fibrils: Consensus versus controversy. *Acc. Chem. Res.* (2013). doi:10.1021/ar300282r

266. Weingarth, M. & Baldus, M. Solid-state NMR-based approaches for supramolecular structure elucidation. *Acc. Chem. Res.* (2013). doi:10.1021/ar300316e

267. Xu, J. *et al.* Bicelle-enabled structural studies on a membrane-associated cytochrome b5 by solid-state MAS NMR spectroscopy. *Angew. Chemie - Int. Ed.* (2008). doi:10.1002/anie.200801338

268. Morag, O., Sgourakis, N. G., Baker, D. & Goldbourt, A. The NMR-Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc Natl Acad Sci U S A* **112**, 971–976 (2015).

269. Yan, S., Suiter, C. L., Hou, G., Zhang, H. & Polenova, T. Probing structure and dynamics of protein assemblies by magic angle spinning NMR spectroscopy. *Acc. Chem. Res.* (2013). doi:10.1021/ar300309s

270. Loquet, A. *et al.* Corrigendum: Atomic model of the type III secretion system needle. *Nature* **486**, 276–279 (2012).

271. Demers, J. P. *et al.* High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nat Commun* **5**, 4976 (2014).

272. Habenstein, B. *et al.* Hybrid Structure of the Type 1 Pilus of Uropathogenic Escherichia coli. *Angew Chem Int Ed Engl* **54**, 11691–11695 (2015).

273. Cady, S. D. *et al.* Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* **463**, 689–692 (2010).

274. Park, S. H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* (2012). doi:nature11580 [pii] 10.1038/nature11580

275. Wang, S. *et al.* Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. *Nat Methods* (2013). doi:nmeth.2635 [pii] 10.1038/nmeth.2635

276. Walti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Abeta(1-42) amyloid fibril. *Proc Natl Acad Sci U S A* **113**, E4976-84 (2016).

277. Tuttle, M. D. *et al.* Solid-state NMR structure of a pathogenic fibril of full-length human alpha-synuclein. *Nat Struct Mol Biol* **23**, 409–415 (2016).

278. Lu, J. X. *et al.* XMolecular structure of β-amyloid fibrils in alzheimer's disease brain tissue. *Cell* (2013). doi:10.1016/j.cell.2013.08.035

279. Petkova, A. T. *et al.* A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16742–16747 (2002).

280. Paravastu, A. K., Leapman, R. D., Yau, W. M. & Tycko, R. Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc Natl Acad Sci U S A* **105**, 18349–18354 (2008).

281. Seuring, C. *et al.* The three-dimensional structure of human β-endorphin amyloid fibrils. *Nat. Struct. Mol. Biol.* 1–7 (2020). doi:10.1038/s41594-020-00515-z

282. Schanda, P. & Ernst, M. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2016). doi:10.1016/j.pnmrs.2016.02.001

283. Hoffmann, B., Löhr, F., Laguerre, A., Bernhard, F. & Dötsch, V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2018). doi:10.1016/j.pnmrs.2017.11.004

284. Guan, X., Chaffey, P. K., Zeng, C. & Tan, Z. New methods for chemical protein synthesis. *Top. Curr. Chem.* **363**, 155–192 (2015).

285. Raibaut, L., El Mahdi, O. & Melnyk, O. Solid phase protein chemical synthesis. *Top. Curr. Chem.* **363**, 103–154 (2015).

286. Castellani, F., van Rossum, B. J., Diehl, A., Rehbein, K. & Oschkinat, H. Determination of solid-state NMR structures of proteins by means of three-dimensional 15N-13C-13C dipolar correlation spectroscopy and chemical shift analysis. *Biochemistry* **42**, 11476–11483 (2003).

287. Loquet, A., Giller, K., Becker, S. & Lange, A. Supramolecular interactions probed by 13C-13C solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132**, 15164–15166 (2010).

288. Loquet, A., Lv, G., Giller, K., Becker, S. & Lange, A. 13C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. *J. Am. Chem. Soc.* (2011). doi:10.1021/ja200066s

289. Habenstein, B., Loquet, A., Giller, K., Becker, S. & Lange, A. Structural characterization of supramolecular assemblies by 13C spin dilution and 3D solid-state NMR. *J. Biomol. NMR* (2013). doi:10.1007/s10858-

012-9691-9

290. Paepe, G. D. *et al.* Heteronuclear proton assisted recoupling. *J. Chem. Phys.* **134**, 95101 (2011).

291. De Paëpe, G., Lewandowski, J. R., Loquet, A., Böckmann, A. & Griffin, R. G. Proton assisted recoupling and protein structure determination. *J. Chem. Phys.* (2008). doi:10.1063/1.3036928

292. Tycko, R. Solid-state NMR studies of amyloid fibril structure. *Annu Rev Phys Chem* **62**, 279–299 (2011).

293. Comellas, G. & Rienstra, C. M. Protein structure determination by magic-angle spinning solid-state NMR, and insights into the formation, structure, and stability of amyloid fibrils. *Annu Rev Biophys* **42**, 515–536 (2013).

294. Meier, B. H., Riek, R. & Bockmann, A. Emerging Structural Understanding of Amyloid Fibrils by Solid-State NMR. *Trends Biochem Sci* **42**, 777–787 (2017).

295. Cohen, M. L. *et al.* Rapidly progressive Alzheimer's disease features distinct structures of amyloid-β. *Brain* (2015). doi:10.1093/brain/awv006

296. Collinge, J. & Clarke, A. R. A general model of prion strains and their pathogenicity. *Science* (2007). doi:10.1126/science.1138718

297. Safar, J. *et al.* Eight prion strains have PrP(Sc) molecules with different conformations. *Nat. Med.* (1998). doi:10.1038/2654

298. Petkova, A. T. *et al.* Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science (80-. ).* (2005). doi:10.1126/science.1105850

299. Gath, J. *et al.* Unlike twins: An NMR comparison of two α-synuclein polymorphs featuring different toxicity. *PLoS One* (2014). doi:10.1371/journal.pone.0090659

300. Bousset, L. *et al.* Structural and functional characterization of two alpha-synuclein strains. *Nat. Commun.* (2013). doi:10.1038/ncomms3575

301. Lu, J. X. *et al.* Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. *Cell* **154**, 1257–1268 (2013).

302. Xiao, Y. *et al.* Abeta(1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat Struct Mol Biol* **22**, 499–505 (2015).

303. Sgourakis, N. G., Yau, W. M. & Qiang, W. Modeling an in-register, parallel "iowa" Aβ fibril structure using solid-state NMR data from labeled samples with Rosetta. *Structure* (2015). doi:10.1016/j.str.2014.10.022

304. Qiang, W., Yau, W.-M., Luo, Y., Mattson, M. P. & Tycko, R. Antiparallel -sheet architecture in Iowa-mutant -amyloid fibrils. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1111305109

305. Ferguson, N. *et al.* General structural motifs of amyloid protofilaments. *Proc. Natl. Acad. Sci. USA* **103**, 16248–16253 (2006).

306. *Solid-State NMR Spectroscopy Principles and Applications*. *Solid-State NMR Spectroscopy Principles and Applications* (2007). doi:10.1002/9780470999394

307. Bennett, A. E., Rienstra, C. M., Auger, M., Lakshmi, K. V & Griffin, R. G. Heteronuclear decoupling in rotating solids. *J. Chem. Phys.* **103**, 6951–6958 (1995).

308. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. *J. Magn. Reson.* **142**, 97–101 (2000).

309. Detken, A., Hardy, E. H., Ernst, M. & Meier, B. H. Simple and efficient decoupling in magic-angle spinning

solid-state NMR: The XiX scheme. *Chem. Phys. Lett.* (2002). doi:10.1016/S0009-2614(02)00335-4

310. Stejskal, E. O., Schaefer, J. & Waugh, J. S. Magic-angle spinning and polarization transfer in proton-enhanced NMR. *J. Magn. Reson.* (1977). doi:10.1016/0022-2364(77)90260-8

311. Hediger, S., Meier, B. H. & Ernst, R. R. Adiabatic passage Hartmann-Hahn cross polarization in NMR under magic angle sample spinning. *Chem. Phys. Lett.* **240**, 449–456 (1995).

312. Verel, R., Ernst, M. & Meier, B. H. Adiabatic dipolar recoupling in solid-state NMR: The DREAM scheme. *J. Magn. Reson.* (2001). doi:10.1006/jmre.2001.2310

313. Kubo, A. & McDowell, C. A. Spectral spin diffusion in polycrystalline solids under magic-angle spinning. *J. Chem. Soc. Faraday Trans. 1 Phys. Chem. Condens. Phases* (1988). doi:10.1039/F19888403713

314. Takegoshi, K., Nakamura, S. & Terao, T. 13C-1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* (2001). doi:10.1016/S0009-2614(01)00791-6

315. Elena, B., Lesage, A., Steuernagel, S., Böckmann, A. & Emsley, L. Proton to carbon-13 INEPT in solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 17296–17302 (2005).

316. Morris, G. a. & Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *J. Am. Chem. Soc.* **233**, 760–762 (1979).

317. Daskalov, A. *et al.* Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLoS Biol* **13**, e1002059 (2015).

318. Loquet, A., Tolchard, J., Berbon, M., Martinez, D. & Habenstein, B. Atomic scale structural studies of macromolecular assemblies by solid-state nuclear magnetic resonance spectroscopy. *J. Vis. Exp.* (2017). doi:10.3791/55779

319. Harris, R. K. *et al.* Further Conventions for NMR shielding and chemical shifts (IUPAC recommendations 2008). *Magn. Reson. Chem.* (2008). doi:10.1002/mrc.2225

320. Bockmann, A. *et al.* Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR* **45**, 319–327 (2009).

321. J. Cavanagh, W. Fairbrother, A. Palmer III, N. Skelton, M. R. *Protein NMR Spectroscopy: Principles and Practice, 2nd*. *Protein NMR Spectroscopy: Principles and Practice,* (2007).

322. Hardy, E. H., Detken, A. & Meier, B. H. Fast-MAS total through-bond correlation spectroscopy using adiabatic pulses. *J. Magn. Reson.* (2003). doi:10.1016/j.jmr.2003.08.003

323. Baldus, M., Iuliucci, R. J. & Meier, B. H. Probing through-bond connectivities and trough-space distances in solids by magic-angle-spinning nuclear magnetic resonance. *J. Am. Chem. Soc.* **119**, 1121–1124 (1997).

324. Biological Magnetic Resonance Data Bank. in *Encyclopedia of Biophysics* (2013). doi:10.1007/978-3-642-16712-6_100106

325. Wang, Y. & Jardetzky, O. Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci.* **11**, 852–861 (2002).

326. Skinner, S. P. *et al.* CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* (2016). doi:10.1007/s10858-016-0060-y

327. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btu830

328. Hong, M. Determination of Multiple ϕ-Torsion Angles in Proteins by Selective and Extensive 13C Labeling and Two-Dimensional Solid-State NMR. *J. Magn. Reson.* (1999). doi:10.1006/jmre.1999.1805

329. Lundstrom, P. *et al.* Fractional 13C enrichment of isolated carbons using [1-13C]- or [2- 13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and side-chain methyl positions in proteins. *J. Biomol. NMR* **38**, 199–212 (2007).

330. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98–102 (2002).

331. Higman, V. A. *et al.* Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins. *J. Biomol. NMR* **44**, 245–260 (2009).

332. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. Nmr* **44**, 213–223 (2009).

333. Bardiaux, B., Malliavin, T. & Nilges, M. ARIA for solution and solid-state NMR. *Methods Mol Biol* **831**, 453–483 (2012).

334. Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).

335. Guerry, P. & Herrmann, T. Comprehensive automation for NMR structure determination of proteins. *Methods Mol Biol* **831**, 429–451 (2012).

336. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003).

337. Schütz, A. K. *et al.* Atomic-resolution three-dimensional structure of amyloid b fibrils bearing the osaka mutation. *Angew. Chemie - Int. Ed.* (2015). doi:10.1002/anie.201408598

338. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19**, 315–316 (2003).

339. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat Protoc* **2**, 2728–2733 (2007).

340. Nilges, M. A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins* **17**, 297–309 (1993).

341. Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.* (1996). doi:10.1002/pro.5560050609

342. Montelione, G. T. *et al.* Recommendations of the wwPDB NMR validation task force. *Structure* (2013). doi:10.1016/j.str.2013.07.021

343. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

344. Davis, I. W. *et al.* MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm216

345. Linser, R. Solid-state NMR spectroscopic trends for supramolecular assemblies and protein aggregates. *Solid State Nucl Magn Reson* **87**, 45–53 (2017).

346. van der Wel, P. C. A. Insights into protein misfolding and aggregation enabled by solid-state NMR spectroscopy. *Solid State Nucl Magn Reson* **88**, 1–14 (2017).

347. Rossini, A. J. *et al.* Dynamic nuclear polarization enhanced solid-state NMR spectroscopy of functionalized metal-organic frameworks. *Angew Chem Int Ed Engl* **51**, 123–127 (2012).

348. Debelouchina, G. T. *et al.* Dynamic nuclear polarization-enhanced solid-state NMR spectroscopy of GNNQQNY nanocrystals and amyloid fibrils. *Phys Chem Chem Phys* **12**, 5911–5919 (2010).

349. van der Wel, P. C. A., Hu, K. N., Lewandowski, J. & Griffin, R. G. Dynamic nuclear polarization of amyloidogenic peptide nanocrystals: GNNQQNY, a core segment of the yeast prion protein Sup35p. *J Am Chem Soc* **128**, 10840–10846 (2006).

350. Stanek, J. *et al.* NMR Spectroscopic Assignment of Backbone and Side-Chain Protons in Fully Protonated Proteins: Microcrystals, Sedimented Assemblies, and Amyloid Fibrils. *Angew Chem Int Ed Engl* **55**, 15504–15509 (2016).

351. Agarwal, V. *et al.* De novo 3D structure determination from sub-milligram protein samples by solid-state 100 kHz MAS NMR spectroscopy. *Angew Chem Int Ed Engl* **53**, 12253–12256 (2014).

352. Andreas, L. B. *et al.* Structure of fully protonated proteins by proton-detected magic-angle spinning NMR. *Proc Natl Acad Sci U S A* **113**, 9187–9192 (2016).

353. Cala-De Paepe, D. *et al.* Is protein deuteration beneficial for proton detected solid-state NMR at and above 100 kHz magic-angle spinning? *Solid State Nucl. Magn. Reson.* (2017). doi:10.1016/j.ssnmr.2017.07.004

354. Struppe, J. *et al.* Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz. *Solid State Nucl. Magn. Reson.* (2017). doi:10.1016/j.ssnmr.2017.07.001

355. Böckmann, A., Ernst, M. & Meier, B. H. Spinning proteins, the faster, the better? *J. Magn. Reson.* (2015). doi:10.1016/j.jmr.2015.01.012

356. Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu Rev Biochem* **86**, 69–95 (2017).

357. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* **8**, 595–608 (2016).

358. Stubbs, G. & Stohr, J. Structural Biology of PrP Prions. *Cold Spring Harb Perspect Med* **7**, (2017).

359. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* (2017). doi:10.1038/nature23002

360. Gremer, L. *et al.* Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy. *Science (80-. ).* **358**, 116–119 (2017).

361. Neary, D., Snowden, J. S. & Mann, D. M. Cognitive change in motor neurone disease/amyotrophic lateral sclerosis (MND/ALS). *J Neurol Sci* **180**, 15–20 (2000).

362. Da Cruz, S. & Cleveland, D. W. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Curr Opin Neurobiol* **21**, 904–919 (2011).

363. Negash, S. *et al.* Cognitive and functional resilience despite molecular evidence of Alzheimer's disease pathology. *Alzheimers Dement* **9**, e89-95 (2013).

364. Ratti, A. & Buratti, E. Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *J Neurochem* **138 Suppl**, 95–111 (2016).

365. Anderson, P. & Kedersha, N. Stress granules. *Curr Biol* **19**, R397-8 (2009).

366. Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).

367. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).

368. Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and

Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell* **69**, 465-479 e7 (2018).

369. Hyman, A. A., Weber, C. A. & Julicher, F. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**, 39–58 (2014).

370. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* (2018). doi:10.1016/j.tcb.2018.02.004

371. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* **36**, 932–941 (2009).

372. Kwiatkowski T. J., J. *et al.* Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science (80-. ).* **323**, 1205–1208 (2009).

373. Kim, H. J. *et al.* Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473 (2013).

374. Johnson, B. S. *et al.* TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *J Biol Chem* **284**, 20329–20339 (2009).

375. Fang, Y. S. *et al.* Full-length TDP-43 forms toxic amyloid oligomers that are present in frontotemporal lobar dementia-TDP patients. *Nat Commun* **5**, 4824 (2014).

376. Chang, C. K. *et al.* The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem Biophys Res Commun* **425**, 219–224 (2012).

377. Sasaguri, H. *et al.* The extreme N-terminus of TDP-43 mediates the cytoplasmic aggregation of TDP-43 and associated toxicity in vivo. *Brain Res* **1647**, 57–64 (2016).

378. Pillai, M. & 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). doi:10.1021/acs.biochem.8b01013

379. Tavella, D., Zitzewitz, J. A. & Massi, F. Characterization of TDP-43 RRM2 Partially Folded States and Their Significance to ALS Pathogenesis. *Biophys. J.* (2018). doi:10.1016/j.bpj.2018.09.011

380. Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R. & Zitzewitz, J. A. 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 (2014).

381. Lagier-Tourenne, C. & Cleveland, D. W. Rethinking ALS: the FUS about TDP-43. *Cell* **136**, 1001–1004 (2009).

382. Kwon, M. J. *et al.* Screening of the SOD1, FUS, TARDBP, ANG, and OPTN mutations in Korean patients with familial and sporadic ALS. *Neurobiol Aging* **33**, 1017 e17–23 (2012).

383. Buratti, E. Functional Significance of TDP-43 Mutations in Disease. *Adv Genet* **91**, 1–53 (2015).

384. Jiang, L. L. *et al.* Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Sci Rep* **6**, 23928 (2016).

385. Luckgei, N. *et al.* The conformation of the prion domain of Sup35p in isolation and in the full-length protein. *Angew Chem Int Ed Engl* **52**, 12741–12744 (2013).

386. Li, Q., Yokoshi, M., Okada, H. & Kawahara, Y. The cleavage pattern of TDP-43 determines its rate of clearance and cytotoxicity. *Nat. Commun.* (2015). doi:10.1038/ncomms7183

387. Seyfried, N. T. *et al.* Multiplex SILAC analysis of a cellular TDP-43 proteinopathy model reveals protein inclusions associated with SUMOylation and diverse polyubiquitin chains. *Mol. Cell. Proteomics* (2010). doi:10.1074/mcp.M800390-MCP200

388. Tang, M., Comellas, G. & Rienstra, C. M. Advanced Solid-State NMR Approaches for Structure Determination of Membrane Proteins and Amyloid Fibrils. *Acc Chem Res* (2013). doi:10.1021/ar4000168

389. Eisenberg, D. *et al.* The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* **39**, 568–575 (2006).

390. Riek, R. The Three-Dimensional Structures of Amyloids. *Cold Spring Harb Perspect Biol* **9**, (2017).

391. Tycko, R. Characterization of amyloid structures at the molecular level by solid state nuclear magnetic resonance spectroscopy. *Amyloid, Prions, Other Protein Aggregates, Pt C* **413**, 103–122 (2006).

392. Matlahov, I. & van der Wel, P. C. A. Hidden motions and motion-induced invisibility: Dynamics-based spectral editing in solid-state NMR. *Methods* (2018). doi:10.1016/j.ymeth.2018.04.015

393. Siemer, A. B., Ritter, C., Ernst, M., Riek, R. & Meier, B. H. High-resolution solid-state NMR spectroscopy of the prion protein HET-s in its amyloid conformation. *Angew. Chem. Int. Ed.* **44**, 2441–2444 (2005).

394. Heise, H. *et al.* 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 (2005).

395. Goormaghtigh, E., Cabiaux, V. & Ruysschaert, J. M. Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. I. Assignments and model compounds. *Subcell Biochem* **23**, 329–362 (1994).

396. Luca, S. *et al.* Secondary chemical shifts in immobilized peptides and proteins: a qualitative basis for structure refinement under magic angle spinning. *J. Biomol. NMR* **20**, 325–331 (2001).

397. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1804**, 1405–1412 (2010).

398. Groenning, M. *et al.* Binding mode of Thioflavin T in insulin amyloid fibrils. *J. Struct. Biol.* (2007). doi:10.1016/j.jsb.2007.06.004

399. Chang, Z., Deng, J., Zhao, W. & Yang, J. Amyloid-like aggregation and fibril core determination of TDP-43 C-terminal domain. *Biochem. Biophys. Res. Commun.* (2020). doi:10.1016/j.bbrc.2020.08.096

400. Barbet-Massin, E. *et al.* Rapid proton-detected NMR assignment for proteins with fast magic angle spinning. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja507382j

401. Andreas, L. B., Le Marchand, T., Jaudzems, K. & Pintacuda, G. High-resolution proton-detected NMR of proteins at very fast MAS. *J Magn Reson* **253**, 36–49 (2015).

402. Li, B. *et al.* Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* (2018). doi:10.1038/s41467-018-05971-2

403. Zhang, W. *et al.* Heparin-induced tau filaments are polymorphic and differ from those in alzheimer's and pick's diseases. *Elife* (2019). doi:10.7554/eLife.43584

404. Fandrich, M., Meinhardt, J. & Grigorieff, N. Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion* **3**, 89–93 (2009).

405. Fitzpatrick, A. W. P. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* (2017). doi:10.1038/nature23002

406. Hughes, M. P. *et al.* Atomic structures of low-complexity protein segments reveal kinked b sheets that assemble networks. *Science (80-. ).* (2018). doi:10.1126/science.aan6398

407. Westermark, G. T., Johnson, K. H. & Westermark, P. Staining methods for identification of amyloid in tissue. *Methods Enzymol.* (1999). doi:10.1016/S0076-6879(99)09003-5

408. Paravastu, A. K., Petkova, A. T. & Tycko, R. Polymorphic fibril formation by residues 10-40 of the Alzheimer's beta-amyloid peptide. *Biophys J* **90**, 4618–4629 (2006).

409. Chiti, F. *et al.* Development of enzymatic activity during protein folding. Detection of a spectroscopically silent native-like intermediate of muscle acylphosphatase. *J Biol Chem* **274**, 20151–20158 (1999).

410. Agrawal, S. *et al.* RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. *Sci. Rep.* **9**, (2019).

411. Franzmann, T. M. & Alberti, S. Prion-like low-complexity sequences: Key regulators of protein solubility and phase behavior. *Journal of Biological Chemistry* (2019). doi:10.1074/jbc.TM118.001190

412. Zacco, E., Martin, S. R., Thorogate, R. & Pastore, A. 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.* (2018). doi:10.3389/fnmol.2018.00372

413. Chang, C. ke *et al.* The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem Biophys Res Commun* **425**, 219–224 (2012).

414. Nishihira, Y. *et al.* Sporadic amyotrophic lateral sclerosis: two pathological patterns shown by analysis of distribution of TDP-43-immunoreactive neuronal and glial cytoplasmic inclusions. *Acta Neuropathol* **116**, 169–182 (2008).

415. Mori, F. *et al.* Maturation process of TDP-43-positive neuronal cytoplasmic inclusions in amyotrophic lateral sclerosis with and without dementia. *Acta Neuropathol* **116**, 193–203 (2008).

416. Meier, B. H. *et al.* Prion fibrils of Ure2p assembled under physiological conditions contain highly ordered, natively folded modules. *J Mol Biol* **394**, 108–118 (2009).

417. Habenstein, B. *et al.* A Native-Like Conformation for the C-Terminal Domain of the Prion Ure2p within its Fibrillar Form. *Angew Chem Int Ed Engl* **51**, 7963–7966 (2012).

418. Bockmann, A. *et al.* The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form. *J. Mol. Biol.* **394**, 119–127 (2009).

419. Morgan, B. R., Zitzewitz, J. A. & Massi, F. Structural Rearrangement upon Fragmentation of the Stability Core of the ALS-Linked Protein TDP-43. *Biophys. J.* (2017). doi:10.1016/j.bpj.2017.06.049

420. Wang, W. *et al.* The inhibition of TDP-43 mitochondrial localization blocks its neuronal toxicity. *Nat Med* **22**, 869–878 (2016).

421. Rutherford, N. J. *et al.* Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS Genet* **4**, e1000193 (2008).

422. Nelson, O. *et al.* Familial Alzheimer disease-linked mutations specifically disrupt Ca2+ leak function of presenilin 1. *J Clin Invest* **117**, 1230–1239 (2007).

423. Li, H. R., Chiang, W. C., Chou, P. C., Wang, W. J. & Huang, J. R. 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 (2018).

424. Pantoja-Uceda, D. *et al.* Phe-Gly motifs drive fibrillization of TDP-43's prion-like domain droplets. *bioRxiv* (2020). doi:10.1101/2020.02.20.957969

425. Brown, J. W. P. *et al.* β-Synuclein suppresses both the initiation and amplification steps of α-synuclein aggregation via competitive binding to surfaces. *Sci. Rep.* (2016). doi:10.1038/srep36010

426. Kroschwald, S. *et al.* Different Material States of Pub1 Condensates Define Distinct Modes of Stress Adaptation and Recovery. *Cell Rep.* (2018). doi:10.1016/j.celrep.2018.05.041

427. Franzmann, T. M. *et al.* Phase separation of a yeast prion protein promotes cellular fitness. *Science (80-. ).* (2018). doi:10.1126/science.aao5654

428. Riback, J. A. *et al.* Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* (2017). doi:10.1016/j.cell.2017.02.027

429. Winklhofer, K. F., Tatzelt, J. & Haass, C. The two faces of protein misfolding: Gain- and loss-of-function in neurodegenerative diseases. *EMBO Journal* (2008). doi:10.1038/sj.emboj.7601930

430. Vogler, T. O. *et al.* TDP-43 and RNA form amyloid-like myo-granules in regenerating muscle. *Nature* (2018). doi:10.1038/s41586-018-0665-2

431. Moreno-Gonzalez, I. & Soto, C. Misfolded protein aggregates: Mechanisms, structures and potential for disease transmission. *Seminars in Cell and Developmental Biology* (2011). doi:10.1016/j.semcdb.2011.04.002

432. Eichner, T. & Radford, S. E. A diversity of assembly mechanisms of a generic amyloid fold. *Mol Cell* **43**, 8–18 (2011).

433. Saupe, S. J. The [Het-s] prion of Podospora anserina and its role in heterokaryon incompatibility. *Seminars in Cell and Developmental Biology* (2011). doi:10.1016/j.semcdb.2011.02.019

434. Ritter, C. *et al.* Correlation of structural elements and infectivity of the HET-s prion. *Nature* **435**, 844–848 (2005).

435. Mathur, V., Seuring, C., Riek, R., Saupe, S. J. & Liebman, S. W. Localization of HET-S to the cell periphery, not to [Het-s] aggregates, is associated with [Het-s]-HET-S toxicity. *Mol Cell Biol* **32**, 139–153 (2012).

436. Daskalov, A. *et al.* Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proc. Natl. Acad. Sci. U. S. A.* (2016). doi:10.1073/pnas.1522361113

437. Murphy, J. M. *et al.* The pseudokinase MLKL mediates necroptosis via a molecular switch mechanism. *Immunity* (2013). doi:10.1016/j.immuni.2013.06.018

438. Adachi, H. *et al.* An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *Elife* (2019). doi:10.7554/eLife.49956

439. Wu, H. Higher-order assemblies in a new paradigm of signal transduction. *Cell* (2013). doi:10.1016/j.cell.2013.03.013

440. Sun, L. *et al.* Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase. *Cell* (2012). doi:10.1016/j.cell.2011.11.031

441. Daskalov, A., Dyrka, W. & Saupe, S. J. Theme and variations: evolutionary diversification of the HET-s functional amyloid motif. *Sci Rep* **5**, 12494 (2015).

442. Wasmer, C. *et al.* Structural similarity between the prion domain of HET-s and a homologue can explain amyloid cross-seeding in spite of limited sequence identity. *J Mol Biol* **402**, 311–325 (2010).

443. Daskalov, A. *et al.* Structural and molecular basis of cross-seeding barriers in amyloids 1 2. *bioRxiv* 2020.07.06.188508 (2020). doi:10.1101/2020.07.06.188508

444. Mompeán, M. *et al.* The Structure of the Necrosome RIPK1-RIPK3 Core, a Human Hetero-Amyloid Signaling Complex. *Cell* (2018). doi:10.1016/j.cell.2018.03.032

445. Bhasne, K. & Mukhopadhyay, S. Formation of Heterotypic Amyloids: α-Synuclein in Co-Aggregation. *Proteomics* (2018). doi:10.1002/pmic.201800059

446. Vasconcelos, B. *et al.* Heterotypic seeding of Tau fibrillization by pre-aggregated Abeta provides potent

seeds for prion-like seeding and propagation of Tau-pathology in vivo. *Acta Neuropathol.* (2016). doi:10.1007/s00401-015-1525-x

447. Ahmed, A. B., Znassi, N., Château, M. T. & Kajava, A. V. A structure-based approach to predict predisposition to amyloidosis. *Alzheimer's Dement.* (2015). doi:10.1016/j.jalz.2014.06.007

448. Bhattacharya, A., Wunderlich, Z., Monleon, D., Tejero, R. & Montelione, G. T. Assessing model accuracy using the homology modeling automatically (HOMA) software. *Proteins Struct. Funct. Genet.* (2008). doi:10.1002/prot.21466

449. Bockmann, A. *et al.* The Molecular Organization of the Fungal Prion HET-s in Its Amyloid Form. *J. Mol. Biol.* **394**, 119–127 (2009).

450. Daskalov, A. *et al.* Contribution of specific residues of the beta-solenoid fold to HET-s prion function, amyloid structure and stability. *PLoS Pathog* **10**, e1004158 (2014).

451. Wan, W. & Stubbs, G. Fungal prion HET-s as a model for structural complexity and self-propagation in prions. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1322933111

452. Wang, J. *et al.* Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science (80-. ).* (2019). doi:10.1126/science.aav5870

453. Wang, J. *et al.* Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science (80-. ).* (2019). doi:10.1126/science.aav5868

454. Udan, M. & Baloh, R. H. Implications of the prion-related Q/N domains in TDP-43 and FUS. *Prion* (2011). doi:10.4161/pri.5.1.14265

455. Derkatch, I. L. *et al.* Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI+] prion in yeast and aggregation of Sup35 in vitro. *Proc Natl Acad Sci U S A* **101**, 12934–12939 (2004).

456. Kurt, T. D. *et al.* Asparagine and glutamine ladders promote cross-species prion conversion. *J. Biol. Chem.* (2017). doi:10.1074/jbc.M117.794107

457. Yuan, S. & Akey, C. W. Apoptosome structure, assembly, and procaspase activation. *Structure* (2013). doi:10.1016/j.str.2013.02.024

458. Hou, F. *et al.* MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response. *Cell* (2011). doi:10.1016/j.cell.2011.06.041

459. Lu, A. *et al.* Unified polymerization mechanism for the assembly of asc-dependent inflammasomes. *Cell* (2014). doi:10.1016/j.cell.2014.02.008

460. Lin, S. C., Lo, Y. C. & Wu, H. Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature* (2010). doi:10.1038/nature09121

461. Tay, S. *et al.* Single-cell NF-B dynamics reveal digital activation and analogue information processing. *Nature* (2010). doi:10.1038/nature09145

462. Loquet, A. & Saupe, S. J. Diversity of amyloid motifs in NLR signaling in fungi. *Biomolecules* (2017). doi:10.3390/biom7020038

463. Benkemoun, L. *et al.* Methods for the in vivo and in vitro analysis of [Het-s] prion infectivity. *Methods* **39**, 61–67 (2006).

464. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinforma.* (2016). doi:10.1002/cpbi.3

465. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* (2004). doi:10.1002/jcc.20084

466. Shen, M. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* (2006). doi:10.1110/ps.062416606

467. Wulf, M. A., Senatore, A. & Aguzzi, A. The biological function of the cellular prion protein: an update. *BMC Biol* **15**, 34 (2017).

468. Sperling, R. & Johnson, K. Biomarkers of alzheimer disease: Current and future applications to diagnostic criteria. *CONTINUUM Lifelong Learning in Neurology* (2013). doi:10.1212/01.CON.0000429181.60095.99

469. Jellinger, K. A., Janetzky, B., Attems, J. & Kienzl, E. Biomarkers for early diagnosis of Alzheimer disease: "ALZheimer ASsociated gene" - A new blood biomarker? *Journal of Cellular and Molecular Medicine* (2008). doi:10.1111/j.1582-4934.2008.00313.x

470. Cai, L., Innis, R. & Pike, V. Radioligand Development for PET Imaging of β-Amyloid (Aβ)-Current Status. *Curr. Med. Chem.* **14**, 19–52 (2006).

471. LeVine, H. Quantification of β-sheet amyloid fibril structures with thioflavin T. *Methods Enzymol.* (1999). doi:10.1016/S0076-6879(99)09020-5

472. Groenning, M. Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *Journal of Chemical Biology* (2010). doi:10.1007/s12154-009-0027-5

473. Hawe, A., Sutter, M. & Jiskoot, W. Extrinsic fluorescent dyes as tools for protein characterization. *Pharmaceutical Research* (2008). doi:10.1007/s11095-007-9516-9

474. Lindgren, M., Sörgjerd, K. & Hammarström, P. Detection and characterization of aggregates, prefibrillar amyloidogenic oligomers, and protofibrils using fluorescence spectroscopy. *Biophys. J.* (2005). doi:10.1529/biophysj.104.049700

475. Buchner, J. & Kiefhaber, T. *Protein Folding Handbook*. *Protein Folding Handbook* (2008). doi:10.1002/9783527619498

476. Howie, A. J. & Brewer, D. B. Optical properties of amyloid stained by Congo red: History and mechanisms. *Micron* **40**, 285–301 (2009).

477. Benditt, E. P., Eriksen, N. & Berglund, C. Congo red dichroism with dispersed amyloid fibrils, an extrinsic cotton effect. *Proc. Natl. Acad. Sci. U. S. A.* **66**, 1044–1051 (1970).

478. Naiki, H. & Gejyo, F. Kinetic analysis of amyloid fibril formation. *Methods Enzymol.* (1999). doi:10.1016/S0076-6879(99)09022-9

479. Kuznetsova, I. M., Sulatskaya, A. I., Maskevich, A. A., Uversky, V. N. & Turoverov, K. K. High Fluorescence Anisotropy of Thioflavin T in Aqueous Solution Resulting from Its Molecular Rotor Nature. *Anal. Chem.* (2016). doi:10.1021/acs.analchem.5b02747

480. Stsiapura, V. I., Maskevich, A. A., Kuzmitsky, V. A., Turoverov, K. K. & Kuznetsova, I. M. Computational study of thioflavin T torsional relaxation in the excited state. *J. Phys. Chem. A* (2007). doi:10.1021/jp070590o

481. Naiki, H., Higuchi, K., Hosokawa, M. & Takeda, T. Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavine T. *Anal. Biochem.* (1989). doi:10.1016/0003-2697(89)90046-8

482. Levine, H. Thioflavine T interaction with synthetic Alzheimer's disease β-amyloid peptides: Detection of amyloid aggregation in solution. *Protein Sci.* (1993). doi:10.1002/pro.5560020312

483. Lindberg, D. J., Wranne, M. S., Gilbert Gatty, M., Westerlund, F. & Esbjörner, E. K. Steady-state and time-resolved Thioflavin-T fluorescence can report on morphological differences in amyloid fibrils formed by Aβ(1-40) and Aβ(1-42). *Biochem. Biophys. Res. Commun.* (2015). doi:10.1016/j.bbrc.2015.01.132

484. Kuznetsova, I. M., Sulatskaya, A. I., Uversky, V. N. & Turoverov, K. K. A new trend in the experimental methodology for the analysis of the thioflavin T binding to amyloid fibrils. *Mol. Neurobiol.* (2012). doi:10.1007/s12035-012-8272-y

485. Sulatskaya, A. I., Maskevich, A. A., Kuznetsova, I. M., Uversky, V. N. & Turoverov, K. K. Fluorescence quantum yield of thioflavin T in rigid isotropic solution and incorporated into the amyloid fibrils. *PLoS One* (2010). doi:10.1371/journal.pone.0015385

486. Brooks, D. J. Imaging amyloid in Parkinson's disease dementia and dementia with Lewy bodies with positron emission tomography. *Movement Disorders* (2009). doi:10.1002/mds.22581

487. Kung, M. P. *et al.* IMPY: An improved thioflavin-T derivative for in vivo labeling of β-amyloid plaques. *Brain Res.* (2002). doi:10.1016/S0006-8993(02)03436-4

488. Klunk, W. E. *et al.* Uncharged thioflavin-T derivatives bind to amyloid-beta protein with high affinity and readily enter the brain. *Life Sci.* **69**, 1471–1484 (2001).

489. Mathis, C. A. *et al.* A lipophilic thioflavin-T derivative for positron emission tomography (PET) imaging of amyloid in brain. *Bioorganic Med. Chem. Lett.* (2002). doi:10.1016/S0960-894X(01)00734-X

490. Groenning, M. *et al.* Study on the binding of Thioflavin T to β-sheet-rich and non-β-sheet cavities. *J. Struct. Biol.* (2007). doi:10.1016/j.jsb.2006.12.010

491. Sabaté, R. & Saupe, S. J. Thioflavin T fluorescence anisotropy: An alternative technique for the study of amyloid aggregation. *Biochem. Biophys. Res. Commun.* (2007). doi:10.1016/j.bbrc.2007.06.063

492. Sabaté, R., Lascu, I. & Saupe, S. J. On the binding of Thioflavin-T to HET-s amyloid fibrils assembled at pH 2. *J. Struct. Biol.* (2008). doi:10.1016/j.jsb.2008.02.002

493. Frieg, B., Gremer, L., Heise, H., Willbold, D. & Gohlke, H. Binding modes of thioflavin T and Congo red to the fibril structure of amyloid-β(1-42). *Chem. Commun.* (2020). doi:10.1039/d0cc01161d

494. Krebs, M. R. H., Bromley, E. H. C. & Donald, A. M. The binding of thioflavin-T to amyloid fibrils: Localisation and implications. *J. Struct. Biol.* **149**, 30–37 (2005).

495. Wu, C. *et al.* The binding of thioflavin T and its neutral analog... [J Mol Biol. 2008] - PubMed result. *J. Mol. Biol.* (2008). doi:10.1016/j.jmb.2008.09.062

496. Vázquez-Fernández, E. *et al.* The Structural Architecture of an Infectious Mammalian Prion Using Electron Cryomicroscopy. *PLoS Pathog.* (2016). doi:10.1371/journal.ppat.1005835

497. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* (2019). doi:10.1038/s41586-019-1026-5

498. Kajava, A. V, Baxa, U., Wickner, R. B. & Steven, A. C. A model for Ure2p prion filaments and other amyloids: the parallel superpleated beta-structure. *Proc Natl Acad Sci U S A* **101**, 7885–7890 (2004).

499. Margittai, M. & Langen, R. Template-assisted filament growth by parallel stacking of tau. *Proc Natl Acad Sci U S A* **101**, 10278–10283 (2004).

500. Pickersgill, R. W. A primordial structure underlying amyloid. *Structure* (2003). doi:10.1016/S0969-2126(03)00009-1

501. Kajava, A. V. & Steven, A. C. β-Rolls, β-Helices, and Other β-Solenoid Proteins. *Advances in Protein Chemistry* (2006). doi:10.1016/S0065-3233(06)73003-0

502. Daskalov, A. *et al.* Identification of a novel cell death-inducing domain reveals that fungal amyloid-controlled programmed cell death is related to necroptosis. *Proc Natl Acad Sci U S A* **113**, 2720–2725 (2016).

503. Sulatskaya, A. I., Lavysh, A. V., Maskevich, A. A., Kuznetsova, I. M. & Turoverov, K. K. Thioflavin T fluoresces as excimer in highly concentrated aqueous solutions and as monomer being incorporated in amyloid fibrils. *Sci. Rep.* (2017). doi:10.1038/s41598-017-02237-7

504. Raj, C. R. & Ramaraj, R. γ-Cyclodextrin induced intermolecular excimer formation of thioflavin T. *Chem. Phys. Lett.* (1997). doi:10.1016/s0009-2614(97)00600-3

505. Wan, W. *et al.* Heterogeneous seeding of a prion structure by a generic amyloid form of the fungal prion-forming domain HET-s(218-289). *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.505511

506. Wasmer, C. *et al.* Infectious and noninfectious amyloids of the HET-s(218-289) prion have different NMR spectra. *Angew. Chem. Int. Ed. Engl.* **47**, 5839–5841 (2008).

507. Schütz, A. K. *et al.* The amyloid-congo red interface at atomic resolution. *Angew. Chemie - Int. Ed.* (2011). doi:10.1002/anie.201008276

508. Schütz, A. K. *et al.* Binding of Polythiophenes to Amyloids: Structural Mapping of the Pharmacophore. *ACS Chem. Neurosci.* (2018). doi:10.1021/acschemneuro.7b00397

509. Xu, Y., Lorieau, J. & McDermott, A. E. Triosephosphate Isomerase: 15N and 13C Chemical Shift Assignments and Conformational Change upon Ligand Binding by Magic-Angle Spinning Solid-State NMR Spectroscopy. *J. Mol. Biol.* (2010). doi:10.1016/j.jmb.2009.10.043

510. Sen, A. *et al.* Mass analysis by scanning transmission electron microscopy and electron diffraction validate predictions of stacked β-solenoid model of HET-s prion fibrils. *J. Biol. Chem.* (2007). doi:10.1074/jbc.M611464200

511. Gowda, C. *et al.* The conformation of the Congo-red ligand bound to amyloid fibrils HET-s(218–289): a solid-state NMR study. *J. Biomol. NMR* (2017). doi:10.1007/s10858-017-0148-z

512. Voropai, E. S. *et al.* Spectral properties of thioflavin T and its complexes with amyloid fibrils. *J. Appl. Spectrosc.* (2003). doi:10.1023/B:JAPS.0000016303.37573.7e

513. Stsiapura, V. I. *et al.* Thioflavin T as a molecular rotor: Fluorescent properties of thioflavin T in solvents with different viscosity. *J. Phys. Chem. B* (2008). doi:10.1021/jp805822c

514. Maskevich, A. A. *et al.* Spectral properties of thioflavin T in solvents with different dielectric properties and in a fibril-incorporated form. *J. Proteome Res.* (2007). doi:10.1021/pr0605567