TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie
Bayerisches NMR Zentrum
Festkörper NMR Spektroskopie

NMR Investigations of Antibody Light Chains Involved in AL-Amyloidosis and the Stress granules forming protein TIA-1

Tejaswini Pradhan

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation

Vorsitzender:	Prof. Dr. Franz Hagn
Prüfer der Dissertation:	1. Prof. Dr. Bernd Reif
	2. Prof. Dr. Johannes Buchner

Die Dissertation wurde am …18.02.2019…… bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am ...12.03.2019… angenommen

Abstract

Protein aggregation can lead to the formation of either disease related amyloids or functional amyloids via misfolding. Out of more than 30 amyloid diseases, antibody light chain amyloidosis (AL amyloidosis) is one such rare systemic disease. In this disease, the precursor protein is antibody free light chain (LC), which is usually composed of variable domain (V_L). It is a plasma cell dyscrasia where monoclonal Immunoglobulin light chain secreted into the blood stream and aggregates into amyloid fibrils in the various organs of the body affecting especially to heart and kidney dysfunction and eventually leading to death. Immunoglobulin's genetics causes repertoire of sequences making this disease difficult to study. Some amyloids however are not involved in disease but has regulatory function such as TIA-1 (T-cell Intracellular Antigen-1) known as stress granule marker. In this thesis, antibody light chain and TIA-1 are the two amyloidogenic proteins are investigated. We have employed combination of solution and solid-state NMR along with other biophysical technique to study these amyloidogenic proteins.

Our main objective is to understand the misfolding process in V_L proteins. Here we studied, S3706 (λIII subgroup) light chain variable domain sequence derived from specific patient whose fibrils were deposited in heart and are associated to AL amyloidosis. It is to be noted that due to antibody diversity every patient sequence is unique. There is very few information available for this protein until now. For comparison, we cloned germline and single point mutated proteins of S3706. We assigned V_L native monomeric state of patient and germline by solution state NMR followed by characterisation of dimer state, which is in equilibrium state that is usually found in all LCs. In S3706 protein, at higher concentration dimer population is high which protect the aggregation process while at lower concentration monomer is dominated that promotes aggregation process. For the first time, we uncovered and characterise oligomeric state via aggregation process using solution state NMR, which is present specifically in patient protein, S3706_patdel. Germline protein, S3706_GL does not show any intermediates states and single point mutation S3706_R50G show similarity to both patient and germline. We also suspect the presence of several oligomeric states are either due to polymorphism or merely represents the stages of aggregation pathway. In addition to support our data, we employed Thioflavin T assay, electron microscopy, dynamic light scattering and Gel electrophoresis technique.

EGCG is polyphenol tea extract, is known to inhibit the fibrils formation for many amyloidogenic proteins. We did the initial studies for the binding of EGCG with S3706 using solution state NMR. In patient protein although the chemical shift perturbations were not large, the changes were localized around proline residues and dimer interface. The intensity was largely decreased due to the precipitation of protein. Our results reveal that patient

ABSTRACT

S3706 is more amyloidogenic than its respective germline protein.

Fibrils are the final most stable state in aggregation process in amyloids. We prepared S3706 and its variants fibrils using *ex-vivo* and *in-vitro* seeds that are further analysed by solid state NMR. We reproduced the fibril preparation and spectral features in S3706_patdel. We found that electrostatic interactions or salt bridges connection from Lys and Arg are specifically in patient fibrils that are weak or absent in germline fibrils. We proposed that Arg at 50^{th} position is essential for salt bridge connection. This was proved by single point mutation, S3706_R50G fibrils where salt bridge connection was found to be broken for both Lys and Arg side chain. We assigned more than 60% rigid residues of S3706_patdel using ^{13}C detected 3D experiments in solid state NMR.

Another protein, we study here is Ribonucleic acid (RNA) binding protein, TIA-1 (T-cell Intracellular Antigen-1) composed of RRMs (RNA recognition motifs) and Q-rich domain. Our research is focused on C-terminal Q-rich domain, which is responsible for stress granule formation and recently found to be involved in Amyotrophic lateral sclerosis (ALS) disease. We investigated specifically how Q-rich domain is interacting with RRM domains in TIA-1 protein. QRD44 comprising of first 44 residues from Q-rich domain can show phase separation and form fibrils independently. Comparison of rigid residues from RRMs and without RRMs with QRD44 domain show only certain segments from RRMs are involved in fibrils. By mutating linker region, we proved that that linker between RRM1 and RRM2 are interacting with Q-rich domain. It was also found that the protein with RRM domains exhibits faster aggregation process. To characterize the fibrils, we used Differential interference contrast (DIC) microscopy, Electron microscopy (EM) and Thioflavin T (ThT) assay. Our findings provide structural insight of TIA-1 protein by expanding the knowledge of domain orientation.

In summary, V_L projects results provide new vision in the aggregation of LCs proteins, which will further help in understanding of AL amyloidosis. This can contribute in the development of therapeutics of this amyloid related disease. In TIA-1 project, the structural insight specifically in C-terminal Q-rich domain will further enable to broad the knowledge of domain orientation in U1 snRNP complex.

Acknowledgements

I would like to acknowledge my supervisor Prof. Bernd Reif for giving me the opportunity to do PhD under his guidance. My sincere gratitude goes to him for his invaluable guidance, encouragement and giving independence to learn and to gain confidence in my work.

I am very grateful to the thesis committee members, Prof. Johannes Buchner and Prof. Franz Hagn for their participation as an examiner and chair respectively.

My deepest appreciation to Dr. Manuel Hora, former PhD student in our group for his continuous assistance. He has not only wholehearted helped me during the project work as well as in thesis proofreading but also boost up my confidence level.

I would like to thank all my collaborators on different projects. First, I would like to thank Prof. Marcus Fändrich and Dr. Karthikeyan Annamalai for providing the patient sequence and fibrils without which the project could not have been started. For V_L project thanks to Prof. Johannes Buchner again for his useful discussion and Dr. Benedikt Weber for his help in V_L project and sharing the plasmids and for TIA-1 project, thanks to Prof. Michael Sattler and Dr. Parvin jagtap for the discussion.

I would like to thank Dr. Michaela Aichler and Gabriele Mettenleiter for their assistance in TEM measurement in HMGU.

I am grateful to Dr. Ute Hegenbert and Dr. Stefan Schönland for the invitation to Heidelberg meetings and helpful emails during all these years.

My sincere thanks to GRK1721 graduate school for providing useful courses that improved in overall development and knowledge for different techniques.

I owe a deep vote of thanks to Dr. Riddhiman Sarkar for helping me in solid-state NMR experiments, teaching, and all discussions. I also want to thank previous lab members our group, Dr. Kai Xue, Dr. Diana, Dr. Vanessa, Dr. Elke, Dr. Maria, Antje for their support in need and sharing their knowledge. I want to acknowledge Dr. Carina Motz and Benita for the lab assistance. I would like to thank Asita Djamschidi for giving information and cooperation in official work. I would like to thank Zheng Niu for all her support, friendly discussion about work and life. Thanks to Munirah for sharing time. Thanks to all our group colleagues Marcus, Arpita, Dr. Saba, Matthias and Natascha for providing nice environment to work in BNMRZ. Thanks to all the people in TUM who somehow helped me during my PhD stay.

I would like to express my deepest gratitude to my husband, Mr. Rohit Kumar Gupta for all his support, motivation and moral strength during my entire stay of PhD. I would also like to thank my parents and brother for their moral strength, affection and motivation.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

INTRODUCTION

1 Introduction

1.1 Protein aggregation and Amyloidosis

A common phenomenon that occurs often in the biochemistry during protein purification is an insoluble clump or accumulation of protein that is problematic. However, occasionally the situation is unlike when we want to study these insoluble proteins that are formed due to specific perturbation in structure. These large insoluble agglomerates are formed due to self-association of protein known as aggregation[1]. The protein aggregation is controlled and regulated by molecular machinery known as molecular chaperones[2]. Aggregation is associated with many infectious diseases such as in prions and amyloid related diseases. Amyloids is derived from Latin amylum and is a general term given where protein aggregates to form insoluble fibrils like structure as a result of protein misfolding. Therefore, protein misfolding event is a major hallmark for all amyloidogenic disease despite the protein precursor is distinct in each disease[3].

The word 'amyloids' was coined by Rudolph Virchow as they stain blue with an iodine stain and were believed to be similar to starch or cellulose by mistake which was not the case, instead they were primarily proteinaceous[4]. The positive iodine stain was later explained by the presence of glycosaminoglycans that is present in mostly all amyloids[5]. For the usage of pathological diagnosis, amyloids are defined as 'extracellular depositions of protein fibrils with characteristic appearance in Electron microscope (EM), typical X-ray diffraction pattern and affinity for Congo red with concomitant green birefringence' [6]. The first case to be observed of amyloid disease was in 1639 by Nicolaus Fontanus in 'waxy liver and white stone-containing spleen[7,8]. The most studied, well known and utmost devastating amyloid disease is Alzheimer's; first described by Alois Alzheimer in 1906[9]. Alzheimer's is one of the neuro-generative diseases where Aβ-amyloid plaques are localized. There are several degenerative diseases where amyloid fibrils are not localized but present in various organs known as systemic amyloidosis[10,11]. (**Table 1)** Other than pathogens, amyloids also found in bio-films[12], harmones[13], which has specific biological role[3,14,15].

In biological systems, proteins tends to fold to their thermodynamically stable native globular state under physiological conditions[16]. However, some proteins are natively unfolded proteins also known as 'intrinsically disordered proteins (IDPs)' having a random coil structure, with different function in biological systems for example Aβ-peptide[17]. It has been demonstrated that aggregation involves the self-assembly of highly native-like folding intermediates[18]. The native state can convert into partially folded states, or oligomers (toxic or non-toxic), which can lead to fibrils state as most stable state (pathogenic or non-pathogenic) or may form amorphous aggregates[19]. It is believed that partially unfolded states trigger the fibrils formation as hydrophobic residue which are buried are exposed for

intermolecular interactions[20]. The transient states formed in aggregation process are found to be different from the intermediates formed during folding[21]. The aggregates formed can be ordered as in case all amyloid fibrils or disordered as in inclusion bodies[22]. These initial aggregates form the basis for nucleation process which can also recruits the normal protein to form fibrils. *In-vitro* seeds play an important role in enhancing fibril growth and removing the lag phase in fibril formation[23]. Other than pathological aggregation, in some cases protein aggregation happens as a biological cellular function like in stress granules and P-bodies where RNA binding protein that possess disordered domains can self-assembled to form fibrils and function as storage or regulation[24,25]. **(Figure 1a)**

Table 1 Amyloid related diseases[3,26]

Disease name	**Precursor protein / peptide**	**Protein structure**	**Localized / Systemic**
Alzheimer's disease	Aβ peptides	Intrinsically disordered	Localized
Parkinson's disease	α-Synuclein	Intrinsically disordered	Localized
Type II diabetes	Amylin or IAPP	Intrinsically disordered	Localized
Haemodialysis related amyloidosis	β2m	β-Sheets and Ig like	Systemic
AL amyloidosis	Light chain Immunoglobulin	β-Sheets and Ig like	Systemic/ Localized
AH amyloidosis	Heavy chain Immunoglobulin	β-Sheets and Ig like	Systemic/ Localized
AA amyloidosis	Serum amyloid A	α- helical	Systemic
ATTR	Transthyretin	β-Sheets	Systemic
Huntington's disease	Huntingtin	Intrinsically disordered	Localized

There are many factors that can be responsible for crossing the kinetic barrier for states of protein like temperature, chemical denaturants, and change of pH. The proper underlying mechanism of protein aggregation and to prevent the aggregation is still not well understood. It has been in debate for 10 years about the presumably end state fibrils can explain the progression of the disease but not the cause. Moreover, oligomer intermediates i.e. the initial phase or the protofibrils were reported to be more cytotoxic to cells than fibrils[27–29]. The kinetics of this process is well studied by biophysical methods and can be defined by three characteristics stages, lag phase, growth phase and a final plateau regime. In the lag phase, there can be small events occurring like primary nucleation, elongation, secondary nucleation and fragmentation[23]**. (Figure 1b)** The cross-seeding effects were shown that could explain the progression of disease where normal protein can also be recruited by preformed amyloid fibrils.

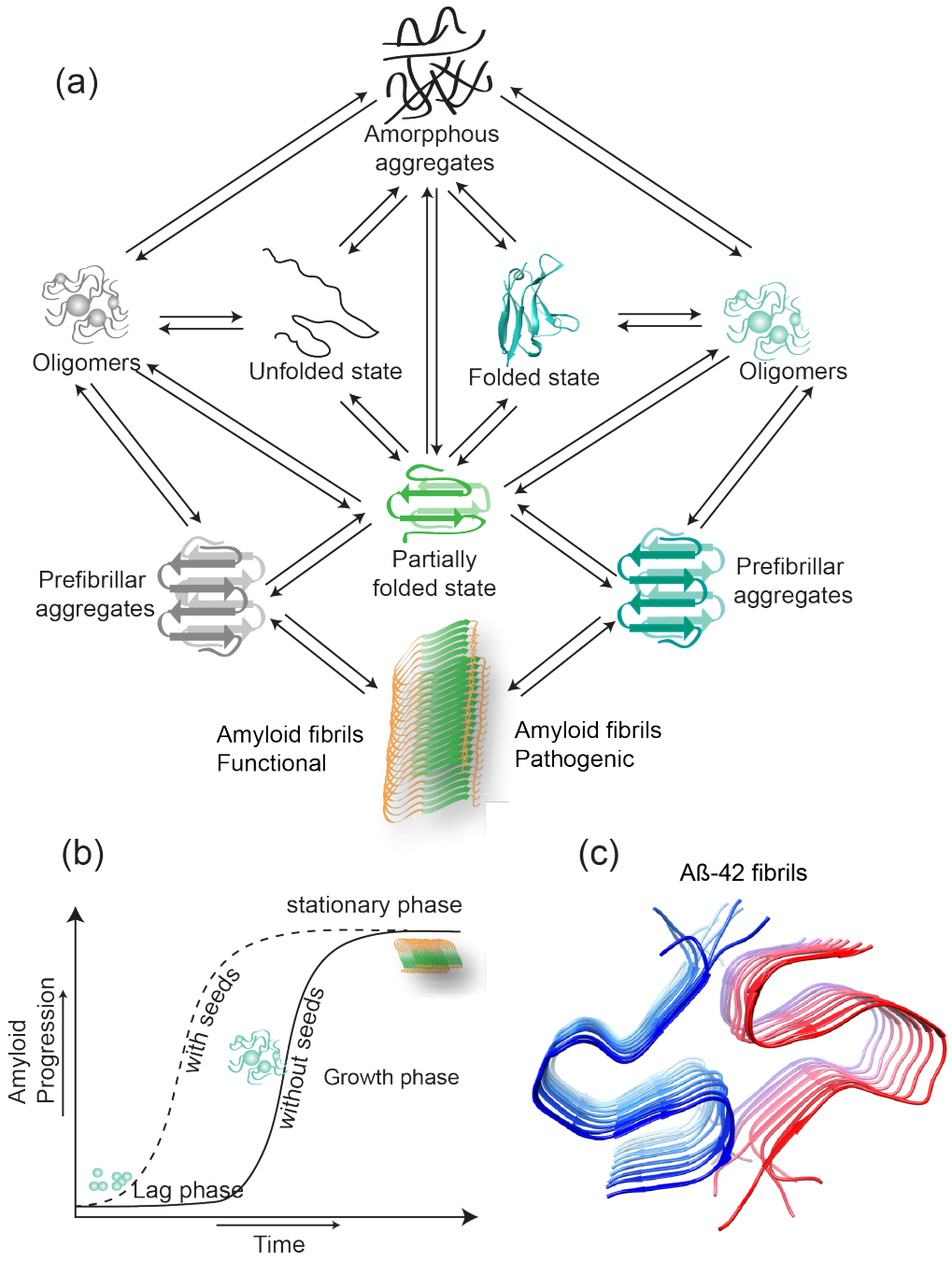


Figure 1: Common pathways in protein aggregation and fibril formation[19]

(a) Different conformational intermediates of protein during aggregation that includes unfolded and globular state. (b) Aggregation kinetics of protein towards amyloid formation with and without seeds that included lag phase monomer, dimer or trimers followed by growth phase with intermediates (partially fooled or oligomers) and in the end, it is stationary phase with mature fibrils. With seeds lag phase is reduced and homogeneous fibrils is obtained[23]*. (c) Aβ-42 fibrils structure solved by ssNMR (pdb: 5KK3)*[30]*.The repetitive subunits consists of two monomers (blue colour and red colour) with β-strands in each layer. From light to dark colour is the fibril axis and, in both direction, fibrils could be extended.*

Despite all differences in the sequence and native structure all amyloids are closely packed and highly ordered structures and contain common 'cross β-sheets'[31–36]. Moreover, solid state NMR (ssNMR) spectroscopy has allowed providing the direct evidence for the pattern of β- sheets in fibrillar environment at atomic level.

Aβ is first and mostly studied amyloid among all amyloid protein due to its severe epidemic. Several groups in ssNMR field, were able to achieve high resolution structure[30,35,37,38]**.** Due to polymorphism that is exhibited by many amyloids, different structure can be possible, like in Aβ shows different polymorph both in *in-vivo* and *in-vitro*. Aβ (1–42) appears to be a more toxic species than Aβ(1–40) with only two amino acid difference in C-terminal[39]. (**Figure 1c)** Other than Aβ, other amyloid structures solved by ssNMR are β2M, TTR, α-synuclein[34,40–42].

1.2 Immunoglobulins

Immunoglobulin's (Ig) or antibodies are globular proteins with Y-shaped structure. There are five main classes of Ig HC, each class defines IgA, IgD, IgE, IgG and IgM isotypes. Immunoglobulin consists of Fc and Fab fragment, which is associated to glycan association and antigen binding respectively. The Fab fragment is composed of two identical light chains (LC) and two heavy chains (HC), which are connected by conserved disulphide bridges. Each LC chain consist of one N-terminal variable domain (V_L) and terminal constant domain (C_L) whereas HC consist of only one variable domain (V_H) and three Constant domain (C_H 1, C_H 2, C_H3) which constitute the Fc fragment (crystalline) [43–45]. The typical LC has approx. 25kDa molecular weight and each V domain or C domain has 110-130 amino acids constituting 12-13 kDa approximately. LC can be of two type κ and λ isotypes which are further divided into subgroups[46]. There are seven subgroups $V\lambda_I$, $V\lambda_{II}$, $V\lambda_{III}$, $V\lambda_{IV}$, $V\lambda_V$, $V\lambda_{VI}$ and $V\lambda_{VII}$ for λ isotypes and 6 subgroup $V\kappa_I$, $V\kappa_{II}$, $V\kappa_{III}$, $V\kappa_{IV}$, $V\kappa_{V,}$ $V\kappa_{VI}$ for κ isotypes[47,48].

The Ig LC variable domain has a Greek fold also known as Ig fold which consists of nine anti-parallel β-strands A, B, C, C', C", D, E, F and G with 4+5 orientation while the HC variable domain consists of seven β-strands A, B, C, D, E, F, G from N-terminus to C-terminus with 4+3 orientation[44]. Both LC and HC have similarity in tertiary structure but differ in β-barrel structure[43]. One of the important hallmark in Ig like fold is the conservation of the disulphide bridge between B and F stand in the hydrophobic core in close proximity to tryptophan[49]. The striking difference between constant and variable domain in LCs is the presence of CDRs (complementarity determining regions) in variable domain, these are the antigen-binding sites[50,51]. The less variable region provides the structural framework (FR1, FR2, FR3) for the CDRs (CDR1, CDR2, CDR3). **(Figure 2)**

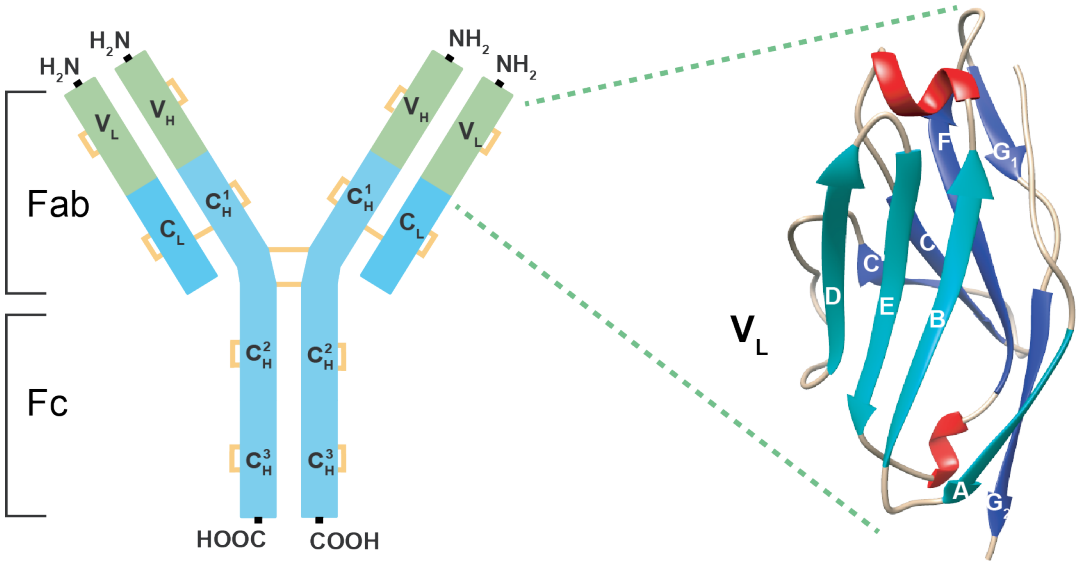


Figure 2: Antibody structure

Antibody structure domain linked by disulphide bonds. Right hand side is the LC variable domain structure with A, B, E, D, C, C', F, G_1, G_2 β- strands.

1.2.1 Immunoglobulin genetics

In this section, the focus is about the antibody diversity which -occurs due the complex genetics involved in immunoglobulins. There are theories about the variability of antibodies 'Germline Theory' first proposed by Lederberg in 1959 and widely accepted 'Somatic recombination theory' proposed by Smithies in 1963[52,53]. LC and HC are each encoded by separate multigene family[44,48]. In humans, first 95-98 amino acids of the light chain are encoded by variable (V) and the following 12-13 amino acids are encoded by joining (J) segments located on chromosome 2 for κ light chain and chromosome 22 for λ light chain. The heavy chain is coded by V, J and additional diversity (D) segments located in chromosome 14. The constant domain of light chain is encode by third C-gene[43,54].

During B-cell development, the gene rearrangement of the DNA occurs both in V_L/V_H and C_H region genes. The gene rearrangement for V_L is illustrated in **Figure 3**. Despite the number of germline-encoded V_L segments are less, they can produce infinite number of antibodies to due to the variable region, which has complementary-binding site. The source of the variability in complementary binding region depends upon somatic hypermutation. Along with somatic hypermutation, there are other sources of variation, which are listed in **Table 2**. V(D)J combination produces the primary repertoire while class-switch recombination (CSR) and somatic hypermutation improves the quality of the B-cell response after antigen initiation[55]. Although there are some variation in amino acid sequence, the Ig fold is same and there are cluster of conserved, buried core residues which can be identified from same families[49].

Table 2 Source of antibody sequence variation[56]

Source of Variation	CDR1	CDR2	CDR3
Sequence encoding	V segment	V segment	VJ/VDJ junction
Junctional flexibility	-	-	+
P- nucleotide addition	-	-	+
N- nucleotide addition	-	-	+
Somatic hypermutation	+	+	+

Table 3 Combinatorial antibody diversity in Humans

Multiple germline segments	HC	LC	
		κ	λ
V	51	40	30
D	27	0	0
J	6	5	4
VDJ and VJ combination	51*27*6 = 8262	40*5 = 200	30*4 = 120
Possible combinatorial associations of HC and LC from single subjects		8262*(200*120) = $2.64*10^6$	

Here, only functional gene segments are listed. Diversity contributed by junctional flexibility, P-region nucleotide addition and somatic hypermutation are not included, so actual numbers exceeds these estimates by several orders of magnitude[54]*.*

Thus, due to variability of antibodies that are discussed above, every single individual has different monoclonal antibody even though the germline gene are same.

1.3 Antibody Light Chain (AL) amyloidosis

In this section, there is brief history of AL amyloidosis followed by detail description of the disease. Magnus-Levy observed that in 31 cases of multiple myeloma was present in B-cells or bone marrow and in 18 cases in other organs[57]. He showed the relation of Bence's Jones protein and amyloidosis which was later showed by Osserman[58], Apitz[59]. Later he demonstrated that the origin of these amyloids is in plasma cells[60]. In 1968, Glenner revealed that in X-ray diffraction, these amyloid filaments are pleated sheet structure and adjacent chain segments were laterally arranged in antiparallel manner [61].

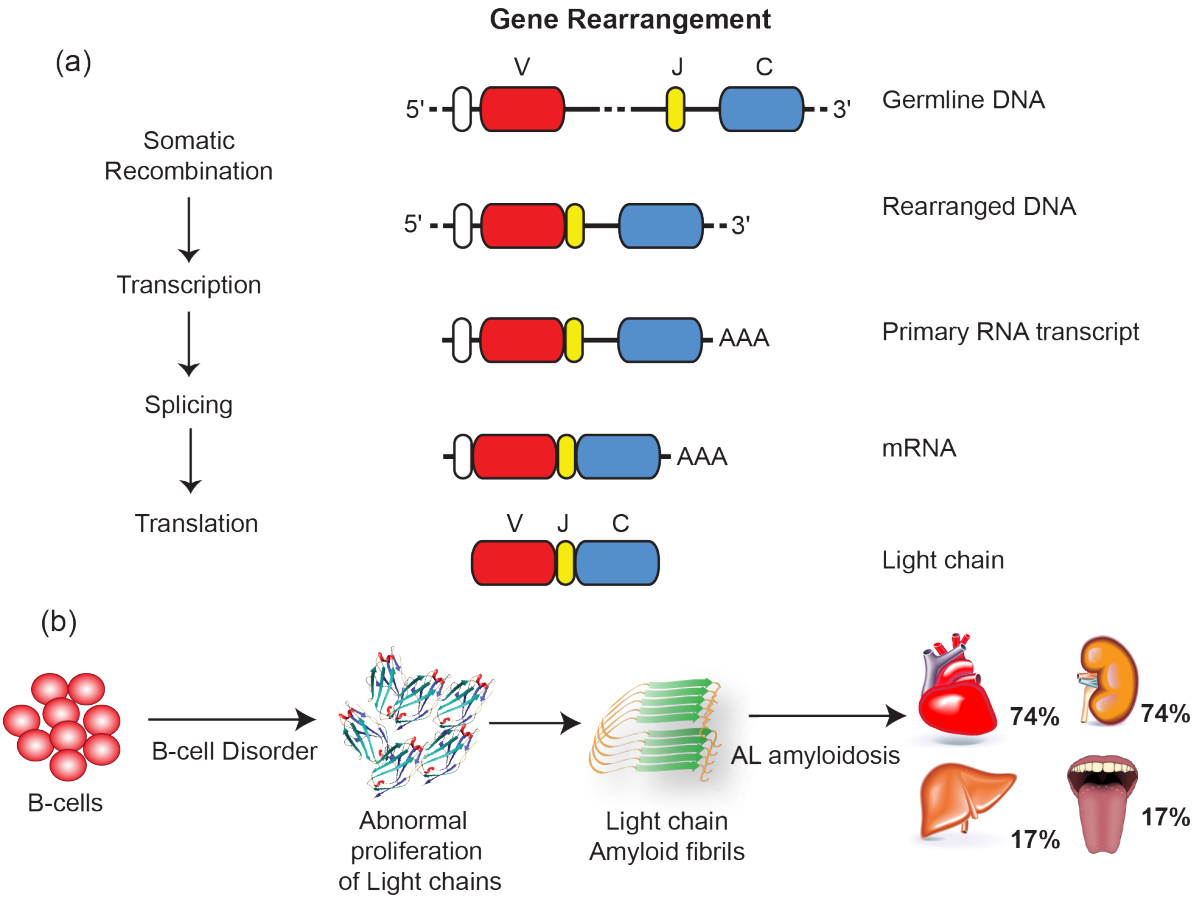


Figure 3: Immunoglobulin genetics and AL amyloidosis.

(a) Gene rearrangement in antibody responsible for its diverse nature[54,56]*. (b) Flowchart depicting the progression of AL amyloidosis. B-cell disorder produces abnormal proliferation of free LCs, which aggregates followed by deposition on various parts of the organs causing AL amyloidosis*[62]*.*

Light chain amyloidosis previously known as primary amyloidosis is a rare disease in which clonal B cell overproduce free Immunoglobulin light chains, which is then secreted into main blood stream where they misfolds into insoluble aggregates or amyloid fibrils. This is systemic fatal plasma cell dyscrasia where fibrils deposit in multiple organs of the body causing the malfunction of the organ. It mainly affects the heart and kidney mostly with 75% and other organs like tongue and liver 25%[63]. **(Figure 3c)** Due to large variation of amino acid sequences, which, occurs during plasma cell differentiation and maturation process as, discussed in Immunoglobulin genetics sections, every sequence is unique. This also makes the study of disease difficult. It is often associated with Multiple myeloma (MM), Light chain disposition disease (LCDD) also plasma cell dyscrasia but they are clinically distinct. In both diseases mentioned above, there is no fibril formation but cells divide leading to cancer in MM while in LCDD the LC precipitation occurs in kidney thus causes Light chain nephropathy[64]. The plasma cell dyscrasia with no amyloid deposition is classified as monoclonal gammopathy of undetermined significance (MGUS).

The concentration of κ and λ in healthy human being is 3.3-19.4 mg/L and 5.7-26.3 mg/L respectively and normal ratio is 0.26/1.65[65]. In AL the κ/λ ratio is 1:3, which suggests that Vλ germ line genes are more prevalent in amyloid formation unlike other plasma cell dyscrasia[66]. Other than LC, there are many other molecules that are commonly present in amyloid deposits like Serum amyloid P (SAP), proteoglycans, extracellular matrix components like collagen IV, elastin[67–69].

Due to immunogenetics discussed in above section every sequence is unique to the patient as every individual has different monoclonal antibody and makes this disease difficult to study. More number of patient sequence studies in future, will allow to determine the possible reason of variable deposition of LCs fibrils in multiple organs[70–72]. It was observed by clinical studies that there is correlation for subgroup of κ and λ LCs with the organs involved in LC amyloid fibrils[66]. In AL amyloidosis patients, it was found that among all subgroups of κ and λ LCs, it was found that $V\lambda_{VI}$ subgroup has preferential in forming fibrils[73]. Other than immunogenetics, factors involved in AL amyloidosis are discussed in following section.

1.3.1 Factors influencing AL amyloidosis

Due to variation in sequences in AL amyloidosis as mentioned in Immunogenetics section, makes difficult to probe the main cause of fibrillogenesis. There are several hypotheses that are reported *in-vitro* studies to probe the fibrillation process[74].

In most of the patients, full V_L domain with small segment of constant is attached was to be found in fibrils, which gives the hint it may be probably, V_L domain is essential for the fibril formation. In regard to this, Poshusta et al. reported that mutation in specific structural region of V_L are associated with AL amyloids[75]. Point mutations can destabilize the native state of the protein towards the fibril formation[76–81]. It was suggested that- specific mutations in germline alter the surface properties of V_L protein BRE may improve the stability of protein and prevent the fibril formation[82]. Point mutation in N- terminal of MAK33 V_L can the overall stability and integrity of LCs[83]. Both thermodynamic and kinetic control play significant role in in AL amyloidosis[84]. It was observed that isotype λ is frequently associated to AL amyloidosis while κ in LCDD disease and among all λ subgroup, λ_{VI} isotype are more prevalent in AL amyloidosis.

Simpson et al. showed that during the folding process in MAK33 V_L domain there can be two intermediates possible which can be the entry point for amyloid pathway[85]. This promotes the idea that more highly unfolded intermediates are the precursor for the fibrillation[86,87]. It was suggested that soluble off pathway oligomeric intermediates are major transient state, with unfolded conformation and form the early events in fibrillation in LEN light chain[88]. This was in agreement with previously published paper about SMA protein that

partially folded states to be as critical precursors[89].

In direction of protein oligomers, domain swapping has been proposed by Benett et. al[90] and Heringa and Taylor[91] which was later extended to pathological oligomerization, as in amyloids[90]. In Ig like domain CLTA-44, show the same feature where strand swapping display cis-trans proline isomerisation and native -like- H-bonding[92]. In addition to thermodynamic and kinetic stability by proline cis-trans isomerization can play major role in amyloid fibril process described in AL-103 LC also recounted in other protein β2-Microglubulin that possess Ig-like fold[84,93]. It was proposed that the structure of oligomers is similar to mature fibrils and not to its native structure and fibril formation via domain swapping is unlikely in MAK33[94]. The soluble oligomers of LCs were reported to be more cytotoxic than its fibrils state[95].

Patients in AL amyloidosis typically contain LCs that are apparently produced by aberrant endo-proteolysis[96,97]. It was hypothesized that full LCs that are kinetically unstable might undergo proteolysis and thus releases amyloidogenic variable fragments[98]. In agreement to this, it was demonstrated that full LC from Bence Jones protein was cleaved by proteolysis using pepsin[99]. However, it is still not clear how proteolysis influences the pathogenesis of the disease. Furthermore, it is still in debate that whether the proteolysis is the first event of the pathogenesis of protein. In addition the role of constant domain has been much in attention because of presence of segments of constant domain in extraction of the amyloid fibrils[100,101].

There are many biomolecules associated with amyloids such as GAGs (glycosaminoglycan's), component of all plasma membranes and extracellular matrix that functions in cell signalling and adhesion. Mode of action of GAGs on AL amyloid fibrils studies suggested that the charge due to sulphate ions and size number due to disaccharide repeats can enhance the process of fibrillation[102]. Post-translational modification is also a common feature in LC amyloids, N-linked glycosylation by mutation and di-sulphide linked dimerization are studied[103,104].

Pepys and his co-workers suggested that circulating SAP which is common component could be the precursor in systemic amyloidosis[69]. Formulating this observation they developed ^{123}I-labeled SAP component scintigraphy technique that is used as diagnosis for the bone marrow involvement in systemic amyloidosis by National amyloidosis centre (NAC)[105–107].

Recruitment of soluble protein from normal antibody repertoire was demonstrated by *in-vitro* studies in which self-seeding and cross-seeding experiments was performed. This shows the self-propagating phenomena of fibrils and reflects the propagation of fibrils in organs[108,109].

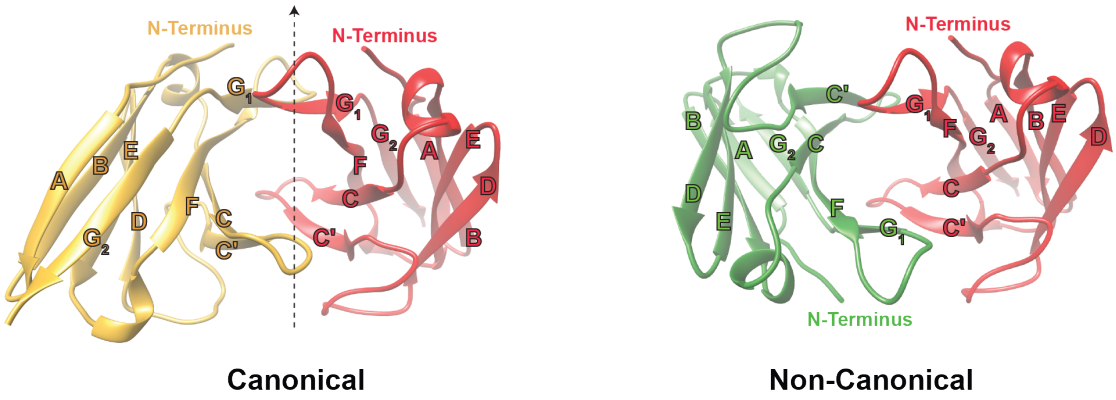


Figure 4: Dimer Interface in V_L proteins.

In canonical the two same strands are opposite to each other and these are mirror images. In non-canonical, one monomer is rotated to 180°.

Another important feature of almost all LCs is the dimer existence either in canonical or non-canonical form. **(Figure 4)** The interface of the V_L-V_L dimer is same as the V_L-V_H dimer in LC quaternary structure[110]. *In-vitro* studies showed that when compared to the germline κIO18/O8 that has canonical form, pathogenic sequence AL09 form altered dimer interface, where one monomer is rotated to 90° or 180°. Later, it was shown that mutating, the amino acids in dimer interface can change dimer interface form i.e. canonical or non-canonical and thus alter the energy landscape which eventually change the amyloid propensity[111–114]. Mutating Tyrosine residue has proved to play significant role in dimer interface in AL09[115]. It has been proved that the protein has to be in monomeric stage for fibril formation rather than dimer which protects the given process[114,116]. The dimer interface can play a crucial role in promoting amyloidogenesis along with other thermodynamic factors.

Ramirez-Alvarado et. al suggested that there is differential dependence of the protein concentration and fibril formation in LCs fibrils. At lower protein concentration there is synergistic effect for recruitment and elongation involves in primary nucleation process and at higher concentration the secondary nucleation is dominant process[117].

1.3.2 Therapeutics for AL amyloidosis

AL amyloidosis is rare disease and it estimated that the incidence of disease occurs 9-14 person per year per million and median survival of the human is less than 10 years from the diagnosis depending on the severity of disease[118]. The amount of serum free light chain assay is done frequently used to probe this disease but not all free light chain leads to the amyloid fibrils in *in-vivo*[119,120]. Other than serum free light chain, cardiac biomarkers NT-proBNP and TroponinT (TnT) forms the basis for prognosis[121]. The major drawback is the late diagnosis due to unspecific common symptoms like weight loss, fatigue, which overlaps with other diseases.

Current treatment includes conventional systemic chemotherapy options for AL amyloidosis like Melphalan-Dexamethasone, Bortezomib- Dexamethasone, Cyclophosphamide-Thalidomide-dexamethasone, Cyclophosphamide- Bortezomib - dexamethasone combinations or stem cell transplantation or organ transplantation depending upon the progression of the disease[122,123]. Bendamustine, Doxycycline are the therapy for which trial is going on[122]. However, these drug treatments have drawbacks too because it is poorly tolerated and develops side effects. Another promising approach is the depletion of SAP by (R)-1-[6-[(R)-2-carboxy- pyrrolidin-1-yl]-6-oxo-hexanoyl] pyrrolidine-2-carboxylic acid (CPHPC) followed by anti-SAP antibody treatment, which binds to all amyloid proteins and its universally present in systemic amyloids[106,124,125].

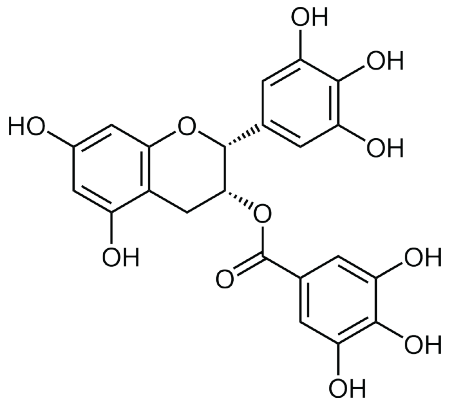


Figure 5: EGCG structure

Another approach to reduces proteotoxicity was to activate unfolded protein response (UPR) transcription factors XBP1s and/or ATF6 which reduces the amyloidogenic LCs secretion in the absence of stress[126]. The novel approach used is to target the amyloid fibrils by using monoclonal antibody NEOD001 (murine 24A)[127,128]. Recently, a bifunctional peptide, p66 has been developed that interacts with mAb 11-1F4 also known as CAEL101, that interacts with AL amyloids and also many other amyloids[129].

Epigallocatechin gallate (EGCG) is a polyphenol extract from green tea, due to its biological importance like anti-oxidative, anti-cancerous, anti-angiogenic, anti-inflammatory, anti-atherogenic and several other health benefits has been studied for almost all amyloid as drug candidate[130–132].**(Figure 5)** Several studies have been done for the EGCG with Aβ, Huntington, Ig LC and α-synuclein as a drug candidate[133–137]. Phase 2 study of EGCG in cardiac patients in TAME-AL amyloidosis in Heidelberg University evaluation shows no significant results.[138] In clinical studied it has displayed improvement in cardiac amyloidosis however, it has some issues due to variable bio-availability and unstability[139,140].

In-vitro studies, suggests that EGCG interacts with amyloids fibrils and converts them into insoluble amorphous aggregates, that are non-toxic to mammalian cells[133]. When added to soluble monomers it prevents the pathways of fibril formation and promotes the formation of unstructured oligomers[133–135]. Apart from EGCG, other tea catechins were also found to have the same effect on amyloid fibrils[141,142]. Oxidation of EGCG drives the remodelling of fibrils by binding with free amines through the formation of Schiff's base and cross-linking but the driving force is still not clear[136]. It interacts with unfolded protein by hydrophobic backbone interactions and H-bonding[135].

In LCs, as the EGCG interaction mechanism involves the unfolded protein that reacts rapidly to the aggregation prone protein to convert insoluble aggregates and thus selects the amyloidogenic protein. Through NMR investigation it was found that, there are two pathways which is possible, one is off-pathway where it binds with proline 59 and other pathway to bind with proline 44 to form irreversible precipitation, which was further confirmed by ssNMR[143]. Recently it has been suggested that in apoA-I, EGCG, depending upon the bioavailability, GAGs and other cofactors may regulates the remodelling present *in-vivo*[134].

1.4 Low Complexity Regions

So far, the introduction describes AL amyloidosis where the monomer protein involved is well structured and folded. There is another category of proteins where it is unstructured in its monomeric form but form structured fibrils. Ribonucleic acid (RNA) binding proteins are such category where one of the is domain random coil structure or unstructured. It typically contains RNA recognition motifs (RRMs) and low complexity motifs or prion like domains (PRD). RRMs are associated to RNA regulation whereas low complexity motifs are involved in RNA granules with cellular functions. (**Figure 6a)**

The low complexity means protein has high occurrence or repetition of specific amino acids in protein sequence such as Fused in Sarcoma (FUS) has 156 amino acids with Serine/ glycine/ glutamine and in T-cell restricted Intracellular Antigen-1 (TIA-1) has PRD domain that contain 23 amino acids with Serine, Threonine or Glutamine. This domain is also known as Q-rich due to high repetition of glutamine residue in this region. These are also termed as 'Intrinsically disordered region' because they possess unstructured random coil structure. Low complexity motifs are characterised by prion like domains according to the prion-forming index. There are 210 human RRM containing proteins out of which 29 possess putative prion domain protein and 12 are top candidates among prions in entire genome[144]. Many RNA binding proteins such as transactive response (TAR) DNA binding protein-43 (TDP-43), Fused in Sarcoma (FUS) are associated with neurodegenerative disease like Amyotrophic lateral sclerosis (ALS) is tabulated in **Table 4.**

Table 4 Few examples of Low complexity proteins and their properties[145]

Protein (amino acids)	Low complexity Motifs	Structured domains	Membrane-less Organelles	morphology	Related Disease
FUS (526)	156aa S/G/Q 55aa RGG 76aa RGG	RRM ZnF	Paraspeckles stress granules	Droplets, Hydrogels	ALS
hnRNP A1/A2 (372/353)	180aa G/S/R/Q	2 RRMs	Stress granules,	Droplets, Hydrogels	ALS, IBM, Paget's, FTLD
RBM14 (669)	300aa A/R/S/Q/P	2 RRMs	Paraspeckles,	Hydrogels	ALS
TDP-43 (414)	43aa G/F/N 9aa A-rich 17aa Q/N 37aa S/G	2 RRMs	Stress granules	Droplets	ALS
TIA-1 (386)	23aa S/T/Q	3 RRMs	Stress granules	Droplets	Welander distal myopathy, ALS

FUS = Fused in Sarcoma; hnRNP = Heterogeneous nuclear Ribonucleoprotein; TDP = transactive response (TAR) DNA binding protein; TIA-1 = T-cell intracellular Antigen-1; ALS = Amyotrophic lateral Sclerosis; IBM = Inclusion body myositis; FTLD = Frontotemporal lobar degeneration

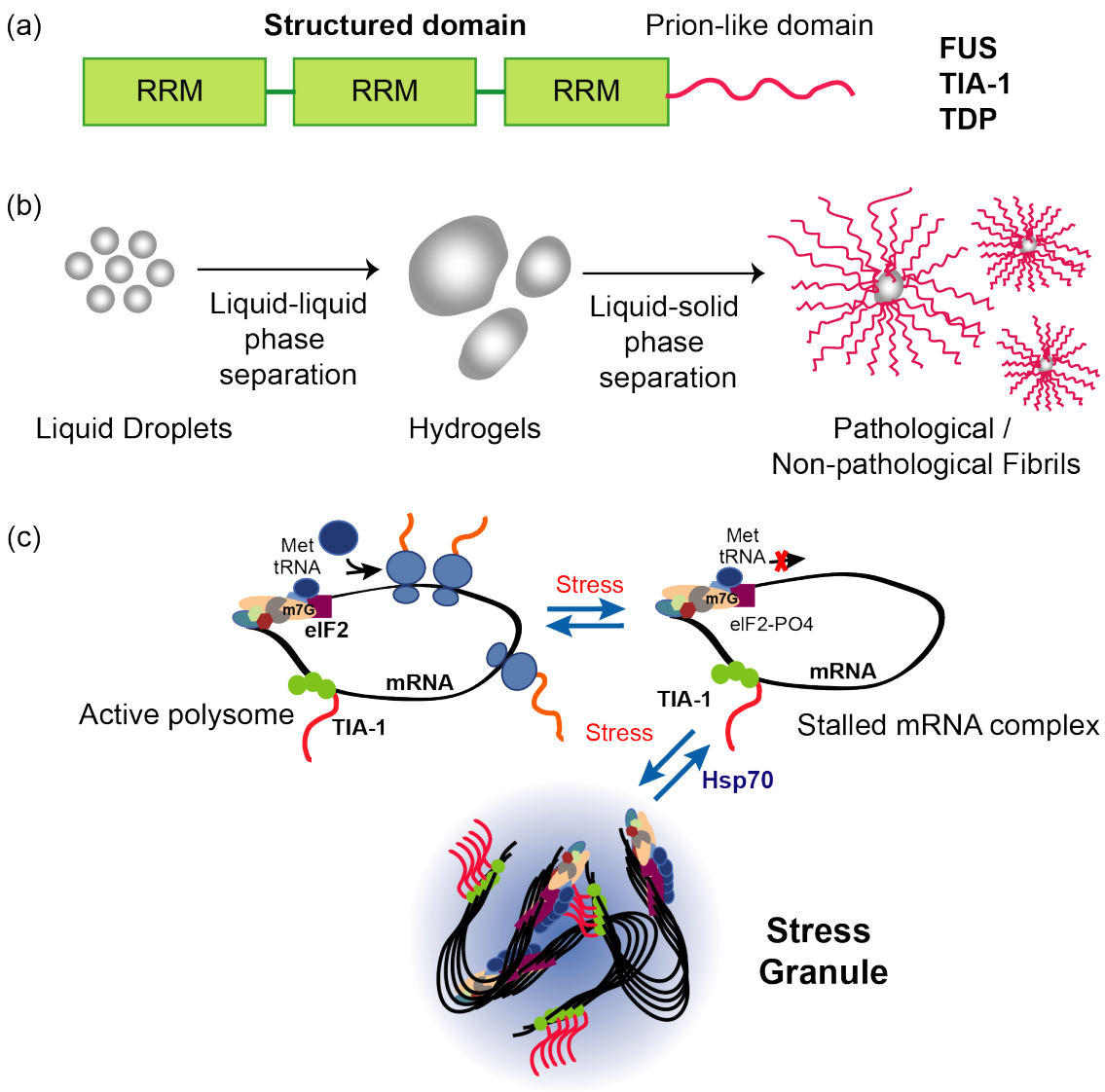


Figure 6: RNA binding proteins in phase separation and stress granules.

(a) RNA binding proteins composed of RRMs and low complexity domains or PRD domains. (b) Phase separation involves formation of liquid droplets followed by fusion of droplets to form gel state that is triggered by high local concentration of PRD domain proteins and further liquid-solid phase separation to form the end fibrils state[146]*. (c) Model of stress granules formation in TIA-1 adapted from Warris et al*[147]*. During stress, phosphorylation occurs at initiation factor eIF2 that leads to the stalling of mRNA and translation stops and concentration of cytoplasmic TIA-1 increases. TIA-1 PRD domain self-assembles to fibril formation and leads to granule formation along with other proteins.*

1.4.1 Phase separation

Phase separation is the phenomena commonly seen in RNA binding protein and in couple of years this topic has developed much interest among scientist is discussed in next section. RNA binding proteins exhibits two-step phase separation; first is liquid-liquid phase separation to form hydrogels and next step is liquid-solid phase separation to form fibrous state. **(Figure 6b)** This phase separation phenomena are observed in protein that exhibits low complexity domains such FUS, TIA-1, hnRNPs[145,146,148–150].

Several studies have been done to probe this interesting phenomenon during recent 10 years. In FUS, transportin act as physiological chaperone and reduces the phase separation and gelation process thereby controlling the protein synthesis. Arginine Hypo methylation promotes the formation of fibrous state that are rich in β-sheet content. Phase separation is dominated by cooperative cation π interaction in PRD domain and Arginine in structured C-terminal domain[151].

1.4.2 Stress granules (SGs)

Eukaryotic cells contain the assemblies of RNAs, and numerous proteins known as RNA granules or membrane-less organelles. These organelles include Stress granules, P-bodies, Para speckles, Stress granules as the name suggest, are formed by cells when they are subjected to environmental stress like heat shock, oxidative stress, viral infection. These granules composed of all initiation factors except eIF2, eIF5, RNA binding proteins, pool of RNA molecules and numerous proteins in the cytoplasm. Stress granule formation is the adaptive defensive action to prevent the cells from apoptosis. In stress, the pre-initiation complex for translation cannot form due to phosphorylation of initiation factor eIF2-PO4. This phosphorylation decreases the ternary complex eIF2-GTP-$tRNA^{Met}$ needed for initiating methionine onto 40s pre-initiation complex to start the translation. Thus, mRNA is sequestered in complex and the concentration of TIA-1 protein increases. This local high concentration of TIA-1 leads to self-aggregation to form fibrils in stress granules. However, this can be regulated by the chaperones HSP70 (Heat shock protein) by disaggregation with ATP dependance[147]. In the absence of stress, normal translation, formation of polysomes occurs.

TIA-1 is an important and known marker for stress granule formation. PRD domain is essential for this granule formation[152]. In TIA-1, *in-vitro* studies demonstrated that when subjected to arsenite stress, HSP70 and HSP40 prevents the formation of fibrils in polyglutamine fibrils[153,154]. **(Figure 6c)** Recently, it has been shown the relation of cytoplasmic aggregates and neurodegenerative diseases. It is suspected that these SGs might may act as seed to propagate the disease[155].

1.5 Nuclear Magnetic Resonance (NMR)

NMR spectroscopy is an essential technique for both chemist and biochemist in the structure determination of almost all the organic and inorganic or biological molecule. It has been more than half decade, apart from other structural tools like X-ray crystallography, NMR technique undergone many recent advances and emerges as an effective structural tool for determining atomic level of protein.

1.5.1 Nuclear spin states

Most atomic nuclei possess spin called as spin quantum number (I), which is the rotation of the nuclei around a given axis. Nuclei with ½ spin quantum number are the more frequently used NMR active nuclei in biomolecules as tabulated in **Table 5.** Although ^{1}H has most natural abundance as compared to ^{13}C and ^{15}N nuclei, the sensitivity of ^{13}C and ^{15}N can be solved by addition of isotopically labelled media during bacterial culture. As the nuclei are the charged particles it creates the magnetic moment (μ). Magnetic moment of the nuclei is directly proportional to the spin angular momentum (I) and the proportionality constant known as gyromagnetic ratio (γ) as shown in equation below[156,157].

$$\mu = \gamma I$$

$$\mu = \frac{\sqrt{\gamma h[I(I+1)]}}{2\pi}$$

where h is Planck's constant.

The magnetic moment of the nuclei splits into two orientations, α-lower energy and β-higher energy states in the presence of magnetic field B_o known as Zeeman splitting. **(Figure 7)** According to the Boltzmann distribution, the magnetic nuclei splits into 2I+1 energy level state[157,158].

Table 5 Nuclei properties important in NMR spectroscopy

Nucleus	Spin Quantum number I	Gyromagnetic ratio γ (10^7 rad T^{-1} s^{-1})	Natural Abundance (%)	NMR frequency (MHz) (B_0=2.3488 T)
^{1}H	1/2	26.7522	99.98	100.00
^{2}H	1	4.1066	0.0156	15.3506
^{13}C	1/2	6.7283	1.108	25.1450
^{14}N	1	1.9338	99.63	7.2263
^{15}N	1/2	-2.7126	0.365	10.1368
^{19}F	1/2	25.1815	100.0	94.0940
^{31}P	1/2	10.8394	100.0	40.4807

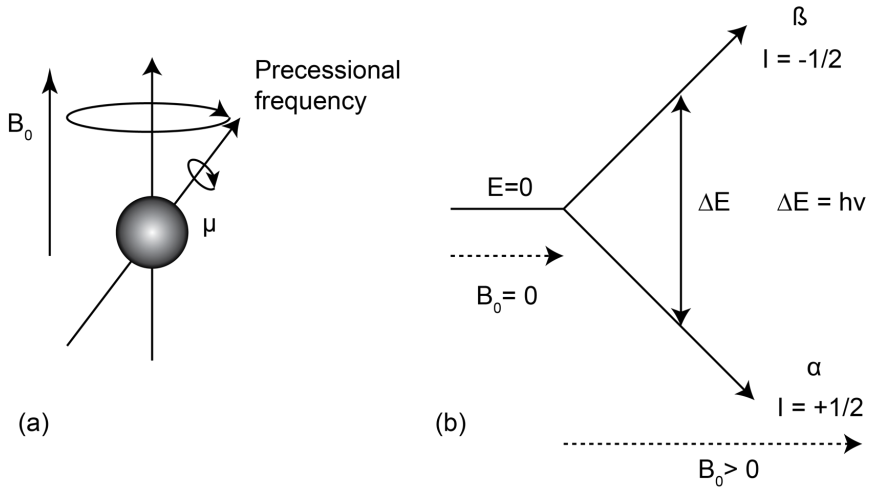


Figure 7: Spin magnet moment and energy level transitions[159]***.***

(a) Pictorial representation of spin magnetic moment μ precession along the applied magnetic field B_0 (b) The energy transition increases between two states increases with increasing applied magnetic field.

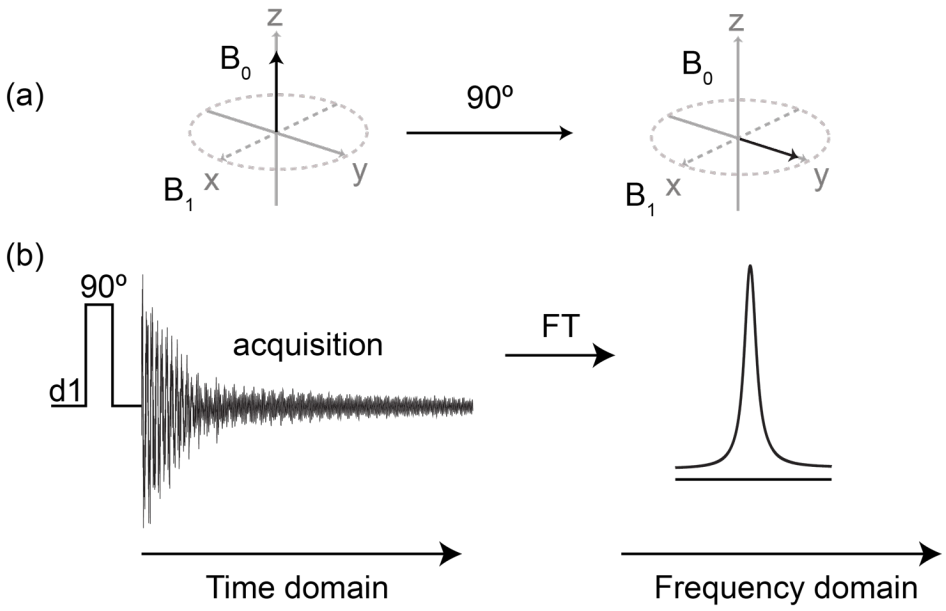


Figure 8: Pulse NMR

(a) Vector representation of Bulk NMR magnetization M_0, when placed in the magnetic field B_0 applied along z-axis before and after application of 90º pulse and B_1 is the RF field strength. (b) Simple pulse sequence with d1 delay and 90º pulse followed by acquisition that provides FID that is FT to NMR spectrum.

1.5.2 Pulsed field NMR

All nuclear spins get polarized under the influence of static magnetic field B_0, resulting in net macroscopic bulk magnetization M_0 along the direction of B_0 at z-axis. When we irradiate radiofrequency (rf) field B_1, M_0 that is at z-axis along the B_0, rotates to xy plane known as transverse magnetization. This magnetization precesses under the influence of B_0 at resonance frequency or Larmor frequency (υ) that induces the electric current in the detection coil. This component is measured as NMR signal. The 90º pulse in NMR means the time period for which the pulse must be applied to rotate the magnetization by 90º. When the magnetization goes back to thermal equilibrium z- direction, signal decays known as Free Induction decay (FID) i.e. time domain function and stored in the computer memory. This is converted by mathematical operation known as Fourier transformation (FT) to frequency domain conventional NMR spectrum[156,157]. **(Figure 8)**

1.5.3 Chemical shift

Chemical shift arises due to the small magnetic field at the nucleus by it surrounding electrons. When the small magnetic field generated is opposite to the direction B_0, the nucleus is said as shielded or upfield and if it is in same direction it is said as de-shielded or downfield. NMR frequency or Larmor frequency of the nucleus (υ) is determined by its gyromagnetic ratio and strength of the magnetic field B_0, which is defined by the equation as given below:

$$\upsilon = \frac{\gamma B_0(1-\sigma)}{2\pi}$$

Where σ is shielding constant. Chemical shift depends on the local electron environment of particular nuclei. Chemical shift δ is usually expressed in parts per million (ppm) by frequency

$$\delta = \frac{\upsilon_{sample} - \upsilon_{reference}}{\upsilon_{reference}} * 10^6$$

Factors influencing chemical shift are electron density, electronegativity of neighbouring groups and anisotropy induced magnetic field effects. For referencing the signal most commonly used is TMS.

1.5.4 Protein Nuclear Magnetic Spectroscopy

1.5.4.1 1D Proton experiments

^{1}H 1D experiments for protein gives the complex spectra in comparison to peptides but still it gives preliminary information about the protein state. In ^{1}H spectrum, at about 6-10 ppm contributes to amide peaks while from 1-5 ppm contributes to aliphatic peaks. Spread in amide region gives important information about the folding state of protein. If the peak dispersion is localized around 6-7 ppm the protein is said to be unfolded otherwise folded when it is fully dispersed is as explained in the **Figure 9.**

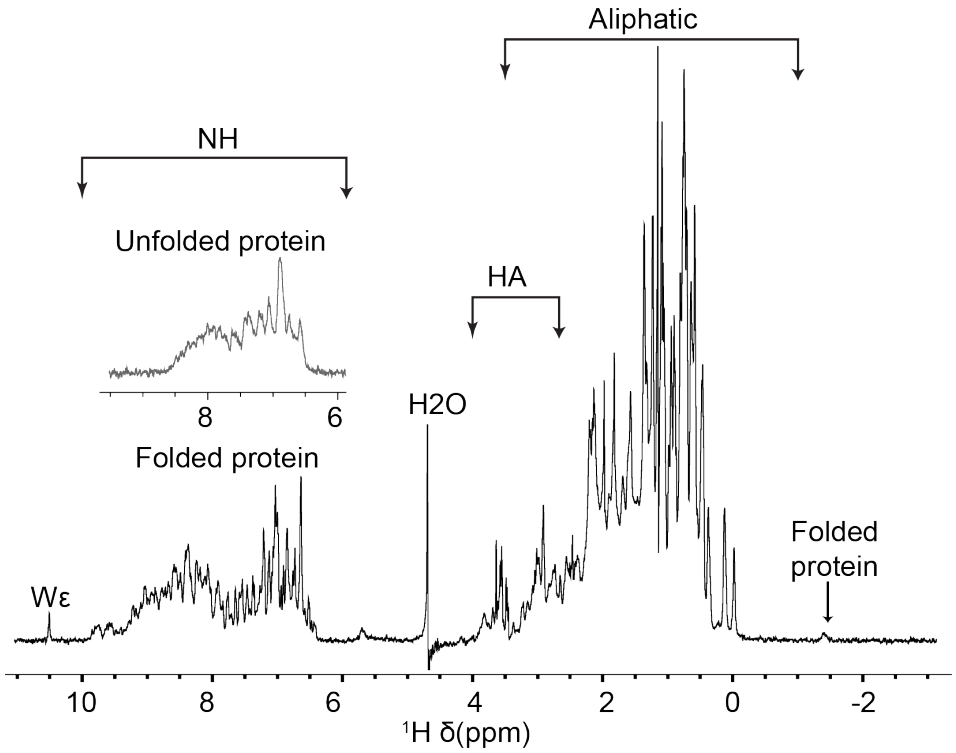


Figure 9: ^{1}H-1D spectrum of Antibody Light chain S3706_patdel protein.

All the important information of that can be derived from 1D protein spectra is depicted. In amide region the peaks dispersion gives the state of the protein whether it is folded or unfolded.

1.5.4.2 2D Heteronuclear single quantum coherence (HSQC)

2D ^{1}H -^{15}N HSQC is the most frequently used experiment in study of protein. Unlike 1D, which gives preliminary information of protein, 2D HSQC gives detailed information. Here, every peak in the spectra, which is the amide resonance, (NH) represents one residue. The only amino acid, which has no hydrogen attached to amide N, does not gives any peak in 2D ^{1}H-^{15}N HSQC is proline. In a fully folded protein, the peaks are sharp, intense and distributed over broad range of chemical while in unfolded protein the peaks are collapsed to narrow range of chemical shift[160]. In unfolded protein, although the dispersion is low, but the peaks can be narrow and sharp with same linewidth as in case of folded protein.

1.5.4.3 Backbone assignment strategies

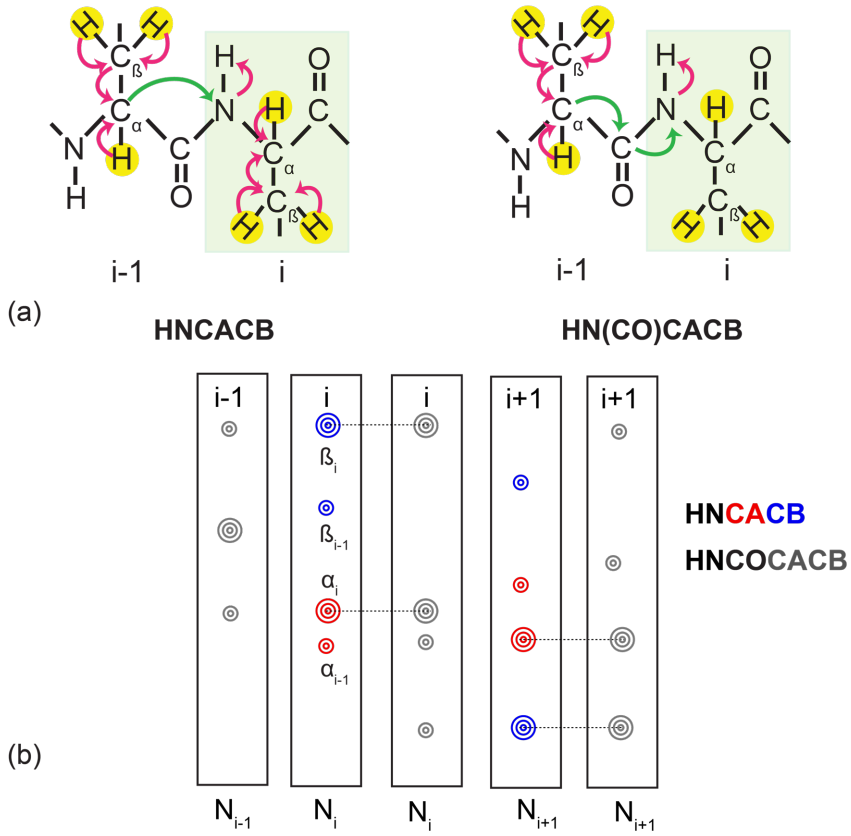


Figure 10: Solution state backbone assignment strategies.

(a) Transfer of magnetization in 3D HNCACB and 3D HNCOCACB resonance experiments[161]*. (b) Schematic representation for sequential connectivity's for assignment.*

Backbone assignment can be done by several ways either of the combination of triple resonance experiments like HNCACB / HN(CO)CACB or HNCA /HN(CO)CA or HNCO /HN(CA)CO[162]. In all these experiments, we can walk through the backbone by i to i-1. In HNCACB, Cα/ Cβ of the same residue appears as strong peaks while the preceding residue Cα/ Cβ appears as weak peaks. Moreover, the Cα/ Cβ peak phases are different as positive and negative respectively which makes these experiments distinct that helps in unambiguous assignments of protein backbone. The HN(CO)CACB is complementary experiment of HNCACB, where Cα/ Cβ strong peaks are from preceding residue and Cα/ Cβ weak peaks are from same residue[163]. **(Figure 10)** The Cα/ Cβ chemical shift gives preliminary information about the secondary structure element i.e. α-helix, β-sheets or random coil of the protein.

1.5.4.4 Chemical exchange

When there are two or more states, which are due to different chemical environment, gives rise to change in chemical shift, scalar coupling or relaxation rate. This phenomenon is called as chemical exchange. It gives the information on the conformational exchange or the reaction rates occurring in the biologically active NMR scale. The change on the NMR spectrum depends on the rate of the exchange process relative to the NMR time scale, which is of the order of milliseconds and signifies the time taken to record one NMR spectrum. If the average lifetime in any one state is longer than the NMR time scale, then distinct peaks for different environments will be observed in the spectrum while if it is short is gives an time average peak broad peak as depicted in **Figure 11**[164].

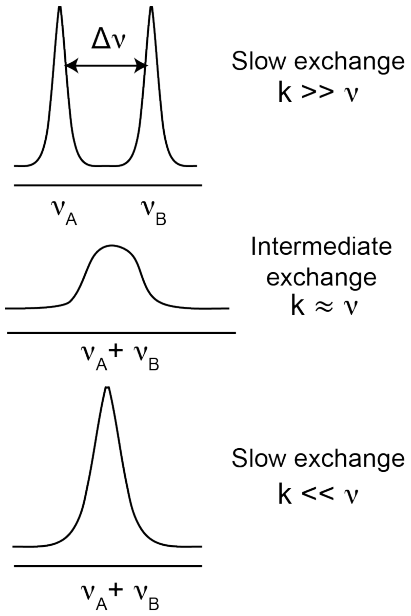


Figure 11: Chemical exchange on NMR time-scale[164]*.*
A and B are the two states, k is reaction rate while ν *is frequency.*

1.5.5 Solid state NMR (ssNMR) spectroscopy

Solution state NMR has limitations due to large protein size; it creates NMR line broadening and spectral overlap making the spectra inappropriate for analysis but in ssNMR due to the absence of tumbling in solids, there is no restriction to the molecular weight or size of the protein. In ssNMR with recent advancements high resolution spectra can be obtained and thus is now applied to biological samples like membranes proteins, large complexes, fibrils that are immobile or undergo slow motion[165–167]. Most of the experiments here are based on heteronuclear ^{13}C detected experiments rather than ^{1}H which is mostly used in solution state NMR. Linewidth and spectral overlap in ^{15}N dimension are the main concern in these experiments. With different labelling schemes 1,3-glycerol, 2-glycerol, NMR active

nuclei, spectral overlap can be reduced. This approach helps in structure calculation by getting distance restraints due to reduced dipolar interaction[168]. For the reasonable spectral dispersion in spectra selective labelling strategies is another way to reduce the NMR active spins and also reducing NMR overlap[169]. Per-deuteration, an alternative novel way allows to detect proton with high sensitivity and resolution[165,170]. More than 50 structure that involved crystalline model spectrin SH3[168], ubiquitin[171], GB1[172]; amyloid fibrils such as transthyretin[173], Aβ peptides[30,37,174,175], alpha synuclein[176], HET's[177]; membranes proteins such as Omp G[178], viruses such as M13 bacteriophage[179] and many other proteins are solved by this technique. Instead of scalar coupling in solution state NMR; here dipolar-coupling is the most significant parameter as there is a reduced dynamic among molecules as compared to solution-state NMR.

1.5.5.1 Dipolar Coupling

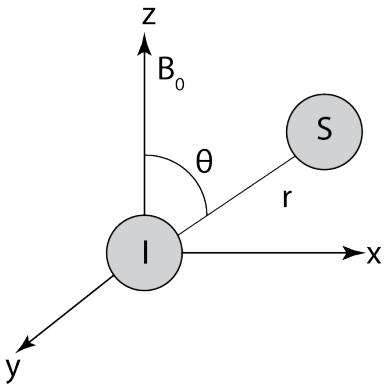


Figure 12: Vector model representing dipolar coupling.
I and S are the two nuclei separated by distance r. B_0 is the main magnetic field.

The interaction between two nuclear spin through space is called as Dipolar coupling or dipole-dipole coupling. The Hamiltonian operator for complete dipolar coupling and homonuclear (under secular approximation) and heteronuclear dipolar coupling is defined as follows:

$$\hat{\mathcal{H}}_D^{complete} = -d\left\{\frac{3}{r^2}\left(\hat{\vec{I}}.\vec{r}\right)\left(\hat{\vec{S}}.\vec{r}\right) - \hat{\vec{I}}.\hat{\vec{S}}\right\}$$

$$\hat{\mathcal{H}}_D^{homo} = -d\frac{1}{2}(3\,cos^2\theta - 1)\left[3\hat{I}_z\hat{S}_z - \hat{\vec{I}}.\hat{\vec{S}}\right]$$

$$\hat{\mathcal{H}}_D^{hetero} = -d(3\,cos^2\theta - 1)\hat{I}_z\hat{S}_z$$

$$d = \hbar\left(\frac{\mu_o}{4\pi}\right)\frac{1}{r^3}\gamma_I\gamma_S$$

While d is dipolar coupling, θ is the orientation of dipole with respect to external magnetic field. $\hat{I}_z\hat{S}_z$ are the spin operators. γ is gyromagnetic ratio of the nuclei. $\hbar$ is the

Planck constant. μ_o is the magnetic permeability and r is the distance between two nuclei.

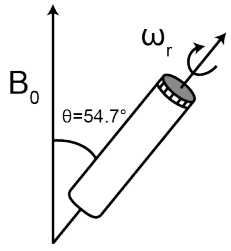
1.5.5.2 Magic angle Spinning (MAS)

In solution state NMR, the interactions are isotropic (orientation independent) while in ssNMR the interactions are anisotropic (orientation dependent) thus possess chemical shift anisotropy (CSA)[180]. The samples in ssNMR exhibits reduced molecular mobility or no motion leading to broad spectra. To reduce the effect of anisotropic interactions, the most robust technique is to rotate the sample at an angle of 54.7º with respect to external magnetic field $B_{0,}$ known as Magic angle spinning (MAS)[180]. **(Figure 13)** This spinning of sample averages out the anisotropic interaction, which leads to narrow line width in ssNMR. Typical Anisotropic interaction found in protein solid state NMR listed in **Table 6**.

Table 6 Chemical Anisotropic interaction in ssNMR found in proteins.

Spins	Type of interaction	
^{1}H-^{1}H	Dipolar (CH_3 group)	60 kHz
^{13}C-^{1}H	Directly bonded	23 kHz
^{15}N-^{1}H	Directly bonded	11 kHz
^{13}C-^{13}C	Directly bonded	3 kHz
^{13}C-^{15}N	Directly bonded	1 kHz

Typically, without magic angle spinning i.e. static field, ^{13}C spectra gives 'powder pake pattern' and as we increase the spinning speed, the spectra become more narrower with satellites side bands. Every rotor size has certain limitation in terms of sample amount and MAS. (**Table 7)**



$$D \alpha 1/ r^3 (3\cos^2\theta - 1)$$

Figure 13: Magic angle spinning (MAS)

where D is the dipolar coupling; θ is the angle between the magnetic field B_o and the vector connecting two spins

Table 7 Rotor parameters used in ssNMR

Rotor size	**0.7 mm**	**1.3 mm**	**1.9 mm**	**2.5 mm Thin wall**	**3.2 mm Thick wall**	**3.2 mm Thin wall**	**4 mm**
Sample volume	0.59 μL	2.5 μL	13.1 μL	12 μL	32.1 μL	46.7 μL	80 μL
ω_r	111 kHz	67 kHz	42 kHz	35 kHz	24 kHz	24 kHz	15 kHz

Where, ω_r is the rotational frequency.

1.5.5.3 Cross polarization (CP)

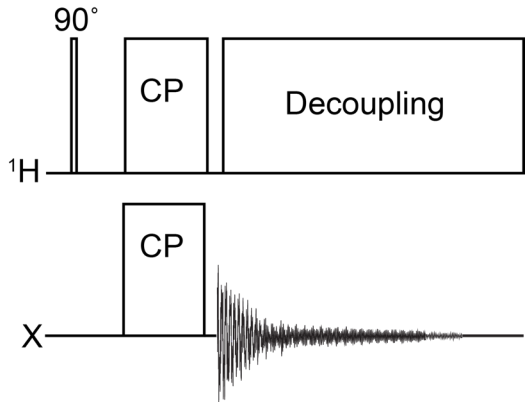


Figure 14: CP Pulse sequence

^{13}C and ^{15}N nuclei have low gyromagnetic ratio therefore gives low signal enhancement, which can be overcome by the polarization transfer from ^{1}H nuclei (abundant spin). In ssNMR, polarization transfer is employed by the heteronuclear dipolar interactions to neighbouring spins by CP as described in **Figure 14.** For efficient magnetic transfer it must fulfil the Hartmann-Hahn condition which is given by the equation below[181]:

$$\gamma_I \omega_I = \gamma_S \omega_S \pm n\omega_r$$

where I and S are the two spins, γ the gyromagnetic ratio, ω_I and ω_S is the spin lock field of the respective spins and ω_r is the rotational frequency.

1.5.5.4 Assignment experiments in solid-state NMR

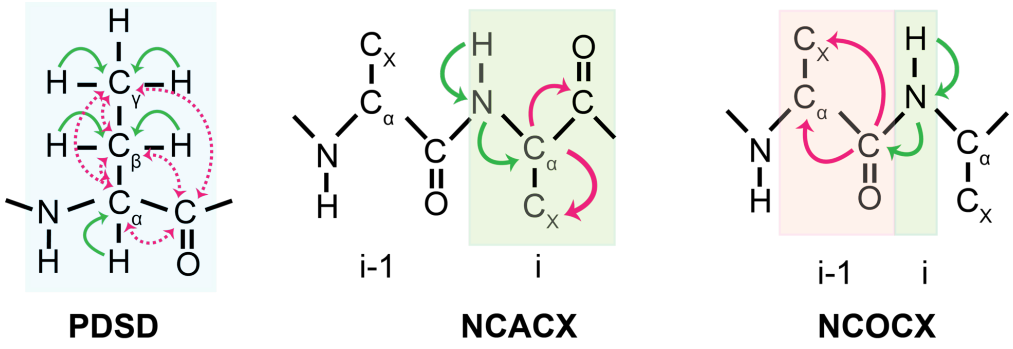


Figure 15: Assignment experiments in ssNMR.

Schematic representation of transfer of magnetization in commonly used experiments in ssNMR, PDSD, NCACX and NCOCX that aid in assignment of protein[182].

For assignment in uniformly labelled sample, C-detected experiments, 2D PDSD (proton driven spin diffusion) or 2D DARR (Dipolar assisted rotational resonance), 3D NCACX and 3D NCOCX are used. 2D PDSD or DARR gives the ^{13}C-^{13}C correlation which is used as the signature of protein similarly like ^{1}H-^{15}N HSQC in solution state NMR. This also gives information about the quality of the sample. In PDSD, the transfer of magnetisation is from ^{1}H (high sensitivity) to ^{13}C (low sensitivity) nuclei and then to neighbouring nuclei depending on the mixing time.

For the sequential assignment of residues in PDSD spectra, NCACX and NCOCX experiments are used as sequential connectivity's. In NCACX and NCOCX, the magnetisation is transferred from ^{1}H to $^{13}C\alpha$ or ^{13}C-CO using cross polarisation followed by PDSD or DARR step to transfer magnetisation further to the neighbouring ^{13}C nuclei. The chemical shift evolution is on ^{15}N and ^{13}C nuclei and finally detection in ^{13}C nuclei[182]. **(Figure 15)** Moreover, 3D CONCA and 3D CANCO helps in further confirming the assignment and eliminating the ambiguity[183]. Depending on the mixing time used in the experiment nearby carbon atoms or long range connectivity's can be observed for the restraints for the structure calculations[166].

1.5.5.5 Rotational-echo double resonance (REDOR)

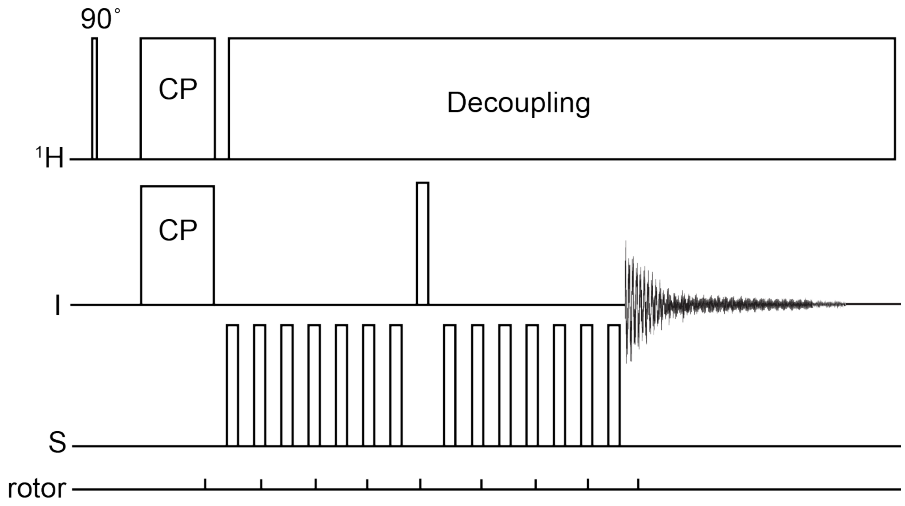


Figure 16**: **REDOR pulse sequence

Thin and thick rectangle represents 90° and 180° pulse respectively. CP is cross polarization transfer.

Rotational-echo double resonance (REDOR) is mainly based on the dephasing of transverse, S-spin magnetization by train of rotor-synchronized π-pulses. **(Figure 16)** It has mostly been used for dipolar coupling measurement between two pairs[184]. This is one of the recoupling schemes where dipolar interaction is reintroduced. It gives important information about the distance restraints. With magic angle, dipolar coupling is given by the equation as given below:

$$d = \frac{\mu_0 \gamma_I \gamma_S h}{4\pi r^3}$$

where μ_0 is the permeability of the free space and r is the inter-nuclear separation. *I* and *S* are the two spins.

1.6 Overview of the Study

This thesis, comprise of two amyloidogenic proteins studies, using solution and solid-state NMR as a main technique. First protein is antibody light chain V_L domain, which is involved in AL amyloidosis. Second protein is TIA-1 primarily Q-rich domain that is involved in stress granule formation.

Next section includes *Materials and Method section* followed by R*esults and Discussion* that includes three chapters. Every chapter starts with aim followed by results, discussion and conclusion. Here is the little overview of the chapters:

- ***Solution state NMR studies of S3706 V_L domain.***

In this chapter, there are two projects that are discussed. First - the aggregation studies in antibody light chain variable domain. Here, we studied the dimer interface and intermediates formed during aggregation process in S3706 variants using solution state NMR and other biophysical techniques. Second – EGCG binding with S3706 V_L variants using solution state NMR.

- ***Solid state NMR investigation of S3706 V_L domain***

This chapter includes the study of S3706 V_L fibrils. We analysed the influence of seeding, mutational effects and assignment that might be important in terms of structural aspect for fibril stability.

- ***Interaction of Q-rich domain with RRMs in TIA-1***

Here we focus on Q-rich domain that has low complexity region and involved in stress granule formation. We study how the Q-rich domain affects the RRM domains using MAS solid state NMR as a main technique.

MATERIALS
&
METHODS

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were supplied by Carl Roth (Karlsruhe, Germany), SERVA (Heidelberg, Germany), Roche Diagnostics (Unterhaching, Germany), Sigma-Aldrich Chemie (Taufkirchen, Germany) and VWR International Company. Epigallocatechin gallate (Sunphenon EGCG) was a kind gift by Stefan Schönland from Heidelberg amyloidosis center (Germany), supplied by Taiyo (Yokkaichi, Japan). $^{15}NH_4Cl$, U-^{13}C D-glucose and D_2O were purchased from Cambridge Isotope Laboratories (Tewksbury, USA) and Eurisotop (Saint-Aubin, France).

2.1.2 Devices

Devices	Supplier
Autoclave Varioclav EP-Z	H+P
Cell Disruption Apparatus Basic Z	Constant systems
Centrifuges	
Avanti J25 and J26 XP	Beckman coulter
Optima XL-A (equipped with FDS)	Beckman coulter (Aviv)
Optima XL-I	Beckman coulter
Rotina 46R	Hettich
Universal R	Hettich
Table top centrifuge 5414 C	Eppendorf
Chromatography systems	
AKTA FPLC	GE Healthcare
AKTA prime	GE Healthcare
DynaPro NanoStar	Wyatt Technology
Fluorescence Spectrometer	Horiba Jobin Yvon
FluoroMax-4	
Frac -950 fraction collector	GE Healthcare
Super loops	GE Healthcare
Eppendorf – thermomixer	Eppendorf
Gel documentation System Biotec II	Biometra
Gel electrophoresis systems Julabo	Serva
Homogeniser Ultra Turrax DIAX900	Heidolph
HPLC systems	GE healthcare
Ice maker	Zieger
Incubator	Thermo scientific
Magnetic stirrer Heidolph MR200	Heidolph
Nuclear Magnetic Resonance	
Solution state & solid-state	Bruker Avance
pH meter	WTW
Electronic balance	
BP 121 S	Sarotius

BL 310	Sarotius
Thermoblock TB	Biometra
Shaker MAX 800	Thermo scientific
Ultra-Filtration cell 8050	Amicon
UV-Vis spectrophotometers	
Nanodrop	Peqlab
Novaspec II	GE Healthcare
Vortex MS2	IKA
Water bath F6-K	Haake

2.1.3 Chromatography Columns

Columns	Company
Desalting column	GE Healthcare
Ni-NTA resin	Qiagen
Q Sepharose Fast Flow	GE Healthcare
Superdex 75 Prep Grade	GE Healthcare
Superdex 75 300/10 Prep Grade	GE Healthcare

2.1.4 Bacterial strain

The pET28b+ vector and *E. coli* BL21 cells was supplied by Novagen (Merck, Darmstadt, Germany). *E. Coli XL-1* blue cells were purchased from Stratagene (Agilent Technologies Deutschland GmbH, Waldbronn, Germany).

2.1.5 Enzymes, Standards and kits

The enzymes and dyes were purchased either from NEB or Promega. The DNA ladder was provided by Serva. For the DNA purification we used Wizard plus Promega supplied SV minipreps.

2.1.6 Buffers and solutions

Table 8 Media and antibiotics used for cultivation of E. coli

Media	Chemical compound	Concentration
LB medium	LB	20 g/l
SOB (super optimal broth) pH 7	Tyrptone	20 g/l
	Yeast extract	5 g/l
	NaCl	10 g/l
	KCl	250mM
	NaOH	5 M

	$MgCl_2$	2 M
M9 (10X)	Na_2HPO_4	60 g
	KH_2PO_4	30 g
	NaCl	5 g
Trace elements (100 X)	EDTA pH 6.5	5 g
	$FeCl_3.6H_2O$	0.83 g
	$ZnCl_2$	84 mg
	$CuCl_2.6H_2O$	
	$CoCl_2.6H_2O$	13 mg
	H_3BO_4	10 mg
	$MnCl_2.6H_2O$	1.6 mg
Minimal Media	M9	100 ml
	Trace elements	10 ml
	$MgSO_4$	1 M
	$CaCl_2$	1 M
	Thiamin- HCl	1 mg/ml
	Biotin	0.1 mg/ml
	Glucose	0.2 %
	$^{15}NH_4Cl$	0.05 %
Antibiotics	Kanamycin	1 mM

All media was sterilized in an autoclave at 121 °C for 20 min. Antibiotics stocks were passed through a sterile filter (0.22 µm) and stored at -20 °C

2.1.7 Software's

Software	Provider
Topspin 3.5 (NUS plugin)	Bruker
Mendeley Desktop 1.19.2	Glyph & Cog, LLC
CcpNMR 2.4.2	Plone and python
Chimera 1.8.1	UCSF
MATLAB 8.4 R2016b	MathWorks
ApE- A plasmid editor	Equi4 software

2.1.8 Databases

PDB	www.rscb.org/pdb
PubMed	www.ncbi.nlm.ni.gov/pubmed
Abysis	http://www.bioinf.org.uk/abysis3.1/
Protoparam	https://web.expasy.org/protparam/
BMRB	http://www.bmrb.wisc.edu/ref_info/statful.htm

2.2 Methods

2.2.1 Molecular Biology

2.2.1.1 Plasmid preparation

DNA plasmids were purified from E. coli XL Blue-1 overnight cultures using the instructions given by the manufacturer in Wizard Plus SV Miniprep kit (Promega Gmbh, Mannheim, Germany). The DNA was sequenced at GATC Biotech AG (Kontanz, Germany) and stored at -20 °C. The concentration of pure plasmid was determined by measuring the UV absorbance at 260 nm.

2.2.1.2 Site directed mutagenesis

Point or fragment mutation was done using polymerase chain reaction (PCR). 20 base pairs upside and downside from the site of mutation were taken as primers. Buffer used for PCR is given in table. The codon used in mutation was considered according to the *E. coli.* The protocol was adapted from Quick-change site – directed mutagenesis kit from stratagene, which is tabulated in **Table 9**.

Table 9 PCR reaction

Reagent	Volume
10x Pfu Ultra reaction buffer	5 μl
Template plasmid	10 ng
Forward primer	2.5
Reverse primer	2.5
Pfu Ultra Polymerase	1 μl
Water	Fill up to 50 μl

The PCR cycle is as follows:

95 °C	5 min	
94 °C	1 min	25X
56 °C	1 min	
68 °C	2 mins/ kbp of template plasmid	
68 °C	10 mins	
4 °C	storage	

The template was digested by adding 1 µl DpnI and incubating for 2 hr at 37 °C after the PCR cycle. The mutants were checked by transformation and sequencing the plasmids.

2.2.1.3 Transformation

The plasmid was transformed into BL21 competent cells. *E. coli*. BL21 cells were thawed on ice and 2 µl of plasmid (50-120 ng /µl) was added, gently mixed and kept on ice for 30 min. This was followed by a heat shock for 30 s at 42°C and subsequent cooling on ice for 10 min. 200 µl of LB was added to cells and incubated for 1 hr at 37°C at 1000 rpm. 50 µl of the suspension was spread evenly on kanamycin-containing LB-agar plates and incubated overnight at 37°C.

2.2.1.4 Protein expression and purification of V_L proteins

V_L proteins were recombinantly expressed in BL21 *E. coli* strain. Few colonies from the LB agar plates of respective protein were inoculated in the pre-culture containing 20 ml of LB and grown overnight at 37°C. Next day, Isotopic labelled protein was expressed in M9 minimal media with kanamycin and $^{15}NH_4Cl$ and ^{13}C-Glucose as nitrogen ad carbon source respectively. The overnight cells were re-suspended in M9 media and transferred to 1-litre culture. For large LB culture, 10 ml of pre-culture was added in 1-litre of LB culture. Cells were grown at 37 °C until OD_{600} reached to 0.6 followed by protein expression induction by 1 mM IPTG addition. The cells were continued to grow at 37 °C for overnight. Next day, the pellets were obtained by centrifugation of cells at 6000 g for 20 min. The supernatant was discarded, and the cells were re-suspended in 50 ml re-suspension buffer containing 1 tablet of EDTA-free protease inhibitor followed by ultra-sonication for lysis. 1 mg/ml DNase was added to the lysate followed by centrifugation at 24000 g at 4 °C for 1 hr to get inclusion bodies (IBs). IBs pellets obtained can be stored at -20 °C until further use.

For purification, IBs were dissolved in 30 ml dissolving buffer for 3 hr at RT with shaking for solubilisation and then centrifuged for 20 mins at 20000 g. The solubilized IBS were subjected to anion exchange chromatography to remove impurities containing low salt buffer using 5 ml Q-Sepharose column. Protein did not bind to the column and comes in flow through. For proper folding, the protein was dialyzed at 4°C overnight against refolding buffer containing redox agents using dialysis tube with MWCO 3.5 kDa. Next day, the refolded protein was concentrated up to 10 ml for size exclusion chromatography, using 120 ml Superdex 75 column. Pure fractions were pooled, and yield was determined by using molar extinction coefficient and molecular weight at 280 nm for S3706 patient sequence. For all construct the values were used according to Protoparam results[185]. The purification protocol was followed that was published previously[85,186]. All the buffers required during purification are written in **Table 10**.

Table 10 Buffer used in expression and purification of V_L protein

Buffer	Reagent	Concentration
Resuspension buffer	Tris	50 mM
	NaCl	10 mM
	EDTA	10 mM
Dissolving buffer (Lysis)	Tris	50 mM
	EDTA	5 mM
	Urea	8 mM
	β-ME	1%
Low salt buffer	Tris	25 mM
(pH 8 at RT)	EDTA	5 mM
	Urea	8 M
Refolding buffer	Tris	250 mM
(pH 8 at 4°C)	EDTA	5 mM
	L-Arginine	100 mM
Freshly prepared	GSSG	1 mM
Freshly prepared	GSH	0.5 mM
Phosphate buffer	[NaH_2PO_4+Na_2HPO_4]	20 mM
(V_L protein) pH 6.5 at RT	NaCl	50 mM

2.2.1.5 Protein expression and purification of TIA-1 proteins

The expression and purification of TIA-1 protein were followed according to published protocol[187]. TIA-1 proteins and its construct were expressed recombinantly in BL21 *E. coli.* Strain. Few colonies from the LB agar plates of respective protein were inoculated in the pre-culture containing 20 ml of LB and grown overnight at 37°C. Next day, Isotopic labelled protein was expressed in M9 minimal media with kanamycin and $^{15}NH_4Cl$ and ^{13}C-Glucose as nitrogen and carbon source respectively. The overnight cells were re-suspended in M9 media and transferred to 1L culture. For large LB culture, 10 ml of pre-culture was added in 1-litre of LB culture. Cells were grown at 37°C until OD_{600} reached to 0.6 followed by expression induction by 0.5mM IPTG addition. The cells were continued to grow at 18°C for overnight. Next day, cells were pelleted by centrifugation at 6000 g for 20 min. The cells can be re-suspended in re-suspension buffer and stored at -20°C until further use.

For purification, the suspended pellets were disrupted by ultra-sonication using 30 ml of cell lysis buffer on ice. The cell walls disrupted, and cell debris was separated from cell lysate by centrifugation at 24000 g for 1 hr at 4°C. The supernatant was loaded in 4 ml Ni column that was pre-equilibrated. The column was washed with approx. 100 ml using same buffer and protein was eluted using elution buffer. To cleave off the expression tag, 2 mg/ml

of TEV protease was added and dialyzed against TEV cleavage buffer at 4°C for overnight using respective MWCO dialysis tube (MWCO= 3.5 kDa for TIA_1 QRD44 and MWCO= 10,000 kDa for Full TIA-1 constructs). Next day, cleaved protein was loaded to 2nd Ni column to remove expression tag protein. The flow through contains the desired protein, which was concentrated using respective amicons tubes (MWCO= 3.5 kDa for TIA-1 QRD44 and MWCO=30,000 kDa for Full TIA-1 constructs). The protein was loaded in Superdex 75 column for size exclusion chromatography, which was pre-equilibrated with gel filtration buffer. The protein fraction was collected, and yield was determined using extinction coefficient and mol. wt. values obtained by protoparam results. The buffers used in the procedure are shown in **Table 11**.

Table 11 Buffer used in expression and purification of TIA-1 proteins

Buffer	Reagent	Concentration
Resuspension buffer (pH 8)	[NaH2PO4+Na2HPO4]	50 mM
	NaCl	300 mM
Cell Lysis buffer	Tris pH 7.5	20 mM
	NaCl	500 mM
	β-ME	1 mM
	Imidazole	10 mM
	Urea (only in case of full TIA-1)	500 mM
Ni column equilibration buffer	Tris pH 7.5	20 mM
	NaCl	500 mM
	β-ME	1 mM
	Imidazole	10 mM
	Urea (only in case of full TIA-1)	500 mM
Ni column elution buffer	Tris pH 7.5	20 mM
	NaCl	500 mM
	β-ME	1 mM
	Imidazole	300-500 mM
	Urea (only in case of full TIA-1)	500 mM
TEV cleavage buffer	Tris	20 mM
	NaCl	150 mM
	β-ME	1 mM
Gel filtration buffer (pH 7.4)	KH2PO4 + K2 HPO4	50 mM
	NaCl	200 mM
	DTT	1 mM

2.2.2 Techniques used for Protein Characterization

2.2.2.1 SDS PAGE

Through SDS PAGE, desired protein purity and degradation was tested. The tricine SDS PAGE was used for low molecular weight protein, The protocol was followed as described by Schagger[188]. For medium and high molecular weight protein sample, normal SDS PAGE was used. For every step of expression and purification, sample was removed and dissolved in sample loading buffer (5X). The samples were incubated at 95°C before putting the gel. The Ultra-low Range Molecular Weight Marker (Sigma-Aldrich) was loaded onto the gels as a reference for molecular weight. The buffer and solution used is described in the **Table 12**.

2.2.2.2 Chemical cross linking

To examine the dimer state, glutaraldehyde was used as chemical cross-linker. Protein was incubated with 2-3% glutaraldehyde was added to 80 µM of S3706_GL and S3706_patdel protein in phosphate buffer for 60 mins. Reaction was stopped by addition of 10µL of Tris–HCl, pH 8 at interval of 10, 20, 30 mins. Cross-linked protein was detected for each interval through SDS–PAGE by addition of non-reducing loading buffer.

2.2.2.3 Dynamic light scattering (DLS)

To detect the macroscopic state of the protein, DLS was run. 50µl of protein sample with 50-100 µM concentrations was used ad subjected to DynaPro NanoStar instrument. Three measurements consisting of 10 acquisitions each were carried out at 25°C. Acquisition times of 60s were used. The data was analysed using DYNAMICS V7 software.

2.2.2.4 Thioflavin T fluorescence

Fibrils formation kinetics was monitored by standard ThT assay[189]. All the samples were done in triplicates. In all the samples 0.02% of sodium azide was added with total 250 µL reaction volumes to avoid bacterial infection. The measurement was done at 440 nm excitation and 480 nm emission wavelengths.

For V_L samples, 50 µM protein with seeds (2.5%) and without seeds was incubated with 25 µM ThT dye in 96 well plates and measured each day in fluorescence microscope with excitation of 440 nm and emission of 480 nm wavelength at 37°C. During incubation, the samples were agitated at 500 rpm at thermoshaker. For TIA-1 proteins, ThT assay was done at 10 µM and 50 µM protein concentration with 25 µM ThT dye with seeds and without seeds. Using 1:1 ratio, the interaction of RNA and protein was monitored. The sample was incubated at 37°C under continuous orbital shaking (350 rpm).

Table 12 Gel electrophoresis

Buffer/Solution	Reagent	Concentration/Composition
Loading buffer	Tris-HCl	500 mM
	DTT	400 mM
	SDS	8%
	Bromophenol Blue	0.4%
	Glycerol	40%
Anode buffer (pH 8.9)	Tris-HCl	1 M
Cathode buffer (pH 8.25)	Tris	1 M
	Tricine	1 M
	SDS	1%
Laemmili buffer (pH 8.45)	Tris	3 M
	HCl (for pH adjusting)	1 M
	SDS	3%
Coomassie staining	Coomassie	0.025%
	Acetic acid	10%
	Water	
Fixing solution	Ethanol	50%
	Acetic acid	10%
De-staining solution	Ethanol	50%
	Acetic acid	10%

2.2.2.5 Transmission electron Microscopy (TEM)

In order to scan the state for protein as monomers, oligomers or fibrils negative stain, TEM was used. Formvar/Carbon 300 mesh copper coated carbon grids (Electron Microscopy Sciences) was kept the in-Argon atmosphere for 10 sec to make it hydrophilic. Grids were prepared by incubating 5 µL protein sample for a minute followed by washing and incubating with 10 µl of uranyl acetate (2%) for staining up to 30 sec. Extra stain was removed from the grid using filter paper. Grids were visualized in TEM employing a Zeiss EM 10 CR (Zeiss, Germany).

2.2.2.6 Preparation of fibrils for ssNMR Experiments

For all S3706 samples, fibrils were prepared in 50 µM protein in PBS buffer at 37°C with constant shaking at 120 rpm in fibril shaker. 2.5-5% seeds was used for seeded samples for homogeneous fibril preparation. For all TIA-1 fibrils, the same condition were as stated for S3706 unless concentration was stated, in PBS buffer at 37°C. All the fibrils are

prepared at physiological pH. The time for fibril incubation was 1 week to 2 weeks depending upon seeded or non-seeded samples.

2.3 NMR spectrometry

2.3.1 Solution state NMR

2.3.1.1 Sample Preparation

All the sample of solution state NMR was measured in shigemi tube (250 µL) or standard NMR tubes (500 µL). 10% D_2O was used in all the samples for locking. All protein samples for NMR contains phosphate buffer at pH 6.5. All the experiments are executed in either 600 MHZ or 500MHz NMR Avance triple resonance cryo/RT probe spectrometer. All the spectra were processed in TOPSPIN software and analysed using CcpNMR software.

2.3.1.2 Backbone Assignment of V_L protein

For backbone assignment of S3706_pat, S3706_patdel and S3706_GL, triple resonance experiment 3D HNCACB, 3D HNCOCACB and 3D HNCA were acquired. All the experiment was done at 298 K of about 1 mM protein concentration. The assignment was done using CcpNMR software.

2.3.1.3 Titration and other NMR studies of VL proteins

For concentration dependent assay, higher to lower concentration of S3706-patdel and S3706-GL $^{1}H^{15}N$-HSQC were recorded with same acquired NMR parameters. For time dependent aggregation analysis, 50 µM proteins was incubated in thermal shaker at 298K with shaking at 250rpm followed by acquiring $^{1}H^{15}N$-HSQC each day for 7-8 days.

For EGCG titration, 20 mM EGCG stock sample was prepared freshly. 0.5X to 20X molar excess was added to 50 µM protein sample and ^{15}N HSQC was recorded at 298 K. For time dependent titration, 10X fold molar excess of EGCG was added to 50 µM protein sample. Chemical shift difference was calculated by equation given below:

$$\Delta\delta_{NH} = \sqrt{(\Delta\delta^{1H})^2 + \frac{1}{25}(\Delta\delta^{15N})^2}$$

2.3.2 Solid state NMR

Using Potassium bromide (KBr) powder sample employed the magic angle setting calibration. ^{13}C and ^{15}N channel referencing was done by using MLF or Adamantane standard sample[190,191]. The chemical shift for referencing by all measurements was done in triple resonance 3.2 mm MAS probe 750 MHz spectrometer.

2.3.2.1 Sample preparation

For solid state NMR sample approx. 15 mg with spacer (teflon) and approx. 22 mg protein sample without spacer were packed in 3.2 mm rotor thin wall/ thick wall ZrO_2 (Bruker, Biospin). Protein aggregates were first centrifuged to approx. 500 µL and sediment into the Bruker NMR rotor using rotor filling tool (Giotto Biotech) in L-100 XP ultracentrifuge (Beckman Coulter) equipped with an SW 32 Ti swinging bucket for 1hr at 28000 rpm. Same procedure was also used filling 1.3 mm rotor using respective amount.

2.3.2.2 Assignment experiments

For fibrils assignment, ^{13}C-^{13}C correlation experiments were acquired either as 2D PDSD or 2D DARR at 10 kHz with 50 ms and 20 kHz with 100ms respectively. Due to spinning side bands around aromatic regions at 10 kHz, MAS was adjusted to 16.5 kHz for aromatic residues. Conventional 3D NCACX and 3D NCOCX were recorded to get sequential connectivity's. In ^{13}C-^{15}N correlations, specific CP based experiments were used to transfer coherence. In all the experiments, 70-80 kHz SPINAL decoupling during acquisition at ^{1}H while in mixing steps 70-100 kHz continuous wave decoupling was used. In addition, 3D CONCA experiment was recorded to confirm and assign ambiguous residues. All the experiments were recorded at 10 kHz MAS or 20k Hz MAS at 0°C.

25% NUS was used in few 3D experiments to gain sensitivity in spectra and to reduce experimental time[192]. Using mdd algorithm in TOPSPIN software with NUS plugin, NUS spectra was processed.

2.3.2.3 2D TEDOR experiments

In order to examine salt bridge, ^{13}C-^{15}N TEDOR experiments were recorded at 16.5 kHz MAS in 750 MHz with short and long mixing in all S3706 variants samples. Magnetization was transferred from highly abundant nuclei ^{1}H to ^{13}C by ramp (90-100%) CP followed by two REDOR blocks on ^{15}N to reintroduce ^{13}C-^{15}N dipolar couplings through rotor-synchronized π-pulses. The first block transfer magnetization from ^{13}C to ^{15}N followed by t1 and the second block transfers the magnetization back to ^{13}C for detection. For short and long mixing 1.9 ms and 15 ms mixing time was used during the experiment respectively.

RESULTS

AND

DISCUSSION

3 Solution state NMR Studies of antibody S3706 V_L domain

3.1 Aggregation studies in S3706 V_L domain

3.1.1 Aim of the project

There are several factors that trigger the monomeric protein into fibrils state to cause AL amyloidosis. The mechanism of how from monomeric soluble folded state transforms into insoluble fibrous state is still not well understood. In regard to this question, our aim is to understand the underlying mechanism of the aggregation process and difference between the patient and germline proteins. So, the first step is the resonance backbone assignment of patient and germline protein using solution state NMR. We characterize the dimer interface, which is common feature exhibit by all V_L protein regardless of kappa or lambda subtype. Furthermore, we characterize the intermediate states that are considered to be the crucial state in AL amyloidosis and analyzed how does it differ from its germline protein and its single point mutant. The main technique we used for characterization of different states of protein is solution state NMR along with some complimentary biophysical methods like TEM, ThT and DLS.

3.1.2 Protein construct and background

S3706 V_L is the light chain variable domain (V_L) of λIII subgroup. It is a specific AL amyloidosis patient derived sequence (heart muscle) and the crystal structure of this protein as dimer already published[193]. In our study, this sequence is designated as S3706_pat. Most of the studies are done in protein S3706_patdel in which we added 2 amino at N-terminal and deleted seven amino acids that does not belong to the variable domain. **(Figure 17)**

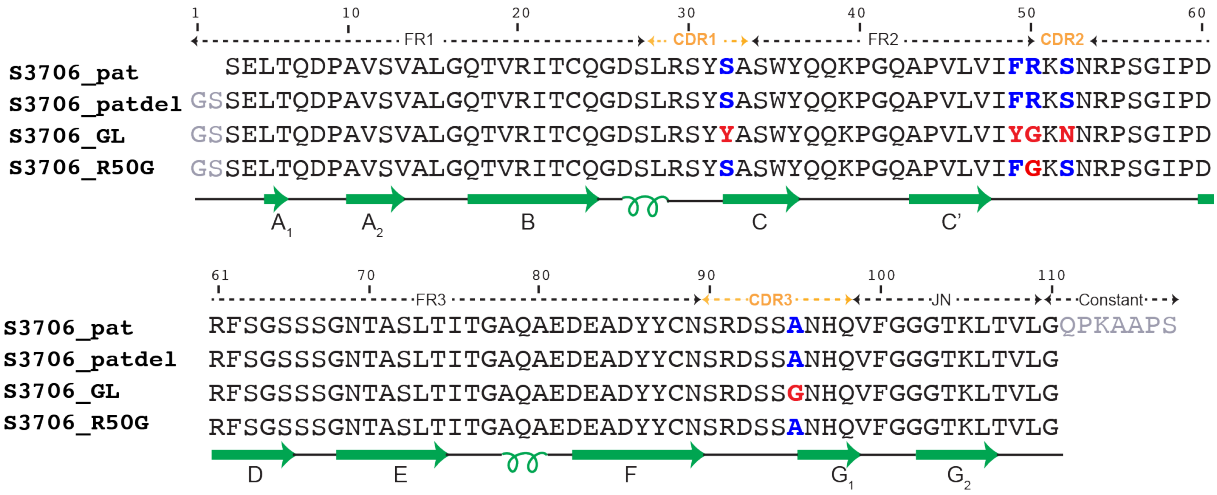


Figure 17: S3706 V_L constructs

The protein possesses the common Ig fold consisting of anti-parallel 4+5 β-strands connected by loops as depicted in **Figure 18a**. Like many other V_L sequences a conserved pair of cysteine (C23-C88) exist, connecting B-strands and F-strand. To stabilise the fold, salt bridges from R61 to E81 and D81 are observed from the crystal structure (PDB: 5L6Q). The C-terminal from constant domain did not show up in the crystal structure. In crystal structure, the protein is in canonical dimer state where strand C' and G1 are in dimer interface[193].

For comparison we used germline sequence, which is derived using Abysis website, where it selects closest antibody germline sequence among all sequences deposited in database. In germline there are 5 mutations that differs from patient sequence. Surprisingly, all the mutations were located to be either close or in the CDRs regions. **(Figure 18b)** Boundaries for framework, CDR regions and β-strands are done according the previously published crystal structure paper of this protein sequence[193].

3.1.3 Resonance assignment of V_L S3706

We assigned 95% of protein backbone in S3706_pat, S3706_patdel and S3706_GL. For sequential connectivity standard 3D HNCACB and 3D HNCOCACB was used. In addition, 3D HNCA was used as complementary for confirming the assignment. There are 110 residues in S3706_patdel out of which 5 are proline residues and the only residues which we could not able to assign are either in the region of dimer interface or in the N-terminal region. The primary sequence contains only one tryptophan, W35, the side chain peak of this tryptophan can be seen around 10 ppm in ^{1}H-^{15}N HSQC. (**Figure 19)** Very high concentration of protein was used for 3D experiments since the protein starts aggregating after three days. The details of aggregation are discussed in later section. The advantage of using high concentration was even after five days of recording the experiments still the original folded peaks are detected in the spectra.

For the assignment of S3706_GL same strategy was used as for S3706_patdel. Here we observed two sets of peaks for some residues that are in proximity with cysteine bridges. Moreover, we could clearly assign the oxidised and reduced cysteine, which undoubtedly tells about the mixture in the sample. We can resolve these two sets of peaks by either adding Hydrogen peroxide or changing the ratio of GSSG/GSH during purification in refolding process.

3.1.4 Effect of deletion of C-terminal constant domain

All the studies in solution and solid-state NMR were done in S3706_patdel construct since it is easy to compare with S3706_GL, which has exactly 5 mutations. It is important to understand the effect of 7 residues that does not belong to variable domain, however it was found in patient sequence[193]. **(Figure 20)**

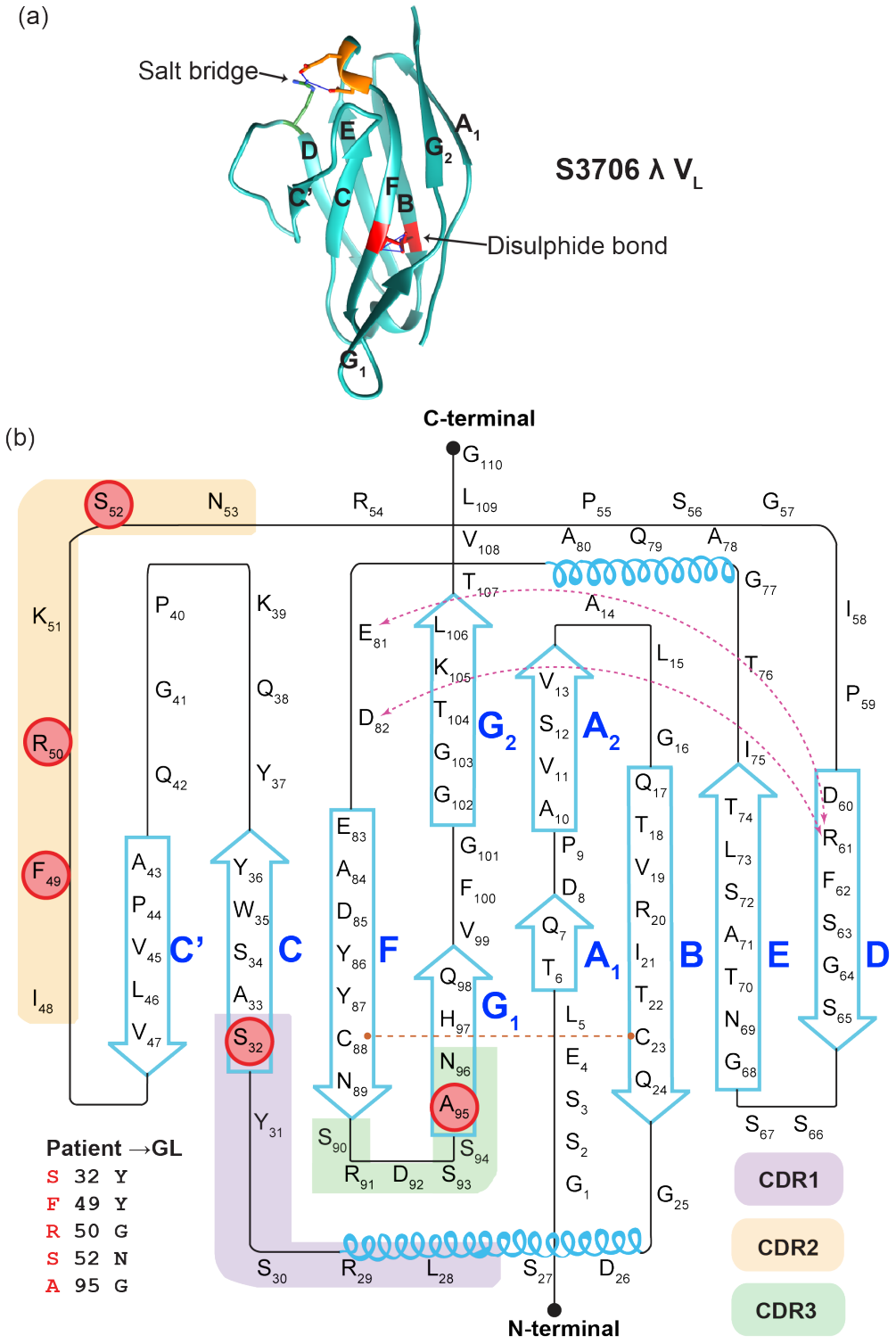


Figure 18: X-ray crystal structure and topology of V_L λ S3706_patdel.

(a) Crystal structure of the monomeric protein demonstrating its globular structure. The disulphide bond, salt bridge and strands nomenclature are marked[193]*. (PDB: 5L6Q) (b) Schematic topology model of S3706_patdel showing full sequence with the secondary structure elements (Ig fold) and complementary determining regions (CDRs). The cysteine disulphide bridge and salt bridge are marked with arrows in orange and pink colour respectively. The 5 mutations from patient to germline are marked in red round circle.*

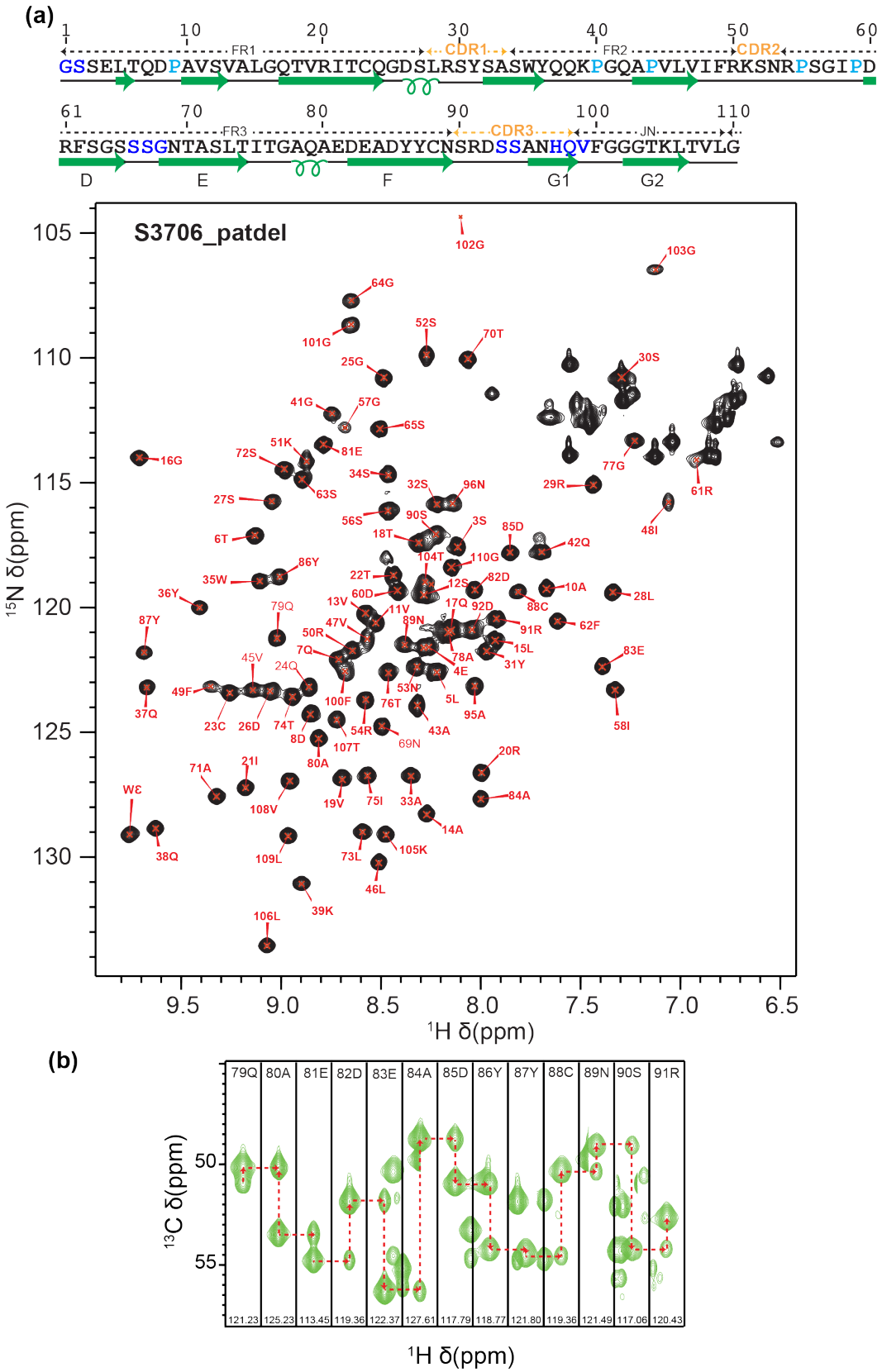


Figure 19: Sequential backbone assignment of S3706_patdel.

(a) ^{1}H-^{15}N HSQC with sequential backbone assignment. In primary sequence the residues marked in black are assigned while marked in dark blue are not assigned. There are 5 proline marked in blue colour. (b) 2D Strip plot extracted from 3D HNCA experiments demonstrating sequential assignment of a segment from S3706_patdel sequence.

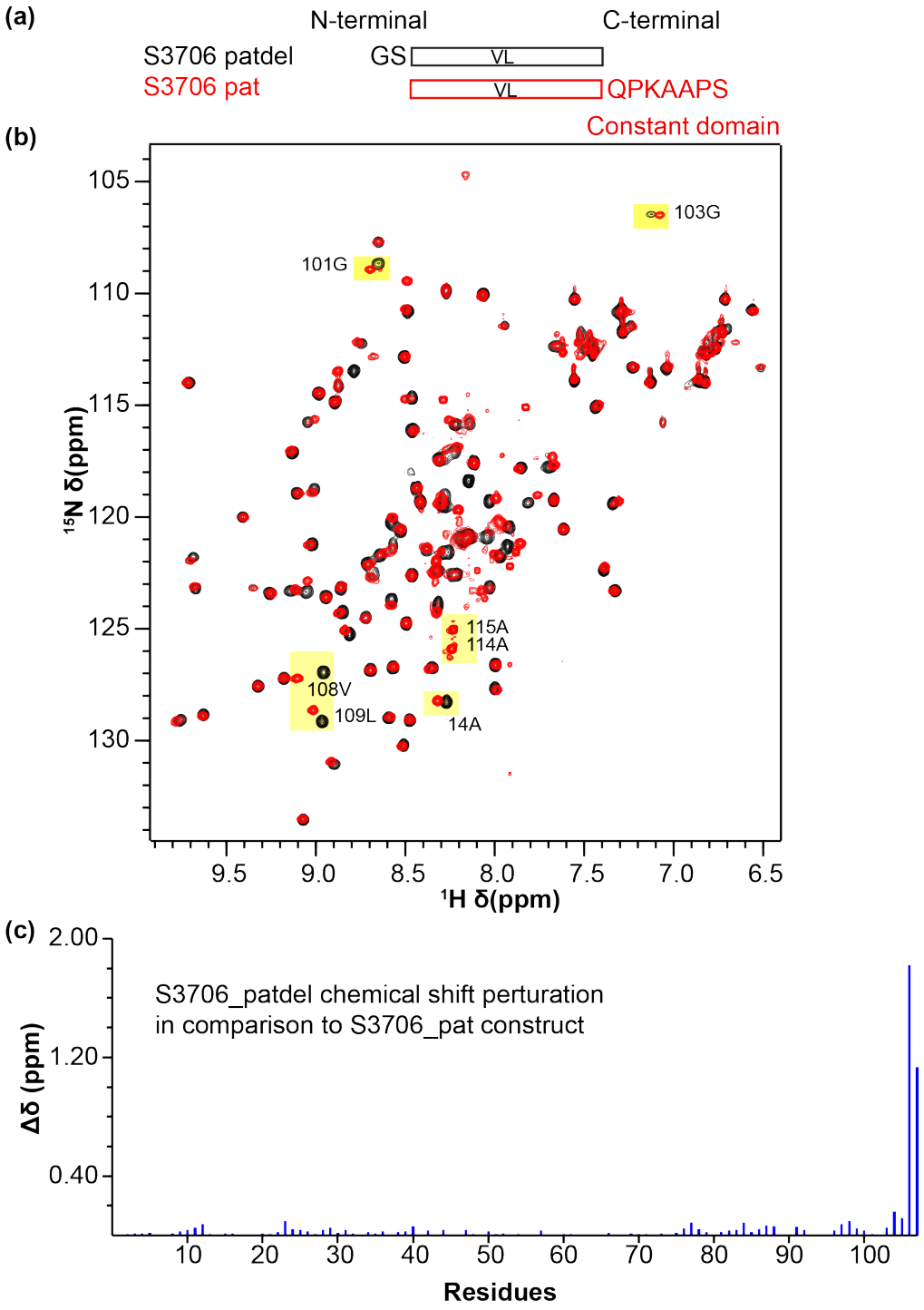


Figure 20: Comparison of S3706_patdel from S3706_pat protein.

(a) Construct in S3706 patient sequence (pat) and modified sequence (S3706_patdel). (b) ^{1}H-^{15}N HSQC spectra overlay of S3706_patdel (Black) and S3706_pat (red). (c) Chemical shift perturbation (CSP) in S3706_patdel on comparison to S3706_pat.

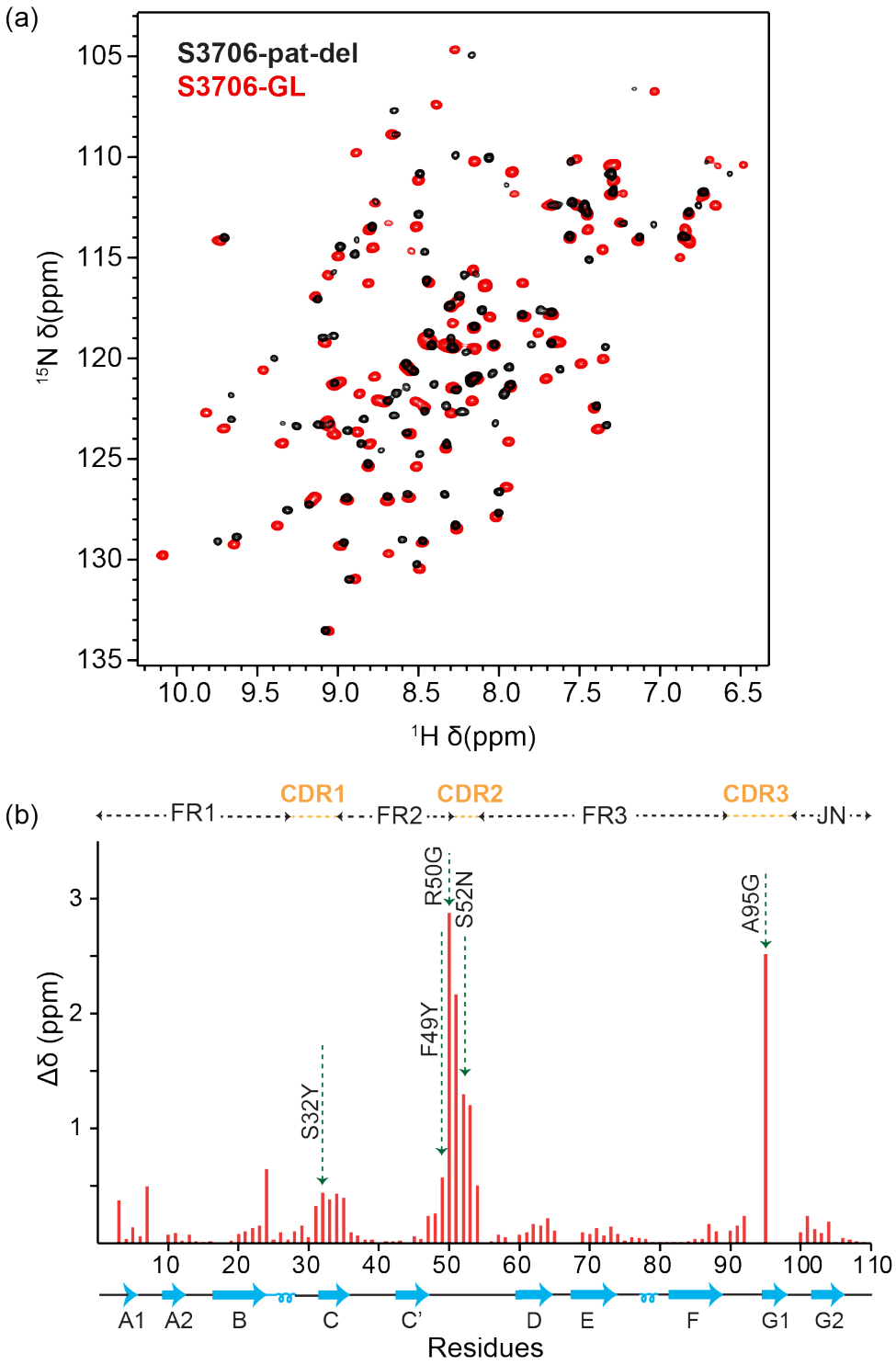


Figure 21: Comparison of S3706_patdel with its germline (S3706_GL).

(a) *^{1}H-^{15}N HSQC spectra overlay of S3706_patdel (Black) and S3706_GL (red). (b) Chemical shift perturbations in S3706_patdel on comparison to S3706_GL. Mutations from patient to germline is marked in arrows.*

3.1.5 Characterization of Dimer interface in V_L protein

In V_L proteins the crystallization typically results in homodimers (V_L-V_L) commonly seen in Bence Jones disease, has the same interface as heterodimers (V_L-V_H)[110]. Mostly all crystal structure, which are reported including our case S3706 were found to be canonical in structure. However, later it was demonstrated that non-canonical altered dimer structure could also exist where one of the monomers is rotated to 90° or 180°. These altered dimer structure were thought to be one of the factors that can promote amyloidogenicity[111,112,115]. Taken into this account it is important to probe the dimer interface in our protein.

Despite sequence variability such as in LEN κ-LC (subgroup I) using relaxation dispersion experiments, MAK33 murine κ-LC (subgroup III), and S3706_patdel λ (subgroup III) by ^{1}H-^{15}N HSQC residue intensities, it was found that most of the protein possess homodimer and same dimer interface. In all the three cases C' and G strands are involved in dimer interface. **(Figure 22)** However, the orientation of the monomers might be different.

To further characterize the dimer properties, concentration dependent titration in solution state NMR was performed for both S3706_patdel and S3706_GL. We used series of dilution from 1.1 mM to 13 µM for S3706_patdel to get 13 dilution points and from 1.3 mM to 18 µM to get 14 points for S3706_GL protein. Due to chemical exchange, line broadening in ^{1}H-^{15}N HSQC spectra is observed that indicates the existence of another molecular species dimer state. Comparing the lowest and highest concentrated spectra, the largest chemical shift change was observed in G102 residue in both proteins. **(Figure 23a)** Mapping all the chemical shift perturbation to residues shows that the large changes are localized in C' and G2 strand in both proteins. Slight differences are observed in the loop region between F and G1 strand in S3706_GL. **(Figure 23b)** Using G102, A43, K39 and F100 chemical shift values from dilution series, we calculated the K_d for dimer interface for each residue and the average K_d was 1737 ± 609 µM in S3706_patdel. Similarly, using G102, A43, K39 and G41 chemical shift values K_d was calculated and the average was 3603-± 322 µM in S3706_GL. **(Figure 23c)**

For the calculation of K_d, we took the following assumption and equation:

$$\left.\begin{aligned} K_d &= \frac{[M]^2}{[D]} \\ x &= [M] + 2D \\ \frac{(CS - CS_{monomer})}{CS_{dimer} - CS_{monomer}} &= \frac{2[D]}{([M] + 2[D]} \end{aligned}\right\} \text{Assumptions}$$

$$CS = (CS_{dimer} - CS_{monomer}) * \frac{\left(\left(K_d^2 + 8xK_d\right)^{0.5} - K_d\right)^2}{(8xK_d)}$$

[M], [D] are the concentration of Monomer and dimer respectively. x is the total concentration. CS is chemical shift.

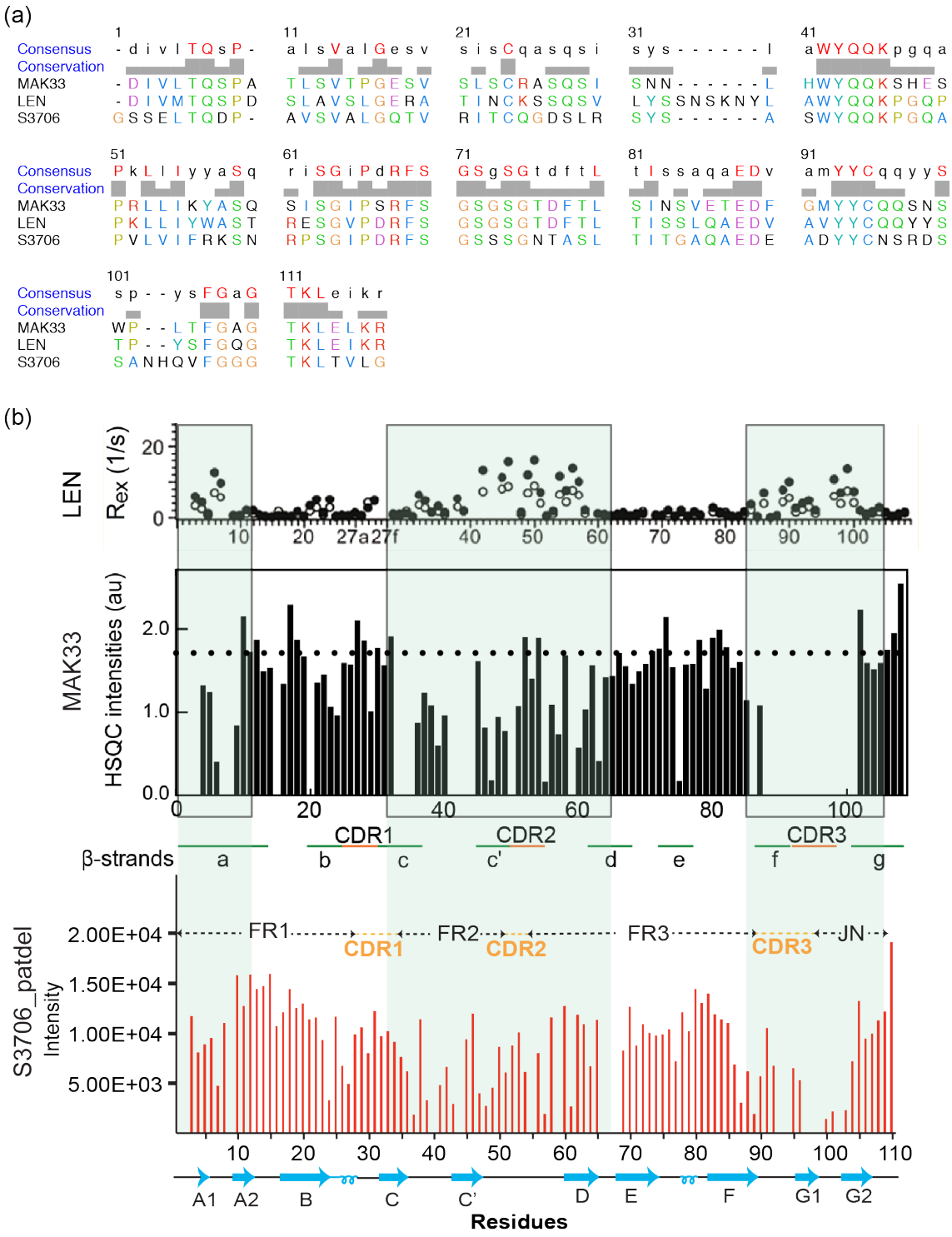


Figure 22: Dimerization in V_L proteins.

(a) Sequence alignment of different V_L sequence, MAK33 murine κ-LC subgroup III, LEN κ-LC subgroup IV[194] and S3706 λIII subgroup, using CLUSTAL W. (b) Comparison of interface region in different V_L sequence obtained by NMR relaxation dispersion experiments in LEN protein and ^{1}H-^{15}N HSQC intensity experiment in MAK33 and S3706_patdel protein.*

* MAK33 ^{1}H-^{15}N HSQC intensities plot provided by our previous colleague Dr. Manuel Christin Hora

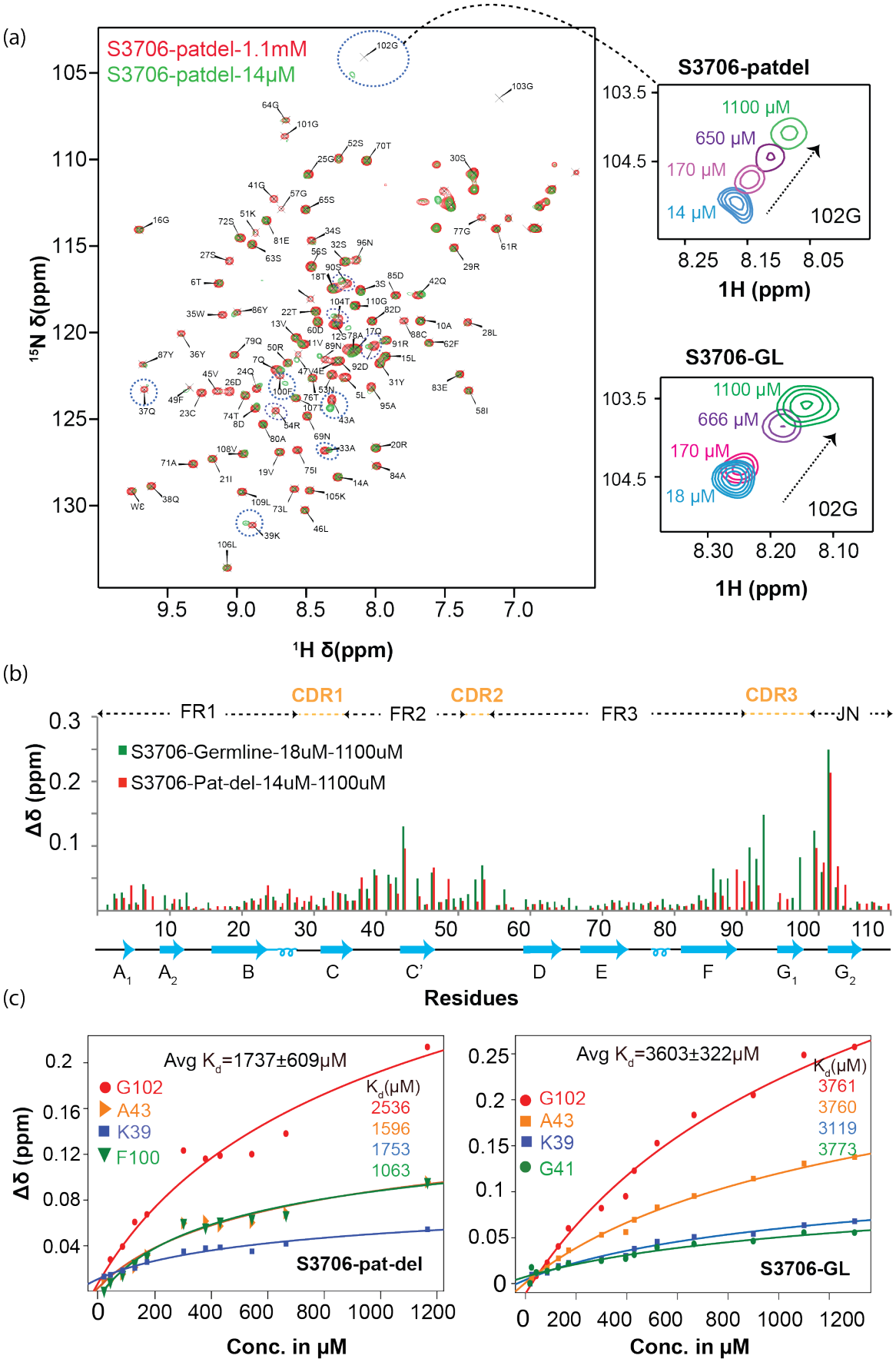


Figure 23: Concentration dependent titrations in V_L proteins.

(a) ^{1}H-^{15}N HSQC overlay of 1.1 mM and 14 µM in S3706_patdel. The residues that shows significant

CSPs are encircled in blue color. The residue shows two state monomer-dimer equilibrium that are in fast exchange on the NMR timescale. Right hand side: Zoom plot of G102 at different concentration was shown for both S3706_patdel and S3706_GL. (b) ^{1}H-^{15}N CSPs in S3706_patdel and S3706_GL for the residues involved in dimer interface by comparing the highest and lowest concentration of protein. (c) Normalized ^{1}H-^{15}N HSQC CSPs plot for residue G102, A43, K39 and F100 in S3706_patdel and G102, A43, K39 and G41 in S3706_GL at series of dilution of protein. Non-linear curve fitting of dilution data gives the average K_d 1737 ± 609μM and 3603 ± 322μM for S3706_patdel and S3706_GL respectively.

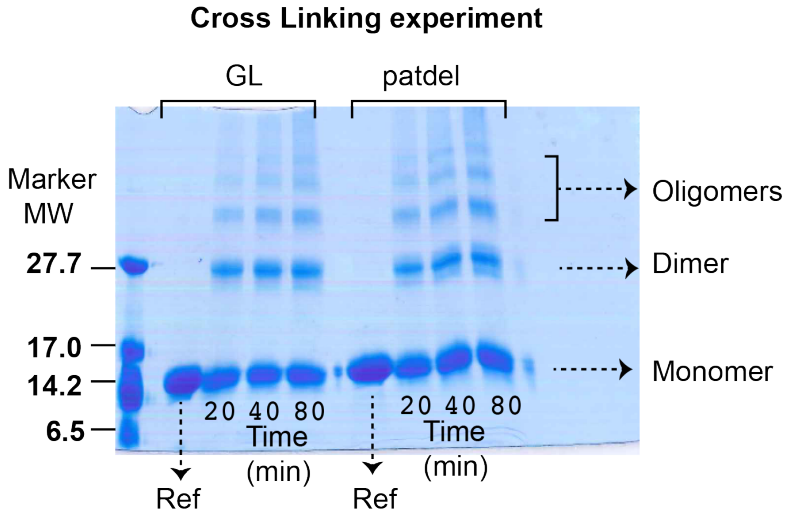


Figure 24: Chemical Cross-linking experiment.

Experiment was done by using chemical cross linker, glutaraldehyde. SDS PAGE reveals that protein solution has monomer, dimer and higher molecular weight oligomers. The molecular weight of protein is ~12kDa for both patient and germline protein. Ref is the reference where no glutaraldehyde is added. MW= Molecular weight in kDa.

As a complementary test we performed cross-linking experiment using glutaraldehyde as a chemical cross linker. We observed the dimer and oligomer bands distinctly in SDS-PAGE gel for both S3706_patdel and S3706_GL protein. The gel results indicate that in both proteins we have a mixture monomer, dimer and high molecular weight oligomers states. **(Figure 24)**

We further investigate about the dimer model that can be possible in V_L patient proteins. In case of patient sequence, the X-ray crystal structure is canonical where G-G and C'-C' strands form the interface, and this is the common observation found in most of the V_L proteins. In addition, we modified our model according to 2KQM and 2KQN (from database) and named as S3706 -KI Y87H-modified and S3706-AL-09 H87Y-modified[112]. S3706 -KI Y87H-modified is canonical and superimposes well with crystal structure with slight extension of G-strand. S3706-AL-09 H87Y-modified is the altered non-canonical structure where one monomer is rotated to 90° and the dimer interface is formed by C'-G strands.

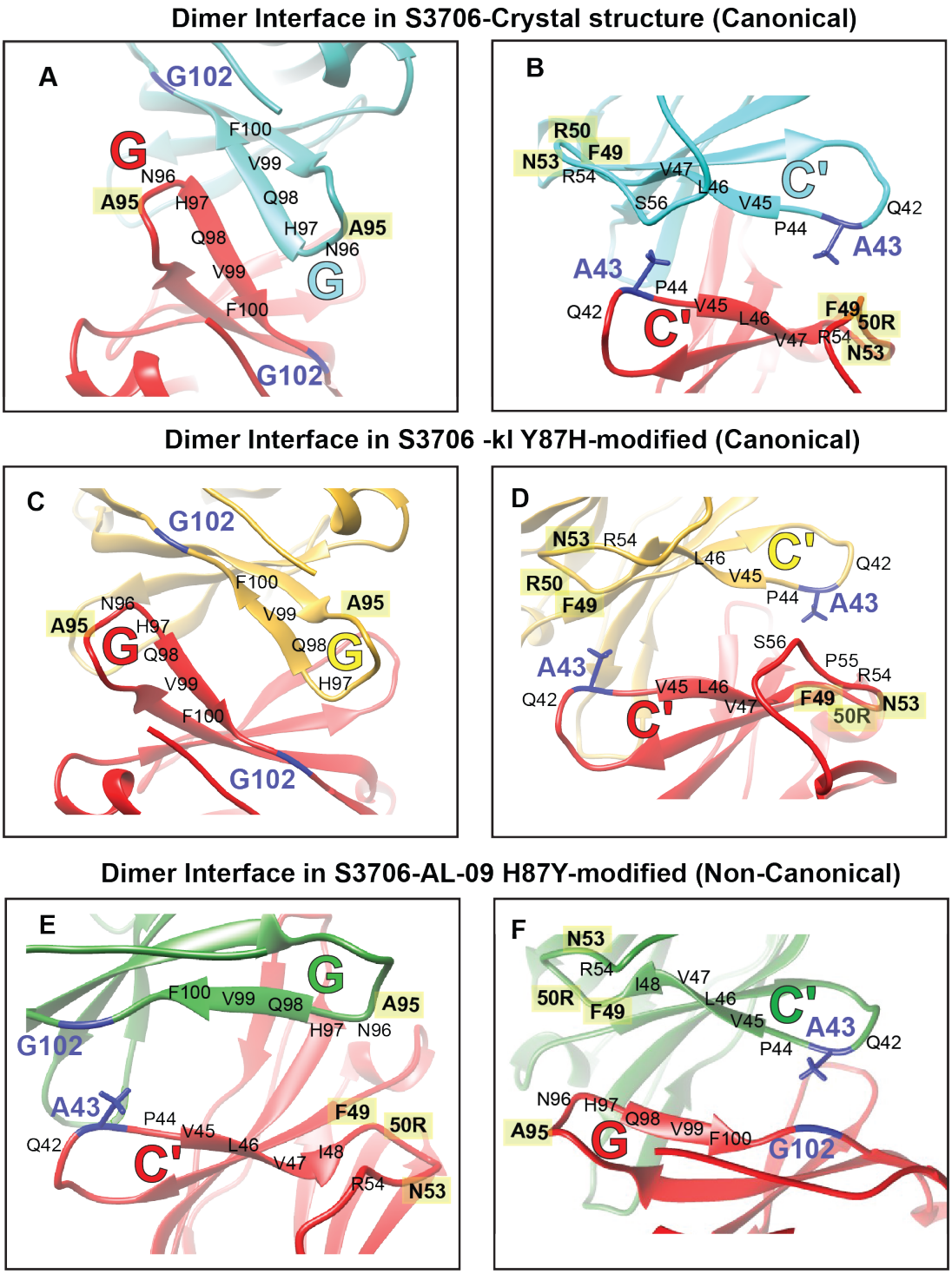


Figure 25: Dimer Interface in different dimer models.

A & B *are the ribbon views of canonical S3706_patel obtained from crystal structure PDB: 5L6Q.* ***C & D*** *are the ribbon views of canonical S3706 -kI Y87H-modified structure that are adapted from PDB: 2KQM.* ***E & F*** *are the ribbon view of non-canonical S3706-AL H87Y-modified structure that are adapted from PDB: 2KQN*[112]*. NMR based concentration dependent titration with significant chemical shift* ***(A43, G102)*** *are shown in purple color.* ***49F, 50R*** *and* ***53N*** *are the mutations in interface region are highlighted in yellow color.*

When CSPs obtained from concentration dependent NMR titration is mapped on the crystal structure (canonical), the residue G102, displaying the most significant change does not fit to the model. The model fits perfectly while we slightly extend the G strand like in the case of S3706 -KI Y87H-modified (canonical). Moreover, in S3706 -KI H87Y-modified model (non-canonical), the NMR data fits perfectly. This indicates that both canonical and non-canonical structural dimer model is possible. The dimer interface of S3706 protein in solution can be different from the interface that is observed by the crystal structure.

3.1.6 Aggregation in Light chain Antibody

In this section, we described the aggregation behavior of V_L proteins by means of NMR complemented by biophysical techniques like CD and ThT. We were able to characterize the oligomeric states of protein by NMR and DLS.

3.1.6.1 Thermodynamic stability

(Provided by Dr. Benedikt Weber from Prof. Dr. Johannes Buchner group)

Thermodynamic stability of the protein was determined by using CD spectroscopy. Thermal unfolding curves shows that the thermal transition is 44.2°C, 55.7°C and 48.6°C for S3706_patdel, S3706_GL and S3706_R50G respectively. Among the three constructs, most unstable protein is S3706_patdel and most stable is its germline protein. **(Figure 26)**

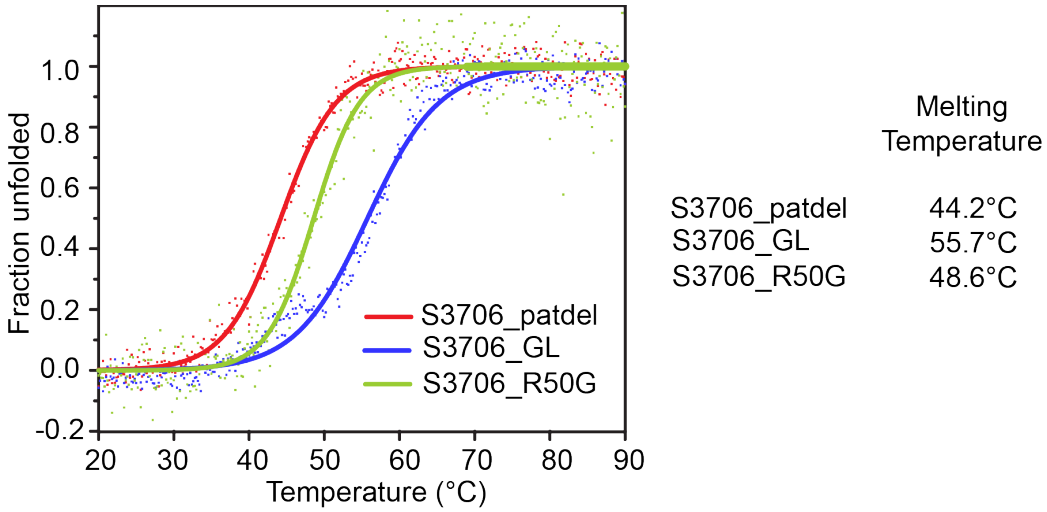


Figure 26: Thermal-unfolding curves.

Thermal Transition shows that S3706_patdel, patient sequence is the most thermodynamically unstable among its variants[*].

[*] CD Data was provided by Dr. Benedikt Weber from Prof. Dr. Johannes Buchner group

S3706_R50G shows the transition temperature in between the patient and germline protein. Since, this protein is single point mutation and the thermal stability of the protein is increased by 5°C, suggesting the crucial role of this amino acid. With five mutations in germline protein, S3706_GL has 10°C higher meting temperature than patient protein S3706_patdel. This indicates that patient protein has high degree of aggregation propensity that is due to lower thermal stability. To further confirm we performed kinetic assay by ThT fluorescence and aggregation kinetics by solution state NMR that are discussed in next sections.

3.1.6.2 Aggregation Kinetics by solution state NMR

We recorded ^{1}H-^{15}N HSQC experiment each day to follow aggregation kinetics of S3706 variants with time of 7-8 days. All the experiments were performed at 25°C using same NMR spectrometer. The experiments were done 3 times independently to check the repetition of aggregation. The drastic changes were observed in patient sequence, S3706_patdel. On onset of 3rd day, along with natively folded peaks, we detected new peaks in unfolded region of the spectra. Following 3rd day, new peaks was observed out of which some are short lived while some stays overtime. On 7th day all the peaks were unfolded with no sign of folded peaks. On the contrary, S3706_GL shows no sign of new peaks and only the intensity from natively folded peaks decreases overtime. In case of single mutation S3706_R50G we observed few peaks on 3rd day onwards but those were found to be short lived and weak in intensity overtime. In this mutant, on 8th day, the spectrum was empty with only the side chain visible. Cross peak intensity from the resonance from three backbone peaks including G16, I58 and E83 versus time plot indicates that germline, S3706_GL is the most stable protein and aggregates slowest with sign no intermediate states while patient protein is most unstable and aggregates fastest exhibiting some intermediates states. Single mutant variant, S3706_R50G is in between the two extreme variants. **(Figure 27)**

To explore further about intermediates states observed in S3706_patdel, we analyzed peak from side chain of tryptophan. In all S3706 constructs it is to be noted that only one tryptophan exists at position 35. This single occurrence of this residue makes our aggregation intermediates analysis convenient. As previously mentioned, nothing unusual was a seen in NMR spectrum of S3706_GL protein during aggregation. Only one peak from folded W35ε peaks was observed which reduces its intensity overtime. In S3706_patdel aggregation kinetics, after 2nd day we detected very weak intensity peak that intensifies till 3th day and then disappeared next day. On third day, we observed two new peaks one with high intensity I_2 and another low intensity I_3 that stays till the end i.e. 7th day of experiment. Similarly, we detected two more intermediates for W35ε at 5th and 6th day. S3706_R50G protein showed both traits of germline and patient. The native W35ε can be followed till 7th

day of the experiment and overtime along with backbone native peaks it diminishes at 8th day. We detected a short-lived intermediate I_1 and long lived I_2 that remain till the end of the experiment. On 8th day as mention before only side of glutamine and tryptophan (Intermediate I_2) can be seen. **(Figure 28)**

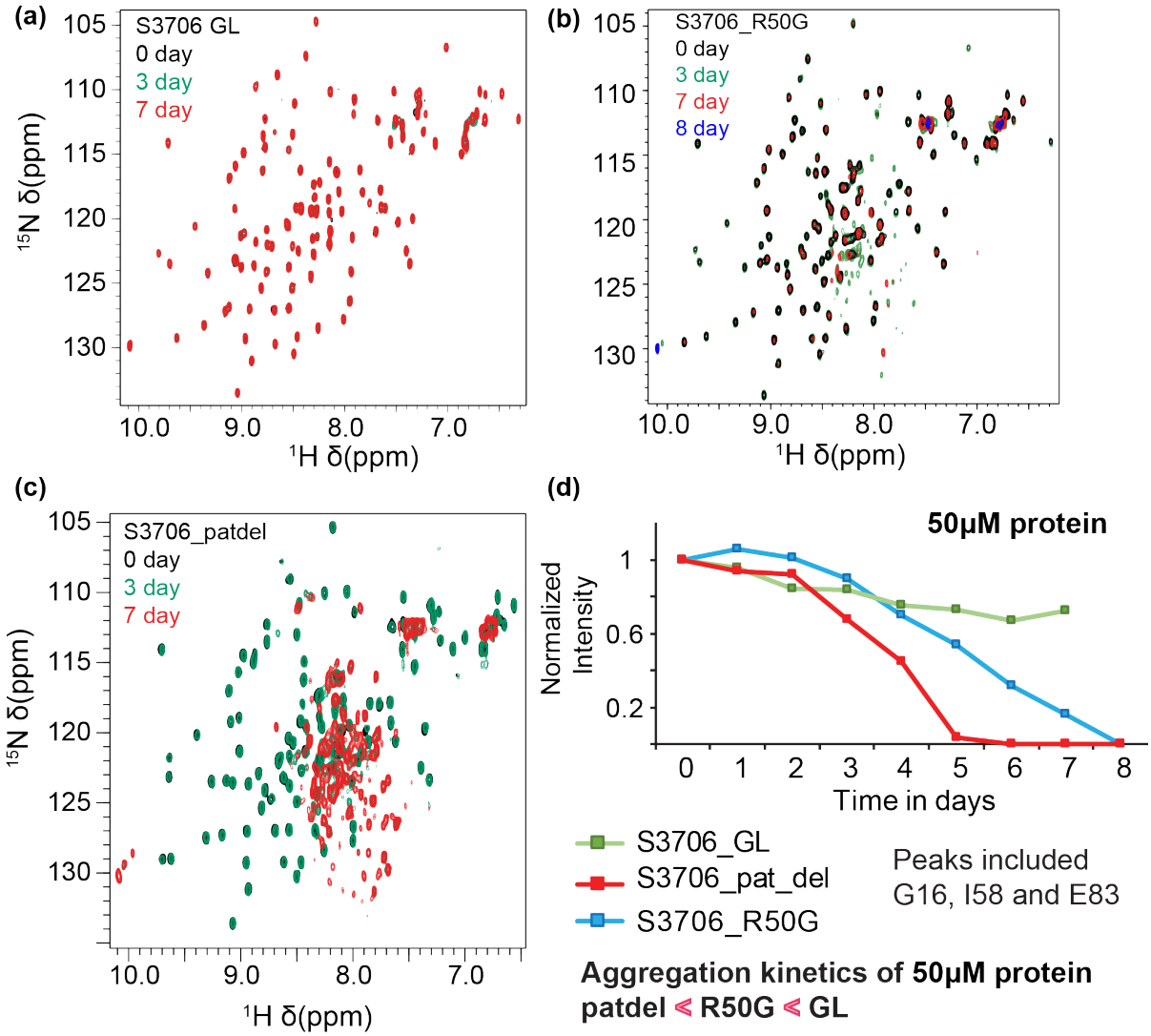


Figure 27: Aggregation kinetics of S3706 variants using solution state NMR.

(a), (b) and (c) are the ^{1}H-^{15}N HSQC overlay representing time dependent aggregation kinetics of 50µM S3706_GL, S3706_R50G and S3706_patdel monomeric protein respectively at 25℃. All the condition was kept same for all variants and experiments were recorded in same NMR spectrometer. In S3706_GL peaks folded peaks intensity decreases overtime with no sign of new peaks. In S3706_R50G, few new peaks are observed overtime, but everything disappears at 8th day leaving only side chains. In S3706_patdel, after 3rd day intermediate peaks can be observed indicating partial unfolding of peaks. (d) Normalized intensity plot for S3706 variants over time in days. Here, folded peaks G16, I58 and E83 backbone peaks are included and averaged. These data confirm that S3706_patdel protein is most unstable and aggregates faster than its variants.

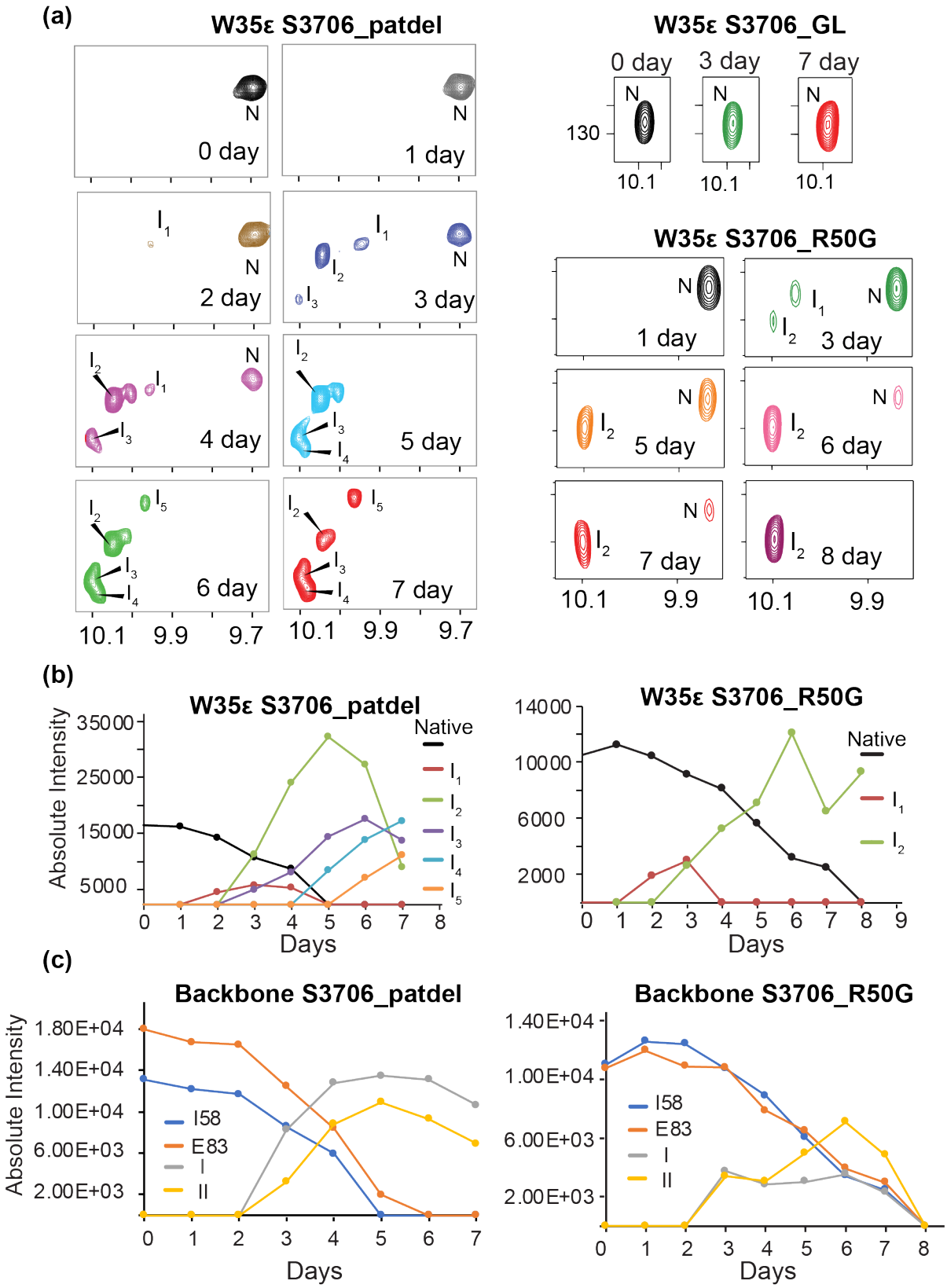


Figure 28: Analysis of tryptophan side chain and intermediates.

(a) Aggregation kinetics of W35ε in S3706 variants at 50 μM concentration. N is the native folded W35ε that disappears overtime in S3706_patdel and S3706_R50G while in S3706_GL peaks was seen till 7th day of experiment. Intriguingly, new peaks in S3706_patdel is observed and named as

intermediates I_{1-5}. In case of S3706_R50G only two intermediates were observed for W35ε whereas S3706_GL shows no intermediate states during time dependent aggregation. (b) Cross peak intensity plot of W35ε versus time is plotted which illustrates the kinetics of native and intermediates W35ε peaks in S3706_patdel and S3706_R50G. (c) Kinetics of folded and unfolded intermediates peaks from protein backbone with time. I56 and E83 residues are folded backbone residues and I and II are the intermediate peaks picked from protein backbone.

To analyse the kinetics of the intermediates in S3706_patdel and S3706_R50G, we took the cross peaks intensity of native and intermediate W35ε and plotted them versus time. Surprisingly, I_2 intermediate showed higher resonance intensity than the native W35ε peak in both proteins. Furthermore, it was certain that the initiation of intermediates formation differs with time; some are short lived while some long. **(Figure 28b)** Like previously stated, we also analysed the backbone folded, unfolded intermediate peaks in S3706_patdel and S3706_R50G. Intermediates II and I shows the similar kinetics as folded side chain W35ε. For backbone peaks analysis we included I56 and E83 residues for both S3706_patdel and S3706_R50G. The intermediate shows the maximum intensity at 5th and 6th day and then reduces with time in case of S3706_patdel and S3706_R50G respectively. **(Figure 28c)** To verify that these intermediates are not due the effect of thermodynamic stability, we recorded the time dependent aggregation again by changing the temperature. For S3706_patdel the temperature was lowered to 288 K, 10 °C lower than the room temperature (298K) while for S3706_GL was increased to 310 K, 10 °C higher than the room temperature (298K). Interestingly, both proteins demonstrated the same behaviour as recorded at 298K. In germline, there were no intermediates and in patient protein, unfolding of peaks and intermediates was observed. **(Figure 29)**

A significant effect was observed on comparing ^{1}H-^{15}N HSQC of S3706_patdel kinetics, at lower (50 μM) and higher concentration (1 mM). One observation was, regardless of protein concentration, the onset of aggregation starts at 3rd day. Unfolded peaks from lower concentration overlaps perfectly with higher concentration protein. Second crucial observation was at 5th day lower concentration protein unfolds completely while in higher concentration protein, folded peaks are still observed along with unfolded peaks. **(Figure 30)** With higher concentration (1 mM), the folded peaks can be observed till approx. one month. Comparison of aggregation kinetics for high and lower concentration shows that the aggregation initiation process is at same instance while the end stage gets prolonged at higher concentration. **(Figure 31)** This observation suggest that at higher concentration, dimer population is high that renders protein to unfold and thus protects it to go further aggregation process to from fibrils.

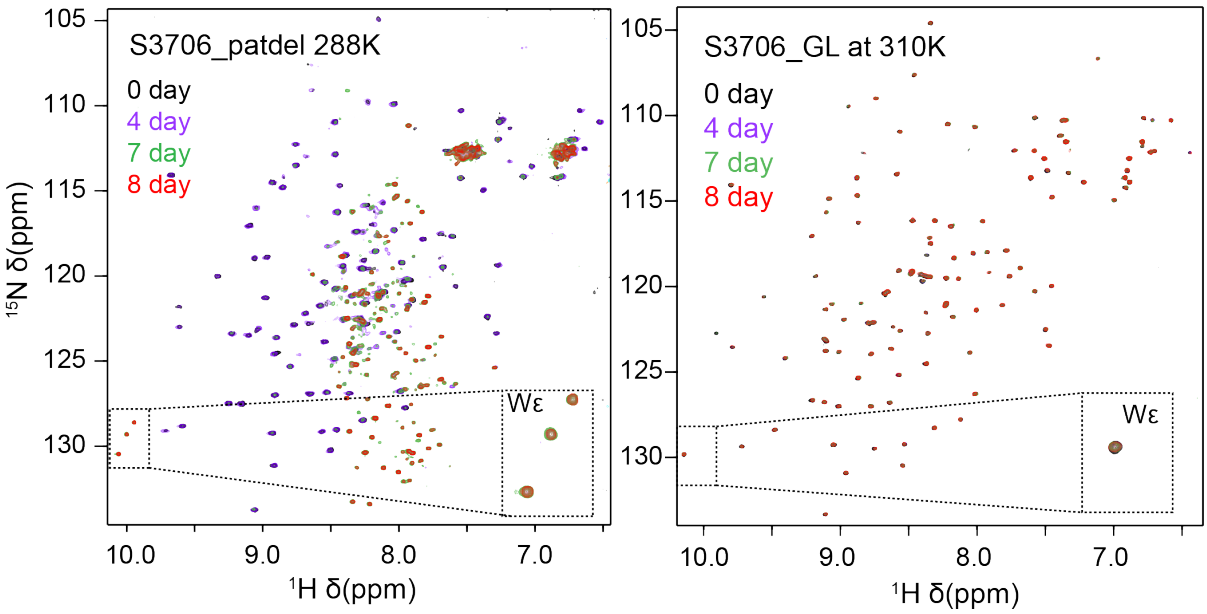


Figure 29: Aggregation kinetics at different temperature.

Time dependent aggregation kinetics was recorded 10°C lower and higher from room temperature (25°C) in S3706_patdel and S3706_GL protein respectively. For S3706_patdel and S3706_GL, even at 288K we observe the same behaviour of W35ε, and backbone residues as recorded at 298K.

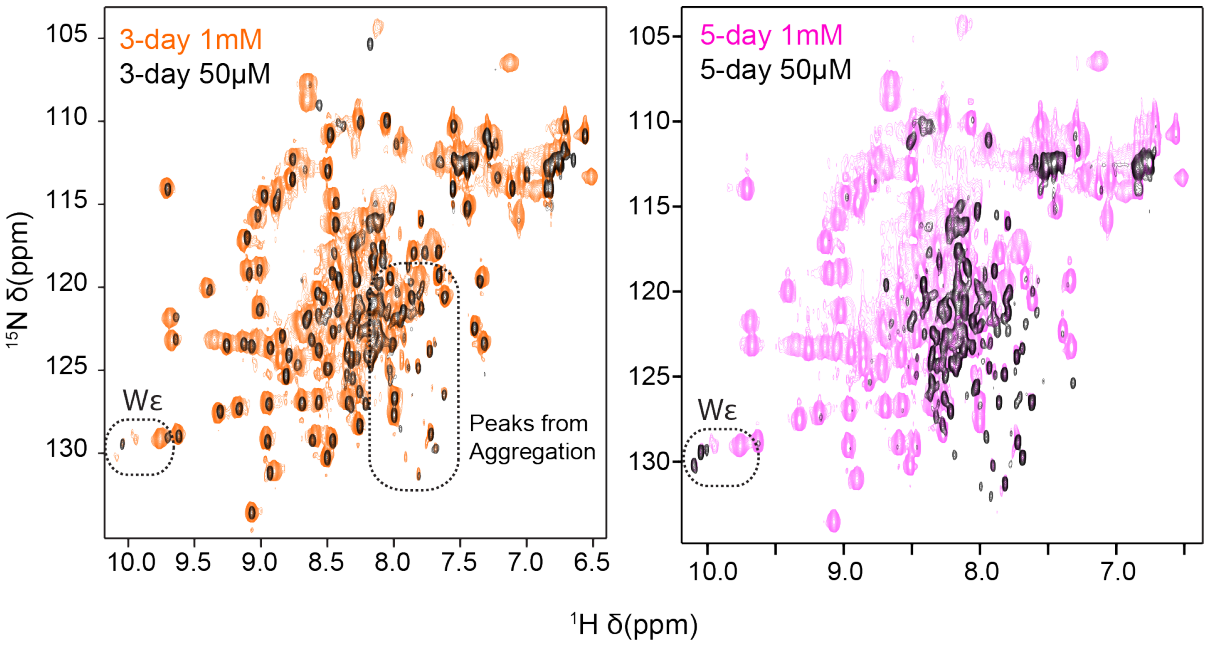


Figure 30: Effect of concentration on aggregation kinetics.

Comparison of ^{1}H-^{15}N HSQC S3706_patdel at lower (50 μM) and higher concentration (1 mM) at 3rd day indicates that initiation of aggregation in protein is at same time regardless of the concentration and both samples are partially unfolded. At 5th day the lower 50 μM protein is fully unfolded in comparison to 1 mM protein. This indicted that at higher concentration the unfolding process is protective.

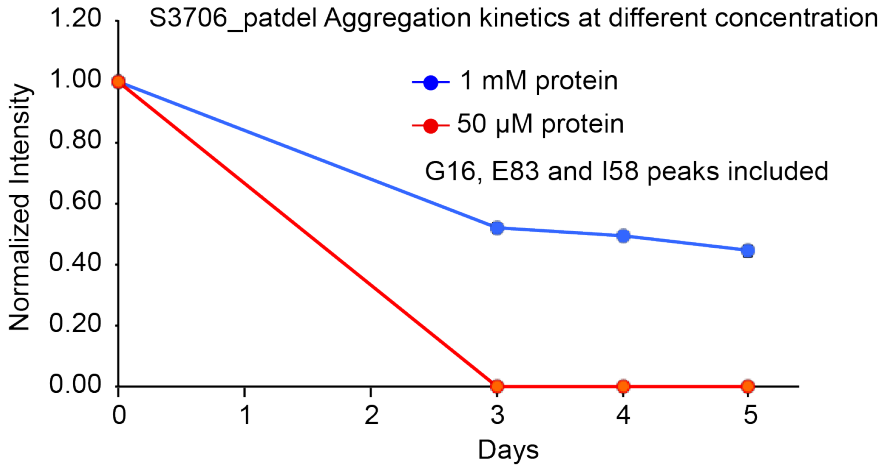


Figure 31: Aggregation kinetics of S3706_patdel at different concentration

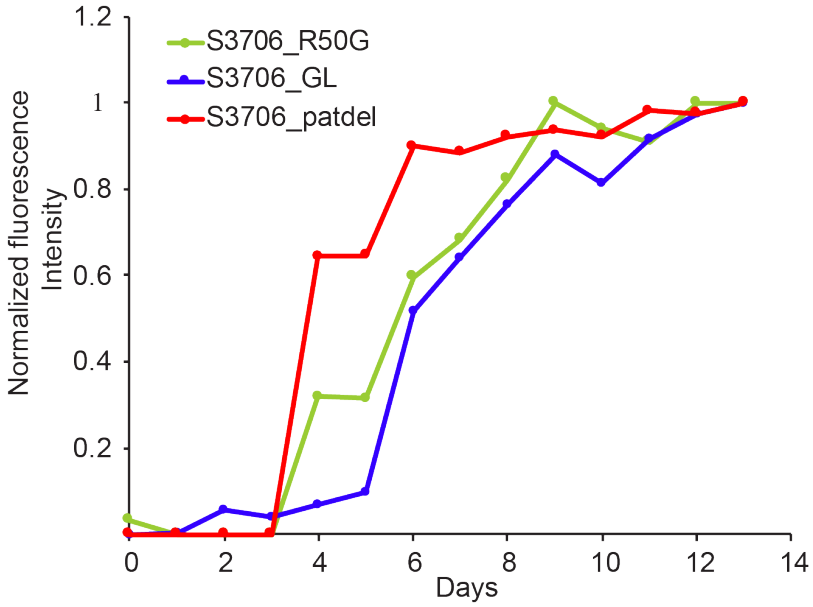


Figure 32: ThT kinetics of S3706 V_L variants.

Experiments were performed in triplicates at 37℃ with orbital shaking. ThT fluorescence is normalised and is the average value of triplicates. The plot shows that S3706_patdel forms fibrils faster than its single mutant R50G and germline with 5 mutations.

3.1.6.3 ThT kinetics

In order to probe the fibril kinetics, we used standard ThT fluorescence assay. ThT dye with positive signal is used to monitor amyloid fibril formation. The kinetic experiment was performed for 2 weeks and the fluorescence measurement was done once in a day for each variant. For 2 weeks protein was incubated at 37°C in 96 well plates with gentle shaking. All the conditions were kept same for all variants during experiment as described in *Material and Methods* chapter under section **2.2.2.4.**

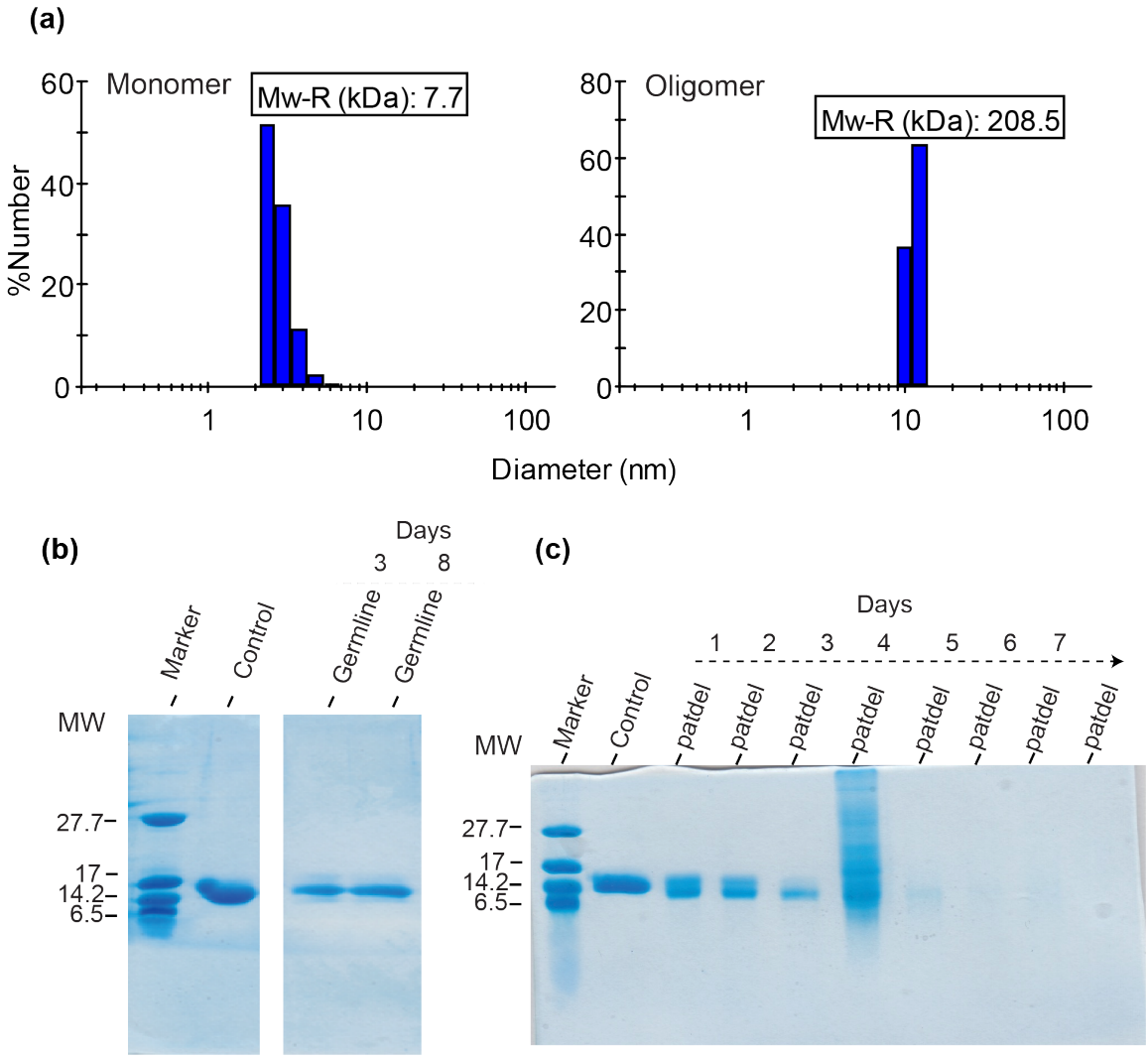


Figure 33: Oligomer characterization by DLS and SDS-PAGE

(a) & (b) depicts size distribution of monomer and oligomeric intermediates species with their % population using DLS experiment respectively. (c) & (d) are SDS PAGE of S3706_GL and S3706_patdel overtime. In germline, monomer band is observed at end of the experiment (8th day) in patient protein, monomer band disappears after 4th day indicating that monomer is changed into higher molecular weight species. There is no degradation observed overtime in both proteins.

In all the variants, we observed the lag phase followed by growth phase and stationary phase. From the ThT kinetics, it was revealed that S3706_patdel aggregates fastest to form fibrils followed by single point mutant S3706_R50G and slowest is the germline protein, S3706_GL. **(Figure 32)** This also supports the thermodynamic and aggregation data that is previously discussed in section **3.1.6.1** and **3.1.6.2** where patient is the most unstable protein.

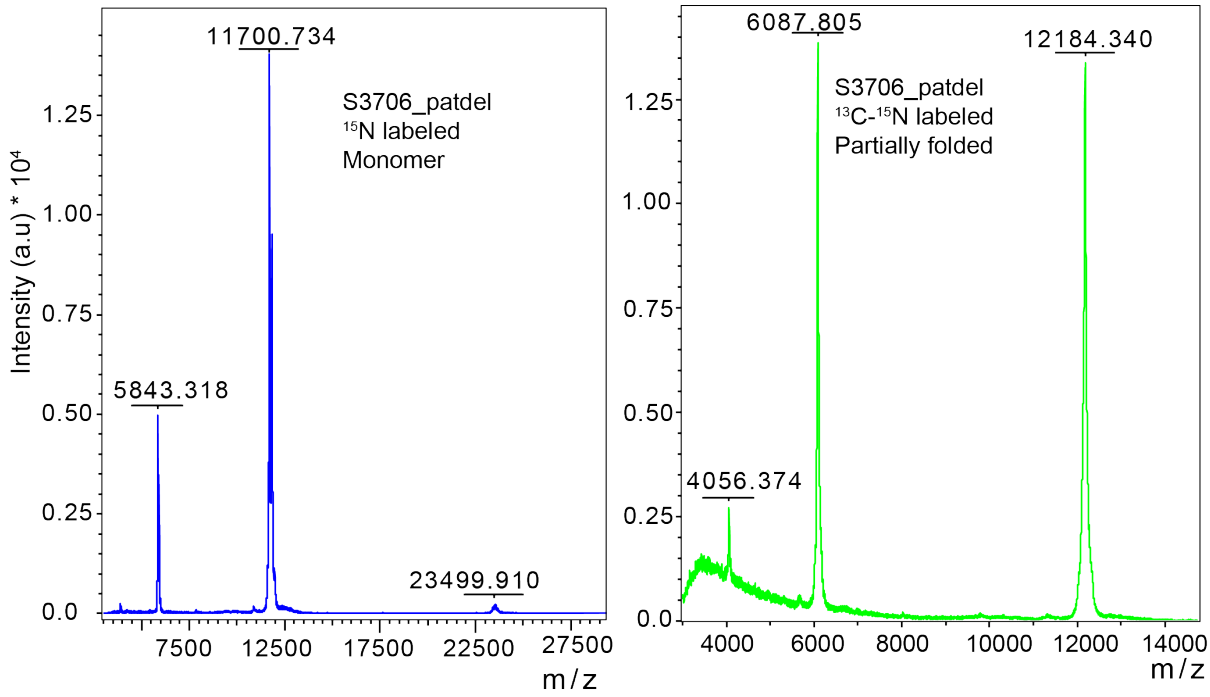


Figure 34: Mass spectrometer data for S3706_patdel

3.1.7 Oligomer characterisation

Intermediates during aggregation were suspected to consist of higher molecular weight species. In order to confirm this, we did Dynamic light scattering (DLS) experiment for monomer and intermediates along with conventional SDS-PAGE gel. In DLS experiments, fresh S3706_patdel protein with no intermediate state in NMR spectra showed low diameter representing smaller, monomeric species while the protein with intermediates states in NMR spectra showed higher diameter species indicating the larger, higher molecular species oligomers. **(Figure 33 a-b)** Through SDS PAGE, we saw disappearance and weakening of monomeric bands and appearance of larger molecular weight bands indicating oligomer formation that occurs overtime in days in S3706_patdel while in S3706_GL there was no change observed over 8 days. The gel results also confirm that there were no degradation bands for both proteins overtime. **(Figure 33 c-d)**

We also check S3706_patdel protein mass to ensure there is no degradation during aggregation process using MALDI. **(Figure 34)** The sample used was differently isotopically labelled in both samples but after deduction of respective mass corresponding to its isotope, the mass number remain same.

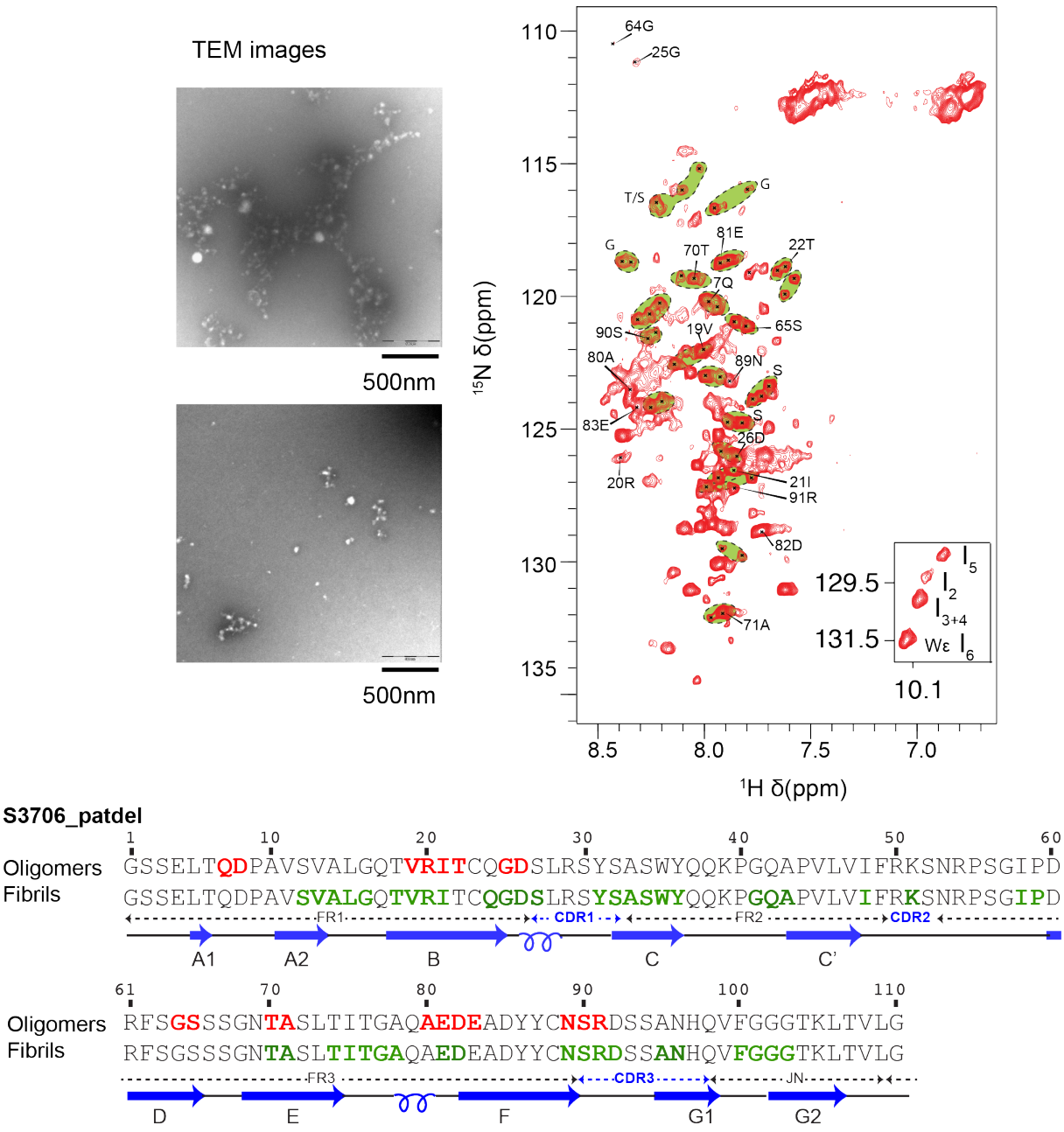


Figure 35: Oligomer characterization by NMR and TEM

From TEM images we observed round structures of various size. Resonance assignment of amide backbone of oligomer species in ^{1}H-^{15}N spectra is displayed on right hand side. The peaks marked in green shadow are the residues that show 2-3 sets of resonances. Four sets of resonances are shown for W35ε side chain. Secondary structure of protein is represented according to crystal structure published recently which is similar to the monomer structure in solution[195].

We also confirm the existence of S3706_patdel oligomers by TEM, which revealed variable round structures, which are larger in size. **(Figure 35)** We further investigate oligomers by NMR by performing conventional 3D experiments to assign protein backbone amide peaks. Because of more than one intermediates were seen overtime, as previously

described in time dependent aggregation kinetics of S3706_patdel, we recorded 3D experiments every day for almost one month in 1 mM protein sample. Here also 2-3 backbone resonances of same peak were observed that shows same resonances in 3D HNCACB for same residues and 3D HNCOCACB for preceding residues. This repetition of peaks made the sequential assignment of backbone amide peaks difficult. Nevertheless, we assign approx. 20 residues from oligomers and compared with fibrils assignment, which will be discussed in next chapter. Although many peaks were strong in 3D assignment spectra, there was no sequential residue available. Therefore, either it was not assigned or assigned only residue type for those residues. Some of assigned segments of oligomers were found similar to assigned segments in fibrils. Although the explanation of similarity in assigned segments in oligomers and fibrils is not so clear at the moment but the sedNMR of the oligomers in solid-state NMR can suggest about the explanation. **(Figure 35)**

3.1.8 Discussion and conclusion

Here we will discuss the dimer and aggregation properties of V_L proteins. We assigned 95% protein backbone except non-prolines residues in S3706 variants. The residues, which are not assigned belongs mostly to dimer interface region.

S3706_pat is the patient derived sequence with seven extra residues, which does not belong to variable domain. We showed that deletion of these residues does not affect the overall integrity of the protein. Furthermore, these additional residues were found to be unstructured by NMR analysis. In X-ray crystallography, the additional residues do not show up, indicating the unstructured segments of protein crystals. In MAK33 Kappa light chain, it has been shown that I2E mutation at 2nd position destabilize the protein that leads to the fibril formation[196]. However, in S3706_pat, a lambda light chain, even with the addition of two residues at N-terminal did not affect the overall protein structure. S3706_pat and S3706_patdel behave similar in terms of structure but kinetically and thermodynamically difference is still not well understood. In all our experiments we used S3706_patdel construct denotes as patient protein and S3706_GL as germline protein.

From CD data, it was very straightforward revealed that S3706_patdel is most unstable and S3706_GL is most stable protein among three variants. Single point mutant S3706_R50G thermal transition data is in between the patient and germline suggesting that this point mutation might be important for changing the properties of germline to patient.

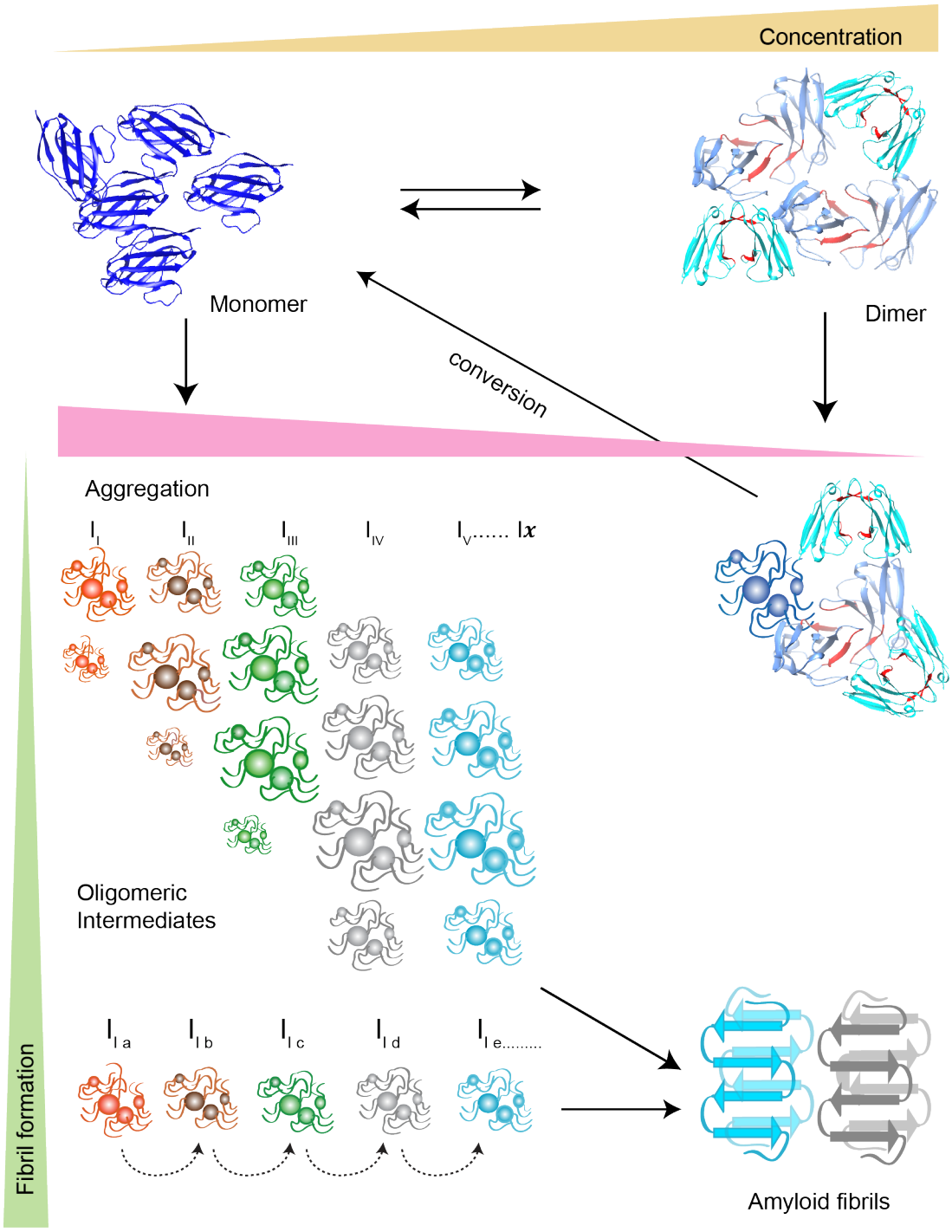


Figure 36: Pictorial representation depicting the relationship between concentration and aggregation in S3706.

Model representing the relationship between concentration and aggregation in S3706 light chain. As we drive from lower to higher concentration in S3706 protein, the monomer-dimer equilibrium shifts towards dimer. Monomeric protein was found to be prone to aggregation while dimer protects the protein towards aggregation. Dimer population can convert into monomeric protein to follow aggregation pathway with time. During aggregation kinetics more than one oligomeric intermediate were formed in patient protein via unfolding out of which eventually only few leads to fibril formation. Different oligomeric state observed are either due to polymorphism or are merely the stages of

intermediates during the pathway to fibrils. These intermediates were found specifically in-patient protein and captured by solution state NMR.

LC proteins is known to exist as dimers and proved by many researchers in past few years[81,197]. Using solution state NMR, we confirm that both patient and germline proteins are in equilibrium between monomer and dimer state. At lower concentration, equilibrium is shifted towards monomer and vice versa. We also confirmed the existence of dimer by cross-linking assay using SDS PAGE together with the presence of small percentage of oligomeric state at lower concentration. In S3706 protein from NMR concentration dependent assay, it was revealed that both canonical and non-canonical is possible. X-ray crystal structure of patient protein was found to be canonical, but it only fits with NMR results with slight modification in G-strand that needs to be extended to fulfill the highest chemical shift changes of G102 residue. Other than canonical structure, altered dimer structure with 180° rotation is also possible with our NMR results. This also holds true for germline protein, as the chemical shift perturbations are similar as patient protein. The high K_D value in both patient and germline suggest that the affinity of dimer association is very weak and were not significantly different.

Our main and interesting finding is the oligomeric intermediates using NMR technique, formed during aggregation pathway, which was specifically observed in patient. The respective germline protein does not form any oligomeric intermediates but still aggregates eventually. Single point mutant S3706_R50G shows intermediate behavior between patient and germline protein. The intermediates formed only in S3706_patdel also reflect the unstablity of protein, which was also revealed in CD data. More than one oligomeric intermediate from NMR aggregation assay distinctly suggest the possibility of polymorphism of fibrils through selection of stable intermediate. One tryptophan W35 in the protein sequence, gives side chain NMR resonance around 9–10 ppm and thus it gives easily the direct evidence of the more than one intermediate in S3706_patdel while no intermediates in S3706_GL. Through TEM and DLS experiments we confirm the presence of large oligomers in S3706_patdel protein.

Combining aggregation and concentration dependent monomer-dimer data extracted from solution state NMR, a pictorial representation is proposed connecting the link between both processes. As we proceed from lower to higher concentration, monomer-dimer equilibrium is shifted towards dimer. Monomeric protein promotes the aggregation process while dimer is protective and thus must convert to monomer to follow aggregation. Several oligomeric intermediates observed during aggregation can be explained either by polymorphism or as a series of changes in structural level i.e. transformation from one form to another. Out of many intermediates only few by competition leads to form stable amyloid fibrils. **(Figure 36**)

The reason of the difference observed during aggregation process between S3706_patdel and S3706_GL is still not well understood but we confirm that single mutant S3706_R50G changes the protein stability as well as aggregation pattern and thus behaves more like germline protein. The important positively charged arginine at 50^{th} position is explained in its fibrils state in next chapter using solid state NMR.

As proteolysis is suspected to be an important factor in LCs fibrils formation, it was essential to check the degradation of protein during aggregation step. SDS-PAGE reveals clearly that there was no degradation of protein in patient as well as in germline protein. Instead we observed disappearance of monomer bands reflecting the formation of larger molecular weight species. From the Mass spectrometry we also confirm that there is no degradation of protein during aggregation process.

Assignment of oligomers reveals the existence of more than one intermediates. One of the important aspects of the LCs oligomers is the toxicity, which has gain lot attention among researchers recently[89,197]. For amyloids it was proposed that oligomers could be cytotoxic than its fibrils state. It will be interesting to know cytotoxicity for our protein, S3706_patdel oligomers that was trap during aggregation process. Another significant question for future is the difference in toxicity of S3706_patdel and S3706_GL fibrils state.

Recently, it has been shown the importance of constant domain in LC and how it can affect the protein dimer properties and stability[198]. It will be interesting to know how the S3706 V_L protein changes on addition of constant domain, as full antibody domain. Moreover, it will be exciting to know the role of extra amino acids at C-terminal that does not belong to variable domain, which is usually seen in patient fibrils sequences and not well understood. In summary, for the first time we provide a direct evidence of oligomers by solution state NMR. Our findings will assist the proper understanding the mechanism of LCs fibril formation.

3.2 Binding of S3706 V_L protein with EGCG

3.2.1 Aim of the project

Green tea polyphenol, (-) epi-gallocatechin 3-gallate (EGCG) is known to inhibit the fibril formation in LC fibrils by the formation of insoluble aggregates[199]. Here, our aim is to analyze the interaction of EGCG with patient derived protein and to understand how it is different to its respective germline protein.

3.2.2 EGCG titration

To probe the interaction of EGCG, we added 5-fold, 10-fold and 20-fold excess of EGCG solution to S3706_patdel and its germline protein S3706_GL. Like all the amyloid both S3706 proteins show precipitation and the solution became turbid. Overall, there were less significant chemical shift changes seen, when we overlay ^{1}H-^{15}N HSQC with 10-fold excess EGCG of both S3706 proteins to its non-EGCG protein. **(Figure 37a)** To analyze accurately, we plotted chemical shift perturbation on addition of 10-fold excess EGCG in S3706_patdel along with its residues. Although the changes are subtle with largest perturbation observed 0.02 ppm, interestingly the changes are localized around C' and G strands. On closer analysis it seemed that 51K, 45V has chemical shift changes that are consistent with different addition of molar excess of EGCG. We also observed that CSPs are localized in places near to proline residue at P9, P40, P44, P45 and P60. However, the CSPs are subtle. When we checked the normalized intensity plot of the same spectra, only 60% of intensity remained and there were no unusual changes seen. For germline we did not show the data since the same effect as in patient.

We also performed time course EGCG titration with 5-fold, 10-fold and 20-fold excess of EGCG in both patient and germline protein. The ^{1}H-^{15}N HSQC spectra were recorded till five days for both protein with all folds of EGCG. During experiments the protein was kept at 25°C with 250 rpm agitation in NMR tube. The experiments were done in triplicates to avoid mistakes and reproducibility. To analyze the data, we used average intensity where 58I, 83E, 16G, 14A and 109L residues are included for both proteins. As expected S3706_patdel show rapid reduced intensity on addition of 20X EGCG in comparison to 10X and 5X fold excess of EGCG. When compared with its germline protein, in S3706_patdel the intensity was reduced quickly after 1 day. 5-fold excess shows that in germline 50% of intensities are still there at 5^{th} day of experiment while in case of S3706_patdel the intensity disappears on 4^{th} day. These results indicate that patient protein precipitates faster than its germline protein.

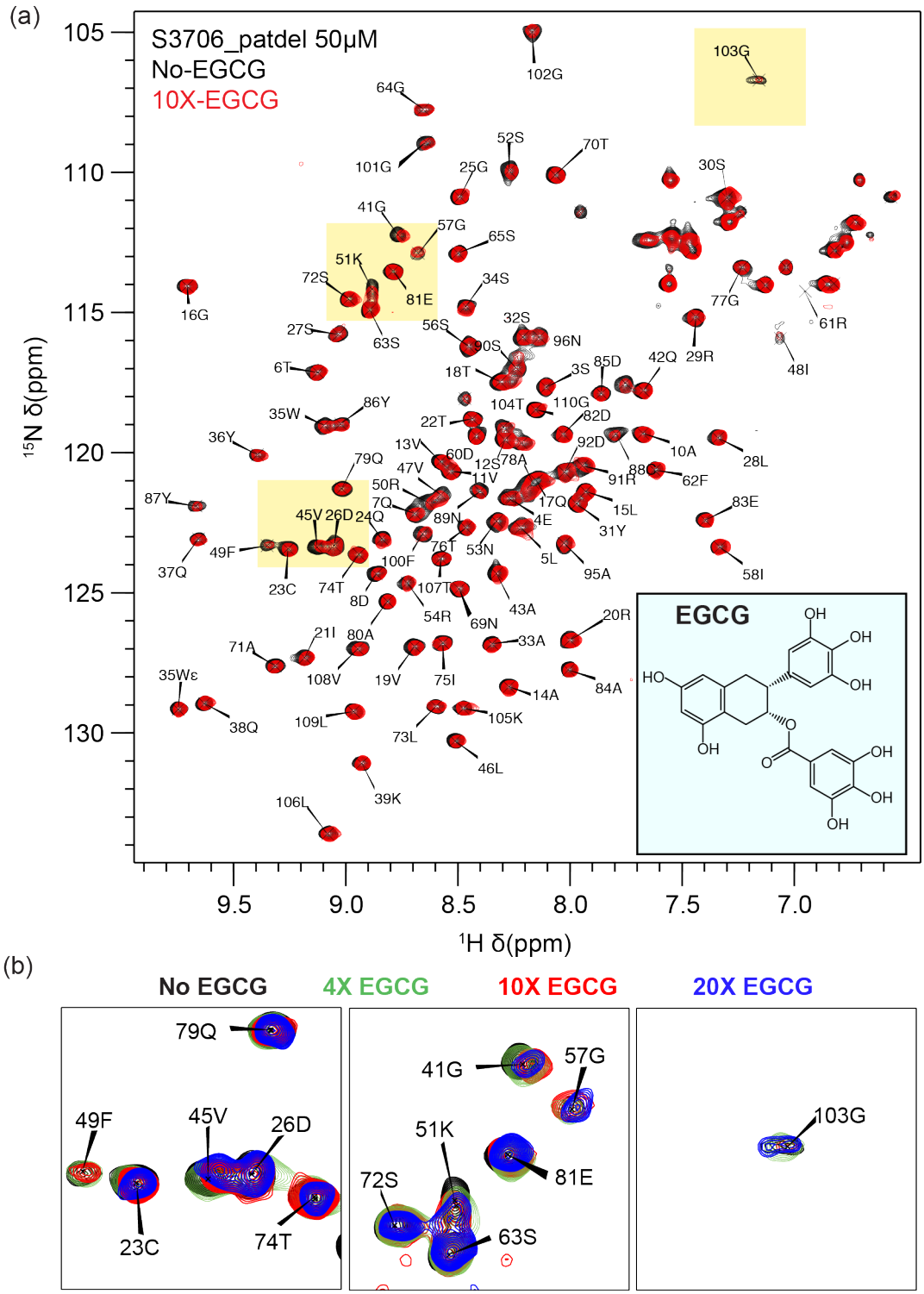


Figure 37: Interaction of EGCG with S3706_patdel.

(a) ^{1}H-^{15}N HSQC spectra overlay of S3706_patdel with and without EGCG. (b) Zoom plot of S3706_patdel upon addition 5-fold excess (green), 10-fold excess (red) and 20-fold excess (blue) of EGCG.

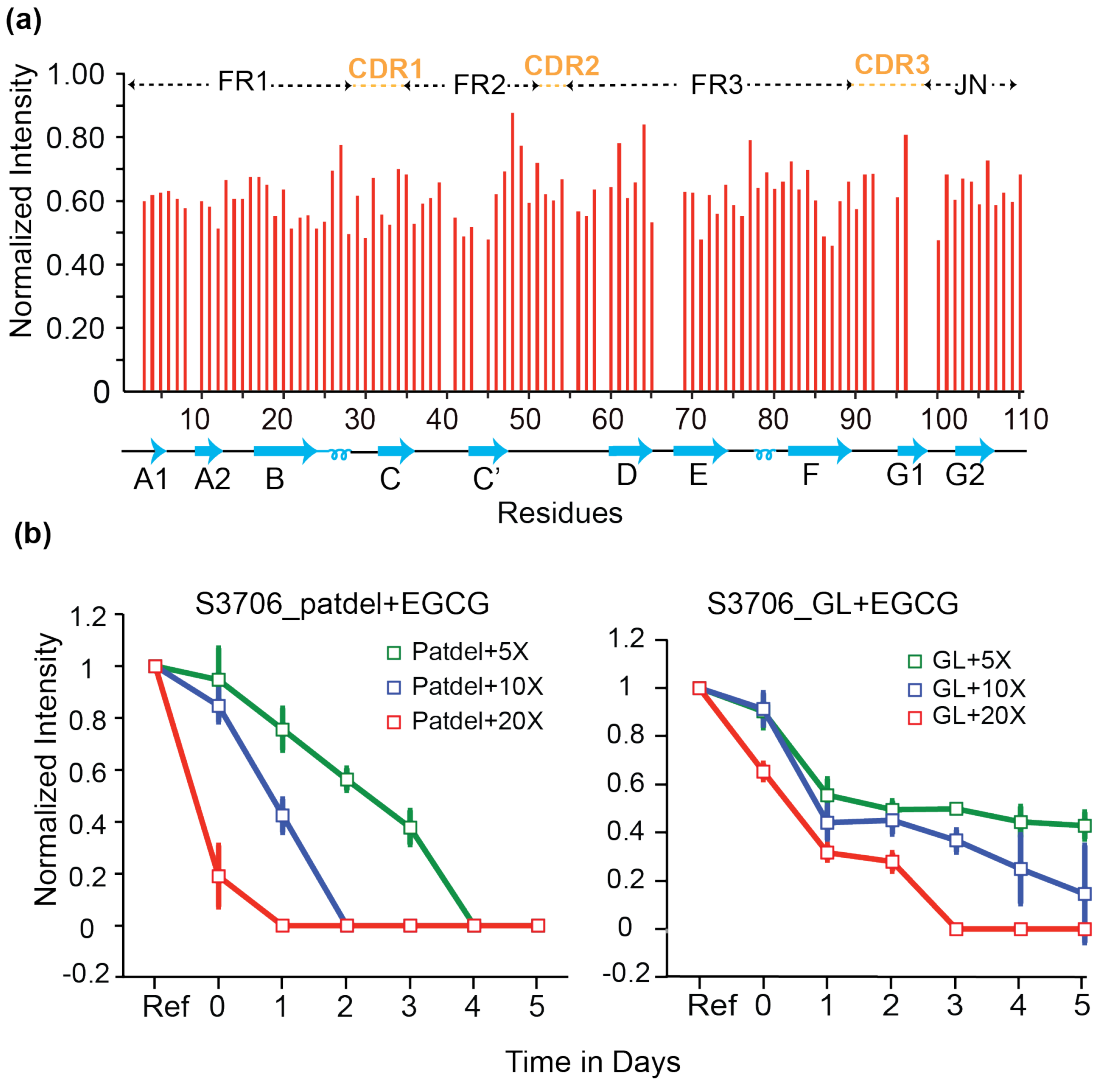


Figure 38: Normalized intensity plots of S3706 V_L protein.

(a) Normalized chemical shift intensity plot of S3706_patdel from of ^{1}H-^{15}N HSQC spectra on addition of 10X EGCG shows overall reduced intensity. (b) & (c) Normalised intensity plot of time dependent EGCG titration in S3706_patdel and S3706_GL on addition of 5-fold, 10-fold and 20-fold molar excess of EGCG. Here is the average of peak intensity of few residues including 58I, 83E, 16G, 14A and 109L for both proteins. Ref is the reference or starting point when no EGCG was added to the protein. S3706_patdel shows higher amyloidogenicity than S3706_GL.

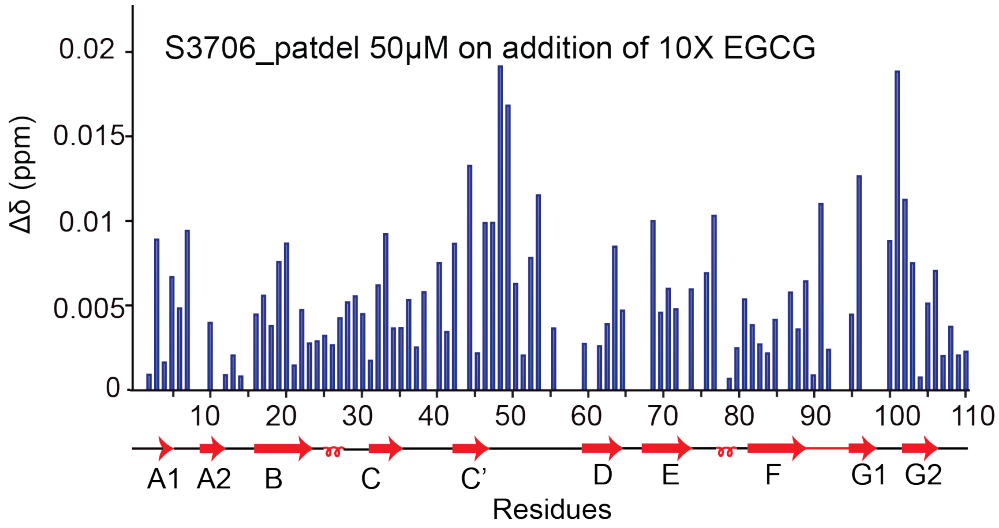


Figure 39: Chemical shift perturbation plot on addition of 10-fold excess of EGCG

3.2.3 Discussion and Conclusion

Previously, our group reported that EGCG binds with more amyloidogenic protein to form insoluble precipitates and proline at 44 and 59 position in MAK33 are the binding sites[199]. In a similar direction, we also proposed that EGCG selects and binds preferentially to protein that is more amyloidogenic. It induces precipitation to form insoluble, unstructured aggregates in both germline and patient, but the kinetics is faster in more amyloidogenic protein.

Although P44 and P59 residues are conserved in patient derived S3706 protein but unlike MAK33 we observed subtle CSPs on EGCG addition. A subtle change in strands was observed around C' and G strands and it is interesting to note that P44 is located in C' strand. There were large changes observed near P59 residue. Due to very little changes seen in chemical shift perturbation it is difficult to say undoubtedly that EGCG binds with P44. The slower kinetics in EGCG titration in germline indicates that this sequence with 5 amino acids changed makes it less amyloidogenic. EGCG distinctly distinguish the protein according to the sequence and binds favorably to form unstructured precipitates and thus inhibit fibril formation. Due to the natural availability and non-toxicity of EGCG, it is a good candidate for inhibiting the amyloidosis disease. Clinical studies are going on in AL amyloidosis patients to detect the effectiveness as a therapeutic agent. Recently, it has been shown have positive effect in cardiac amyloidosis[138]. However, in phase II clinical study under TAME-AL, the EGCG effect was shown to be insignificant that reflects issues like bioavailability and instability[139,140]. In future more laboratorial and clinical experiments on EGCG will expand the knowledge and development of therapeutics strategies.

4 Solid-state NMR investigation of antibody V_L domain

4.1 Aim of the project

AL amyloidosis disease is caused by light chain fibrils deposition especially variable domain in organs. For better understanding of the disease, it important to know the structure fibrils formed by light chains. Recently, atomic structure of light chain fibrils from patient is published using cryo EM technique where they show the oxidized cysteine pair that stabilizes the amyloid fibrils[200]. Solid state NMR technique is in initial stage to study the LCs amyloid fibrils. Until now nothing is published regarding the structure of LC fibrils using solid state NMR. Few groups has published the assignment of fibrils, mouse LC MAK33 (κIV subgroup)[94], patient derived LC AL09 (κI subgroup)[201,202] and very recently 6aJL2-R24G (λVI subgroup) reported in-register parallel cross β-sheets structure, using solid state NMR. It was previously claimed the secondary structure is retained by fibrils state as monomer by domain swapping mechanism[92] that was overruled later[94,203]. Only these three LC proteins out of numerous sequences till now are studied and published within 2-3 years using MAS solid-state NMR technique as a main technique. The research in LC fibrils is quite new and more effort towards structure is needed for identification of core in fibrils structure as the sequence in AL amyloidosis patients has high degree of variability.

In this light, our aim is to characterize patient derived S3706 (λIII subgroup), light chain fibrils and to identify the principal interactions that might be involved in stabilizing the amyloid fibril formation by solid state NMR. We prepared different samples of fibrils of S3706_patdel and compared with its germline S3706_GL and single point mutation mutant S3706_R50G. We want to characterize the main features of fibrils core and degree of variability amongst S3706 variants.

4.2 Results

4.2.1 S3706 light chain fibrils preparation

Isotopically ^{13}C-^{15}N labelled fibrils were prepared described in method section **2.2.2.6** for solid-state NMR measurement. Sensitivity and resolution are two major factors that governs the spectral quality in solid-state NMR. In order to gain maximum sensitivity, we packed 3.2 mm rotor and for better resolution, seeds were used to obtain homogeneous fibrils. As discussed before, this sequence was obtained from patient suffering from AL amyloidosis and whose heart was transplanted, which is named as S3706. The amyloid fibrils was extracted from S3706 patient heart fibrils[195] and these were used as seeds to prepare solid state NMR samples in order to mimic same fibrils molecular structure as in patient[204].

S3706 fibrils were characterized by ThT assay and TEM measurement. ThT assay

protocol is discussed in detail in section **3.1.6.3** previously. TEM images of S3706_pat *ex-vivo* and S3706_patdel fibrils prepared from with *ex-vivo* seeds show well mature fibrils depicted in **Figure 40**.

Reproducibility is the huge issue in amyloid fibrils due to polymorphism phenomenon, therefore we prepared two samples S3706_patdel_01 and S3706_patdel_02, both seeded with *ex-vivo* material using exactly same preparation method. In order to understand the effect of seeding we prepared another sample with no seeds. Furthermore, we prepared two more sample of S3706_GL and S3706_R50G both seeded with *ex-vivo* material, so that it adopts same molecular structure as patient and the changes are caused only due to the mutational effect.

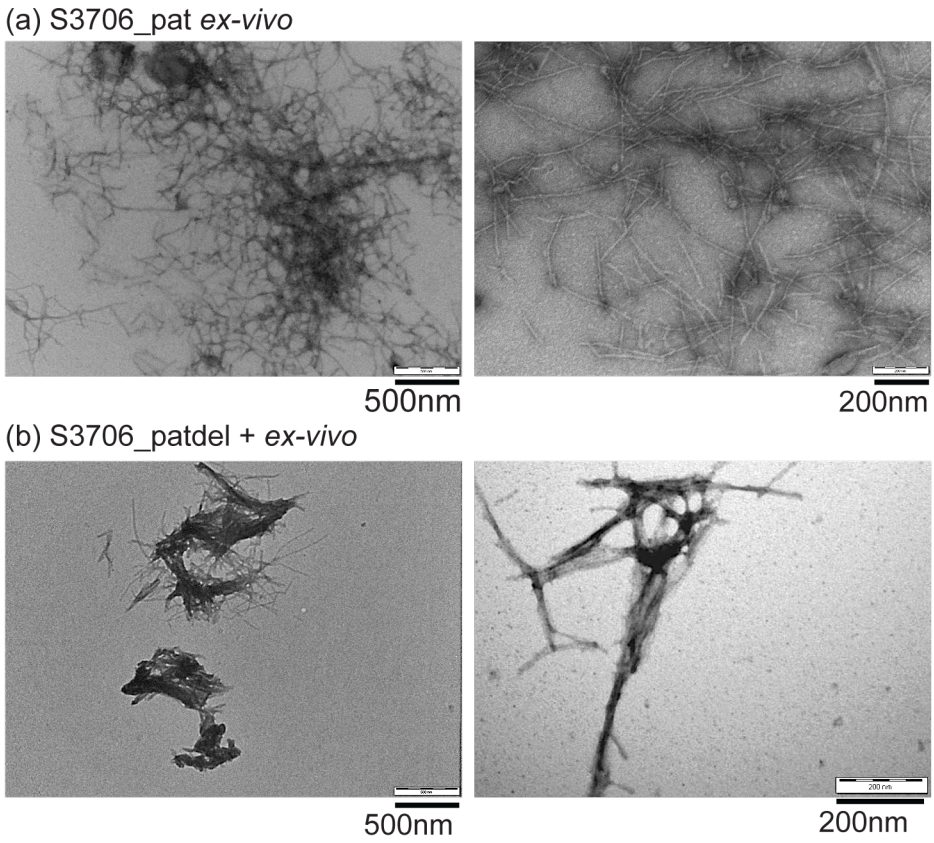


Figure 40: S3706 Fibrils image by TEM measurements

(a) [*]S3706_pat ex-vivo patient fibrils extracted from the amyloids deposited in heart[193]*. (b) ^{13}C-^{15}N S3706_patdel fibrils prepared by seeding ex-vivo material as seeds.*

Another curiosity in terms of fibrils preparation was what if we seeded with *in-vitro* fibrils (own seeds), hence we prepared sample of S3706_patdel and S3706_R50G each

[*] Kindly provided by Dr. Karthikeyan Annamalai from Prof. Dr. Marcus Fändrich group.

seeded with their own *in-vitro* fibrils. All the condition was same for both samples as mentioned in *Materials and method* section **2.2.2.6**.

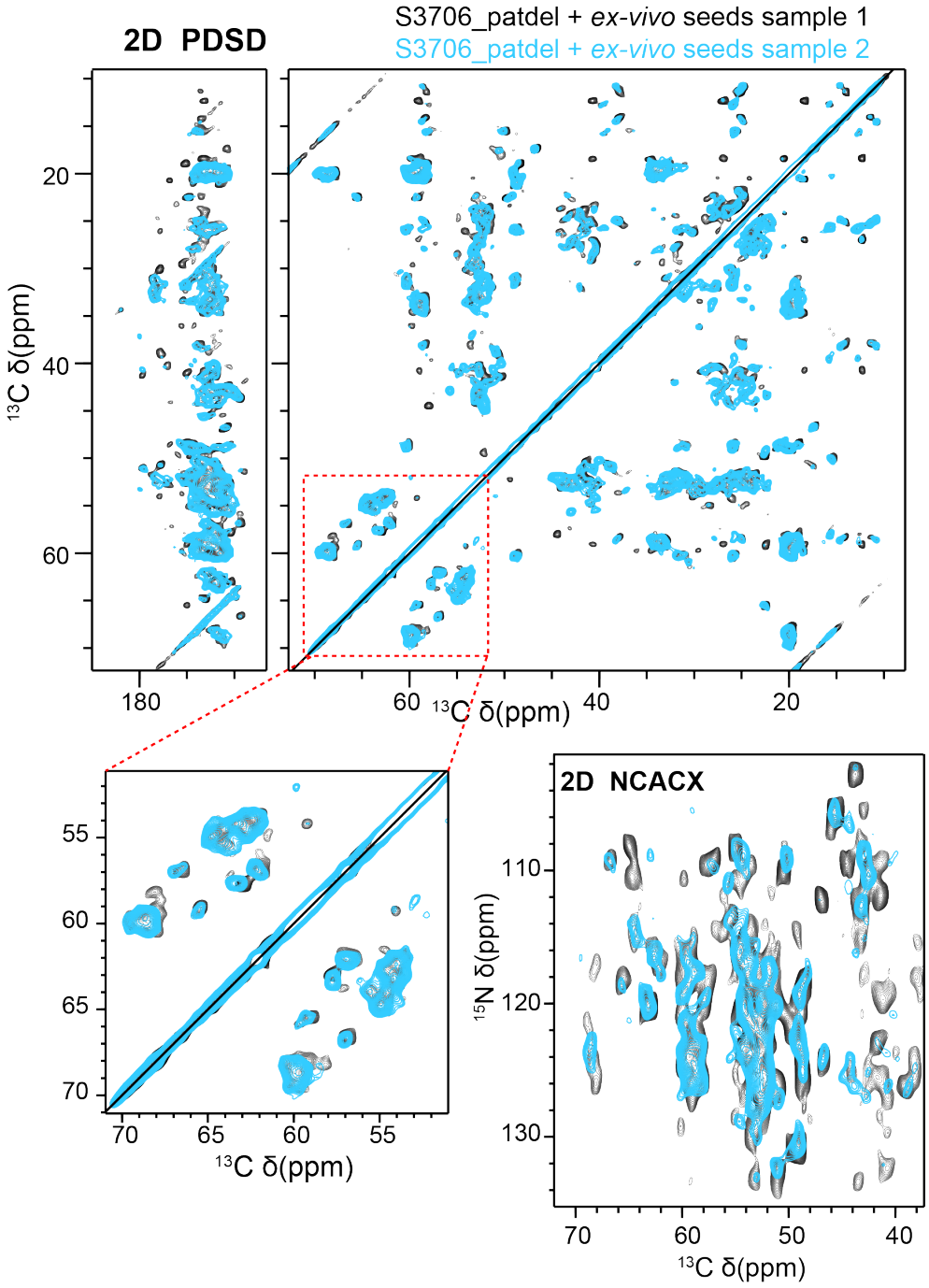


Figure 41: Reproducibility in S3706_patdel fibrils

2D PDSD and 2D NCACX spectra overlay focusing on Ser/Thr region and Cα-region respectively, in two samples of S3706_patdel. Both protein fibrils were seeded with ex-vivo material and using same preparation method.

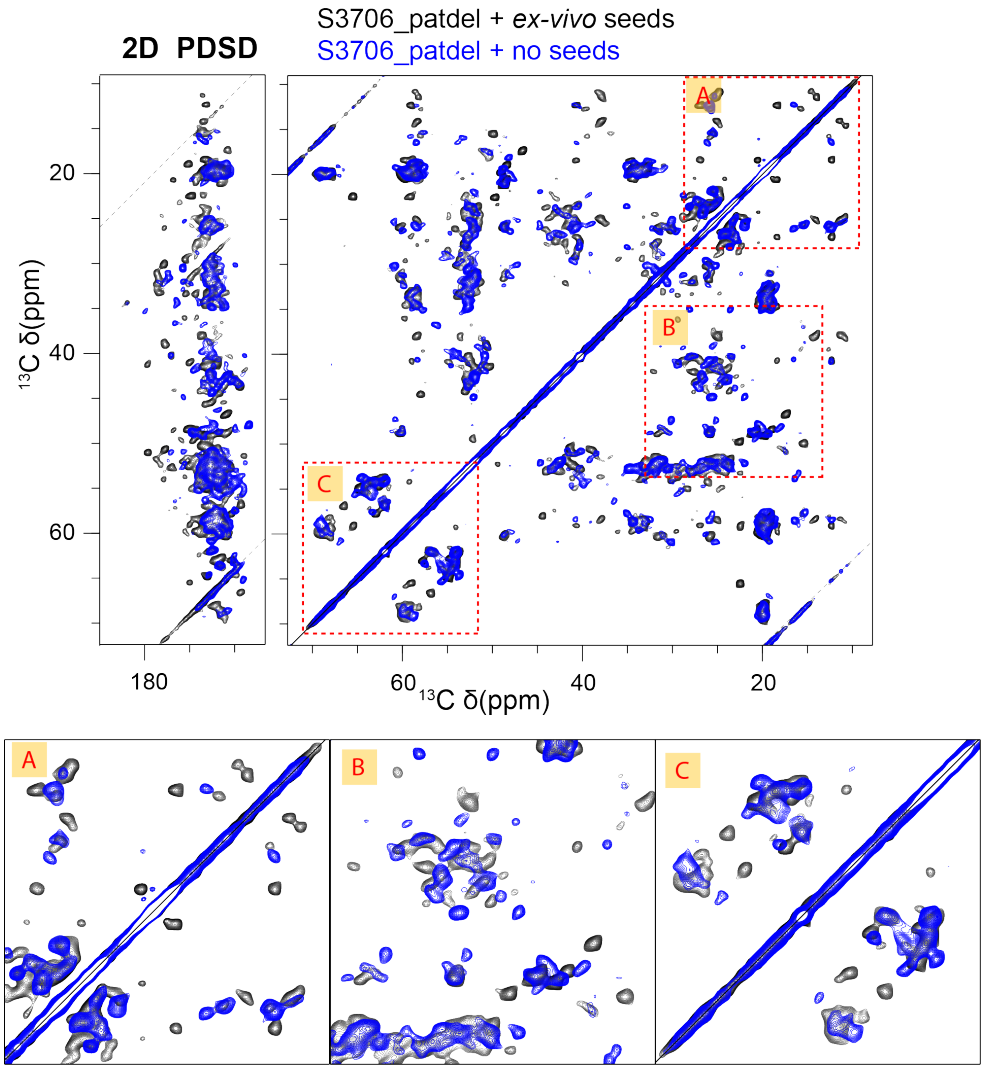


Figure 42: 2D PDSD superposition depicting the effect of seeding in S3706_patdel

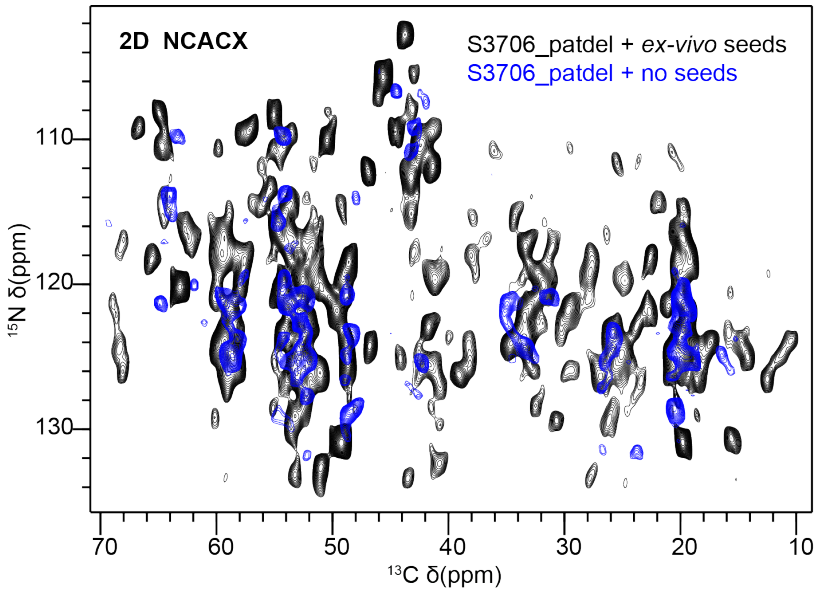


Figure 43: 2D NCACX superposition depicting the effect of seeding

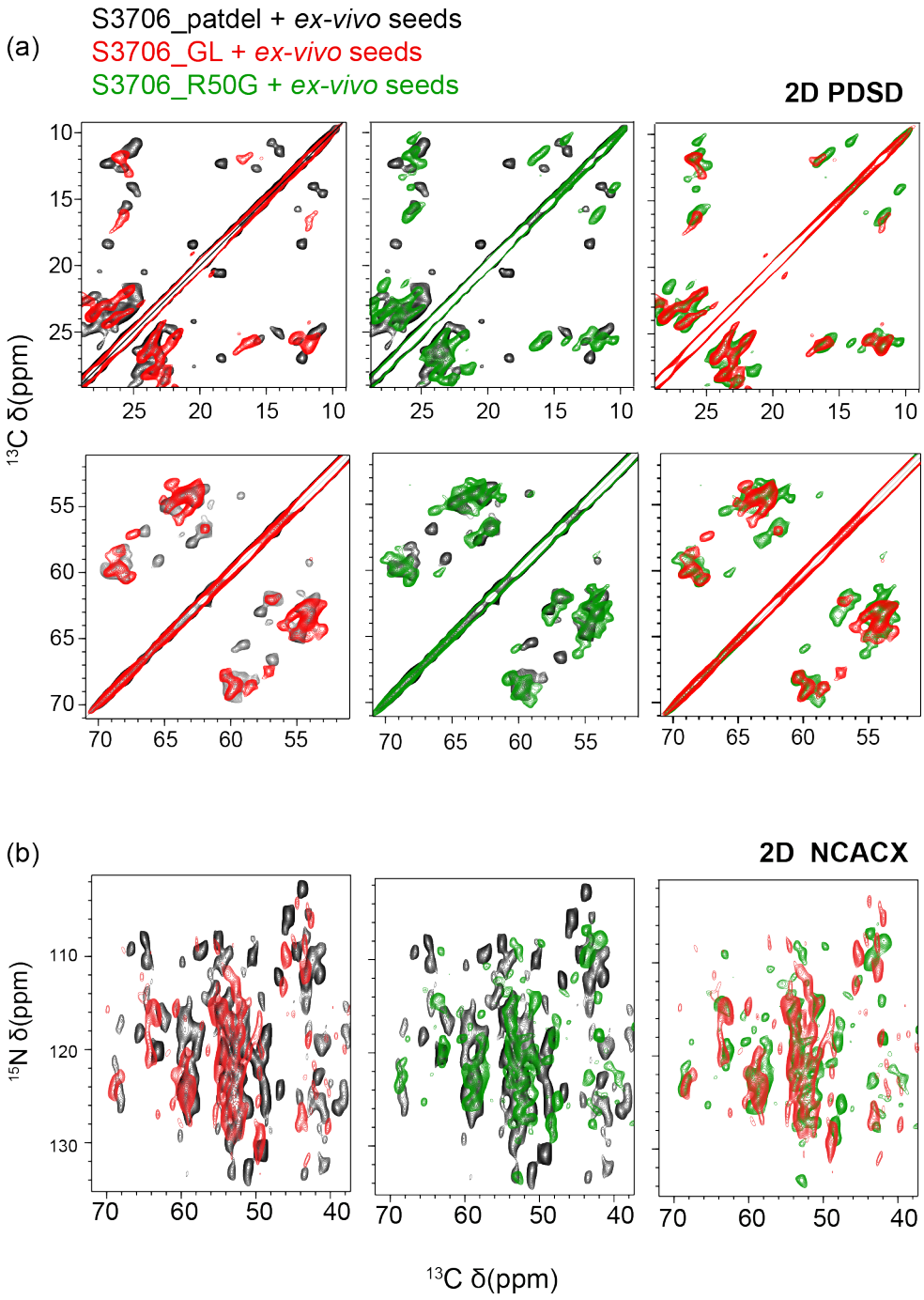


Figure 44: Comparison of 2D PDSD and 2D NCACX in S3706 variants.

(a) Superposition of Iso/Val and Ser/Thr region from ^{13}C-^{13}C correlations in S3706 variants. (b) Superposition of Cα-region from ^{13}C-^{15}N correlations in S3706 variants. S3706_ patdel, S3706_GL and S3706_R50G are colour coded with black, red and green respectively. All the samples were seeded with ex-vivo seeds that are extracted from heart fibrils of AL amyloidosis patient.

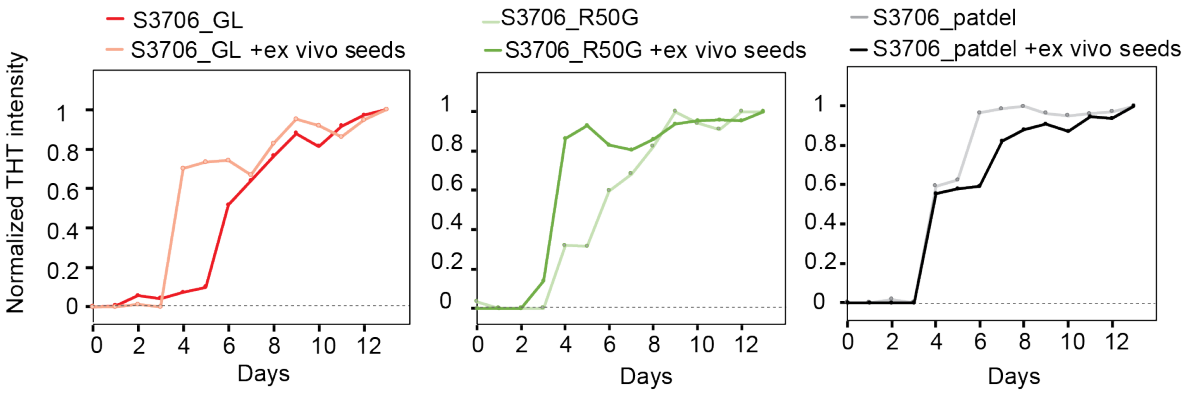


Figure 45 ThT assay for S3706_patdel

Comparison of with and without ex-vivo seeds in S3706 variants using ThT assay. 50 µM protein with 25 µM ThT dye were used during assay and samples were agitated at 120 rpm at 37°C. All the samples were subjected to same condition.

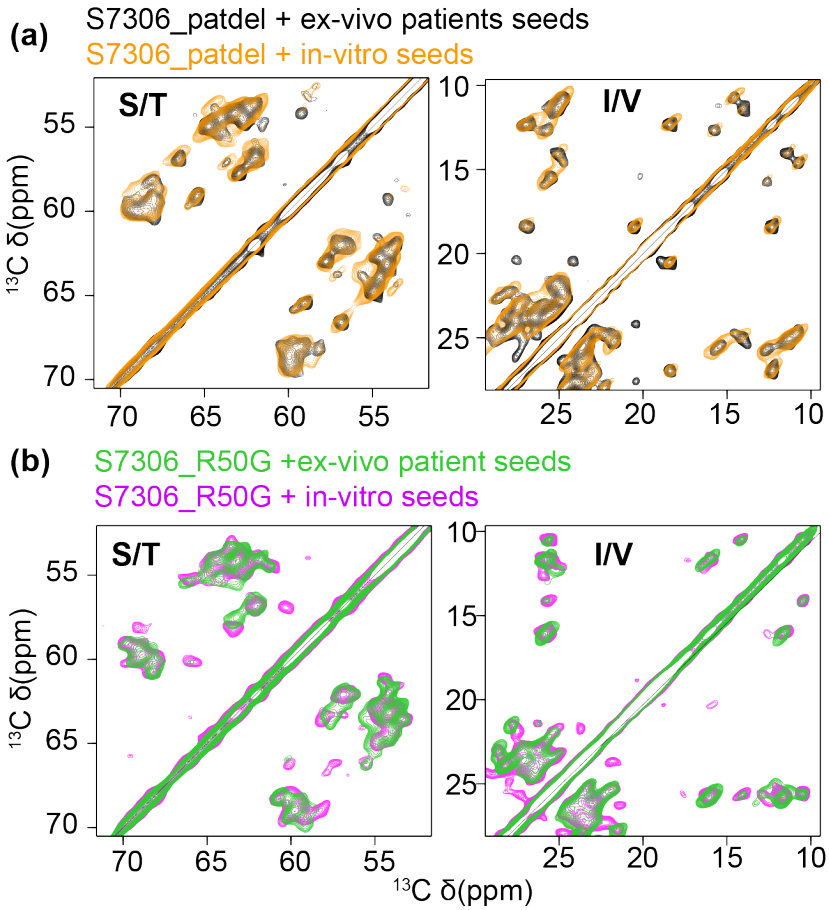


Figure 46: Comparison of ex-vivo and in-vitro seeded samples in S3706 variants. *(a) & (b) Superposition of ex-vivo patient seeded and in-vitro seeded fibrils of S3706_patdel and S3706_R50G in black/orange and green/pink colour respectively. In both samples, the perfect overlapping between the two spectra indicates regardless of type of seeds, S3706 populates same polymorph.*

4.2.2 Comparison of S3706 light chain fibrils variants by Solid-state NMR

We recorded ^{13}C-^{13}C correlation spectra for all the fibrils samples as the first measurement which typically solid-state NMR do for analyzing the amyloid fibrils and can be considered similar as ^{1}H-^{15}N HSQC spectra of protein in solution state NMR. This correlation conveys the sensitivity, quality of fibrils in the rotor. We also recorded ^{13}C-^{15}N correlations to check the sensitivity in N-dimension for all samples. All the spectra for comparison were recorded at 10KHz MAS using same NMR parameters at 750MHz spectrometer at 273K. For all samples, 50 ms mixing time, and 80kHz decoupling in ^{1}H -dimension were used to record 2D PDSD and 2D NCACX spectra.

For reproducibility, on comparing S3706_patdel sample 1 and 2 seeded with *ex-vivo* material, we obtain exactly same spectra in 2D PDSD and 2D NCACX. This indicates that we populate same stable polymorph every time if we use same seeds. In the sample 2, the fibrils material was less, which account for loss peak intensity. It is also possible that the peaks missing are from N-terminal or C-terminal region that are dynamic in nature. **(Figure 40)**

The peaks were narrow with linewidth of 100 Hz in 2D PDSD spectra. Comparison of seeded with non-seeded samples shows very large shifts of peaks in 2D PDSD. In 2D NCACX for this sample, the sensitivity was very low due to heterogeneity in fibrils caused by polymorphism**. (Figure 41)**

We compared 2D PDSD of S3706 variants, S3706_patdel, S3706_GL and S3706_R50G that are all seeded with *ex-vivo* material. Single mutant S3706_R50G mutant spectra show the similarity for both patient and germline. For illustration of above discussed comparison, we focused on Iso and Ser/Thr region from 2D PDSD as these amino acids has a specific non-overlapped chemical shift region. **(Figure 44a)** We also compared the Cα-region from 2D NCACX region for S3706 variants. The peaks overlapping in the spectra from S3706 variants suggest the possibility of common core in fibrils.

To examine the difference between patient fibrils and recombinantly produced *in-vitro* fibrils, we compared S3706_patdel and S3706_R50G sample seeded with *ex-vivo* patient fibrils with the samples seeded with *in-vitro* fibrils. In both 2D PDSD and 2D NCACX, the spectra were found to be similar for both samples. The two regions from 2D PDSD focusing on Ser/Thr and Iso/val chemical shift are depicted in **Figure 46**. The perfect superposition of peaks in both fibril samples indicates that regardless of *ex-vivo* from patient or *in-vitro* seeds, similar polymorph is populated during seeding in fibril formation. For further investigation, we used only the samples that are seeded with *ex-vivo* patient material as these are closer to patient fibrils and has potential to contribute the most significant information.

We also performed the ThT assay in S3706_variants to see the effect of seeding. Although the lag period is still seen all the variants but from solid-state NMR experiments, the

fibrils were found to be homogeneous. **(Figure 45)**

4.2.3 Electrostatic interaction in S3706 fibrils

Electrostatic interactions or salt bridges in amyloid fibrils are important in stabilizing the protein, fibrils formation and its stability[205]. Salt bridges are formed by non-covalent interaction from the oppositely charged residues that are close to each other in space. It mostly arises from the cationic ammonium (RNH_3^+) from lysine or guanidinium ($RNHC(NH_2)_2^+$) from arginine and the anionic carboxylate $RCOO^-$ of either aspartic acid or glutamic acid.

From 2D NCACX spectra, we observe all the correlation from lysine (Nξ) and Arginine side (Nε) chain correlation to all aliphatic ^{13}C atoms. In S3706_patdel, out of three Lys, for one Lys we see the entire cross peaks generated from Nξ-C correlation at ~35 ppm with very good resolution. In addition, for Arg out of six in sequence we observed only one to show the entire cross peaks at ~80 ppm, generated from Nε-C correlation. **(Figure 47b)** This observation prompted us to run the experiment to examine salt bridge connection in S3706 fibrils.

We performed 2D TEDOR experiments to acquire ^{13}C-^{15}N correlations with larger spectral window to observe the positively charged lysine and arginine side chain cross peaks in S3706_patdel, S3706_GL and S3706_R50G fibrils samples. Two 2D TEDOR spectra of each sample was recorded using short (1.9ms) and long mixing time (15ms) to observe the electrostatic interaction in fibrils. All the experiments for were acquired using 16.5 kHz MAS in 750MHz spectrometer at 273K.

We recorded the short mixing 2D TEDOR in S3706_GL and single mutant S3706_R50G to probe the lys and arg side chain. In lys region, prominent chemical shift change was observed in both variants in comparison to S3706_patdel. Since, in S3706_R50G mutant the mutated R50 is adjacent to K51, the two lysines, which show largest chemical shift perturbation can be K51 and K39. In Arg side chain, the spectra were quite different so only superposition leaves some ambiguity in for the assignment.

In S3706_patdel from short mixing 2D TEDOR experiment we observed three nicely resolved Nξ-Cε cross peaks for all three lysine residues from sequence. Out 6 arginine's from sequence, we observed two strong and two weak Nε-Cδ cross peaks. **(Figure 47a)** For long-range correlation, 2D TEDOR with 15 ms mixing was recorded. From lysine and arginine, we observed one peak at ~180 ppm indicating the salt bridge connection. We mapped the new peak from long mixing 2D TEDOR to 2D PDSD carbonyl region to obtain the salt bridge connection from Lys to Asp ($N\xi^+ \ldots C\gamma^-$) and Arg to Asp ($N\varepsilon^+ \ldots C\gamma^-/C\delta^-$). For both lysine and arginine, it is undoubtedly the Asp side chain that makes the salt bridge. **(Figure 47b)** These results indicate that the two salt bridges observed in S3706_patdel can stabilize the amyloid fibrils.

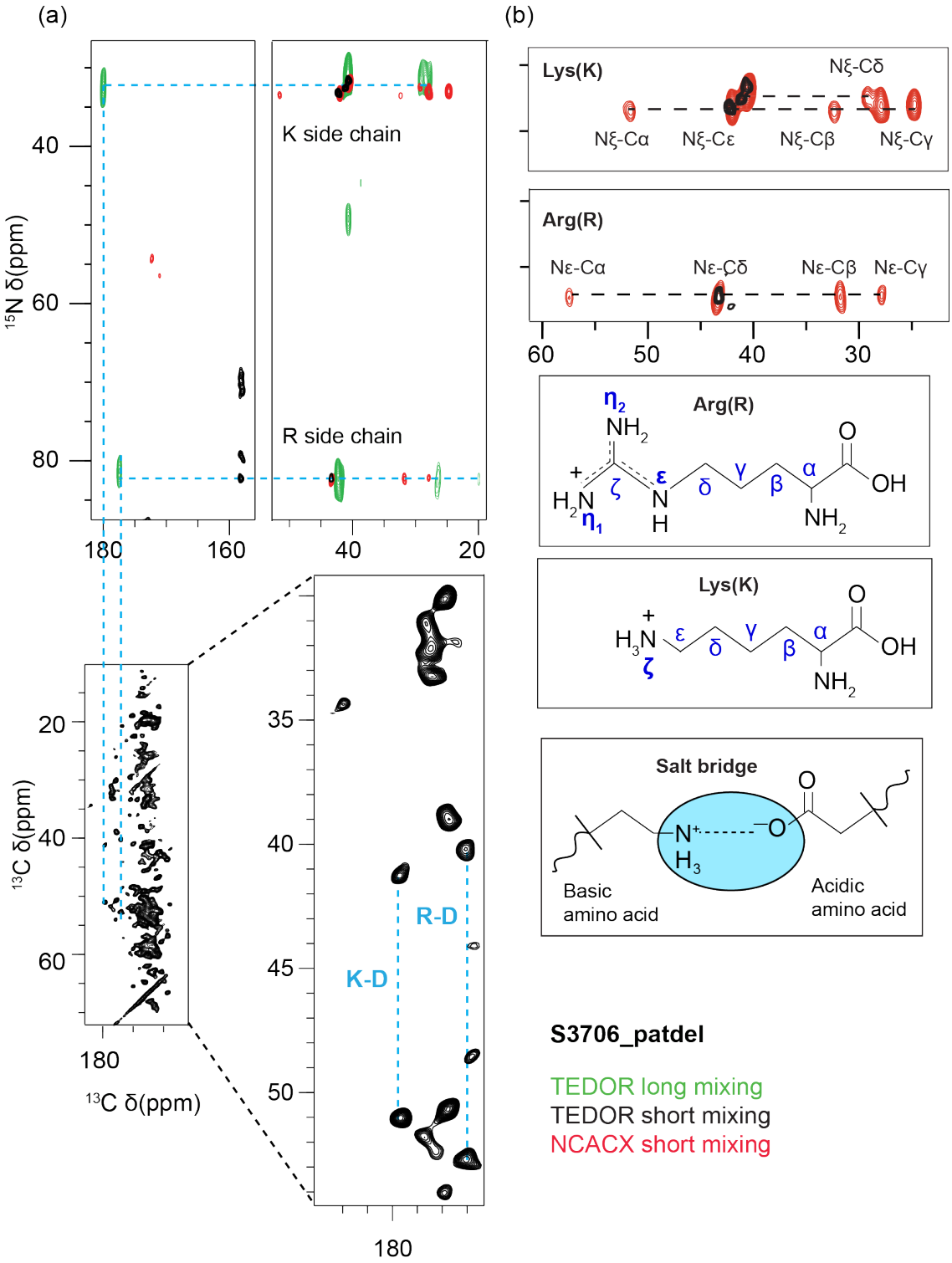


Figure 47: Long range contact in S3706_patdel by 2D TEDOR experiment.

(a) Superposition of long (green) and short mixing (black) 2D TEDOR experiment focusing on side chain in S3706_patdel. 2D NCACX (Red) is overlaid with showing all the ^{13}C-^{15}N shift of Lysine and Arginine side chain. In S3706_patdel, patient sequence shows long range contacts depicting salt bridge between positively charged lysine and arginine with negatively charged aspartic acid and glutamic acid marked with blue line. (b) Superposition of S3706_patdel 2D NCACX (red) and 2D TEDOR shot mix (black) depicting the ^{13}C-^{15}N correlation of lysine and arginine side chain along with the amino acid structure.

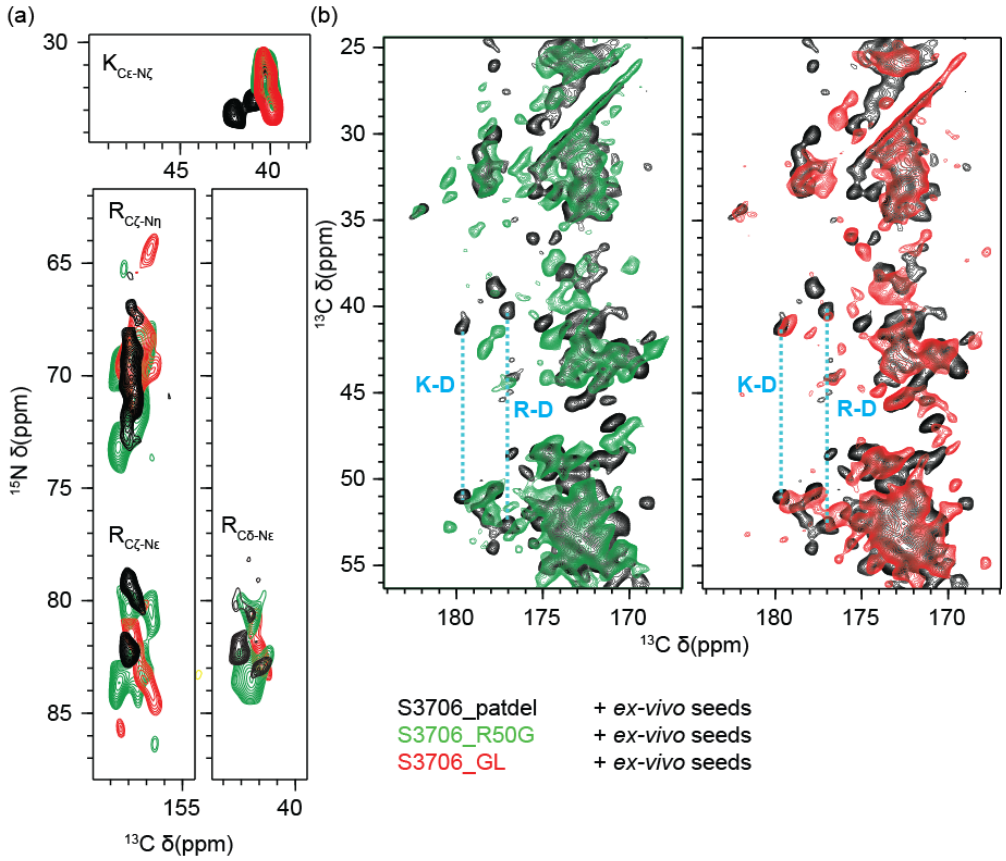


Figure 48: Comparison of salt bridge donor and acceptor in S3706 variants

(a) Comparison of Lys and Arg side chain in S3706 variants using ^{13}C-^{15}N correlations. (b) Comparison of Asp side chain involved in salt bridge connection in S3706 variants using ^{13}C-^{13}C correlations.

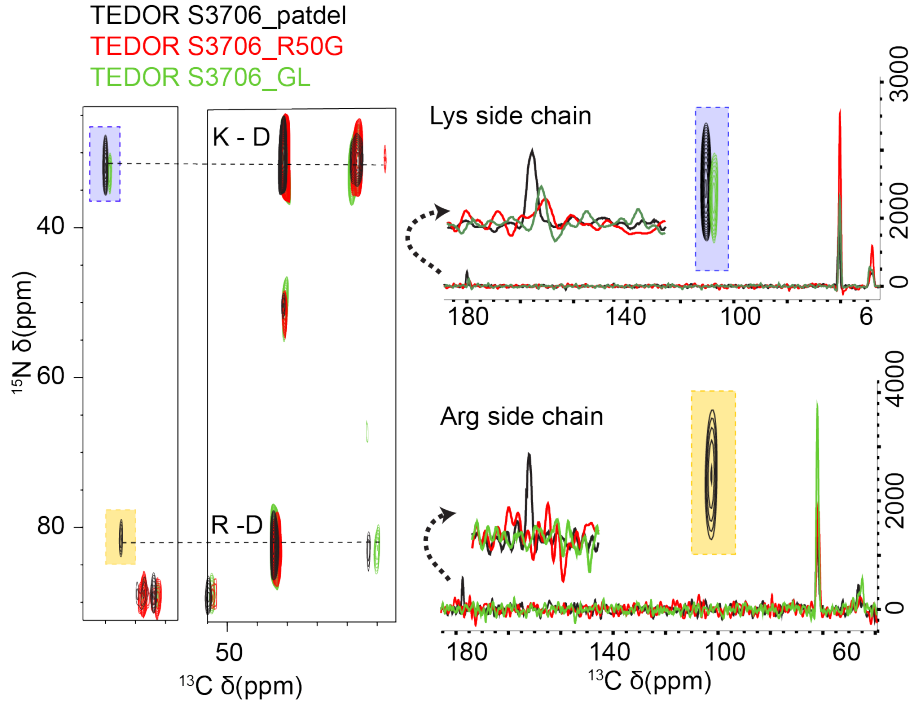


Figure 49: Electrostatic interactions in S3706 variants.

2D TEDOR overlay of S3706_patdel with S3706_GL and S3706_R50G, focusing on the presence of

long-range contacts. Right hand side is the 1D row extracted from 2D TEDOR illustrating the salt bridge connection between cationic ammonium (RNH_3^+) from lysine and the guanidinium ($RNHC((NH_2)_2^+$) from arginine to carboxylic group of aspartic acid. S3706_patdel shows the strong salt bridge connection while in S3706_R50G there is no connection observed. A very weak connection from lysine to carboxylic group was observed in S3706_GL.

In order to probe the salt bridge in germline, S3706_GL and single mutant S3706_R50G, we compared with long mixing 2D TEDOR experiment with patient sequence S3706_patdel. Interestingly, in S3706_GL the salt bridge connection is lost in Arg and very weak connection is observed from Lys side chain. The more remarkable observation was in single mutant S3706_R50G, where there was no connection observed. We extracted 1D projection from Lys and Arg row from 2D TEDOR long mixing and compared all three constructs. This comparison also shows that patient fibrils possess two salt bridge connections while the connection is either weak or broken in S3706_GL and S3706_R50G. **(Figure 49)**

4.2.4 Solid state NMR assignment of rigid residues in S3706_patdel fibrils

Unlike solution state NMR, the assignment experiments are not well established and needs many complementary experiments along with thorough optimisation. Due to low sensitivity in ^{13}C-dimension and large linewidth makes the sequential experiment difficult. The detail of the experiments in discussed in section **2.3.2.2.**

Using solid state NMR, we assigned rigid residues from S3706_patdel fibrils seeded with *ex-vivo* material. As these fibrils can mimic the *in-vivo* patient fibrils makes this study most exciting. In order to get sequential assignment of the rigid residues in fibrils, we recorded ^{13}C detected resonance assignment experiments, which comprise of 3D NCACX and 3D NCOCX. In assistance we also recorded 3D CONCA, 2D NCACX and 2D NCOCX as a complement experiment. Typically, 3D experiments for sequential assignments are used but due to large linewidth in ^{15}N- dimension and limitation in indirect dimension acquisition time makes the assignment difficult. 2D experiments with large number of scans along with more acquisition time in N-dimension makes the spectra well dispersed and thus assist the sequential assignment.

We assigned 40 residues out of 60 peaks that are observable in Cα region of S3706_patdel. The assignment of residues is shown in 2D PDSD aliphatic region and 2D TEDOR Cα region. 2D strips were extracted from 3D NCACX and 3D NCOCX to illustrates the sequential assignment of S3706_patdel fibrils rigid core. **(Figure 50)**

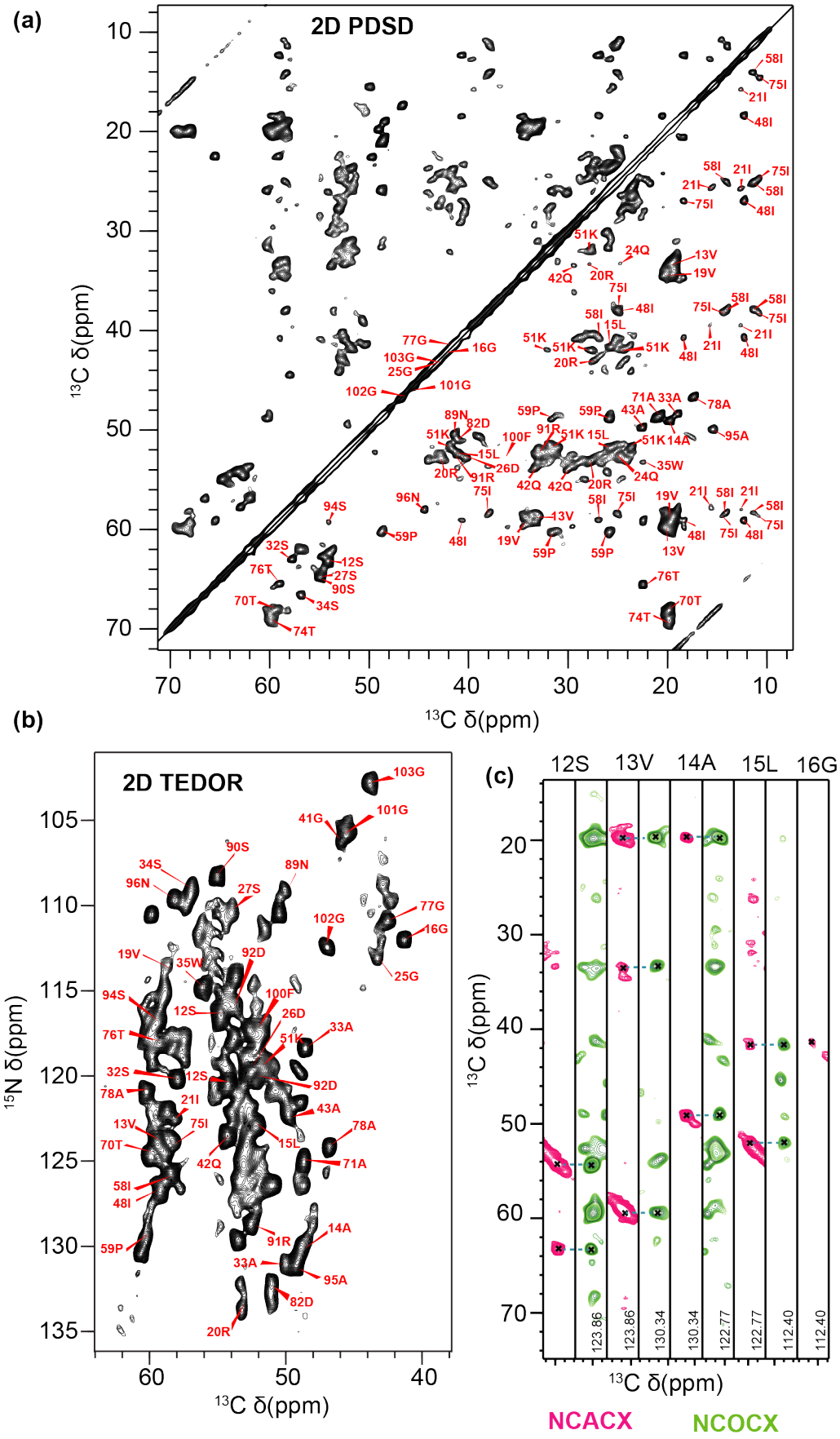


Figure 50: Assignment of S3706_patdel fibrils by solid state NMR

(a) & (b) are 2D ^{13}C-^{13}C correlation and 2D ^{15}N-^{13}C correlation depicting the assignment of rigid

residues in S3706_patdel fibrils seeded with ex-vivo material respectively. (c) 2D strip extracted from 3D NCACX and 3D NCOCX experiment illustrating the sequential assignment of a small segment from S3706_patdel fibrils.

4.2.5 Flexible region of S3706_patdel fibrils

Since in S3706_patdel 2D PDSD and 2D NCACX (CP based) that provides rigid residues information, does not show up peaks for all 110 residues. This indicates that all the residues are not rigid and indicated the presence of flexible region. In order to probe flexible region, we ran 2D INEPT experiment that is similar as solution state experiment.

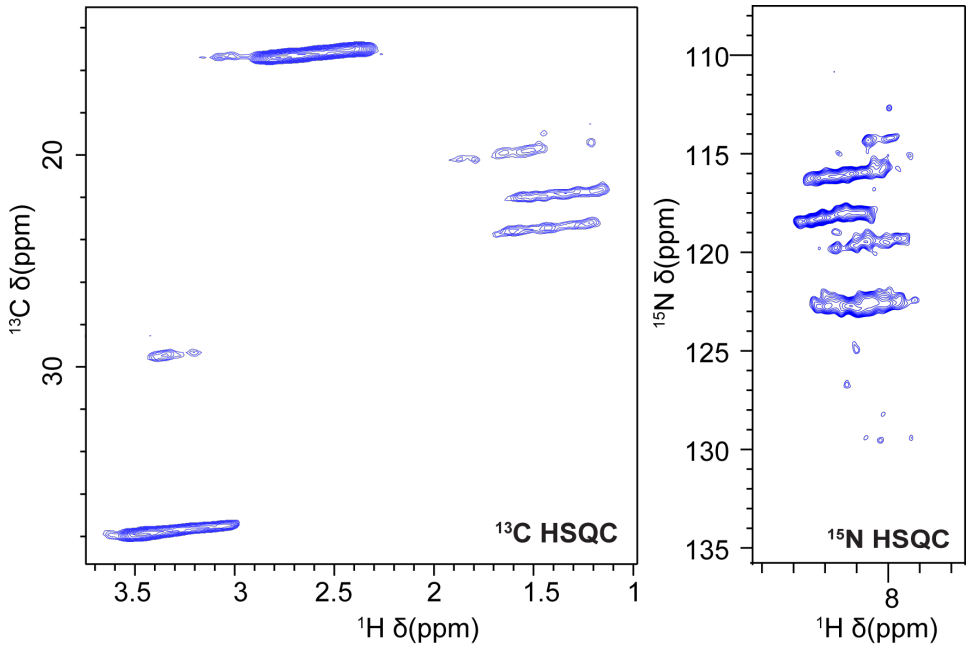


Figure 51: 2D INEPT in ssNMR to probe flexible part in S3706_patdel fibrils.

From 2D ^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC spectra we observe few broad peaks that corresponds to flexible. **(Figure 51)** This provides us the reasoning of missing residues form the fibrils in ^{13}C detected experiments that is for determining the hydrophobic core of the fibrils. Since the peak's linewidth were very broad, it was not worthwhile to run the 3D experiment for sequential assignment.

4.3 Discussion and conclusion

S3706_patdel is the patient derived V_L sequence, thus investigation of this protein fibrils will help in understanding the structure of amyloid plaques deposited in AL amyloidosis. In this study, we used solid-state NMR method to understand crucial interactions that are essential in the fibril formation. For comparison, we prepared the germline and point mutation to understand how they are different from patient fibrils.

Polymorphism and heterogeneity are a common phenomenon that occurs in amyloid fibrils formation. To get homogeneous fibrils, we used *ex-vivo* seeds that are the fibrils extracted from patient heart. The fibrils from patient act as seeds and template to obtain homogeneous sample and thus avoiding polymorphism. This is also reported in Aβ fibrils where seeds were used as template for self-assembly[204].

We successfully reproduce the same spectral features by preparing two samples using exactly same preparation. Our results indicate that if we use no seeds another polymorph is obtained with some of spectral features retained as in seeded spectra. Furthermore, the comparison of germline and point mutation fibrils with patient fibrils spectra shows the chemical shift perturbation is only due to mutational effect, keeping other spectral features is retained from patient. To achieve this, we used *ex-vivo* seeds from patients to prepare germline and single point mutation fibrils. We speculate that the core of the fibrils is same and during disease progression, the normal protein is recruited by the seeding effect.

Another interesting finding in terms of fibrils preparation was to obtain the same spectra irrespective of the seeds from *ex-vivo* or *in-vitro*. The same spectral features in 2D PDSD and 2D NCACX illustrates that during the fibrils formation the same polymorph was formed regardless of seeds prepared *in-vitro* from recombinant protein or fibrils extracted from patient. Thus, these results suggest that in future scientist don't have to rely on patient fibrils, as it is difficult to obtain.

Salt bridge formation is important information for that can stabilise the protein as well as fibrils. Lys and Arg side chain have specific chemical shift region in ^{13}C-^{15}N correlation spectra that can be easily identified. In patient fibrils, both Lys and Arg side chain were nicely resolved and observed all correlation that indicates the salt bridge existence. Indeed, 2D TEDOR experiment revealed unambiguous cross peak for Lys and Arg side chain connection to carboxylic acid group of Asp and Glu. These electrostatic interactions strongly appear in patient fibrils while in point mutation it was totally broken. In germline fibrils, the interaction was weak in Lys while from Arg connection was broken. In S3706_R50G, the broken salt bridge connection reveals that Arg at 50^{th} position is important for electrostatic interaction and for stabilising the fibrils. Till now nothing is published regarding the salt bridge for LC fibrils by solid state NMR. Thus, we hypothesize that these interactions were essential for the stabilisation of fibrils, which is only observed in patient fibrils. These salt bridges could be inter-molecular between two sheets or intra-molecular interaction within one cross β-unit as seen in Aβ fibrils[37]. To identify the salt bridge the exact position in the sequence, obviously the next step was the assignment of fibrils. Since the patient sequence was most important sequence, we only assigned this construct, and this can be used as template for assigning the spectra in point mutation and germline fibrils.

Although the sequence is unique in LCs, it is important to compare the rigid segment,

which are assigned by solid state NMR and we believe the structure and region should be similar in AL amyloid fibrils. As we stated earlier not much is known about the fibrils state and the topic is quite new in research, only three fibrils assignment are published till date. We compared S3706 fibrils assignment rigid residues with all the sequences published till now i.e. murine MAK33 (κIII), patient derived AL09 (κI) and very recent protein patient derived 6aJL2 (λVI)[94,201,203]. 6aJL2 and S3706 protein sequence both belongs to λ-type and thus share 60% identity. The comparison shows that there is variability in assigned segments within all four sequences. In MAK33 C-terminal residues were mostly assigned and nothing can be assigned in N-terminal region, which is in contrast to AL 09. 6aJL2 is closer to S3706 protein in terms of sequence as both belongs to λ subgroup shows few regions with similar assignment. **(Figure 52)**

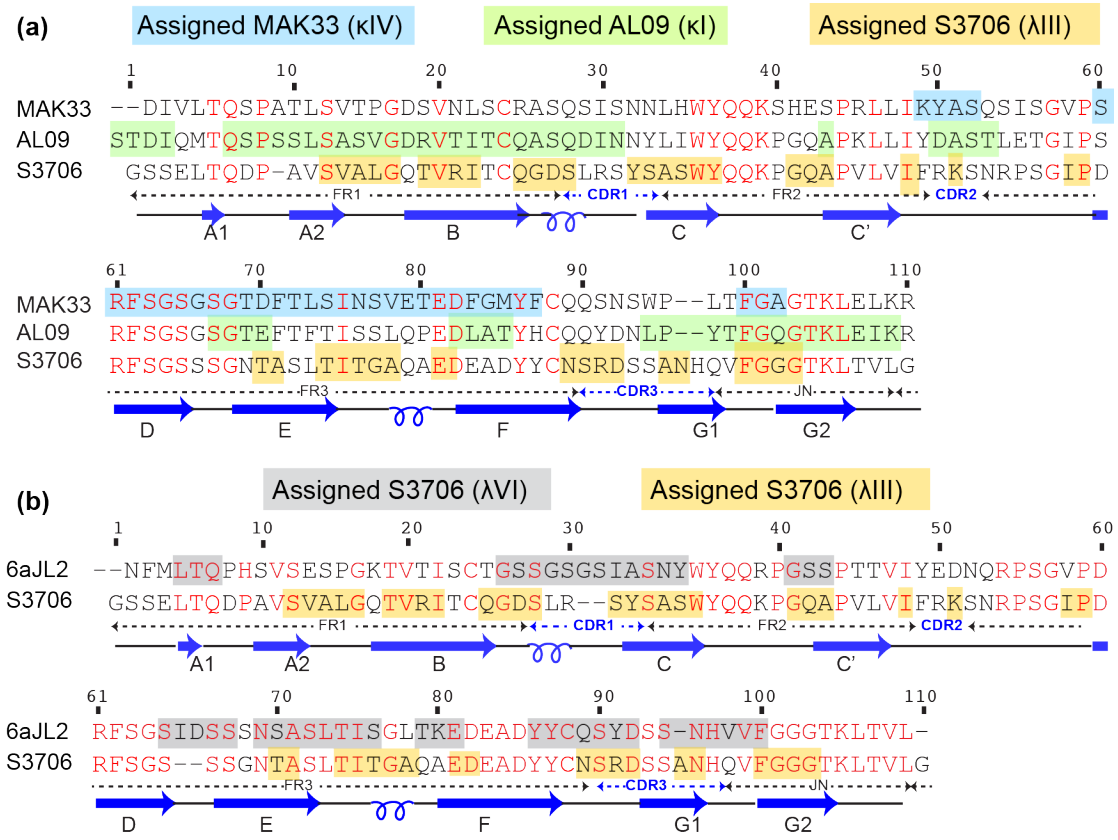


Figure 52: Comparison of assigned rigid regions in V_L fibrils using ssNMR.

(a) MAK33 (murine) and AL09 (patient derived) are the two LC V_L sequences that belongs to κIV and κI subgroup respectively[94,201]. Assigned residues in MAK33, AL09 and S3706 are in blue, green and yellow colour respectively. (b) 6aJL2 is LC V_L sequence that belongs to λVI subgroup[203]. The assigned residue is highlighted in grey. In (a) and (b) we compare other V_L sequences to the rigid residues assigned by solid state NMR in S3706 fibrils that belongs to λIII subgroup are highlighted in yellow. Sequence similarity is colored in red. Numbering of residues is according to the S3706_patdel sequence. Secondary structure is marked in blue arrows are from soluble monomeric state in solution.

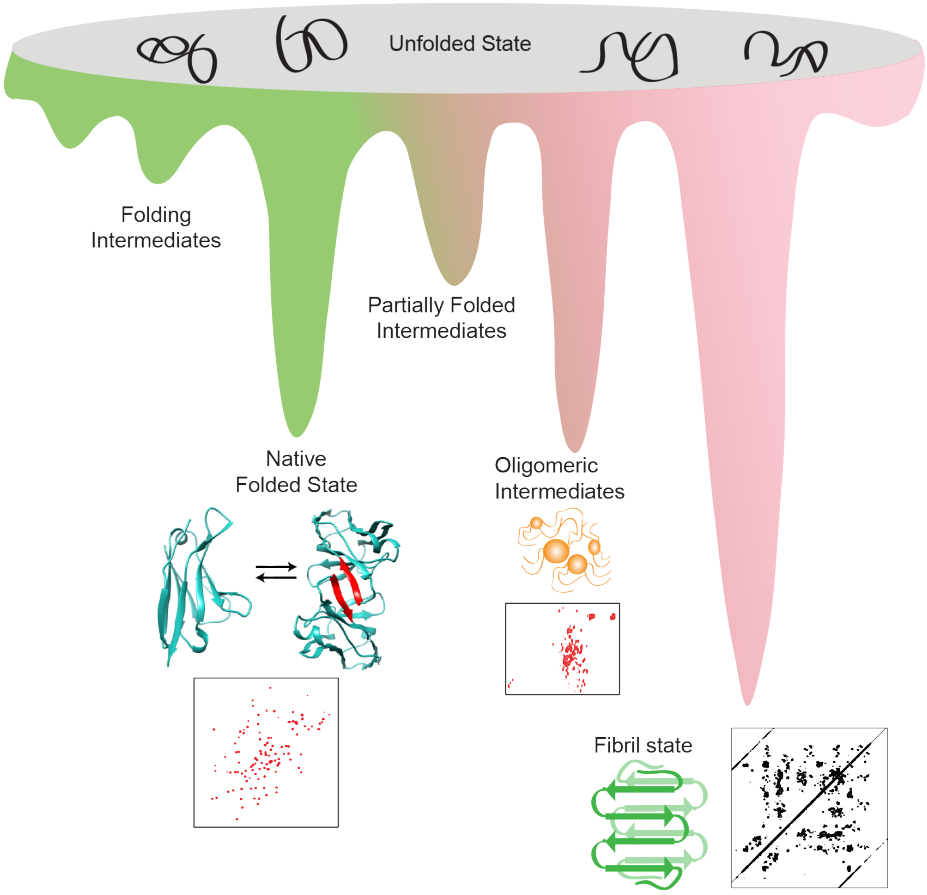


Figure 53: Energy landscapes of V_L protein states detected by NMR method

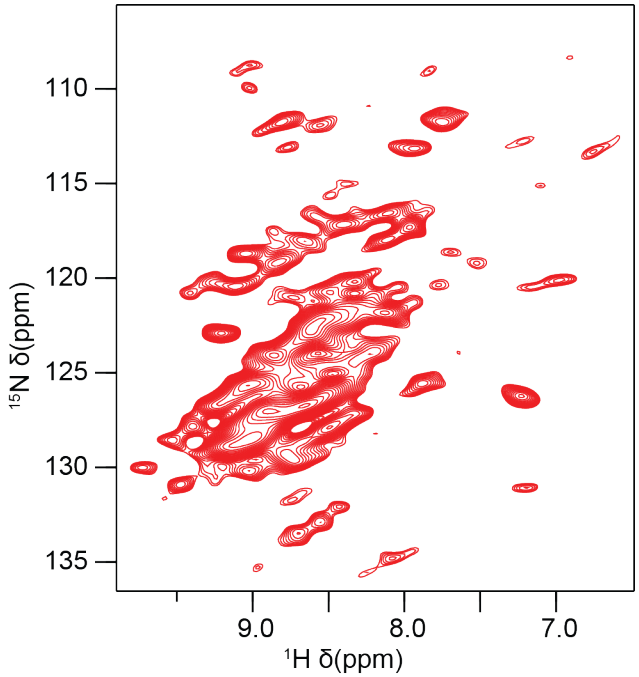


Figure 54: ^{1}H-^{15}N HSQC of deuterated S3706_patdel

In native monomeric state, oxidised cysteines are conserved in all light chains and it is one of stabilising factors in theses LC proteins. It will be interesting to know what happen to the cysteine states in fibrils state. Recently it has been shown by cryo EM that oxidised cystine bridge is conserved in AL 55 fibrils state[200]. Oxidised state in cysteine was also assigned by solid state NMR in 6aJL2 fibrils[203]. Although we could not assign the cysteine, but we assigned the residues, which are in proximity and we believe that cysteine to be rigid and in oxidised state. It was also in debate whether the secondary structures are conserved in fibrils state as in native monomers or a rearrangement occurs upon fibrillation. 6aJL2 fibrils assignment show that there is conversion of globular state to fibrils state and their results also suggested that the β-strands are stacked in-register parallel in fibrils[203].

To conclude, we characterise the protein states including monomeric globular state, oligomeric state and final stable fibril state that are seen in the energy landscape of protein, using solution and solid-state NMR method. **(Figure 53)** As fibrils preparation is important in solid-state NMR, we successfully able to regenerate the spectra and to show the cross-seeding effect. We proposed Arg at 50^{th} position is important for the salt bridge formation, which might be important in stabilising fibrils. In future, complete assignment of rigid residues fibrils will support in proper understanding the role of electrostatic interaction. We prepared deuterated V_L fibrils samples and to account amide contacts using 3D experiments using 1.3 mm probe in future and these will be later used in structural calculation. **(Figure 54)** Our findings may help in understanding the misfolding of native folded state protein giving new insight to the LC fibril. Due to lack of understanding in AL amyloidosis fibrils in atomic structural level, it is important to have more research in this direction in future. For proper understanding of the disease and accomplishing a reliable conclusion, a larger number of LC protein sequences are required to be investigated.

5 Interaction of Q-rich domain with RRMs in TIA-1

5.1 Background

T-cell Intracellular antigen-1 (TIA-1) is Ribonucleic acid (RNA) binding protein composed of three RNA recognition motifs (RRMs) and glutamine rich motif (Q-rich) at C-terminal. It is associated with Welander distal myopathy in which there is mutation at position 384 where E is mutated to K in patients in Q-rich domain[206]. It acts as apoptosis promoting factor and involved in alternative mRNA splicing, cytoplasmic RNA metabolism and translational regulator[207]. Human TIA-1 promotes the splicing of human FAS pre-mRNA to insoluble FAS ligand and triggers cell death by inclusion of FAS exon 6. This inclusion is promoted through the recruitment of snRNP by phosphorylation of TIA-1 by FAST-K enzyme[208]. It was shown that RRM-2 and RRM-3 of TIA-1 strongly binds with pyrimidine rich FAS pre-mRNA and U9 or U15 in nano-molar affinity[187]. Upon RNA binding, TIA-1 facilitates splicing site recognition by employing U1 small nuclear ribonucleoprotein (snRNP) complex to the splicing site by U1C interaction, which is snRNP-associated protein. It was reported that TIA-1 Q-rich domain interacts with U1C which is aided by TIA-1 RRM-1 domain[209].

Mutation in Q-rich domain can lead to the ALS and frontotemporal Dementia, promotes the phase separation and alter the SGs dynamics[210]. Recently it was revealed that Zinc divalent cation can accelerates, the phase separation, the self-assembly and Arsenite induced stress granules and these processes can be reversed by chelator TPEN in TIA-1[211].

5.2 Aim of the study

QRD domain are the low complexity motifs that exhibits phase separation as described in Introduction section **1.4.** Our aim is to characterize the Q-rich domain (QRD) and its interaction with other RRMs domains in TIA-1 by solid state NMR. We used other biophysical methods like DIC ThT and TEM, which further help in characterization of Q-rich domain. This may help in understanding the role of this domain in snRNP complex and stress granules.

QRD-44 are first 44 residues from Q-rich domain, a fragment capable of forming fibrils. We studied another construct RRM123_QRD44 along with QRD44. RRM123_QRD44 includes all RRM domains with QRD44 fragment. To understand the domain interaction, we mutated the linker between RRM-1 and RRM-2 that is named as RRM123_QRD44_mut. We used DIC, TEM and ThT assay methods to characterise the fibrils and protein.

5.3 Results

5.3.1 Solution state NMR data of RRM1 and Q-rich domain

(Kindly provided by Pravin Jagtap from Prof. Michael Sattler group)

When RRM-1-QRD44 (artificial construct) domain was compared with RRM-1 domain only, few chemical shifts were difference was observed in ^{1}H-^{15}N HSQC. These are mapped in homology model of RRM-1 domain, which indicates that it is localized to certain region at N-terminal and C-terminal of RRM-1 domain. In context to full TIA-1, the chemical shift perturbations were localized in the linker between RRM1 and RRM2 and N-terminal. **(Figure 56c)** This suggests that RRM1 domain interacts with Q-rich domain, which further helps in interaction with U1C protein as described in model in **Figure 56b**. This supports the previously observed data for Q-rich which indicates that it mediates with RRM-1 domain to facilitates the protein-protein interaction with U1C for the recruitment in U1snRNP[209].

TIA-1 Q-44 ^{1}H-^{15}N HSQC spectra shows that it is random coil which is as expected from the literature. The peaks are dispersed within the narrow range from 7.5 - 8.5 ppm in ^{1}H scale, representing a typical characteristic of unstructured or disordered protein. **(Figure 55)** When the protein was kept outside at room temperature, it was found to be unstable and peaks disappeared from the spectra. To further probe the aggregation process, TEM and ThT assay was done which is discussed in later section.

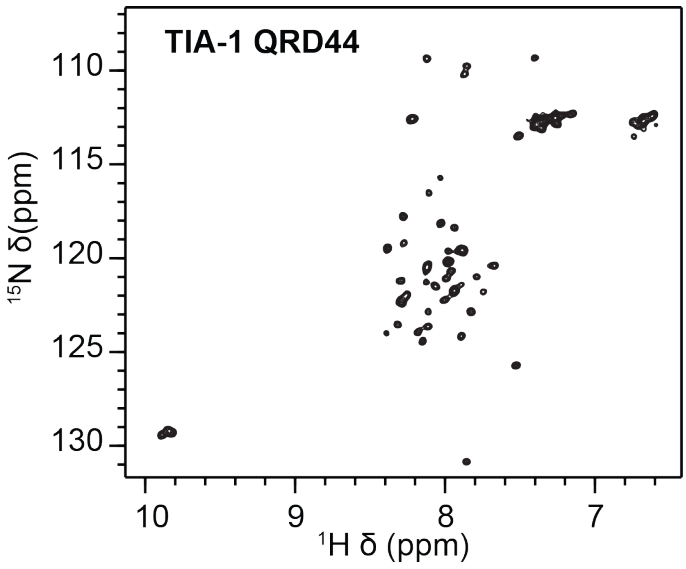


Figure 55: Solution state NMR of TIA-1 QRD44

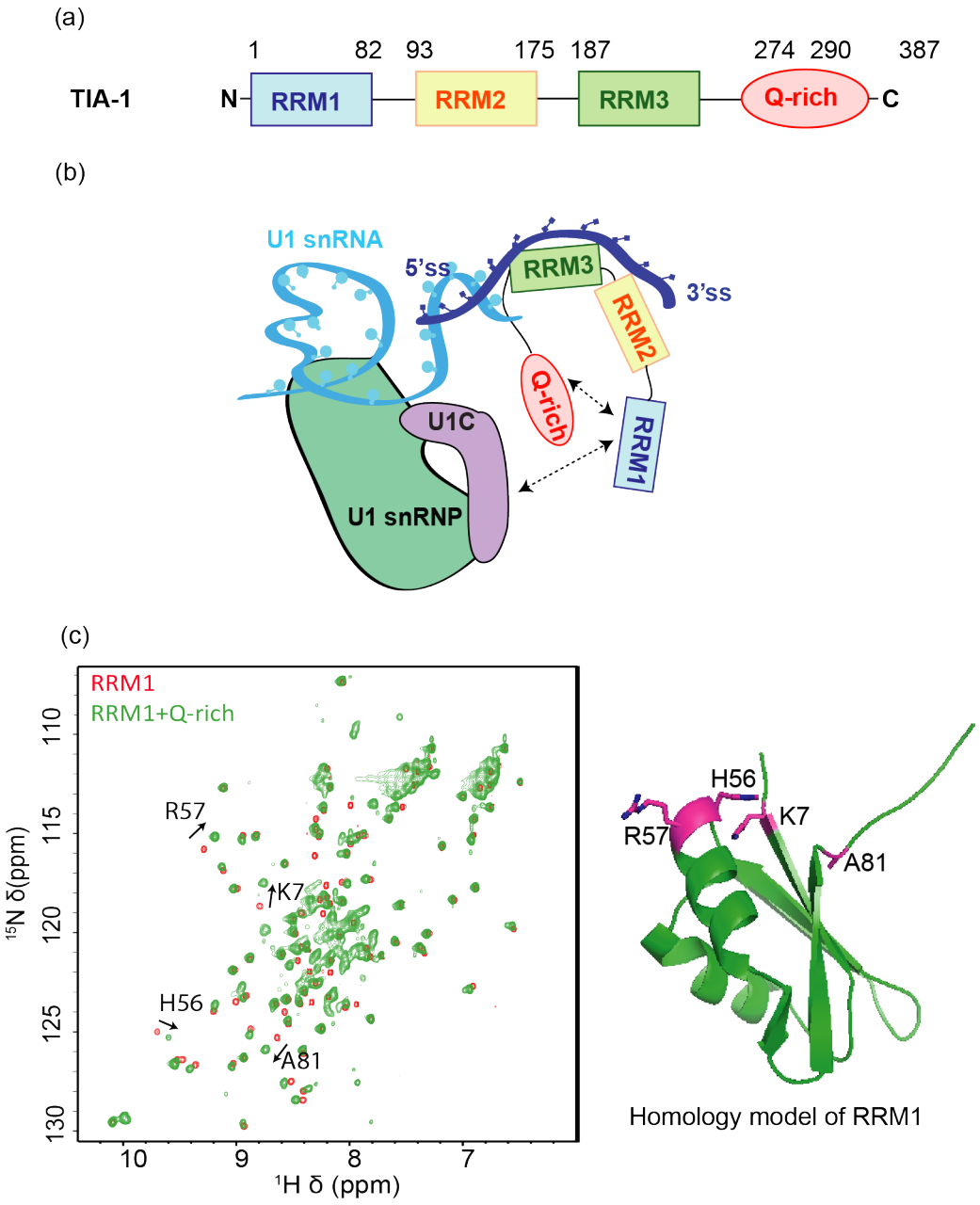


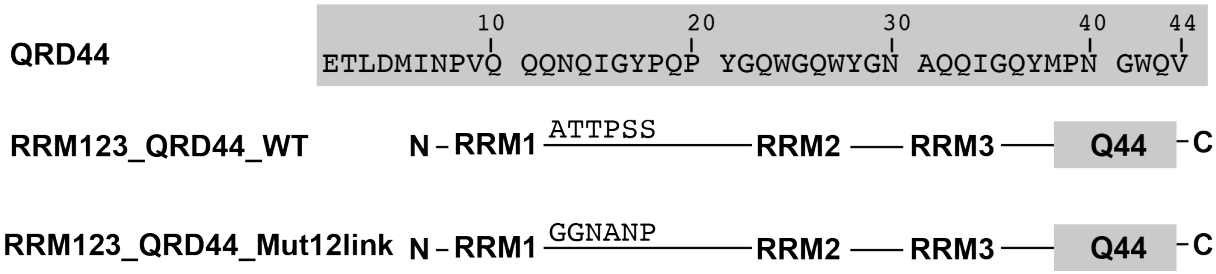
Figure 56: TIA-1 domains and their domain interactions.

(a) Schematic representation of domains in TIA-1 protein. (b) Model of U1 snRNP complex including TIA-1, other proteins and RNA elements presents in U1 snRNP. Two arrows show the protein-protein interaction between TIA-1 and U1C and domain interactions between RRM1 and Q-rich domain. (c) Left hand Side: ^{1}H-^{15}N HSQC overlay of RRM1 domain with RRM1+Q-rich (artificial construct) depicting the residues which are shifted. Right hand side: CSPs from ^{1}H-^{15}N HSQC mapped in homology model of RRM1 domain. (Kindly provided by Pravin Jagtap from AG Sattler).

5.3.2 TIA-1 constructs

First 44 residues from Q-rich domain were able to form fibrils and amenable for solid-state NMR studies. RRM123_QRD44_WT and RRM123_QRD44_Mut construct was used to probe the interaction between Q-rich domain and RRM motifs so, it includes all RRMs. In RRM123_QRD44_Mut construct, the first 6 residues were mutated from linker region between RRM1-RRM2. The full sequence of the constructs is written in **Appendix III** and all constructs used in this study are described below:

TIA-1 contructs



All the constructs have His_6 tag for Nickel purification, but the solubility tag is different. In QRD44, as a solubility tag, thioredoxin was used whereas in RRM123_QRD44_WT and RRM123_QRD44_Mut12 construct GST tag. Purification protocol is described in *Materials and Method* section.

5.3.3 Characterization of fibrils by ThT assay

In order to follow the kinetics of protein ThT assay was performed. We used 10 μM TIA-1 protein from all the three constructs QRD44, RRM123_QRD44_WT and RRM123_QRD44_Mut. RRM123_QRD44_WT and its mutant have less significant difference in fibril formation. TIA-1 RRMs constructs shows faster kinetics when compared with QRD44. For QRD44, the fibrils were obtained in 12 hrs. while for RRMs construct, kinetics was quite fast, and the fibrils were obtained in 2 hrs. Among all the constructs, QRD44 was the slowest in forming the fibrils. This indicates that RRMs promotes the faster aggregation. Furthermore, we added 10 μM RNA to the two RRMs constructs to probe the effect having prior knowledge of binding of RNA with RRMs. We did not find any significant difference in kinetics of RRM123_QRD44_WT and RRM123_QRD44_Mut on addition of RNA. **(Figure 57)**

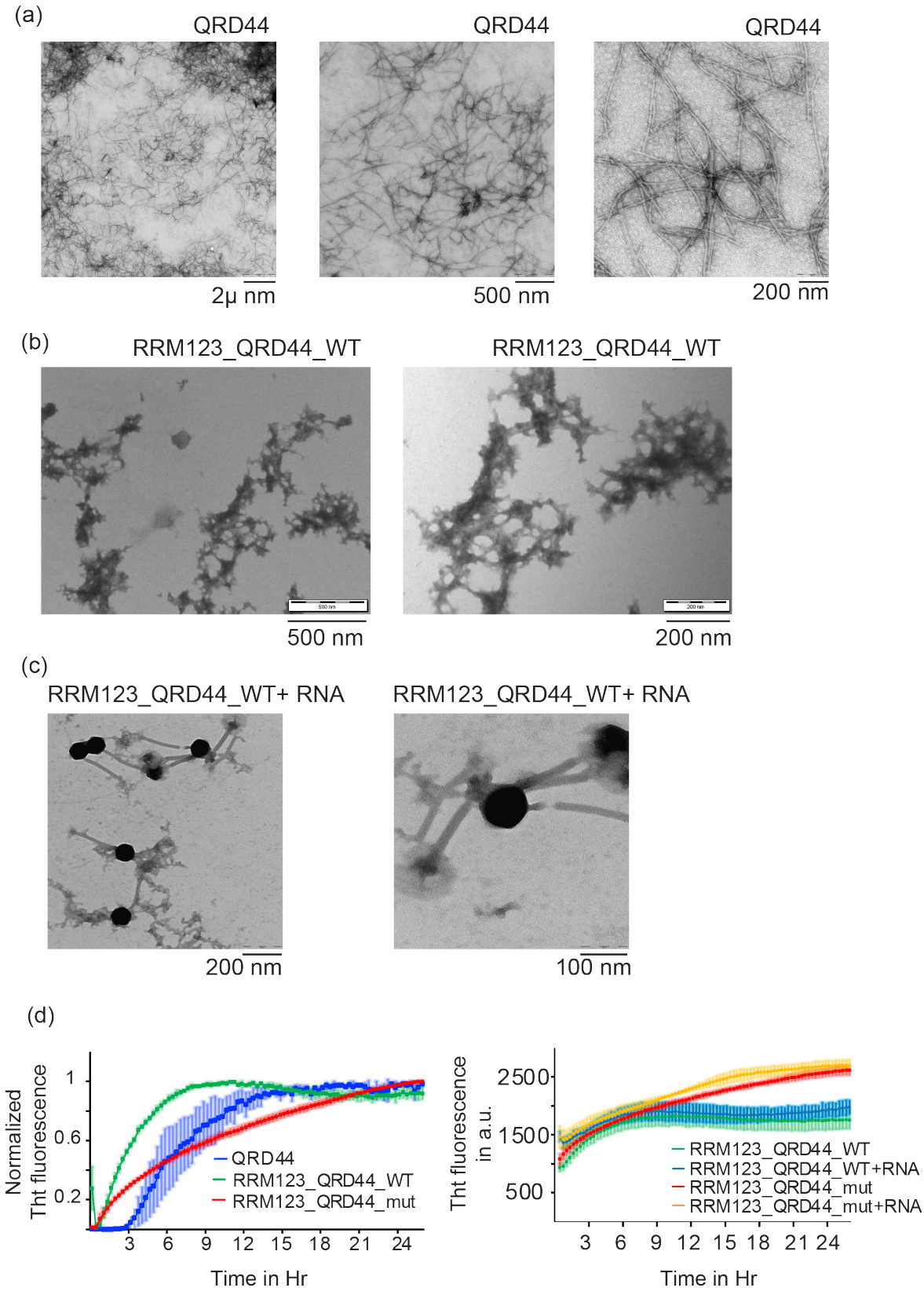


Figure 57: Fibrils characterization using TEM and ThTassay.

(a), (b) and (c) TEM images of QRD44 alone, RRM123_QRD44_WT and its mutnat respectively at different magnification. It shows slightly different morphology in the presence of RRMs. Moreover in presence of RNA, it shows diffuse morphology at the ends of the fibrils. d) ThT assay plots showing comparison of QRD44 alone, with WT RRM domains and its mutant (RRM1-RRM2 linker).

5.3.4 Morphology of fibrils by TEM

For checking the morphology of the fibrils, negative staining TEM was performed. The details of the experiments are described in *Material and Methods* section. In case of TIA-1 QRD44, nice distinct fibrils were obtained which are well dispersed as seen by TEM. On higher magnification, we observed the twist morphology in fibrils. In case of RRM123_QRD44_WT, the fibrils were entangled, and morphology was not so clear**. (Figure 57a-c)** The fibrils where RNA was added to the protein, we observed fibrils clump together with the diffuse end in the fibrils. In few cases, it seemed like many fibrils are attached to each other.

5.3.5 Phase separation using Differential interference contrast (DIC) Microscopy

Phase separation was detected using images generated by DIC microscope instrument. As expected, we observed phase separation in TIA-1 proteins with Q-rich domain. In QRD44 at lower concentration we observe the round shaped droplets, which could self-assemble together to form gels followed by star shaped fibrils structure. To examine the β-sheet structure of protein, we added ThT dye to the protein sample with droplets and fibrils and detected in ThT wavelength, we observed blue fluorescence in both samples thus gives the positive indication. **(Figure 58a)** At very high concentration, we observe the gel state of the protein. Ongoing detail of the image we observed the fused droplets and fibrils like structure inside the gels. **(Figure 58b)** In gel state also, ThT test was positive indicating the β-sheet structure (data not shown). In case of RRM123_QRD44_Mut we did not observe the droplet formation due to the fast kinetics of protein, but we observed the gels and star shape fibrils. We do not have data from RRM123_QRD44_WT but we expect the same results as in the case of its mutant**. (Figure 58c)**

5.3.6 Solid state NMR of QRD44 domain

In order to get homogeneous sample, different concentration of protein like 50 µM (with and without seeding) and 10 µM (with seeding) was prepared and recorded with MAS in 750 MHz in 3.2 mm rotors. Despite using seeds all the samples show heterogeneity, only advantage was the resolution as increased and spectra were reproducible. The sequential assignment was not possible because of overlapped spectra in ^{13}C-^{15}N correlations and heterogeneity is sample just aided to difficulty. Another reason was glutamines are overlapped to each other at same position that makes the mapping of sequence difficult. Among the three uniformly labeled samples, the seeded spectra (10 µM) were used for the comparison in all figures since it has the good resolution. Non-seeded spectra and seeded in 50 µM QRD44 protein are shown in appendix.

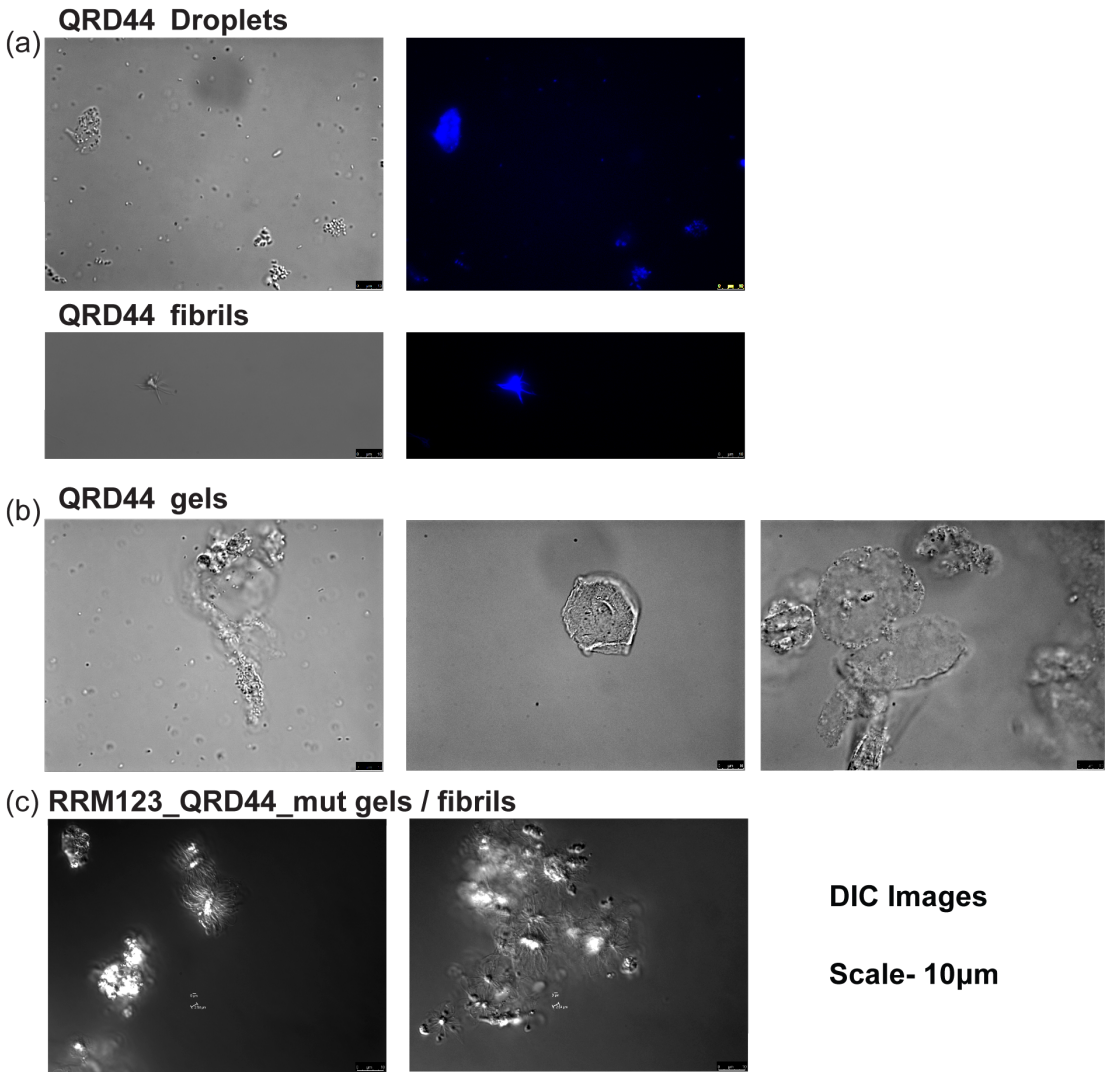


Figure 58: DIC images of TIA-1 protein.

(a) DIC image for droplets formed by QRD44 at lower concentration. Below are the star shaped fibrils formed after fusion of droplets. Right side: After addition of ThT dye, blue fluorescence was observed in both droplets and fibrils state. (b) QRD44 gel formation at very high concentration of protein. Droplets come close to each other to form gels and closer look show fibrils like structure inside the gels (c) Gels as well as star shaped fibrils are observed in RRM123_QRD44_mut sample which was incubated for two weeks at 37°C in fibril shaker.

In QRD44 2D PDSD, that gives ^{13}C-^{13}C correlations we assign the spin system of amino acids. Isoleucine, Alanine and threonine has very distinct region and nice dispersion can be seen. In glutamine region, there is blob due to overlapping of at least 11 glutamines but due to polymorphism one expects more peaks. In carbonyl region, carboxylic group from glutamic acid side chain has distinct region around 180 ppm. Among four prolines we observe 3 strong and 3 weak peaks and all peaks are in trans conformation as the difference between Cγ and Cβ is around 5[212]. According to secondary structure probability, all the peaks in aliphatic region correspond to sheet region depicted in pink shade. Since the sequence contains 4 tyr and 3 trp, it was worthwhile to check aromatic region and assigned the spin system. **(Figure 59)**

To further investigate in QRD44 fibrils, we recorded 2D NCACX and 2D TEDOR that gives the ^{13}C-^{15}N correlations. While probing the glycine residues ~45 ppm at C-dimension, it is found that the number of peaks observed (11) are approx. double of number of peaks expected (6) in both 2D NCACX and 2D REDOR spectra. For proline residues 2D TEDOR was analyzed, the number of peaks appears (4) were twice as number of peaks expected (8) as stated above. In summary, the observation in ^{13}C-^{15}N correlations were similar to ^{13}C-^{13}C correlations indicating polymorphism in fibrils. **(Figure 60)**

Due to the overlapping glutamine region and polymorphism, 3D experiments exhibit low resolution therefore non-analyzable and thus it was not possible to do the sequential assignment of this protein.

Deuterium labeling with different preparation methods was also tried in addition in QRD44 construct to get insight of structural information, but the spectra was not analyzable due to low signal to noise and overlapping. (Data not shown)

5.3.7 Long range distance contacts in QRD44 fibrils

Long mixing 2D PDSD was recorded to observe long-range contacts and compared with short mix 2D PDSD In QRD44 fibrils. In long mixing, new peaks appeared along with peaks from short mixing which is mapped to residue to show connections in both aliphatic and aromatic regions. In aliphatic region, connection from Glycine to Glutamine and in aromatic region, connection from tyrosine to tryptophan and tyrosine to isoleucine were very distinguishable. **(Figure 60)**

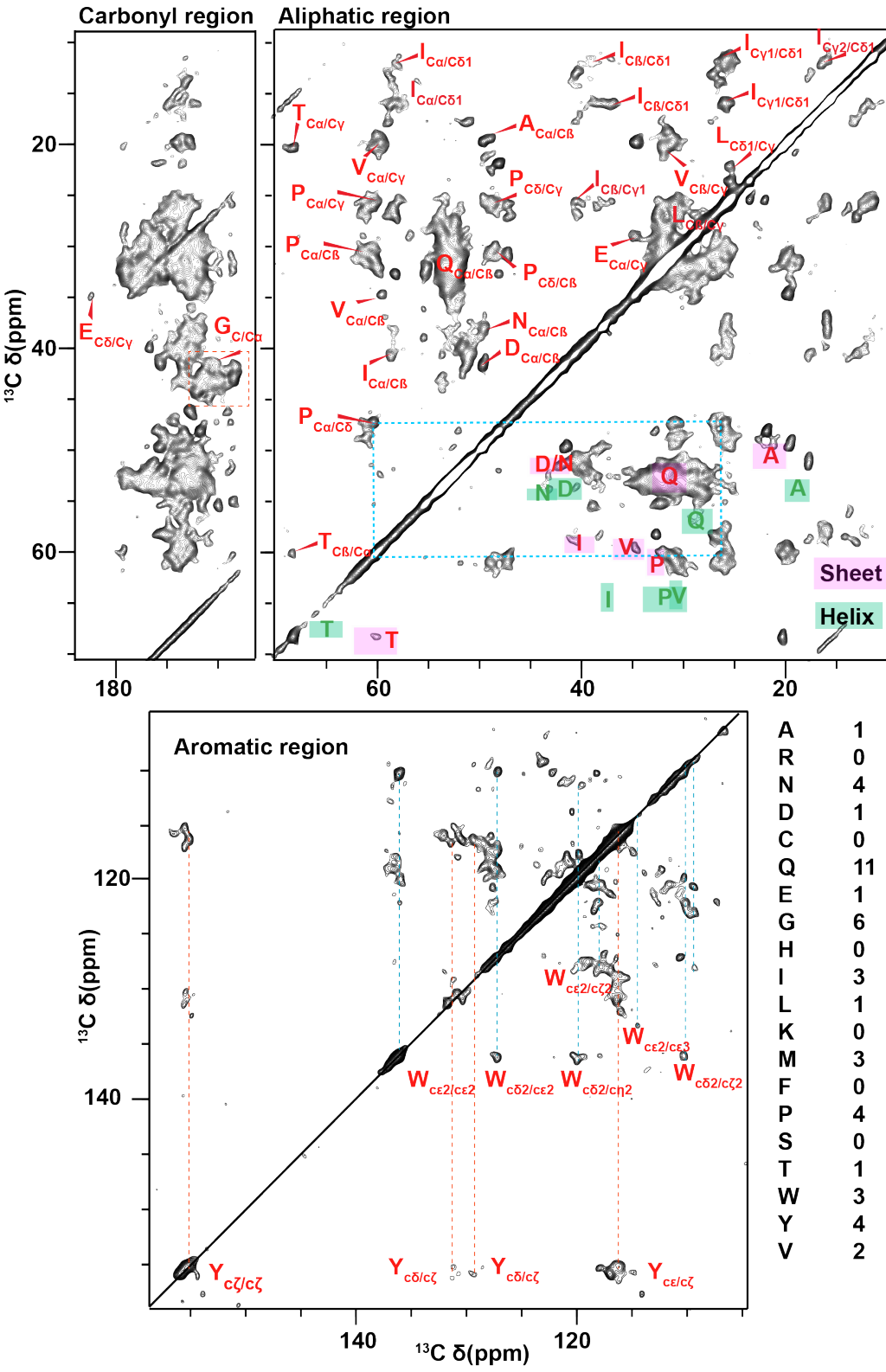


Figure 59: ^{13}C-^{13}C correlations of TIA-1 QRD44 fibrils.

one side of the diagonal amino acids spin system in aliphatic region is indicated and on the other side secondary region, helix region (Green shade) and sheet region (Pink shade) is indicated. All the peaks in 2D PDSD belongs to the sheet region in TIA-1 QRD44 fibrils. Proline spin system is indicated in blue line. Below is the aromatic region of the same fibrils with spin system indicated. Amino acids with their number of occurrences in primary sequence are displayed in right hand side.

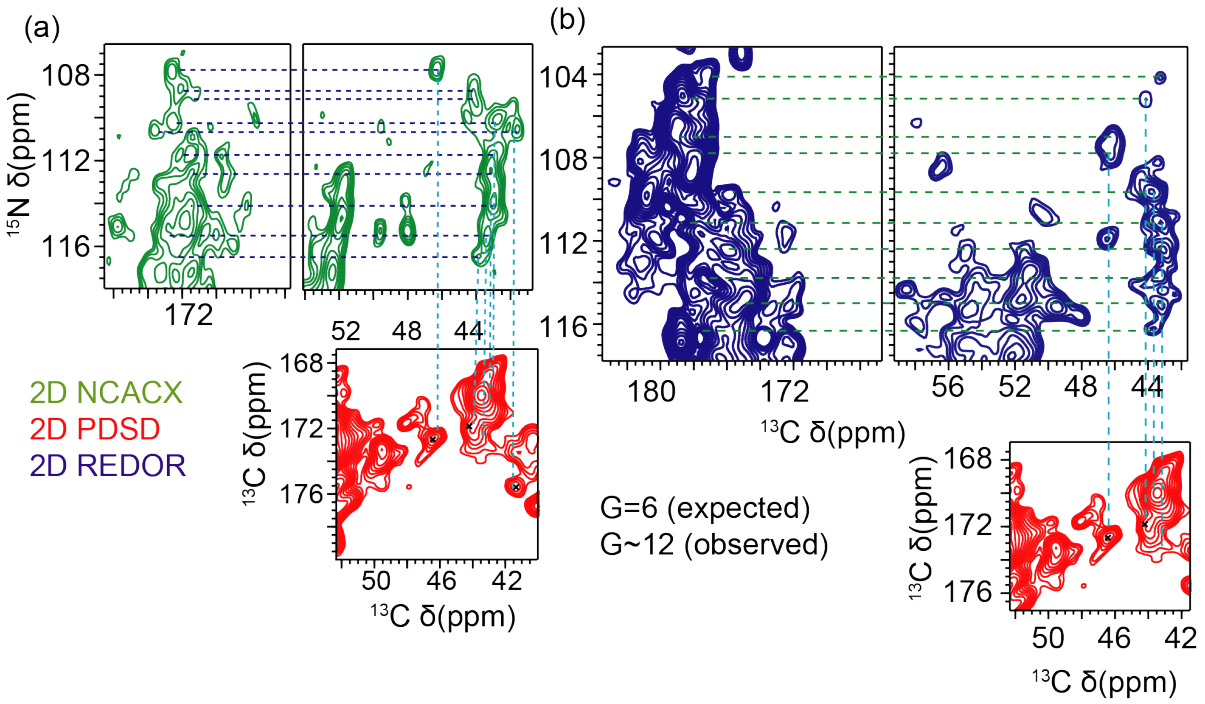


Figure 60: ^{13}C-^{15}N correlations of TIA-1 QRD44 fibrils.

(a) and (b) Correlation showing glycine Cα to carbonyl in 2D NCACX and 2D PDSD and 2D TEDOR and 2D PDSD respectively. Expected glycine peaks from sequence are 6 while observed from spectra are approx. 12.

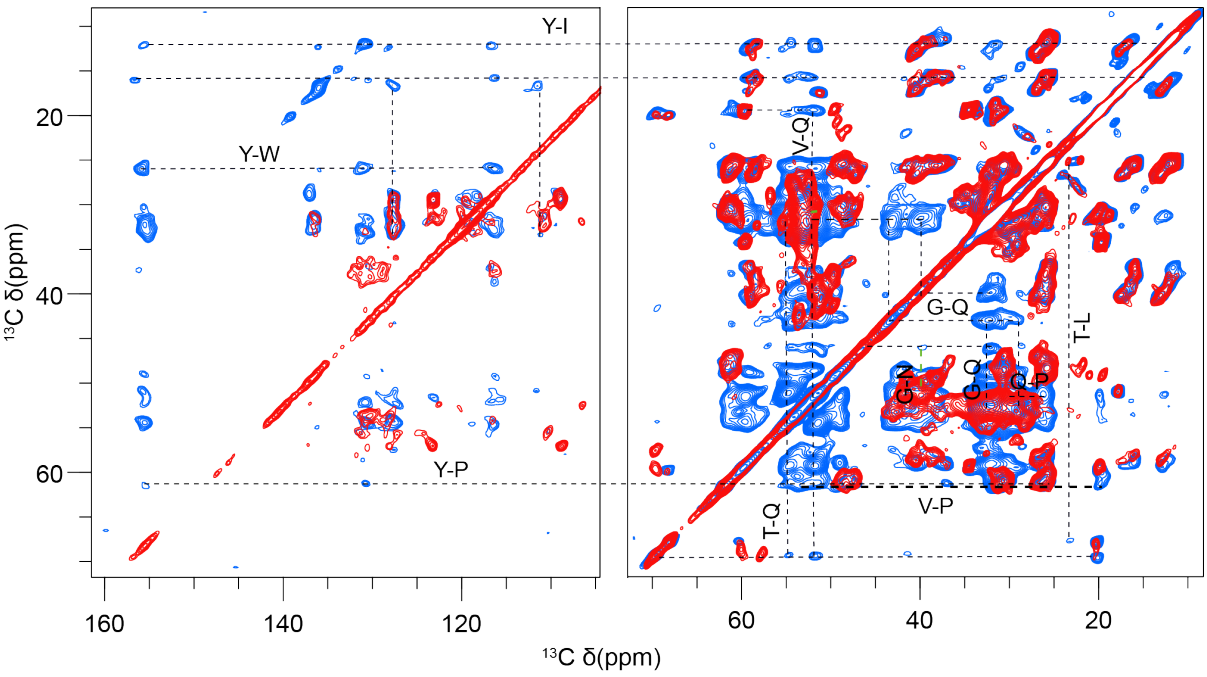


Figure 61: Long range contacts in TIA-1 QRD44 fibrils.

^{13}C-^{13}C Correlation with long mixing (Blue) and short mixing (Red) 2D PDSD depicting long-range contacts in aliphatic and aromatic region.

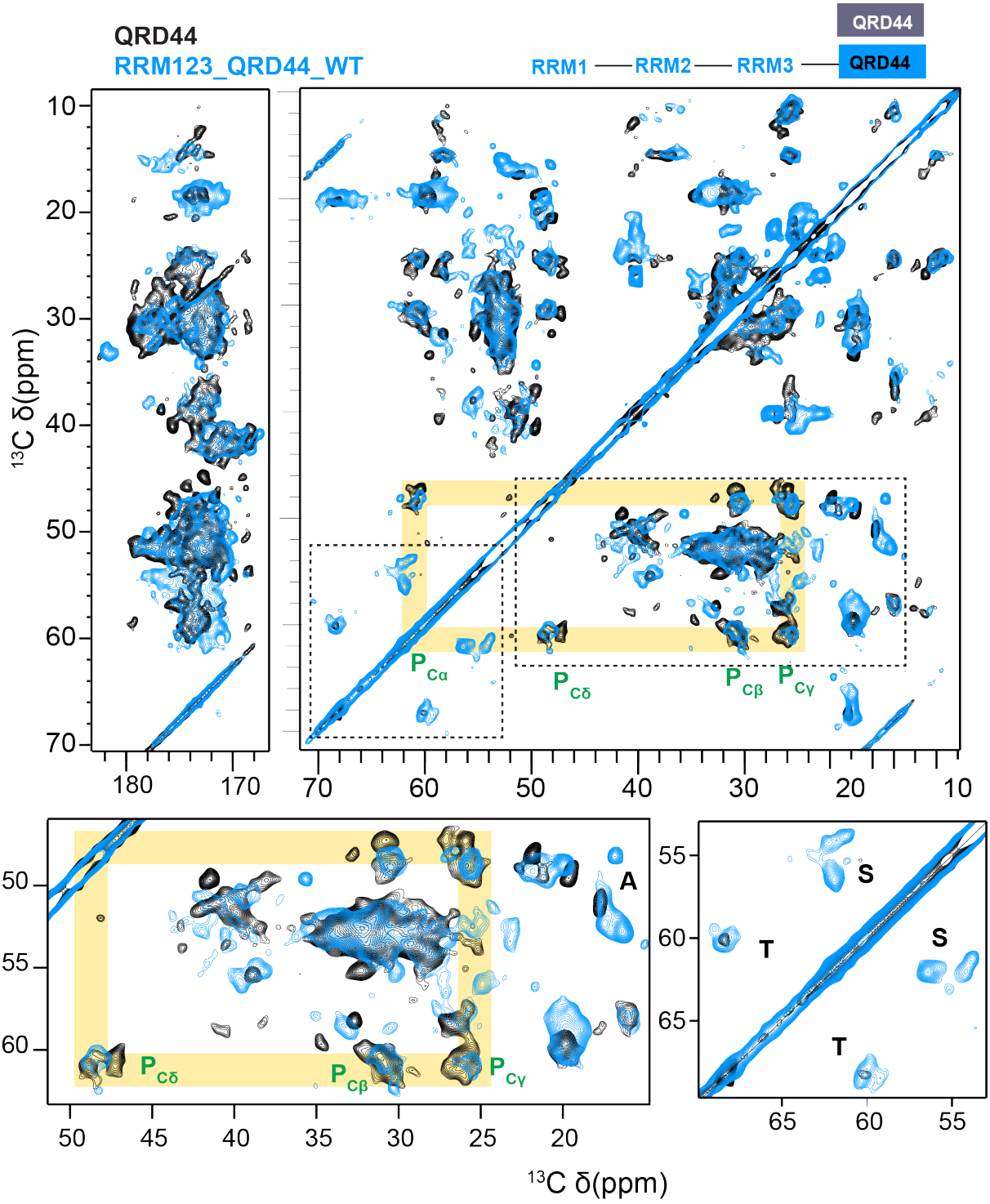


Figure 62: Comparison between TIA-1 QRD44 with and without RRM domains.

Superposition of 2D ^{13}C-^{13}C correlations from QRD44 (Black) and RRM123_QRD44_WT (Light Blue) fibrils depicting aliphatic and carbonyl regions in shown above. Below are the zoom plots, which are marked in dotted square in above figure. Extra peaks can be seen in the blue spectra in alanine, serine and threonine region. Proline spin system is marked in yellow shade.

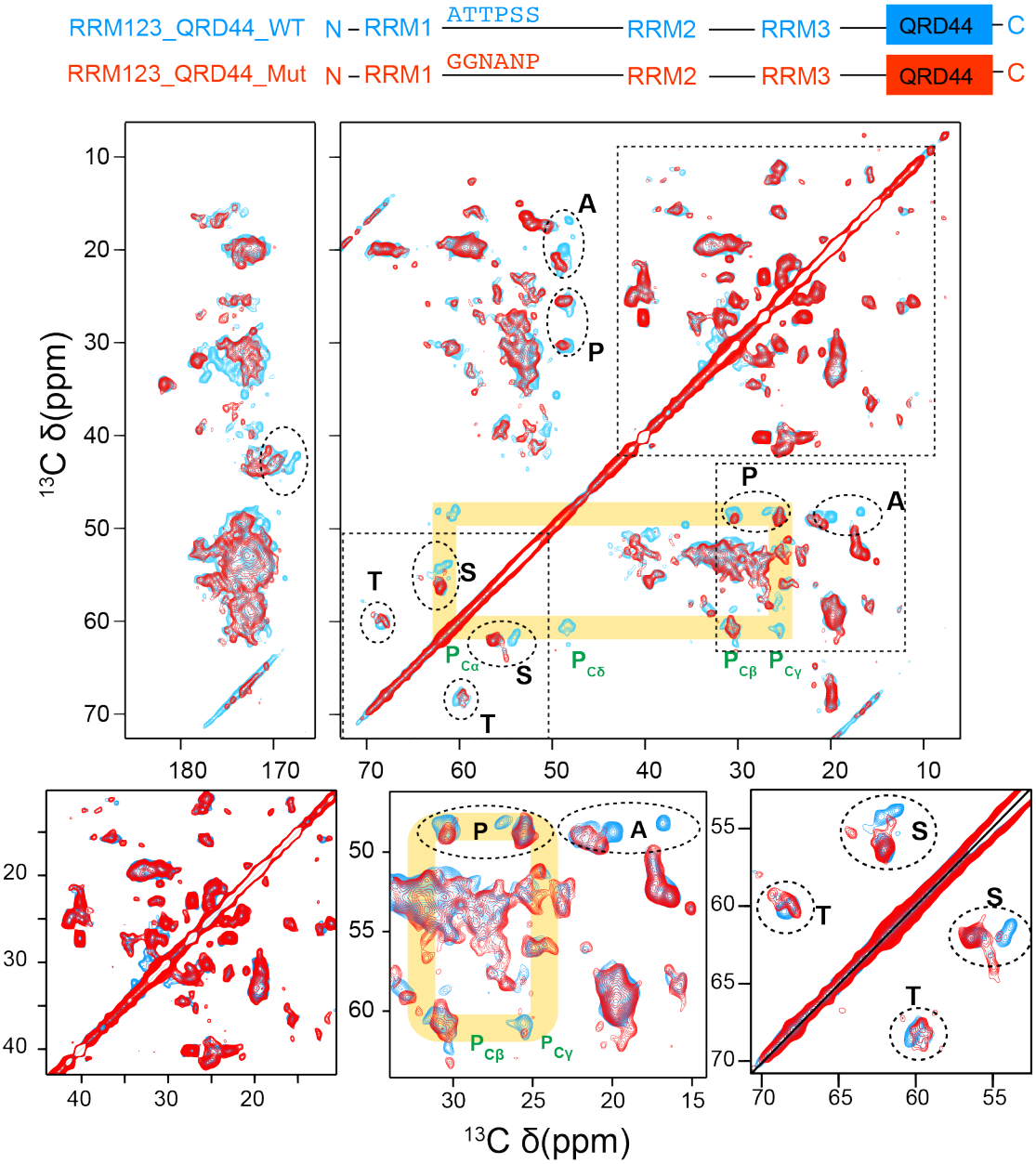


Figure 63: Comparison between TIA-1 WT and its mutant fibrils.

Superposition of ^{13}C-^{13}C correlations from RRM123_QRD44_WT (Blue) and RRM123_QRD44_mut (Red). Changes are marked in dotted round circle and with their amino acid type. Proline spin system is marked in light green shade with its spin system. Dotted square is zoom section that are shown below. Zoom plot from 2D PDSD with no changes is in the first spectra followed by the aliphatic region with changes in Pro/Ala and Ser/Thr region.

5.3.8 Interaction of Q rich domain and RRM123_WT domain

QRD44 alone and RRM123_QRD44_WT fibrils was compared to analyze the effect of RRM domains. From 2D PDSD overlay, it was observed that overlapped peaks are from QRD44 domain and non-overlapped peaks are certainly from RRM domains. Considering the number of peaks appeared in 2D PDSD from RRM123_QRD44_WT sample which has total 318 residues out of which 44 residues are from QRD44 and other 274 peaks are from RRM domains, the spectra was not crowded. Although the number of residues in RRM123_QRD44_WT fibrils are more as compared to QRD44, all the residues peaks does not appear in this sample. It indicates that only some region of RRMs appeared in 2D PDSD. Peaks appeared in serine region, gives the clear indication that it must be from RRM domain because QRD44 construct has no serine in the sequence. In the proline region, the peaks were as QRD44 spectra, all were in trans conformation. In threonine and alanine position, we observed some new peaks, which must be emerging from RRMs. Overall the spectra do not change from QRD44 to RRM123_QRD44_WT with some extra peaks added in the spectra. **(Figure 62)** Based on the observation, we mutated the linker region from RRM-1-RRM-2, which has serine, threonine and alanine.

5.3.9 Effect of linker region in TIA-1 fibrils

RRM123_QRD44_mut is the construct where we mutated 6 amino acids ATTPSS to GGNANP from RRM-1-RRM-2 linker region. **(Figure 63)** We prepared the fibrils using the same conditions as its WT fibrils and compared this construct to RRM123_QRD44_WT using 2D PDSD in solid state NMR. All the spectra were recorded using same spectrometer and MAS spinning. On comparison of 2D PDSD spectra from the two constructs, it was observed that only few peaks disappeared in RRM123_QRD44_mut sample but the overall spectra remain same. **(Figure 63)** In side chain region of Isoleucine, Valine and Lysine region, the peaks are perfectly overlapping. In alanine, serine and threonine region, we observed see some peaks are disappeared very clearly. **(Figure 63)** Moreover, in proline spin system, one spin system disappeared in RRM123_QRD44_mut construct. We observed few more peaks missing in RRM123_QRD44_mut that does not correspond to amino acids included in mutation.

5.4 Discussion

It is important to know the characteristics feature of low complexity motifs in RRM binding proteins as it is involved in regulatory mechanism of cells like stress granule formation[154,213]. It is known that Q-rich domain in TIA-1 has the properties of phase separation and granule formation and recently it was demonstrated that zinc cation can

enhance this process[211]. It was proposed that Q-rich domain accelerates the protein-protein interaction with U1C, U1 snRNP specific protein, which further helps in recruitment of U1 snRNP[209].

From our results, it was found that QRD44 i.e. first 44 amino acids from Q-rich domain can form fibrils independently. To see the effect of RRM domains we took the RRM123_QRD44_WT where all three RRMs are included and RRM123_QRD44_mut where 6 amino acids were mutated from RRM1-RRM2 linker region.

From solution state NMR, it was revealed that QRD44 possess random coil structure and behaves as intrinsically disordered protein. RRM123_QRD44_WT and RRM123_QRD44_mut were very unstable protein as compared to QRD44 and in solution state NMR it shows very few peaks and non-analysable spectra. (Data not shown) RRM123_QRD44_WT and RRM123_QRD44_mut aggregates faster than QRD44 alone.

To further prove, we performed ThT kinetics, which also supports the solution state NMR data that shows that RRM123_QRD44_WT aggregates faster than QRD44. This indicates that RRM domains accelerate the fibril formation. Comparing between WT and its mutant, it reveals that changing the linker, lowers the rate of formation of fibrils. It also gives a hint the linker between RRM1-RRM2 play an important role in structure of fibrils.

TEM microscopy assay revealed that the morphology of fibrils formed by QRD44 are distinct and non-overlapped in case of QRD44 construct as compared to RRM123_QRD44_WT and RRM123_QRD44_mut, where fibrils were diffuse and clump together. Moreover, in case of RRM123_QRD44_mut we observed the diffuse structure in tail of the fibrils which indicates that the presence of RRMs domain. We assumed and expect that the diffuse structure might be because of RRM domains, which are flexible, soluble and not involved in fibril formation.

From DIC microscopy, we observed the phase separation that is the characteristic feature of all RNA binding protein containing low complexity region as discussed in *Introduction* section. In QRD44, at lower concentration, liquid droplets were seen, and higher concentration, gels is observed. Although we did not perform time course study in DIC for all constructs, but we assumed that these liquid droplets fused together form gel like structure which are less mobile and later they form large fibrils as seen in RRM123_QRD44_mut. We speculate that these droplets and gels possess β-sheet propensity, which matches to positive ThT results.

For the solid-state NMR samples, despite trying lower concentration and seeding protocol, all samples in QRD44 shows polymorphism as we observed 2 set of peaks in few regions in ^{13}C-^{13}C correlation. Sequential assignment was not possible in this protein due to poor quality of spectra in ^{13}C-^{15}N correlation spectra. Two reasons can be speculated for this observation: a) due to heterogeneous sample b) high occurrence glutamine amino acid.

Nevertheless, we assign the spin system of amino acids and all possess β-sheet propensity.

With respect to RRMs, from 2D PDSD we speculate that few regions in RRMs are rigid along with Q-rich domain that stabilises the fibrils structure. We confirmed this as not so many peaks appeared in RRM123_QRD44_WT as compared to Q-rich construct. We also proved that linker between RRM-1 and RRM-2 play crucial role by solid-state NMR experiments of mutated RRM123_QRD44_mut. Comparing mutant and WT constructs, the overall spectrum is very similar so these mutated amino acid does not affect the core structure of fibrils. It could be possible that whole linker along with some RRM1 residues are involved in fibril structure.

We proposed the initial model for the domain arrangement in the TIA-1 fibrils. We speculate that RRM1 domain along with first linker to be involved in the fibril core. **(Figure 64)** In this model we speculate that linker between RRM1 and RRM2 named as linker 12 is close to Q-rich fibrils while the other domains and regions are flexible or more dynamic**.** From previous papers we know that RRM2 and RRM3 has high affinity for the RNA and weak affinity of RRM1 for TIA-1 protein[187].

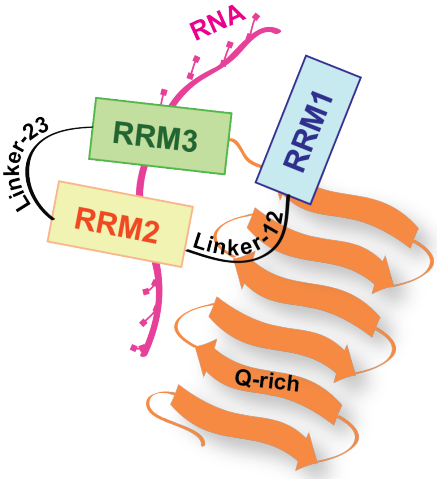


***Figure 64**: Model for relative orientation of RRMs with respect to fibrils in TIA-1*

5.5 Conclusion and Outlook

To conclude, we characterize the Q-rich domain in monomer and fibril state with and without RRM domains. We proved that linker between RRM1 and RRM2 near Q-rich domain is involved in fibril which might stabilize the fibrils. We proposed the model about domain arrangement with respect to one another. This can help in understanding the overall orientation of TIA-1 protein in U1 snRNP complex. There are still few open questions which need to address like role of RNA in fibril formation and how does RNA affect the fibrils structure. Furthermore, it will be interesting to know how HSP70 dismantles the fibrils in structural level, which is known for stress granules regulation.

6 References

1. Ross, C. a & Poirier, M. a. Protein aggregation and neurodegenerative disease. *Nature medicine* **10 Suppl,** S10-7 (2004).
2. Doyle, S. M., Genest, O. & Wickner, S. Protein rescue from aggregates by powerful molecular chaperone machines. *Nature Reviews Molecular Cell Biology* **14,** 617–629 (2013).
3. Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A. & Radford, S. E. A new era for understanding amyloid structures and disease. *Nature Reviews Molecular Cell Biology* **19,** 755–773 (2018).
4. Eberth, C. J. Zur Amyloidfrage. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* **84,** 111–118 (1881).
5. Snow, A. D., Willmer, J. & Kisilevsky, R. Sulfated glycosaminoglycans: a common constituent of all amyloids? *Laboratory investigation; a journal of technical methods and pathology* **56,** 120–3 (1987).
6. Sipe, J. D. *et al.* Amyloid fibril proteins and amyloidosis: chemical identification and clinical classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid : the international journal of experimental and clinical investigation : the official journal of the International Society of Amyloidosis* **23,** 1–5 (2016).
7. Kyle, R. A. Amyloidosis: A convoluted story. *British Journal of Haematology* **114,** 529–538 (2001).
8. Sipe, J. D. & Cohen, A. S. Review: History of the Amyloid Fibril. *Journal of Structural Biology* **130,** 88–98 (2000).
9. Möller, H. J. & Graeber, M. B. The case described by Alois Alzheimer in 1911. Historical and conceptual perspectives based on the clinical record and neurohistological sections. *European archives of psychiatry and clinical neuroscience* **248,** 111–22 (1998).
10. Fändrich, M. On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cellular and Molecular Life Sciences* **64,** 2066–2078 (2007).
11. Knowles, T. P. J., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nature reviews. Molecular cell biology* **15,** 384–96 (2014).
12. Morris, V. K. *et al.* Solid-state NMR spectroscopy of functional amyloid from a fungal hydrophobin: A well-ordered ??-sheet core amidst structural heterogeneity. *Angewandte Chemie - International Edition* **51,** 12621–12625 (2012).
13. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in

pituitary secretory granules. *Science* **325,** 328–332 (2009).

14. Furukawa, Y. *et al.* Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1832,** 1271–1278 (2017).
15. Fowler, D. M. *et al.* Functional amyloid formation within mammalian tissue. *PLoS Biology* **4,** 0100–0107 (2006).
16. Dobson, C. M., Šali, A. & Karplus, M. Protein folding: A perspective from theory and experiment. *Angewandte Chemie - International Edition* **37,** 868–893 (1998).
17. Na, J. H., Lee, W. K. & Yu, Y. G. How do we study the dynamic structure of unstructured proteins: A case study on nopp140 as an example of a large, intrinsically disordered protein. *International Journal of Molecular Sciences* **19,** (2018).
18. Jahn, T. R., Parker, M. J., Homans, S. W. & Radford, S. E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nature structural & molecular biology* **13,** 195–201 (2006).
19. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–90 (2003).
20. Canet, D. *et al.* Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature Structural Biology* (2002). doi:10.1038/nsb768
21. Silow, M. & Oliveberg, M. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 6084–6 (1997).
22. Fink, A. L. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design* **3,** R9–R23 (1998).
23. Arosio, P., Knowles, T. P. J. & Linse, S. On the lag phase in amyloid fibril formation. *Physical Chemistry Chemical Physics* **17,** 7606–7618 (2015).
24. Li, X., Rayman, J. B., Kandel, E. R. & Derkatch, I. L. Functional Role of Tia1/Pub1 and Sup35 Prion Domains: Directing Protein Synthesis Machinery to the Tubulin Cytoskeleton. *Molecular Cell* **55,** 305–318 (2014).
25. Lopez de Silanes, I. *et al.* Identification and Functional Outcome of mRNAs Associated with RNA-Binding Protein TIA-1. *Molecular and Cellular Biology* **25,** 9520–9531 (2005).
26. Dobson, C. M., Swoboda, B. E. P., Joniau, M. & Weissman, C. The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* **356,** 133–145 (2001).
27. Salahuddin, P., Fatima, M. T., Abdelhameed, A. S., Nusrat, S. & Khan, R. H. Structure of amyloid oligomers and their mechanisms of toxicities: Targeting amyloid oligomers using novel therapeutic approaches. *European Journal of Medicinal Chemistry* **114,** 41–58 (2016).

28. Sikkink, L. A. & Ramirez-Alvarado, M. Cytotoxicity of amyloidogenic immunoglobulin light chains in cell culture. *Cell Death and Disease* **1,** e98–e98 (2010).
29. Fändrich, M. Oligomeric intermediates in amyloid formation: Structure determination and mechanisms of toxicity. *Journal of Molecular Biology* **421,** 427–440 (2012).
30. Colvin, M. T. *et al.* Atomic Resolution Structure of Monomorphic A??42 Amyloid Fibrils. *Journal of the American Chemical Society* **138,** 9663–9674 (2016).
31. Tycko, R. Proteopathic Seeds and Neurodegenerative Diseases. *Proteopathic Seeds and Neurodegenerative Diseases* 19–28 (2013). doi:10.1007/978-3-642-35491-5
32. Rambaran, R. N. & Serpell, L. C. Amyloid fibrils: abnormal protein assembly. *Prion* **2,** 112–7 (2008).
33. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–457 (2007).
34. Fitzpatrick, A. W. P. *et al.* Atomic structure and hierarchical assembly of a cross-amyloid fibril. *Proceedings of the National Academy of Sciences* **110,** 5468–5473 (2013).
35. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proceedings of the National Academy of Sciences of the United States of America* **113,** E4976-84 (2016).
36. Heise, H. Solid-state NMR spectroscopy of amyloid proteins. *ChemBioChem* **9,** 179–189 (2008).
37. Petkova, A. T. *et al.* A structural model for Alzheimer's -amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences* **99,** 16742–16747 (2002).
38. Schütz, A. K. *et al.* Atomic-resolution three-dimensional structure of amyloid b fibrils bearing the osaka mutation. *Angewandte Chemie - International Edition* (2015). doi:10.1002/anie.201408598
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. Lim, K. H. *et al.* Solid-State NMR Studies Reveal Native-like β-Sheet Structures in Transthyretin Amyloid. *Biochemistry* **55,** 5272–5278 (2016).
41. Iwata, K. *et al.* 3D structure of amyloid protofilaments of beta2-microglobulin fragment probed by solid-state NMR. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 18119–24 (2006).
42. Lv, G. *et al.* Structural comparison of mouse and human α-synuclein amyloid fibrils by solid-state NMR. *Journal of Molecular Biology* **420,** 99–111 (2012).
43. Davies, D. R. & Segal, E. A. avid M. Three-dimensional structure of immunoglobulins.

Annual review of biochemistry 961–97 (1975). doi:https://doi.org/10.1146/annurev.bi.44.070175.003231

44. Mix, E., Goertsches, R. & Zett, U. K. Immunoglobulins - Basic considerations. *Journal of Neurology* **253,** 9–17 (2006).
45. Späth, P. J. Structure and function of immunoglobulins. *Sepsis* **3,** 197–218 (1999).
46. Williams, S. C. *et al.* Sequence and evolution of the human germline V lambda repertoire. *Journal of molecular biology* **264,** 220–232 (1996).
47. Solomon, A. & Weiss, D. T. Structural and Functional Properties of Human -Light-Chain Variable-Region Subgroups. **2,** 387–394 (1995).
48. Smith, G. P., Hood, L. & Fitch, walter M. Antibody Diversity. *Annual Review of Biochemistry* **40,** 969–1012 (1971).
49. Lesk, A. M. & Chothia, C. Evolution of proteins formed by β-sheets. II. The core of the immunoglobulin domains. *Journal of Molecular Biology* **160,** 325–342 (1982).
50. Feige, M. J. & Buchner, J. Principles and engineering of antibody folding and assembly. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1844,** 2024–2031 (2014).
51. Bork, P., Holm, L. & Sander, C. The immunoglobulin fold. Structural classification, sequence patterns and common core. *Journal of molecular biology* **242,** 309–20 (1994).
52. Seidman, J., Leder, A., Nau, M., Norman, B. & Leder, P. Antibody diversity. *Science* **202,** 11–17 (1978).
53. SMITHIES, O. Gamma-Globulin Variability: A Genetic Hypothesis. *Nature* **199,** 1231–1236 (1963).
54. Kindt, T. J., Goldsby, R. A., Osborne, B. A. & Kuby, J. *Kuby immunology.* (W.H. Freeman, 2007).
55. de Villartay, J.-P., Fischer, A. & Durandy, A. The mechanisms of immune diversification and their disorders. *Nature reviews. Immunology* **3,** 962–72 (2003).
56. Janeway, C. *Immunobiology 5 : the immune system in health and disease*. (Garland Pub, 2001).
57. MAGNUS-LEVY, A. Amyloidosis in multiple-myeloma; progress noted in 50 years of personal observation. *Journal of the Mount Sinai Hospital, New York* **19,** 8–9
58. OSSERMAN, E. F., TAKATSUKI, K. & TALAL, N. MULTIPLE MYELOMA I. THE PATHOGENESIS OF "AMYLOIDOSIS. *Seminars in hematology* **1,** 3–85 (1964).
59. Apitz, K. Die Paraproteinosen. *Virchows Archiv für Pathologische Anatomie und Physiologie und für Klinische Medizin* **306,** 631–699 (1940).
60. MAGNUS-LEVY, A. Multiple Myelome (XII). *Acta Medica Scandinavica* **95,** 217–280 (2009).

61. EANES, E. D. & GLENNER, G. G. X-RAY DIFFRACTION STUDIES ON AMYLOID FILAMENTS. *Journal of Histochemistry & Cytochemistry* **16,** 673–677 (1968).
62. Merlini, G. *et al.* Systemic immunoglobulin light chain amyloidosis. *Nature Reviews Disease Primers* **4,** 38 (2018).
63. Milani, P., Merlini, G. & Palladini, G. Light chain amyloidosis. *Mediterranean Journal of Hematology and Infectious Diseases* **10,** 1–14 (2018).
64. Sanders, P. W. & Booker, B. B. Pathobiology of cast nephropathy from human Bence Jones proteins. *The Journal of clinical investigation* **89,** 630–9 (1992).
65. Katzmann, J. A. *et al.* Serum reference intervals and diagnostic ranges for free kappa and free lambda immunoglobulin light chains: relative sensitivity for detection of monoclonal light chains. *Clinical chemistry* **48,** 1437–44 (2002).
66. Abraham, R. R. S. *et al.* Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain–associated amyloidosis (AL). *Blood* **101,** 3801–3808 (2003).
67. De Beer, F. C. & Pepys, M. B. Isolation of human C-reactive protein and serum amyloid P component. *Journal of immunological methods* **50,** 17–31 (1982).
68. Likó, I. *et al.* Evidence for an extended interacting surface between beta-amyloid and serum amyloid P component. *Neuroscience letters* **412,** 51–5 (2007).
69. Baltz, M. L. *et al.* Circulating serum amyloid P component is the precursor of amyloid P component in tissue amyloid deposits. *Clinical and experimental immunology* **66,** 691–700 (1986).
70. Enqvist, S., Sletten, K., Stevens, F. J., Hellman, U. & Westermark, P. Germ line origin and somatic mutations determine the target tissues in systemic AL-amyloidosis. *PLoS ONE* **2,** (2007).
71. Del Pozo Yauner, L. *et al.* Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains. *Proteins: Structure, Function and Genetics* **72,** 684–692 (2008).
72. Perfetti, V. *et al.* Analysis of V(lambda)-J(lambda) expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r (lambdaIII) as a new amyloid-associated germline gene segment. *Blood* **100,** 948–953 (2002).
73. Wall, J. *et al.* Thermodynamic instability of human λ6 light chains: Correlation with fibrillogenicity. *Biochemistry* (1999). doi:10.1021/bi991131j
74. Stevens, F. J. *et al.* A Molecular Model for Self-Assembly of Amyloid Fibrils: Immunoglobulin Light Chains. *Biochemistry* **34,** 10697–10702 (1995).
75. Poshusta, T. L. *et al.* Mutations in Specific Structural Regions of Immunoglobulin Light Chains Are Associated with Free Light Chain Levels in Patients with AL Amyloidosis. *PLoS ONE* **4,** e5169 (2009).

76. Baden, E. M., Randles, E. G., Aboagye, A. K., Thompson, J. R. & Ramirez-Alvarado, M. Structural insights into the role of mutations in amyloidogenesis. *Journal of Biological Chemistry* **283,** 30950–30956 (2008).
77. Weber, B. *et al.* A single residue switch reveals principles of antibody domain integrity. *Journal of Biological Chemistry* **293,** 17107–17118 (2018).
78. Hernández-Santoyo, A. *et al.* A single mutation at the sheet switch region results in conformational changes favoring λ6 light-chain fibrillogenesis. *Journal of Molecular Biology* **396,** 280–292 (2010).
79. Bhavaraju, M. & Hansmann, U. H. E. Effect of single point mutations in a form of systemic amyloidosis. *Protein Science* **24,** 1451–1462 (2015).
80. Hurle, M. R., Helms, L. R., Li, L., Chan, W. & Wetzel, R. A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proceedings of the National Academy of Sciences of the United States of America* **91,** 5446–5450 (1994).
81. Mukherjee, S., Pondaven, S. P., Hand, K., Madine, J. & Jaroniec, C. P. Effect of amino acid mutations on the conformational dynamics of amyloidogenic immunoglobulin light-chains: A combined NMR and in silico study. *Scientific Reports* **7,** 10339 (2017).
82. Kobayashi, Y. *et al.* Decreased amyloidogenicity caused by mutational modulation of surface properties of the immunoglobulin light chain BRE variable domain. *Biochemistry* **53,** 5162–5173 (2014).
83. Nokwe, C. N. *et al.* A residue-specific shift in stability and amyloidogenicity of antibody variable domains. *The Journal of biological chemistry* **289,** 26829–46 (2014).
84. Blancas-Mejía, L. M. M. *et al.* Kinetic control in protein folding for light chain amyloidosis and the differential effects of somatic mutations. *Journal of Molecular Biology* **426,** 347–61 (2014).
85. Simpson, E. R., Herold, E. M. & Buchner, J. The folding pathway of the antibody V(L) domain. *Journal of molecular biology* **392,** 1326–38 (2009).
86. Qin, Z., Hu, D., Zhu, M. & Fink, A. L. Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. *Biochemistry* **46,** 3521–3531 (2007).
87. Chiti, F. *et al.* Detection of two partially structured species in the folding process of the amyloidogenic protein beta 2-microglobulin. *Journal of molecular biology* **307,** 379–391 (2001).
88. Souillac, P. O., Uversky, V. N. & Fink, A. L. Structural transformations of oligomeric intermediates in the fibrillation of the immunoglobulin light chain LEN. *Biochemistry* **42,** 8094–8104 (2003).
89. Khurana, R. *et al.* Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* **40,** 3525–3535 (2001).

90. Liu, Y. & Eisenberg, D. 3D domain swapping: As domains continue to swap. *Protein Science* **11,** 1285–1299 (2002).
91. Heringa, J. & Taylor, W. R. Three-dimensional domain duplication, swapping and stealing. *Current Opinion in Structural Biology* (1997). doi:10.1016/S0959-440X(97)80060-7
92. Sonnen, A. F. P. *et al.* Domain metastability: A molecular basis for immunoglobulin deposition? *Journal of Molecular Biology* **399,** 207–213 (2010).
93. Sakata, M. *et al.* Kinetic coupling of folding and prolyl isomerization of beta2-microglobulin studied by mutational analysis. *Journal of molecular biology* **382,** 1242–55 (2008).
94. Hora, M. *et al.* MAK33 antibody light chain amyloid fibrils are similar to oligomeric precursors. *PLoS ONE* **12,** 1–14 (2017).
95. Marin-Argany, M. *et al.* Cell Damage in Light Chain Amyloidosis. *Journal of Biological Chemistry* **291,** 19813–19825 (2016).
96. Dasari, S. *et al.* Proteomic detection of immunoglobulin light chain variable region peptides from amyloidosis patient biopsies. *Journal of Proteome Research* **14,** 1957–1967 (2015).
97. Glenner, G. G., Harbaugh, J., Ohms, J. I., Harada, M. & Cuatrecasas, P. An amyloid protein: The amino-terminal variable fragment of an immunoglobulin light chain. *Biochemical and Biophysical Research Communications* (1970). doi:10.1016/0006-291X(70)90227-5
98. Morgan, G. J. & Kelly, J. W. The Kinetic Stability of a Full-Length Antibody Light Chain Dimer Determines whether Endoproteolysis Can Release Amyloidogenic Variable Domains. *Journal of Molecular Biology* **428,** 4280–4297 (2016).
99. Glenner, G. G. *et al.* Creation of "amyloid" fibrils from Bence Jones proteins in vitro. *Science (New York, N.Y.)* **174,** 712–4 (1971).
100. Klimtchuk, E. S. *et al.* The critical role of the constant region in thermal stability and aggregation of amyloidogenic immunoglobulin light chain. *Biochemistry* **49,** 9848–9857 (2010).
101. Blancas-Mejía, L. M. *et al.* Thermodynamic and fibril formation studies of full length immunoglobulin light chain AL-09 and its germline protein using scan rate dependent thermal unfolding. *Biophysical Chemistry* **207,** 13–20 (2015).
102. Mclaughlin, R. W., De Stigter, J. K., Sikkink, L. A., Baden, E. M. & Ramirez-Alvarado, M. The effects of sodium sulfate, glycosaminoglycans, and Congo red on the structure, stability, and amyloid formation of an immunoglobulin light-chain protein. *Protein Science* (2006). doi:10.1110/ps.051997606
103. Blancas-Mejía, L. M., Hammernik, J., Marin-Argany, M. & Ramirez-Alvarado, M.

Differential effects on light chain amyloid formation depend on mutations and type of glycosaminoglycans. *The Journal of biological chemistry* **290,** 4953–65 (2015).

104. Ren, R. *et al.* Role of glycosaminoglycan sulfation in the formation of immunoglobulin light chain amyloid oligomers and fibrils. *Journal of Biological Chemistry* **285,** 37672–37682 (2010).
105. Hawkins, P. ., Lavender, J. ., Myers, M. . & Pepys, M. . DIAGNOSTIC RADIONUCLIDE IMAGING OF AMYLOID: BIOLOGICAL TARGETING BY CIRCULATING HUMAN SERUM AMYLOID P COMPONENT. *The Lancet* **331,** 1413–1418 (1988).
106. Richards, D. B. *et al.* Therapeutic Clearance of Amyloid by Antibodies to Serum Amyloid P Component. *New England Journal of Medicine* **373,** 1106–1114 (2015).
107. Gillmore, J. D. *et al.* Sustained pharmacological depletion of serum amyloid P component in patients with systemic amyloidosis. *British Journal of Haematology* **148,** 760–767 (2010).
108. Blancas-Mejía, L. M. & Ramirez-Alvarado, M. Recruitment of Light Chains by Homologous and Heterologous Fibrils Shows Distinctive Kinetic and Conformational Specificity. *Biochemistry* **14,** acs.biochem.6b00090 (2016).
109. Martin, E. B. *et al.* Differential recruitment efficacy of patientderived amyloidogenic and myeloma light chain proteins by synthetic fibrilsA metric for predicting amyloid propensity. *PLoS ONE* **12,** 1–18 (2017).
110. Novotný, J. & Haber, E. Structural invariants of antigen binding: comparison of immunoglobulin VL-VH and VL-VL domain dimers. *Proceedings of the National Academy of Sciences of the United States of America* **82,** 4592–6 (1985).
111. Baden, E. M. *et al.* Altered dimer interface decreases stability in an amyloidogenic protein. *Journal of Biological Chemistry* **283,** 15853–15860 (2008).
112. Peterson, F. C., Baden, E. M., Owen, B. A. L., Volkman, B. F. & Ramirez-Alvarado, M. A single mutation promotes amyloidogenicity through a highly promiscuous dimer interface. *Structure* **18,** 563–570 (2010).
113. Kaplan, B., Livneh, A. & Sela, B.-A. Immunoglobulin free light chain dimers in human diseases. *TheScientificWorldJournal* **11,** 726–735 (2011).
114. Brumshtein, B. *et al.* Formation of amyloid fibers by monomeric light chain variable domains. *Journal of Biological Chemistry* **289,** 27513–27525 (2014).
115. DiCostanzo, A. C., Thompson, J. R., Peterson, F. C., Volkman, B. F. & Ramirez-Alvarado, M. Tyrosine residues mediate fibril formation in a dynamic light chain dimer interface. *Journal of Biological Chemistry* **287,** 27997–28006 (2012).
116. Arndt, K. M., Müller, K. M. & Plückthun, A. Factors influencing the dimer to monomer transition of an antibody single-chain Fv fragment. *Biochemistry* **37,** 12918–12926

(1998).

117. Blancas-Mejía, L. M., Misra, P. & Ramirez-Alvarado, M. Differences in Protein Concentration Dependence for Nucleation and Elongation in Light Chain Amyloid Formation. *Biochemistry* **56,** 757–766 (2017).
118. Kyle, R. *et al.* Incidence and natural history of primary systemic amyloidosis in Olmsted County, Minnesota, 1950 through 1989 [see comments]. *Blood* **79,** (1992).
119. Matsuda, M. *et al.* Serum levels of free light chain before and after chemotherapy in primary systemic AL amyloidosis. *Internal medicine (Tokyo, Japan)* **44,** 428–433 (2005).
120. Kumar, S. *et al.* Serum immunoglobulin free light chain in primary amyloidosis: Prognostic value and correlations with clinical features. *Blood* **112,** 2733a (2008).
121. Dispenzieri, A. *et al.* Serum cardiac troponins and N-terminal pro-brain natriuretic peptide: A staging system for primary systemic amyloidosis. *Journal of Clinical Oncology* (2004). doi:10.1200/JCO.2004.03.029
122. Gertz, M. A. Immunoglobulin light chain amyloidosis diagnosis and treatment algorithm 2018. *Blood Cancer Journal* **8,** 44 (2018).
123. Kastritis, E. & Dimopoulos, M. A. Recent advances in the management of AL Amyloidosis. *British Journal of Haematology* **172,** 170–186 (2016).
124. Pepys, M. B. *et al.* Targeted pharmacological depletion of serum amyloid P component for treatment of human amyloidosis. *Nature* **417,** 254–259 (2002).
125. Pepys, M. New therapeutic perspectives – amyloid removal. *Orphanet Journal of Rare Diseases* **10,** I17 (2015).
126. Cooley, C. B. *et al.* Unfolded protein response activation reduces secretion and extracellular aggregation of amyloidogenic immunoglobulin light chain. *Proceedings of the National Academy of Sciences* **111,** 13046–13051 (2014).
127. Wall, J. S. *et al.* AL Amyloid Imaging and Therapy with a Monoclonal Antibody to a Cryptic Epitope on Amyloid Fibrils. *PLoS ONE* (2012). doi:10.1371/journal.pone.0052686
128. Gertz, M. A. *et al.* First-in-human phase I/II study of NEOD001 in patients with light chain amyloidosis and persistent organ dysfunction. *Journal of Clinical Oncology* **34,** 1097–1103 (2016).
129. Wall, J. S. *et al.* Bifunctional amyloid-reactive peptide promotes binding of antibody 11-1F4 to diverse amyloid types and enhances therapeutic efficacy. *Proceedings of the National Academy of Sciences* 201805515 (2018). doi:10.1073/pnas.1805515115
130. Kim, H.-S., Quon, M. J. & Kim, J.-A. New insights into the mechanisms of polyphenols beyond antioxidant properties; lessons from the green tea polyphenol, epigallocatechin 3-gallate. *Redox biology* **2,** 187–195 (2014).

131. Bastianetto, S., Yao, Z., Papadopoulos, V. & Quirion, R. Neuroprotective effects of green and black teas and their catechin gallate esters against beta-amyloid-induced toxicity. *The European journal of neuroscience* **23,** 55–64 (2006).
132. Azam, S., Hadi, N., Khan, N. U. & Hadi, S. M. Prooxidant property of green tea polyphenols epicatechin and epigallocatechin-3-gallate: Implications for anticancer properties. *Toxicology in Vitro* **18,** 555–561 (2004).
133. Bieschke, J. *et al.* EGCG remodels mature alpha-synuclein and amyloid-beta fibrils and reduces cellular toxicity. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 7710–5 (2010).
134. Townsend, D. *et al.* Epigallocatechin-3-gallate remodels apolipoprotein A-I amyloid fibrils into soluble oligomers in the presence of heparin. *Journal of Biological Chemistry* **293,** 12877–12893 (2018).
135. Ehrnhoefer, D. E. *et al.* EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. *Nature structural & molecular biology* **15,** 558–66 (2008).
136. Palhano, F. L., Lee, J., Grimster, N. P. & Kelly, J. W. Toward the molecular mechanism(s) by which EGCG treatment remodels mature amyloid fibrils. *Journal of the American Chemical Society* **135,** 7503–10 (2013).
137. Hunstein, W. Epigallocathechin-3-gallate in AL amyloidosis: a new therapeutic option? *Blood* **110,** 2216 (2007).
138. A Randomised Trial for the Treatment of Cardiac AMyloid Light-chain Amyloidosis With the Green Tea Compound Epigallocatechin-3-gallate (TAME-AL) - ICH GCP - Clinical Trials Registry. Available at: https://ichgcp.net/clinical-trials-registry/NCT02015312. (Accessed: 23rd November 2018)
139. Mereles, D. & Hunstein, W. Epigallocatechin-3-gallate (EGCG) for Clinical Trials: More Pitfalls than Promises? *International journal of molecular sciences* **12,** 5592–603 (2011).
140. Lee, M.-J. *et al.* Pharmacokinetics of tea catechins after ingestion of green tea and (-)-epigallocatechin-3-gallate by humans: formation of different metabolites and individual variability. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **11,** 1025–32 (2002).
141. Grelle, G. *et al.* Black tea theaflavins inhibit formation of toxic amyloid-β and α-synuclein fibrils. *Biochemistry* **50,** 10624–36 (2011).
142. He, J., Xing, Y.-F., Huang, B., Zhang, Y.-Z. & Zeng, C.-M. Tea catechins induce the conversion of preformed lysozyme amyloid fibrils to amorphous aggregates. *Journal of agricultural and food chemistry* **57,** 11391–6 (2009).
143. Hora, M. *et al.* Solid- and solution-state nuclear magnetic resonance spectroscopic

studies on antibody light chain amyloid formation and interactions with epigallocatechin gallate. *Amyloid* **24,** 10 (2017).

144. King, O. D., Gitler, A. D. & Shorter, J. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Research* **1462,** 61–80 (2012).
145. Courchaine, E. M., Lu, A. & Neugebauer, K. M. Droplet organelles? *The EMBO Journal* **35,** 1603–1612 (2016).
146. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).
147. Waris, S., Wilce, M. C. J. & Wilce, J. A. RNA recognition and stress granule formation by TIA proteins. *International Journal of Molecular Sciences* **15,** 23377–23388 (2014).
148. Brangwynne, C. P. Phase transitions and size scaling of membrane-less organelles. *The Journal of cell biology* **203,** 875–881 (2013).
149. Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular Cell* **60,** 231–241 (2015).
150. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell* **60,** 208–219 (2015).
151. Qamar, S. *et al.* FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **173,** 720–734.e15 (2018).
152. Kedersha, Gupta, M., Li, W., Miller, I. & Anderson, P. eIF-2 ␣ to the Assembly of Mammalian Stress Granules. *The Journal of Cell biology* **147,** 1431–1441 (1999).
153. Muchowski, P. J. *et al.* Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 7841–6 (2000).
154. Gilks, N. *et al.* Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Molecular biology of the cell* **15,** 5383–98 (2004).
155. Dobra, I., Pankivskyi, S., Samsonova, A., Pastre, D. & Hamon, L. Relation Between Stress Granules and Cytoplasmic Protein Aggregates Linked to Neurodegenerative Diseases. **7,** (2018).
156. Cavanagh, J. *Protein NMR spectroscopy : principles and practice*. (Academic Press, 2007).
157. Keeler, J. *Understanding NMR Spectroscopy*. (2002).
158. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*. *Systems Biology* (2007). doi:10.1093/cid/cis040
159. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J. & Rance, M. *Protein*

NMR Spectroscopy. *Protein NMR Spectroscopy* **8,** (2007).

160. Ferella, L., Rosato, A. & Turano, P. What Can be Learned About the Structure and Dynamics of Biomolecules from NMR. *NMR of Biomolecules: Towards Mechanistic Systems Biology* 33–45 (2012). doi:10.1002/9783527644506.ch3
161. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89,** 496–514 (1990).
162. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy* **34,** 93–158 (1999).
163. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89,** 496–514 (1990).
164. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1814,** 942–968 (2011).
165. Linser, R., Fink, U. & Reif, B. Proton-detected scalar coupling based assignment strategies in MAS solid-state NMR spectroscopy applied to perdeuterated proteins. *Journal of Magnetic Resonance* **193,** 89–93 (2008).
166. 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).
167. Shevelkov, V. Development of MAS solid state NMR methods for structural and dynamical characterization of biomolecules Dissertation.
168. Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420,** 98–102 (2002).
169. Lacabanne, D., Meier, B. H. & Böckmann, A. Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *Journal of Biomolecular NMR* **71,** 141–150 (2018).
170. Reif, B. Ultra-high resolution in MAS solid-state NMR of perdeuterated proteins: Implications for structure and dynamics. *Journal of Magnetic Resonance* **216,** 1–12 (2012).
171. Manolikas, T., Herrmann, T. & Meier, B. H. Protein structure determination from 13C spin-diffusion solid-state NMR spectroscopy. *Journal of the American Chemical Society* (2008). doi:10.1021/ja078039s

172. Nieuwkoop, A. J. & Rienstra, C. M. Supramolecular protein structure determination by site-specific long-range intermolecular solid state NMR spectroscopy. *Journal of the American Chemical Society* (2010). doi:10.1021/ja100992y
173. Jaroniec, C. P. *et al.* High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences* (2004). doi:10.1073/pnas.0304849101
174. 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
175. 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
176. Tuttle, M. D. *et al.* Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nature Structural and Molecular Biology* **23,** 409–415 (2016).
177. Van Melckebeke, H. *et al.* Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state nmr spectroscopy. *Journal of the American Chemical Society* (2010). doi:10.1021/ja104213j
178. Retel, J. S. *et al.* Structure of outer membrane protein G in lipid bilayers. *Nature Communications* (2017). doi:10.1038/s41467-017-02228-2
179. Morag, O., Sgourakis, N. G., Baker, D. & Goldbourt, A. The NMR–Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proceedings of the National Academy of Sciences* (2015). doi:10.1073/pnas.1415393112
180. Barbet-Massin, E. & Pintacuda, G. Biomolecular Solid-State NMR/Basics. *NMR of Biomolecules: Towards Mechanistic Systems Biology* 345–364 (2012). doi:10.1002/9783527644506.ch20
181. Hediger, S., Meier, B. H., Kurur, N. D., Bodenhausen, G. & Ernst, R. R. NMR cross polarization by adiabatic passage through the Hartmann—Hahn condition (APHH). *Chemical Physics Letters* **223,** 283–288 (1994).
182. Pauli, J., Baldus, M., Van Rossum, B., De Groot, H. & Oschkinat, H. Backbone and side-chain13C and15N signal assignments of the α-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. *ChemBioChem* **2,** 272–281 (2001).
183. Higman, V. A. Solid-state MAS NMR resonance assignment methods for proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **106–107,** 37–65 (2018).
184. Gullion, T. & Schaefer, J. Rotational-echo double-resonance NMR. *Journal of Magnetic Resonance (1969)* **81,** 196–200 (1989).
185. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (2005). doi:10.1385/1-59259-890-0:571

186. Feige, M. J. *et al.* Dissecting the alternatively folded state of the antibody Fab fragment. *Journal of molecular biology* **399,** 719–30 (2010).
187. Wang, I. *et al.* Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1. *Nucleic Acids Research* **42,** 5949–5966 (2014).
188. Schägger, H. Tricine–SDS-PAGE. *Nature Protocols* **1,** 16–22 (2006).
189. Hyberts, S. G., Arthanari, H. & Wagner, G. Applications of non-uniform sampling and processing. *Topics in current chemistry* **316,** 125–48 (2012).
190. Morcombe, C. R. & Zilm, K. W. Chemical shift referencing in MAS solid state NMR. *Journal of Magnetic Resonance* **162,** 479–486 (2003).
191. Rienstra, C. M. *et al.* De novo determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy. *Proceedings of the National Academy of Sciences* **99,** 10260–10265 (2002).
192. Hyberts, S. G., Arthanari, H. & Wagner, G. Applications of non-uniform sampling and processing. *Topics in current chemistry* **316,** 125–48 (2012).
193. Annamalai, K. *et al.* Common Fibril Structures Imply Systemically Conserved Protein Misfolding Pathways In Vivo. *Angewandte Chemie - International Edition* **56,** 7510–7514 (2017).
194. Mukherjee, S., Pondaven, S. P. & Jaroniec, C. P. Conformational flexibility of a human immunoglobulin light chain variable domain by relaxation dispersion nuclear magnetic resonance spectroscopy: implications for protein misfolding and amyloid assembly. *Biochemistry* **50,** 5845–5857 (2011).
195. Annamalai, K., Liberta, F., Vielberg, M., Close, W. & Lilie, H. Systemically conserved pathways of protein misfolding in vivo.
196. Nokwe, C. N. *et al.* A Stable Mutant Predisposes Antibody Domains to Amyloid Formation through Specific Non-Native Interactions. *Journal of Molecular Biology* **428,** 1315–1332 (2016).
197. Andrich, K. *et al.* Aggregation of Full Length Immunoglobulin Light Chains from AL Amyloidosis Patients Is Remodeled by Epigallocatechin-3-gallate. *The Journal of biological chemistry* **292,** 2328–2344 (2016).
198. Rennella, E., Morgan, G. J., Kelly, J. W. & Kay, L. E. Role of domain interactions in the aggregation of full-length immunoglobulin light chains [Biophysics and Computational Biology]. *Proceedings of the National Academy of Sciences of the United States of America* 1–10 (2018). doi:10.1073/pnas.1817538116
199. Hora, M. *et al.* Epigallocatechin-3-gallate preferentially induces aggregation of amyloidogenic immunoglobulin light chains. *Scientific Reports* **7,** 41515 (2017).
200. Swuec, P. *et al.* Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain (AL) amyloidosis patient. doi:10.1101/444901

201. Piehl, D. W., Blancas-Mejía, L. M., Ramirez-Alvarado, M. & Rienstra, C. M. Solid-state NMR chemical shift assignments for AL-09 VL immunoglobulin light chain fibrils. *Biomolecular NMR Assignments* 1–6 (2016). doi:10.1007/s12104-016-9718-3
202. Piehl, D. W. *et al.* Immunoglobulin Light Chains Form an Extensive and Highly Ordered 2 Fibril Involving the N- and C - Terminus 1. *ACS Omega* **2,** 712–720 (2017).
203. Lecoq, L. *et al.* A Substantial Structural Conversion of the Native Monomer Leads to In-Register Parallel Amyloid Fibril Formation in Light Chain Amyloidosis. *ChemBioChem* (2018). doi:10.1002/cbic.201800732
204. Petkova, A. T. Self-Propagating, Molecular-Level Polymorphism in Alzheimer's -Amyloid Fibrils. *Science* **307,** 262–265 (2005).
205. Tomlinson, J. H., Ullah, S., Hansen, P. E. & Williamson, M. P. Characterization of salt bridges to lysines in the protein G B1 domain. *Journal of the American Chemical Society* **131,** 4674–4684 (2009).
206. Hackman, P. *et al.* Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. *Annals of Neurology* **73,** 500–509 (2013).
207. Förch, P. *et al.* The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Molecular Cell* **6,** 1089–1098 (2000).
208. Izquierdo, J. M. & Valcárcel, J. Fas-activated Serine/Threonine Kinase (FAST K) synergizes with TIA-1/TIAR proteins to regulate Fas alternative splicing. *Journal of Biological Chemistry* **282,** 1539–1543 (2007).
209. Förch, P., Puig, O., Martínez, C., Séraphin, B. & Valcárcel, J. The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5′ splice sites. *EMBO Journal* **21,** 6882–6892 (2002).
210. Mackenzie, I. R., Nicholson, A. M., Boylan, K. B., Taylor, J. P. & Rademakers Correspondence, R. TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics In Brief. *Neuron* **95,** 808–816.e9 (2017).
211. Rayman, J. B., Karl, K. A. & Kandel, E. R. TIA-1 Self-Multimerization, Phase Separation, and Recruitment into Stress Granules Are Dynamically Regulated by Zn2+. *Cell Reports* **22,** 59–71 (2018).
212. Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on13C chemical shift statistics. *Journal of Biomolecular NMR* **24,** 149–154 (2002).
213. Kedersha, N. & Anderson, P. Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. *Biochemical Society transactions* **30,** 963–969 (2002).

REFERENCES

7 Appendix I

Abbreviations

1D, 2D, 3D	One, Two, Three-dimensional
β-ME	Beta-mercaptoethanol
AL amyloidosis	Antibody Light chain amyloidosis
BRMB	Biological Magnetic Resonance Data bank
CDR	Complementarity determining region
CP	Cross Polarization
CSA	Chemical shift anisotropy
CSI	Chemical shift Index
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetra acetic acid
EGCG	(-) Epigallocatechin 3-gallate
FID	Free Induction decay
GL	Germline
GSH	Glutathione (reduced)
GSSG	Glutathione disulphide (oxidised)
HPLC	High performance liquid chromatography
HSQC	Heteronuclear Single Quantum Coherence
IBs	Inclusion bodies
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
K_d	Dissociation constant
kDa	Kilo Dalton
LB	Lysogenic broth
LC	Light chain
MAS	Magic angle spinning
MW	Molecular weight
MWCOs	Molecular weight cut offs
NMR	Nuclear magnetic resonance
NUS	Non-uniform Sampling
OD_{600}	Optical density at 600nm
PDB	Protein data bank
PDSD	Proton driven spin diffusion

ppm	parts per million
Pre-mRNA	Precursor messenger RNA
REDOR	Rotational echo double resonance
Rpm	rotation per minute
RNA	Ribonucleic acid
RT	Room temperature 25^{o}C
S/N ratio	signal-to-noise ratio
SDS-PAGE electrophoresis	Sodium dodecyl sulphate polyacrylamide gel
SedNMR	Sedimentation NMR
SPECIFIC-CP	Spectrally induced filtering in combination with cross polarisation
SPINAL-64 decoupling steps	Small phase incremental alternation decoupling with 64
T_m	Melting temperature
TEDOR	Transferred echo double resonance
TEM	Transmission electron microscopy
TEV	Tobacco Etch Virus
TIA-1	T-cell Intracellular Antigen-1
ThT	Thioflavin T
UV	Ultraviolet
V_L	Variable domain of Immunoglobulin light chain

Appendix II

Primers

Protein	Forward Primer	Reverse Primer
Y32S- S3706 (V_L) WT-del	GTGATAGCCTGCgtAGCTATTCTGCA AGCTGGTATCAGCAGAA	TTCTGCTGATACCAGCTTGCAGAATA GCTacGCAGGCTATCAC
Y49-G50R-N52S-S3706 (V_L) WT-del	AGGCACCGGTTCTGGTTATTTTTCGT AAATCTAATCGTCCGAGCGGTATTCC	GGAATACCGCTCGGACGATTAGATTT ACGAAAAATAACCAGAACCGGTGCCT
G95A- S3706 (V_L) WT-del	GTAATAGCCGTGATAGCAGTGCAAAC CATCAGGTTTTTGGTGG	CCACCAAAAACCTGATGGTTTGCACT GCTATCACGGCTATTAC
R50G- S3706 (V_L) WT-del	CACCGGTTCTGGTTATTTTTGGTAAA TCTAATCGTCCGAGCGG	CCGCTCGGACGATTAGATTTACCAAA AATAACCAGAACCGGTG
TIA-1 RRM123 QRD44	GGGTAAGGAAGTCAAAGTGAATTGGg gtggtaacgcaaacccgCAAAAGAAA GATACAAGCAATCATT	AATGATTGCTTGTATCTTTCTTTTGc gggtttgcgttaccaccCCAATTCAC TTTGACTTCCTTACCC

Appendix III

Sequences

V_L protein

S3706_pat

SELTQDPA VSVALGQTVR ITCQGDSLRS YSASWYQQKP GQAPVLVIFR KSNRPSGIPD RFSGSSSGNT ASLTITGAQA EDEADYYCNS RDSSANHQVF GGGTKLTVLG PKAAPS$_{115}$

S3706_patdel

GSSELTQDPA VSVALGQTVR ITCQGDSLRS YSASWYQQKP GQAPVLVIFR KSNRPSGIPD RFSGSSSGNT ASLTITGAQA EDEADYYCNS RDSSANHQVF GGGTKLTVLG$_{110}$

S3706_GL

GSSELTQDPA VSVALGQTVR ITCQGDSLRS YYASWYQQKP GQAPVLVIYG KNNRPSGIPD RFSGSSSGNT ASLTITGAQA EDEADYYCNS RDSSGNHQVF GGGTKLTVLG$_{110}$

S3706_R50G

GSSELTQDPA VSVALGQTVR ITCQGDSLRS YSASWYQQKP GQAPVLVIFG KSNRPSGIPD RFSGSSSGNT ASLTITGAQA EDEADYYCNS RDSSANHQVF GGGTKLTVLG$_{110}$

TIA-1 protein

QRD44

ETLDMINPVQ QQNQIGYPQP YGQWGQWYGN AQQIGQYMPN GWQV$_{44}$

RRM123_QRD44_WT

MEDEMPKTLY VGNLSRDVTE ALILQLFSQI GPCKNCKMIM DTAGNDPYCF VEFHEHRHAA AALAAMNGRK IMGKEVKVNW ATTPSSQKKD TSNHFHVFVG DLSPEITTED IKAAFAPFGR ISDARVVKDM ATGKSKGYGF VSFFNKWDAE NAIQQMGGQW LGGRQIRTNW ATRKPPAPKS TYESNTKQLT YDEVVNQSSP SNCTVYCGGV TSGLTEQLMR QTFSPFGQIM EIRVFPDKGY SFVRFNSHES AAHAIVSVNG TTIEGHVVKC YWGKETLDMI NPVQQQNQIG YPQPYGQWGQ WYGNAQQIGQ YMPNGWQV$_{318}$

RRM123_QRD44_mut

MEDEMPKTLY VGNLSRDVTE ALILQLFSQI GPCKNCKMIM DTAGNDPYCF VEFHEHRHAA AALAAMNGRK IMGKEVKVNW GGNANPQKKD TSNHFHVFVG DLSPEITTED IKAAFAPFGR ISDARVVKDM ATGKSKGYGF VSFFNKWDAE NAIQQMGGQW LGGRQIRTNW ATRKPPAPKS TYESNTKQLT YDEVVNQSSP SNCTVYCGGV TSGLTEQLMR QTFSPFGQIM EIRVFPDKGY SFVRFNSHES AAHAIVSVNG TTIEGHVVKC YWGKETLDMI NPVQQQNQIG YPQPYGQWGQ WYGNAQQIGQ YMPNGWQV$_{318}$

Appendix IV

Solution state NMR assignment of S3706_patdel

RESIDUE NUMBER	AMINO ACID	H	N	CA	CB
3	Ser	8.12	117.55	55.96	61.43
4	Glu	8.28	121.56	54.37	28.33
5	Leu	8.22	122.60	50.92	41.33
6	Thr	9.13	117.11	58.92	68.21
7	Gln	8.71	122.08	51.41	31.27
8	Asp	8.85	124.33	50.52	37.90
10	Ala	7.66	119.19	49.75	18.47
11	Val	8.54	120.53	58.24	33.15
12	Ser	8.29	119.57	53.41	63.43
13	Val	8.58	120.23	56.29	32.93
14	Ala	8.27	128.29	48.54	16.52
15	Leu	7.92	121.26	54.08	40.00
16	Gly	9.71	113.99	42.32	
17	Gln	8.16	120.79	51.96	25.95
18	Thr	8.31	117.46	59.23	67.73
19	Val	8.69	126.86	56.12	32.43
20	Arg	8.00	126.60	51.56	30.12
21	Ile	9.18	127.22	57.96	37.44
22	Thr	8.44	118.71	59.22	68.91
23	Cys	9.26	123.42	51.10	44.01
24	Gln	8.86	123.17	51.42	30.56
25	Gly	8.49	110.71	43.58	
26	Asp	9.05	123.34	55.35	37.97
27	Ser	9.04	115.75	58.70	59.80
28	Leu	7.34	119.39	53.02	37.33
29	Arg	7.43	115.08	55.47	27.52
30	Ser	7.30	110.80	56.31	62.06
31	Tyr	7.96	121.73	54.35	37.40
32	Ser	8.22	115.86	55.53	61.74
33	Ala	8.34	126.77	49.52	17.32
34	Ser	8.46	114.67	54.76	63.80
35	Trp	9.11	118.94	53.21	30.83
36	Tyr	9.41	120.00	53.77	39.81
37	Gln	9.67	123.19	51.25	31.79
38	Gln	9.63	128.85	52.22	29.45
39	Lys	8.90	131.05	51.85	29.67
41	Gly	8.74	112.23	42.94	
42	Gln	7.69	117.78	50.80	29.16
43	Ala	8.32	124.01	47.91	14.61
45	Val	9.14	123.29	58.31	32.80
46	Leu	8.51	130.22	52.36	39.86

47	Val	8.57	121.26	59.44	30.49
48	Ile	7.06	115.78	53.93	
49	Phe	9.35	123.16	51.63	39.82
50	Arg	8.64	121.72	54.79	26.22
51	Lys	8.87	114.14	-	
52	Ser	8.27	109.88	55.09	63.36
53	Asn	8.32	122.45	50.75	36.29
54	Arg	8.72	124.35	50.62	30.43
56	Ser	8.46	116.11	57.32	60.46
57	Gly	8.68	112.80	42.25	
58	Ile	7.33	123.30	52.56	33.98
60	Asp	8.42	119.37	52.63	36.94
61	Arg	6.92	114.20	54.30	26.43
62	Phe	7.62	120.55	55.21	37.54
63	Ser	8.89	114.87	54.84	62.68
64	Gly	8.65	107.70	41.63	
65	Ser	8.50	112.83	54.47	63.20
69	Asn	8.50	124.60	49.53	35.12
70	Thr	8.06	110.07	58.20	70.24
71	Ala	9.32	127.57	48.63	19.84
72	Ser	8.98	114.46	54.82	63.61
73	Leu	8.59	128.99	49.92	39.13
74	Thr	8.94	123.56	58.70	67.01
75	Ile	8.57	126.74	57.45	36.48
76	Thr	8.45	122.59	58.34	66.40
77	Gly	7.23	113.31	43.74	
78	Ala	8.15	120.92	51.05	16.22
79	Gln	9.02	121.23	50.29	29.34
80	Ala	8.81	125.23	53.41	14.92
81	Glu	8.79	113.45	54.81	25.91
82	Asp	8.03	119.36	51.77	37.66
83	Glu	7.39	122.37	56.40	27.05
84	Ala	8.00	127.61	48.69	18.17
85	Asp	7.85	117.79	51.05	40.40
86	Tyr	9.01	118.77	54.34	39.27
87	Tyr	9.69	121.80	54.55	39.50
88	Cys	7.81	119.36	50.29	41.17
89	Asn	8.38	121.49	49.02	40.62
90	Ser	8.22	117.06	54.27	62.70
91	Arg	7.92	120.43	52.72	29.13
92	Asp	8.04	120.87	50.62	39.32
95	Ala	8.03	123.15	49.89	16.16
96	Asn	8.13	115.87	51.01	35.63
100	Phe	8.68	122.57	54.29	39.81
101	Gly	8.65	108.67	42.03	
102	Gly	8.10	104.35	39.76	
103	Gly	7.12	106.48	42.06	

104	Thr	8.28	119.00	58.46	70.18
105	Lys	8.47	129.10	53.51	29.25
106	Leu	9.07	133.52	51.31	42.00
107	Thr	8.58	123.79	59.22	67.97
108	Val	8.96	126.95	57.98	29.71
109	Leu	8.97	129.25	51.92	39.81
110	Gly	8.14	118.38	43.25	

Solution state NMR assignment of S3706_GL

RESIDUE NUMBER	AMINO ACID	H	N	CA	CB
2	Ser	8.09	116.24	56.27	61.44
3	Ser	8.46	117.90	55.77	61.14
4	Glu	8.26	121.47	54.61	28.13
5	Leu	8.12	121.77	50.92	41.64
6	Thr	9.05	116.98	59.09	68.12
7	Gln	8.72	122.06	51.26	31.38
8	Asp	8.83	124.15	50.49	37.76
10	Ala	7.61	118.97	49.72	18.57
11	Val	8.55	120.38	58.23	33.13
12	Ser	8.28	119.36	53.65	63.70
13	Val	8.55	120.41	56.11	33.29
14	Ala	8.25	128.37	48.65	16.48
15	Leu	7.92	121.32	54.25	39.93
16	Gly	9.71	114.04	42.42	
17	Gln	8.13	120.79	51.84	25.90
18	Thr	8.30	117.23	59.68	67.63
19	Val	8.68	126.92	56.25	32.29
20	Arg	7.98	126.48	51.46	30.11
21	Ile	9.17	127.53	58.02	37.26
22	Thr	8.50	119.24	58.68	68.92
23	Cys	9.45	124.75	53.63	27.96
24	Gln	8.65	126.21	51.07	30.41
25	Gly	8.54	110.88	43.83	
26	Asp	9.05	123.75	55.52	37.94
27	Ser	9.06	115.62	58.44	59.74
28	Leu	7.34	119.94	52.71	36.50
29	Arg	7.35	114.60	55.36	27.38
30	Ser	7.29	111.07	56.30	62.24
31	Tyr	7.70	120.90	53.99	37.83
32	Tyr	8.06	117.94	54.05	36.24
33	Ala	9.14	16.86	49.16	16.62
34	Ser	8.66	116.32	62.96	54.79
35	Trp	8.96	120.94	53.06	30.12
36	Tyr	9.51	120.41	53.94	39.87
37	Gln	9.69	123.37	51.23	31.74

38	Gln	9.61	129.01	52.29	29.36
39	Lys	8.91	131.09	51.82	29.58
41	Gly	8.76	112.15	43.07	
42	Gln	7.66	117.64	50.86	29.01
43	Ala	8.33	124.21	47.80	14.45
45	Val	9.04	122.90	58.31	32.82
46	Leu	8.48	130.25	52.44	39.60
47	Val	8.74	120.84	59.58	30.40
48	Ile	6.86	114.84	54.04	39.42
49	Tyr	8.86	121.68	52.80	38.92
50	Gly	8.37	107.35	43.77	
51	Lys	7.95	123.99	53.47	27.72
52	Asn	7.86	116.00	51.05	35.10
53	Asn	8.14	116.90	50.19	36.12
54	Arg	8.30	122.61	50.02	30.52
56	Ser	8.44	116.12	57.60	60.33
57	Gly	8.68	113.16	42.31	
58	Ile	7.37	123.37	52.56	33.62
60	Asp	8.43	118.96	52.78	36.97
61	Arg	6.82	113.97	54.29	26.53
62	Phe	7.49	120.16	55.23	37.76
63	Ser	8.78	114.40	54.97	63.32
64	Gly	8.65	108.69	41.53	
65	Ser	8.48	113.11	54.63	63.01
66	Ser	8.47	115.25	62.73	56.05
67	Ser	8.15	115.50	55.34	61.11
69	Asn	8.49	125.11	49.25	35.01
70	Thr	8.13	110.00	58.12	70.07
71	Ala	9.37	128.10	48.77	20.53
72	Ser	8.97	114.77	54.72	63.49
73	Leu	8.68	129.35	49.92	39.04
74	Thr	9.01	123.69	58.84	66.98
75	Ile	8.55	126.80	57.36	36.33
76	Thr	8.46	122.33	58.20	66.36
77	Gly	7.23	113.13	43.85	
78	Ala	8.16	120.84	52.04	16.28
79	Gln	8.99	121.10	50.34	29.47
80	Ala	8.80	125.26	53.49	14.95
81	Glu	8.78	113.45	54.73	25.87
82	Asp	8.02	119.34	51.88	37.67
83	Glu	7.39	122.36	56.31	27.04
84	Ala	7.99	127.57	48.77	18.25
85	Asp	7.83	117.80	50.88	40.44
86	Tyr	9.02	119.04	54.29	39.10
87	Tyr	9.83	122.68	54.79	39.61
88	Cys	7.78	118.56	50.30	41.95
89	Asn	8.47	122.15	48.74	40.86

90	Ser	8.20	117.35	54.24	63.17
91	Arg	7.58	118.97	52.26	29.10
92	Asp	8.13	119.52	50.21	38.70
93	Ser	8.55	114.57	50.22	60.16
95	Gly	7.90	110.58	42.50	
97	His	8.31	119.23	51.29	36.99
98	Gln	8.31	117.97	52.33	27.88
99	Val	8.26	119.88	52.85	27.67
100	Phe	8.55	122.35	54.14	39.28
101	Gly	8.55	109.42	41.81	
102	Gly	8.22	103.81	43.32	
103	Gly	7.00	106.57	42.03	
104	Thr	8.15	118.43	58.75	70.29
105	Lys	8.49	129.18	53.70	29.43
106	Leu	9.05	133.30	51.07	41.94
107	Thr	8.55	123.65	59.14	67.92
108	Val	8.93	126.93	58.17	29.67
109	Leu	8.97	129.18	51.70	39.77
110	Gly	8.15	118.20	43.21	

Appendix V

Samples for solid state NMR

	Fibrils Sample	Rotor Size	Spacer
	V_L S3706 samples		
1	S3706_patdel + *ex-vivo* patient seeds_1	3.2 mm	1 bottom
2	S3706_patdel + *ex-vivo* patient seeds_2	3.2 mm	1 bottom
3	S3706_patdel + no seeds	3.2 mm	1 bottom
4	S3706_patdel + own *in-vitro* seeds	3.2 mm	-
5	S3706_GL + *ex-vivo* patient seeds	3.2 mm	1 bottom
6	S3706_R50G + *ex-vivo* patient seeds	3.2 mm	1 bottom
7	S3706_R50G + own *in-vitro* seeds	3.2 mm	1 bottom
8	S3706_patdel+ *ex-vivo* patient seeds (Deuterated)	1.3 mm	-
	TIA-1 samples		
9	QRD44 + no seeds	3.2 mm	-
10	QRD44 + seeds	3.2 mm	1 bottom
11	QRD44 + seeds	3.2 mm	1 bottom
12	RRM123_QRD44_WT	3.2 mm	1 bottom
13	RRM123_QRD44_mut	3.2 mm	1 bottom
14	QRD44 + seeds (Deuterated)	1.3 mm	-
15	QRD44 + seeds (Deuterated) preparation 1	1.3 mm	-
16	QRD44 + seeds (Deuterated) preparation 2	1.3 mm	-
17	RRM123_QRD44_WT + RNA	1.9 mm	-

8 Publications

1) Weber, B., Brandl, M.J., Pulido Cendales, M.D., Berner, C., Feind, G.M**., Pradhan, T.**, Zacharias, M., Reif., B., Buchner, J. (2018). A single residue switch reveals principles of antibody domain integrity. Journal of Biological Chemistry (JBC).

In preparation

2) **Pradhan, T.,** Annamalai, K., Fändrich, M., Buchner, J., Reif., B. (2019). Influence of seeding on the conformational properties of an immunoglobulin light chain protein probed by MAS solid-state NMR, **in preparation**.

3) **Pradhan, T.,** Weber, B., Fändrich, M., Buchner, J., Reif., B. (2019). Conformational analysis of oligomeric aggregation intermediate states of an immunoglobulin light chain protein, **in preparation**.

4) **Pradhan, T.,** Sarkar R., Annamalai, K., Fändrich, M., Buchner, J., Reif., B. (2019). MAS solid-state NMR structural analysis of amyloid fibrils formed by an immunoglobulin light chain protein, **in preparation**.

5) **Pradhan, T.**, Fändrich, M., Buchner, J., Fändrich, M., Reif., B. (2019). Interaction of EGCG with amyloidogenic light chain, **in preparation**.

6) **Pradhan, T.,** Jagtap, P., Sattler, M., Reif., B. (2019). Solid-state NMR studies of fibrillar forms of the T-cell Intracellular Antigen-1 (TIA-1): Implications for stress granule formation**, in preparation**.