Assembly and Structure Characterization of Intermediate Filament Low Complexity Domain

By

HILLARY SUTTON

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dylan Murray, Chair

Andrew Fisher

Kassandra Ori-Mckenney

Committee in Charge

2022

ACKNOWLEDGEMENTS

I want to sincerely thank my advisor, Dylan Murray, for his guidance, patience, and mentorship that enabled me to grow as a scientist and to navigate my alternative path. I would also like to thank the members for the Murray lab for their encouragement and lending me their knowledge time after time. I am very grateful for Andy Fisher and Kassie Ori-Mckenney's service on my committee. Last, I want to acknowledge all my wonderful family, friends and especially my husband, Dustin, for loving me and supporting me for all that I am and choose to do. I am very grateful to be cushioned in so much love.

ABSTRACT

Intermediate filament (IF) proteins self-assemble into coiled-coil structures via a central rod domain. Disordered head or tail domains of low complexity flank this region and are required for assembly into unit length filament bundles of eight or more coiled-coil tetramers that anneal end to end to make fully formed filaments. These disordered regions can be responsible for many cellular regulation processes and are implicated in numerous diseases that are associated with mutations in these regions. However, it is difficult to characterize these disordered regions in the context of whole protein and assemblies using high resolution techniques such as crystallography and solution state nuclear magnetic resonance (NMR). To pursue a high-resolution structure of an IF protein, solid state NMR is used to address the limitations of many low-resolution techniques and other high-resolution techniques that are incompatible with large or disordered proteins. The head and tail domains in full-length vimentin and the tail domain fibrils of atypical isoform of tropomyosin Tm1 I/C are investigated using CC DARR, NCACX and NCOCX NMR experiments with isotopic labeling techniques such as paramagnetic relaxation enhancement, segmental labeling, 50:50 mixed labeling and glycerol labeling. The distance restraints from these experiments will contribute to a calculation of a high-resolution structure. Understanding details of the structure of the disordered domains in the tail domain polymers and fully assembled filaments can give insight to the assembly mechanism and function of other intermediate filament proteins.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

IF	Intermediate filament
LC	Low complexity
NMR	Nuclear magnetic resonance
MF	Microfilament
MT	Microtubule
ULF	Unit length filament
NFL	Neurofilament triplet proteins
EPR	Electron paramagnetic resonance
FUS	Fused in sarcoma
TEM	Transmission electron microscopy
MAS	Magic angle spinning
DARR	Dipolar assisted rotational resonance
PRE	Paramagnetic relaxation enhancement
IPTG	Isopropyl-β-D-thiogalactoside
BME	β-mercaptoethanol

1. INTRODUCTION

Intermediate filaments (IF) are intracellular structural proteins that contain disordered regions of low complexity that are difficult to characterize in high resolution and are essential for the stability and function of these proteins. Low complexity (LC) refers to the high prevalence of a small number of different amino acid types within the sequence.[1] Often, these regions are disordered, lacking stable and folded three-dimensional structure. They predominantly occupy random coil conformations, but they can adopt many structural conformations depending on the environment. Despite not having activities such as catalysis or synthesis typical of globular, well-folded proteins, these regions are not functionless and can be responsible for many cellular regulation processes. Additionally, many diseases are associated with mutations in disordered LC protein regions that result in misregulation of biological processes.[2] However, characterizing these disordered regions, particularly in context of whole protein structure and assemblies, can be difficult using high resolution techniques such as crystallography and solution state nuclear magnetic resonance (NMR). Currently, there is no high-resolution structure of a full-length IF protein. Solid state NMR can address the shortcomings of many low-resolution techniques and other high-resolution techniques that are incompatible with large or disordered proteins. This is valuable because characterizing the behavior of these disordered regions can provide insight into the assembly mechanisms underlying biological processes that are disrupted due to mutated forms in of the proteins in diseased states.[3] Ultimately, understanding these mechanisms in a specific IF can extend insight to other IFs, which are one of the 100 largest protein families in humans, as well as the properties of other disordered proteins.[4] Solid state NMR has the potential

to characterize the disordered regions of IF proteins in the context of the high-resolution structure of fully assembled IF.

1.1 Intermediate Filaments

IF proteins are one three major classes of intracellular filaments and are an important component of the cellular cytoskeleton. They are found in the cytoplasm and the nucleus, and they provide cellular structure and motility by assembling into diverse networks of polymers of various flexibility depending on differential expression of distinct IF proteins.[5] Though they have been studied extensively in context of cell structure, more recently they have been found to participate in other significant cellular functions such as gene regulation, cell signal transduction pathways, and organization and assembly of the other filament classes, microtubule and microfilaments.[6] These functions are mostly transient and due to rapid changes from cell signaling and regulation, and the disordered, non-globular regions of these proteins are responsible for much of this activity and harbor many disease mutations.[7]

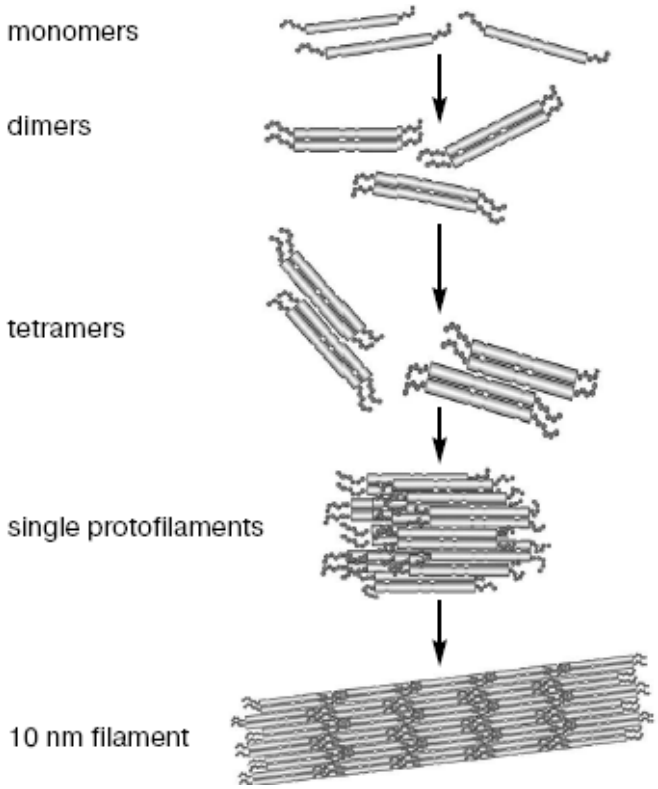
1.1.1 Characteristics and Assembly

IFs are distinguished from the actin microfilament (MF) and microtubule (MT) cytoskeletal proteins by their diameters. Actin MF have diameters around 7 nm, MT around 25 nm, and IF are in the middle at around 10 nm, which is the origin of the name intermediate filament.[8] IF monomers consist of α-helical rods and linkers domains flanked by disordered LC amino-terminal head and carboxyl-terminal tail regions. Additionally, they are self-assembling and do not require any external energy molecules or cofactors for filament formation in contrast to actin and microtubules.[9]

Assembly begins with the formation of parallel dimers that associate half-staggered and antiparallel to form a tetramer subunit. This structure is the cytoplasmic subunit of IF and has

remarkable stability in up to 5 M urea due to the hydrophobic interactions of the common interface between the coiled-coil dimers.[10,8] This structure differs from globular MF and MT assembly units in that IFs lack polarized ends in their tetrameric form and the tetrameric blocks can anneal and detach from either end with other tetramers. With increasing ionic strength, the tetramers associate laterally into an eight-tetramer-bundle, which forms the unit-length filament (ULF). ULFs anneal end-to-end to extend into a full-length filament, followed by radial compaction into the mature filament.[11] A schematic of this process is shown in figure 1.[12]

Figure 1: Intermediate filament assembly process. Monomers associate to form parallel dimers that associate in a staggered antiparallel arrangement to form tetramer subunits. Eight tetramers bundle laterally to form unit length filaments (or protofilaments), then anneal end to end to form filaments, and finally radially compact to 10 nm to form mature filaments. This figure was reproduced from Minin and Moldaver, 2008.

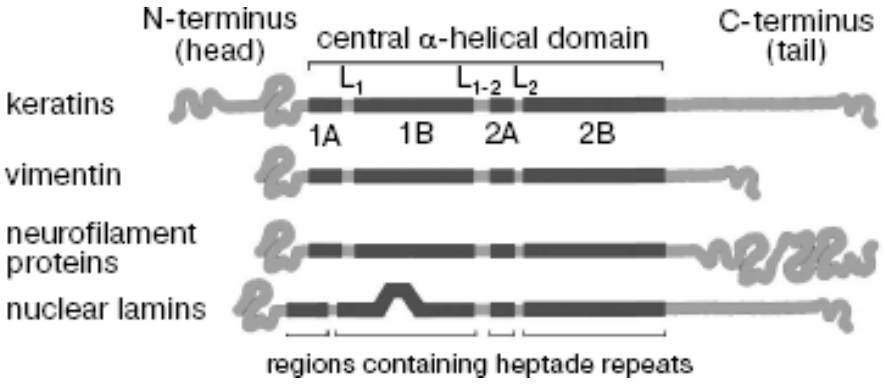


Because the tetramer subunits are not polarized, the exchange of subunits allows the cytoskeletal network to be flexible and dynamic. They can reach their subunit tetrameric form in seconds, a

unit length filament in one minute, and a mesh network can form within ten minutes.[8] Despite their flexibility, they are very stable in the tetrameric subunit and, depending on the environment, the ULF and the mature filament network can be as well.[9] Their resilience under mechanical stress is much higher than for MTs and MFs.[13] IFs are also unique in that they are primarily but not exclusively regulated by phosphorylation, whereas MF and MT are regulated by associated proteins and other post-translational modifications such as acetylation, polyglutamylation, tyrosination, and polyglycylation.[14,15]

Across many higher order species and throughout many tissues from muscle to neuron, IFs have a highly conserved sequence of coiled rods and disordered linkers, whereas the disordered LC head and tail domains vary greatly in sequence and length (figure 2).[12] From an evolutionary perspective, the highly conserved coiled-coil domains are consistent with the ubiquitous structural role of IF proteins, while the variability of the head and tail sequences and lengths among the members of this diverse class of proteins highlight their roles in broad functional activity.

Figure 2: Examples of variability of intermediate filament protein molecules. The coiled-coil domain is highly conserved among different IF classes and proteins; however, the head and tail domains are exceptionally diverse in sequence and length. This figure was reproduced from Minin and Moldaver, 2008.

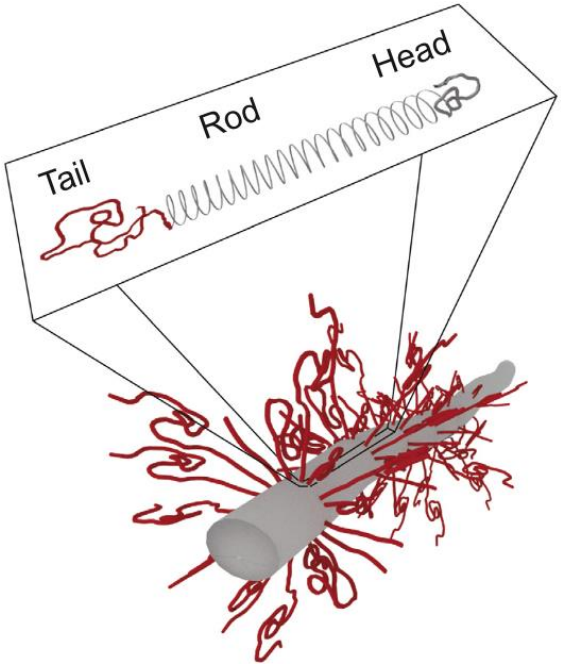


IF are associated with many tissue-specific diseases due to mutations, misregulations and as markers, and this specificity reflects the differential expression of IF. These diseases are predominantly based on modulating normal regulation of assembly which is directly related to the head and tail domains.[16,17] A mutation in the head domain of the low molecular weight subunit of the IF neurofilament triplet proteins (NFL) is linked to neurodegenerative Charcot-Marie-Tooth disease by disrupting assembly in neuronal cells.[18] There is a high prevalence of disease mutations in the disordered LC head and tail regions, and the lack of structure of these domains suggests new and exciting modes of disruption.[3] The Human Intermediate Filament Database has compiled 119 distinct diseases with 547 sequence and allelic variants found in the head and 533 in the tail.[19] The disordered nature of these domains makes it challenging to characterize their role in disease, so being able to probe these disruptions in their fully assembled form can give greater insight to the disease mechanism and aid treatment development.

Generally, the head is necessary for assembly into full length filaments, but the tail is not.[20] Small X-ray scattering was used to address the lack of details about intermediates in filament assembly and showed that the fold of the head domain back onto the coiled rod stabilizes the formation of tetramers. Removal of basic arginine residues indicated that different segments of the head domain are necessary for tetramer and filament formation. Although the tail domain is not required for assembly, it is suspected to provide stability to the fully assembled filament because without it, the filament diameter varies, and the structures become much more labile.[21] Post translational modifications also regulate assembly and are most prevalent in the head and tail domains.[17] The vimentin tail has a conserved β-turn near multiple phosphorylation sites suggesting that this region is involved in regulation of filament and network stability.[22] In the fully assembled filament, the tail domains protrude from the core and resemble a bottle brush

(figure 3).[3] These bristles form flexible extensions that connect filaments and other structures in the cytoplasm through the mediation of phosphorylation, pH or divalent metal cations.[23]

Figure 3: Schematic of the bottle brush shape produced by tail domain protrusions in a fully assembled intermediate filament. Tail domains are involved in cell signaling, stabilizing organelle anchors, and assembling networks through mediation of phosphorylation, pH and divalent metal cations. This figure was reproduced from Korneich et al, 2015.



The tail domains are also important in the formation of unique cytoskeletal networks in various cell types.[24] In addition to providing scaffolding for the cell, these structures anchor organelles like mitochondria and Golgi, and also participate in cell signaling and protein targeting.[25]

1.1.2. Vimentin

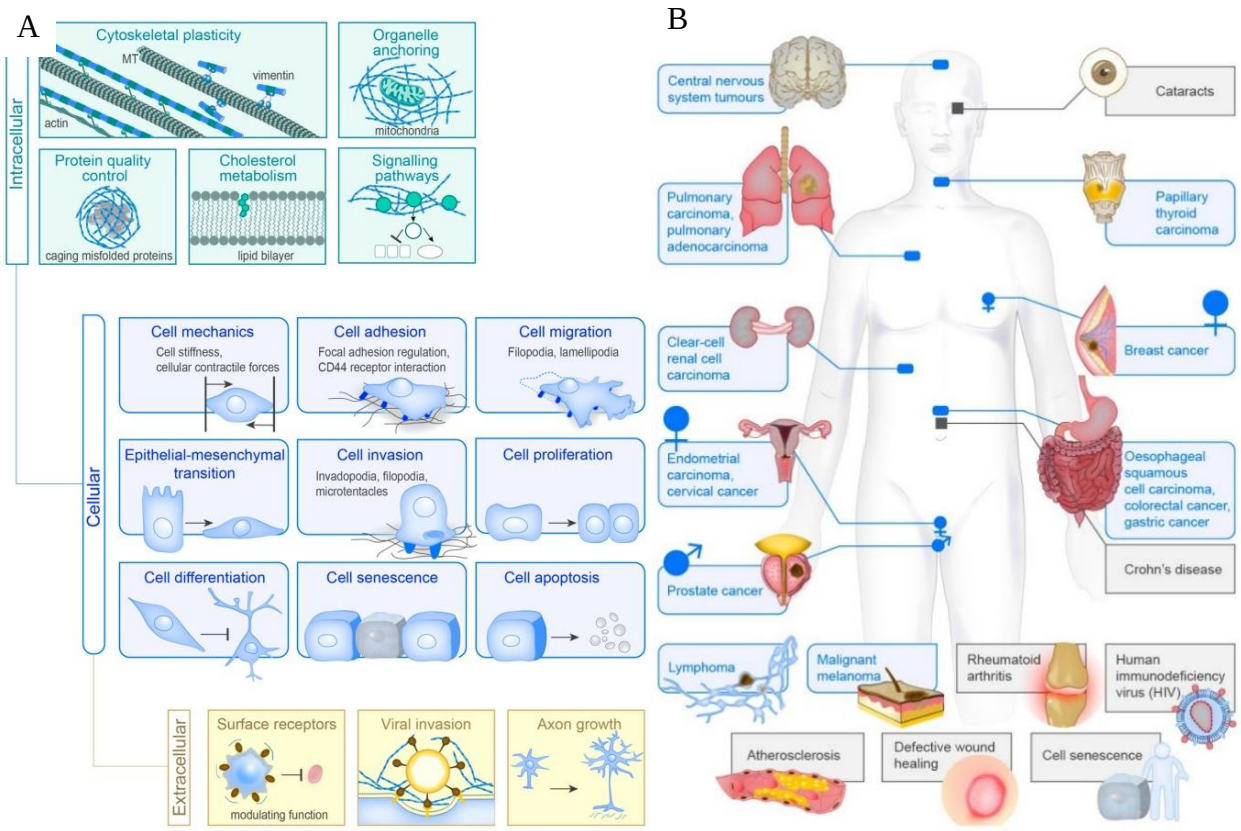
Vimentin is an excellent model for characterizing IFs because it has been extensively characterized within the constraints of traditional structure-probing techniques, it is the most widely distributed filament across tissue types and species with numerous structural and regulatory roles, and, consequently, it is related to many diseases. Although the entire length of

the protein has not been characterized in the fully assembled filament or during the assembly process, a significant portion of the coiled-coil core has. The central rod and linkers have been established using low resolution techniques or using high resolution techniques in segments. Crystallography has been used to establish the central rod domain parameters.[26] Using site-directed spin labeling and electron paramagnetic resonance (EPR), the orientation of the dimers and tetramers was revealed, the span of the central coiled rods was confirmed, and the disordered linkers were characterized in assembly.[27] The whole central domain structure was completed by combining experimental data with molecular dynamic modeling.[28] Using EPR, the head has been shown to be dynamic and folded back on the coil to associate with the neighboring rod domain during assembly.[29] This well-characterized structure of rods and linkers can be used as a reference and built upon with higher resolution data to elucidate the structure and function of the head and tail domains at an atomic level in the context of the fully assembled filament.

Vimentin is expressed in leukocytes, blood vessel endothelial cells, some epithelial cells, and mesenchymal cells. Although initially considered to have roles limited to providing structure in cells, evidence of broader functions continues to emerge. Because vimentin expression is prevalent in diverse cell types, its differentiated network formations can provide cell-type-specific-skeletal crosstalk that can supply cells with mechanisms related to their shape and physiological activities.[30] Vimentin expression has been shown to disrupt the stability of triglycerides in distinct cell types and could be involved in the metabolism of specific lipid components.[31] Vimentin has also been long overlooked for its regulatory roles because of unobvious phenotypes from mice-deficient studies, but it is now known to organize crucial proteins involved in attachment, migration and cell signaling.[32] With the numerous functions that vimentin has in cell structure, cell-cell interactions and regulation of signaling pathways shown

in figures 4.A, there consequently are also several ailments that can be attributed to vimentin mutations and misregulation (figure 4.B).[33]

Figure 4: Mutations and misregulation disrupt normal functions of vimentin and are associated with many diseases. (A) Examples of vimentin intracellular (top), cellular (middle) and extracellular (bottom) functions. (B) examples of human diseases in various tissues linked to vimentin. These figures were reproduced from Danielsson et. al, 2018.



Vimentin is related to many diseases including cataracts, Crohn's disease, HIV, and cancer.[33] Most post-translational modifications occur in the head and tail domain and are responsible for the tissue-specific roles.[12] Characterizing the structure of the disordered head domain has valuable potential in understanding misfunction because the head domain is necessary for assembly, most phosphorylation sites reside on the head, and many disease-causing mutations reside in the beginning of the central rod that the head associates with in assembly.[34,14,35] Characterizing the fully assembled structure can elucidate mechanisms of mutations and

misregulation which progresses towards developing treatment or prevention for their linked diseases.

1.1.3. Tm1 I/C

Another protein family that has been increasingly exposed for having a diverse set of functions is tropomyosin. Tropomyosins are proteins found in all animals that form head-to-tail polymers of coiled dimers down the length of actin filaments and are most known for playing a critical role in regulating the access of binding to actin filaments in muscle contraction.[36] Recently, a non-canonical isoform of tropomyosin produced from an alternative splicing process was found in the Tm1 locus of *Drosophila* that has many similarities to IFs.[37] The Tm1 I/C isoform has LC N- and C-terminal domains flanking a coiled coil, spontaneously assembles into filaments between the diameters of actin MF and MT, and it does not associate with actin as canonical tropomyosin does. A notable difference between this isoform and traditional IF proteins is that the carboxyl-end tail domain is similar in function to the head domain in typical IFs, which have both have been shown to form cross-β structures similar to β amyloid, and the amino-end head domain behaves similarly to IF tail domains.[38] The central rod domain consists of a single coiled coil in Tm1 I/C, while there are three coiled coils broken up by flexible linkers in IF rod domains.

One of the aspects of Tm1 I/C that has been supposed to mimic IF function is interactions with RNA and RNA binding proteins to transport and stabilize biomolecules. Localization of the germ plasm to the posterior pole of *Drosophila* oocytes is necessary for germ cell formation, and this polarization is dependent on oskar mRNA localization.[39] Because germ granule deposition in the proper location of egg cells is dependent on Tm1 I/C, it is speculated that the tail domain

might interact with or stabilize oskar mRNA in localization.[40] Repeating LC domains of assembled IFs have been shown to associate with RNA binding proteins along the length of the filament in intervals equal to the distance between these domains. In the case of IF vimentin and RNA-binding protein fused in sarcoma (FUS), this interaction is dependent on the LC head domains.[41] *Drosophila* inexplicably do not express cytoplasmic IF proteins, so the IF-like isoform of tropomyosin could be an evolutionary make up for the lack of IF's distinctive cytoskeletal role in the species. Investigating the function of this adaptive protein isoform could give insight to other analogous functions of IF proteins.

1.2. High-Resolution Structure Characterization

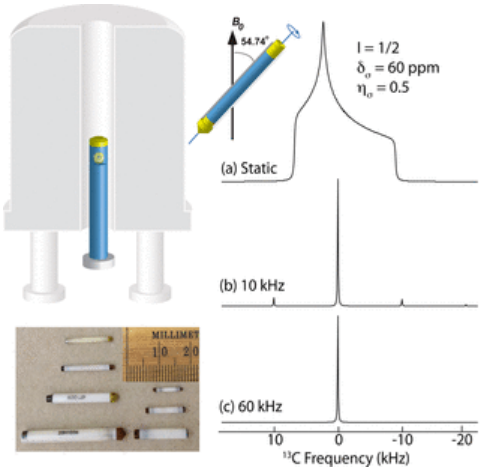
Although IFs are among the most abundant proteins in many organisms, there is much that is still speculated about their assembly mechanism because of the lack of insightful data at the atomic level. The basis of what is understood about these proteins has been gathered using low resolution techniques like EPR, atomic force microscopy, transmission electron microscopy (TEM) and immunofluorescence or in a 'divide and conquer' approach of high-resolution techniques from isolated segments of protein using crystallography and solution state nuclear magnetic resonance (NMR).[42] Although the disordered nature of the head and tail domains enables IFs to perform their transient functions, this property makes it difficult to analyze the fully assembled and functional protein structure in high resolution, which is also the circumstance for many other proteins with disordered regions.[7] Proteins with disordered regions are highly flexible with conformational heterogeneity resulting in an ensemble of structures under physiological conditions rather than a single one. Because disordered regions lack a persistent, dominantly stable three-dimensional structure, crystallography is not a viable method to determine the relevant conformations and interactions of these regions.[43] Even in isolated

segments of the more rigid coil domains, producing crystals of IF fragments that can be resolved crystallographically has shown to be exceedingly difficult and has only been successful about one in ten times.[42] Solution NMR has overcome the challenges of studying the population of structures for proteins with intrinsically disordered regions.[44] However, there is a limit for proteins size around 100 kDa, and larger assemblies like IFs are too bulky for sufficient isotropic tumbling to resolve signals.[45] To calculate structures of IFs at the molecular level, a more versatile NMR technique is required.

1.2.1. Magic Angle Spinning Solid State Nuclear Magnetic Resonance

Magic angle spinning (MAS) solid state NMR is a high-resolution method that can overcome the limitations of solution NMR and probe IF proteins in their fully assembled and other physiologically relevant conformations using magic angle spinning. Fast, random tumbling averages out anisotropic interactions of samples in solution NMR which results in sharp, resolved peaks. In solid state NMR, these orientation-dependent interactions are observed in broad, unresolved peaks called powder patterns because of insufficient tumbling.[46]

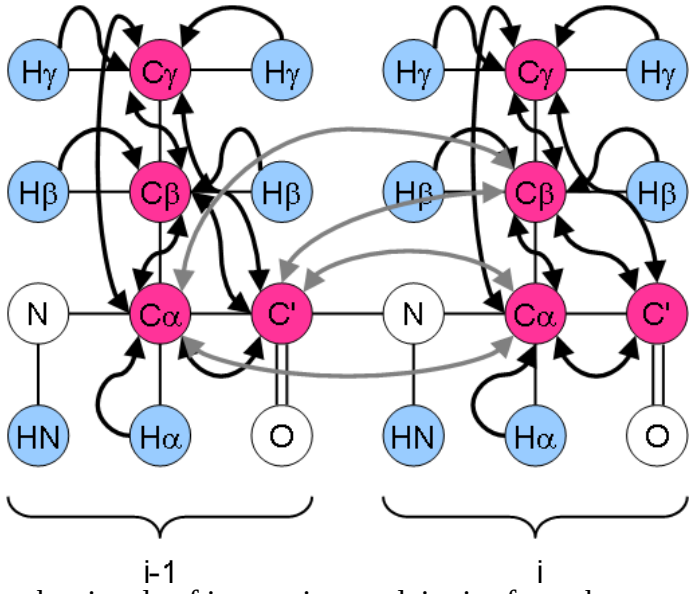
Figure 5: Anisotropic powder pattern is resolved by magic angle spinning. (Left) Schematic of sample placement in MAS NMR; the rotor is spun at an angle of 54.74° with respect to the applied magnetic field and various sizes of rotors. (A) The anisotropy of static sample results in powder patterns and unresolved signal. (B, C) When the sample is spun rapidly at the magic angle and exceeds the anisotropic interaction, the signal averages out into an isotropic peak. This figure was reproduced from Polenova et. al, 2015.



To circumvent this, the sample is spun rapidly on an axis at the magic angle of 54.74° relative to the fixed magnetic field to average out the dipolar coupling interactions and resolve the signals without isotropic tumbling. MAS solid state NMR can accommodate the protein assemblies of IFs that are too large for adequate isotopic tumbling. Acquiring high resolution data from the disordered regions in the context of the full assembly can complement the existing data using other various techniques to obtain a high-resolution characterization of the full-length filament.

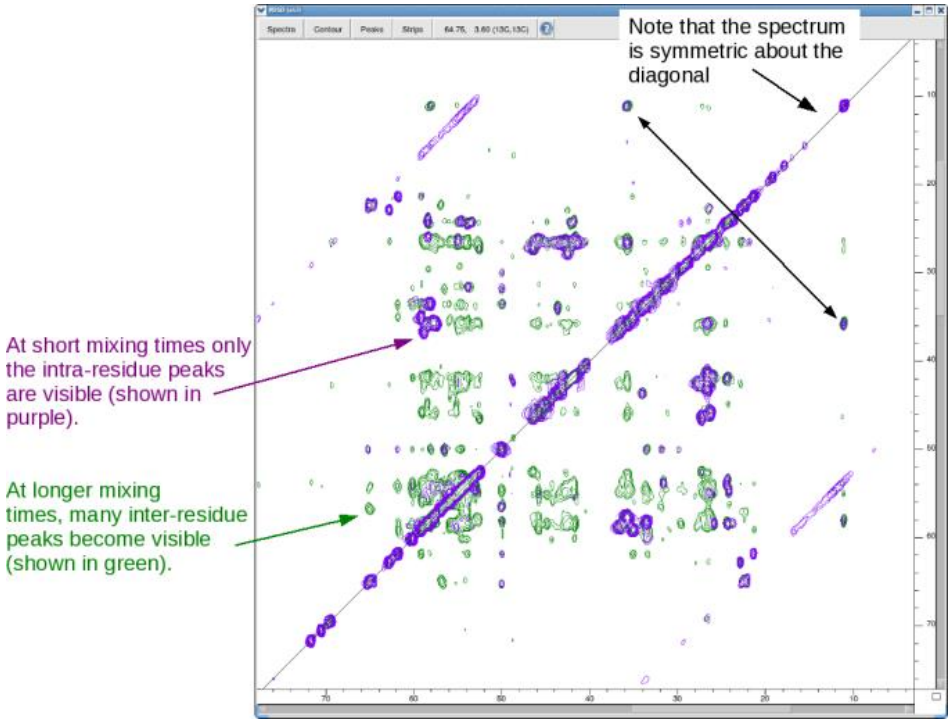
Specific experiments that are useful in assigning amino acid residues along the backbone are dipolar assisted rotational resonance (DARR), NCACX and NCOCX in ^{15}N and ^{13}C uniformly labeled samples. These experiments are especially useful for this task because they involve the transfer of magnetization from one nucleus to other nuclei to create a multidimensional spectrum which can help identify individual amino acids. DARR is a technique that is useful to first determine the quality of spectra possible with the sample.[47] This technique overcomes the relatively weak dipolar coupling between two carbon nuclei by transferring the magnetization from a ^{1}H nucleus to a ^{13}C nucleus by cross polarization which then transfers more efficiently to other ^{13}C nuclei that are close in space (figure 6).[48]

Figure 6: Scheme of DARR magnetization transfer. Magnetization is transferred from a ^{1}H nucleus to a ^{13}C nucleus by cross polarization. This is useful for assigning resonances to amino acids in the protein primary sequence. This figure was reproduced from Higman 2012.



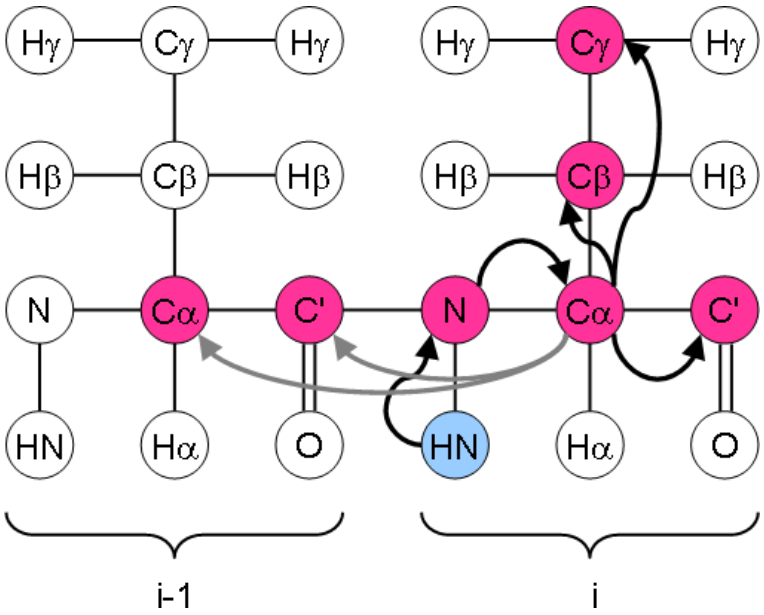
In DARR experiments, the signals of interacting nuclei arise from the associated chemical shifts on each axis that are correlated along the diagonal. An example of this type of spectrum is shown in figure 7.[48] Each signal is identified by the typical chemical shifts of a specific carbon on an amino acid type, in a particular secondary structure. For example, in an α-helix, lysine has average chemical shifts of 58.9 ppm, 32.3 ppm & 178.4 ppm for Cα, Cβ & CO respectively.[49] This experiment can be done at varying mixing times of transferring magnetization between the carbon nuclei to obtain more specific information; for example, shorter times like 50 ms identify nearby intra-residue contacts and longer times like 250–500 ms can identify longer range inter-residue contacts.

Figure 7: ^{13}C-^{13}C DARR Experiment. The correlation spectrum helps to identify interactions between ^{13}C nuclei on an amino acid to help assign residues in the protein primary sequence. This figure was reproduced from Higman 2012.



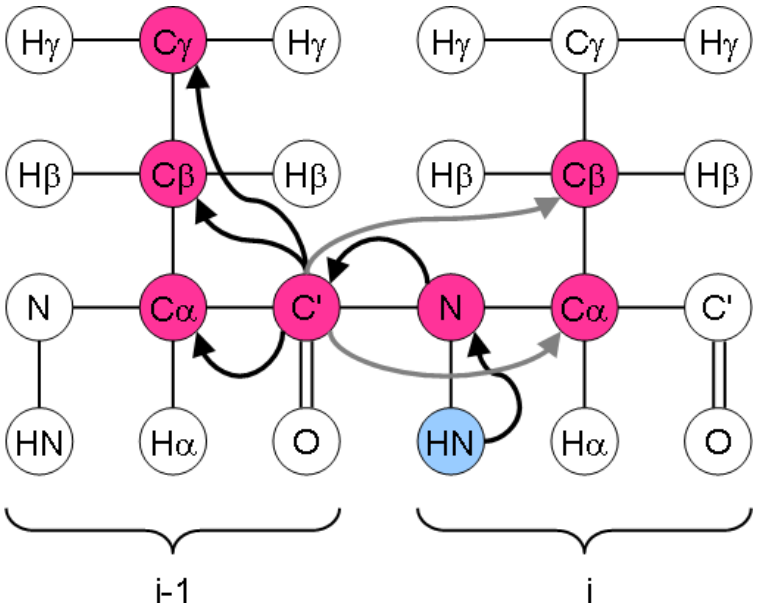
NCACX is useful in identifying the signals arising from spin interactions within a single residue. This is done by transferring the magnetization from a backbone amino ^{1}H to the ^{15}N nucleus via cross polarization, then specifically to the Cα carbon of that residue. Then, a DARR pulse is applied to transfer magnetization to other nearby ^{13}C nuclei, which is detailed in figure 8.[50] This results in a correlation spectrum between nitrogen and other carbons on a single amino acid residue and helps to identify amino acid types by correlating the chemical shifts as described previously for the CC spectrum.

Figure 8: Scheme of NCACX magnetization transfer. Magnetization is transferred from a nitrogen ^{1}H nucleus to the ^{15}N nucleus via cross-polarization, and then specifically to the Cα nucleus on the same residue. The magnetization is transferred from this to other nearby ^{13}C nuclei. This is useful for assigning resonances to amino acids in the protein primary sequence. This figure was reproduced from Higman 2012.



Finally, NCOCX is used similarly to NCACX, except that it specifically relates two neighboring amino acid residues along a sequence by transferring magnetization from the amino hydrogen to the ^{15}N and then explicitly to the carbonyl ^{13}C nucleus on the i-1 residue as shown in figure 9.[51] This links assignments of amino acid types made in the NCACX so that a map of the backbone can be created to assign the observed NMR signals to specific residues in the protein sequence. In addition to two-dimensional spectra, the experiments can be collected in three dimensions to correlate three interacting nuclei and further distinguish the signals due to higher resolution in the three-dimensional spectra.

Figure 9: Scheme of NCOCX magnetization transfer. Magnetization is transferred from a nitrogen ^{1}H nucleus to the ^{15}N nucleus via cross-polarization, and then specifically to the CO nucleus on the preceding residue. The magnetization is transferred from this to other nearby ^{13}C nuclei. This is useful for assigning resonances to amino acids in the protein primary sequence. This figure was reproduced from Higman 2012.

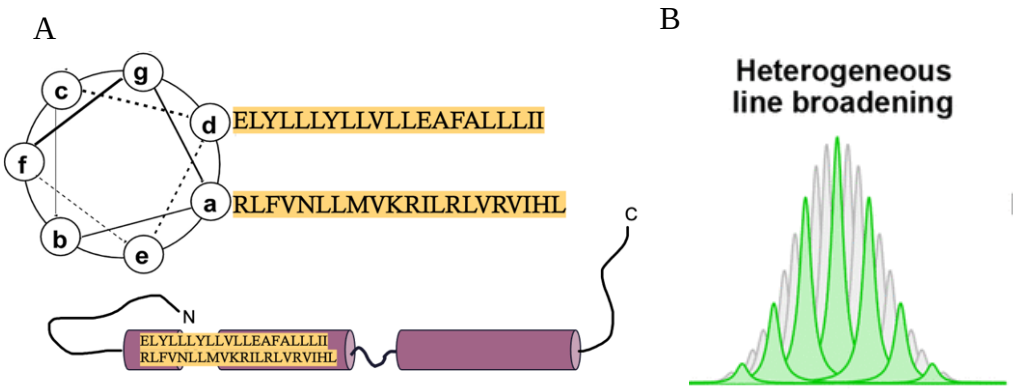


These sets of multi-dimensional NCACX, NCOCX and DARR data are useful to correlate individual residue interactions to their neighboring residues and help identify the protein backbone to build a map of the through-space structural interactions.

Even in the solid state, analyzing large, fully labeled proteins using NMR is difficult. Vimentin, for example, is a 54 kDa monomeric protein, so there is a large number of observed NMR signals that must be assigned to specific amino acid residues. An NMR spectrum of this large IF protein results in broad and difficult to distinguish signals. This is because of the heptad repeat of the coiled coils in the rod domains.[52] The heptad repeat is a sequence of amino acids following the pattern $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$, where a and d are commonly nonpolar residues, and gives rise to an α-helical conformation that forms the basic structural unit of the rod domains. As you travel down the rod of the protein, there is a common surface of nonpolar residues.[53] This results in similar chemical environments for repeated amino acids on various positions of the backbone, as shown in figure 10.A. The observed signals then overlap on the NMR spectrum in what is known as heterogeneous line broadening (figure 10.B), and they become challenging to resolve from one another and subsequently assign to specific residue positions. This problem requires a

more specialized approach to make NMR assignments to acquire a high-resolution structural information for a fully assembled IF protein.

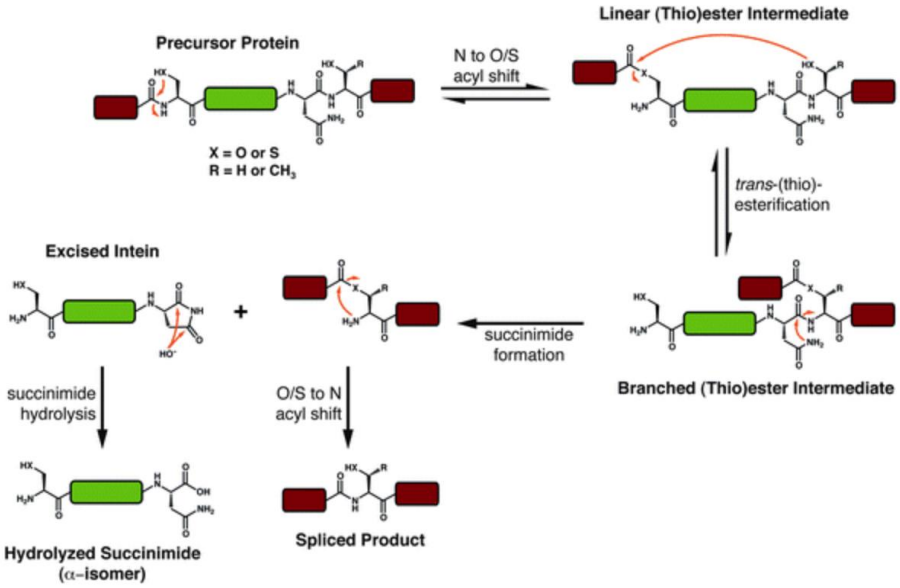
Figure 10: Similar chemical environments of different nuclei result in heterogeneous line broadening that creates a challenge for assigning resonances. (A) amino acid residues in heptad repeat along the side of a coiled rod of IF protein. (B) Scheme of overlapping NMR signals in heterogeneous line broadening. This image was reproduced from De Roo et. al, 2018.[54]



1.2.2. Segmental Labeling

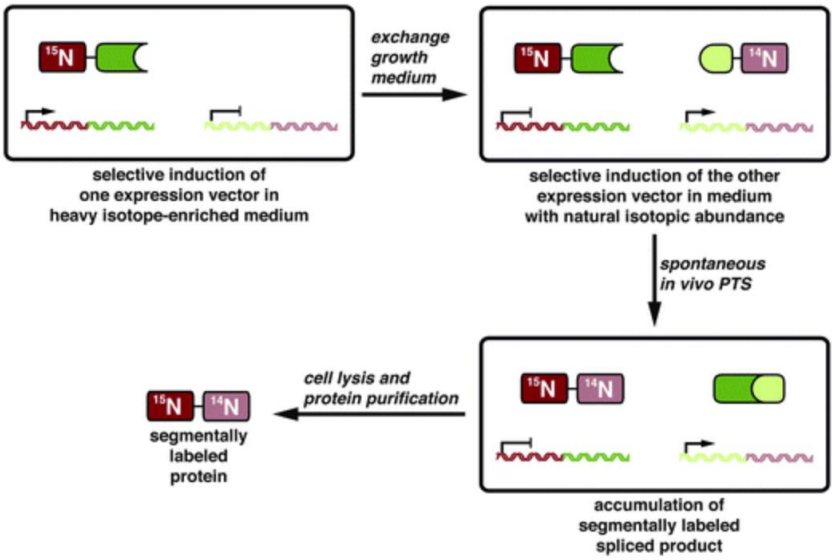
Segmental labeling is a technique that can be exploited to circumvent the challenge of resolving a fully labeled IF protein using solid state NMR. This can target the observed NMR signals to specific regions of the protein, which simplifies the spectrum and helps identify location-specific interactions. There are many signals that are difficult to distinguish on large proteins like IFs. Reducing the overall information in the spectrum by selectively labeling just the domain of interest can significantly improve the clarity of the data and help to gather useful information about the interactions of that domain. This is possible using split-intein chemistry. Intein refers to an *int*ervening prot*ein* that carries out an autoprocessing event where the C- and N-intein fragments associate in a process called protein trans splicing, and then remove themselves from the two exteins (*ex*ternal pro*tein* or protein of interest) by cleaving two peptide bonds, and finally splicing together the desired protein fragments by forming a new peptide bond, as is depicted in a general intein chemistry mechanism in figure 11.[55]

Figure 11: Protein splicing mechanism using intein chemistry. In the case of split inteins, the green domain represents the N- and C-inteins already associated in trans-splicing. The exteins in red are ligated as the intein excises itself by breaking the two flanking peptide bonds and forming a new peptide bond between the exteins in the autoprocessing mechanism. This image was reproduced from Shah and Muir, 2014.



It is helpful to use the split intein technique to investigate a target domain of an IF, like the disordered head or tail domain for example, in the context of the intact filament assembly. Using cloning, the sequence of the target domain on, for example, the N-terminus end of the IF protein can be inserted into a vector just before the N-intein sequence which would result in a single protein sequence of the target fragment and part of the intein. Similarly, the remaining IF sequence on the C-terminus end can be inserted directly after the C-intein sequence in a separate vector to yield a second single protein sequence of the other target fragment and the other part of the intein. The vector with the target domain can be expressed in an isotopically labeled medium to result in an NMR active fragment of the IF, and the other fragment in an unlabeled medium to remain unactive. When combined, the two fragments fuse together in the spontaneous protein splicing process and can then be purified as depicted in figure 12.

Figure 12: Segmentally labeled split intein scheme. The N-terminal fragment of the protein is independently expressed in isotopically labeled media, and the C-terminal fragment is not. This results in a segmentally labeled protein after the split intein autoprocessing. This image was reproduced from Shah and Muir, 2014.

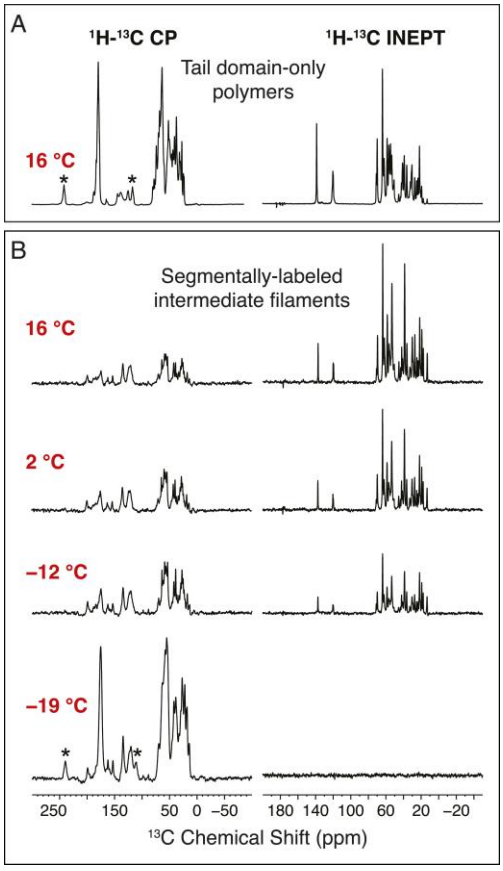


The result is a full-length protein that includes NMR active nuclei only on the desired region of the protein, such as the tail domain. Data obtained using this design can give insight into the interactions of the tail domain in assemblies formed by the full-length protein without the signals from the well-characterized rod domain crowding the spectrum.

This technique was implemented to investigate the structure of the tail domain of Tm1 I/C, and the results suggest that the tail forms the same β-sheet structure when in isolated fragments and in the full-length form.[38] When the disordered tail fragments are purified and incubated, they assemble into amyloid-like polymers called fibrils similar to the LC domain of Ifs including desmin and neurofilament light NFL proteins as well as RNA binding proteins like fused in sarcoma.[56,57] The comparable β-strand secondary structures of the uniformly labeled LC domain only fibrils and the full-length segmentally labeled LC domain for all of these proteins have been demonstrated using solid state NMR. The one-dimensional CP- and INEPT-based ss-

NMR spectra of the Tm1 I/C uniformly labeled tail domain polymer and segmentally labeled IF in figure 13 show the relative mobility of each at various temperatures.

Figure 13: One-dimensional CP- and INEPT-based ss-NMR spectra of the Tm1-I/C tail domain illustrate the immobile and mobile regions of the domain respectively at varying temperatures. A) Tail domain-only polymers at 16°C. B. Segmentally labeled IF at temperatures ranging from 16°C to -19°C. These data were reproduced fromSysoev et. al, 2020.

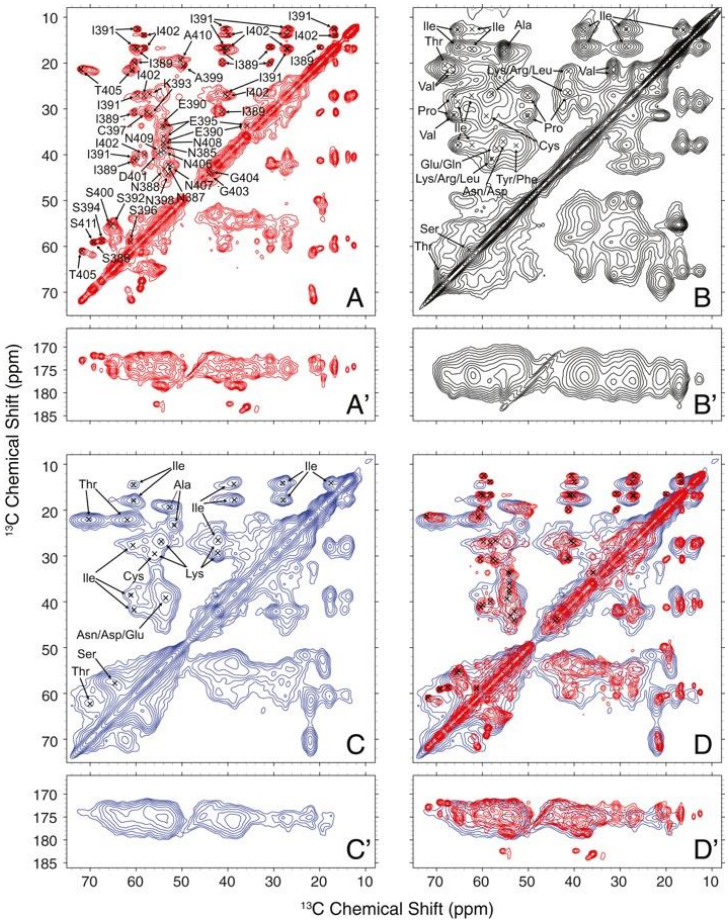


The ^{1}H-^{13}C CP experiment detects signals from the immobilized regions, and the ^{1}H-^{13}C INEPT from the mobilized regions. At 16°C, the tail in the segmentally labeled intermediate filaments is more mobile than in the tail domain-only polymers, indicated by the relatively weaker CP signals. As the temperature decreases to −19°C and the protein movement slows, the CP signal of the IF increases and resembles those of the polymers at 16°C. In the INEPT spectrum, the mobilized region shows slowed movement as signal decreases with decreased temperature.

These data suggest that each form of the tail domain (i.e. segmentally labeled filaments and tail domain-only fibrils) can adopt well-ordered conformations and contains regions of significant mobility and disorder.

The structural similarity is further supported in the overlay of more detailed two-dimensional experiments that highlight the similarities and differences of these structures. The ^{13}C-^{13}C DARR experiments measure the carbon interactions of rigid sites of the protein one to three atoms away at 50 ms mixing time. The spectrum in figure 14.A shows the isolated tail domain-only polymer at 16 °C. The spectrum in figure 14.C shows the tail domain in the segmentally labeled full length filament at −19 °C in figure. These two spectra are overlayed in figure 14.D, and it is evident that they have similar chemical shifts for the rigid portions of the tail domain. Figure 14.B shows the spectrum of the cold-ether precipitated tail domain-only polymer sample to verify that the chemical shifts of the others are a result of true structural conformations rather than precipitated protein in amorphous conformations. The sharp signals from the tail domain only polymers were assigned to the amino acids of the rigid portion of the tail domain using NCACX and NCOCX experiments, and the chemical shifts of these residues correlate with β-strand conformations in both structures.

Figure 14: Two-dimensional ^{13}C-^{13}C DARR experiments for (A) Tm1 I/C tail domain-only polymer. (B) Cold-ether precipitated tail domain-only polymer. (C) Segmentally labeled intermediate filaments. (D) Overlay of A and C shows the similar resonances of the tail domain-only polymer and tail domain in the segmentally labeled IF. These data were reproduced from Sysoev et. al, 2020.



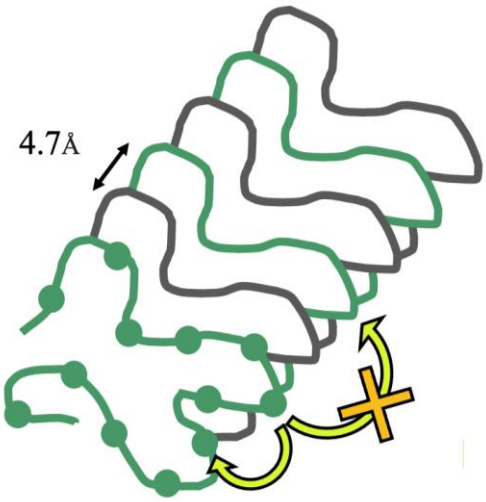
Segmental labeling helped to show the similarity of the Tm1 I/C tail domain NMR data in the tail domain-only polymers and in the full intermediate filament. This allows inferences to be made about the similarity of the tail domain structure in these two different assemblies. Inferring that the tail domain in the two assemblies occupy the same secondary β structure grants simpler analysis of this fragment. Instead of further investigating the tail domain in the context of the fully assembled filament, which involves additional preparation steps such as split intein chemistry and is limits NMR experiments to solid state only, the isolated tail in the polymeric fibrils can be utilized instead.

1.2.3. 50:50 Mixed Labeling

Disordered protein regions can form a cross-β fibril core composed of β sheets that run parallel to the axis of the fibril.[58] The observed NMR chemical shifts of the Tm1 I/C tail domain

are consistent with β-strand structures, and TEM images suggest the fibrils adopt a cross β-like structure like the scheme in figure 15.[38] The sheets are stabilized by intermolecular hydrogen bonds formed from the amide hydrogen and carbonyl oxygen of the protein backbone that space each molecule by 4.7A.[59]

Figure 15: Schematic of intermolecular and intramolecular nuclei interactions of 50:50 mixed ^{13}C labeled and ^{15}N labeled Tm1 I/C tail domain fibrils. The average distance of every other labeled molecule exceeds the observable internuclear distance using NMR, and limits observable signals to intramolecular interactions.



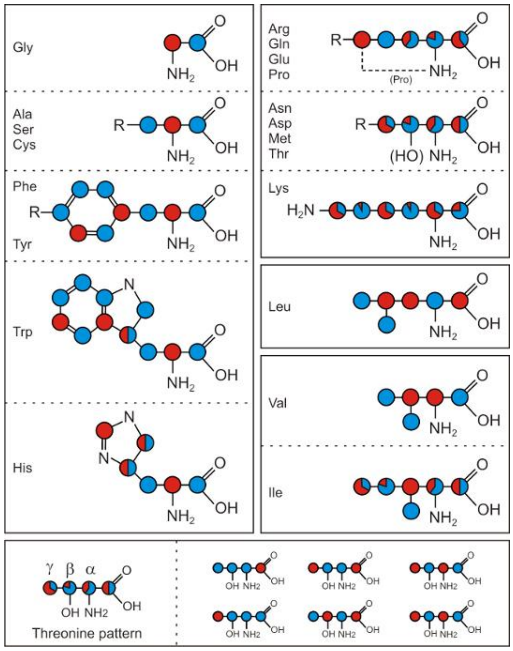
Because the structure of the tail domain in the fibril is inferred to be similar to its conformation in the full-length filament, the interactions in the fibril can be used to estimate how the tail domain folds in the IF by measuring intramolecular contacts in the fibril using solid state NMR. With the amino acid residues of the fibril core assigned to specific resonances, additional experiments can be used to estimate the intramolecular distances that can characterize the folds of the β-strands within each sheet. However, it is difficult to observe the intramolecular distances because there are competing nearby intermolecular signals from the neighboring molecules. Because the molecules in cross-β fibrils stack parallel along the long axis, the same amino acid residues on the primary sequence will be close in space on neighboring molecules, and they give rise to intermolecular magnetization transfers that compete with the desired intramolecular

magnetization transfer. To reduce these competing signals, a mix of 50% ^{13}C isotopically labeled molecules and 50% ^{15}N labeled molecules is prepared to create a fibril with no ^{13}C intermolecular contacts observable by NMR. This approach makes the average distance between labeled molecules around 9.4 Å, which is larger than the 3-8 Å measurable distance limit via dipolar coupling, and reduces intermolecular interactions.[60] By limiting the competing intermolecular interactions, fewer signals arise, and the intramolecular signals are relatively stronger and easier to assign.

1.2.4. Glycerol Labeling

Glycerol labeling can be used to achieve stronger sensitivity from distant spins in a complex system such as proteins by reducing the spectral overlap and the effects of dipolar truncation.[61] Dipolar truncation occurs when the strong dipolar coupling interactions from nearby spins attenuates the dipolar interactions with more distant spins.[62] Glycerol labeling is a labeling technique that reduces the number of labeled nuclei in the amino acids of proteins synthesized using bacterial expression systems. The glycolytic and pentose phosphate pathways can be exploited to synthesize amino acids in unique patterns of labeling that result in 'all or nothing' motifs of side chain carbons. Similarly, the citric acid cycle can be used to create amino acids with distinct populations of labeling called isotopomers. A map of these labeling patterns is shown in figure 16 when culturing bacteria in media with isotopically labeled $^{15}NH_4Cl$ and 2-^{13}C glycerol or 1,3 ^{13}C glycerol.

Figure 16: Glycerol labeling using 2-^{13}C glycerol shown in red and 1,3-^{13}C glycerol shown in blue. Labeling bacterially expressed proteins with 2 or 1,3-^{13}C glycerol takes advantage of the bacteria's metabolic pathways to create distinct labeling patterns called isotopomers. This helps to reduce the effects of dipolar truncation and resolve signals more easily. This figure was reproduced from Higman, 2012.



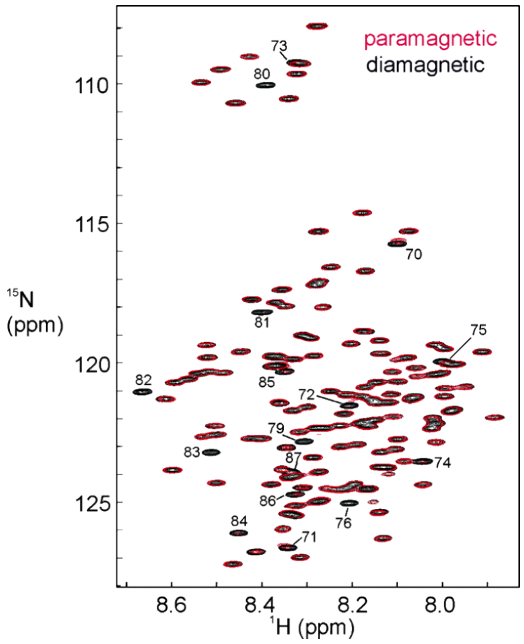
By sparsely labeling the spins on amino acids, the effects of dipolar truncation can be reduced to observe longer-range contacts more easily between nuclei on the backbone and the amino acid side chain. In addition to this, the reduced number of signals decreases the spectral overlap and results in more resolved signals.

1.2.5. Paramagnetic Relaxation Enhancement

Once residue assignments are made, paramagnetic relaxation enhancement (PRE) can be useful for determining protein structure using NMR. PRE overcomes the limit of measuring interatomic distance restraints for spacings larger than 5 Å between nuclei.[63] This is done by attaching a paramagnetic label at a carefully placed cysteine residue which creates strong hyperfine couplings between the unpaired electron of the label and the neighboring nuclei. In the resulting NMR spectrum, this is seen as relaxation enhancements. The broadening of observed resonances from spins nearby the paramagnetic label can occur up to 20 Å away. Changes in line width, relaxation rates or intensity can be compared to the relative intensities of corresponding signals from a spectrum recorded with a diamagnetic version of the paramagnetic label. These

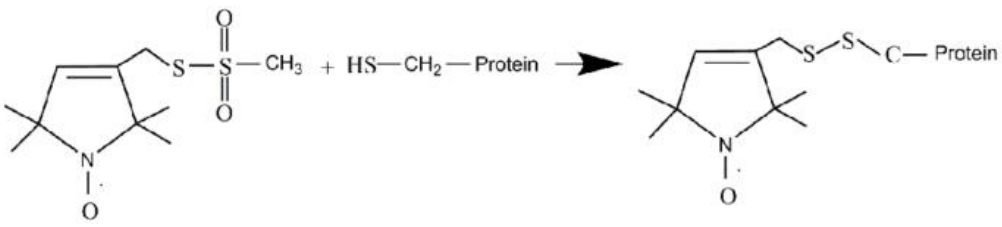
differences can be converted into structural parameters, such as the spatial proximity of spins. Figure 17 shows an example of data from a PRE experiment where the paramagnetic spin-labeled sample in red is overlayed with the reduced diamagnetic sample in black, and the subset of PRE-broadened signals are labeled.

Figure 17: ^{15}N-HSQC spectra of spin-labeled protein. The signals of the paramagnetic samples are in red and the signals of the reference diamagnetic sample is in black These data were reproduced from Dyson and Wright, 2004.



To implement this technique, the backbone chemical shifts of the protein need to be assigned first to interpret the distance restraints of the spectrum acquired from the spin-labeled sample.[64] An MTSSL nitroxide spin label (S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate) is commonly used for PRE NMR experiments because it is simple to use, relatively small, inexpensive and commercially available. The label is attached to a single cysteine residue via a disulfide bond, and the reaction is shown in figure 18.[65]

Figure 18: Reaction of MTSL with the thiol group on a cysteine residue from the protein of interest to form a disulfide bond. This figure was reproduced from Sahu and Lorigan, 2015.



This residue can be in the native protein sequence or substituted using site-directed mutagenesis, and all other cysteines should be removed. The substitution location needs to be solvent-exposed and in a defined secondary structure to not compromise the strength of the technique. It must also be verified that the cysteine substitution or removal does not alter the structure prior to spin-labeling. The label is attached to the ^{1}H-^{15}N-labeled and protein purified and maintained in a reducing agent like dithiothreitol to avoid aggregation. Once the spectrum of the labeled sample has been obtained, the intensities of the disrupted spin in the paramagnetic sample and the original signal in diamagnetic sample can be compared to calculate distant restraints. Multiple experiments can be done at various spin labeled sites to build up sufficient data to help establish a structure.

1.3. Research Goals

IF proteins are a challenge to study because of their typically disordered low-complexity head and tail domains. As a result, there is no high-resolution structure of any type of these proteins in their full-length form. Using solid state NMR techniques like PREs, segmental labeling, mixed labeling, and glycerol labeling to characterize disordered IF domains in their assembled state, it is promising that significant information can be gathered to progress towards a high-resolution structure, and to further characterize the IF assembly mechanism. This is important for investigating IF function beyond their long-known contribution to cell structure and can be used to better understand roles of many types across various tissues and species, which can ultimately lead to understanding IF in disease for treatment and prevention.

2. METHODS

2.1. $^{13}C/^{15}N$ Labeled Full Length Vimentin Preparation

2.1.1. Protein Expression

Full-length human vimentin protein was expressed in BL21(DE3) cells. Bacteria cells were grown overnight in a starter culture of LB media with 100 mg/L ampicillin, then diluted to 4 L LB media with ampicillin and grown at 37°C with shaking until an OD_{600} ~1.0 was reached. The cells were harvested by centrifugation at 6000 g for 10 minutes, and then transferred to 1 L M9 media with ^{13}C glucose and ^{15}N ammonium chloride for 30 minutes. Protein expression was induced using 0.5 mM isopropyl-β-D-thiogalactoside (IPTG), and cells were grown for 3 hours, harvested by centrifugation at 6000 g for 10 minutes, flash frozen and stored at -80°C

2.1.2. Inclusion Body Purification

The bacterial cells were resuspended in 40 mL Buffer #1 (50 mM Tris-HCl pH 7.5, 50 mM glucose, 10 mM EDTA, 10 mg/mL lysozyme), incubated at 37°C for 15 minutes and then split into two portions. To each portion, 20 ml Buffer #2 (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA, 1% v/v Triton X-100), 10 µg/mL DNAasel, 10 µg/mL RNAaseA, and 10 mM $MgCl_2$ was added, then the solutions were incubated at 37°C for 15 minutes and centrifuged on a tabletop centrifuge at 5000 g for 10 minutes. The pellets were resuspended in 10 mL Buffer #3 (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.5% v/v Triton X-100), combined, and centrifuged on a tabletop centrifuge at 5000 g for 10 minutes. The pellet was resuspended in 20 mL Buffer #4 (10 mM Tris-HCl pH 7.5, 1.5M KCl, 1 mM EDTA, 0.5% v/v Triton X-100) and centrifuged in a tabletop centrifuge at 5000 g for 10 minutes. The pellet was resuspended in 20 mL Buffer #5 (10 mM Tris-HCl pH 7.5, 0.15 M KCl, 1 mM EDTA, 0.5% v/v Triton X-100) and centrifuged in a tabletop centrifuge at 5000 g for 10 minutes. The pellet was resuspended in 20 mL Buffer #3

and centrifuged in a tabletop centrifuge at 5000 g for 10 minutes. The pellet was resuspended in 20mL Buffer #6 (8 M urea, 1 mM EDTA, 10 mM Tris-HCl pH 7.5) and centrifuged in a floor centrifuge at 20,000 g for 10 minutes. The supernatant was frozen at -80°C or immediately purified.

2.1.3. Protein Purification

The protein was purified using size exclusion chromatography on a HiLoad 16/60 Superdex 75 column using Buffer #6. The sample was divided into 5.5mL injections over three separate runs at a flow rate of 1 mL/minute. The sample fractions for all runs were collected and combined to a total of 36mL. The combined fractions were purified in a single anion exchange run on a Bio-Rad UNO Q6 column that eluted on a gradient from Buffer #6 to 100% Buffer #7 (10 mM Tris-HCl pH 7.5, 1 M NaCl, 8 M urea, 1 mM EDTA) at 4.5 mL/minute. This amount of sample was excessive and cause the column to clog, so further purification attempts used injections no larger than 24mL.

2.1.4. Assembly of Intermediate Filaments and Tetramers

Purified protein was dialyzed at room temperature under various conditions to assemble filaments and tetramers. Condition set #1 was dialyzed overnight for tetramers in 2 mM sodium phosphate, and filaments in 2 mM sodium phosphate and 200 mM KCl. Condition set #2 was conducted stepwise beginning with both tetramers and filaments dialyzed in 8 M urea and 2 mM sodium phosphate for 2 hours, then 4 M urea and 2 mM sodium phosphate for 2 hours, then 2 M urea and 2 mM sodium phosphate for 2 hours, and finally for 2 hours for tetramers in 2 mM sodium phosphate and filaments in 2 mM sodium phosphate and 200 mM KCl. Condition set #3 was conducted stepwise starting with both tetramers and filaments dialyzed overnight in 5 M urea and 2 mM sodium phosphate, then 4 hours in 2.5 M urea and 2 mM sodium phosphate, and

finally overnight for tetramers in 2 mM sodium phosphate and filaments in 2 mM sodium phosphate and 100 mM KCl. Condition set #4 was conducted stepwise and began with both tetramers and filaments dialyzed overnight in 5 M urea and 5 mM Tris-HCl pH 8.4, then for 4 hours in 2.5 M urea, 5 mM Tris-HCl pH 8.4, and finally overnight for tetramers in 5 mM Tris-HCl pH 8.4 and for filaments in 5 mM Tris-HCl pH 8.4 and 160 mM NaCl. Condition #5 was conducted stepwise beginning with both tetramers and filaments dialyzed in 8 M urea and 2 mM sodium phosphate for 2 hours, then 6 M urea and 2 mM sodium phosphate for 2 hours, then 4 M urea and 2 mM sodium phosphate for 2 hours, and finally for 2 hours for tetramers in 2 mM sodium phosphate and filaments in 2 M urea and 2 mM sodium phosphate. The filament harvest was spiked to 100 mM KCl. Condition #6 was conducted for just filaments and dialyzed overnight in 25 mM Tris-HCl pH 7.5 and 160 mM NaCl.

2.2. $^{13}C/^{15}N$ Labeled Tm1 I/C Tail Domain Preparation

2.2.1. Protein Expression

His-tagged Tm1 I/C Tail, in a pHis-parallel plasmid containing residues 373-441 of *Drosophila melanogaster* was expressed in BL21(DE3) cells. Bacteria cells were grown overnight in a starter culture of LB media with 100 mg/L ampicillin, then diluted to 4 L LB media with ampicillin and grown at 37°C with shaking until an OD_{600} ~1.0 was reached. The cells were harvested by centrifugation at 6000g for 10 minutes, and then transferred to 1 L M9 media with ^{13}C glucose and ^{15}N ammonium chloride for 30 minutes. Protein expression was induced using 0.5 mM IPTG, and cells were grown for 3 hours, harvested by centrifugation at 6000g for 10 minutes, flash frozen and stored at -80°C.

2.2.2. Protein Purification

The bacterial pellet was thawed over ice, and the cells were lysed using lysis buffer (10 mg/mL lysozyme, EDTA-free proteolytic inhibitor, 6 M guanidinium, 500 mM NaCl, 50 mM Tris HCl pH 7.5, 1% v/v Triton-X 100) and sonification.

The lysate was then purified using nickel-affinity chromatography on a BioRad Bio Scale Mini Nuvia IMAC Ni Charged 5 mL column. The sample was equilibrated in equilibration buffer (8 M urea, 20 mM HEPES pH 7.5, 500 mM NaCl, 5 mM β-mercaptoethanol (BME)), washed in wash buffer (equilibration buffer + 20 mM imidazole), then eluted in elution buffer (equilibration buffer + 200 mM imidazole) on a gradient from 0-100% at 2 mL/minute.

Two fractions were further purified using hydrophobic interaction chromatography on a Millipore Hi Trap Phenyl FF (HS) 5 mL column. The samples were combined and dialyzed into Buffer A (20 mM Tris HCl pH 7.5, 6 M urea, 5 mM BME, and 1 M ammonium sulfate) overnight at room temperature. The sample was eluted through the column in Buffer A followed by a gradient to 100% Buffer B (buffer A without ammonium sulfate) at 2 mL/minute to wash.

2.2.3. Preparation of Tail Domain Polymers

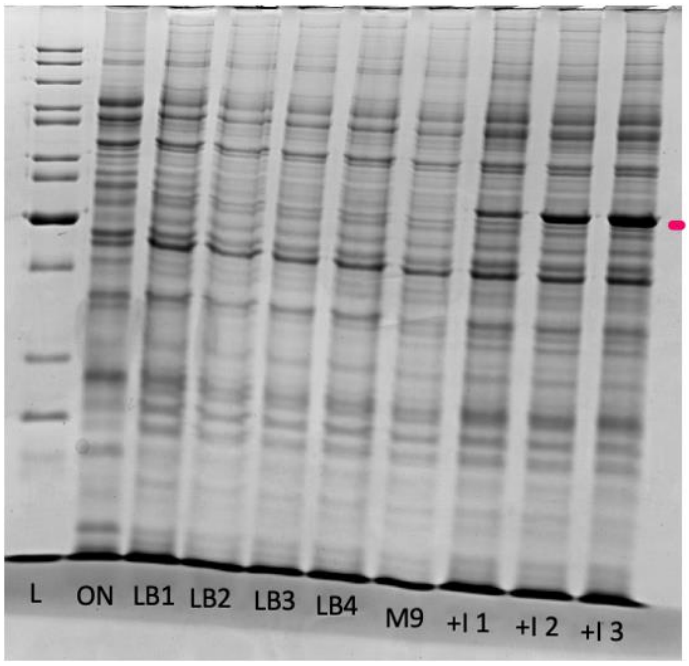
Purified protein was concentrated to 10 mg/mL and then dialyzed against gelation buffer (20 mM Tris HCl pH 7.5, 200 mM NaCl, and 5 mM BME) overnight at room temperature. The solution was centrifuged in a tabletop centrifuge at 3000g for 20 minutes to remove insoluble particles, sonicated for 5 minutes, and then concentrated to 50 mg/mL. The solution incubated for 6 days at 4°C to form fibril polymers.

3. RESULTS AND DISCUSSION

3.1. Uniformly Labeled Vimentin

To attempt assignment of amino acid residues signals for fully assembled vimentin, a full-length, uniformly labeled $^{15}N/^{13}C$ C328S vimentin variant was synthesized using bacterial expression and was monitored using SDS-PAGE electrophoresis (figure 19).

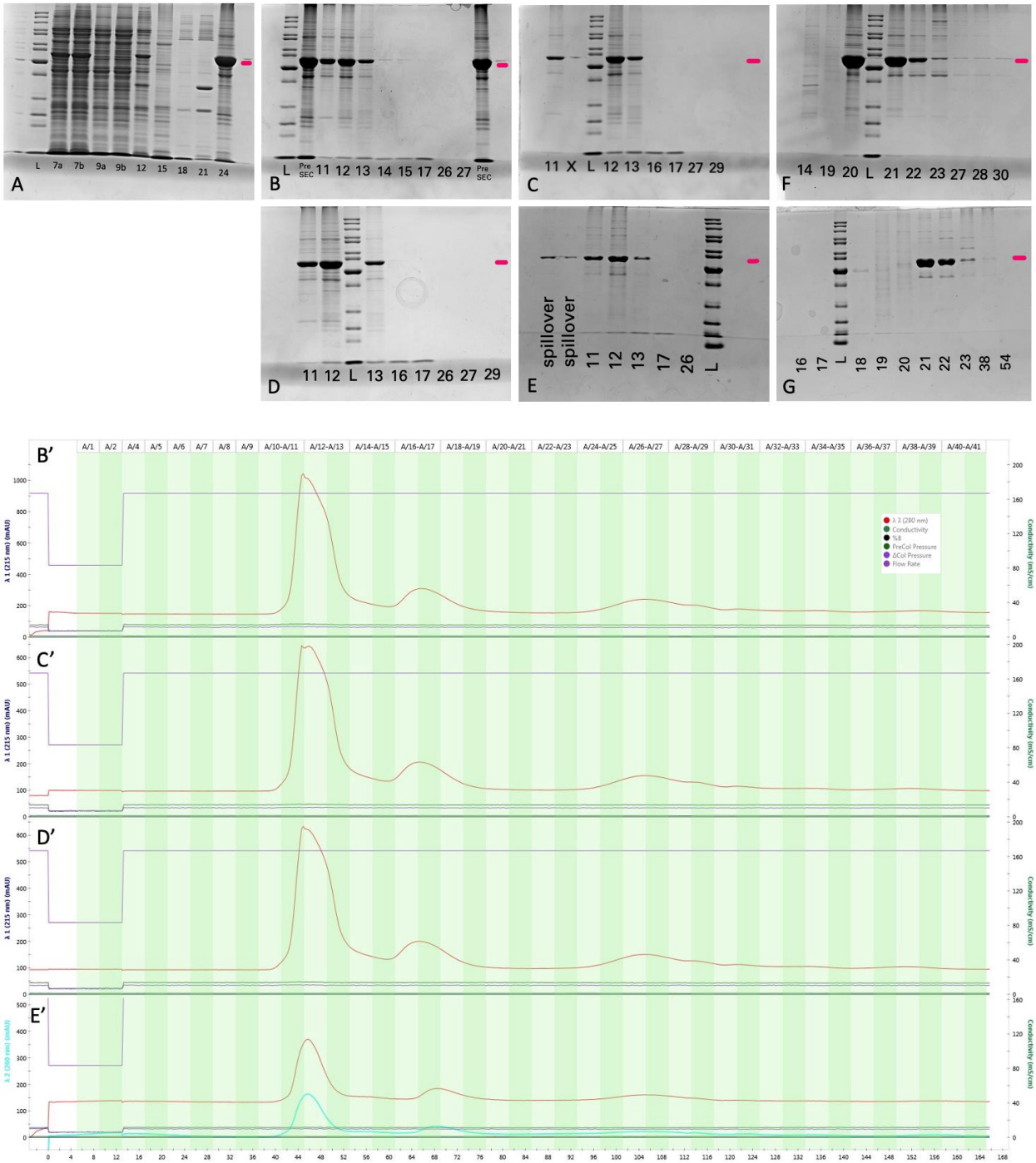
Figure 19: SDS PAGE gel of expressions of $^{15}N/^{13}C$ C328S vimentin. The protein size is 54 kDa and is marked in pink. Ladder (L); overnight culture (ON). IPTG at 1, 2, and 3 hour time points (+I 1, 2, 3).

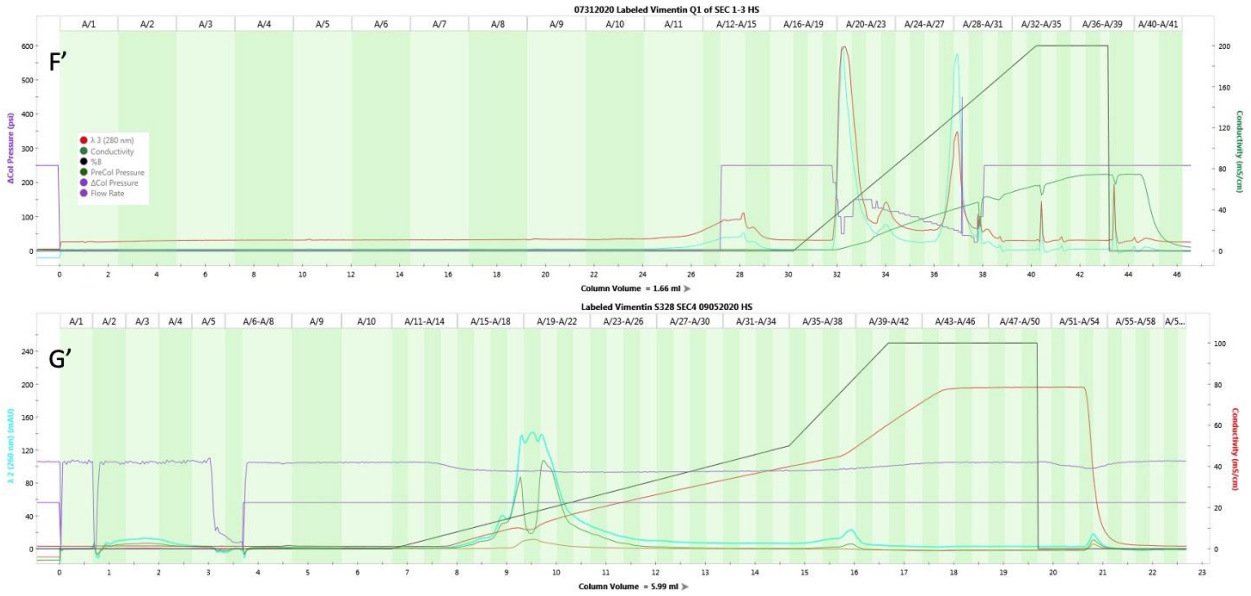


The serine substitution is used to eliminate the need for reducing conditions necessary to prevent cystine disulfide crosslinking. The protein was purified by isolating inclusion bodies, solubilizing them in denaturant, size exclusion chromatography (SEC), and anion exchange chromatography (figure 20). The fractions collected from three size exclusion runs were loaded onto a single anion exchange run and resulted in overpressure and clogged the column (figure F'). The sample was purified, but it required a new column as significant efforts to revive the original failed (figure G'). The total yield was around 35 mg.

Figure 20: Purification of $^{15}N/^{13}C$ C328S vimentin. SDS PAGE gel of the inclusion body lysis (A), four runs of SEC (B-E) and two runs of anionic exchange chromatography (F, G) of $^{15}N/^{13}C$ C328S vimentin. The protein size is 54

kDa and is marked in pink. Ladder (L); numbers represent lysis steps and fraction numbers; spillover and ‘x’ are rejected lanes. Chromatograms of SEC (B’-E’) and anion exchange (F’, G’).





Once purified, a stepwise dialysis was performed to conditions suitable to form tetramers and filaments most closely resembling physiological environments (condition set #1 in methods) and monitored using TEM (figure 21). Previously reported TEM data shows the expected morphology of vimentin tetramers is around 55 nm in length and unit length filament morphology is around 100 nm (figure 22.A), and fully assembled filaments much longer and uniform (figure 22.B). Nevertheless, in figure 21.A, the vimentin structures that were dialyzed into condition set #1 (methods) for tetramers have lengths that are much shorter than expected around 20 nm, and in figure 21.B, the structures dialyzed into filament conditions are a little longer than 100 nm, which is unexpected. Additional conditions for tetramer and filament assembly that were previously published were also tested, and they all yielded similar results. As it is important to verify that the vimentin protein is indeed in the properly assembled filament for the purposes of the NMR experiments, the conditions for assembling vimentin tetramers and filaments need to be further optimized.

Figure 21: TEM images of vimentin assemblies dialyzed into condition set #1 (A) tetramers in 2 mM sodium phosphate (B) filaments in 2 mM sodium phosphate, 200 mM KCl.

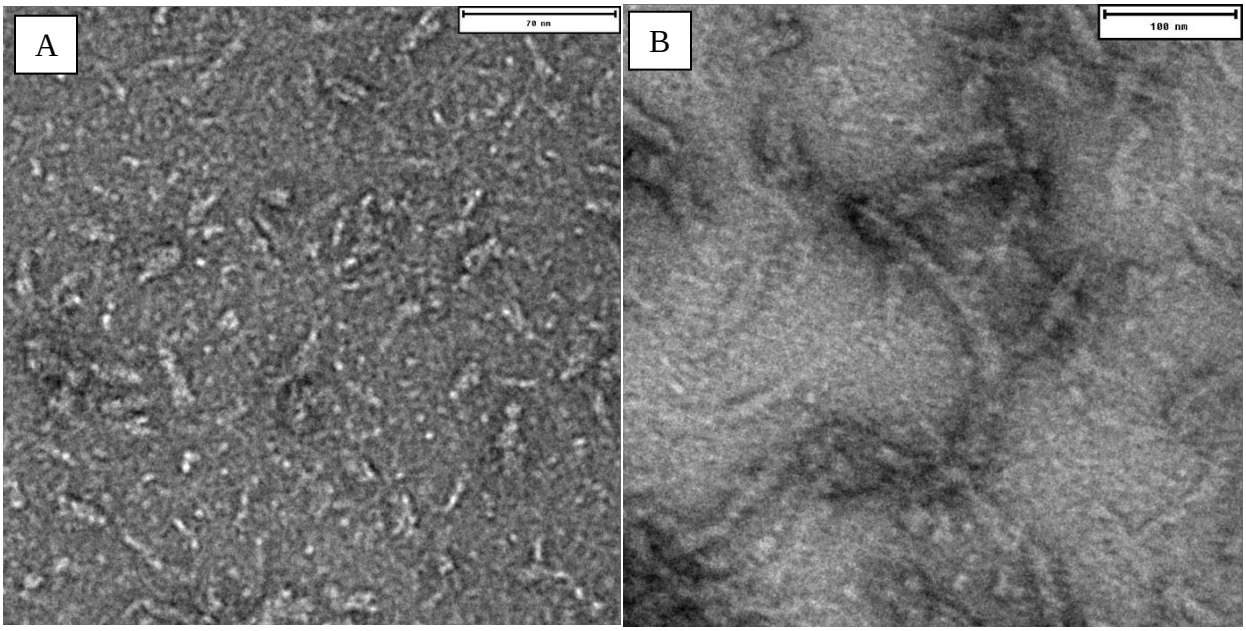
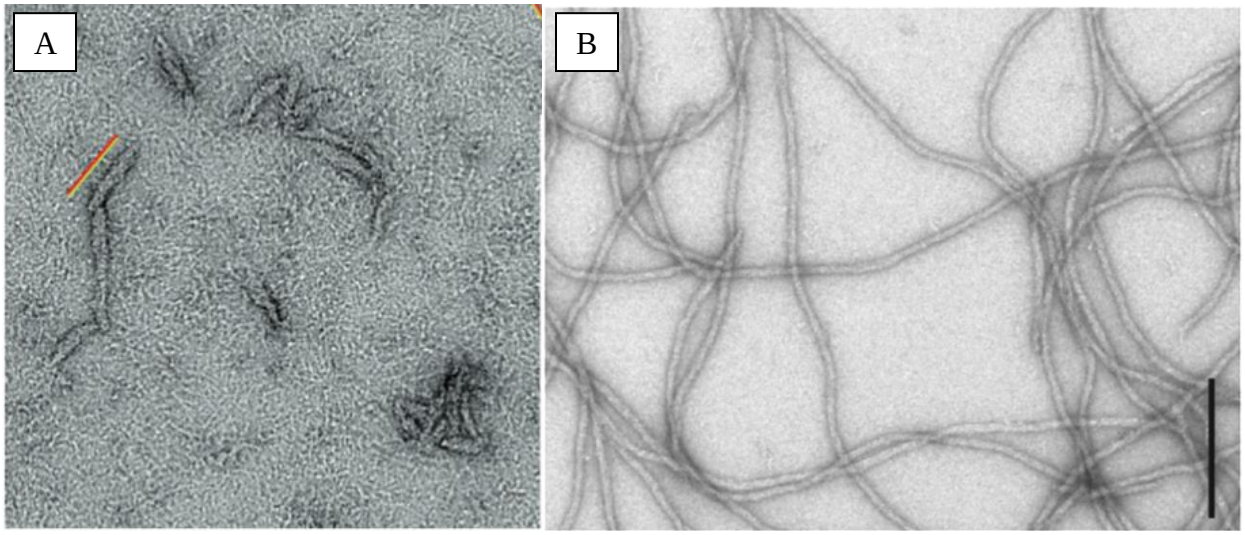


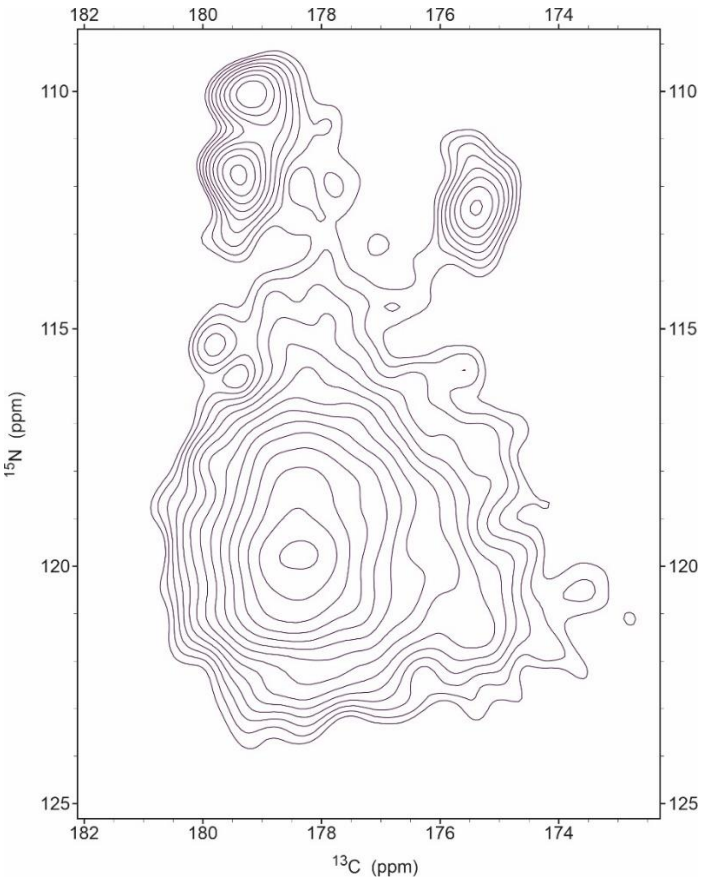
Figure 22: EM images of Vimentin assemblies in phosphate assembly buffer (A) formation of unit length filaments stopped almost immediately after initiation using chemical fixation; scale bar = 100 nm. This image is reproduced from Kirmse et. Al, 2010. (B) formation of extended filaments after 1 hour; scale bar = 200 nm. This image is reproduced from Winhein et. al, 2011.



NMR experiments of the vimentin samples were pursued to identify the structured residues and make sequence specific assignments. However, the residues were too difficult to

distinguish because of heavily overlapped signals, possibly resulting from the heptad repeat inherently present in the central coiled coil of intermediate filament assemblies. In the NCOCX spectrum shown in figure 23 the nitrogen chemical shifts around 95-130 ppm form cross peaks with the i-1 backbone carbonyl carbon chemical shifts between 170-182 ppm which arise from coupling interactions between these nuclei. Because these nuclei are in similar chemical environments along the coiled coil of the rod domain, the signals are heavily overlapped from heterogenous line broadening, and the individual residues are not able to be distinguished for assignment. To circumvent this issue and reduce the number of signals, samples with alternative labeling techniques need to be pursued.

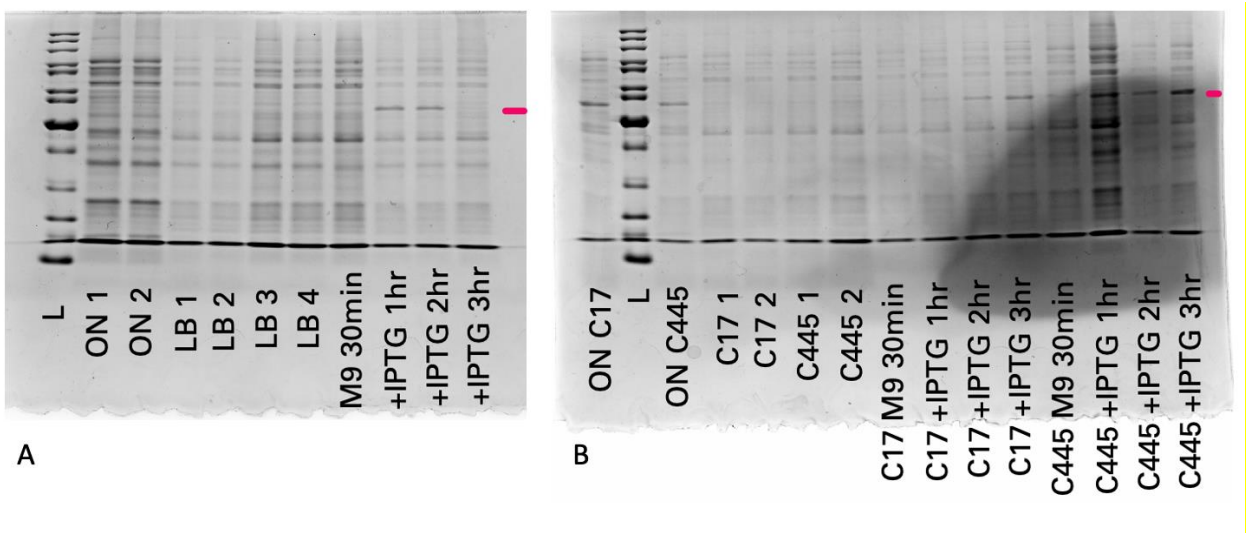
Figure 23: Uniformly labeled vimentin filament NCOCX NMR spectrum in the carbonyl region. Nuclei in similar chemical environments creates heterogeneous line broadening and overlap and creates a challenge for assigning residues.



3.2. Vimentin Sample Preparation for PRE

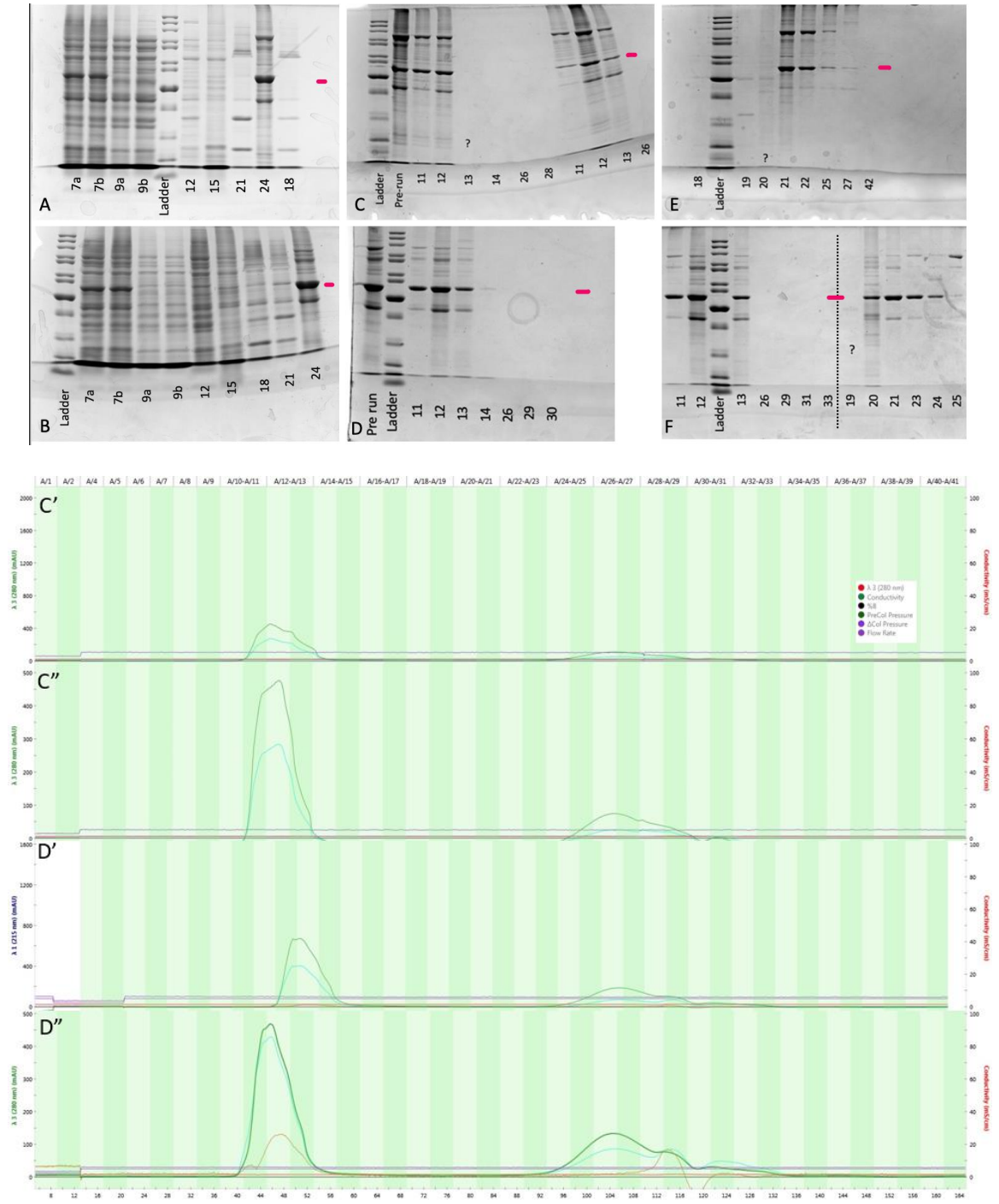
To prepare for after assignments are established, three vimentin recombinants were synthesized containing a single cysteine residue including wild type (wt) vimentin which has a single cysteine residue at 328, and mutants G17C and K445C for structure determination using PRE (figure 24).

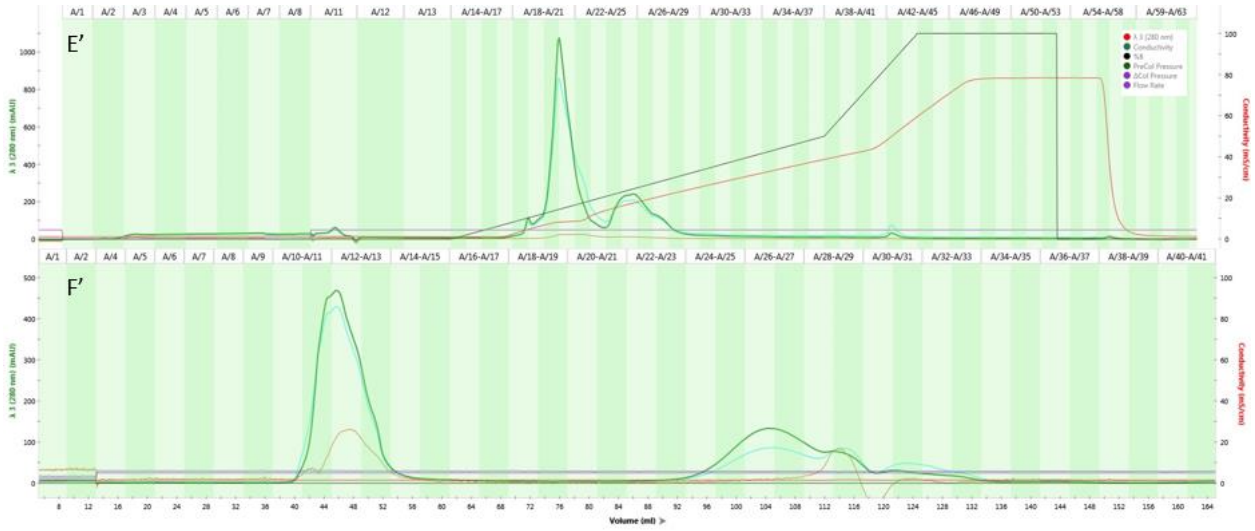
Figure 24: SDS PAGE gels of expressions of wt vimentin (C328) (A), and mutants G17C and K445C (B, C). The protein size is 54 kDa and is marked in pink. Ladder (L); overnight culture (ON).



The C17C and K445C samples were purified using an inclusion body lysis, SEC and anion exchange chromatography, and the yield was 13.1 mg and 11.7 mg respectively (figure 25). After purification, a spin label would be attached to the cysteine residue via a disulfide linkage and each variant would be individually used in PRE experiments to observe a reduction and broadening of signals from nuclei nearby in space to the PRE label. Accumulation of experiments to observe these changes among multiple samples could yield a significant amount of information to reveal the interactions of the disordered regions with the spin labels and provide insight to the structure in the fully assembled form**.**

Figure 25: Purification of $^{15}N/^{13}C$ vimentin mutants G17C and K445C. Gels of the inclusion body purification for G17C (A), and K445C (B); two runs of SEC for G17C (C), and K445C (D, F left); anion exchange chromatography for G17C (E), and K445C (F right). Chromatograms of the SEC for G17C (C', C"), and K445C (D', D"); anion exchange chromatography for G17C (E'), and K445C (F'). The protein size is 54 kDa and is marked in pink. Numbers represent lysis steps and fraction numbers.



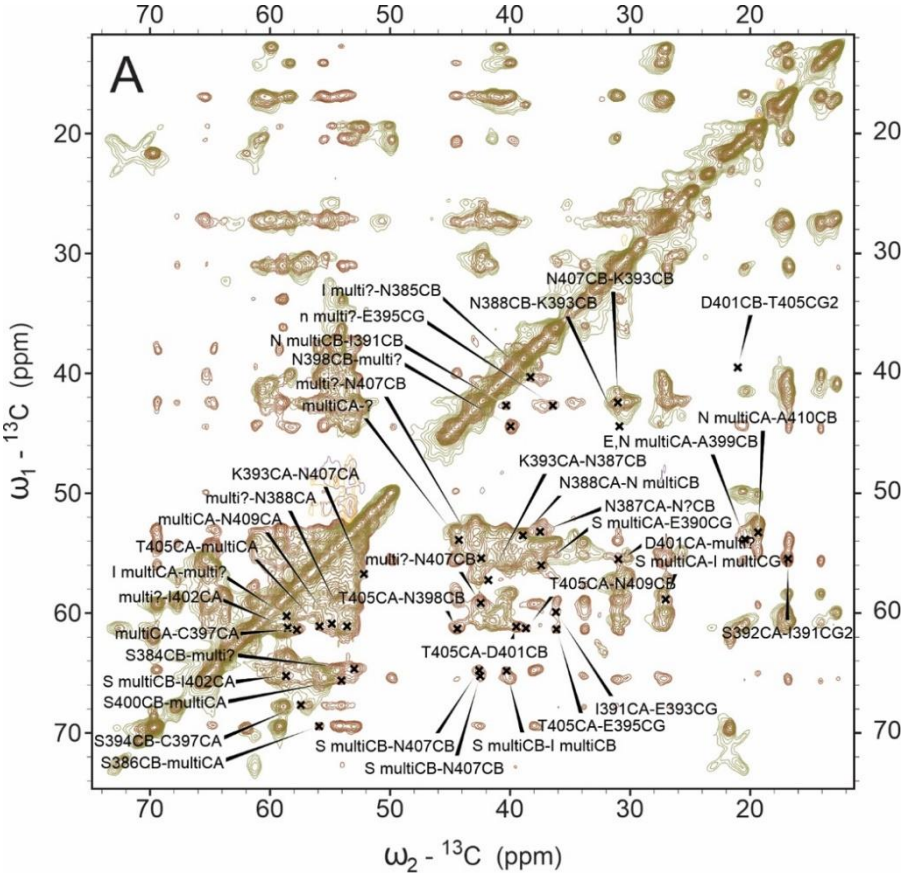


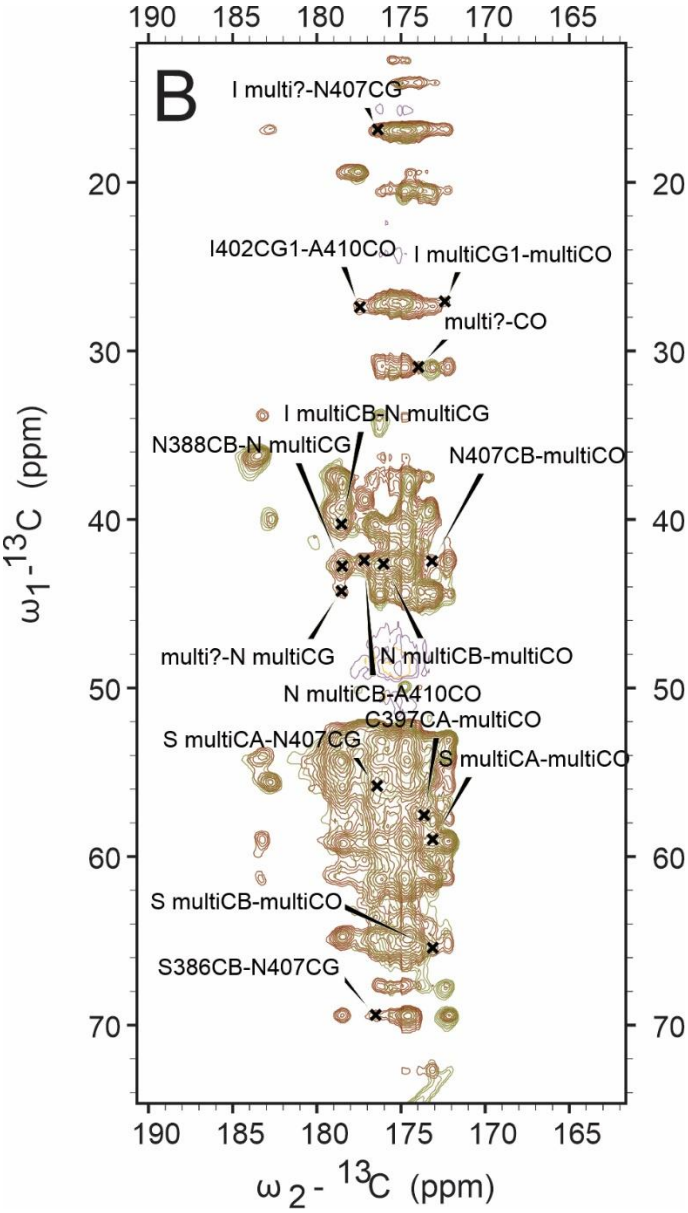
3.3. Tm1 I/C Tail Domain Structure Calculation

Assignments of tail domain only fibrils were previously made using NCACX and NCOCX experiments previously published by Sysoev et al in 2020. Additional data was collected using mixed labeling and glycerol labeling to help identify structural interactions, and the NCACX assignment data were used as a reference to identify residues in the new experiments.

Figure 26 shows the overlay of the CC DARR spectrum collected at a short 50 ms mixing time with the spectrum collected at 500 ms for the 50:50 mixed sample experiments. The measurement using short mixing time only reveals strong through-space interactions among nuclei nearby on the same residue or protein backbone. Therefore, the presence of a new signals in the measurement using longer mixing times indicates longer range magnetization transfers between nuclei, some of which are between different amino acids far apart in the protein sequence but relatively close together in space. These signals are most important for constraining the three-dimensional structure of the Tm1 monomer in the fibril. In figure 26 these signals are labeled according to which residues participated in the magnetization transfer. These signals are interpreted as distance constraints between 2 and 8 Å.[57]

Figure 26: Spectrum of Tm1 I/C Tail Domain Only Polymers 50:50 Mixed Labeling CC DARR 50 ms (green) overlay with 500 ms (brown). Signals observed in the 500 ms spectrum that are not present in the 50 ms spectrum and correlate with assigned residues three or more positions apart on primary sequence are considered for structural calculations. (A) Aliphatic region. (B) Carbonyl region.

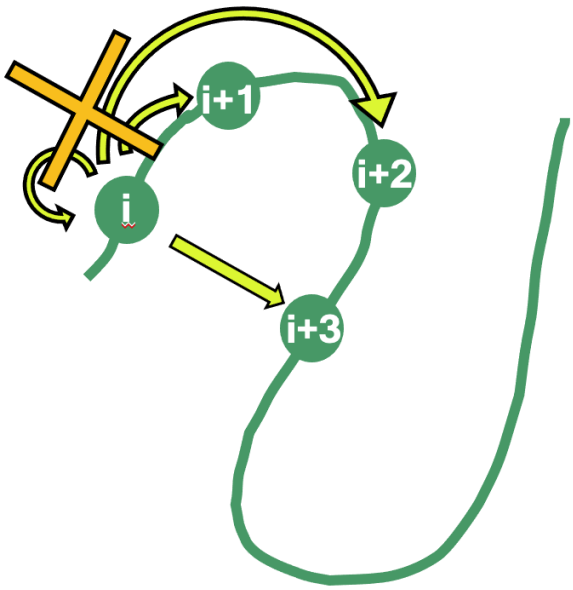




If the nuclei giving rise to a signal arose from amino acid residues within two residues of each other in the protein primary sequence, that signal was removed from consideration. This is because two carbon atoms the distance of two amino acid residues apart, which is about ~4-5 Å, can still be in a straight line and satisfy the distance constraint, and therefore these signals do not provide new structural information. However, if the signal arises from nuclei in residues that are more than two positions away in the primary sequence (i+3 or more), they are important structural constraints. Figure 27 illustrates the significance of a signal arising between nuclei that

are two or more residues apart; interactions beyond the distance constraint of residues on a straight line indicate a fold in the protein.

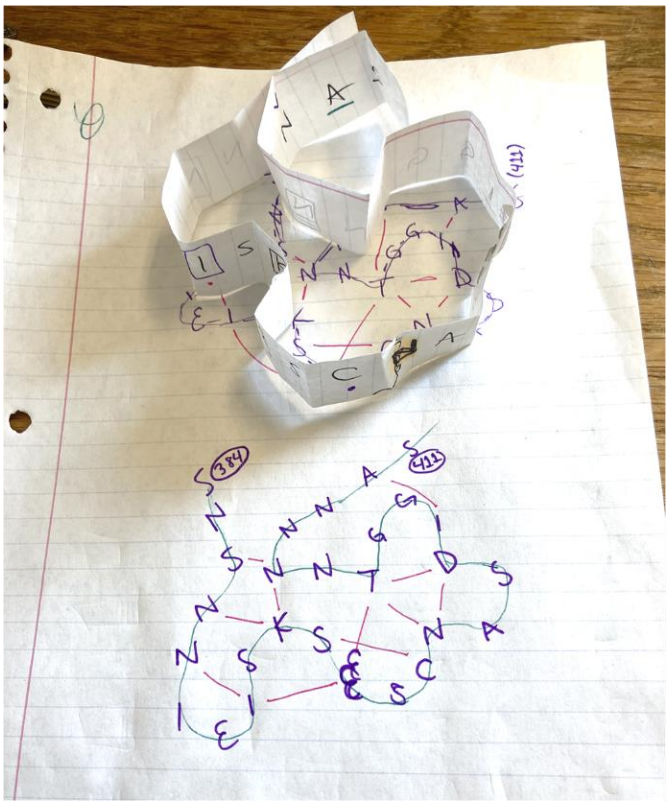
Figure 27: Schematic of amino acid residue distances along backbone and through space. Residues within two positions along the backbone are not considered due to proximity on the primary sequence and lack of structural information. Signals from residues more than two positions apart in the protein primary sequence give rise to important structural information



The distance constraints involving residues three or more positions away in the protein primary sequence are organized in table A1 according to ambiguity; if the signal had at least one chemical shift associated with multiple residues, then it was labeled ‘ambiguous’ and if it had only one residue for each chemical shift, then it was labeled ‘unambiguous.’ These distance constraints are currently being used to computationally determine a structure by another member of the lab, which can utilize both the ambiguous and unambiguous distance constraints. However, a structure consistent with the unambiguous contacts can be constructed using a strip of paper to represent the protein polypeptide backbone (figure 28). The amino acid pairs with unambiguous contacts were color or symbol coded on the sequence of the assigned tail domain. Then, the paper strip was folded to bring the pairs close in proximity until all of the contacts

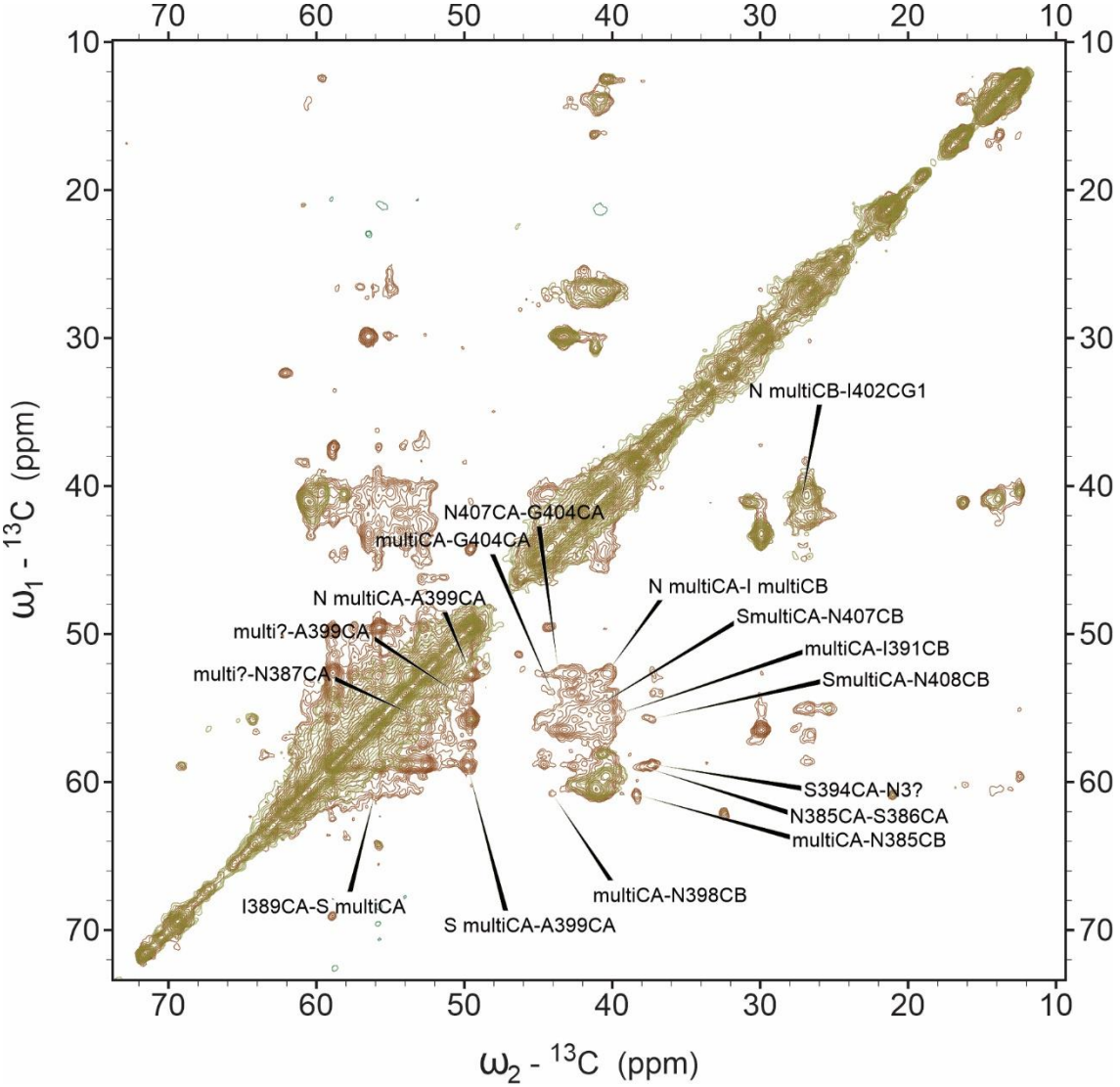
were accommodated in a reasonable structure. While this is not precise, it is similar to the preliminary computational data, which is very exciting.

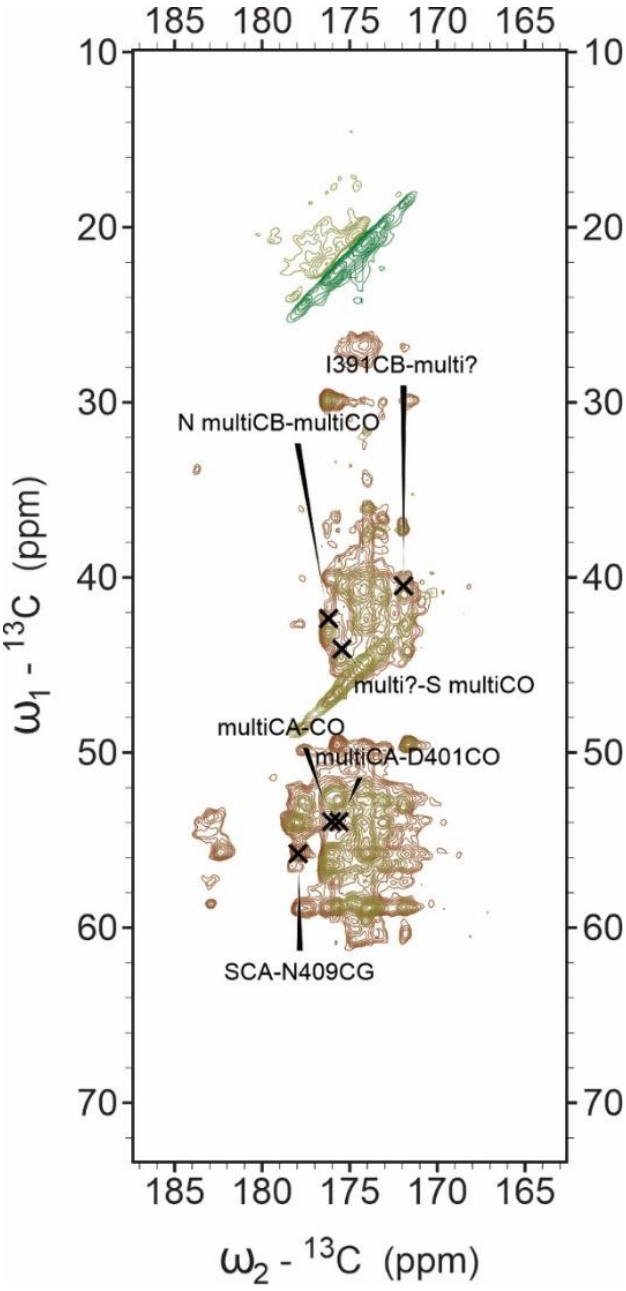
Figure 28: Preliminary model of the rigid portion of the Tm1 I/C tail domain fibril structure based on unambiguous NMR contacts.



Glycerol labeling lessens the effects of dipolar truncation by reducing strong, nearby couplings, and therefore allows strong magnetization transfers between distantly coupled nuclei that would have otherwise not been detected. These transfers give rise to signals in the spectrum of the glycerol labeled sample that are not present in spectrum of the fully labeled sample and helps to identify other long range contacts that can improve confidence in structural calculations. The relevant interactions of 2-glycerol labeled Tm1 I/C tail domain fibrils in CC DARR were tabulated the same way as for the mixed labeling experiment (figure 29, table A2).

Figure 29: Spectrum of Tm1 I/C Tail domain only polymers 2-glycerol labeling. The CC DARR 50 ms mixing time spectrum overlayed with the 500 ms spectrum. (A) Aliphatic region (B) Carbonyl region:



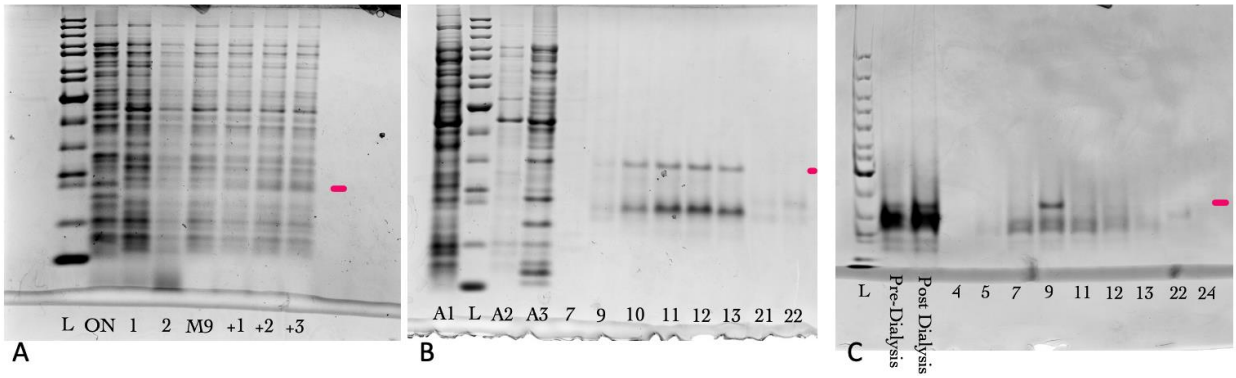


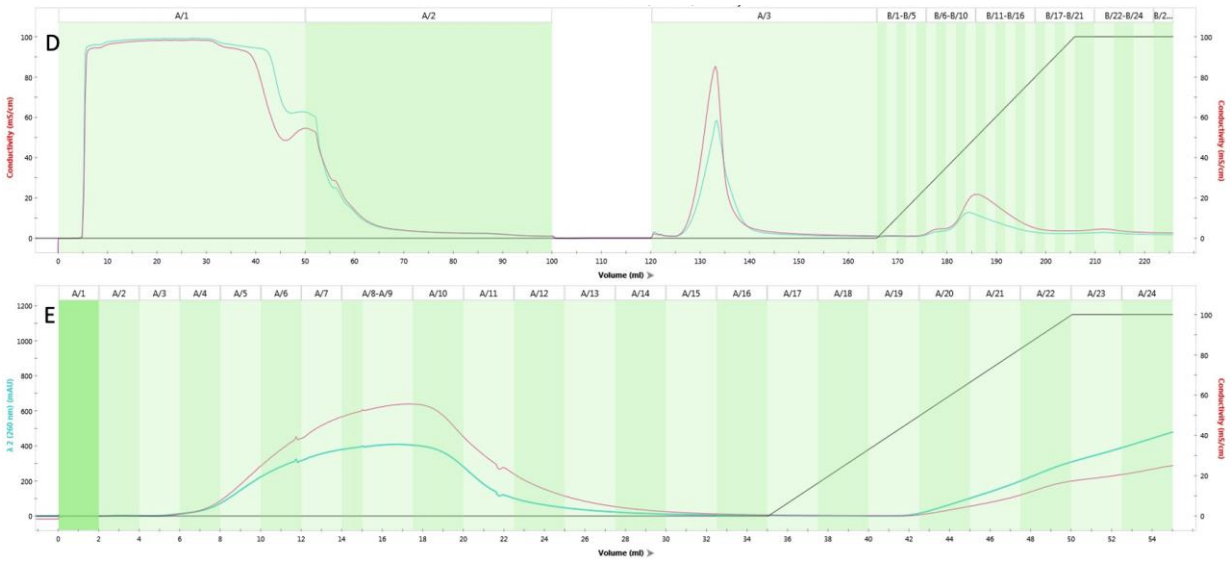
NCA and NCACX experiments on fully labeled samples correlate ^{15}N and ^{13}C interactions within an amino acid residue. The NCOCX experiment instead correlates the ^{15}N chemical shift of residue i with the ^{13}C chemical shifts of residue i-1. Similar to the procedure described previously for the mixed labeling samples, signals that were outside of the i+2 range were identified. In NCA and NCACX experiments, a comparison of the 2-glycerol labeled spectrum with the corresponding uniformly labeled spectrum was made to identify signals that

were not present in the uniformly labeled (figures A1-3, tables A3,4). This was also completed for 1,3-glycerol labeled samples measured in NCACX and NCOCX experiments (figures A4-7, tables A5,6). The CC DARR data for the 1,3-glycerol sample was collected and there are many contacts to sort out, especially in the asparagine region which the tail domain is rich in; however, this data is not completed yet. These distance constraints are also currently being incorporated into the structure calculation for the Tm1 I/C tail domain. There were many contacts observed from the NMR experiments on the mixed and glycerol labeled samples for the Tm1 I/C tail domain fibril. Among them, there were There were over 50 unambiguous contacts observed which came mostly from the 50:50 mixed labeled CC DARR, 1,3-glycerol labeled NCACX, and 1,3-glycerol labeled NCOCX experiments. These data are promising for calculating a high resolution of the disordered domain of this IF.

To probe the Tm1 I/C tail domain structure further, additional protein was synthesized using bacterial expression and purified using nickel affinity chromatography followed by hydrophobic interaction chromatography (figure 30).

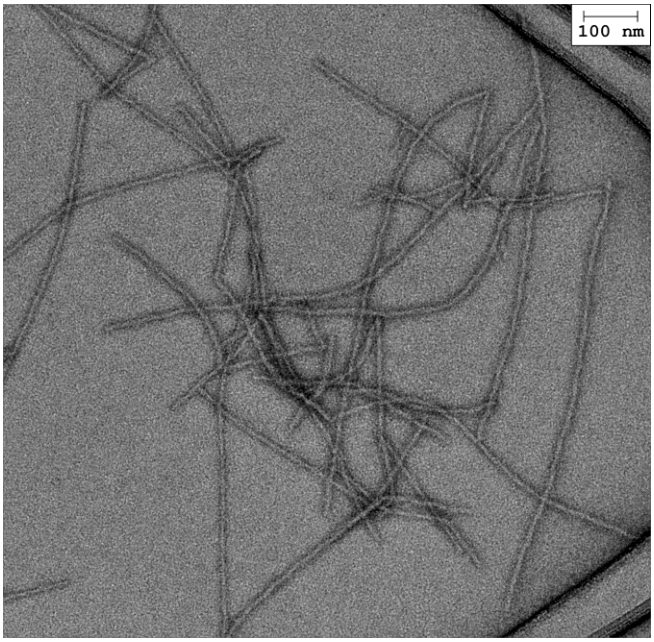
Figure 30: SDS PAGE gels of Tmi I/C tail domain expression (A) purification using nickel affinity chromatography (B) and hydrophobic interaction chromatography (C). Chromatograms of nickel affinity chromatography (D) and hydrophobic interaction chromatography. The protein size is 10.6 kDa and is marked in pink.





Fibrils were prepared from solubilized protein by dial ysis into gelation buffer and then concentrated to 50 mg/mL form fibrils, which were imaged using TEM. A representative micrograph of the fibrils is shown in figure 31. The fibrils are relatively straight, are mostly unbundled, and have a distinct twist.

Figure 31: TEM image of Tm1 I/C tail domain fibrils concentrated to 50 mg/mL in 20 mM Tris HCl pH 7.5, 200 mM NaCl, and 5 mM BME.



These fibrils are being used in measurements to determine the fibril mass per length analysis and to perform a cryo-electron microscopy image reconstruction. These experiments can determine other important structural constraints, such as the number of molecules per layer and a and low-resolution electron density map for the protein in the fibril.

4. SUMMARY

Using solid state NMR, the structures and assembly mechanisms of disordered LC protein domains can be probed and characterized on an atomic level. This is important for large assemblies of proteins with disordered regions like intermediate filaments, which cannot sufficiently isotropically tumble in solution NMR or form crystals suitable for crystallography. The large IF family has many implications in various tissue-specific diseases and there is ever-emerging evidence of functions beyond their well-known role of providing cellular structure. To investigate the structure of IF LC disordered domains in the context of assembly, the head and tail domains of vimentin and tropomyosin isoform Tm1 I/C were probed. Uniformly labeled full-length vimentin was analyzed using NMR to attempt to assign residues. Additional samples were prepared with cysteine mutations to be used in PRE experiments after assignments are made, The Tm1 I/C tail domain fibril was prepared using various isotopic labeling methods including 50:50 mixed, 2-glycerol and 1,3 glycerol labeling. These methods were used in an array of solid-state NMR experiments including CC DARR, NCACX and NCOCX to calculate distance restraints and are currently being used to computationally determine atomic structural information. Further analysis of these samples and data can contribute to the characterization of the still elusive high-resolution structure of intermediate filament protein.

REFERENCES

(1) Mier, P.; Paladin, L.; Tamana, S.; Petrosian, S.; Hajdu-Soltész, B.; Urbanek, A.; Gruca, A.; Plewczynski, D.; Grynberg, M.; Bernadó, P.; Gáspári, Z.; Ouzounis, C. A.; Promponas, V. J.; Kajava, A. V.; Hancock, J. M.; Tosatto, S. C. E.; Dosztanyi, Z.; Andrade-Navarro, M. A. Disentangling the Complexity of Low Complexity Proteins. *Brief. Bioinform.* **2020**, *21* (2), 458–472. https://doi.org/10.1093/bib/bbz007.

(2) Kulkarni, P.; Uversky, V. N. Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules* **2019**, *9* (4), 147. https://doi.org/10.3390/biom9040147.

(3) Kornreich, M.; Avinery, R.; Malka-Gibor, E.; Laser-Azogui, A.; Beck, R. Order and Disorder in Intermediate Filament Proteins. *FEBS Lett.* **2015**, *589* (19PartA), 2464–2476. https://doi.org/10.1016/j.febslet.2015.07.024.

(4) Hesse, M.; Magin, T. M.; Weber, K. Genes for Intermediate Filament Proteins. 7.

(5) Herrmann, H.; Hesse, M.; Reichenzeller, M.; Aebi, U.; Magin, T. M. Functional Complexity of Intermediate Filament Cytoskeletons: From Structure to Assembly to Gene Ablation. In *International Review of Cytology*; Academic Press, 2002; Vol. 223, pp 83–175. https://doi.org/10.1016/S0074-7696(05)23003-6.

(6) Mierke, C. T. *Physics of Cancer, Volume 3 (Second Edition): Experimental Biophysical Techniques in Cancer Research*; IOP Publishing, 2021. https://doi.org/10.1088/978-0-7503-3115-9.

(7) Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013**, *18* (9), 10802–10828. https://doi.org/10.3390/molecules180910802.

(8) Herrmann, H.; Aebi, U. Intermediate Filaments: Structure and Assembly. *Cold Spring Harb. Perspect. Biol.* **2016**, *8* (11), a018242. https://doi.org/10.1101/cshperspect.a018242.

(9) Robert, A.; Hookway, C.; Gelfand, V. I. Intermediate Filament Dynamics: What We Can See Now and Why It Matters. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **2016**, *38* (3), 232–243. https://doi.org/10.1002/bies.201500142.

(10) Bernot, K. M.; Lee, C.-H.; Coulombe, P. A. A Small Surface Hydrophobic Stripe in the Coiled-Coil Domain of Type I Keratins Mediates Tetramer Stability. *J. Cell Biol.* **2005**, *168* (6), 965–974. https://doi.org/10.1083/jcb.200408116.

(11) Mücke, N.; Kämmerer, L.; Winheim, S.; Kirmse, R.; Krieger, J.; Mildenberger, M.; Baßler, J.; Hurt, E.; Goldmann, W. H.; Aebi, U.; Toth, K.; Langowski, J.; Herrmann, H. Assembly Kinetics of Vimentin Tetramers to Unit-Length Filaments: A Stopped-Flow Study. *Biophys. J.* **2018**, *114* (10), 2408–2418. https://doi.org/10.1016/j.bpj.2018.04.032.

(12) Minin, A. A.; Moldaver, M. V. Intermediate Vimentin Filaments and Their Role in Intracellular Organelle Distribution. *Biochem. Mosc.* **2008**, *73* (13), 1453–1466. https://doi.org/10.1134/S0006297908130063.

(13) Wagner, O. I.; Rammensee, S.; Korde, N.; Wen, Q.; Leterrier, J.-F.; Janmey, P. A. Softness, Strength and Self-Repair in Intermediate Filament Networks. *Exp. Cell Res.* **2007**, *313* (10), 2228–2235. https://doi.org/10.1016/j.yexcr.2007.04.025.

(14) Omary, M. B.; Ku, N.-O.; Tao, G.-Z.; Toivola, D. M.; Liao, J. 'Heads and Tails' of Intermediate Filament Phosphorylation: Multiple Sites and Functional Insights. *Trends Biochem. Sci.* **2006**, *31* (7), 383–394. https://doi.org/10.1016/j.tibs.2006.05.008.

(15) Westermann, S.; Weber, K. Post-Translational Modifications Regulate Microtubule Function. *Nat. Rev. Mol. Cell Biol.* **2003**, *4* (12), 938–948. https://doi.org/10.1038/nrm1260.

(16) Omary, M. B. “IF-Pathies”: A Broad Spectrum of Intermediate Filament–Associated Diseases. *J. Clin. Invest.* **2009**, *119* (7), 1756–1762. https://doi.org/10.1172/JCI39894.
(17) Ku, N.-O.; Zhou, X.; Toivola, D. M.; Omary, M. B. The Cytoskeleton of Digestive Epithelia in Health and Disease. *Am. J. Physiol.-Gastrointest. Liver Physiol.* **1999**, *277* (6), G1108–G1137. https://doi.org/10.1152/ajpgi.1999.277.6.G1108.
(18) Liem, R. K. H.; Messing, A. Dysfunctions of Neuronal and Glial Intermediate Filaments in Disease. *J. Clin. Invest.* **2009**, *119* (7), 1814–1824. https://doi.org/10.1172/JCI38003.
(19) *Human Intermediate Filament Database*. http://www.interfil.org/statistics.php (accessed 2022-05-09).
(20) Parry, D. A. D.; Strelkov, S. V.; Burkhard, P.; Aebi, U.; Herrmann, H. Towards a Molecular Description of Intermediate Filament Structure and Assembly. *Exp. Cell Res.* **2007**, *313* (10), 2204–2216. https://doi.org/10.1016/j.yexcr.2007.04.009.
(21) Hess, J. F.; Budamagunta, M. S.; Aziz, A.; FitzGerald, P. G.; Voss, J. C. Electron Paramagnetic Resonance Analysis of the Vimentin Tail Domain Reveals Points of Order in a Largely Disordered Region and Conformational Adaptation upon Filament Assembly. *Protein Sci.* **2013**, *22* (1), 47–55. https://doi.org/10.1002/pro.2182.
(22) Chou, Y.-H.; Opal, P.; Quinlan, R. A.; Goldman, R. D. The Relative Roles of Specific N- and C-Terminal Phosphorylation Sites in the Disassembly of Intermediate FIlament in Mitotic BHK-21 Cells. 10.
(23) Wu, H.; Shen, Y.; Wang, D.; Herrmann, H.; Goldman, R. D.; Weitz, D. A. Effect of Divalent Cations on the Structure and Mechanics of Vimentin Intermediate Filaments. *Biophys. J.* **2020**, *119* (1), 55–64. https://doi.org/10.1016/j.bpj.2020.05.016.
(24) Laser-Azogui, A.; Kornreich, M.; Malka-Gibor, E.; Beck, R. Neurofilament Assembly and Function during Neuronal Development. *Curr. Opin. Cell Biol.* **2015**, *32*, 92–101. https://doi.org/10.1016/j.ceb.2015.01.003.
(25) Toivola, D. M.; Tao, G.-Z.; Habtezion, A.; Liao, J.; Omary, M. B. Cellular Integrity plus: Organelle-Related and Protein-Targeting Functions of Intermediate Filaments. *Trends Cell Biol.* **2005**, *15* (11), 608–617. https://doi.org/10.1016/j.tcb.2005.09.004.
(26) Conserved Segments 1A and 2B of the Intermediate Filament Dimer: Their Atomic Structures and Role in Filament Assembly. *EMBO J.* **2002**, *21* (6), 1255–1266. https://doi.org/10.1093/emboj/21.6.1255.
(27) Hess, J. F.; Budamagunta, M. S.; Shipman, R. L.; FitzGerald, P. G.; Voss, J. C. Characterization of the Linker 2 Region in Human Vimentin Using Site-Directed Spin Labeling and Electron Paramagnetic Resonance. *Biochemistry* **2006**, *45* (39), 11737–11743. https://doi.org/10.1021/bi060741y.
(28) Gae, D. D.; Budamagunta, M. S.; Hess, J. F.; McCarrick, R. M.; Lorigan, G. A.; FitzGerald, P. G.; Voss, J. C. Completion of the Vimentin Rod Domain Structure Using Experimental Restraints: A New Tool for Exploring Intermediate Filament Assembly and Mutations. *Structure* **2019**, *27* (10), 1547-1560.e4. https://doi.org/10.1016/j.str.2019.07.011.
(29) Aziz, A.; Hess, J. F.; Budamagunta, M. S.; Voss, J. C.; FitzGerald, P. G. Site-Directed Spin Labeling and Electron Paramagnetic Resonance Determination of Vimentin Head Domain Structure *. *J. Biol. Chem.* **2010**, *285* (20), 15278–15285. https://doi.org/10.1074/jbc.M109.075598.
(30) Chang, L.; Goldman, R. D. Intermediate Filaments Mediate Cytoskeletal Crosstalk. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (8), 601–613. https://doi.org/10.1038/nrm1438.

(31) Schweitzer, S. C.; Evans, R. M. Vimentin and Lipid Metabolism. *Subcell. Biochem.* **1998**, *31*, 437–462.
(32) Ivaska, J.; Pallari, H.-M.; Nevo, J.; Eriksson, J. E. Novel Functions of Vimentin in Cell Adhesion, Migration, and Signaling. *Exp. Cell Res.* **2007**, *313* (10), 2050–2062. https://doi.org/10.1016/j.yexcr.2007.03.040.
(33) Danielsson, F.; Peterson, M. K.; Caldeira Araújo, H.; Lautenschläger, F.; Gad, A. K. B. Vimentin Diversity in Health and Disease. *Cells* **2018**, *7* (10), 147. https://doi.org/10.3390/cells7100147.
(34) Herrmann, H.; Häner, M.; Brettel, M.; Müller, S. A.; Goldie, K. N.; Fedtke, B.; Lustig, A.; Franke, W. W.; Aebi, U. Structure and Assembly Properties of the Intermediate Filament Protein Vimentin: The Role of Its Head, Rod and Tail Domains. *J. Mol. Biol.* **1996**, *264* (5), 933–953. https://doi.org/10.1006/jmbi.1996.0688.
(35) Hess, J. F.; Budamagunta, M. S.; FitzGerald, P. G.; Voss, J. C. Characterization of Structural Changes in Vimentin Bearing an Epidermolysis Bullosa Simplex-like Mutation Using Site-Directed Spin Labeling and Electron Paramagnetic Resonance. *J. Biol. Chem.* **2005**, *280* (3), 2141–2146. https://doi.org/10.1074/jbc.M412254200.
(36) Schevzov, G.; Whittaker, S. P.; Fath, T.; Lin, J. J.; Gunning, P. W. Tropomyosin Isoforms and Reagents. *Bioarchitecture* **2011**, *1* (4), 135–164. https://doi.org/10.4161/bioa.1.4.17897.
(37) Cho, A.; Kato, M.; Whitwam, T.; Kim, J. H.; Montell, D. J. An Atypical Tropomyosin in Drosophila with Intermediate Filament-like Properties. *Cell Rep.* **2016**, *16* (4), 928–938. https://doi.org/10.1016/j.celrep.2016.06.054.
(38) Sysoev, V.; Kato, M.; Sutherland, L.; Hu, R.; Mcknight, S.; Murray, D. Dynamic Structural Order of a Low-Complexity Domain Facilitates Assembly of Intermediate Filaments. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*. https://doi.org/10.1073/pnas.2010000117.
(39) Ephrussi, A.; Lehmann, R. Induction of Germ Cell Formation by Oskar. *Nature* **1992**, *358* (6385), 387–392. https://doi.org/10.1038/358387a0.
(40) Dimitrova-Paternoga, L.; Jagtap, P. K. A.; Cyrklaff, A.; Vaishali; Lapouge, K.; Sehr, P.; Perez, K.; Heber, S.; Löw, C.; Hennig, J.; Ephrussi, A. Molecular Basis of MRNA Transport by a Kinesin-1–Atypical Tropomyosin Complex. *Genes Dev.* **2021**, *35* (13–14), 976–991. https://doi.org/10.1101/gad.348443.121.
(41) Lin, Y.; Mori, E.; Kato, M.; Xiang, S.; Wu, L.; Kwon, I.; McKnight, S. L. Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **2016**, *167* (3), 789-802.e12. https://doi.org/10.1016/j.cell.2016.10.003.
(42) Chernyatina, A. A.; Hess, J. F.; Guzenko, D.; Voss, J. C.; Strelkov, S. V. Chapter One - How to Study Intermediate Filaments in Atomic Detail. In *Methods in Enzymology*; Omary, M. B., Liem, R. K. H., Eds.; Intermediate Filament Proteins; Academic Press, 2016; Vol. 568, pp 3–33. https://doi.org/10.1016/bs.mie.2015.09.024.
(43) Tompa, P. On the Supertertiary Structure of Proteins. *Nat. Chem. Biol.* **2012**, *8* (7), 597–600. https://doi.org/10.1038/nchembio.1009.
(44) Yuwen, T.; Skrynnikov, N. R. CP-HISQC: A Better Version of HSQC Experiment for Intrinsically Disordered Proteins under Physiological Conditions. *J. Biomol. NMR* **2014**, *58* (3), 175–192. https://doi.org/10.1007/s10858-014-9815-5.
(45) Danmaliki, G. I.; Hwang, P. M. Solution NMR Spectroscopy of Membrane Proteins. *Biochim. Biophys. Acta BBA - Biomembr.* **2020**, *1862* (9), 183356. https://doi.org/10.1016/j.bbamem.2020.183356.

(46) Polenova, T.; Gupta, R.; Goldbourt, A. Magic Angle Spinning NMR Spectroscopy: A Versatile Technique for Structural and Dynamic Analysis of Solid-Phase Systems. *Anal. Chem.* **2015**, *87* (11), 5458–5469. https://doi.org/10.1021/ac504288u.
(47) PDSD | Protein NMR.
(48) DARR | Protein NMR.
(49) Wishart, D. S. Interpreting Protein Chemical Shift Data. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58* (1), 62–87. https://doi.org/10.1016/j.pnmrs.2010.07.004.
(50) NCACX | Protein NMR.
(51) NCOCX | Protein NMR.
(52) Parry, D. Coiled-Coils in ??-Helix-Containing Proteins: Analysis of the Residue Types within the Heptad Repeat and the Use of These Data in the Prediction of Coiled-Coils in Other Proteins. *Biosci. Rep.* **1982**, *2*, 1017–1024. https://doi.org/10.1007/BF01122170.
(53) Steinert, P. M.; Steven, A. C.; Roop, D. R. The Molecular Biology of Intermediate Filaments. *Cell* **1985**, *42* (2), 411–419. https://doi.org/10.1016/0092-8674(85)90098-4.
(54) De Roo, J.; Yazdani, N.; Drijvers, E.; Lauria, A.; Maes, J.; Owen, J. S.; Van Driessche, I.; Niederberger, M.; Wood, V.; Martins, J. C.; Infante, I.; Hens, Z. Probing Solvent–Ligand Interactions in Colloidal Nanocrystals by the NMR Line Broadening. *Chem. Mater.* **2018**, *30* (15), 5485–5492. https://doi.org/10.1021/acs.chemmater.8b02523.
(55) Shah, N. H.; Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chem. Sci.* **2013**, *5* (2), 446–461. https://doi.org/10.1039/C3SC52951G.
(56) Zhou, X.; Lin, Y.; Kato, M.; Mori, E.; Liszczak, G.; Sutherland, L.; Sysoev, V. O.; Murray, D. T.; Tycko, R.; McKnight, S. L. Transiently Structured Head Domains Control Intermediate Filament Assembly. *Proc. Natl. Acad. Sci.* **2021**, *118* (8), e2022121118. https://doi.org/10.1073/pnas.2022121118.
(57) Murray, D. T.; Kato, M.; Lin, Y.; Thurber, K. R.; Hung, I.; McKnight, S. L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171* (3), 615-627.e16. https://doi.org/10.1016/j.cell.2017.08.048.
(58) Fitzpatrick, A. W. P.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Müller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P. J.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V.; Griffin, R. G.; Dobson, C. M. Atomic Structure and Hierarchical Assembly of a Cross-β Amyloid Fibril. *Proc. Natl. Acad. Sci.* **2013**, *110* (14), 5468–5473. https://doi.org/10.1073/pnas.1219476110.
(59) Gorkovskiy, A.; Thurber, K. R.; Tycko, R.; Wickner, R. B. Locating Folds of the In-Register Parallel β-Sheet of the Sup35p Prion Domain Infectious Amyloid. *Proc. Natl. Acad. Sci.* **2014**, *111* (43), E4615–E4622. https://doi.org/10.1073/pnas.1417974111.
(60) Williamson, P. T. F.; Verhoeven, A.; Ernst, M.; Meier, B. H. Determination of Internuclear Distances in Uniformly Labeled Molecules by Rotational-Resonance Solid-State NMR. *J. Am. Chem. Soc.* **2003**, *125* (9), 2718–2722. https://doi.org/10.1021/ja028210u.
(61) 1,3-13C- and 2-13C-Glycerol | Protein NMR.
(62) Bayro, M. J.; Huber, M.; Ramachandran, R.; Davenport, T. C.; Meier, B. H.; Ernst, M.; Griffin, R. G. Dipolar Truncation in Magic-Angle Spinning NMR Recoupling Experiments. *J. Chem. Phys.* **2009**, *130* (11), 114506. https://doi.org/10.1063/1.3089370.
(63) Dyson, H. J.; Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **2004**, *104* (8), 3607–3622. https://doi.org/10.1021/cr030403s.

(64) Sjodt, M.; Clubb, R. T. Nitroxide Labeling of Proteins and the Determination of Paramagnetic Relaxation Derived Distance Restraints for NMR Studies. *Bio-Protoc.* **2017**, *7* (7), e2207. https://doi.org/10.21769/BioProtoc.2207.
(65) Sahu, I.; Lorigan, G. Biophysical EPR Studies Applied to Membrane Proteins. *Phys. Chem. Biophys.* **2015**, *5*. https://doi.org/10.4172/2161-0398.1000188.

APPENDIX

Table A1: Contacts of Tm1 I/C tail domain only polymers 50:50 mixed labeling CC DARR 500ms. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**
42.44	31.02	N407 CB	K393 CB
39.5	21.04	D401 CB	T405 CG2
42.71	40.38	N388 CB	I391 CB
67.68	57.44	S394 CB	C397 CA
56.75	52.19	K393 CA	N407 CA
61.13	39.46	T405 CA	D401 CB
61.24	38.67	T405 CA	N409 CB
59.9	36.2	I391 CA	E395 CG
61.34	36.14	T405 CA	E395 CG
57.26	41.83	K393 CA	N387 CB
56.75	52.19	K393 CA	N407 CA
27.42	177.4	I402 CG1	A410 CO
69.41	176.5	S386 CB	N407 CG

(B)

Ambiguous							
w1 (ppm)	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**	**w1 (ppm)**	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**
64.65	52.99	S384 CB	N398 CA	58.86	27.04	S394 CA	I391 CG1
			N387 CA			S386 CA	I402 CG1
60.25	58.6	I398 CA	S394 CA			S411 CA	
		I391 CA	I402 CA			S396 CA	
61.22	58.51	T405 CA	I402 CA	65.64	54.07	S400 CB	N406 CA
		S396 CB					N385 CA
61.08	53.58	T405 CA	N388 CA				E390 CA
		S396 CB					N408 CA
59.13	42.44	S386 CA	N407 CB				E395 CA
		S411 CA		65.24	58.66	S411 CB	I402 CA
		S394 CA				S392 CB	
		S396 CA		42.69	40.31	N388 CB	I391 CB

55.43	42.42	S392 CA S400 CA D401 CA	N407 CB			N407 CB	
53.52	38.94	N388 CA	N409 CB N406 CB	65.41	173.1	S392 CB S400 CB	I389 CO C403 CO T405 CO
53.85	20.42	E395 CA N408 CA E390 CA N406 CA N388 CA	A399 CB	57.75	175.4	C397 CA	S384 CO S392 CO N388 CG
53.26	19.31	N398 CA N387 CA	A410 CB	58.99	173.1	S394 CA S386 CA S411 CA S396 CA	I389 CO G403 CO T405 CO
44.41	39.98	N398 CB	D401 CB I402 CB	44.23	178.5	G403 CA N398 CB	N385 CG N406 CG N408 CG N409 CG
40.31	38.31	I402 CB I391 CB	N385 CB	42.44	177.2	N407 CB N387 CB	A410 CO
42.66	36.42	N407 CB N388 CB	E395 CG	42.66	176	N407 CB N388 CB	E395 CO D401 CO
42.68	37.9	N407 CB N388 CB	N385 CB	42.48	173.2	N407 CB	I389 CO G403 CO T405 CO
64.79	42.58	S384 CB S411 CB	N407 CB				
65.24	42.48	S392 CB S411 CB	N407 CB	40.25	178.5	I402 CB I391 CB	N385 CG N406 CG N408 CG N409 GG
64.8	40.31	S384 CB S411 CB	I391 CB I402 CB				
60.87	54.85	I389 CA T405 CA	N409 CA	42.77	178.5	N388 CB	N385 CG N406 CG N408 CG
61.08	55.85	T405 CA	D401 CA S400 CA S384 CA	30.96	173.9	I389 CG1 C397 CB K393 CB	S400 CO G404 CO
53.92	44.25	N388 CA E395 CA N408 CA N406 CA E390 CA	N398 CB G403 CA	27.06	172.4	I391 CG1 I402 CG1	N388 CO S394 CO
55.5	30.99	D401 CA	K393 CB C397 CB I389 CG1	16.86	176.4	I389 CD I402 CG1 I391 CG1	N407 CG
56.01	37.4	S384 CA S400 CA	E390 CG	57.53	173.6	C397 CA	I389 CO S400 CO G404 CO
61.38	57.77	T40G CA	C397 CA	55.84	176.4	S400 CA S384 CA	N407 CG

		I389 CA		175.9	178.6	E395CO	N385 CG
69.42	55.89	S386 CB	S400 CA			D401 CO	N406 CG
			D401 CA			N398 CG	N408 CG
55.17	33.86	N409 CA	E395 CB				
		D401 CA	E390 CB				

Table A2: Contacts of Tm1 I/C tail domain only polymers 2-glycerol labeling CC DARR 50ms overlay with 500ms. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**
52.59	43.64	N407 CA	G404 CA

(B)

Ambiguous							
w1 (ppm)	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**	**w1 (ppm)**	**w2 (ppm)**	**AA1 (C)**	**AA2 (C)**
58.9	49.78	S386 CA	A399 CA	55.74	37.38	S400 CA	N408 CB
		S394 CA				S384 CA	
		S411 CA		55.73	40.55	D401 CA	I391 CB
		S396 CA				S400 CA	
54.43	49.85	N385 CA	A399 CA			S384 CA	
		N408 CA		60.85	38.36	T405 CA	N385 CB
		E390 CA				I389 CA	
52.74	49.55	N407 CA	A399 CA	55.75	177.9	S400 CA	N409 CG
		N393 CA				S384 CA	
		N387 CA		53.96	175.6	N388 CA	D401 CO
55.79	52.75	S384 CA	N387 CA			N408 CA	
		S400 CA				N406 CA	
		D401 CA				E390 CA	
55.8	42.52	S400 CA	N407 CB			N385 CA	
		S384 CA				E395 CA	
52.66	40.4	N387 CA	I391 CB	42.36	176.2	N407 CB	E395 CO
		N398 CA	I402 CB			N387 CB	D401 CO
42.01	27.41	N387 CB	I402 CG	40.47	171.9	I391 CB	N388 CO
		N407 CB					N398 CO
60.51	55.98	I389 CA	S384 CA				N387 CG
			S400 CA	40.23	173.1	I402 CB	I389 CO
55.07	43.57	N409 CA	G404 CA				T405 CO

		S392 CA		175.6	173.1	D401 CO	I389 CO
60.78	44.07	T405 CA I389 CA	N398 CB				T405 CO

Figure A1: Spectrum of Tm1 I/C tail domain only polymers 2-glycerol labeling NCA overlay with uniformly labeled NCACX:

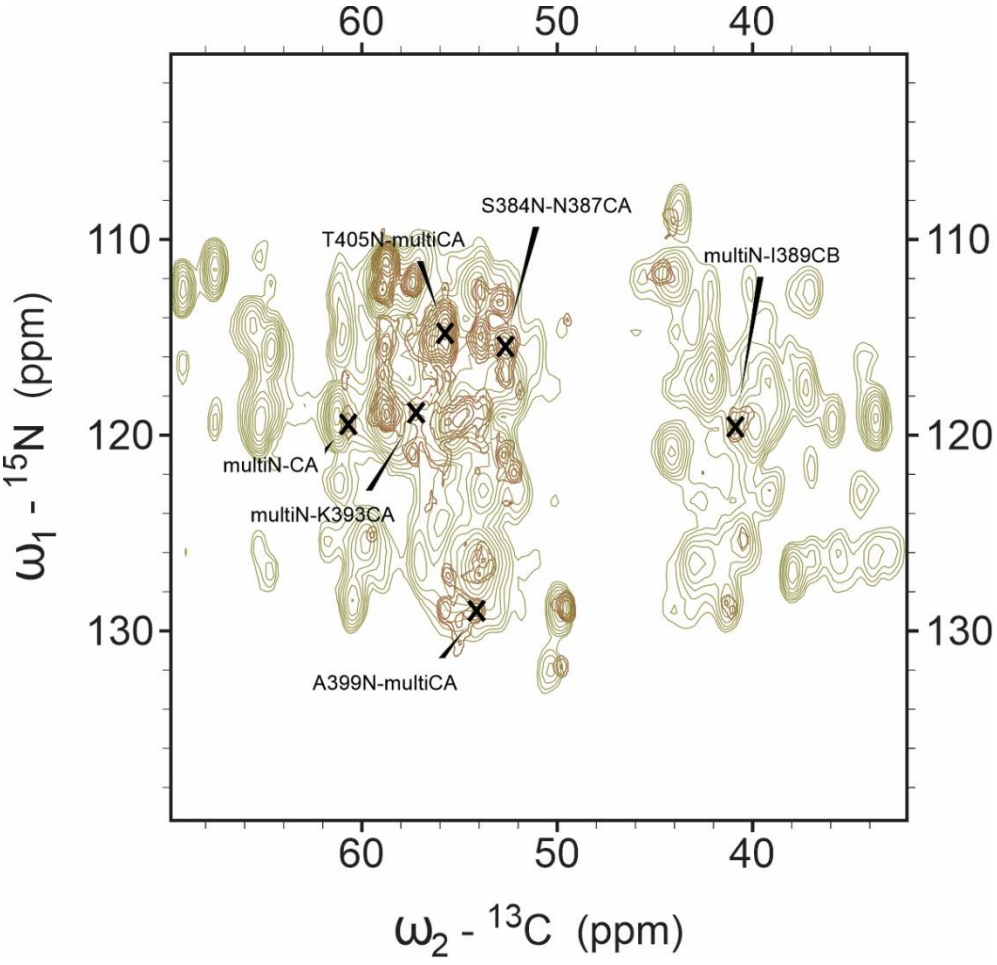


Table A3: Contacts of Tm1 I/C tail domain only polymers 2-glycerol labeling NCA overlay with uniformly labeled NCACX. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (CA)**

115.5	52.67	S384	N387

(B)

Ambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (CA)**
119.4	60.69	S396 E395 S392	I389 T405
114.8	55.76	T405	S400 S384 D401
118.8	57.24	S396 N409	K393
129	54.13	A399	N385 E390 N406 E395 N408
119.6	40.87	G403 D401 E395 S392	I398 CB

Figure A2: Spectrum of Tm1 I/C tail domain only polymers 2-glycerol labeling NCACX overlay with uniformly labeled NCACX aliphatic region:

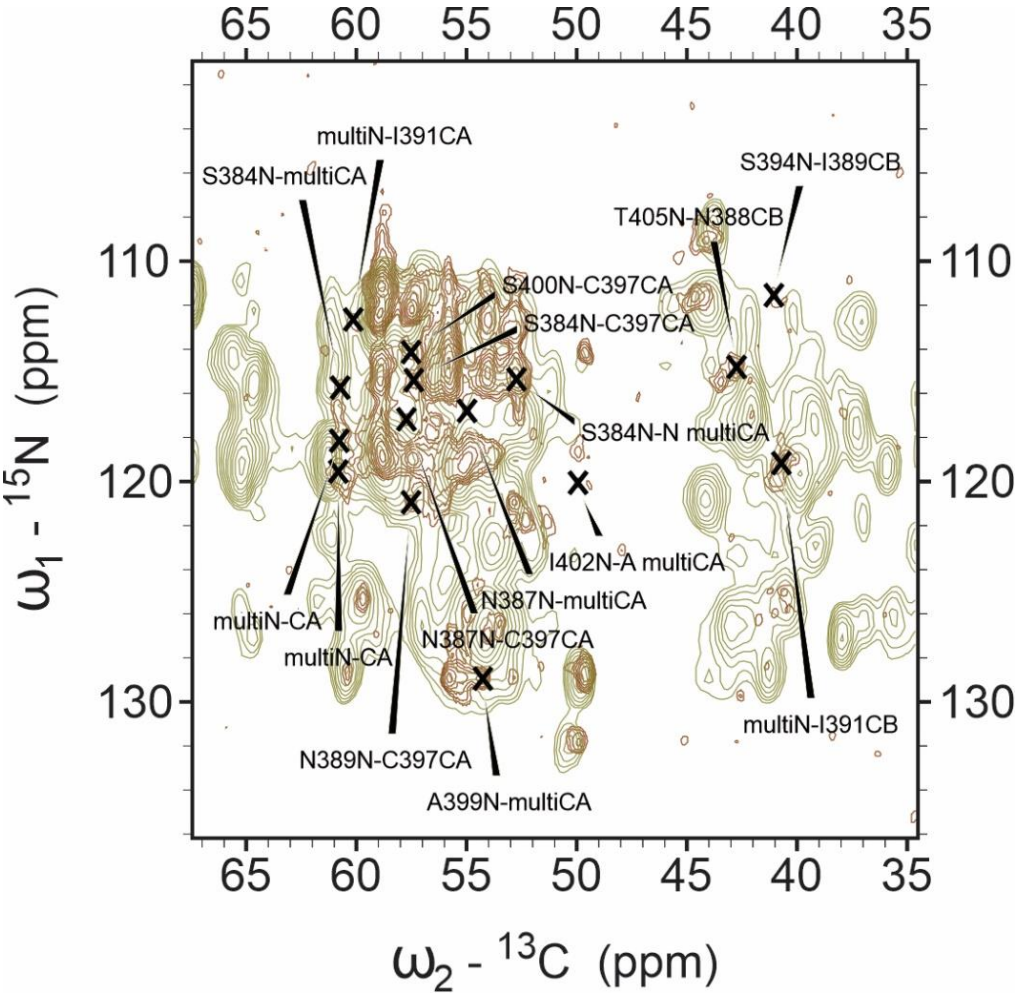


Figure A3: Spectrum of Tm1 I/C tail domain only polymers 2-glycerol labeling NCACX overlay with uniformly labeled NCACX carbonyl region:

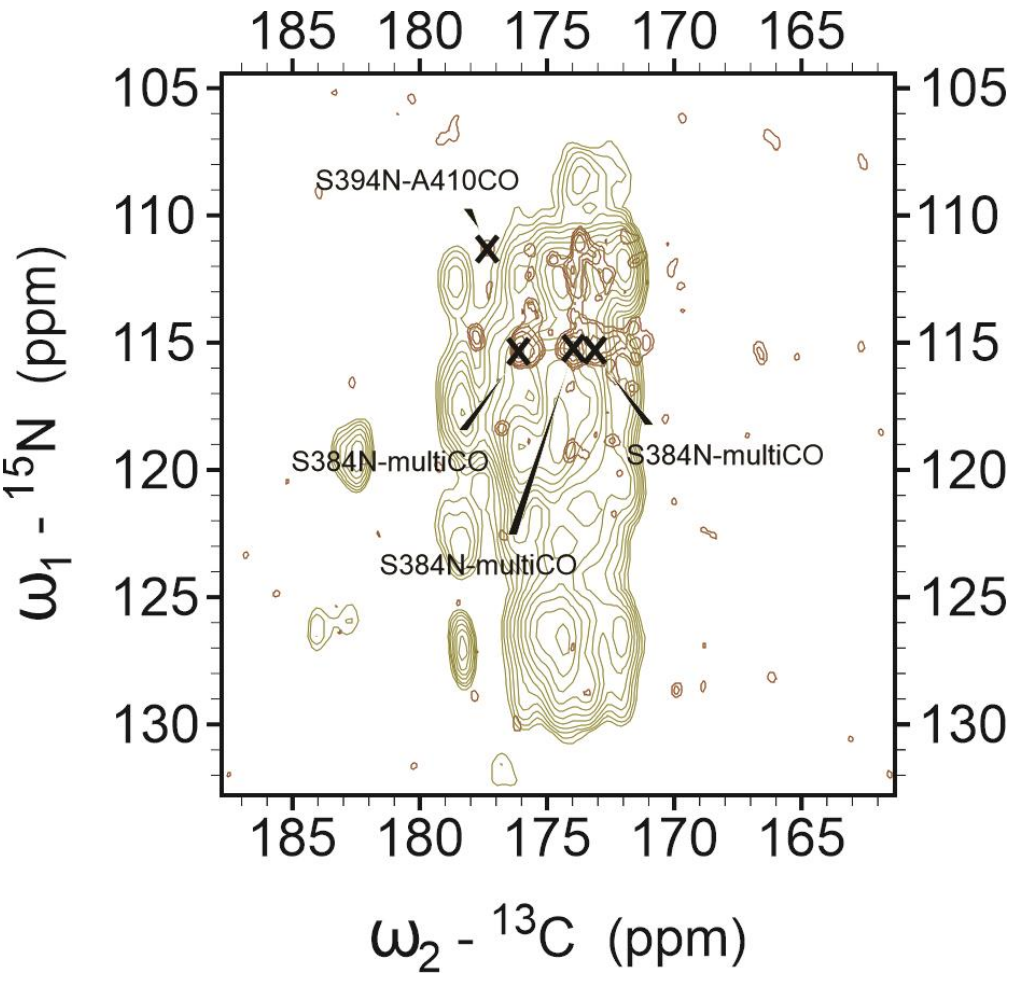


Table A4: Contacts of Tm1 I/C tail domain only polymers 2-glycerol labeling NCACX overlay with uniformly labeled NCACX. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
114.8	42.73	T405	N388 CB
115.4	57.4	S384	C397 CA
117.2	57.72	N387	C397 CA
114.1	57.53	S400	C397 CA
111.5	41.06	S394	I389 CB
120.9	57.33	N389	C397 CA

(B)

Ambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
129	54.28	A399	N408 CA E395 CA N406 CA E390 CA N385 CA
119.2	40.71	E395 S396 D401	I391 CB
115.4	176.1	S384	E395 CO D401 CO N3893 CO
115.3	173.1	S384	T405 CO I389 CO
116.8	54.99	N387	S392 CA N409 CA
108.9	58.84	G404	S394 CA S386 CA S411 CA S396 CA
112.6	60.6	S386 C397 N408	I391 CA
115.7	60.74	S384	T405 CA I389 CA
118.2	60.79	S411 N409	T405 CA I389 CA
120.1	49.93	I402	A399 CA A410 CA

Figure A4: Spectrum of Tm1 I/C tail domain only polymers 1,3-glycerol labeling NCACX overlay with uniformly labeled NCACX aliphatic region:

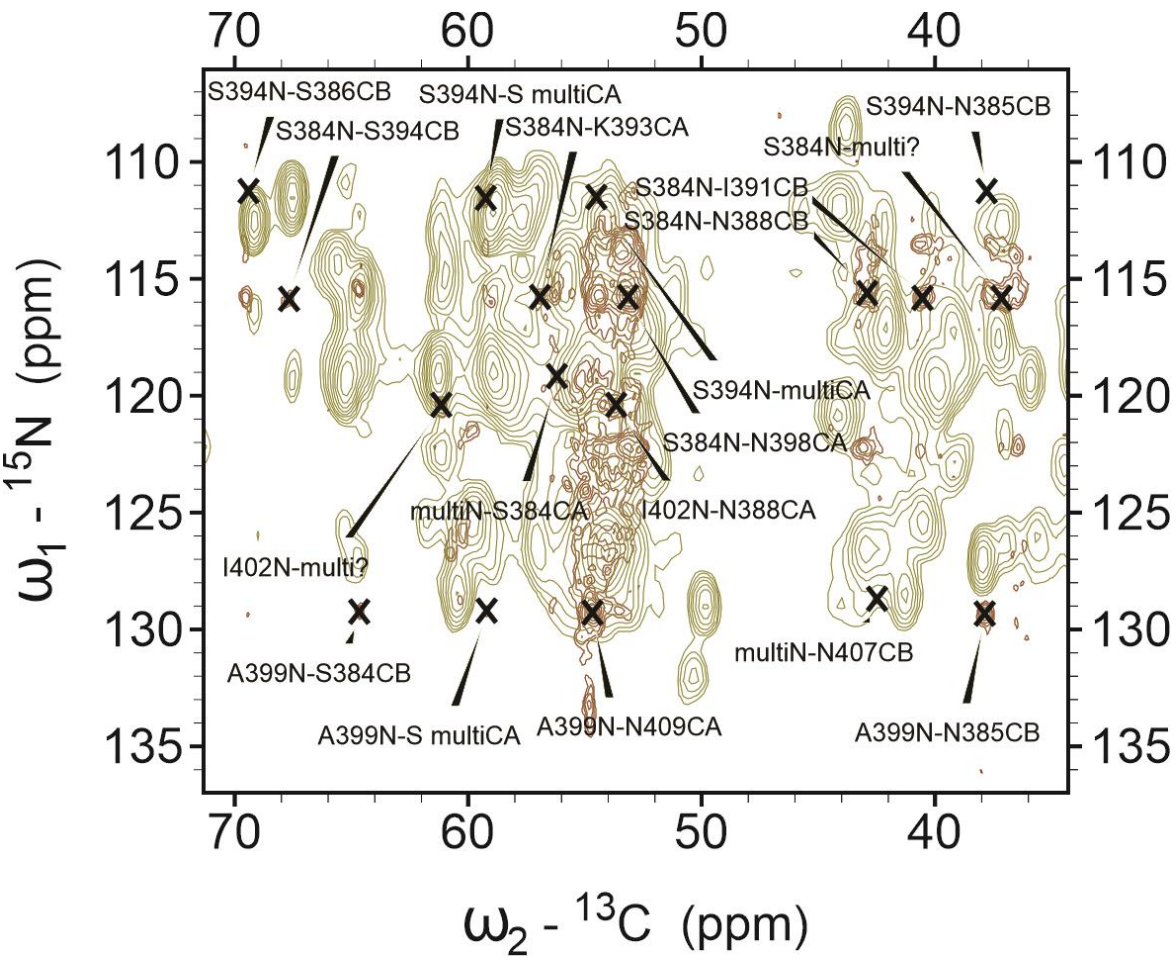


Figure A5: Spectrum of Tm1 I/C tail domain only polymers1,3-glycerol labeling NCACX overlay with uniformly labeled NCACX carbonyl region:

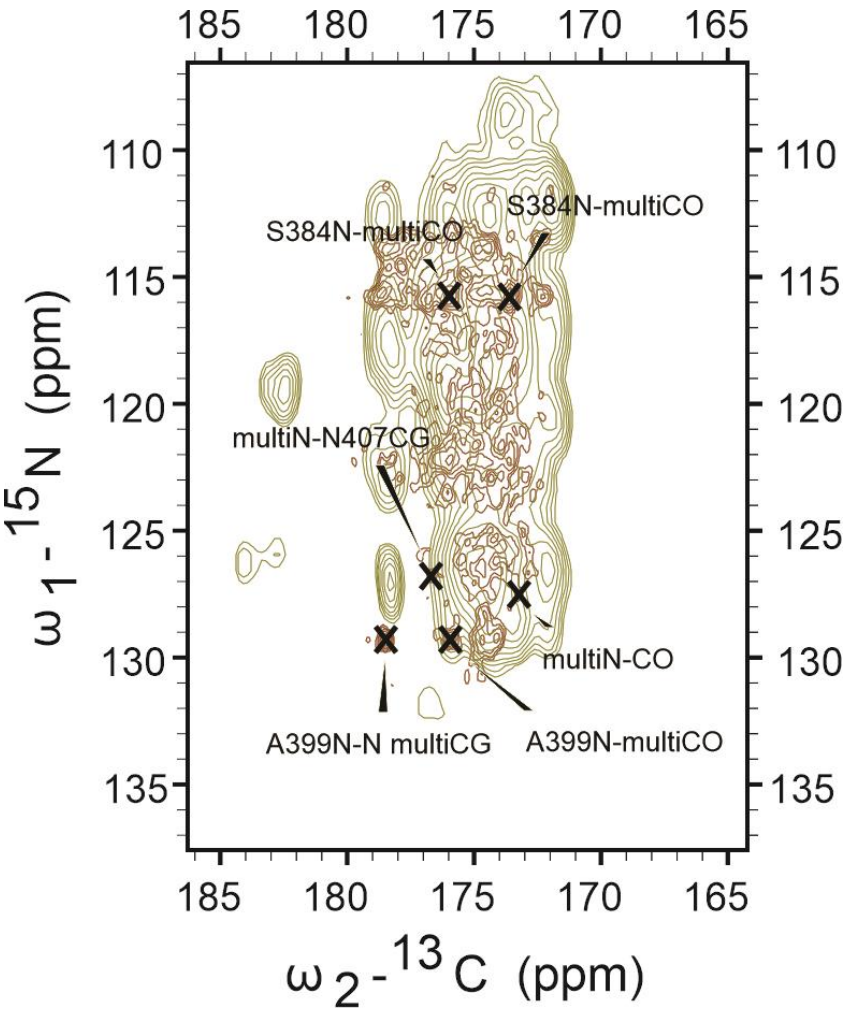


Table A5: Contacts of Tm1 I/C tail domain only polymers1,3-glycerol labeling NCACX overlay with uniformly labeled NCACX. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
115.8	56.93	S384	K393 CA
115.8	53.16	S384	N398 CA
120.4	53.64	I402	N388 CA
129.3	54.67	A399	N409 CA
115.9	67.68	S384	S394 CB
129.2	64.66	A399	S384 CB
111.2	69.4	S394	S386 CB
111.3	37.79	S394	N385 CB
115.6	42.91	S384	N388 CB

129.3	37.85	A399	N385 CB
123.1	13.84	N406	I402 CD1
128.5	13.94	I389	I402 CD1

(B)

Ambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
115.8	173.6	S384	S400 CO G404 CO 1389 CO
115.7	176	S384	N398 CO D401 CO E395 CO
126.8	176.7	E390 N388	N407 CG
127.5	173.2	N385 K393	G403 CO T406 CO
129.3	175.9	A399	N398CG D401 CO E395 CO
129.3	178.5	A399	N409 CG N385 CG N406CG N408 CG
111.5	54.52	S394	E390 CA E385 CA N409 CA
120.4	61.16	I402	T405 CA S396 CB
115.8	37.14	S384	N408 CB E390 CG
128.7	42.5	I389 A399	N407 CB

Figure A6: Spectrum of Tm1 I/C tail domain only polymers 1,3-glycerol labeling NCOCX overlay with uniformly labeled NCOCX aliphatic region:

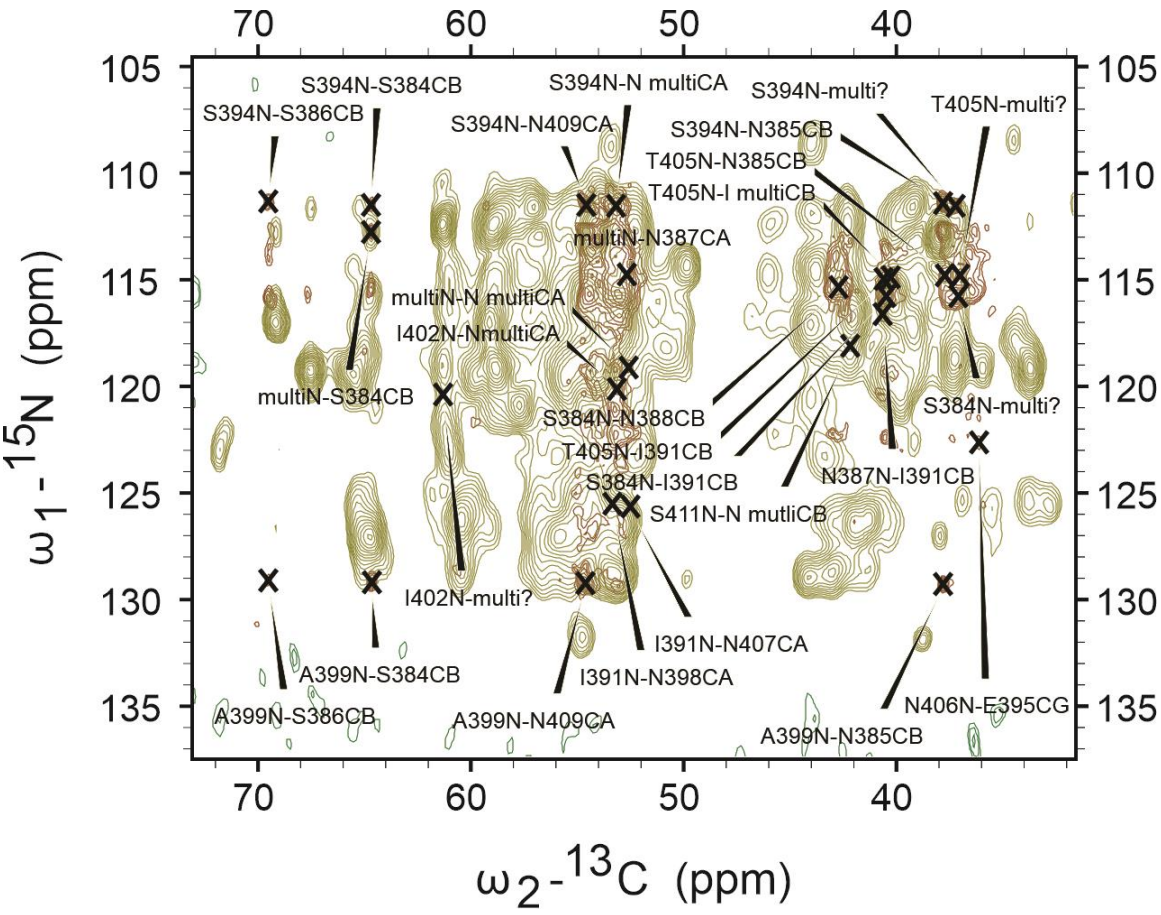


Figure A7: Spectrum of Tm1 I/C tail domain only polymers1,3-glycerol labeling NCOCX overlay with uniformly labeled NCOCX carbonyl region:

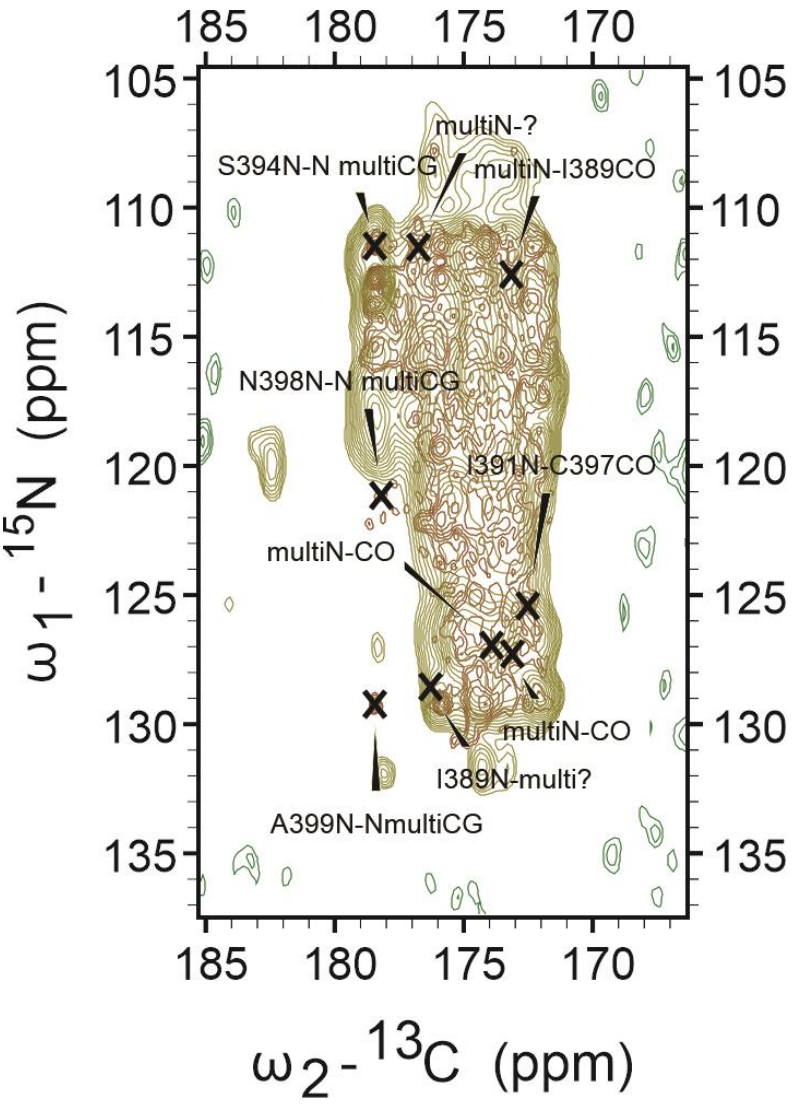


Table A6: Contacts of Tm1 I/C tail domain only polymers1,3-glycerol labeling NCOCX overlay with uniformly labeled NCOCX. Amino acid #1 (AA1), amino acid #2 (AA2). (A) Unambiguous contacts. (B) Ambiguous contacts.

(A)

Unambiguous			
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
125.4	172.5	I391	C397 CO
111.3	69.4	S394	S386 CB
111.5	64.68	S394	S384 CB
129.2	64.66	A399	S384 CB
129.1	69.51	A399	S386 CB
111.5	54.59	S394	N409 CA

125.5	53.33	I391	N398 CA
125.7	52.51	I391	N407 CA
129.2	54.62	A399	N409 CA
115	40.6	T405	I391 CB
115.4	42.7	S384	N388 CB
115.8	40.49	S384	I391 CB
116.6	40.63	N387	I391 CB
122.6	36.13	N406	E395 CG1
114.8	37.74	T405	N385 CB
129.3	37.79	A399	N385 CB
111.4	37.79	S394	N385 CB

(B)

Ambiguous							
w1 (ppm)	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**	**w1 (ppm)**	**w2 (ppm)**	**AA1 (N)**	**AA2 (C)**
111.6	176.8	S394 G403	N407 CG A410 CO	112.8	64.7	C397 N408	S384 CB
112.6	173.2	S386 C397 N408	I389 CO	120.4	61.3	I402	T405 CA S396 CB
				111.5	53.16	S394	N387 CA N398 CA
111.5	178.5	S394	N409 CG N385 CG N408 CG N404 CG	114.7	52.67	S400 T405	N387 CA
				119.1	52.58	S396 E395	N387 CA N407 CA
121.2	178.2	N398	N385 CG N409 CG	120.1	53.13	I402	N387 CA N398 CA
127	173.9	K393 N385 E390 N388	G404 CO S400 CO	114.9	40.26	T405	I402 CB I391 CB
				118.1	42.19	S411	N387 CB N407 CB
127.3	173.1	K393 N385	T405 CO G402 CO I389 CO	115.8	37.09	S384	N408 CB E390 CG
128.5	176.3	I389	E395 CO N398 CG	114.8	37.01	T405	N408 CB E390 CG
129.2	178.4	A399	N409 CG N385 CG	111.5	37.22	S394	N408 CB E390 CG