Probing the Hydrogen Bond Network of FUS Fibril Assemblies

Rebecca Rafique*

Mentor: Dr. Dylan T. Murray
Department of Chemistry at UC Davis

Introduction

The majority of eukaryotic proteins consist of both structured and intrinsically disordered regions (IDRs) (Dunker et al., 2013). The structured regions adopt well-defined tertiary structures and have well-defined functions. The intrinsically disordered regions, on the other hand, do not have a well-defined structure yet are still important for protein functions including regulation of transcription and translation, cellular signal transduction and the ordered assembly of macromolecular machines (Dyson et al., 2005). An important class of IDRs are the low complexity domains (LCD) of RNA-binding proteins that help regulate many stages of RNA processing (Calabretta, Richard, 2015). These domains are highly biased toward a small subset of the twenty naturally occurring amino acids. These regions do not contain a large fraction of small aliphatic and bulky hydrophobic residues like most globular proteins, but consist mainly of charged and polar amino acids.

IDRs are also present in yeast prion proteins. In prion diseases, normally well folded proteins adopt a detrimental conformation. An IDR in the prion protein catalyzes the transition from a functional state to the misfolded state observed in prion diseases. Spongiform Encephalopathy,

famously known as mad cow disease, is one of the many diseases caused by these infectious prion proteins. Prion-like domains are predicted to be found in around 250 human proteins, the majority being found in DNA and RNA binding proteins (Kim, et al., 2013). Several RNA binding proteins containing prion-like domains enriched in uncharged polar amino acids are associated with neurodegenerative diseases (Harrison, et al., 2017). Prion-like domains are genetically linked to many protein misfolding diseases and result in the formation of proteinaceous aggregates in patient tissues.

Fused in sarcoma (FUS) is an RNA binding protein involved in transcription and DNA repair that contains a prion-like domain. The prion-like domain of FUS is a LCD with an abundance of Tyrosine, Glycine, Serine and Glutamine residues. During cell stress, this domain promotes the formation of membraneless organelles called RNA granules. RNA granules have properties similar to liquid droplets which are able to effectively concentrate the FUS protein in a dynamic phase. Mutant FUS lacking the N-terminal prion-like domain fails to form liquid droplets, confirming that the prion-like domain is essential for forming phase-separated liquid droplets (Patel, et al., 2015).

Genetic mutations in the primary amino acid sequence of FUS causes the liquid droplet assemblies to transition to a less dynamic gel-like phase, composed of fibrils of the FUS protein. The fibril structures are more ordered than the RNA granules and are similar to the protein fibrils of Tau and Amyloid Beta that make up the amyloids observed in Alzheimer's. The fibrils are thought to represent the pathogenic aggregated state of FUS found in the cells of patients afflicted with Amyotrophic Lateral Sclerosis (ALS) (Kwiatkoswki, et al. 2009). Mutations in FUS play a large role in the formation of these pathogenic inclusions. Comparing the liquid droplet formation by wild

type FUS and FUS containing a G156E mutation using fluorescence microscopy reveals that the mutation makes FUS less dynamic in the droplets. After 8 hours, wild type FUS remains in liquid droplet form. For the mutated FUS protein, the droplets had transitioned into more rigid fibril structures. A solution containing wild type FUS liquid droplets will eventually convert into fibrous structures: mutations such as G156E greatly accelerate the rate of fibril conversion. Recent work by solid state nuclear magnetic resonance revealed an atomic resolution structure for fibrils formed by the FUS LCD (Protein Data Bank (PDB) code 5W3N; Murray, et al., 2017). This structure suggests that interactions between polar amino acids play a significant role in the stabilization of the FUS fibrils, but this has not been investigated experimentally.

Based on the FUS fibril structure, we hypothesize that there is an intramolecular hydrogen bond network that stabilizes FUS in fibril form. To investigate this, we recombinantly expressed and purified wild type FUS-LCD and used this material to prepare fibrils with the intent of probing the thermodynamic stability of this structure using calorimetry and a urea denaturation assay. PCR site directed mutagenesis will allow us to implement specific mutations in the primary amino acid sequence designed to disrupt the potential hydrogen bond network in the fibrils. We can then quantitatively determine the thermodynamic effect of the mutations, which will help explain why the protein is prone to fibril formation. Our work is a starting point for understanding ALS-linked mutations in the FUS protein that favor the formation of fibrils.

Methods

FUS Expression

The phis FUS expression plasmid was transformed into Invitrogen BL21(DE3) cells. 2 µl of the plasmid DNA was

incubated with 50 µl of cells on ice for 15 min. The cells were heat shocked at 42 °C for 60 s and then incubated on ice for 2 min, 500 µL of Luria-Broth (LB) media was added, and the cells were incubated at 37 °C with 220 RPM shaking for 30 min. 25 µl and 100 µl of the cells were spread onto LB/agar plates containing 100 µg/ml ampicillin and placed in the incubator overnight. The next afternoon, a 50 ml LB culture with 100 µg/ml ampicillin was inoculated from a single colony on the plate and incubated overnight at 37 °C with 220 RPM shaking.

A 10-fold dilution of the overnight was prepared with LB media. OD_{600} values measured with a 1 cm pathlength was 6.62. 7.5 µl of the overnight culture was added to 1 L of LB media containing 100 µg/ml ampicillin in a 4 L baffled flask. The 1 L culture was incubated at 37 °C with 210 RPM shaking. 3 hr later, the OD_{600} measured on an undiluted sample of the culture at a path-length of 1 cm was 0.676. Expression of the FUS-LCD protein was initiated by adding 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG), and the culture was incubated at 37 °C with 210 RPM shaking for three hr. The OD_{600} measured on a 10-fold LB media dilution sample of the culture at 1 cm path-length was 2.96. The cells were harvested by centrifugation at 6,000 g for 15 min. The cell pellet was split in half, scraped into two 50 ml tubes and stored at – 80 °C.

FUS Purification (Ni^{2+} Affinity Chromatography)

The purification of the FUS-LCD utilized three different buffers: Equilibrium buffer (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 500 mM NaCl, 6 M urea), Wash buffer (20 mM HEPES, 500 mM NaCl, 20 mM imidazole, 6 M urea) and Elution buffer (20 mM HEPES, 500 mM NaCl, 200 mM imidazole, 6 M urea).

The cell pellet was removed from the − 80 °C freezer and resuspended in 30 ml lysis buffer (6 M guanidinium hydrochloride, 50 mM Tris pH 7.5, 500 mM sodium chloride, 1% v/v Triton X- 100) with 0.25 mg/ml hen egg white lysozyme. Three tablets of Thermo Scientific Pierce Protease Inhibitor, EDTA-Free were added to the lysis solution and the mixture was tip sonicated using a Branson Sonifier 250 with a ¼ inch micro tip for 0.3 s at an amplitude of 30% with 3.0 s rest between pulses for a total on time of 1 min. The lysed cells were centrifuged at 75,600 RCF for 25 min.

The supernatant was loaded onto a Bio-Rad 5 ml Bio-scale Mini Nuvia IMAC Ni Charged column and equilibrated in equilibration buffer at a flow rate of 2 ml/min. The column was washed with 50 ml of equilibration buffer followed by 150 ml of wash buffer at a flow rate of 2 ml/min, each time until the A_{280} absorbance returned to the baseline value. The protein was eluted with a gradient from 100% wash buffer to 100% elution buffer at a flow rate of 1 ml/min followed by 20 ml of 100% elution buffer.

SDS-PAGE gel samples were made from all fractions of the purification using 4x SDS loading Buffer. The volume of each fraction added to the gel sample was normalized to the wash buffer fraction volume. The gel samples were heated for 10 min at 90 °C and centrifuged before loading onto the SDS PAGE gel.

The SDS PAGE gel had a 12% resolving gel and 4% stacking gel. The gel was run in 1X SDS buffer (1:9 dilution of 10x tris-glycine-SDS running buffer with ultra-pure water) at 80 V for 15 min and 35 min at 200 V. The gel was stained using Coomassie Blue G-250 methanol-acetic acid stain and imaged using a Bio-Rad Chemi-Doc MP Imaging System. A Fisher

BioReagents EZ-Run Rec Protein Ladder was used for a molecular weight marker.

FUS Size Exclusion Chromatography

A 24 ml Bio-Rad SEC 650 column was used for size exclusion chromatography. The column was equilibrated and run using filtered and degassed buffer containing 20 mM HEPES, 500 mM sodium chloride and 6 M urea. 0.4 µl of the 2.66 mg/ml elution sample was injected. The column was run at a flowrate of 1.0 ml/ min and 2 ml fractions were collected. SDS-PAGE gel samples were made from wash fractions exhibiting non-zero absorbance at 280 nm. The remaining procedure for the SDS-PAGE gel was the same as used for the Ni^{2+} affinity chromatography.

Fibril Formation and Imaging

3 ml of purified FUS-LCD protein at a concentration of 2.52 mg/ml was dialyzed against 20 mM HEPES buffer. 1.4 ml was recovered from dialysis and mixed with a 200 µl solution containing 0.26 mg FUS-LCD fibril seeds. The seeds were prepared by tip sonication with a $1/8^{th}$ inch microtip using a Branson Sonifier 250 operated at an amplitude of 10% with 0.1 s on time and 1 s off time for a total of 10 min. The purified FUS-LCD protein and seeds were incubated on the bench top for two days.

The 5 µl of the FUS-LCD sequence solution was deposited on a glow discharged 200 mesh carbon copper grid and incubated for 2 min, followed by two 5 µl washes with water for 10 s and staining with 5 µl of 3% uranyl acetate. A JEOL JEM-1230 electron microscope was used to record negatively stained TEM images of the FUS fibrils.

PCR Site Directed Mutagenesis

The PCR reactions contained: 30 µl autoclaved H_2O, 1.5 µl DMSO, 2 µl FUS template DNA (87 ng/µl), 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM forward and reverse T47A primer (sequence provided in Fig. 1 of Supplemental Information), 10 µl of 5x Phusion GC Buffer and 0.5 µl Phusion DNA Polymerase. The protocol on the thermocycler was one cycle of 98 °C for 30 s, one cycle of 98 °C for 7 s, one cycle of 70 °C for 20 s, then 31 cycles of 72 °C for 2 min 42 s, followed by one cycle of 72 °C for 8 min. The reaction was then kept at 4 °C overnight. The next day, 1 µl of the New England Bio-Lab (NEB) DpnI enzyme was added to the reaction tube and incubated for 1 hr at 37 °C. The PCR product was transformed into NEB DH5α chemically competent *E. coli* cells. 4 µl of the PCR product was incubated with 50 µl of cells on ice for 30 min. The cells were heat shocked at 42 °C for 30 s and then incubated on ice for 5 min, 950 µl of SOC media was added, and the cells were incubated at 37 °C with 250 RPM shaking for 1 hr. Next, two 10-fold serial dilutions in SOC media were performed. 100 µl of the neat and diluted cells were spread onto their respective LB/agar plates containing 100 µg/ml ampicillin and placed in the incubator overnight. Overnights of the colonies were purified using a QIAprep Spin miniprep kit. The purified plasmid was sent to GeneWiz for sequencing.

To troubleshoot the PCR reaction, the conditions for the PCR reaction were altered. Three different conditions were tested. For these tests, the FUS template DNA (87 ng/µl) volume was reduced to 1 µl. The first condition increased the DMSO volume to 3 µl. The second condition used the Phusion HF Buffer. The third condition used 2.5 µl of .5mM magnesium chloride. The thermocycler protocol followed a similar protocol to the previous one but replaced the 1 cycle of 70 °C for 20 s with one cycle of 68 °C for 20 s.

Three more PCR conditions were tested, each with a different starting amount of the FUS plasmid template (171 ng/µl). Each PCR reaction contained 30 µl autoclaved H_2O, 1.5 µl DMSO, 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM forward and reverse T47A primer, 10 µl of 5x Phusion GC Buffer and 0.5 µl Phusion DNA Polymerase. The amounts of FUS Template DNA varied from 2 µl, 4 µl and 6 µl. The thermocycler conditions followed the same protocol as the first one but the 31 cycles of 72 °C for 2 min 42 s were changed to 31 cycles of 72 °C for 3 min.

Results

Expression and Purification of FUS

The *E. coli* recombinant protein expression system produced milligram quantities of highly pure wild type FUS protein with an N-terminal 6X His tag, having the amino acid sequence shown in Fig. 1A. The chromatogram from the single-step nickel affinity purification procedure is shown in Fig. 1B. The His-tagged FUS protein has a strong affinity for the Ni^{2+} resin column. The imidazole used in the purification buffer out-competes the FUS protein for binding to Ni^{2+} resin. A moderate 20 mM imidazole was sufficient to remove most impurities. Elution with an imidazole gradient from 20mM to 200mM revealed a single peak, suggesting a single molecular species interacts strongly with the Ni^{2+} resin. The SDS PAGE analysis of the purification fractions shown in Fig. 1C shows a strong band near a molecular weight of 70 kDa. There are faint bands consistent with molecular weights of 45 and 30 kDa also present in the elution fractions. Our result is consistent with a previous study of this FUS construct (Murray, et al., 2017). Our procedure yields 5.5 mg of His-tagged FUS-LCD from 3.9 g of wet *E. coli* cells at a purity greater than 90%.

```
ASNDYTQQA  TQSYGAYPTQ PGQGYSQQSS QPYGQQSYSG YSQSTDTSGY GQSSYSSYGQ 60
SQNTGYGTQS TPQGYGSTGG YGSSQSSQSS YGQQSSYPGY GQQPAPSSTS GSYGSSSQSS 120
SYGQPQSGSY SQQPSYGGQQ QSYGQQQSYN PPQGYGQQNQ YNSSSGGGGG GGGGGNYGQD 180
QSSMSSGGGS GGGYGNQDQS GGGGSGGYGQ QDRQ 214
```

Figure 1A. Residues 2-214 of the human FUS-LCD sequence. The FUS protein used in experiments contains N-terminal His-tag with the sequence MSYYHHHHHHDYDIPTTENLYFQGAMPD.

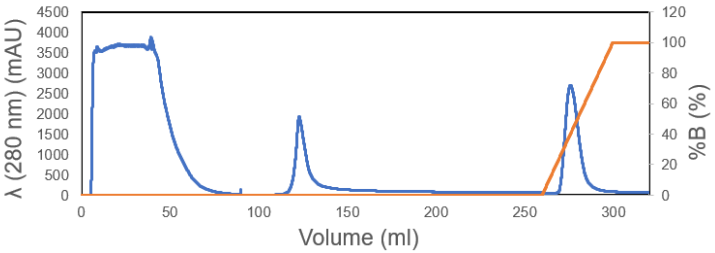

Figure 1B. The chromatogram obtained from the Ni^{2+} affinity chromatography purification of the FUS protein. The blue line is the A_{280} trace and the orange line is the percent buffer B.

The first and most broad peak in Figure 1B is the flow-through from the column loading, or the proteins that did not bind to the column. The second peak is the impurities that are washed out with 20 mM imidazole, or the proteins that bind to the column with moderate affinity. The third peak is the elution fractions, or the proteins that interact strongly with the column. Gel samples were prepared for the flow-through, wash, and elution fractions with significant absorbance at 280 nm.

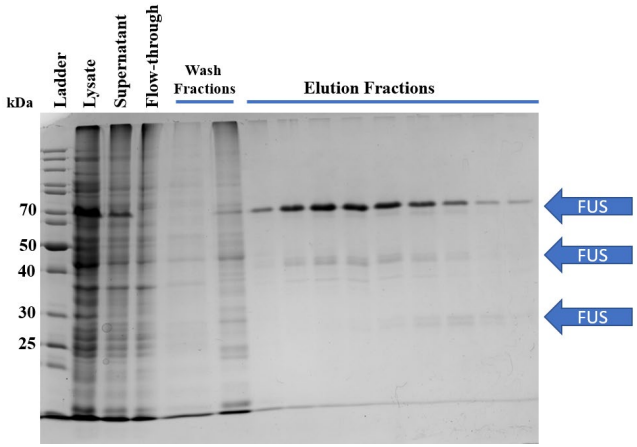

Figure 1C. The SDS PAGE gel of the fractions collected from the Ni^{2+} Affinity Chromatography.

The FUS protein has a molecular weight of 25 kDa. For the elution fractions, the most prominent bands are observed around 75 kDa, 50 kDa and 25 kDa, indicated by the 3 blue arrows in Figure 1C. The SDS-PAGE gel indicates that the FUS protein in the elution fractions exists in trimer, dimer and monomeric forms.

A gel filtration, or size exclusion chromatography (SEC) column was run on the FUS protein from the Ni^{2+} purification. The SEC chromatogram is shown in Fig. 2A. There is a single, relatively narrow, peak in the chromatogram. The location of this peak it consistent with a molecular weight species of approximately 50-100 kDa based on the manufacturer`s documentation. There is no evidence of multiple molecular weight species in the chromatogram. The fractions from this peak were used to prepare gel samples for an SDS PAGE. The bands observed in the SDS PAGE gel (Fig. 2B) are similar to the 75kDa and 50 kDa bands observed in the elution fractions in the gel in Fig. 1C.

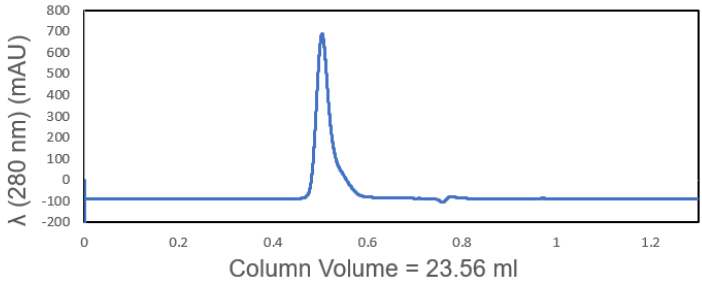

Figure 2A. The chromatogram from the size exclusion chromatography of the Ni^{2+} purified FUS protein. The blue line is the A_{280} trace.

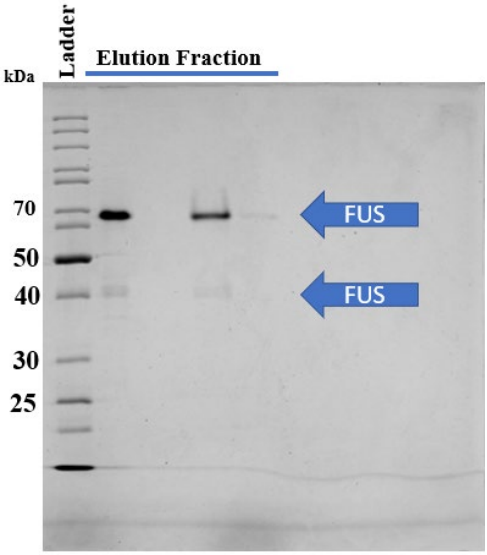

Figure 2B. An SDS PAGE gel of the size exclusion chromatography of the FUS-LCD protein.

Fibril Formation by FUS

A seeding procedure was used to form fibrils of the FUS protein. Elution fractions 2, 3 and 4 were used to create the fibrils. Urea was removed from the fractions by dialysis. Fibril seeds were prepared by tip sonicating preformed FUS fibrils and mixed with the dialyzed proteins. After 2 days, straight, unbranched fibrils formed. A transmission electron microscope (TEM) image of negatively stained FUS-LCD fibrils is shown in Fig. 3 and is consistent with previously published images of FUS fibrils (Murray, et al., 2017).

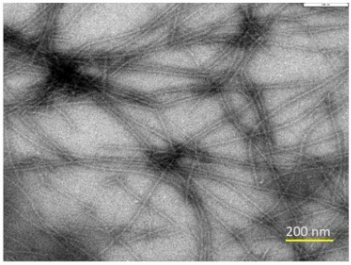

Figure 3. Negatively stained TEM image of FUS fibrils obtained at 25,000 magnification.

PCR Site Directed Mutagenesis on FUS

Residues Serine 77, Threonine 47, Serine 48 and Threonine 71 were identified as possible hydrogen bond pairs in the published structure of the FUS fibrils (Fig. 4A, PDB code 5W3N). PCR conditions for the site directed mutagenesis procedure were difficult to optimize. Most PCR reactions returned the nucleotide sequence corresponding to the wild type protein sequence shown in Fig. 1A. PCR reactions that did produce the desired mutation resulted in mis-alignment of the remaining amino acid sequence of the FUS-LCD sequence (Fig. 4B).

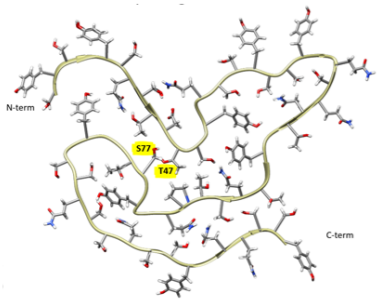

Figure 4A. The repeating units in the FUS LCD fibril structure (PDB code: 5W3N). The fibril core contains many polar residues in close proximity capable of hydrogen bonding. The putative pair of S77 and T47 are highlighted. The image was created with the UCSF Chimera program.

```
PASNDYTQQA TQSYGAYPTQ PGQGYSQQSS QPYGQQSYSG YSQSTDASGY GPVHGRFRLC  60
QSTDASGYQS TDASGYGQST DASGYGQSTD ASGYGQSTDA SGYGQSTDAS GVHGRFRLWP 120
SPRTLQAMAQ STDASGYGQS TDXSGYGQSS YSSYGQSQNT GYGTQSTPQG YGSTGG    176
```

Figure 4B. Sequencing results of the FUS-LCD sequence PCR product with the desired T47A mutation, highlighted in yellow and a misalignment of the remaining sequence starting at the underlined residue 52.

Discussion

In this paper, we have shown that 5.5 mg of pure His-tagged FUS was produced from our *E. coli* expression system and Ni^{2+} affinity chromatography purification protocol. This was confirmed through an SDS PAGE analysis. The FUS protein shows up in the SDS-PAGE gel at molecular weights of approximately 75, 50 and 25 kDa, suggesting that the FUS-LCD protein is capable of aggregation or assembly even in 6M urea. A SEC analysis confirmed that the elution fractions contained pure FUS in several oligomeric states. The protein obtained from our procedure was used to prepare fibrils that gave rise to TEM images similar to published TEM images of the FUS-LCD.

The bands at 75 and 50 kDa in the SDS PAGE gels for both the Ni^{2+} affinity chromatography and the SEC suggest that the FUS-LCD can aggregate or self-assemble in urea. These bands are consistent with trimer and dimer assemblies of the FUS-LCD. It is surprising that an intrinsically disordered protein like

the FUS-LCD still assembles into well-defined oligomeric states in a strong denaturing agent. The lack of higher order oligomeric states indicates that this assembly is well-defined and not amorphous aggregation of the protein.

The FUS-LCD is most devoid of hydrophobic residues. In the context of the protein transitioning from a state of intrinsic disorder to a well folded-protein fibril, the absence of hydrophobic residues suggests a novel balance of forces that rely mainly on hydrogen bonding interactions. How urea interferes with this process warrants further investigation. PCR site directed mutagenesis was used to introduce T47A mutation intended to disrupt a putative hydrogen bond in the FUS-LCD fibril core. Our efforts led to the T47A mutation in the primary amino acid sequence of FUS-LCD, but there was a misalignment of the remaining amino acids in the sequence This was most likely due to the high GC content and repetitive sequences of the FUS-LCD. Alternative methods of site directed mutagenesis or a more comprehensive screen of primers and PCR conditions are required to affect mutations in the FUS-LCD sequence.

Our future efforts will quantitatively probe the thermodynamic stability of the FUS-LCD fibrils using calorimetry and a urea denaturation assay. Additional efforts will focus on obtaining mutations in the primary amino acid sequence of the FUS-LCD. Together, these efforts will elucidate the role of hydrogen bond networks in the fibril formation of the FUS-LCD protein. These measurements will then be extended to understand how ALS-linked mutations in the FUS-LCD favor the formation of pathogenic FUS fibrils.

Supplemental Information

5'- CAG TCC ACG GAC ACT TCA GGC TAT GGC $_{5438}$ - 3'

5'- GCC ATA GCC TGA AGC GTC CGT GGA CTG - 3'

5'- GCC AGT CCA CGG ACG CTT CAG GCT ATG - 3'

Supplemental Figure 1. The first sequence is the template of FUS-LCD with nucleotide bases 5412-5438. The 2nd sequence is the T47A forward primer and the 3rd sequence is the T47A reverse primer. Each primer has one base modification (in red).

As shown if Supplemental Fig. 1, the codon ACT in the FUS-LCD codes for Threonine. The T47A forward primer is complimentary to the FUS-LCD forward template (5'-3') with one base modification shown in red. The reverse primer is complimentary to the FUS-LCD reverse template (3'-5') with one base modification. The base modification codes for Alanine.

References

(1) Calabretta, S.; Richard, S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends in Biochemical Sciences*. 2015. https://doi.org/10.1016/j.tibs.2015.08.012.

(2) Dunker, A. K.; Babu, M. M.; Barbar, E.; Blackledge, M.; Bondos, S. E.; Dosztányi, Z.; Dyson, H. J.; Forman-Kay, J.; Fuxreiter, M.; Gsponer, J.; et al. What's in a Name? Why These Proteins Are Intrinsically Disordered. *Intrinsically Disord. Proteins* **2013**. https://doi.org/10.4161/idp.24157.

(3) Dyson, H. J.; Wright, P. E. Intrinsically Unstructured Proteins and Their Functions. *Nature Reviews Molecular Cell Biology*. 2005. https://doi.org/10.1038/nrm1589.

(4) Harrison, A. F.; Shorter, J. RNA-Binding Proteins with Prion-like Domains in Health and Disease. *Biochemical Journal*. 2017. https://doi.org/10.1042/BCJ20160499.

(5) Kim, H. J.; Kim, N. C.; Wang, Y. D.; Scarborough, E. A.; Moore, J.; Diaz, Z.; MacLea, K. S.; Freibaum, B.; Li, S.; Molliex, A.; et al. Mutations in Prion-like Domains in HnRNPA2B1 and HnRNPA1 Cause Multisystem Proteinopathy and ALS. *Nature* **2013**. https://doi.org/10.1038/nature11922.

(6) Kwiatkowski TJ Jr1, Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T, Valdmanis P, Rouleau GA, Hosler BA, Cortelli P, de Jong PJ, Yoshinaga Y, Haines JL, Pericak-Vance MA, Yan J, Ticozzi N, Siddique T, B. R. J. Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science (80-.). 323* (5918), 1205–1208.

(7) 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**. https://doi.org/10.1016/j.cell.2017.08.048.

(8) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M.; et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**. https://doi.org/10.1016/j.cell.2015.07.047.