

CHEMICAL FOOTPRINTING OF POLYMERIC STRUCTURE OF HNRNPA2  
LOW COMPLEXITY DOMAIN

APPROVED BY SUPERVISORY COMMITTEE

---

Steven McKnight, Ph.D.

---

Deepak Nijhawan, M.D, Ph.D.

---

Michael Rosen, Ph.D.

---

Steven Kliewer, Ph.D.

---

Hongtao Yu, Ph.D.

CHEMICAL FOOTPRINTING OF POLYMERIC STRUCTURE OF HNRNPA2  
LOW COMPLEXITY DOMAIN

by

SIHENG XIANG

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2016

CHEMICAL FOOTPRINTING OF POLYMERIC STRUCTURE OF HNRNPA2  
LOW COMPLEXITY DOMAIN

SIHENG XIANG, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2016

Supervising Professor: STEVEN L. MCKNIGHT, Ph.D.

Many DNA and RNA regulatory proteins contain polypeptide domains that are unstructured when analyzed in cell lysates. These domains are typified by an over-representation of a limited number of amino acids and have been termed prion-like, intrinsically disordered or low complexity domains. These low complexity sequences have been shown to induce phase transition in low salt buffer. When incubated at high concentration, certain of these low complexity domains polymerize into labile, amyloid-like fibers. I developed a chemical footprinting method to probe solvent accessible residues in the low complexity domain

polymers. By acetylating protein side chains with N-acetylimidazole, and comparing the acetylation in native and denatured conformation by use of SILAC mass spectrometry, I generated an NAI footprint for hnRNPA2 polymers. I deployed this footprinting technique to probe the structure of the native hnRNPA2 protein present in isolated nuclei, and offered evidence that its low complexity domain exists in a similar conformation as that described for recombinant polymers of the protein. To study the structure of the low complexity sequence in liquid-like droplets, I systematically mutated individual tyrosine or phenylalanine residues to serine, assayed the ratio of these mutants that partitioned into the droplet phase, and compared the results with their abilities to grow polymeric fibers from wild-type seeds. The same region which contained mutations impeding fiber growth were found to display decreased partitioning into liquid-like droplets. Additionally, the NAI footprint of hnRNPA2 in these liquid-like droplets appeared to be similar to the footprint found in fibers. These observations suggest that the hnRNPA2 low complexity domain adopts a similar structure in amyloid-like fibers and liquid-like droplets. Combining these results, my studies favor the perspective that cross- $\beta$  polymerization commonly drives the formation of hydrogels, the retention of low complexity domains trapped by hydrogels, the formation of liquid-like droplets, the partitioning of low complexity domains into existing liquid-like droplets, and the formation and maturation of RNA granules. In other words, my results provide evidence that the involvement of low complexity domains in the formation of RNA granules, liquid-like droplets and hydrogels all rely on one in the same phenomenon – cross- $\beta$  polymerization.

## TABLE OF CONTENTS

TITLE .....	i
TITLE PAGE .....	ii
ABSTRACT .....	iii
TABLE OF CONTENTS .....	v
PRIOR PUBLICATIONS .....	ix
LIST OF FIGURES .....	x
LIST OF TABLES .....	xiii
LIST OF APPENDICES .....	xiv
LIST OF ABBREVIATIONS.....	xv
CHAPTER ONE Introduction .....	1
Low Complexity (LC) Sequences and Non-membrane Bound Sub-cellular Organelles.....	1
RNA Granules and Nuclear Bodies Are non-Membrane Bound Sub-Cellular Structures .....	1
Transcription Factors Contain LC Sequences .....	2
RNA Regulatory Proteins Contain G/S-Y-G/S Repeats in Their LC Domain.....	4
LC Sequences Form Labile Amyloid-like Fibers .....	10
LC Domain of the FUS Protein Forms Hydrogels .....	10
Hydrogels are Composed of Labile Amyloid-like Fibers .....	12
Hydrogel Retention as an Assay of Fiber Polymerization .....	14

Serine:Arginine Domains of pre-mRNA Splicing Factors Bind hnRNPA2	
Polymers .....	15
Correlative Relationship between Fiber Formation and Biological Activities of LC	
Sequences .....	25
Tyrosine Residues in FUS LC Domain Critical for Stress Granule Recruitment .....	25
Correlation of Hydrogel Retention and Transcriptional Activation of FUS LC	
Domain Mutants.....	26
CHAPTER TWO The LC Domain of hnRNPA2 Adopts Similar Conformations in	
Hydrogel Polymers, Liquid-like Droplets and Nuclei .....	32
Development of a Chemical Footprinting Method .....	33
Optimizing Reaction Conditions for Protein Acetylation by NAI .....	34
NAI Footprint of GST Enzyme .....	36
NAI Footprint of Nuclear PARP1 .....	40
NAI Footprint of hnRNPA2 LC Domain in Recombinant Fibers .....	51
Determination of the NAI Footprint of Recombinant hnRNPA2 Fibers .....	51
Mutations in the NAI-protected Region of the hnRNPA2 LC Domain Impede	
Hydrogel Binding.....	55
Human Genetics Reveals Causative Mutations of Familial Neurodegenerative	
Diseases in the NAI-protected Region of the hnRNPA2 LC Domain.....	57
Correlative relationship of hnRNPA2 Footprints between Recombinant and	
Nuclear Forms of the Protein .....	70
NAI Footprints of Endogenous hnRNPA2 from Mammalian Cell Nuclei .....	70

Co-expression of hnRNPA2 with Peptidyl-prolyl cis-trans Isomerase Causes Tyrosine 324 to Become NAI-accessible in Recombinant Polymers .....	72
PTB:hnRNPA2 LC Forms Liquid-like droplet by LC domain polymerization .....	79
Mutations in the LC Domain of hnRNPA2 Act Correlatively on Hydrogel Binding and Partitioning into Liquid-like Droplets .....	80
Liquid-like Droplets Display the NAI Footprint Found in Hydrogel Polymers and Nuclear hnRNPA2 .....	82
CHAPTER THREE Discussions and Conclusions.....	91
Discussions .....	91
Conclusions.....	100
CHAPTER FOUR Methodology .....	107
Materials .....	107
Experimental Procedure.....	107
Expression and Purification of mCherry or eGFP Fusion Proteins.....	107
Expression and Purification of His <sub>6</sub> -tagged Proteins. ....	108
Purification of Glutathione S-Transferase (GST). ....	109
Hydrogel-binding Assays.....	109
Heavy Protein Preparation.....	110
Acetylation of Recombinant Proteins.....	111
Measuring Acetylation of Tyrosine Side Chains. ....	112
Mass Spectrometry Analysis and Data Processing. ....	112
Calculating Solvent Accessible Surface.....	114

293T Cell Culture and Nuclei Preparation. ....	114
Acetylation of 293T Nuclei. ....	115
2D-HPLC Mass Spectrometry.....	116
Identification of Hydrogel Binding Proteins from Heavy Cell Lysates .....	117
Purification of MBP:PTB:hnRNP LC Fusion Protein. ....	118
Liquid-like Droplet Formation by MBP:PTB:hnRNPA2 LC and NAI Footprinting.....	119
eGFP:hnRNP A2 Recruitment by Liquid-like Droplet. ....	119
Localization of PPIA to Stress Granules. ....	120
Liquid-Like Droplets Formation by His <sub>6</sub> -tagged FUS LC.....	121
Transmission Electron Microscopy of mCherry:hnRNPA2 Fibers. ....	121
Immunoelectron Microscopy of SR Bound FUS Fibers. ....	122
APPENDIX A Heavy Medium for Bacterial Cultures .....	124
APPENDIX B Heavy DMEM Medium for Mammalian Cell Cultures .....	125
APPENDIX C Mass Spectrum of Representative Peptides of GST Enzyme.....	126
APPENDIX D Mass Spectrum of Representative Peptides of hnRNPA2 .....	132
Bibliography .....	134

## PRIOR PUBLICATIONS

- Xiang, S.**, Kato, M., Wu, C. L., Lin, Y., Ding, M., Zhang, Y., Yu, Y. & McKnight, S. L. (2015). The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets and Nuclei. *Cell* **163**, 829–839.
- Kwon, I., **Xiang, S.**, Kato, M., Wu, C. L., Theodoropoulos, P., Wang, T., Kim, J., Yun, Y., Xie, Y. and McKnight, S.L. (2014). Poly-dipeptides encoded by the C9ORF72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**:1139-1145.
- Kwon, I., Kato, M., **Xiang, S.**, Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-regulated Binding of RNA Polymerase II to Fibrous Polymers of Low Complexity Domains. *Cell* **155**:1049-1060.

## LIST OF FIGURES

FIGURE 1-1. RNA Granules.....	7
FIGURE 1-2. Nuclear Bodies.....	8
FIGURE 1-3. Low Complexity Sequences of FUS and hnRNPA2.....	9
FIGURE 1-4. Hydrogels Bind eGFP-Tagged LC Sequences of RNA Regulatory Proteins.....	18
FIGURE 1-5. Hydrogel Retention of RNA Requires both LC And RNA-Binding Domains .....	19
FIGURE 1-6. Electron Microscopy and X-ray Diffraction Experiment of Hydrogels .....	20
FIGURE 1-7. Propensity of LC Domains to Form $\beta$ Structure .....	21
FIGURE 1-8. Amyloid-like Fibers Formed by LC Domains Are Sensitive to SDS .....	22
FIGURE 1-9. GFP:SR Reversibly Binds hnRNPA2 Hydrogels .....	24
FIGURE 1-10. Mutations of FUS LC Domain Affect Stress Granule Recruitment .....	29
FIGURE 1-11. Correlation of Hydrogel Retention and Transcriptional Activation of FUS LC Domain Mutants .....	31
Figure 2-1. Stability of NAI Compound in Buffers.....	43
Figure 2-2. Acetylation of BSA .....	44
Figure 2-3. Differing Patterns of Acetylation of Folded and Denatured Samples of Glutathione-S-transferase Mediated by N-acetylimidazole .....	45
Figure 2-4. NAI Footprint of GST Enzyme Correlates with Solvent Accessible Surface.....	46

Figure 2-5. Establishment of an NAI Footprint for the Poly-ADP-ribose Polymerase Enzyme within Nuclei Isolated from Mammalian Cells .....	48
Figure 2-6. Chromatographic Separation and Acetylation Site Identification of hnRNPA2 LC .....	60
Figure 2-7. MS1 Spectra of Native Peptides from Acetylated hnRNPA2.....	61
Figure 2-8. Quantification of Native Peptides from SILAC Samples .....	62
Figure 2-9. Mass Spectra of Peptides Acetylated on the Same Serine Residue .....	63
Figure 2-10. Footprint of NAI-mediated Acetylation of Recombinant hnRNPA2 Polymeric Fibers .....	65
Figure 2-11. Correlative Relationship between Binding of Mutated Variants of the LC Domain of hnRNPA2 to NAI Accessibilities .....	67
Figure 2-12. Genetic Mutations Associate with Familial MSP .....	68
Figure 2-13. NAI Footprints of Recombinant and Nuclear hnRNPA2 LC .....	75
Figure 2-14. PPIA Binds hnRNPA2 and Localizes to Stress Granules.....	77
Figure 2-15. Correlative Relationship between the NAI Footprint of Endogenous hnRNPA2, Recombinant Protein, and Recombinant Protein Co- Expressed with PPIA.....	78
Figure 2-16. Liquid-Like Droplet Formation Driven by LC Domains.....	86
Figure 2-17. Correlative Relationship between Liquid-Like Droplets Partitioning and Hydrogel Retention .....	87
Figure 2-18. Liquid-Like Droplets Display the Same NAI Footprint as Found in Polymeric Fibers .....	89

Figure 2-19. Graphical Representation of Conversion of Soluble  
MBP:PTB:hnRNP LC Fusion Protein into Liquid-like Droplet  
State..... 90

Figure 3-1. Liquid-like Droplets Formed by His<sub>6</sub>-tagged FUS LC ..... 99

Figure 3-2. NAI Footprint Probes Structure of hnRNPA2 LC ..... 106

## LIST OF TABLES

TABLE 2-1 Footprint Deduced by NAI-Mediated Acetylation of GST .....	49
TABLE 2-2 Footprint Deduced by NAI-Mediated Acetylation of PARP .....	50
TABLE 2-3 Footprint Deduced by NAI-Mediated Acetylation of hnRNPA2 .....	69

## LIST OF APPENDICES

APPENDIX A. Heavy Medium for Bacterial Cultures .....	124
APPENDIX B. Heavy DMEM Medium for Mammalian Cell Cultures .....	125
APPENDIX C. Mass Spectrum of Representative Peptides of GST Enzyme.....	126
APPENDIX D. Mass Spectrum of Representative Peptides of hnRNPA2 .....	132

## LIST OF ABBREVIATIONS

ALS	Amyotrophic lateral sclerosis
ATP	Adenosine triphosphate
BME	2-Mercaptoethanol
BSA	Bovine serum albumin
BuGZ	BUB3-interacting and GLEBS motif-containing protein ZNF207
CaMKII	Ca <sup>2+</sup> /calmodulin-dependent protein kinase II
CDK8	Cyclin-dependent kinase 8
CFP	Cyan fluorescent protein
CID	Collision-induced dissociation
CIRBP	Cold inducible RNA binding protein
CLK	Cdc2-like kinase 1
CPEB2	Cytoplasmic polyadenylation element binding protein 2
CTD	C-terminal domain of the largest subunit of PolIII
CV	Column volume
Cy5.5	Cyanine 5.5
DAPI	4',6-diamidino-2-phenylindole
DCP2	mRNA-decapping enzyme 2
DDX	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide
DMEM	(Dulbecco's modified Eagle medium
DMSO	Dimethyl sulfoxide

DNA-PK	DNA-dependent protein kinase
DTT	Dithiothreitol
DYRK	Dual specificity tyrosine-phosphorylation-regulated kinase
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EGTA	Ethylene glycol tetraacetic acid
eIF4GII	Eukaryotic translation initiation factor 4G II
EWS	Ewing sarcoma
FET	FUS/EWS/TAF15
FMRP	Fragile X mental retardation protein
FRAP	Fluorescence recovery after photobleaching
FTD	fronto- temporal dementia
FUS	Fused in sarcoma
GS4B	Glutathione Sepharose 4B
GST	Glutathione S-transferase
HEPES	HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hnRNP	Heterogeneous nuclear ribonucleoparticule
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
KH Domain	K Homology (KH) domain
LAF-1	Lethal and feminized-1
LB	Lysogeny broth
LC	Low complexity

Lsm4	U6 snRNA-associated Sm-like protein LSm4
m/z	Mass over charge ratio
MBK	Minibrain kinase
MBP	Maltose-binding protein
MEF	Mouse embryo fibroblast
MEG	Maternally expressed gene
MS	Mass spectrometry
MS1	The first stage of MS (for precursor ions)
MS2	The second stage of MS (for fragmented ions)
MSP	Multisystem proteinopathy
NAI	N-acetylimidazole
Nck	Non-catalytic region of tyrosine kinase adaptor protein 1
Ni-NTA	Nickel charged nitrilotriacetic acid (NTA)
NLS	Nuclear localization sequence
NMR	Nuclear magnetic resonance
N-WASP	Neural Wiskott-Aldrich syndrome protein
P Body	Processing body
PAR	Poly (ADP-ribose)
PARP1	Poly (ADP-ribose) polymerase
PBS	Phosphate-buffered saline
PDB	Paget's disease of the bone
PDB ID	Protein data bank identifier

PGL-1	P granule abnormality protein 1
PML Body	Promyelocytic leukemia (PML) body
PMSF	Phenylmethylsulfonyl fluoride
PolII	RNA polymerase II
PP2A	Protein phosphatase 2
PPIA	Peptidyl-prolyl cis-trans isomerase A
PRM	Proline-rich motifs
PRP4	Pre-mRNA processing factor 4
PSKH1	Protein serine kinase H1
PTB	Polypyrimidine tract-binding protein
Pub1	Nuclear and cytoplasmic polyadenylated RNA-binding protein
RBM14	RNA Binding Motif Protein 14
RP	Reverse phase
RRM	RNA recognition motif
SDD-AGE	Semi-denaturing detergent agarose gel electrophoresis
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH3	SRC homology 3
SILAC	Stable isotope labeling by amino acids in cell culture
SR	Serine:Arginine
SRSF2	Serine/arginine-rich splicing factor 2
TAF15	TATA-box binding protein associated factor

TBP	TATA-box binding protein
TDP43	TAR DNA-binding protein 43
TIA1	Nucleolysin TIA-1 isoform p40
TRIS	2-Amino-2-(hydroxymethyl)-1,3-propanediol
UAS	Upstream activating sequence
UTR	Untranslated region
VCP	Valosin-containing protein
YP	Yeast extract peptone

# **CHAPTER ONE**

## **Introduction**

### **LOW COMPLEXITY (LC) SEQUENCES AND NON-MEMBRANE BOUND SUB-CELLULAR ORGANELLES**

#### **RNA Granules and Nuclear Bodies Are non-Membrane Bound Sub-Cellular Structures**

In eukaryotic cells, RNA-binding regulatory proteins recruit their cognate RNAs into ribonucleoparticles called RNA granules. RNA granules are microscopically visible, non-membrane bound cytoplasmic organelles (Figure 1-1). In germ cells, P granules serve as storage organelles of mRNAs that specify germ line differentiation (Voronina et al., 2011). Asymmetric segregation of P granules also helps maintain uneven distribution of mRNAs in early embryos, which allows for the establishment of morphogen gradients required for cell fate determination (Strome and Wood, 1983). In somatic cells, environmental stresses such as heat shock or oxidative damage trigger translational arrest of a subset of mRNAs and sequester them into stress granules (Buchan and Parker, 2009). Processing bodies (P bodies) contain the basic enzymatic machinery for mRNA decay and miRNA repression (Buchan and Parker, 2009). RNA granules in neurons help to shuttle translationally silent mRNAs along dendrites for translation locally near synapse (Knowles and Kosik, 1997).

The cell nucleus contains many types of non-membrane bound sub-cellular organelles, which are also referred to as nuclear bodies (Figure 1-2) (Dundr and Misteli, 2010). These nuclear bodies represent molecularly distinct compartments believed to organize and

optimize various nuclear processes. Nuclear speckles, for example, store and process splicing factors (Lamond and Spector, 2003); transcription factors and RNA polymerase enzymes are also compartmented in discrete loci in the nucleoplasm, called transcription factories (Ghamari et al., 2013).

In this chapter, I will describe the role of low complexity (LC) sequences in RNA granules and nuclear bodies. I will then focus on the LC sequences of RNA regulatory proteins and review biochemical studies of these LC sequences.

### **Transcription Factors Contain LC Sequences**

Prototypic gene specific transcription factors contain both a DNA binding domain and a transcriptional activation domain. DNA binding domains recognize specific sequences via structurally ordered states, including zinc fingers, homeoboxes, helix-loop-helix domains and leucine zipper domains (Pabo and Sauer, 1992).

Unlike DNA binding domains, transcriptional activation domains are largely unstructured. Activation domains of certain transcription factors contain an over-representation of acidic amino acids (Hope et al., 1988). In the context of gene specific transcription factors, these structurally disordered domains have been termed “acid blobs” or “negative noodles” (Sigler, 1988), and other conceptualizations invoking biological function in the absence of folded protein structure.

Not all activation domains associated with gene specific transcription factors are acidic. Some are enriched in glutamine residues, others in proline residues (Triezenberg,

1995). Common properties, however, among the majority of activation domains is the over-representation of one or a small grouping of amino acids. Instead of utilizing a balanced proportion of all 20 amino acids, these domains are of low complexity in nature. Nucleic acids deploy a four-lettered code, proteins a 20 letter code. LC domains operate via the deployment of a highly skewed distribution of amino acids, and would appear to be much more DNA- and RNA-like in the nature of their code.

In addition to the transcriptional activation domains of gene-specific transcription factors, the basal transcription apparatus is also enriched in proteins with low complexity sequences. The transcription apparatus mainly consists of general transcription factors (GTFs), RNA polymerase II (Pol II), the mediator complex, chromosome remodeling factors and transcriptional activators (Lee and Young, 2000; Lorch and Kornberg, 2015). The largest subunit of the RNA polymerase enzyme, for example, contains a LC tail with 52 repeats of heptapeptide YSPTSPS (Corden et al., 1985), which is extensively modified in the transcription cycle (Buratowski, 2009; Phatnani and Greenleaf, 2006). Components of the mediator complex, chromosome remodeling factors and the TATA-box binding protein (TBP), all contain long stretch of polyglutamine sequences (Allen and Taatjes, 2015; Liu et al., 2006).

It has been well accepted that expansion of polyglutamine repeats leads to neurodegenerative diseases, such as spinal bulbar muscular atrophy, Huntington disease, several spinocerebellar ataxias and dentatorubral-pallidoluysian atrophy (Shao and Diamond, 2007; Zoghbi and Orr, 2000). By contrast, the physiological function of polyglutamine sequences found in many complexes associated with transcriptional regulation remain largely

unknown. The polyglutamine sequences in mediator subunits are present in the kinase module which contains CDK8, cyclin C, Med12, and Med13 (Borggreffe et al., 2002; Knuesel et al., 2009). Genetic experiments in yeasts, *Caenorhabditis elegans*, and *Drosophila melanogaster* showed that these subunits are each required for organism to survive, but dispensable for the viability of individual cells (Loncle et al., 2007; van de Peppel et al., 2005; Wang et al., 2004). In TBP, for example, forty consecutive glutamine residues sit in the middle of a vertebrate specific low complexity domain. The vertebrate specific polyglutamine domain of TBP is not required for fundamental cellular processes, but is necessary for maternal immunotolerance of implantation in mice (Hobbs et al., 2002; Schmidt et al., 2003).

### **RNA Regulatory Proteins Contain G/S-Y-G/S Repeats in Their LC Domain**

RNA regulatory proteins are composed of two functional domains. RNA binding domains are able to bind RNA via structurally ordered KH domains, RNA recognition motifs (RRM) or pumilio domains (Lunde et al., 2007). Most RNA regulatory proteins also contain polypeptide domains that lack structural order when purified from cellular lysates. Bioinformatic studies of the disordered domains of RNA regulatory proteins show that these LC domains are enriched with G/S-Y-G/S repeats (Kato et al., 2012). The LC domain of FUS (residues 2-214) contains 27 of such G/S-Y-G/S motifs, interspaced with stretches enriched in glycine and glutamine residues (Figure 1-3A). Similarly, the LC domain of

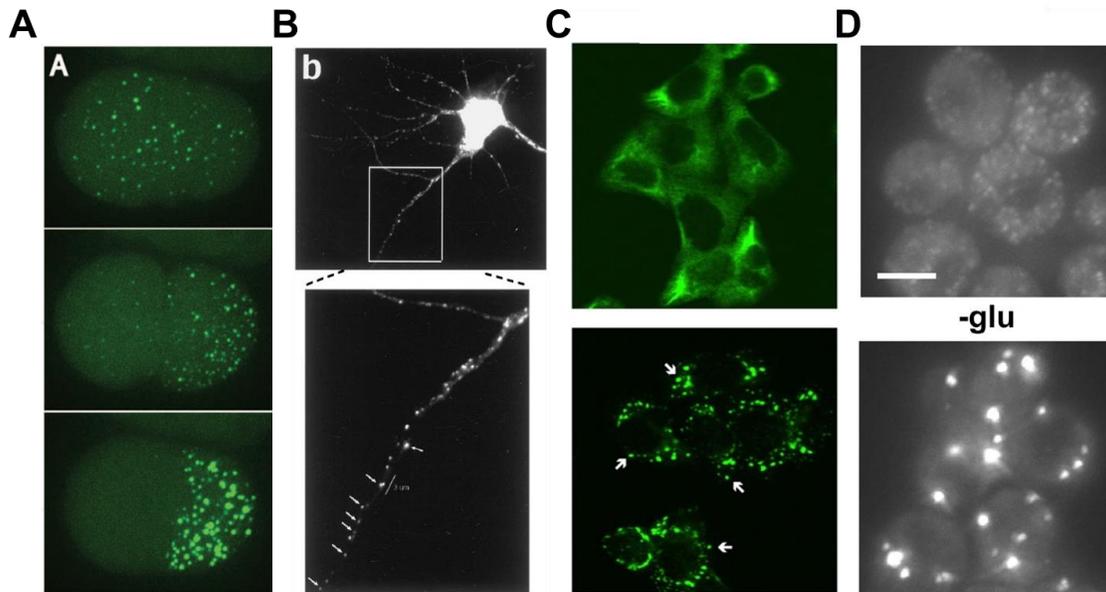
hnRNPA2 (residues 181-341) has 25 of such repeats, eight of which exchange phenylalanine for tyrosine (Figure 1-3B).

Compared with gene-specific transcription factors and their activation domains, less attention has been paid to the LC domains associated with RNA regulatory proteins. Some degree of attention has been focused on the LC domains associated with the FET family of RNA binding proteins, including fused in sarcoma (FUS), Ewing's sarcoma (EWS) and TAF15. The amino terminal LC domains of these three proteins can be translocated onto DNA binding domains as the causative event in many forms of human cancer (Riggi et al., 2007). In the context of these fusion proteins, the LC domains of the FET proteins function as potent transcriptional activation domains (Bertolotti et al., 1999; May et al., 1993; Sánchez-García and Rabbitts, 1994).

In cases of familial neurodegenerative diseases, human genetic studies have linked neuronal cell death to mutations in the genes encoding TDP-43, FUS and several different hnRNPs (Bentmann et al., 2013; Kim et al., 2013). Single amino acid substitution mutations have been found in the LC domains of the TDP-43, FUS and several hnRNP proteins (Bentmann et al., 2013; Kim et al., 2013). The causative mutations leading to neurodegenerative disease have been shown to trigger the aberrant aggregation of these proteins.

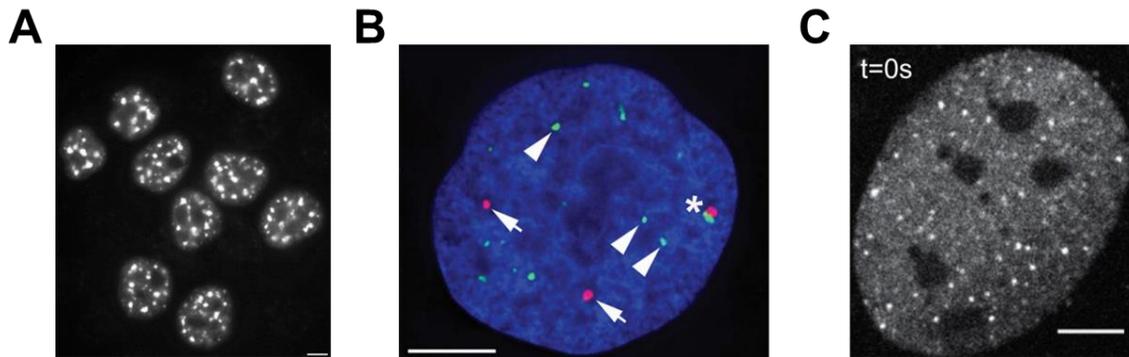
Little is known about the physiological role of these LC sequences. Previous studies have shown that the LC domain of FUS is required for its recruitment to stress granules. Mutations of tyrosine residues in the LC sequence to serine abolish stress granule recruitment (Han et al., 2012). The authors proposed that RNA regulatory proteins are recruited to RNA

granules by their LC sequences, which in turn recruit mRNAs to the granules via RNA binding domains.



**Figure 1-1. RNA Granules**

- (A) Time-lapse images of *Caenorhabditis elegans* zygotes expressing GFP::PGL-1 (P granules). The green colored puncta show embryonic P granules. From top to bottom, images are maximum projections of confocal Z-stacks spanning 8  $\mu\text{m}$  (~half of embryo depth) taken at mitotic exit, interphase and anaphase, respectively (Gallo et al., 2010).
- (B) GFP-labeled CaMKII $\alpha$  3'UTR RNA granules in a KCl-treated cell (Neuronal granules). Primary hippocampal neurons from pregnant embryonic day 18 (E18) Sprague Dawley rat were transiently transfected with two constructs to directly visualize neuronal RNA granules: the fluorescent eGFP-MS2-NLS under CMV promoter, and the RNA component with mouse CaMKII $\alpha$  3'UTR fused to eight copies of the small RNA hairpin-binding element under an RSV promoter. The transfected cells were treated 30 minutes with 10 mM KCl (Rook et al., 2000).
- (C) Stress granules in arsenite treated cells. Stress granules stained by antibody against G3BP1 in untreated HeLa cells and cells treated with 1 mM arsenite for 30 minutes at 37°C (Wehner et al., 2010).
- (D) Processing bodies in *Saccharomyces cerevisiae*. Localization of Dcp2:GFP during exponential growth in YP containing glucose (+glu) or after being deprived of glucose for 10 minutes while being aerated (-glu). Images are maximum intensity projections of a z series compilation of 15 images made using Image J (Decker et al., 2007).



**Figure 1-2. Nuclear Bodies**

- (A) Nuclear Speckles. HeLa cells showing splicing factors localized in a speckled pattern as well as being diffusely distributed throughout the nucleoplasm. Splicing factors were visualized by immunofluorescent microscopy using mouse monoclonal antibody against SC35 (Spector and Lamond, 2011).
- (B) Cajal bodies and promyelocytic leukemia protein (PML) bodies in human cells. Anti-coilin antibody labeled CBs were shown in red. PML bodies were stained using an anti-PML antibody, shown in green. DNA was stained with DAPI (blue) (Cioce and Lamond, 2005).
- (C) Transcription factories in primary MEF cells. Live cell image of primary MEF cells isolated from a CDK9-mCherry knock-in mouse was obtained by spinning-disk confocal microscope. Active transcription sites were detected as distinct fluorescent foci in nucleus (Ghamari et al., 2013).

<b>A FUS LC domain</b>		<b>B hnRNPA2 LC domain</b>	
5-DYTQQATQ		181-MQEVQSSRSRGGG	
SYG	<b>Amino acid composition</b>	NFGFGDSRGGGG	<b>Amino acid composition</b>
AYPTQPGQ	52 G Glycine	NFGPGPGS	71 G Glycine
GYSQQSSQ	50 S Serine	NFRGGSD	17 Y Tyrosine
PYGQQ	43 Q Glutamine	GYGSGR	16 N Asparagine
SYS	27 Y Tyrosine	GFGD	15 S Serine
GYSQSTDTS		GYN	
GYGQS	11 P Proline	GYGGPGGG	9 R Arginine
SYS	10 T Threonine	NFGGSP	9 P Proline
SYGQSQNT	6 N Asparagine	GYGGGRG	8 F Phenylalanine
GYGTQSTPQ	4 D Aspartic acid	GYGGGGP	5 D Aspartic acid
GYGSTG	3 A Alanine	GYGNQGG	5 Q Glutamine
GYGSSQSSQS	1 M Methionine	GYGG	3 M Methionine
SYGQQS		GYD	1 E Glutamic acid
SYP	0 E Glutamic acid	NYGGG	1 K Lysine
GYGQQPAPSSSTSG	0 K Lysine	NYGSG	1 V Valine
SYGSSSQSS	0 R Arginine	NYN	
SYGQPQSG	0 C Cysteine	DFG	0 A Alanine
SYSQQP	0 H Histidine	NYNQQPS	0 C Cysteine
SYGGQQQ	0 V Valine	NYGPMKSG	0 H Histidine
SYGQQQ	0 L Leucine	NFGGSRNMGG	0 I Isoleucine
SYNPPQ	0 I Isoleucine	PYGGG	0 L Leucine
GYGQQN	0 W Tryptophan	NYGPGGSGGSG	0 T Threonine
QYNSSSGGGGGGGGG	0 F Phenylalanine	GYGGRS	0 W Tryptophan
NYGQDQSSMSSGGGSGG		RY-341	
GYGNQDQSGGGGSG			
GYGQQ-211			

**Figure 1-3. Low Complexity Sequences of FUS and hnRNPA2**

Low complexity sequences of FUS (A) and hnRNPA2 (B). The LC sequences of the two RNA regulatory proteins were aligned by their G/S-Y/F motifs. The FUS LC sequence contains 27 G/S-Y-G/S motifs, while hnRNPA2 LC contain both G/S-Y-G/S and G/S-F-G/S motifs. Both LC sequences are compositionally biased. FUS LC has 52 glycine, 50 serine, 43 glutamine and 27 tyrosine residues. The four most abundant amino acids compose 83% of the whole LC sequence, and 10 out of 20 amino acids are absent in the sequence.

## LC SEQUENCES FORM LABILE AMYLOID-LIKE FIBERS

### LC Domain of the FUS Protein Forms Hydrogels

Recombinant LC sequence of FUS (residues 2-214), a FET family RNA binding protein, adopted a hydrogel state when concentrated and stored at cold temperature. The LC domain of FUS is characterized by repetitive tyrosine residues flanked by glycine or serine. These repetitive G/S-Y-G/S motifs are reminiscent of the FG-repeat domains of certain nucleoporins that also adopt a hydrogel-like state. The hydrogels formed by FG-repeat domains were hypothesized to mimic the molecular permeability barrier of the nuclear pore (Frey and Gorlich, 2007; Frey et al., 2006). Large macromolecules such as mCherry protein are incapable of penetrating FG-hydrogels. Proteins with FG-repeats, however, accumulate in the gel and can carry associated cargo proteins with them. Inspired by the FG-hydrogel work, Dr. Masato Kato, associate professor in the McKnight laboratory, designed a fluorescence-based experimental system to study proteins interacting with the hydrogels formed by FUS LC sequence or hnRNPA2 LC sequence (Kato et al., 2012). For this assay, Dr. Kato concentrated various mCherry:LC fusion proteins, and incubated them in a sealed glass-bottom plate until they formed hydrogels on the glass. He challenged the pre-formed hydrogels with eGFP-tagged soluble proteins of interest and evaluated binding between the soluble proteins and hydrogels by the eGFP signals co-localized with mCherry signals from the hydrogels. Purified eGFP protein was not retained by the hydrogel and instead remained

diffuse throughout the chamber even after days of incubation. By contrast eGFP-FUS LC localized to the edge of mCherry:FUS hydrogel, indicating strong binding to the hydrogel (Figure 1-4). Furthermore, Dr. Kato tested several other LC sequences from RNA binding proteins (hnRNPA1, TDP-43, CIRBP, CPEB2, FMRP) and observed hydrogel binding. Finally, Han et al. showed that full-length versions of TIA1 and hnRNPA1 are retained by FUS hydrogels as are their LC domains alone. Deletion of the LC sequences caused the proteins to remain diffuse throughout the fluidic chamber, similar to eGFP alone (Han et al., 2012).

Han et al. showed in a proof of principle experiment that RNAs can also be retained by mCherry-FUS gel if recruited by an RNA-binding protein containing both a low complexity domain for interaction with the gel and an RNA-binding domain to bring with it RNA (Han et al., 2012). mCherry:FUS LC hydrogels were prepared on glass-bottom dishes, and challenged with Cy5.5-labeled hairpin RNA with either CFP alone, CFP fused to the LC sequence of FUS, CFP linked to MS2 (a prototypical RNA binding domain which binds the hairpin RNA) or both the LC sequence and the MS2 RNA binding domain. As shown in Figure 1-5, CFP alone or CFP:MS2 could not bind to the FUS hydrogel, CFP:FUS LC trapped in the hydrogel, but no RNA was recruited. When challenging the mCherry:FUS LC hydrogel with hairpin RNA and CFP:FUS LC:MS2, both CFP-tagged protein and Cy5.5-tagged RNA were trapped by the hydrogel. These experiments were interpreted to mimic the ability of RNA binding proteins to partition into RNA granules and to recruit mRNAs.

## Hydrogels are Composed of Labile Amyloid-like Fibers

Transmission electron microscopy of negatively stained hydrogels formed by mCherry:FUS LC or mCherry:hnRNPA2 LC revealed morphologically uniform fibrous structure 20-30 nm in width (Figure 1-6, A and B). To obtain structural insights of these fibrous structures, Kato et al. dialyzed the fibers against water, lyophilized the samples, and then subjected them to x-ray diffraction analysis. As shown in Figure 1-6 (C and D), fibers formed by both mCherry:FUS LC and mCherry:hnRNPA2 LC displayed diffraction patterns with prominent diffraction rings at 4.6-4.7 Å and at 10 Å. These diffraction rings are prototypical for cross- $\beta$  structure, in which the 4.6-4.7 Å diffraction represents the main chain – main chain distance of adjacent  $\beta$ -strands, and the 10 Å diffraction ring reflects the 10 Å distance in between  $\beta$ -sheets (Astbury et al., 1935; Geddes et al., 1968; Sunde and Blake, 1997). The x-ray diffraction of mCherry alone did not show these reflections. In combination with electron microscopic studies, these data provide strong evidence that LC sequence hydrogels are composed of amyloid-like polymers.

Cross- $\beta$  structures have been linked to pathological amyloid fibers that are extremely stable (Dobson, 2003; McKinley et al., 1983; Meersman and Dobson, 2006). Lindquist and colleagues performed a proteome-wide survey for prion-forming proteins in *S. cerevisiae*, and identified 19 new proteins that form ultrastable amyloid fibers (Alberti et al., 2009). Many of these prionogenic sequences were derived from transcription factors and RNA-

binding proteins. Unlike the prionogenic sequences studied previously, LC sequences in FUS and hnRNPA2 are enriched in glycine, proline, asparagine and glutamine residues, which are not likely to present in  $\beta$ -strands. I performed bioinformatic analysis of the likelihood of LC sequences to form  $\beta$ -structure, using the protein secondary structure prediction method described by Costantini et al. (Costantini et al., 2006). I calculated the probability of the LC sequence to fold as  $\beta$ -strand by summing up the  $\beta$ -propensities of each residue, and compared the  $\beta$ -propensities of LC sequences with the distribution of  $\beta$ -propensities of all proteins in human proteome (Figure 1-7). The  $\beta$ -propensities of all proteins in human proteome centered at 0.99, with a standard deviation of 0.058. The LC sequences of TDP43 and hnRNPA2, respectively, have  $\beta$ -propensities of 0.81, which are lower than 99.4% of proteins in human proteome. LC sequence of FUS has a  $\beta$ -propensity of 0.84, which is lower than 98.5% of proteins. Proteins known to form pathogenic amyloids, including Tau, A $\beta$ , and insulin, have higher  $\beta$ -propensities.

One of the criteria Lindquist group exploited to define prionogenic proteins is the insensitivity of the aggregates to the solubilizing effect of SDS (Alberti et al., 2009). SDS sensitivities of the amyloid-like fibers formed by mCherry:FUS LC and mCherry:hnRNPA2 LC were tested with semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) (Kato et al., 2012). Fiber samples were sonicated to break down big aggregates, and incubated for 10 minutes at 37 °C in presence of no SDS, 0.5%, 1% or 2% SDS, and loaded onto agarose gel to separate monomer and fiber species. Both fibers formed by mCherry:FUS LC and mCherry:hnRNPA2 LC were almost fully depolymerized by the 0.1% SDS added in gel running buffer (Figure 1-8). In contrast, the fibers formed by yeast Sup35

(ySup35) protein (Wickner, 1994) was resistant to even the highest concentration of SDS tested (2%). It appears that despite sharing morphological similarities and the common cross- $\beta$  structural signatures, the fibers formed by LC sequences of FUS and hnRNPA2 are labile.

### **Hydrogel Retention as an Assay of Fiber Polymerization**

Having observed that hydrogels are composed of amyloid-like polymers, Kato et al. (Kato et al., 2012) proposed two possible mechanisms by which exogenously supplied LC monomers might be trapped by hydrogels. Exogenously supplied monomers might bind laterally on the polymer surfaces. Once polymerized, the surface loops of LC sequences arrange in a repetitive manner that may favor the binding of other repetitive sequence. Alternatively, when incubating monomeric LC proteins with a hydrogel, the densely packed polymer ends might facilitate co-polymerization of the exogenous LC monomers into existing polymers within the hydrogel. Kato et al. incubated eGFP:FUS LC monomers with pre-formed mCherry:FUS LC polymeric seeds, and observed by fluorescence microscopy that the red colored mCherry:FUS LC seeds were extended bi-directionally by green colored eGFP:FUS LC. The seed regions of these co-polymers were exclusively red in color, without sign of lateral binding of eGFP:FUS LC.

The LC sequence of FUS contains 27 variants of the G/S-Y-G/S motif. Mutated variants of the eGFP:FUS LC were constructed to replace 5, 9, 15 or all 27 of tyrosine residues with serine. These mutants were then tested for binding to hydrogel droplets formed

by wild type mCherry:FUS LC (Kato et al., 2012). After overnight incubation, reduced hydrogel retention was observed when five tyrosine residues were mutated to serine; even weaker hydrogel binding was observed for the eGFP:FUS LC bearing nine tyrosine-to-serine mutations; and little or no hydrogel retention was observed for eGFP:FUS LC with 15 or more tyrosine-to-serine mutations. The same set of eGFP:FUS LC mutations were also tested in the fiber extension experiments. After incubating the monomeric eGFP:FUS LC mutations with polymeric seeds formed by wild type mCherry:FUS LC, the resulting fibers were examined by fluorescent microscope. Both eGFP:FUS LC mutants with five or nine tyrosine residues mutated were able to grow from the seeds, while no fiber extension was observed for eGFP:FUS LC bearing 15 or more tyrosine-to-serine mutations.

### **Serine:Arginine Domains of pre-mRNA Splicing Factors Bind hnRNPA2 Polymers**

Many pre-mRNA splicing factors contain long repeats of the dipeptide sequence serine:arginine (SR). Focusing on a member of the SR protein family that has been studied extensively, serine:arginine splicing factor 2 (SRSF2), Kwon et al. (Kwon et al., 2014) constructed the fusion protein that appended its SR domain to enhanced green fluorescent protein (eGFP). Unlike the LC sequences of FUS or hnRNP A2, the eGFP:SR does not form amyloid-like polymers, but does bind hydrogels formed by mCherry:hnRNPA2 LC (Figure 1-9).

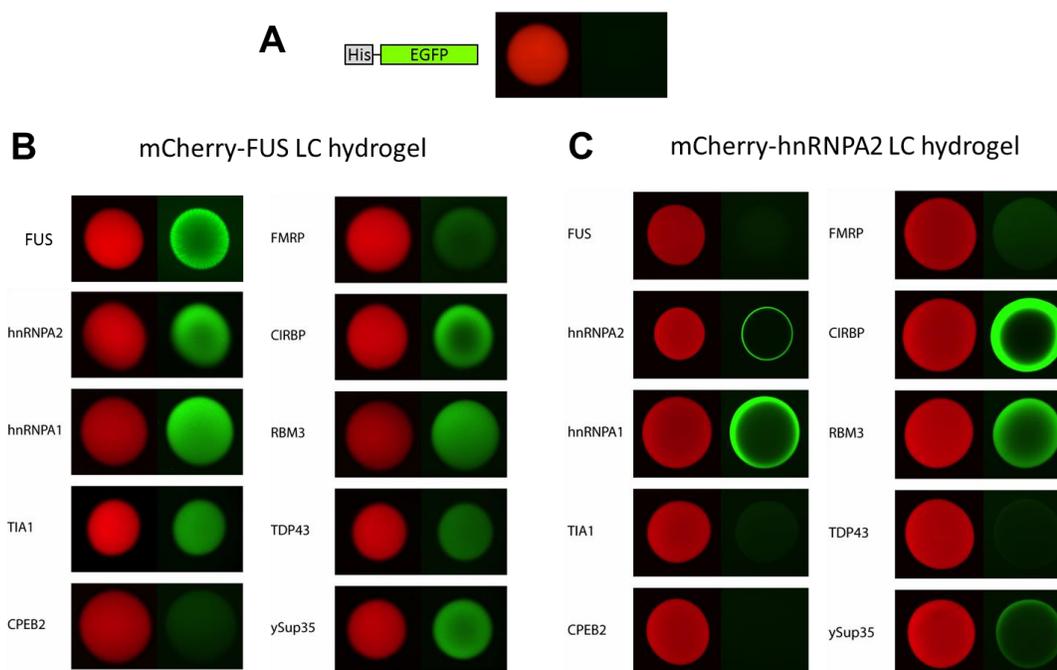
In eukaryotic cells, SR domain containing pre-mRNA splicing factors localized to interchromatin regions called nuclear speckles. Nuclear speckles are dynamic structures that

are roughly 1-3  $\mu\text{m}$  in diameter. They contain small nuclear ribonucleoprotein particles (snRNPs), spliceosome subunits and other non-snRNP protein splicing factors (Lamond and Spector, 2003; Spector and Lamond, 2011). Several kinases (such as CLK/STY, PRP4 and PSKH1) and phosphatases (such as protein phosphatase 1) are also localized to the speckles, and phosphorylation of SR domain release the SR-containing proteins from the speckles (Aubol et al., 2013; Colwill et al., 1996; Duncan et al., 1998; Menegay et al., 2000). In addition, the CLK enzymes themselves contain SR domains, and their SR domains are necessary for the kinase specificity of CLK (Keshwani et al., 2015).

Kwon et al. studied the phosphorylation of eGFP:SR by CLK1 *in vitro* (Kwon et al., 2014). The authors pre-bound the eGFP:SR fusion protein to hydrogel droplets formed by mCherry:hnRNPA2 LC and then exposed the droplets to ATP alone, CLK1 enzyme alone, or a mix of ATP and CLK enzyme. eGFP:SR was released from the hydrogel only in presence of both ATP and CLK enzyme (Figure 1-9A). In attempts to model the phosphorylation reaction within nuclear speckles, which contain both SR domain pre-mRNA splicing factors and the CLK kinases, hydrogel droplets were incubated overnight with eGFP:SR along with a derivative of CLK2 enzyme containing an SR domain. In this case, exposure to ATP alone released GFP:SR from the hydrogel droplet (Figure 1-9C), presumably because the hnRNPA2 LC polymer trapped the CLK2 enzyme via its SR domain and held it in proximity to the eGFP:SR substrate.

To investigate the mechanism by which SR proteins bind polymers formed by mCherry:FUS LC, eGFP:flag:SR was incubated with FUS LC fibers, and SR binding to fibers was visualized by immunoelectron microscopy. The FUS LC fibers with or without

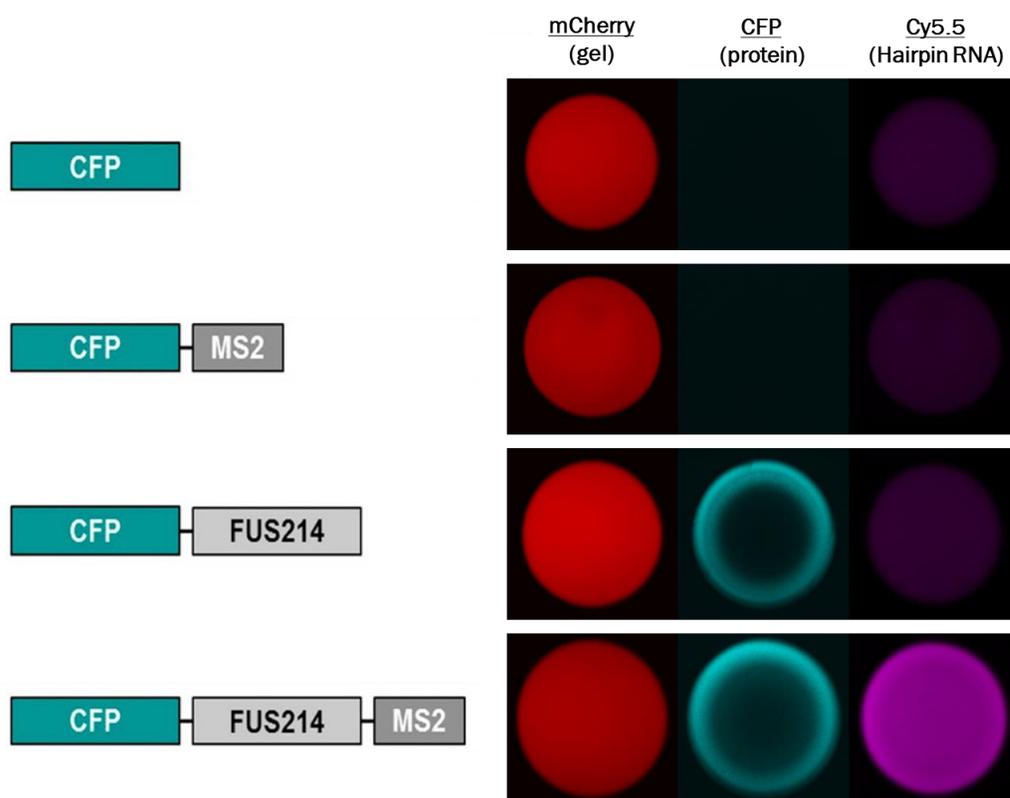
SR binding were absorbed to carbon copper grids, and stained with a mouse monoclonal antibody against the flag-epitope, followed by incubation with gold particles conjugated to secondary antibody. The mCherry:FUS LC polymers showed a repetitive helical pattern when negatively stained by methylamine vanadate, with no gold particles trapped on the fibers. When incubated together with SR proteins, the FUS LC polymers were decorated with gold particles periodically along the length (Figure 1-9D). These results suggested that, unlike the homotypic co-polymerization of FUS LC proteins, eGFP:flag:SR binds laterally to the surface loops of mCherry:FUS LC fibers.



**Figure 1-4. Hydrogels Bind eGFP-Tagged LC Sequences of RNA Regulatory Proteins**

(Kato et al., 2012)

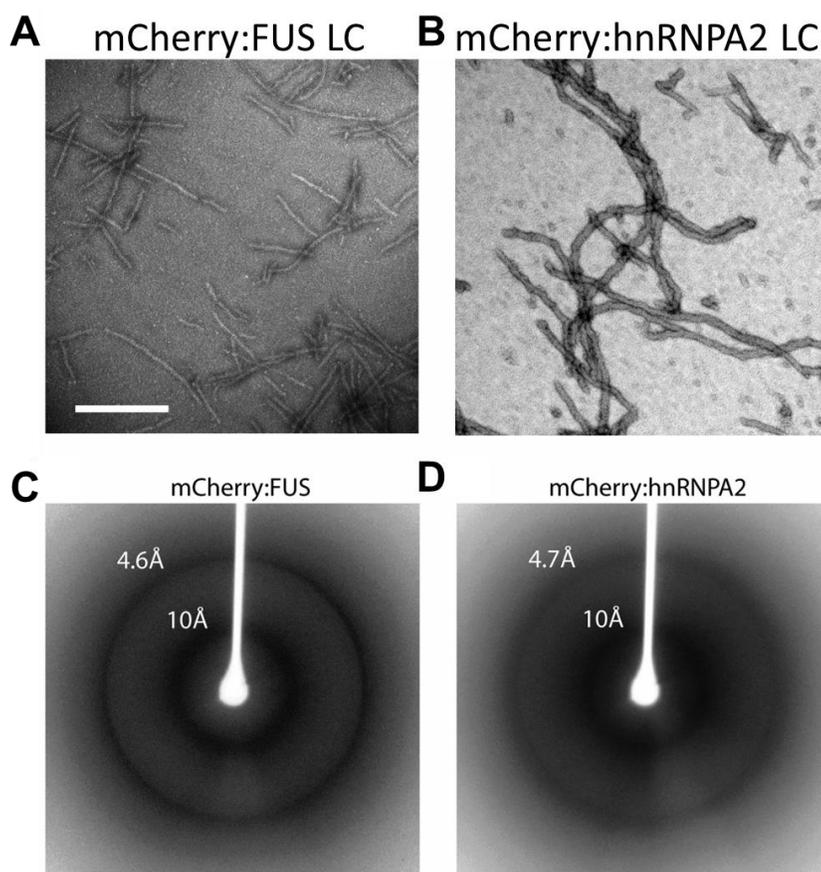
mCherry:FUS LC (A, B) or mCherry:hnRNPA2 LC (C) were concentrated and incubated in glass bottom plates to form hydrogel. The gels were then challenged with solution of eGFP or eGFP-tagged LC domain proteins. As shown in (A), eGFP alone cannot bind hydrogel and diffuses freely, while eGFP:LC selectively binds hydrogels formed by mCherry:FUS LC (B) or mCherry:hnRNPA2 LC (C). The intensities of eGFP signals indicates the abilities of the LC proteins in binding hydrogel. Ring shaped signal such as that eGFP:hnRNPA2 LC binding the hnRNPA2 hydrogel indicates strong binding, in which the eGFP-tagged protein was depleted by binding to edge of hydrogels.



**Figure 1-5. Hydrogel Retention of RNA Requires both LC and RNA-Binding Domains**

(Han et al., 2012)

- (A) Schematic diagrams of fusion proteins used for the hydrogel retention experiments. FUS214 is the LC domain of FUS. The labeled RNA is a high affinity substrate for the MS2 RNA binding domain.
- (B) Hydrogel retention assay with both CFP fusion protein and the RNA substrate. The Cy5.5-RNA (the third column) was only recruited to the mCherry:FUS LC hydrogel when co-incubated with protein containing both LC and MS2 domains. The RNA itself cannot bind hydrogel. The CFP:FUS LC is trapped in the hydrogel itself, but cannot recruit the RNA substrate.

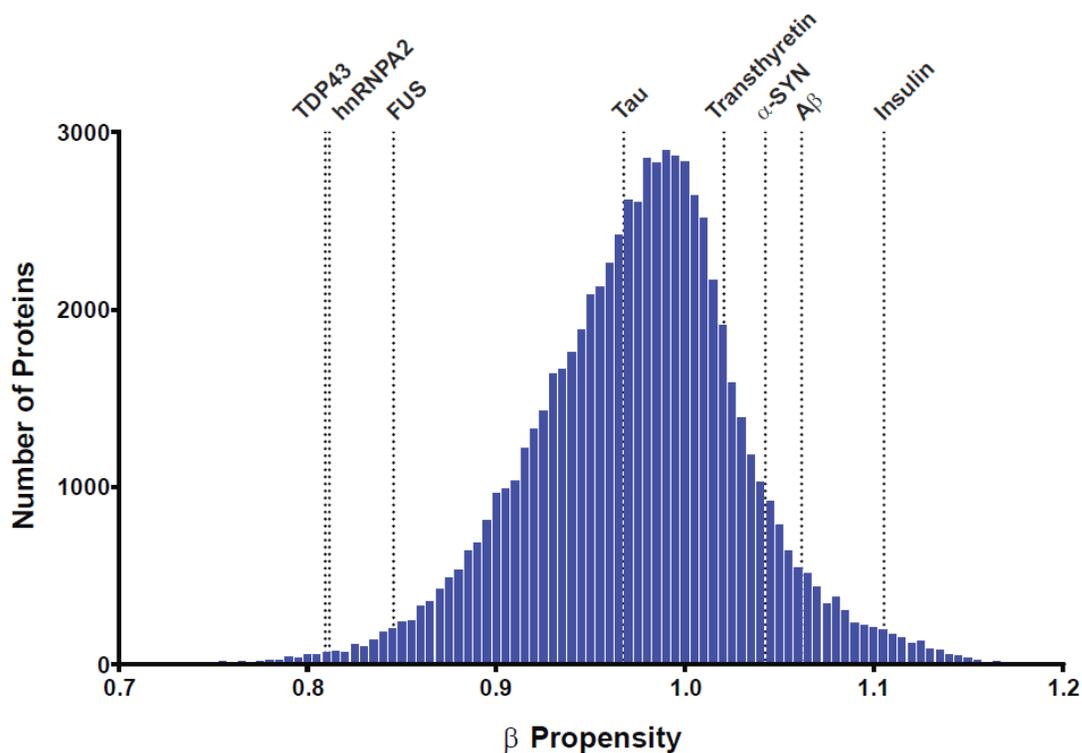


**Figure 1-6. Electron Microscopy and X-ray Diffraction Experiment of Hydrogels**

(Kato et al., 2012)

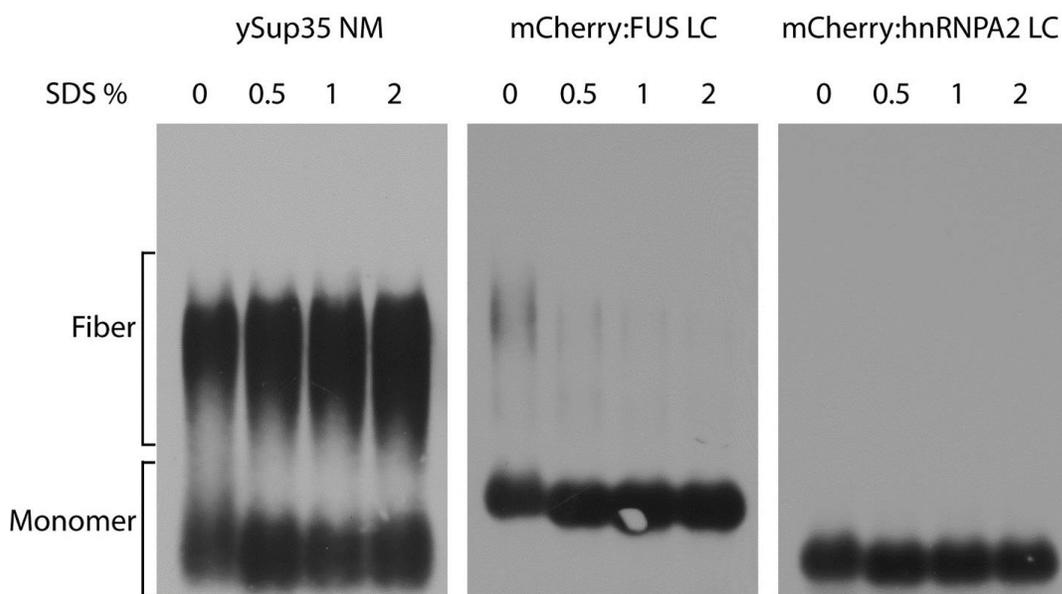
(A and B) Hydrogel droplets formed by mCherry:FUS LC (A) and mCherry:hnRNPA2 LC (B) were negatively stained by 2% uranyl acetate and visualized by transmission electron microscopy. Both samples contains polymeric, amyloid-like fibers.

(C and D) Hydrogel droplets formed by mCherry:FUS LC (C) and mCherry:hnRNPA2 LC (D) were dialyzed against water, lyophilized, and subjected to x-ray diffraction. Reflection rings were observed at 4.6-4.7 Å and at 10 Å.



**Figure 1-7. Propensity of LC Domains to Form  $\beta$  Structure**

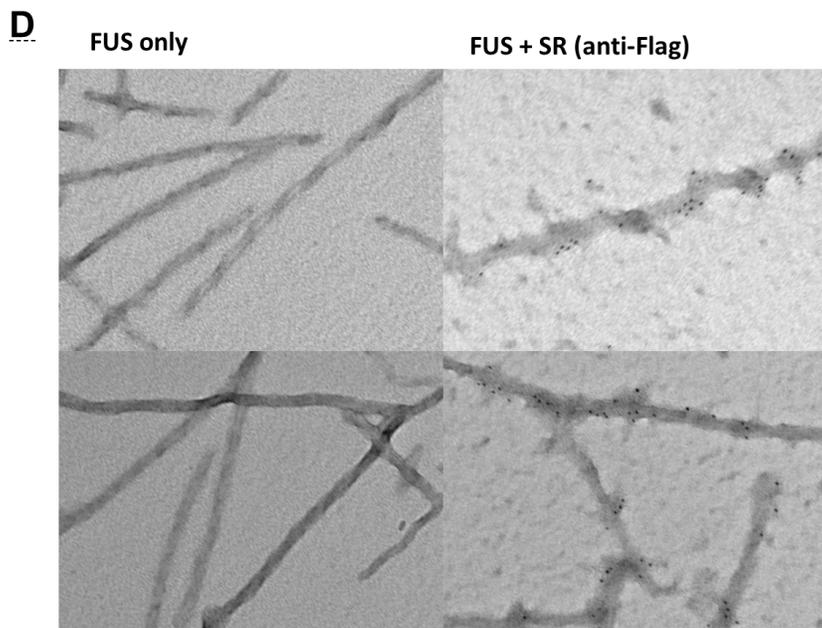
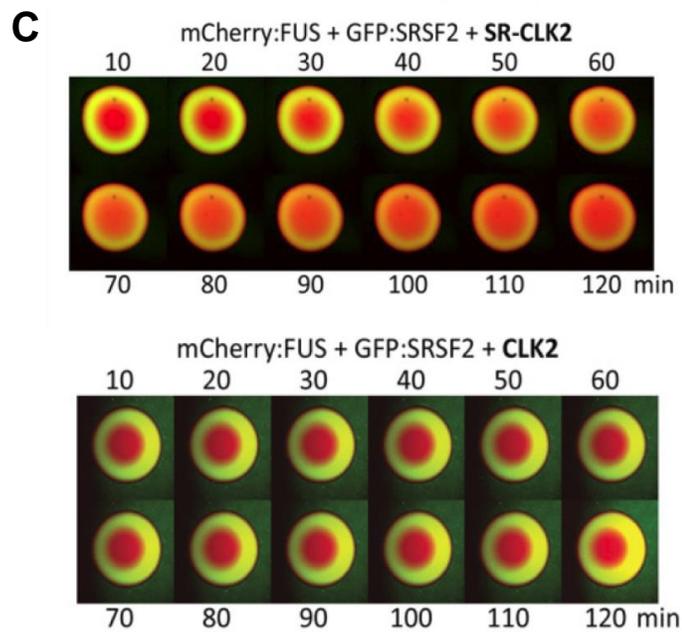
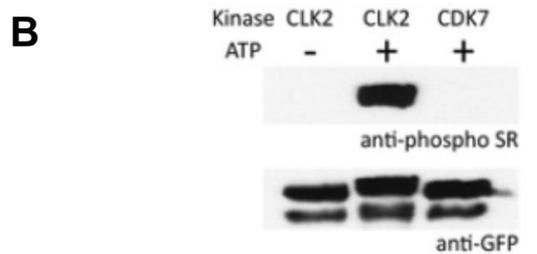
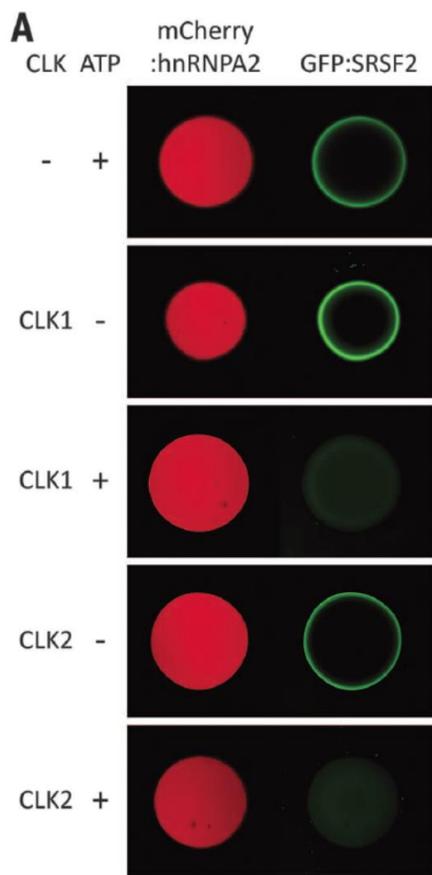
The propensity of each protein to form  $\beta$ -structure in human proteome was calculated by averaging  $\beta$ -propensities of all composing residues, and the frequencies of their  $\beta$ -propensities were plotted as blue bars. The resulting distribution centered at 0.99. LC sequences of FUS, hnRNPA2 and TDP43 have lower  $\beta$ -propensities than average proteins, while prototypic pathological amyloid proteins, such as  $\alpha$ -synuclein, insulin and A $\beta$  have higher  $\beta$ -propensities than average.



**Figure 1-8. Amyloid-like Fibers Formed by LC Domains Are Sensitive to SDS**

(Kato et al., 2012)

Agarose gel electrophoresis of polymers formed by yeast Sup35 (ySup35) NM, mCherry:FUS LC and mCherry:hnRNPA2 LC. Sample fibers were incubated at 37 °C for 10 minutes in absence or in presence of 0.5%, 1% or 2% SDS. After incubation with SDS, the fibers and monomers were separated by semi-denaturing detergent agarose gels running in a buffer containing 0.1% SDS. Fibers formed by LC domains of FUS and hnRNPA2 were almost fully depolymerized by the 0.1 % SDS in the running buffer alone. In contrast, the fibers formed by yeast Sup35 NM was not affected by the highest SDS concentration tested.



**Figure 1-9. GFP:SR Reversibly Binds hnRNPA2 Hydrogels**

(Kwon et al., 2014)

- (A) eGFP:SR was incubated with the mCherry:hnRNPA2 LC hydrogels and exposed to CLK1/2 kinase. eGFP:SR strongly bound hydrogels formed by hnRNPA2 LC, as shown in the green signal at the edge of hydrogels. Overnight incubation of the hydrogels with the CLK1/2 enzyme released eGFP:SR from the hydrogels in an ATP dependent manner.
- (B) Western blot assays using antibodies specific to either eGFP (lower blot) or to the phosphorylated state of SR repeats. Strong phosphorylation of SR was observed when eGFP:SR was exposed to the CLK2 enzyme, but no phosphorylation was observed when the fusion protein was exposed to the CDK7 protein kinase enzyme.
- (C) mCherry:hnRNPA2 LC hydrogel droplets were exposed to both eGFP:SR and a form of the CLK2 protein kinase carrying an SR domain. Upon exposed to ATP alone, the hydrogel-bound eGFP:SR was released (top panel). When the same experiment was conducted using a form of CLK2 that lacked the SR domain, provision of ATP failed to release the bound eGFP:SR (bottom panel).
- (D) eGFP:SR bound lateral surface of mCherry:FUS LC fibers. mCherry:FUS LC fibers were incubated with eGFP:SR or buffer alone and imaged by immunoelectron microscopy. The SR proteins on the fiber were recognized by monoclonal antibody against flag epitope and labeled by gold conjugated secondary antibody.

## **CORRELATIVE RELATIONSHIP BETWEEN FIBER FORMATION AND BIOLOGICAL ACTIVITIES OF LC SEQUENCES**

### **Tyrosine Residues in FUS LC Domain Critical for Stress Granule Recruitment**

FUS protein has been reported to localize to stress granules under oxidative stress (Bosco et al., 2010). To study the biological relevance of polymeric fibers formed by the FUS LC, Kato et al. expressed FUS bearing a variety of mutations in the LC sequence, and studied partitioning of the variants to stress granules.

Flag-tagged FUS wild type protein lacking the nuclear localization signal ( $\Delta$ NLS) was transiently expressed in U2OS cells. After cells were stressed by sodium arsenite, the tagged FUS protein predominantly localized to stress granules, as shown by co-localizing with the stress granule marker, TIA1. Kato et al. constructed variants of eGFP:FUS LC with five, nine, fifteen or all 27 tyrosine residues mutated to serine, and tested the abilities of the mutants to bind mCherry:FUS LC hydrogel. Variants of flag-tagged FUS  $\Delta$ NLS containing the same set of tyrosine-to-serine mutations in the LC domain were transiently expressed in U2OS cells and tested for the stress granule recruitment under oxidative stressed conditions (Figure 1-10). The FUS mutant harboring five tyrosine to serine mutations displayed similar staining pattern to the native protein. The FUS construct with nine tyrosine mutated to serine, which could only bind to the mCherry:FUS hydrogel weakly, exhibited smaller perinuclear puncta, and weak diffused signal in cytoplasm. As mentioned previously, the eGFP:FUS LC mutants with fifteen or more tyrosine residues mutated to serine could not bind the

mCherry:FUS LC hydrogel. Concordantly, no granular aggregates were observed in cells transfected with FUS constructs bearing the same mutations, and the flag:FUS signal remained largely diffuse. As control, cells transfected with the FUS construct containing 27 tyrosine-to-serine mutations were still able to assemble TIA1-positive granules under stress, thereby ruling out the possible failure of stress granule formation. Having observed that the effect of tyrosine to serine mutations in the FUS LC domain on hydrogel retention correlates with stress granule recruitment in U2OS cells, the authors hypothesized that the dynamic recruitment of RNA binding proteins to stress granule by their LC sequences is driven primarily by formation of amyloid-like polymers.

### **Correlation of Hydrogel Retention and Transcriptional Activation of FUS LC Domain Mutants**

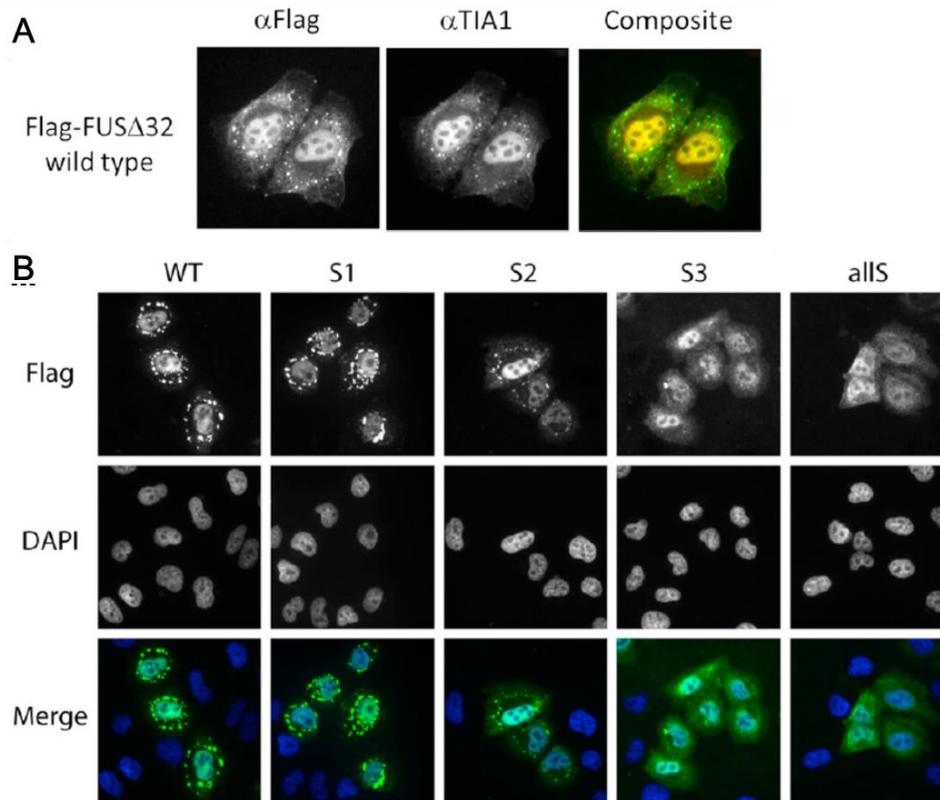
The fused in sarcoma (FUS) and Ewings sarcoma (EWS) genes were first identified as fusion genes produced by cancer-causing chromosome translocations. Since then, numerous forms of cancer have been discovered hosting chromosome translocations with amino terminal low complexity domains of FUS, EWS or TAF15 (FET) RNA regulatory proteins fused to DNA binding domains (Andersson et al., 2008; Arvand and Denny, 2001; Guipaud et al., 2006). The proteins expressed by the fusion genes were later identified as being necessary for the cancer formation, in which the low complexity sequences serve as constitutive transcriptional activation domains that force active transcription of genes supporting cell proliferation (Bertolotti et al., 1999; Crozat et al., 1993; Rabbitts et al., 1993).

Kwon et al. studied the transcriptional activation by FUS LC sequences (Kwon et al., 2013). The LC sequence of FUS (1-266) fused to the CHOP DNA binding domain in sarcoma was linked to the carboxyl terminus of the GAL4 DNA binding domain. The hybrid fusion was transiently transfected to 293T cells, together with a firefly luciferase reporter construct containing five tandem copies of the GAL4 binding site on its promoter region. Transcription activation was assayed by the luciferase activity in cell lysate the day after transfection. The same lysate was subjected to SDS-PAGE and western blot analysis to confirm GAL4:FUS LC expression. Transcriptional activation by the wild type FUS LC was compared with 43 mutants bearing between one and nine tyrosine residues randomly mutated to serine. As shown in Figure 1-11A, the transcriptional activation by the FUS LC constructs negatively correlates with the number of tyrosine-to-serine mutations. FUS LC constructs with one or two tyrosine residues mutated to serine appeared to be as active as the wild type FUS. Decreased activation of luciferase expression was observed for FUS LC constructs bearing three to six tyrosine residues mutated to serine, and no detectable transcription activation was observed for FUS LC derivatives carrying seven or more mutations.

The same set of mutations in FUS LC were also assayed for their ability to bind mCherry:FUS LC hydrogel. The mutated FUS LC (residues 2-214) were fused to eGFP, and tested for their binding to mCherry:FUS LC hydrogel. As shown in figure 1-11B, similar negative correlative relationship was observed for hydrogel binding and the number of tyrosine residues mutated to serine. The correlative plot shown in Figure 1-11C, compares transcriptional activation and hydrogel retention for all 40 mutants. A strong correlation

(Pearson's  $r = 0.7$ ) was observed between the two activities, with one exceptional variant (2A mutant) that transactivated more strongly than the wild type FUS LC.

Kwon et al. studied the gain of function phenotype of the 2A mutant. Purified recombinant 2A mutants showed no distinguishable difference with wild type LC domain protein in binding to mCherry:FUS LC hydrogels. Hydrogels formed by wild type or 2A mutant form of mCherry:FUS LC were challenged with eGFP fused to the carboxyl terminal domain of the largest subunit of RNA polymerase II (CTD). The eGFP:CTD bound hydrogels formed by the 2A mutant significantly more strongly than of the wild type LC domain of FUS. These observations give evidence that the 2A mutant achieves stronger transcriptional activation potential than the wild type LC by acquiring better binding to the CTD. In summary, by comparing transcriptional activation and hydrogel retention of FUS LC mutants, Kwon et al. observed strong correlation between the two activities, and discovered one outlier as a gain-of-function mutation that binds the CTD better than wild type FUS LC. These data give evidence that transcriptional activation capacity of the FUS LC domain may require its ability to form polymeric fibers.

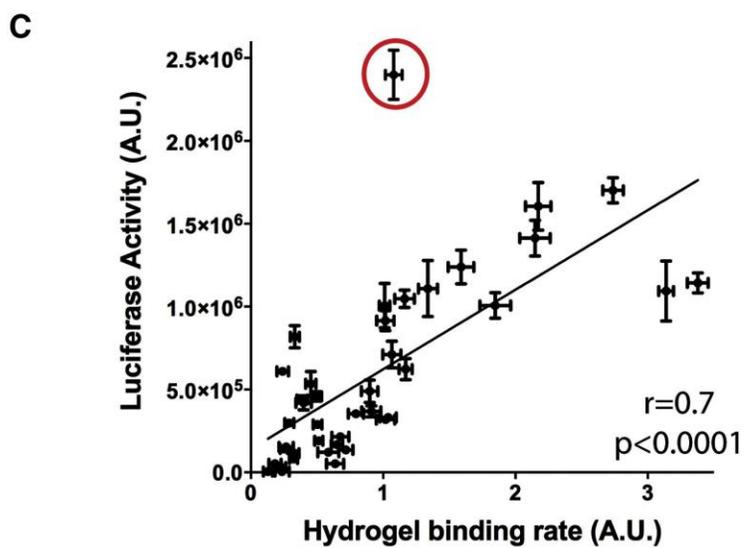
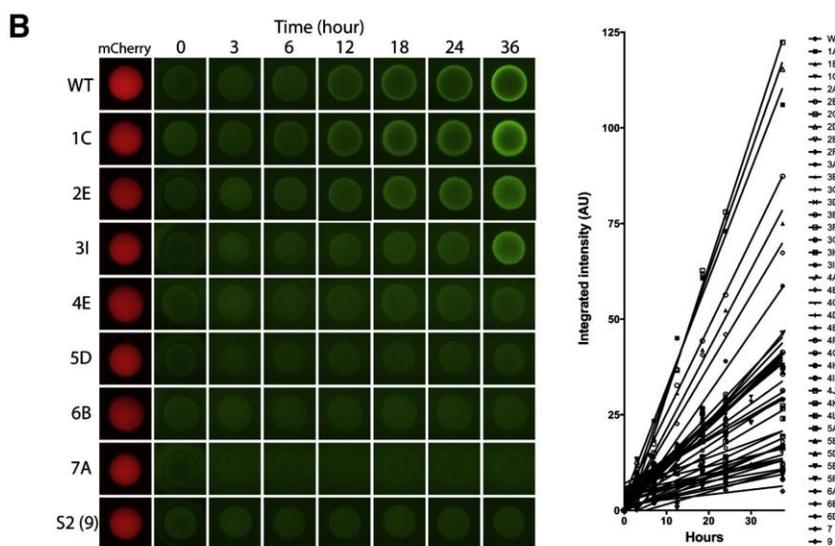
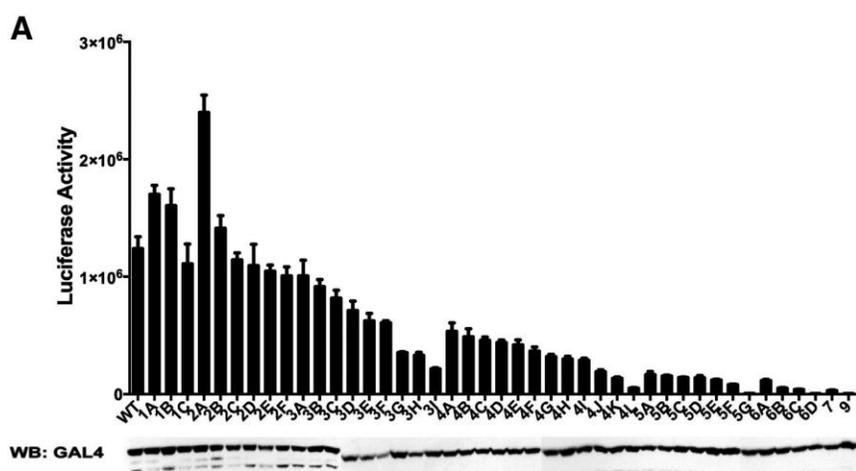


**Figure 1-10 Mutations of FUS LC Domain Affect Stress Granule Recruitment**

(Kato et al., 2012)

(A) Wild type FUS proteins lacking the nuclear localization sequence were recruited to stress granules in oxidative stress condition, as shown by co-localization with the stress granule marker, TIA1.

(B) In vivo stress granule recruitment assay for the tyrosine-to-serine mutants of FUS. Flag fusion vectors expressing FUS wild type or mutants lacking their nuclear localization sequences were transfected to U2OS cells. The cells were stressed with arsenite, fixed, and stained by monoclonal antibody against flag epitope. In the mutated vectors, S1, S2, S3 have 5, 9, 15 tyrosine mutated to serine, respectively.



**Figure 1-11. Correlation of Hydrogel Retention and Transcriptional Activation of FUS****LC Domain Mutants.**

(Kwon et al., 2013)

- (A) The FUS LC domain (1-266) was linked to the carboxyl terminus of GAL4 DNA binding domain, and tested for transcriptional activation. The GAL4:FUS LC wild type or mutated constructs were transiently transfected to 293T cells, together with a luciferase reporter construct with five tandem copies of GAL4 DNA-binding sites on its promoter region. Transcriptional activation by the wild type FUS LC was compared with 43 mutants with one to nine tyrosine residues randomly mutated to serine. Expression levels of each mutants were confirmed by western blotting against GAL4 DNA-binding domain.
- (B) Pre-formed mCherry:FUS LC wild type hydrogels were challenged by monomeric eGFP:FUS LC mutants, and time-lapse confocal images were recorded to obtain the initial binding rate of eGFP.
- (C) Correlative relationship between transcriptional activation and initial hydrogel binding rates. The Pearson correlation coefficient of the correlation is 0.7. The one significant outlier circled is the mutant 2A.

## **CHAPTER TWO**

### **The LC Domain of hnRNPA2 Adopts Similar Conformations in Hydrogel Polymers, Liquid-like Droplets and Nuclei**

The McKnight laboratory observed polymerization of the LC domains of the FET proteins as well as certain hnRNP proteins (Kato et al., 2012). When incubated at high concentration, the recombinant LC domains of these proteins polymerize into reversible amyloid-like fibers. X-ray diffraction revealed that the amyloid-like fibers were assembled by cross- $\beta$  structures. Transiently expressed FUS protein lacking the nuclear localization signal readily partitions into stress granules in oxidatively stressed U2OS cells. Certain variants of FUS protein bearing tyrosine-to-serine mutations in their LC sequences lost their abilities of forming amyloid-like fibers, and the same mutations impede FUS from being recruited to stress granules (Kato et al., 2012). By comparing the effects of mutations in the FUS LC domain on both transcriptional activation capacity and polymerization, a strong correlative relationship gave evidence that polymerization might be of critical importance to the function of LC domains in living cells (Kwon et al., 2013). However, the direct evidence of the structure of endogenous LC sequences remain uncertain. In attempts to address this shortcoming, I developed a chemical footprinting approach to probe the solvent accessible surfaces of proteins. I exploited the footprinting strategy on fibers formed by the LC domain of hnRNPA2, and provided evidence that the LC domain of hnRNPA2 exists in the same structural state in recombinant fibers, cell nuclei, and liquid-like droplets.

## DEVELOPMENT OF A CHEMICAL FOOTPRINTING METHOD

A variety of methods have been developed to obtain structural information by mass spectrometry, especially for the systems that are difficult to study by direct measurements with x-ray crystallography or nuclear magnetic resonance (NMR). Typically, the structural information is encoded into mass-to-charge ratio ( $m/z$ ) of proteolytic fragments from target proteins by altered structure-dependent modifications. To achieve this, multiple modifications have been used, including hydrogen/deuterium exchange (Katta and Chait, 1991; Konermann et al., 2011), chemical cross-linking (Back et al., 2003; Sinz, 2006), non-covalent attachment (Ly and Julian, 2006) and covalent labeling. Due to the repetitive nature of low complexity sequences, it would have been difficult to analyze cross-linked peptides. In order to practice the structure-dependent modification of LC domain in both test tube and cell nuclei, I focused on covalent labeling, which generates stable modification in  $m/z$  value that tolerate subsequent protein processing and purification. Covalent modification can be achieved either by chemicals modifying specific amino acids, as reviewed by Mendoz and Vachet (Mendoza and Vachet, 2009), or by broad-spectrum modifiers that react generally with many amino acids, as in the case of hydroxyl radicals (Tullius and Dombroski, 1986; Wang and Chance, 2011).

The LC sequence of hnRNPA2 has a strongly biased amino acid composition. In the LC domain containing a total 161 residues, 74% of the sequence is composed of only four different amino acids, including 71 glycine, 17 tyrosine, 16 asparagine and 15 serine side chains. Due to the high occurrence of tyrosine and serine throughout the LC sequence,

chemical reagents that specifically modify hydroxyl groups of tyrosine and serine side chains were anticipated to provide structure information with high residue resolution. N-acetylimidazole (NAI) is a reactive chemical that is capable of acetylating hydroxyl side chains in proteins (Riordan et al., 1965a; Timasheff and Gorbunoff, 1967). In this section, I outline optimization of buffer condition for the acetylation reaction, and describe development of the NAI footprint method with recombinant glutathione S-transferase (GST), and tests of the method with endogenous poly (ADP-ribose) polymerase 1 (PARP1) by acetylating intact cell nuclei purified from 293T cells.

### **Optimizing Reaction Conditions for Protein Acetylation by NAI**

Fifty years ago, NAI was first used by Simpson et al. to modify tyrosyl residues in bovine pancreatic carboxypeptidase A (Simpson et al., 1963). A subsequent study from the same group reported on the use of NAI in the determination of free tyrosyl residues in proteins (Riordan et al., 1965b). NAI readily hydrolyzes in water, generating imidazole and acetic acid (Jencks and Carriuolo, 1959). The rate of hydrolysis of acetyl imidazole increases sharply above and below neutrality, so stability of the reagent becomes a significant problem in buffers optimal for protein storage.

I tested the stability of NAI in a variety of commonly used protein buffers. To measure the UV spectra of NAI and its hydrolyzed product at physiological pH, 10 mM of both chemicals were dissolved in a buffer containing 50 mM HEPES pH 7.4 and 150 mM NaCl. As shown in Figure 2-1, the UV spectrum of NAI has two absorbance peaks, with the

major peak centered at 245 nm. The imidazole solution, in contrast, mainly absorbs UV light below 230 nm in wavelength, and absorbs UV only minimally at 245 nm. Having observed the UV spectra of the two chemicals, I monitored the hydrolysis of NAI by optical density at 245 nm. NAI is relatively stable in buffer condition containing 50 mM HEPES pH 7.4. The UV absorbance at 245 nm decreased linearly in a ten-minute time course (Figure 2-1C). High concentration of nucleophilic species, such as TRIS, destabilizes NAI, and higher rate of NAI hydrolysis was observed at alkaline pH. TRIS pH 8.8 fully hydrolyzed NAI within two minutes, and was used to quench remaining NAI from the acetylation reactions in future experiments.

To test reaction conditions for protein acetylation, 100  $\mu$ l of 1 mg/ml Bovine Serum Albumin (BSA) was acetylated in urea buffer (50 mM HEPES 7.4, 150 mM NaCl, 8 M urea). Denatured BSA solutions with additional imidazole or NaCl were acetylated with 200 mM NAI at room temperature for one hour, and the same volume of 1.5 M TRIS pH 8.8 was added to the reaction to quench remaining NAI. The solutions were then concentrated to 50  $\mu$ l with 10 kDa centrifugal filters, and diluted ten-fold with urea buffer. I repeated the procedure five times to clean up small molecules from protein solutions, and measured the number of tyrosine residues acetylated per molecule by differences in UV absorbance, as described by Riordan and Vallee (Riordan and Vallee, 1972). In basal buffer condition with urea buffer, NAI acetylated around nineteen out of twenty-one tyrosine in each BSA molecule. NaCl did not have a detectable effect on the acetylation of BSA (Figure 2-2B). 33 or 100 mM imidazole, however, impeded acetylation slightly, while 300 mM imidazole inhibited acetylation of BSA by 70%.

Different NAI concentrations were then tested, with the same 1 mg/ml BSA in urea buffer. 10, 30, 50, 100, 200 or 400 mM NAI was reacted with BSA for twenty minutes at room temperature. 400 mM NAI fully acetylated BSA protein, as shown in Figure 2-2C. 200, 100, 50, 30 or 10 mM of NAI acetylated 13, 4, 2, 0.8, 0.1 tyrosine residues per molecule, respectively. The information obtained from the acetylation experiments is reliable only if the original protein structure is preserved during the reaction. Otherwise, if the modification on a surface tyrosine perturbs the structure, any modification afterwards will no longer reflect the initial protein structure. Minimal amount of modification should be used to avoid distortion of original protein structure, but certain minimum level of acetylated peptide would be required for quantification by mass spectrometry. In future experiments, multiple NAI concentrations from 10 to 50 mM were tested to meet both requirements.

### **NAI Footprint of GST Enzyme**

Reasoning that the ability of NAI to modify certain amino acid sides chain might be influenced by the structural state of a protein, I compared modification of glutathione-S-transferase (GST) as a function of its folded versus unfolded state. GST enzyme was prepared under conditions of isotopic labeling with  $^{13}\text{C}_6$ -labeled tyrosine to produce a “heavy” protein sample. This sample was denatured with a chaotropic reagent, guanidine thiocyanate, and exposed to NAI under conditions leading to roughly one modification per polypeptide. The reaction was quenched with 0.8 M TRIS pH 8.8 which inactivates the NAI reagent. A corresponding “light” sample of GST was exposed to NAI in its folded state,

under conditions resulting in roughly one modification per polypeptide, and again quenched with TRIS to terminate the reaction. The heavy and light samples were mixed at a 1:1 ratio, digested with chymotrypsin, and then evaluated by Stable Isotope Labeling by Amino acids in Cell culture (SILAC) mass spectrometry (Figure 2-3A).

The peptides containing the same fragment of GST protein, but with different residues acetylated are termed position isomers. Position isomers have similar chemical properties, identical molecular weight, and will generate similar fragmented ions after collision-induced dissociation, which is challenging for chromatographic separation and modification site identification. I optimized the chromatography condition and programmed the mass spectrometer to target only peptides modified by single acetyl group. I searched the mass spectrometry dataset against sequence of GST enzyme, with search parameters that tolerate a mass error of 5 ppm for precursor ion, and 0.8 Da for fragmented ions. To examine the reactivity of NAI to different side chains, I included all amine groups and hydroxyl groups as potential acetylation site. I searched for chymotryptic peptides with a static modification of 57.02146 Da on cysteine, and a dynamic modification of acetylation (42.01056 Da) on tyrosine, serine, threonine, lysine, arginine, glutamine or asparagine, oxidation (15.99491 Da) on methionine, stable isotope (6.02013 Da) on tyrosine, respectively. Acetylation by NAI will not stop chymotrypsin from cleavage after tyrosine residues, and chymotryptic peptides with acetylated tyrosine at very carboxyl terminus was observed, e.g. peptide LLEYLEEKYEEHLY\* (Appendix C, Spectrum F. The asterisk indicates the residue acetylated). Consistent with previous literatures, NAI acetylates tyrosine residues, as shown in peptide LLEY\*LEEKYEEHLY (Appendix C, Spectrum E);

and lysine residues as shown in peptide LLEYLEEK\*YEEHLY (Appendix C, Spectrum G). Unexpectedly, NAI also acetylate serine and threonine residues, as shown in peptides MS\*PILGYW (Appendix C, Spectrum J) and LNGDHVT\*HPDFMLY (Appendix C, Spectrum H), respectively. Although less reactive, certain amine groups in arginine (in KKR\*IEAIPQIDKYL. Appendix C, Spectrum E), asparagine (in LN\*GDHVTHPDFMLY. Appendix C, Spectrum I) or glutamine side chains were also acetylated.

The chromatographic separation of a sample chymotrypsin fragment, KKRIEAIPQIDKYL, is shown in Figure 2-3C. The elution profile was extracted from mass spectrometry dataset by intensities of any precursor ions with m/z value within the range of  $588.35 \pm 5$  ppm. The first peak, with a retention time around 31 minutes, corresponds to a peptide with a charge of +4, which is different from the fragment of interest. The other four peaks, labeled with 1, 2, 3 and 4, are all position isomers of single acetylated peptide from the region. Acetylation sites of those four peptides were interpreted from the fragmented mass spectrum (MS2, Appendix C, Spectra A-D), and the averaged precursor mass spectrum (MS1) at the peak region for each peptide was used for quantification. The patterns of NAI reactivity with the denatured versus folded states of GST were different. As shown at the right side of Figure 2-3A, the MS1 spectrum of peptide KKR\*IEAIPQIDKYL was derived by averaging precursor ion spectra in the time window of peak 4, and the monoisotopic peak of light and heavy species, whose intensities were used for heavy-to-light ratio calculation, were marked with square and circle, respectively. The spectrum shows similar amounts of heavy and light peptide, suggesting that the arginine residues are acetylated to similar extent in the native and denatured states, meaning that the arginine residue is solvent accessible in

folded GST enzyme. The MS1 spectrum of peptide KKRIEAIPQIDKY\*L, however, derived from peak 2, shows more heavy peptide than light, suggesting that the residue is buried in the structure.

The NAI accessibility of amino acid side chains are summarized in Figure 2-4A (See also Table 2-1). Certain amino acid side chains reacted similarly in the two protein samples (Y23, K27, K40, K64 and Y111), whereas others were acetylated to a lesser extent in the folded sample compared with denatured GST (Y7, Y57, Y58 and Y192). I then compared this footprint with the degree of surface exposure of NAI-modified side chains as deduced from the x-ray crystal structure of the enzyme (Rufer et al., 2005) (PDB ID: 1Y6E). A strong correlative relationship was observed between NAI accessibility and solvent exposure in the structure (Figure 2-4B). Surface-exposed residues tended to be NAI-accessible, whereas residues buried within the core of the enzyme tended to be NAI-inaccessible. The NAI footprints show stronger negative correlation with solvent accessibility for residues with larger solvent accessible surface, while residues with near zero solvent accessible surface show a wide range of variety in NAI footprint. Presumably, buried residues with high structural flexibility can transiently move to the protein surface and become acetylated by NAI. I therefore compared NAI footprint of the buried residues with their temperature factor (B-factor) in the crystal structure (Figure 2-4C). Temperature factor of an atom describes the degree to which the electron density is spread out, which can be used as a measurement of dynamic mobility (Parthasarathy and Murthy, 1997; Radivojac et al., 2004). The B-factors of these buried residues show strong negative correlation with the NAI footprint, supporting the hypothesis that residues with near zero solvent accessible surface areas can be acetylated

if they are in a flexible state. Having observed the correlative match between NAI-accessibility and protein structure, I concluded that the NAI footprint is properly reflective of protein structure.

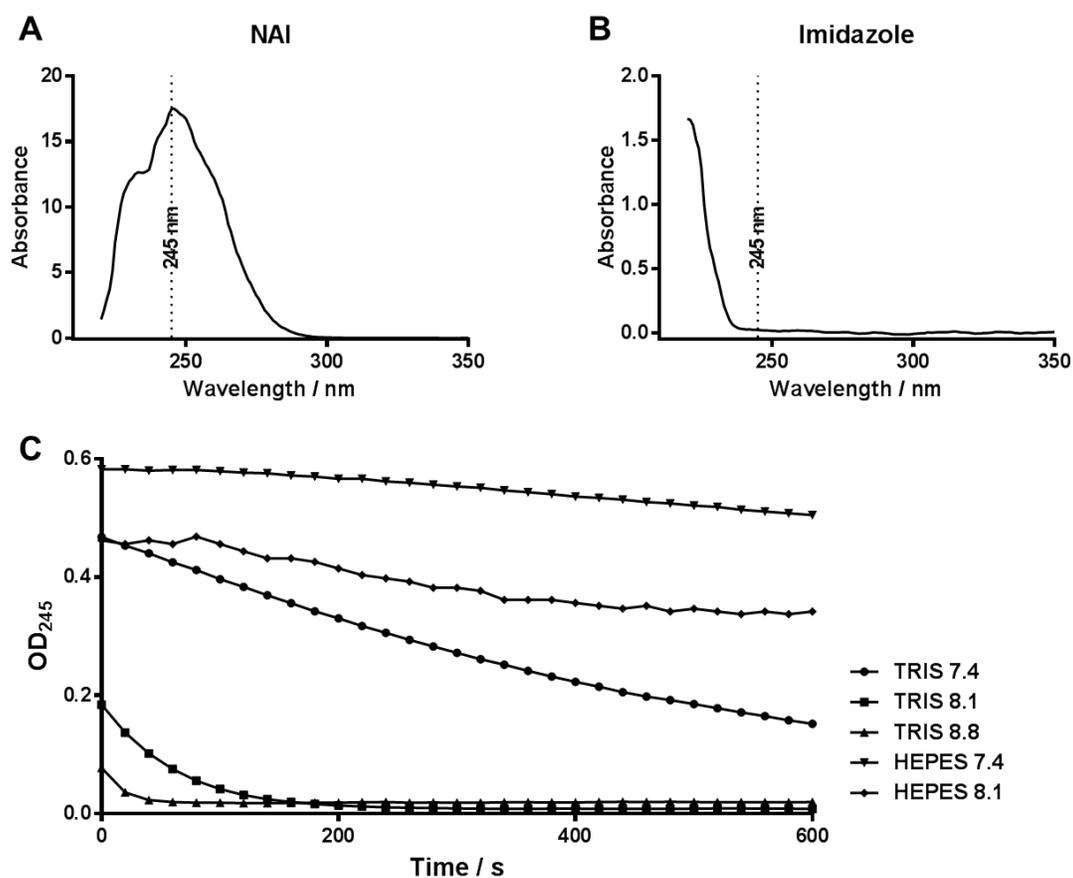
### **NAI Footprint of Nuclear PARP1**

Proceeding from a recombinant protein sample to a native protein within mammalian nuclei, I evaluated the difference in NAI modification of the poly-ADP-ribose polymerase 1 (PARP1) enzyme as a function of its folded versus denatured state. Heavy [ $^{13}\text{C}_6$ ]-tyrosine labeling was achieved by culturing cells in medium deprived of tyrosine and supplemented with an isotopically labeled form of the amino acid (heavy medium, see Appendix B). Incorporation of the heavy amino acids was confirmed by LC-MS/MS analysis of total cell lysate.  $2 \times 10^8$  of either heavy 293T cells or cells that had been grown in normal tissue culture medium (light) were harvested by trypsin digestion, washed and incubated on ice with hypotonic buffer (10 mM HEPES pH 7.4, 10 mM KCl, 2 mM  $\text{MgCl}_2$ , 5 mM BME with protease inhibitors). Cells were broken by use of a tissue homogenizer, and intact nuclei were recovered by centrifuging at 200x g for 10 minutes. The light sample of nuclei was resuspended in native nuclei buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 2 mM  $\text{MgCl}_2$  and protease inhibitor), exposed to a 30 mM level of NAI for 15 minutes before quenching with TRIS. The heavy sample was denatured in nuclei buffer with additional 5 M guanidine thiocyanate and sonicated prior to exposure to the same level of NAI, then also quenched with TRIS. The samples were combined, and fully denatured by heating to 95 °C for five

minutes in buffer containing 5 M guanidine thiocyanate. The mixed sample was then treated with benzonase in a buffer containing 1% SDS and MgCl<sub>2</sub>, digested with chymotrypsin overnight in 2 M urea, and desalted on a C18 column. In order to achieve deep coverage of the nuclear extract, the chymotryptic peptides from total nuclear protein were first fractionated on Agilent 300 Extended C18 column in acidic buffers, and each of the seventeen fractions was lyophilized and analyzed by LC-MS/MS.

14 single acetylated peptides from PARP1 were observed with maximum ion current of  $5 \times 10^4$  or higher, and were used to derive the levels of acetylation on 14 amino acid side chains in the native and denatured forms of PARP1. To subtract differences in post-translational modifications in cell, the differences in NAI modification of the PARP1 enzyme as a function of its folded versus denatured state were normalized by abundances of the corresponding native peptides with no modification (Figure 2-5B, see also Table 2-2). In the resulting NAI footprint, six residues were modified by NAI far more extensively in the denatured sample than the intact enzyme (K621, T799, K802, Y817, S902 and S904), five residues were modified slightly more extensively in the denatured sample relative to the intact enzyme (K571, S782, S783, S808 and K816), and three residues were modified equally in the two samples (K616, K903 and Y907). I again observed a correlation between NAI accessibility and protein structure (PDB ID: 3GJW) (Figure 2-5C). The three side chains that were modified equally in the two samples show a high level of predicted solvent accessibility in the x-ray crystal structure of PARP1 (Miyashiro et al., 2009). Likewise, five of the six residues observed to be highly protected from NAI modification are predicted to be solvent inaccessible by the crystal structure of the enzyme.

Analysis of three consecutive residues in the polypeptide chain of PARP1 is particularly revealing. Serine residue 902 was protected from NAI modification in nuclear PARP and buried beneath the surface of the enzyme. Lysine residue 903 is surface exposed and NAI accessible in the folded form of PARP1. Finally, serine residue 904 is NAI inaccessible in the folded enzyme, and buried beneath the surface in the PARP1 crystal structure. Here I offer that the correlative relationship between NAI accessibility and the predicted level of surface exposure of a given amino acid side chain validates this means of probing protein structure both in a recombinant protein and a native enzyme present in nuclei of mammalian cells.

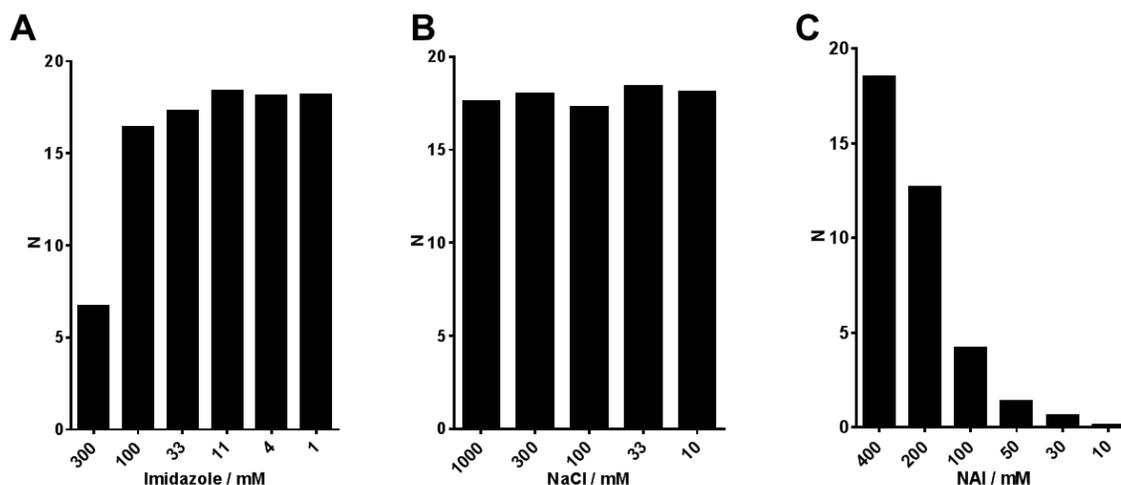


**Figure 2-1. Stability of NAI Compound in Buffers.**

(A) UV spectrum of N-acetylimidazole at pH 7.4.

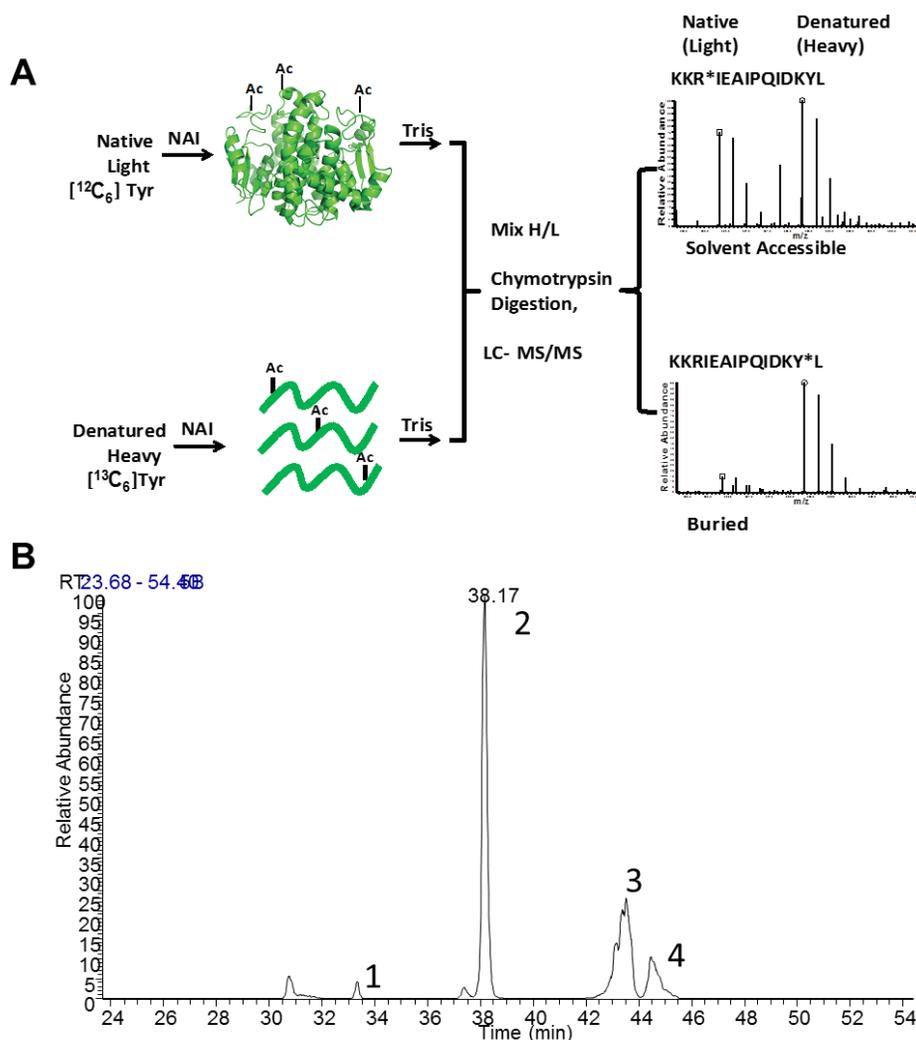
(B) UV spectrum of imidazole at pH 7.4.

(C) Hydrolysis of NAI recorded by UV absorbance at 245 nm. NAI is not stable in alkaline buffer, nor in presence of TRIS.



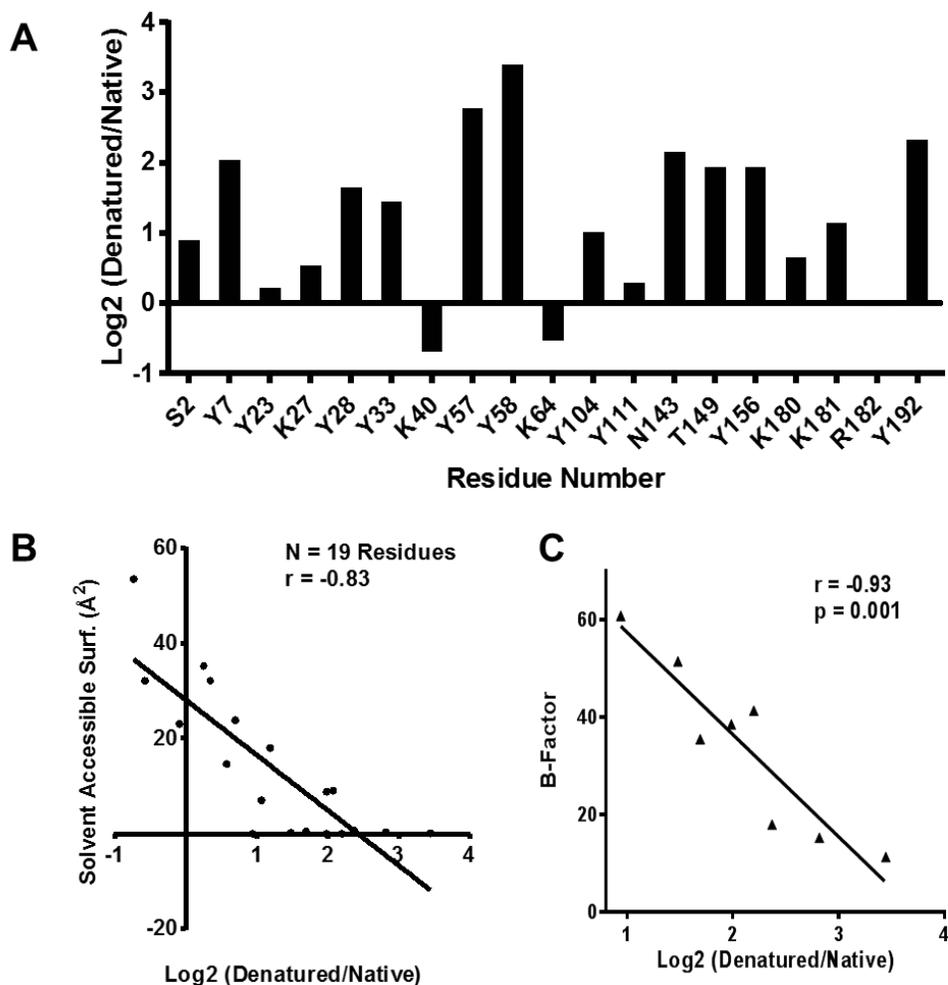
**Figure 2-2. Acetylation of BSA**

- (A) Average number of tyrosine residues acetylated per molecule in buffer with imidazole. The number of tyrosine modified was calculated based on the decrease of UV absorbance at 278 nm. Decreased level of tyrosine acetylation in BSA was observed in buffers with 300 mM or 100 mM imidazole.
- (B) Average number of tyrosine residues acetylated per molecule in different concentration of NaCl salt. NaCl with concentrations up to 1 M does not affect acetylation of tyrosine residues in BSA.
- (C) Number of tyrosine residues acetylated per molecule with a variety of NAI concentrations.



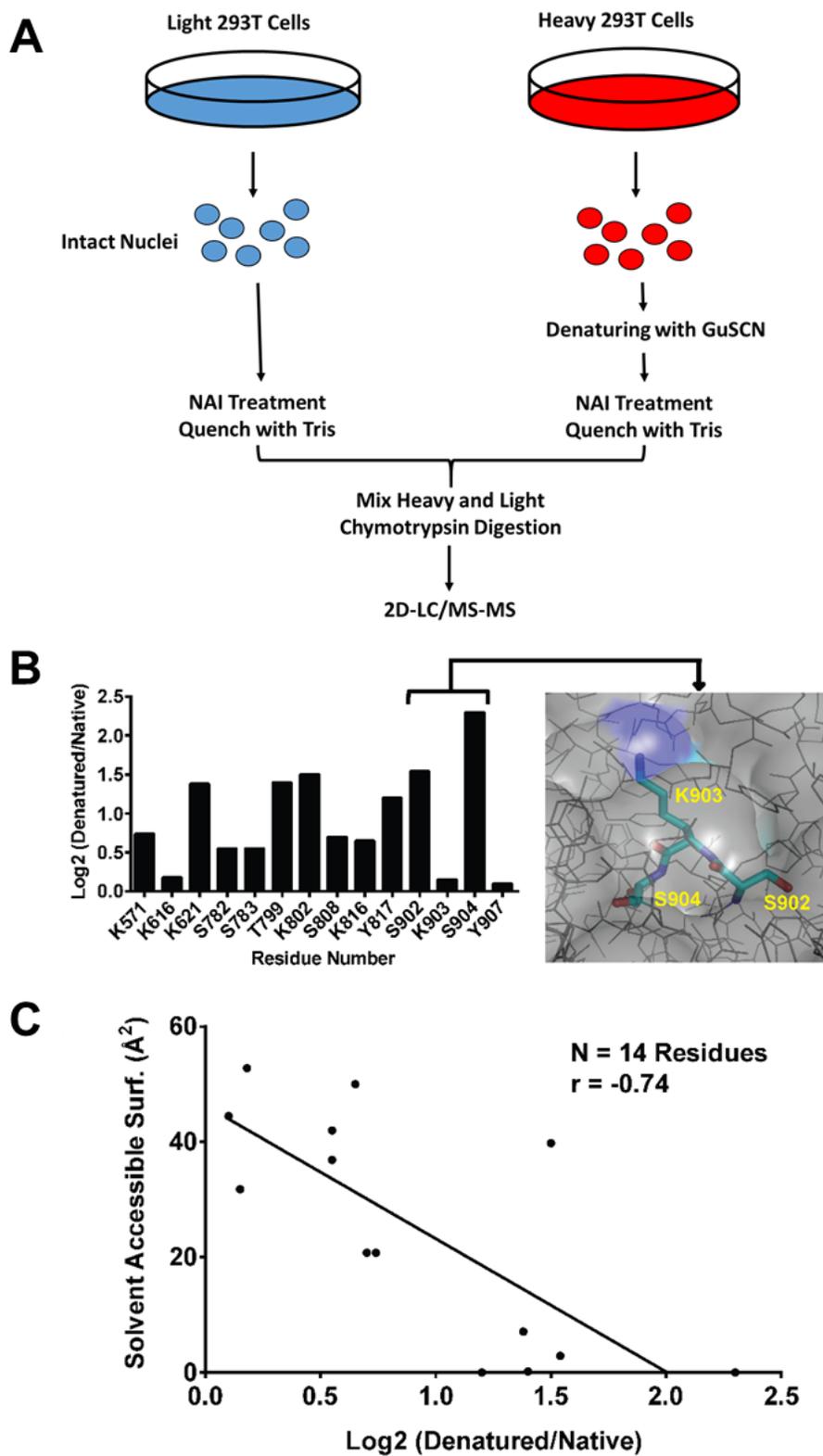
**Figure 2-3. Differing Patterns of Acetylation of Folded and Denatured Samples of Glutathione-S-transferase Mediated by N-acetylimidazole.**

- (A) Folded glutathione-S-transferase (GST) was exposed to N-acetylimidazole (NAI) under conditions leading to roughly one modification per polypeptide chain, with the reaction quenched by the addition of 0.8 M TRIS. A separate batch of GST grown in bacterial cells supplemented with  $^{13}\text{C}$ -labeled tyrosine was denatured in 5 M guanidine thiocyanate prior to NAI treatment. Following quenching with TRIS, the two samples were mixed, digested with chymotrypsin and subjected to SILAC mass spectrometry. MS spectra of two location isomers were shown on the right.
- (B) Extracted chromatography with  $m/z = 588.35$ , the mass of chymotryptic fragment KKRIEAIQIDKYL with one acetyl group attached. Four chromatographic peaks, labeled with 1-4, were identified as K\*KRIEAIQIDKYL, KKRIEAIQIDKY\*L, KK\*RIEAIQIDKYL, KKR\*IEAIPQIDKYL, respectively, with the acetylated residues marked by an asterisk. See Appendix C for their MS2 spectra.



**Figure 2-4. NAI Footprint of GST Enzyme Correlates with Solvent Accessible Surface**

- (A) Nineteen acetylated side chains were scored for abundance in the two samples, yielding an NAI footprint. The degree of residue protection from NAI modification in the folded state, relative to the denatured state, is measured on the Y axis as log<sub>2</sub> values (Data summarized in Table 2-1).
- (B) Plot showing the negative relative relationship (Pearson's  $r = -0.83$ ) between the degree of protection from NAI in the folded state, relative to the denatured state (X-axis), and the measured level of solvent accessibility determined from the x-ray crystal structure of GST (Y-axis). Note that residues with zero solvent accessible surface (analyzed in C) show a wide range of variety in NAI accessibilities.
- (C) Correlation plot comparing B-factor assigned in the x-ray crystal structure (Y-axis), to the degree of protection from NAI (X-axis) for residues with solvent accessible surface area less than 1 Å<sup>2</sup>.



**Figure 2-5. Establishment of an NAI Footprint for the Poly-ADP-ribose Polymerase Enzyme within Nuclei Isolated from Mammalian Cells.**

- (A) Nuclei isolated from 293T cells grown in normal tissue culture medium were exposed to NAI and quenched with TRIS. Nuclei isolated from 293T cells grown in medium supplemented with  $^{13}\text{C}$ -labeled tyrosine were denatured in 5 M guanidine thiocyanate, exposed to NAI and quenched with TRIS. The two samples were mixed at a 1:1 ratio, digested with chymotrypsin and subjected to SILAC mass spectrometry.
- (B) Quantitation of relative NAI protection ( $\log_2$  denatured/native) is shown on X-axis for each of 14 acetylated amino acids (Y-axis). Insert at right shows a portion of the surface of the PARP1 enzyme as resolved by x-ray crystallography, highlighting surface exposure of lysine residue 903 and buried states of serine residues 904 and 902 (Data Summarized in Table 2-2).
- (C) Plot showing correlative relationship between the NAI accessibility of 14 acetylated amino acids of PARP1 (X-axis) relative to their calculated solvent accessibility as deduced from the X-ray crystal structure of the enzyme (Y-axis).

**Table 2-1. Footprint Deduced by NAI-Mediated Acetylation of GST.**

<b>Residue Number</b>	<b>Log<sub>2</sub>(Denatured/Native)</b>	<b>Solvent Accessible Surface (Å<sup>2</sup>)</b>
S2	0.94	0
Y7	2.07	9.1
Y23	0.25	35.3
K27	0.58	14.7
Y28	1.69	0.5
Y33	1.48	0.2
K40	-0.74	53.5
Y57	2.82	0.3
Y58	3.45	0.1
K64	-0.58	32.1
Y104	1.06	7
Y111	0.34	32.1
N143	2.20	0
T149	1.99	0
Y156	1.99	8.8
K180	0.70	23.8
K181	1.19	18
R182	-0.09	23.1
Y192	2.37	0.7

**Table 2-2. Footprint Deduced by NAI-Mediated Acetylation of PARP.**

<b>Residue Number</b>	<b>Log<sub>2</sub> (Denatured/Native)</b>	<b>Solvent Accessible Surface (Å<sup>2</sup>)</b>
K571	0.74	20.8
K616	1.54	2.9
K621	0.15	31.8
S782	2.3	0
S783	1.4	0.2
T799	1.5	39.8
K802	0.7	20.8
S808	0.65	50
K816	1.2	0
Y817	0.1	44.5
S902	0.18	52.8
K903	1.38	7.1
S904	0.55	42
Y907	0.55	36.9

## **NAI FOOTPRINT OF HNRNPA2 LC DOMAIN IN RECOMBINANT FIBERS**

### **Determination of the NAI Footprint of Recombinant hnRNPA2 Fibers**

Hydrogel droplets were formed using a fusion protein linking mCherry to the LC domain of hnRNPA2 (Kato et al., 2012). This protein sample was exposed to 3, 10, 30, 100 or 300 mM NAI for 15 minutes at room temperature, then quenched with TRIS (Experimental Procedures). Similarly prepared hnRNPA2 polymers were formed using protein isotopically labeled with heavy tyrosine. The latter sample was denatured with guanidine thiocyanate prior to NAI-mediated modification, followed by quenching with TRIS. The two samples were combined at a 1:1 ratio, fully denatured by heating in guanidine thiocyanate, digested with chymotrypsin in presence of 2 M urea, and then analyzed by mass spectrometry.

Given the repetitive nature of hnRNPA2, multiple tyrosine or serine residues, which are all highly reactive to NAI, could be present in each peptide. HPLC chromatography was successful in separating position isomers, as shown in the example in Figure 2-6. Extracted chromatography of acetylated peptide derived from the residues 279-294 of hnRNPA2 showed five major peaks (Figure 2-6A). The peptide first eluted was acetylated on serine residue 285, at 38 minute; followed by three consecutive peaks, each separated one minute apart, which were identified as peptides with acetyl group on residues asparagine 282, tyrosine 283 and tyrosine 288, respectively. A peptide acetylated at its chymotryptic cleavage site, tyrosine 294, was eluted one and a half minutes later. Identification of the

residue acetylated was achieved by targeting only precursor ions of twenty-one different  $m/z$  values corresponding to chymotryptic peptides with at least one tyrosine residue, and modified by a single acetyl group (Figure 2-6B), for fragmented mass spectrometry.

Intensities of four native peptides (with no acetylation) were examined to confirm consistent chymotrypsin digestion (Figure 2-7). Peptides with cleavage after both tyrosine (B, C, D) and phenylalanine (A) residues were tested. Peptide GGSRNMGGPY exhibited no miss-cleavage, while peptides NQQPSNYGPMKSGNF (A), GPMKSGNFGGSRNMGGPY (D), and RGGSDGYGSGRFGDGYNGY (C) had one, one and three miss-cleavage, respectively. Despite a wide range of ion current, all four peptides showed heavy to light ratios close to one, which suggested that both heavy and light proteins were consistently digested. Robust chymotrypsin digestion generating similar heavy-to-light ratios for all native peptides, which is essential for SILAC quantification, was achieved by additional denaturation after mixing heavy and light proteins.

Intensities of heavy and light forms of peptide GPMKSGNFGGSRNMGGPY were also examined with samples exposed to different amount of NAI. The target peptide contained the only lysine residue in the LC domain of hnRNPA2. Assuming that the peptide has at least some buried sides chain in native structure, NAI treatment will acetylate more peptides from denatured heavy protein, leaving less heavy peptide unmodified. This preference in acetylation of heavy proteins will affect heavy-to-light ratio of unmodified peptide when significant portion of proteins were acetylated. Sample groups treated with 3, 10 or 30 mM NAI have similar amount of heavy and light form of the target peptide, indicating that only a small amounts of peptide acetylated. In sample groups treated with 100

mM and 300 mM NAI, however, heavy-to-light ratios of the target peptide were 0.55 and 0.35, respectively, suggesting that a large portion of peptide was acetylated. Samples with less acetylation were favored, as they were interpreted to preserve the native structure of polymers formed by hnRNPA2 LC, and experimental dataset from fibers acetylated with 10 or 30 mM NAI were included to calculate the footprint.

Chymotrypsin digestion generated peptides with miss-cleavage, and each residue was present in multiple peptides, according to different miss-cleavages. As shown in Figure 2-9A, both peptides GPMKSGNFGGS\*RNMGGPY (residues 302-319 of hnRNPA2, with acetylation on serine residue 312) and GGS\*RNMGGPYGGGNYGPGGSGGSGGY (residues 310-331) were acetylated at the same site, serine residue 312. MS1 spectrum of the first chymotryptic fragment showed that the heavy peptide is 2.5-fold more abundant than the light peptide (Figure 2-9B), suggesting that serine residues 312 is 2.5 folds more accessible in denatured condition than its native state. MS1 spectrum of the second chymotryptic fragment revealed the same heavy-to-light ratio, despite differences in peptide lengths, signal intensities and charge states. Having tested multiple examples each with different miss-cleaved peptides containing a same acetylated residue, I conclude that the peptide studied to compare its acetylation in native and denatured states does not affect resulting NAI accessibility ratio of other amino acid side chains.

Twenty-three amino acid side chains in hnRNPA2 LC were evaluated for NAI accessibility. Twelve amino acids appeared to be equally accessible to NAI-mediated modification in the two samples, and eleven appeared to be less accessible in the native fibers relative to the denatured protein sample (Figure 2-10). As an example, three of these

acetylated amino acid residues could be identified in the same peptide spanning amino acids 302-319 of the hnRNPA2 polypeptide. HPLC chromatography was successful in separating variants of this peptide acetylated at lysine 305, serine 306, or serine 312 (Figure 2B). The peptide variant acetylated at K305 was found at equal abundance in both light and heavy samples, indicative of the ability of NAI to modify this residue irrespective of whether the protein was in the fibrous or denatured state. The variant acetylated at S306 was considerably less abundant in the light sample than the heavy sample, giving evidence of its protection from NAI modification in the fibrous state. Finally, the variant acetylated at S312 was slightly less abundant in the light sample relative to the heavy sample, consistent with partial protection from NAI modification when the LC domain of hnRNPA2 existed in the polymeric state.

The pattern of protection from NAI modification in polymeric fibers of the hnRNPA2 LC domain, or lack thereof, can be described in the following way. An amino-terminal forty residue region contains no tyrosine and thus cannot be labeled. In the middle of LC sequence, an extensive region of the protein was equally acetylated by the chemical probe irrespective of the fibrous or denatured state, suggesting a flexible, solvent accessible stretch. An equally extensive segment corresponding to a more C-terminal region of the LC domain was protected in the polymeric state, relative to the denatured state, at 11 out of 12 acetylated residues. Right within the middle of this apparently ordered region of the LC domain, lysine residue 305 was found to be equally accessible in both the polymeric and denatured states of the protein, while the subsequent serine residue 306 was acetylated more in the denature state over native, polymeric state. Finally, the three most C-terminal residues scored in the assay

were all equally accessible under both fibrous and denatured states, suggesting a disordered carboxyl terminal tail of the LC domain.

## **Mutations in the NAI-protected Region of the hnRNPA2 LC Domain Impede Hydrogel**

### **Binding**

Is the NAI footprint telling us anything of functional relevance to the LC domain of hnRNPA2? To address this question, Dr. Masato Kato prepared mutated variants of the LC domain of hnRNPA2 wherein all 25 phenylalanine and tyrosine residues were individually mutated to serine. eGFP fusion proteins representing wild type hnRNPA2 and all of the individual mutants were expressed in bacterial cells, purified, and stored with 2 M urea to be kept in their monomeric state. All eGFP:hnRNPA2 LC variants were diluted to 1  $\mu$ M and assayed at the same time for their retention to hydrogel droplets formed by wild-type mCherry:hnRNPA2 LC (Figure 2-11A).

Of the 25 mutants, six were found to substantially impede binding to hydrogel droplets formed from mCherry fused to the wild type LC domain of hnRNPA2, which was presumably due to decreased ability to polymerize. Five of the six tyrosine- or phenylalanine-to-serine mutations that substantially impede hydrogel binding occur within the region of the LC domain that is protected from NAI modification in the fibrous state (Y278S, Y283S, F291S, F309S and Y319S). The sixth mutant that was significantly impeded in hydrogel binding, Y264S, occurs on the amino terminal side of the NAI protected region within a span where I failed to find acetylated peptide with good intensity – a dead

zone in the footprint (residues 258-282). I tentatively conclude that these six residues which are important for the fiber formation of hnRNPA2 LC become buried in the polymeric state, and that polymerization causes NAI protection.

The remaining 19 mutants fell into two categories with respect to hydrogel binding. Twelve mutants bound to hydrogels in a manner indistinguishable from wild type hnRNPA2. Two of these mutants, Y335S and Y341S, were located in the very C-terminal region of the LC domain, concordant with a small region that was fully accessible to NAI modification irrespective of whether the protein was in a polymeric or denatured state. Seven of these phenotype-void mutants, F95S, F197S, F207S, F215S, Y222S, F228S and Y250S were located in the amino terminal region of the LC domain that was widely accessible to NAI modification irrespective of structural state. The remaining three mutations that had no discernible effect on hydrogel binding, F244S, Y257S, Y275S, were all localized in the dead zone of the NAI footprint. Finally, seven mutants, including Y235S, Y250S, Y271S, Y288S, Y274S, Y301S and Y324S, mildly affected binding to mCherry:hnRNPA2 LC hydrogels. These seven mutants mapped randomly across the LC domain of hnRNPA2.

A direct comparison of NAI footprint and hydrogel retention is shown in Figure 2-11B. The black squares with error bars are the NAI footprint, plotted with Log<sub>2</sub> scale on the left Y-axis, while the hydrogel retention were plotted in blue curve as ratio of hydrogel binding of wild type hnRNPA2 LC over the binding of mutant, on the right Y-axis. The two overlapping plots reveal similar trends, qualitatively. I conclude that tyrosine- and phenylalanine-to-serine mutations in NAI protected regions impede hydrogel binding,

whereas those in NAI accessible regions do not impede hydrogel binding. This conclusion favors functional significance of the NAI footprint.

### **Human Genetics Reveals Causative Mutations of Familial Neurodegenerative Diseases in the NAI-protected Region of the hnRNPA2 LC Domain**

Ubiquitin-positive inclusions that contain RNA-binding proteins, such as TDP-43, hnRNPA1, and hnRNPA2B1, have been shown to associate with a group of inherited pleiotropic degenerative disorder that can affect muscle, bone, and the nervous system, including Paget's disease of the bone (PDB), fronto-temporal dementia (FTD), limb-girdle muscular dystrophy, and amyotrophic lateral sclerosis (ALS) (Emery, 2002; Taylor, 2015). The name multisystem proteinopathy (MSP) has been adopted to describe these disease collectively, emphasizing their histopathological character. Some MSPs have been associated with mutations in valosin-containing protein (VCP), an AAA+ family ubiquitin-dependent segregase (Watts et al., 2004). The Taylor group studied clinically indistinguishable families with VCP-related MSP, and found certain families that have no variation in their VCP gene (Kim et al., 2013).

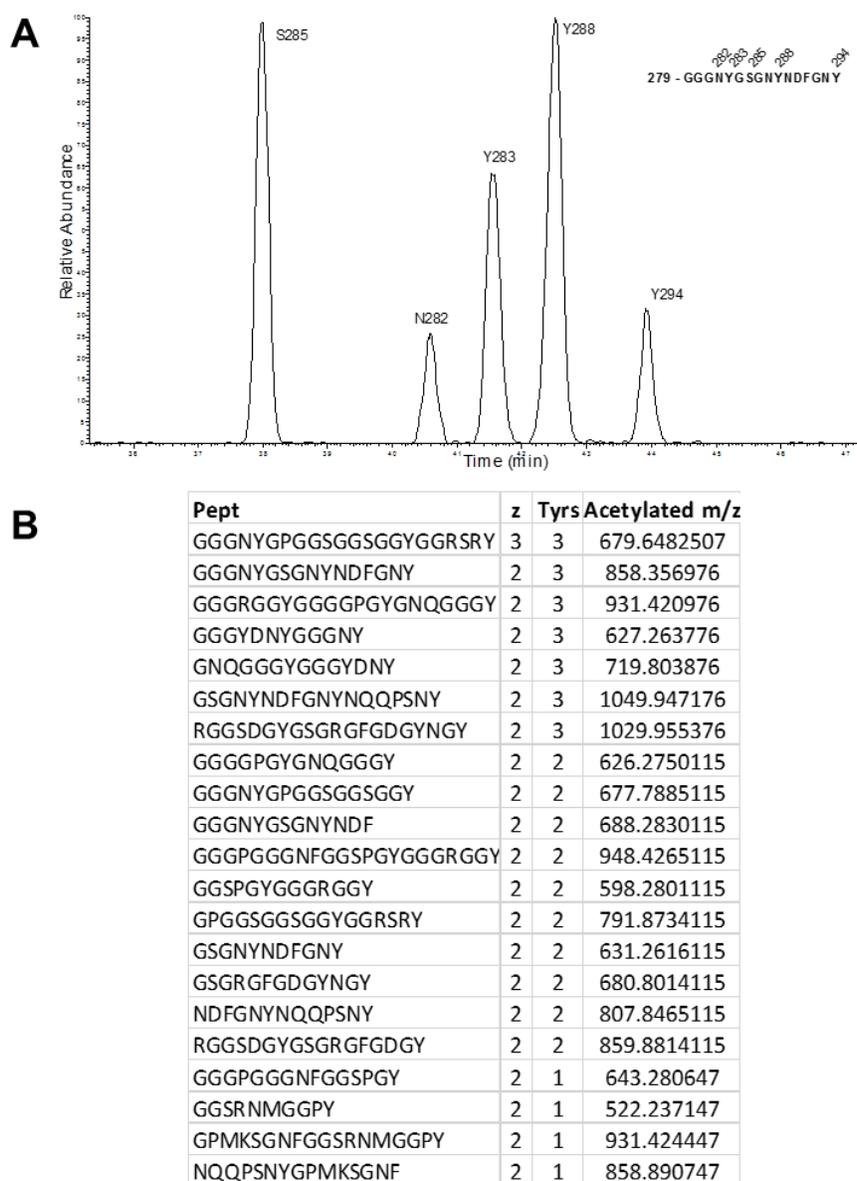
To study the genetic basis of those VCP-like MSP, Taylor group performed exome sequencing and linkage analysis on affected families. From three different families, the authors identified mutations D290V in the gene encoding hnRNPA2B1, and D262V, D262N in the gene encoding hnRNPA1 (Figure 2-12A). Soon after, a separate group discovered a similar aspartate to valine mutation in hnRNPD1 as causative mutation of a

familial case of limb-girdle muscular dystrophies (Vieira et al., 2014). The three hnRNPs are paralogues, and in all four cases the same highly conserved aspartate residue was found to be mutated (Figure 2-12B).

I tested the stabilities of hnRNP LC domains carrying MSP associated aspartate to valine mutations, and compared them with the wild type proteins. Inducible bacteria expression constructs with His<sub>6</sub>-tagged wild type or mutant LC sequences of hnRNPA2, hnRNPA1 and hnRNPD1 were transformed to BL21 cells, induced, and extracted by buffer containing 6 M guanidine HCl. The over-expressed LC proteins were purified by Ni-NTA column in 6 M guanidine HCl, diluted to 0.1 mg/ml with 8 M urea, and dialyzed against gelation buffer to form fibers. The fibers were sonicated and incubated with 0.1%, 0.5%, 1%, 2% SDS or no SDS, and fiber stabilities in SDS containing buffer were assayed by agarose gel electrophoresis followed by western blotting against His<sub>6</sub>-tag. All three wild type LC domains migrated as monomers, even with no SDS treatment prior to gel loading. By contrast, only a small portion of fibers formed by hnRNPA2 D290V started to dissociate after incubation with 2% SDS, and no fiber depolymerization was observed for mutated LC sequence from hnRNPA1 and hnRNPD1, even under the highest SDS concentration tested (2%). I conclude that the aspartate to valine mutations in the LC domain of hnRNPs stabilize their polymeric fibers, and make the fibers insensitive to SDS.

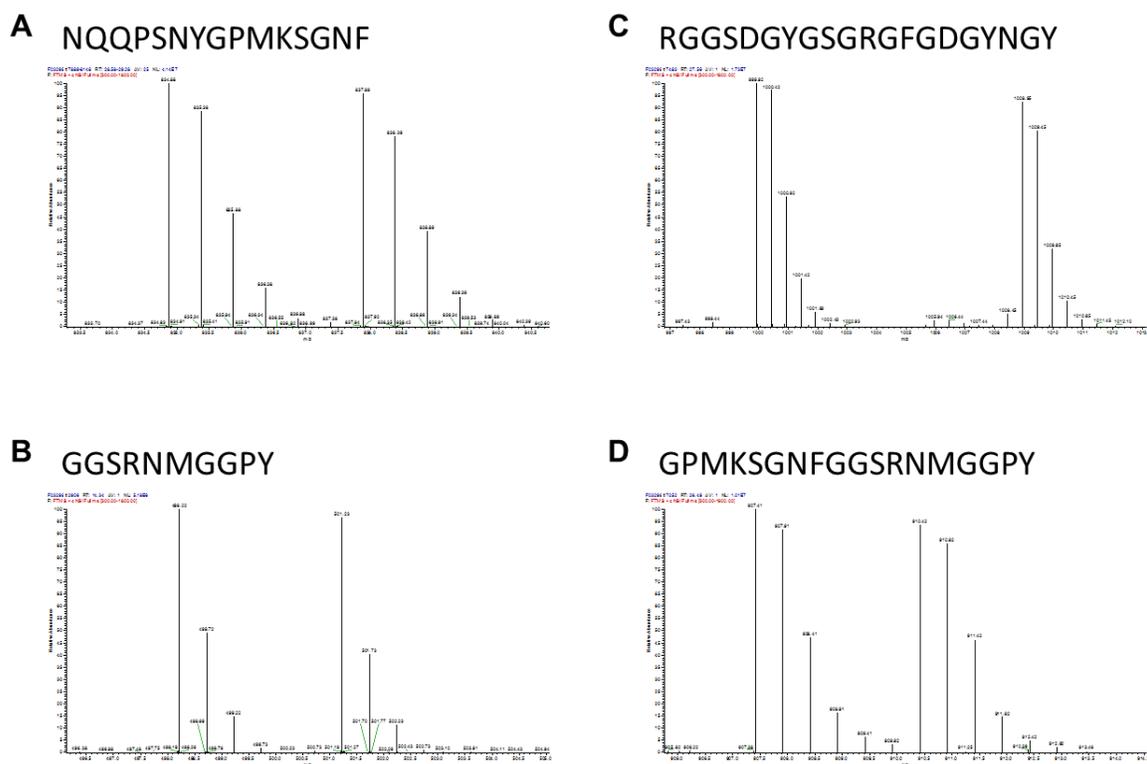
The identification of these aspartate mutations in familial MSP are intriguing. First, aspartate to valine mutations in all three gene products stabilize their polymeric fibers, and fibers formed by the mutated LC domains are no longer SDS sensitive (Figure 2-12B). Second, aspartate residues are the least likely amino acid to be in a  $\beta$ -strand, while valine

residue has the highest  $\beta$  propensity of all twenty amino acids (Koehl and Levitt, 1999). Evolution has clearly favored aspartate, the amino acid that least likely to form a  $\beta$  structure, in the native LC sequence of each of the three different hnRNPs. In multiple cases, the disease-causing mutation changes this aspartate to valine, which favors  $\beta$  structure (Costantini et al., 2006; Koehl and Levitt, 1999). Finally, the aspartate residues sit in the center of the NAI protected region of hnRNPA2 LC. The immediately adjacent phenylalanine residue 291 was identified as the most important tyrosine/phenylalanine residue for polymeric fiber formation (Figure 2-12). I tentatively conclude that the relevant aspartate residues are of unusual importance in creating the proper balance of cross- $\beta$  polymerization:depolymerization.



**Figure 2-6. Chromatographic Separation and Acetylation Site Identification of hnRNPA2 LC**

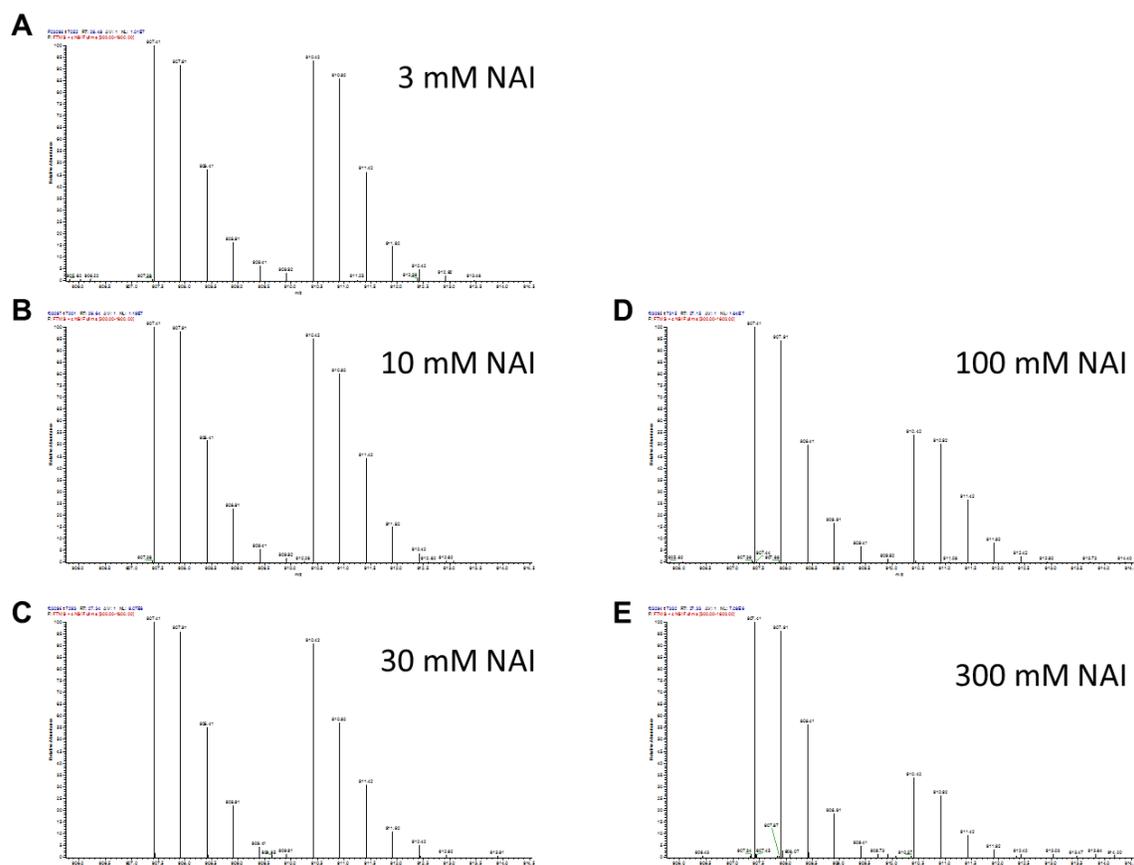
- (A) Elution profile extracted from LC-MS/MS data. Position isomers of acetylated peptides from hnRNPA2 residues 279-294 (GGGNYGSGNYNDFGNY). The amino acid side chain acetylated are labeled on top of each peak.
- (B) Chymotryptic fragments targeted for fragmentation and MS2 analysis. m/z values listed are theoretical m/z for heavy tyrosine containing peptides that acetylated on one single side chain.



**Figure 2-7. MS1 Spectra of Native Peptides from Acetylated hnRNPA2**

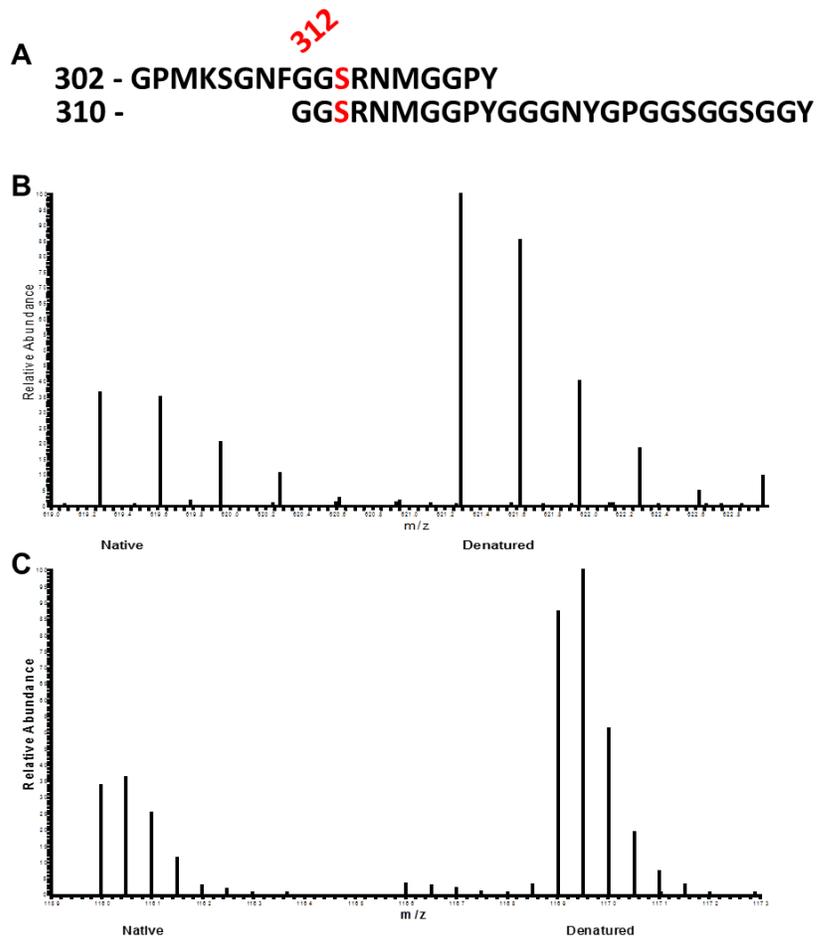
Acetylated heavy and light hnRNPA2 were digested by chymotrypsin and analyzed by LC-MS/MS. Heavy-to-light ratio were observed 1:1 for all four chymotryptic fragments, suggesting a consistent chymotrypsin digestion on both heavy and light parts of the sample. SILAC pairs were shown from

- (A) NQQPSNYGPMKSGNF,
- (B) GGSRNMGGPY,
- (C) RGGSDGYGSGRGFVDGYNGY, and
- (D) GPMKSGNFGGSRNMGGPY.



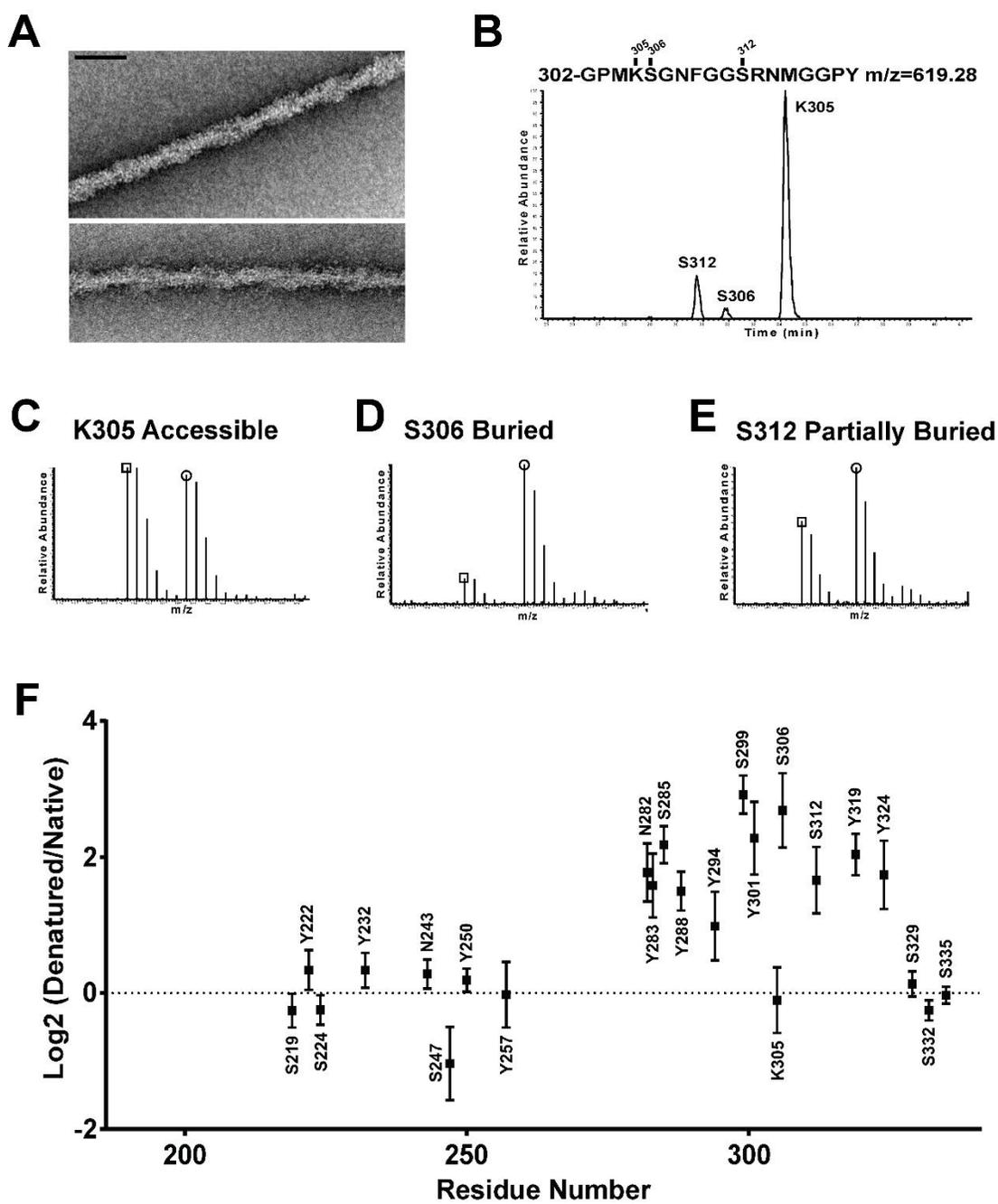
**Figure 2-8. Quantification of Native Peptides from SILAC Samples**

Protein samples were exposed to 3 (A), 10 (B), 30 (C), 100 (D) or 300 (E) mM NAI for 15 minutes at room temperature, then quenched with TRIS (Experimental Procedures). The heavy sample was denatured with guanidine thiocyanate prior to NAI-mediated modification, followed by quenching with TRIS. The two samples were combined at a 1:1 ratio, fully denatured by heating in guanidine thiocyanate, digested with chymotrypsin in presence of 2 M urea, and then analyzed by mass spectrometry. The SILAC pairs from native peptide GPMKSGNFGGSRNMGGPY of all five samples were shown. Similar amount of heavy and light forms of this peptide were observed in samples treated with 3, 10 or 30 mM NAI, while the heavy-to-light ratio decreased in samples treated with 100 mM NAI or more.



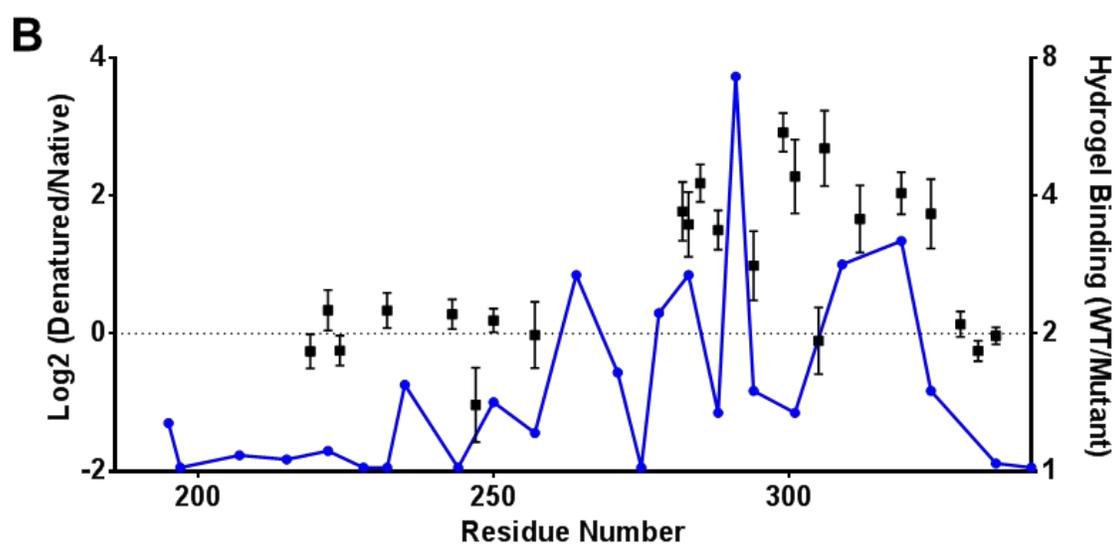
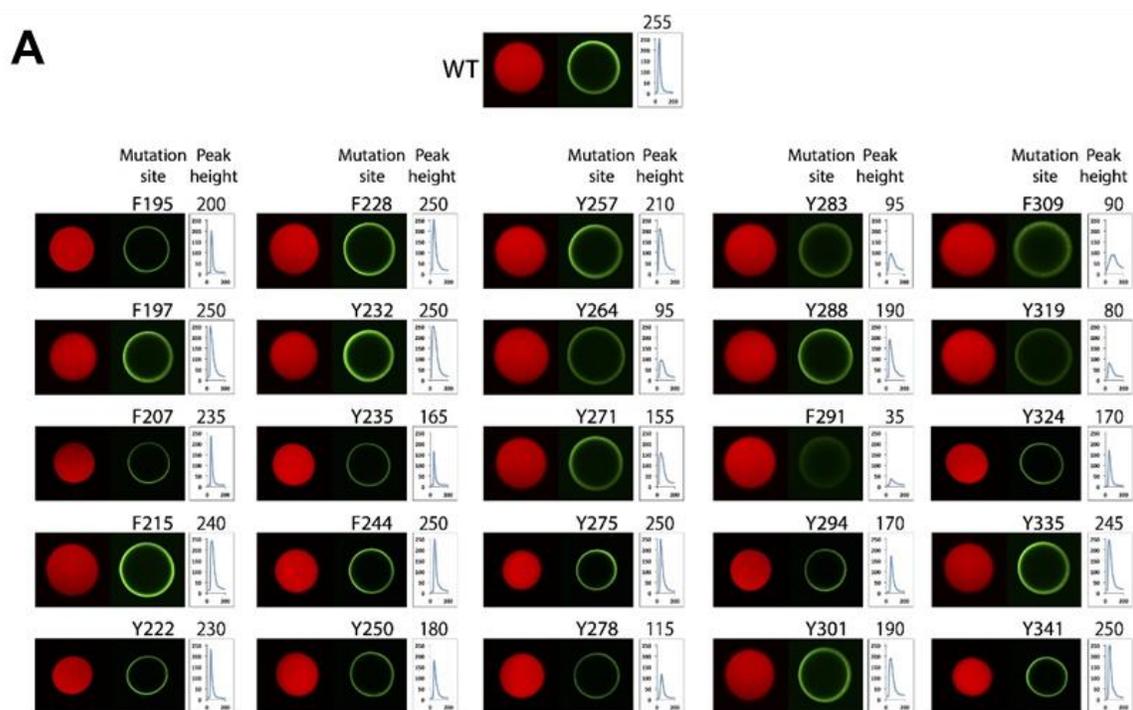
**Figure 2-9. Mass Spectra of Peptides Acetylated on the Same Serine Residue**

- (A) Two peptides generated by chymotrypsin miss-cleavage, with the same serine residue 312 acetylated.
- (B) SILAC pair of peptide GPMKSGNFGGS\*RNMGGPY (asterisk showing the amino acid side chain acetylated).
- (C) SILAC pair of peptide GGS\*RNMGGPYGGGNYGPGGSGGSGGY.



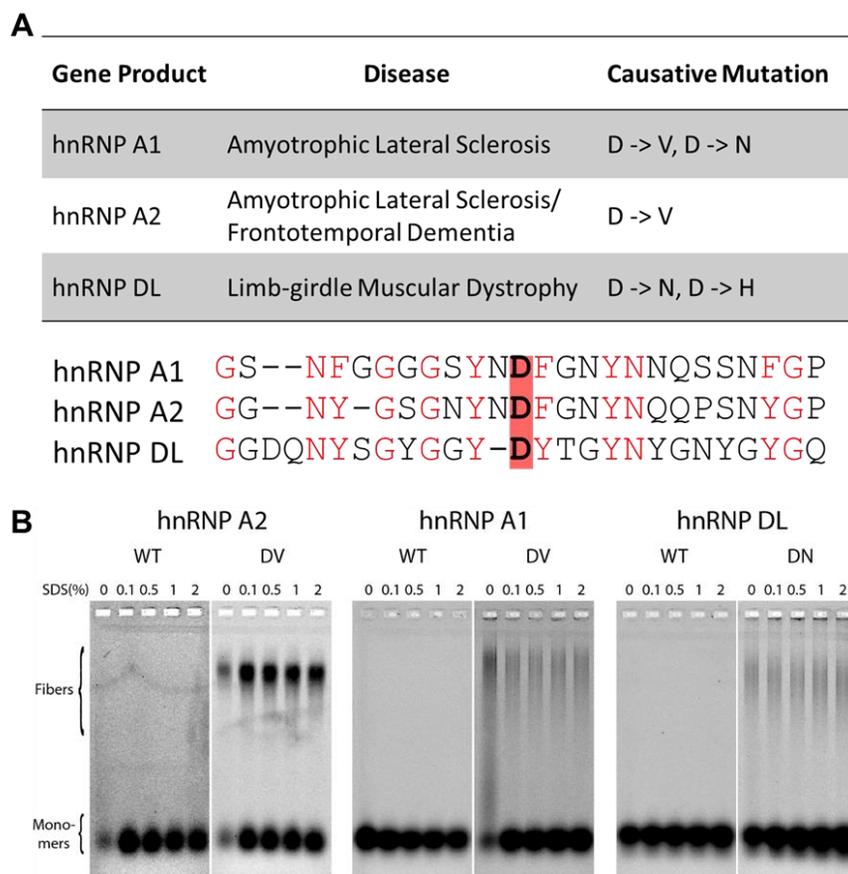
**Figure 2-10. Footprint of NAI-mediated Acetylation of Recombinant hnRNPA2 Polymeric Fibers**

- (A) Electron micrographs of negatively stained polymeric fibers formed from an mCherry:hnRNPA2 LC fusion protein. Scale bar: 70 nm.
- (B) HPLC separation of chymotryptic digestion products of the LC domain of hnRNPA2 corresponding to residues 302-319. The S312 acetylated peptide eluted earlier from the column than the S306 acetylated peptide, which – in turn – eluted earlier than the K305 acetylated peptide.
- (C) Relative abundances of the K305 acetylated peptides in folded versus denatured samples.
- (D) Relative abundances of the S306 acetylated peptides in folded versus denatured samples.
- (E) Relative abundances of the S312 acetylated peptides in folded versus denatured samples.
- (F) NAI footprint of the LC domain of hnRNPA2 (All data are presented as means  $\pm$  SD).



**Figure 2-11. Correlative Relationship between Binding of Mutated Variants of the LC Domain of hnRNPA2 to NAI Accessibility**

- (A) All phenylalanine and tyrosine residues within the LC domain of hnRNPA2 were individually mutated to serine, expressed as GFP fusion proteins, purified and tested for binding to mCherry:hnRNPA2 LC hydrogel droplets. Top figures show images of hydrogel binding by GFP linked to the native LC domain of hnRNPA2 (wild type). Confocal images were scanned to yield the signal intensity of bound GFP, yielding the 25 scans on the right side of each hydrogel image. X-axis indicates the scanned distance in  $\mu\text{m}$ , and Y-axis indicates the GFP signal intensity in arbitrary units.
- (B) Comparison of NAI footprint to hydrogel binding. The black squares with error bars are the NAI footprint, plotted with Log<sub>2</sub> scale on the left Y-axis. The hydrogel retention were plotted as blue curve as ratios of hydrogel binding of wild type hnRNPA2 LC over the binding of mutants, on the right Y-axis.



**Figure 2-12. Genetic Mutations Associate with Familial MSP.**

- (A) The conserved aspartate residue in LC domains of hnRNPA2, hnRNPA1 or hnRNPD1 were mutated to valine or asparagine in families with MSP.
- (B) Semi-denaturing detergent agarose gel electrophoresis showing that hnRNP LC domains with MSP causative aspartate to valine mutations are insensitive to SDS.

**Table 2-3. Footprint Deduced by NAI-Mediated Acetylation of hnRNPA2.**

Residue Number	mCherry:hnRNPA2 Hydrogel	Nuclei	PPIA WT	PPIA R55A	MBP:PTB:LC no TEV	PTB:LC Droplet 10 min	PTB:LC Droplet 2 hours	PTB:LC Droplet 18 hours
S219	-0.26	-0.11	-0.05	0.09	0.13	0.22	0.19	0.21
Y222	0.34	0.09	0.55	0.59	0.25	-0.12	0.44	0.13
S224	-0.25							
Y232	0.34	0.21	0.27	0.14	0.05	0.25	0.15	0.28
N243	0.28							
S247	-1.03	-0.15	-0.29	-0.13	0.32	-0.03	0.19	-0.08
Y250	0.19	0.01	0.26	0.28	-0.13	0.28	0.22	0.45
Y257	-0.02							
N282	1.77							
Y283	1.59	0.46	1.52	1.52	0.35	1.08	1.25	1.63
S285	2.18	0.50	2.30	2.56	0.65	1.52	1.89	2.32
Y288	1.50	0.61	1.79	1.77	0.46	1.03	1.58	1.66
Y294	0.98	0.33	1.56	1.63	0.48	0.84	1.23	1.32
S299	2.92	0.32	2.82	2.63	0.74	1.05	2.68	2.54
Y301	2.28	0.51	2.10	2.00	0.32	1.08	1.88	2.31
K305	-0.10	0.02	-0.03	-0.01	0.1	-0.05	-0.14	0.08
S306	2.69							
S312	1.66	0.49	1.54	1.63	0.43	1.02	1.2	1.72
Y319	2.04	0.57	1.80	1.65	0.65	1.33	1.58	1.65
Y324	1.74	-0.16	-0.26	1.74	-0.18	-0.28	0.13	0.05
S329	0.14	-0.04	0.03	0.04	0.08	0.14	0.08	0.17
S332	-0.25	-0.04	-0.01	0.02	0.05	-0.07	-0.03	-0.18
Y335	-0.03	-0.07	0.03	0.02	-0.18	-0.12	0.19	0.2

## **CORRELATIVE RELATIONSHIP OF HNRNPA2 FOOTPRINTS BETWEEN RECOMBINANT AND NUCLEAR FORMS OF THE PROTEIN**

### **NAI Footprints of Endogenous hnRNPA2 from Mammalian Cell Nuclei**

Using a similar methods described for determining an NAI footprint for the nuclear form of the PARP enzyme (Figure 2-5), I probed the structure of native hnRNPA2 present in nuclei freshly prepared from 293T cells (Experimental Procedures). Briefly, nuclei from  $2 \times 10^8$  isotopically labeled heavy cells was probed under the denaturing conditions of 5 M guanidine thiocyanate. The same number of light nuclei were probed via exposure to the NAI chemical reagent. Following quenching with TRIS, the samples were combined, heated, sonicated briefly, and filtered through a 100 kDa membrane to remove DNA. The flow-through was concentrated by regenerated cellulose filter (Amicon Ultra) with 30 kDa cutoff, digested with chymotrypsin overnight, and desalted on a C18 column. The resulting peptides were fractionated using an off-line two dimensional RP-RP-HPLC (basic reverse phase coupled with acidic reverse phase) protocol (Gilar et al., 2005) and eluted material was assayed by tandem mass spectrometry. The fold differences of acetylated peptides in their folded versus denatured state were normalized by abundances of the corresponding native peptides with no modification, and plotted on Log<sub>2</sub> scale (Figure 2-13).

The NAI footprint observed for native, nuclear hnRNPA2 could be scored for 18 of the 23 acetylated residues observed in the footprint derived from recombinant hnRNPA2, and the two footprints were qualitatively similar (Figure 2-13). Of the acetylation events detected

in both footprints, all nine residues that were equally accessible to NAI-mediated acetylation in both polymeric and denatured samples of recombinant hnRNPA2 protein were also acetylated equally in the native hnRNPA2 irrespective of structural state. Seven of the eight residues that were preferentially protected from acetylation as a function of the fibrous state of recombinant hnRNPA2 protein were also preferentially protected in the native hnRNPA2 protein relative to nuclear protein that had been denatured with 5 M guanidine thiocyanate. The single qualitative difference between the two footprints was tyrosine residue 324. This residue was preferentially protected from NAI-mediated acetylation in the fibrous form of recombinant hnRNPA2, yet was equally accessible to the chemical probe in native hnRNPA2 irrespective of whether nuclei were left intact or denatured.

Despite displaying qualitative similarities, the NAI-generated footprints for recombinant and native hnRNPA2 differed quantitatively in a consistent manner. The NAI protected residues observed in recombinant hnRNPA2 yielded an average of roughly 3-fold ( $\log_2 \sim 1.8$ ) difference when comparing peptide abundance in the light (fibrous) and heavy (denatured) samples. Turning to the native hnRNPA2 assayed in either intact or denatured nuclei, the average difference in peptides revealing NAI protected residues was roughly 1.5-fold ( $\log_2 \sim 0.5$ ). Interpreted most simply, this difference gives indication that a smaller fraction of the native hnRNPA2 present in nuclei may exist in the structurally ordered state than the fraction deduced by studies of recombinant hnRNPA2 polymeric fibers.

## **Co-expression of hnRNPA2 with Peptidyl-prolyl Cis-trans Isomerase Causes Tyrosine 324 to Become NAI-accessible in Recombinant Polymers**

The NAI footprint observed in recombinant hnRNPA2 polymeric fibers was qualitatively similar to that observed for native hnRNPA2 in intact nuclei. Among 18 residues defining the footprint, tyrosine 324 was the single amino acid that was clearly different in the two samples. This residue was protected from NAI-mediated acetylation in fibrous preparations of recombinant hnRNPA2, but not in the native hnRNPA2 present in intact nuclei.

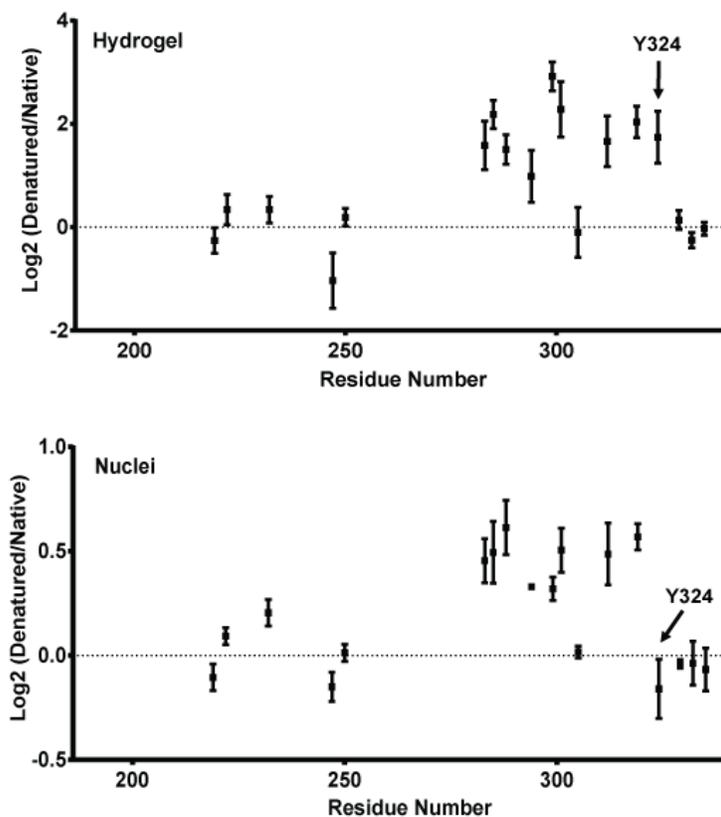
I hypothesized that endogenous hnRNPA2 was modified by nuclear enzymes which affect the structure of hnRNPA2 LC in mammalian cell nuclei. To identify proteins associate with LC domain of hnRNPA2, I cultured heavy cells in DMEM media with lysine and arginine uniformly labeled with  $^{13}\text{C}$  and  $^{15}\text{N}$ . The cell lysate from heavy amino acids labeled cells were prepared in lysis buffer containing 0.1% Triton X-100 (25 mM HEPES pH 7.4, 150 mM NaCl, 0.1% Triton X-100 and protease inhibitor cocktail) by sonication, and incubated overnight with hydrogel droplets formed by mCherry:hnRNPA2 LC. The hydrogel was then washed three times with lysis buffer, and dissolved with 6 M guanidine hydrochloride. The mCherry:hnRNPA2 LC forming the hydrogels was cleared from the solution via tandem Ni-NTA chromatography and the remaining hydrogel binding proteins were analyzed by mass spectrometry. The resulting list of hydrogel binding proteins were sorted by mass spectral count per residue (peptide-spectrum matches divided by number of amino acid in the protein) (Figure 2-14A). The top proteins on the list were cytoskeleton

proteins, followed by hnRNPA2 itself, presumably trapped on the hydrogel by homotypic polymer extension. Following those were single-strand binding protein, Hsp70A and 14-3-3 subunits. The next protein on the list (the 9<sup>th</sup> most abundant protein trapped in the hydrogel, highlighted), is PPIA, the most abundant isoform of a family of peptidyl-prolyl cis-trans isomerase enzymes. PPIA stood out for two reasons: first, it has enzymatic activity that affect conformation of peptide bond before proline residue; second, proline residues are found six positions on the amino terminal side of tyrosine 324, and two positions on its carboxyl terminal side (Figure 2-14B). PPIA has been reported to interact with RNA granule proteins upon biochemical fractionation (Lauranzano et al., 2015), and antibodies to the enzyme revealed co-localization with stress granules marker in U2OS cells after heat shock (Figure 2-14C). I thus reasoned that the PPIA enzyme might affect the structure of hnRNPA2 fibers by facilitating cis-trans interconversion of the peptide bonds of proline residue 319 or 326 of the hnRNPA2 polypeptide.

To test this hypothesis, mCherry:hnRNPA2 LC was co-expressed with either the native form of PPIA or a catalytically inactive mutant (Zydowsky et al., 1992). mCherry:hnRNPA2 LC co-expressed with the active form of PPIA shows better solubility than protein expressed without the enzyme. Moreover, PPIA-exposed protein led to fiber formation that more slowly than mCherry:hnRNPA2 LC expressed alone. Co-expression with the catalytic dead mutant form of PPIA did not affect solubility of hnRNPA2 LC. Following purification of the mCherry:hnRNPA2 protein, polymeric fibers were formed and exposed to the NAI probe under either the polymeric or denatured state. Co-expression of hnRNPA2 LC with the active form of PPIA yielded an NAI footprint wherein tyrosine

residue 324 was equally accessible to acetylation irrespective of fibrous or denatured state (Figure 2-15A). By contrast, co-expression with the catalytically inactive form of PPIA yielded a footprint indistinguishable from that seen for recombinant hnRNPA2 never exposed to the enzyme (Figure 2-15B).

Figures 2-15 C and D correlatively compared the NAI footprints of hnRNPA2 observed in native protein within intact nuclei with that of recombinant protein expressed in either absence or presence of PPIA. The Pearson's correlation coefficient of the native and recombinant footprints was 0.76, which increased to 0.89 when the recombinant hnRNPA2 had been co-expressed with PPIA. The improved correlative relationship between NAI accessibility in fibers formed by mCherry:hnRNPA2 LC co-expressed with PPIA and endogenous hnRNPA2 in cell nuclei suggests that the PPIA enzyme in cell nuclei facilitates folding of hnRNPA2 LC to its native structure.



**Figure 2-13. NAI Footprints of Recombinant and Nuclear hnRNPA2 LC**

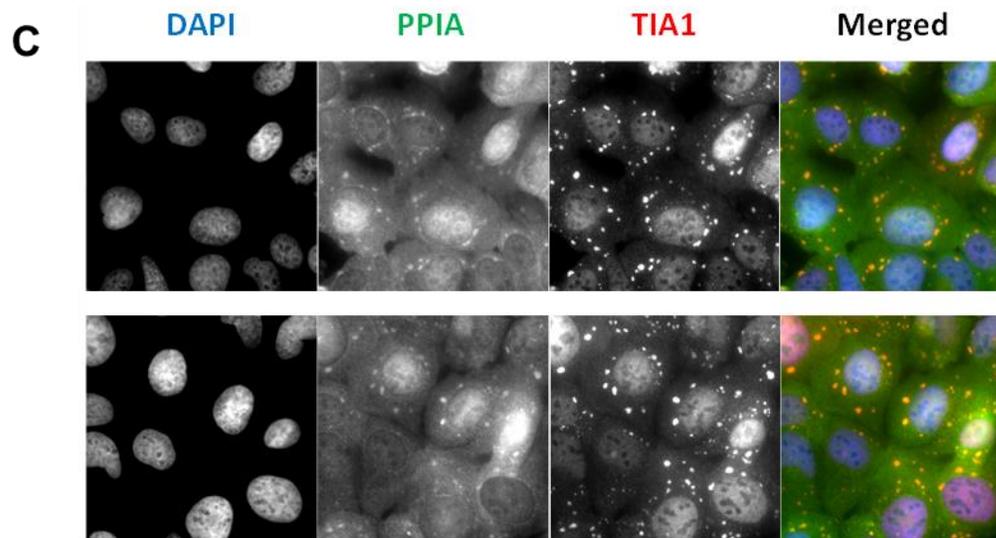
NAI footprint of recombinant hnRNPA2 fibers as described in Figure 2-10 (upper footprint) compared with NAI footprint deduced from native, nuclear hnRNPA2 (lower footprint). Note that tyrosine 324 (arrow) is protected from NAI modification in the folded form of hnRNPA2 in the recombinant form of hnRNPA2, but not in the footprint deduced from the native, nuclear protein. See also Table 2-3.

**A**

Description	Length (AA)	PSMs	% Coverage
ACTG	375	186	72.80
TBA1B	451	163	62.10
ROA2	353	123	61.80
SSBP	148	50	58.80
HSP71	641	166	70.20
1433Z	245	62	66.50
1433E	255	63	81.20
I3L397	147	34	63.90
PPIA	165	35	67.30
BTF3	206	41	75.30
IF4A1	406	74	51.70
PDCD5	125	21	81.60
GRP75	679	114	42.60
C1QBP	282	47	46.10
CH60	573	91	56.70
GRP78	654	100	46.60
1433B	246	37	66.70
F2Z2Y6	67	10	92.50
HNRPK	463	67	63.10

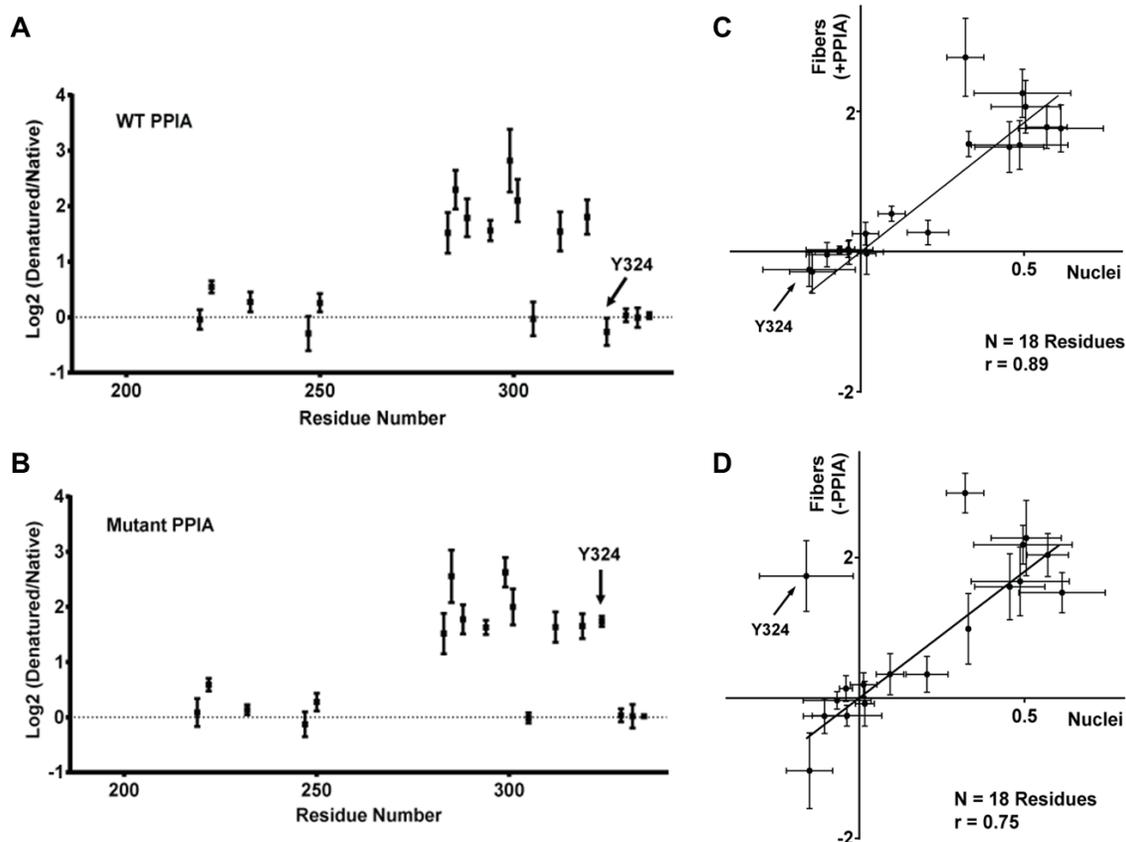
**B**

181-MQEVQSSRSRGGG  
194-NFGFGDSRGGGG  
206-NFGPGPGS  
214-NFRGGSD  
221-GYGSGR  
227-GFGD  
231-GYN  
234-GYGGPGGG  
243-NFGGSP  
249-GYGGGRG  
256-GYGGGGP  
263-GYGNQGG  
270-GYGG  
274-GYD  
277-NYGGG  
282-NYGS  
287-NYN  
290-DFG  
293-NYNQQPS  
300-NYGPMSKSG  
308-NFGGSRNMGG  
318-EYGGG  
323-NYGGGGSGGSG  
334-GYGGRS  
340-RY



**Figure 2-14. PPIA Binds hnRNPA2 and Localizes to Stress Granules**

- (A) List of proteins trapped by mCherry:hnRNPA2 LC hydrogel. Lysate from heavy amino acids (K+6, R+10) labeled 293T cells was incubated with mCherry:hnRNPA2 LC hydrogels (light). The hydrogels were washed and dissolved, and the proteins trapped in hydrogels were identified by mass spectrometry. The protein list was sorted by spectral counts of heavy labeled peptides, normalized by protein size.
- (B) The amino acid sequence of the LC domain of hnRNPA2 (residues 181-341). Tyrosine and phenylalanine residues (in red) are aligned vertically to indicate repetitiveness of these amino acids. Tyrosine residue 324 that became more solvent accessible in the NAI footprinting when the protein was co-expressed with PPIA is highlighted with a green rectangle. Two proline residues neighboring to tyrosine 324 are highlighted with cyan rectangles as potential substrates for PPIA.
- (C) Localization of PPIA in stress granules. U2OS cells were grown on cover slips, heat shocked at 44 °C for 45 minutes, fixed and co-stained with DAPI (blue color), and antibodies to TIA1 (red color) and PPIA (green color). Image shown on right panel merges those of left three panels.



**Figure 2-15. Correlative Relationship between the NAI Footprint of Endogenous hnRNPA2, Recombinant Protein, and Recombinant Protein Co-Expressed with PPIA**

- (A) NAI footprint of recombinant hnRNPA2 LC co-expressed with active PPIA enzyme.  
 (B) NAI footprint of hnRNPA2 LC co-expressed with a catalytically inactive form of the PPIA (R55A). Note that co-expression of hnRNPA2 with the active form of PPIA causes tyrosine 324 to become exposed to NAI modification in the polymeric state.  
 (C and D) Plots showing the correlative relationship of the NAI footprints of recombinant hnRNPA2 LC to that of the native, nuclear form of the protein. Correlation plot (C) compares the footprint of recombinant hnRNPA2 LC not exposed to the PPIA enzyme with the nuclear hnRNPA2 footprint. Correlation plot (D) compares the footprint of recombinant hnRNPA2 LC co-expressed with the active PPIA enzyme with the nuclear hnRNPA2 footprint.

See also Table 2-3.

## **PTB:HNRNPA2 LC FORMS LIQUID-LIKE DROPLET BY LC DOMAIN POLYMERIZATION**

Several groups have described liquid-like behavior of RNA granules, having observed that the granules are round in shape, undergo fusion and fission, and distortion in response to shear forces. Brangwynne et al. applied shear force to *C. elegans*, and studied germline P granules by live cell imaging. It has been shown that the P granules on the periphery of cell nuclei drip down from nuclei and merge with each other (Brangwynne et al., 2009; Brangwynne et al., 2011; Buchan and Parker, 2009; Wang et al., 2014). Fluorescence recovery after photobleaching (FRAP) analyses showed that proteins can move within RNA granules on timescales of seconds, suggesting a liquid-like, highly dynamic structure (Brangwynne et al., 2009; Buchan and Parker, 2009).

A number of investigators have recently reported that LC domains from a variety of proteins, including FUS, hnRNPA1, DDX4 and BuGZ, can prompt formation of liquid-like droplets (Altmeyer et al., 2015; Jiang et al., 2015; Lin et al., 2015; Molliex et al., 2015; Nott et al., 2015; Patel et al., 2015). Liquid-like behavior, such as fission and fusion, and fast movement of proteins, has been well accepted, but less attention has been paid to the mechanism of phase separation. It is of potential importance to know whether the physical forces leading to hydrogel formation (polymerization of LC domains) are the same or different from those leading to liquid like droplets.

## **Mutations in the LC Domain of hnRNPA2 Act Correlatively on Hydrogel Binding and Partitioning into Liquid-like Droplets**

Parker and Rosen groups created RNA-dependent liquid-like droplets driven by the LC domain to mimic the context of RNA granules (Lin et al., 2015). A triple fusion protein was prepared linking the LC domain of a variety of RNA binding proteins on the C-terminal side of the PTB RNA binding domain, which was linked to maltose binding protein (MBP), with a TEV protease cleavage site between the MBP and PTB domains (Figure 2-16A). The MBP:PTB:LC domain fusion further contained a His<sub>6</sub> tag at its C-terminus as well as a SNAP tag for dye-labeling on the N-terminal side of the PTB domain. Upon TEV cleavage and addition of RNA substrate, all six SNAP:PTB:LC fusion proteins tested were observed to phase separate at protein concentrations of 1.25–2.5 μM (Figure 2-16B), producing liquid-like droplets that containing concentrated levels of both protein and RNA. As a control, SNAP:PTB with no LC sequence only phase separated at a 50 μM concentration in the presence of RNA. Liquid-like droplets formed by PTB:LC are round in shape, and have FRAP recovery times of 10-180 seconds. The droplets mature over time, becoming irregular in shape and less dynamic after 24 hours incubation, with filamentous structures extending outside of the droplet bodies (Lin et al., 2015). Parker and Rosen groups proposed that LC domains drive liquid-like droplets formation with relatively low specificity, and the high local concentration of LC sequences in droplets facilitates polymeric fiber formation in absence of proper regulatory factors.

I followed the same procedure of Parker and Rosen groups and created liquid-like droplets driven by the LC domain of hnRNPA2. Following co-expression with PPIA, purification via Ni-NTA column and amylose resin chromatography, the fusion protein of MBP:PTB:hnRNPA2 LC was mixed with a synthetic RNA containing five copies of PTB binding site and exposed to TEV protease. Within ten minutes liquid-like droplets could be observed by light microscopy (Figure 2-17A). I then deployed a droplet partitioning assay to assess whether eGFP:hnRNPA2 LC domain fusion proteins could be incorporated into the liquid-like droplets. Wild type PTB:hnRNPA2 LC was mixed with the repetitive RNA substrate and TEV protease, and incubated at room temperature for 20 minutes allowing liquid-like droplets to form and merge together. 0.1  $\mu$ M of eGFP linked wild-type or mutant hnRNPA2 LC proteins were added to the droplets suspension formed by 5  $\mu$ M PTB:hnRNPA2 LC, and incubated between cover slides for 5 minutes. Liquid-like droplets partitioning for each eGFP:hnRNPA2 LC variant was then assayed by fluorescent microscopy. The partition ratio of eGFP:hnRNPA2 protein was calculated by dividing the signal inside the droplet by the signal outside. Partition ratios were then normalized with the Alexa 549 signal from PTB:hnRNPA2 LC protein in the droplet, to eliminate effect of size differences of the droplets. Recombinant eGFP-alone protein did not partition into these liquid-like droplets relative to the surrounding buffer. The eGFP fusion linked to the wild type LC domain of hnRNPA2 was rapidly incorporated into liquid-like droplets. Mutation F215S, which does not affect hydrogel binding, showed no detectable difference in liquid-like droplets partitioning comparing with wild type protein. In experiments with the other two mutants shown in Figure 2-17A, Y271S, which showed 40% decrease in hydrogel retention,

partitioned less readily in liquid-like droplets; while the mutation that affect hydrogel binding most acutely, F291S, was barely recruited into liquid-like droplets at all. Using this assay I evaluated liquid-like droplets partitioning of all 25 mutants that had been scored for hydrogel binding (Figures 2-17B).

Six mutants were impeded by more than 50% with respect to partitioning into liquid like droplets (Y257S, Y264S, Y278S, F291S, F309S and Y319S), another eight mutants were partially impeded (F195S, F207S, Y235S, Y250S, Y283S, Y288S Y294S and Y301S), and the remaining mutants were incorporated into liquid-like droplets in a manner indistinguishable from the native LC domain (Figure 2-17B). The correlation plot shown in Figure 2-17C gives evidence of a strong concordance ( $r = 0.83$ ) between the effects of mutations on hydrogel binding and partitioning into liquid-like droplets. I offer that this concordance gives evidence that similar regions of the protein promote both hydrogel binding and partitioning into liquid-like droplets, and that the chemical interactions that drive both processes are likely to be the same.

### **Liquid-like Droplets Display the NAI Footprint Found in Hydrogel Polymers and Nuclear hnRNPA2**

If the mutational effects driving hydrogel binding and liquid-like droplets correlate, it is possible that the LC domain of hnRNPA2 might adopt similar structures in both states. To address this question I performed NAI footprinting on the LC domain of hnRNPA2 in the context of the MBP:PTB:hnRNP LC domain fusion protein before TEV cleavage, ten

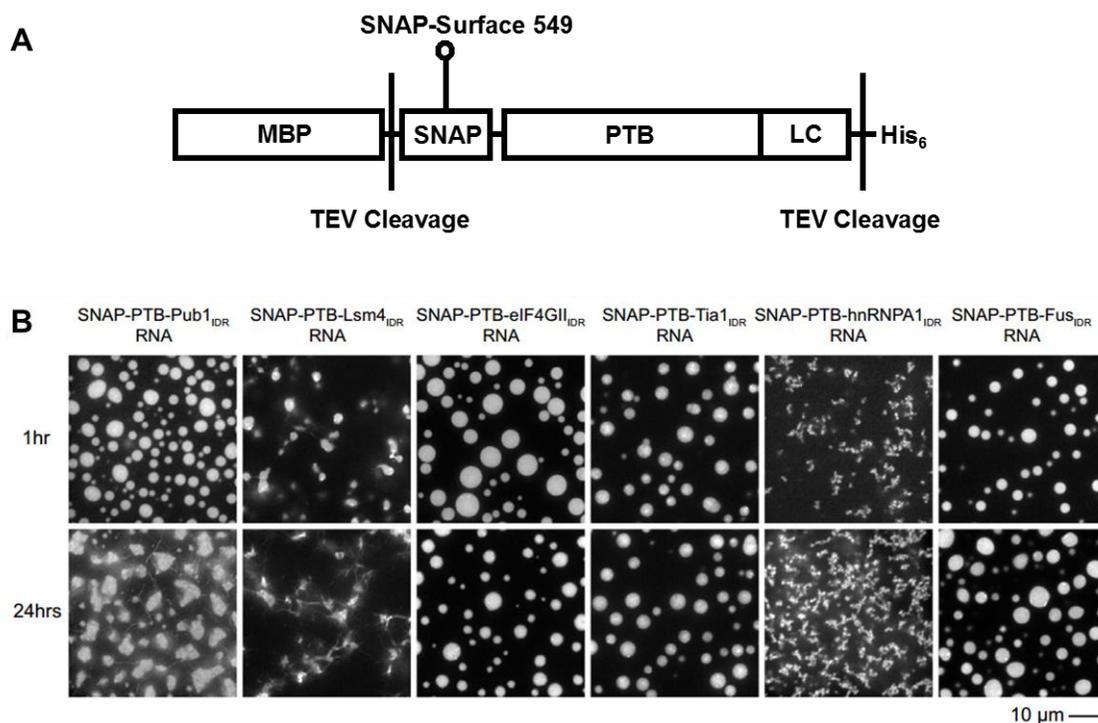
minutes after TEV cleavage, two hours after TEV cleavage, and 18 hours after TEV cleavage. The same mixture of hnRNPA2 LC, TEV protease and RNA substrate was exposed to NAI in the presence of guanidine thiocyanate as a control of reactivity of amino acid side chains. The MBP:PTB:hnRNP LC was ultracentrifuged at 100,000 x g for one hour prior to the experiment to remove potential aggregates. The protein solution was checked under fluorescent microscope, and no aggregates or fibers were detected. Microscopic liquid-like droplets formed 10 minutes after addition of TEV and RNA substrate. The droplets merged with each other, grew bigger in size after incubation for two hours, and became irregular in shape after 18 hours of incubation, with fibrous extension clearly visible on the periphery of the droplets.

As shown in Figure 2-18, evidence of the canonical NAI footprint on the hnRNPA2 LC domain could be detected even before TEV protease cleavage. The quantitative intensity of the footprint became sequentially enhanced at each of the later time points. Specifically, the degree of difference in NAI protected residues between native and denatured samples was – across all protected residues – most pronounced in the 18 hour sample, less so in the two hour sample, further reduced in the 10 minute sample, and least pronounced in the sample assayed prior to TEV protease cleavage. For the 18 hour time point, the degree of protection from NAI-mediated acetylation of buried side chains was indistinguishable between liquid-like droplets and hydrogels. A simplified view of these footprints can be described as an averaging result of a mixture of fully polymerized amyloid-like fibers with the same NAI footprint as hydrogels and random coil which was solvent accessible for all amino acid side chains. Portions of PTB:hnRNPA2 LC side chains that were in a fiber-like

state relative to random coil were calculated based on NAI accessibility of each residues, and the ratio of PTB:hnRNPA2 LC proteins in their polymeric state were plotted as an average of all 18 residues analyzed (Figure 2-18C). About 45% of LC proteins were in fibrous state in protein solution prior to TEV cleavage, this number increase dramatically to close to 80% within 10 minutes of addition of TEV protease and RNA substrate. About 90% of the protein was estimated to adopt a polymeric structure after two-hour incubation, when the droplets formed by SNAP:PTB:hnRNPA2 LC remained round and liquid-like. When the droplets matured after 18 hours of incubation, the NAI accessibility reveal no difference to that deduced from hydrogel samples, and more than 99% of LC sequences were estimated to adopt amyloid-like fiber structure.

Concordant with the observations of others who have studied LC domain partitioning to liquid-like droplets (Jiang et al., 2015; Lin et al., 2015; Molliex et al., 2015; Patel et al., 2015), I conclude that as a function of time, LC polymerization is progressively enhanced within liquid-like droplets. Gorlich and colleagues have reported similar observations as a function of maturation of liquid-like droplets formed from FG repeats associated with nucleoporin proteins (Petri et al., 2012). In summary, mutational studies of the LC domain of hnRNPA2 give evidence that similar forces drive both hydrogel retention and partitioning into liquid-like droplets, and NAI footprinting studies reveal evidence that the LC domain of hnRNPA2 adopts a similar structure in both settings. Thus I propose the following mechanistic model of liquid-like droplets formation (Figure 2-19). The fusion protein of MBP:PTB:hnRNPA2 LC remains soluble and partially polymerized via the LC domain (red sheets) prior to TEV cleavage and exposure to synthetic RNA. Following TEV cleavage and

exposure to synthetic RNA substrate containing five PTB binding sites (yellow rectangle), MBP is left in solution and PTB:hnRNPA2 LC fusion protein partitions into liquid-like droplet (grey shading) in a state of enhanced polymerization; prolonged incubation then yields longer polymeric fibers in the liquid-like droplets, eventually leading to solidification of the droplets via profound fiber polymerization.

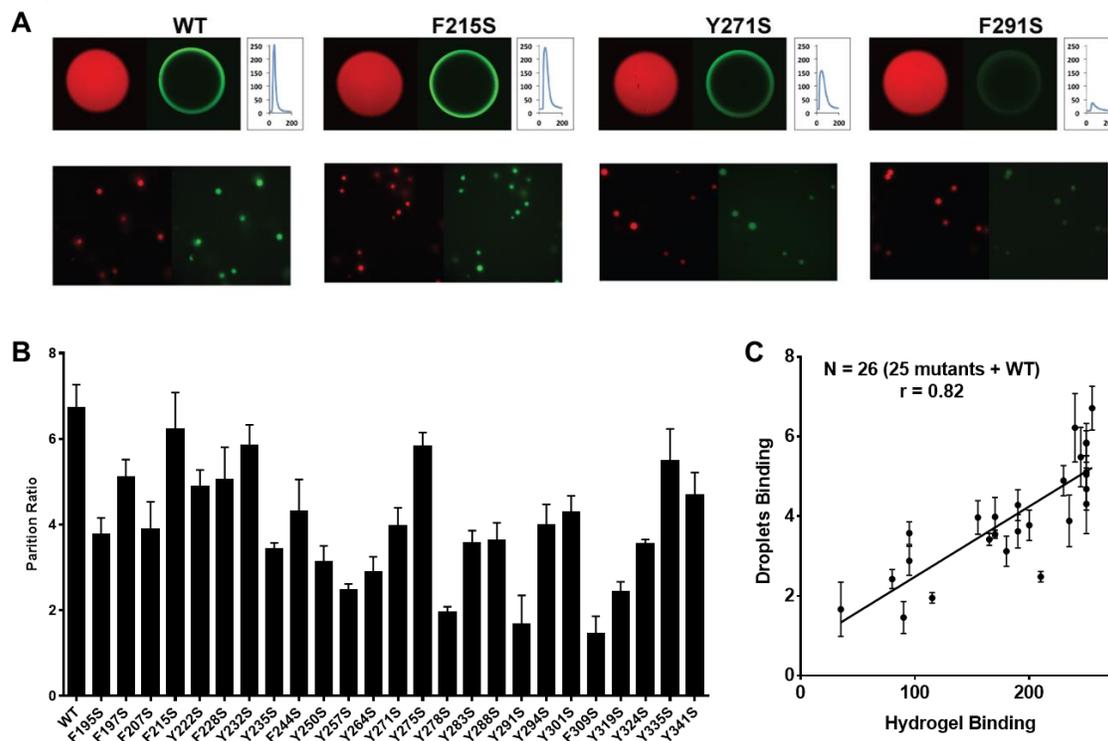


**Figure 2-16. Liquid-Like Droplet Formation Driven by LC Domains**

(A) A diagram showing the domain structure of the MBP:PTB:LC construct. Removable MBP was added to the amino terminus to keep proteins from aggregating. The presence of a SNAP-tag allowed the PTB:LC fusion protein to be appended with a red dye.

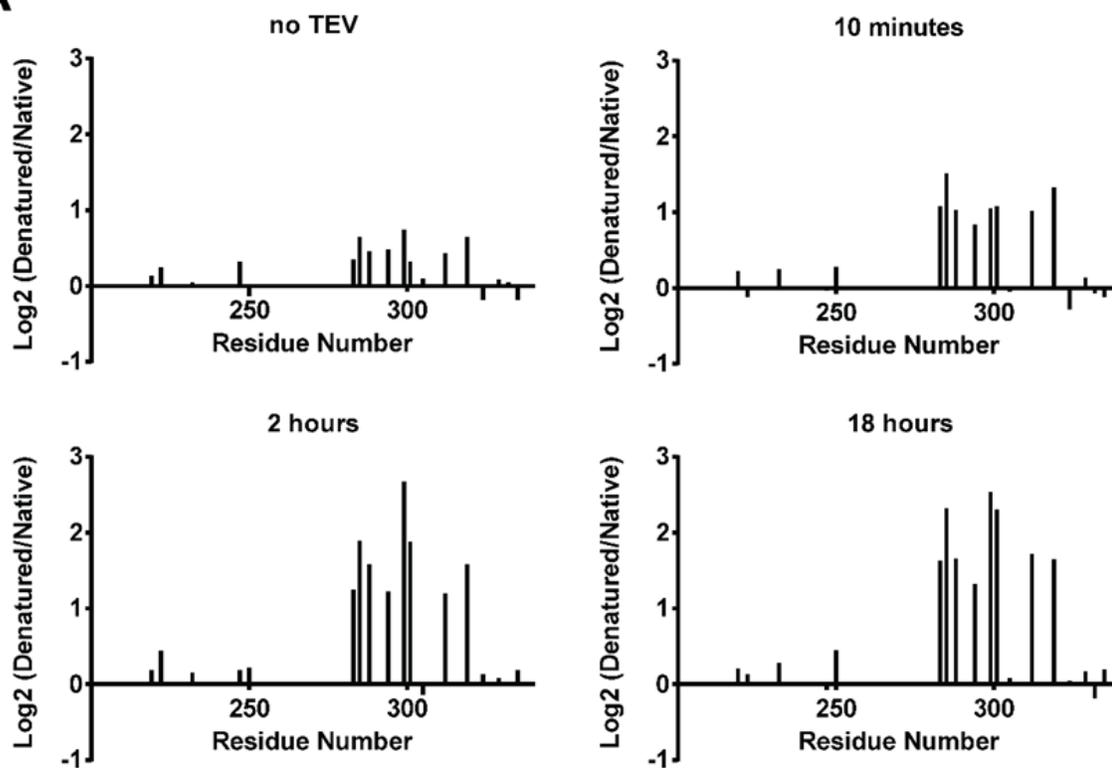
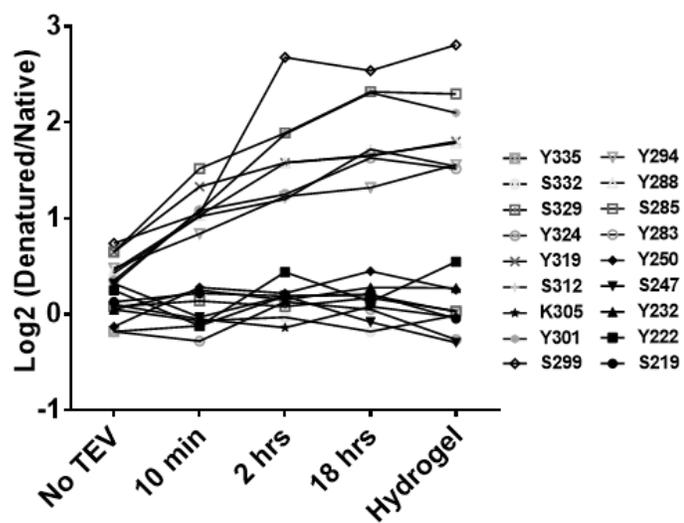
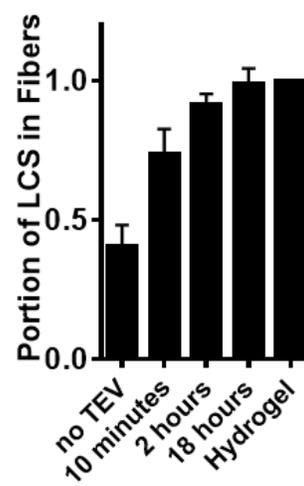
(B) (Lin et al., 2015)

Liquid-like droplets formed upon MBP removal and binding of a PTB:LC fusion protein to a synthetic RNA containing five copies of the PTB recognition sequence. Newly formed liquid-like droplets are round and dynamic. The droplets matured over time, became irregular in shape and less dynamic after 24 hours incubation, with filamentous structures extending outside of the droplet bodies.



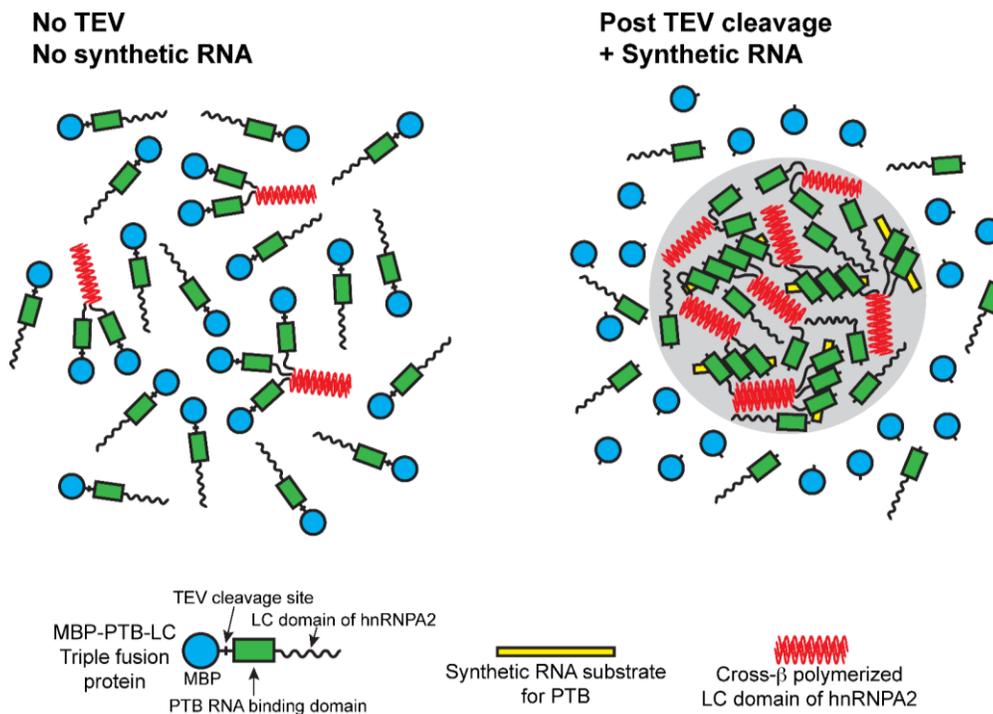
**Figure 2-17. Correlative Relationship between Liquid-Like Droplets Partitioning and Hydrogel Retention**

- (A) Liquid-like droplets formed upon binding of a PTB:hnRNPA2 fusion protein to a synthetic RNA containing five copies of the PTB recognition sequence. When exposed to eGFP alone, no partitioning into liquid-like droplets was observed. When exposed to eGFP fused to the native LC domain of hnRNPA2 (WT), clear evidence of partitioning was observed within minutes. Certain phenylalanine- or tyrosine-to-serine mutants partitioned well into liquid-like droplets (F215S), whereas others did not (Y271 and F291S).
- (B) Partitioning into liquid-like droplets was quantified for all phenylalanine- and tyrosine-to-serine mutants that had been constructed and assayed for binding to mCherry:hnRNPA2 LC hydrogel droplets (Figure 2-11A). Histogram shows relative levels of partitioning of eGFP linked to the native (WT) LC domain of hnRNPA2 as compared with the 25 individual mutants.
- (C) Plot showing the correlative relationship between hydrogel binding and partitioning into liquid-like droplets for eGFP linked to the native (WT) LC domain of hnRNPA2 along with 25 individual phenylalanine- and tyrosine-to-serine mutants.

**A****B****C**

**Figure 2-18. Liquid-Like Droplets Display the Same NAI Footprint as Found in Polymeric Fibers**

- (A) A fusion protein linking maltose binding protein (MBP) to the RNA binding domains of PTB and the LC domain of hnRNPA2 was co-expressed with the peptidyl-prolyl cis-trans isomerase enzyme (PPIA), purified, and mixed with a synthetic RNA containing five PTB binding sites. Addition of TEV protease triggered the rapid formation of liquid-like droplets (Figure 2-17). Protein samples were footprinted with the NAI reagent as a function of time before and after TEV protease cleavage. Hints of the NAI footprint could be seen in the protein sample before exposure to TEV protease, and the intensity of the footprint was sequentially enhanced at the 10 minute and 2 and 18 hours post-cleavage time points.
- (B) The Log<sub>2</sub> ratio of NAI protection for all of the 18 acetylated amino acids is plotted on the Y-axis as a function of time post-exposure to TEV protease (X-axis). See also Data Table 2-3.
- (C) Ratio of PTB:hnRNPA2 LC proteins in their polymeric state were plotted as an average of all 18 residues detected. About 45% of LC proteins were in fibrous state in protein solution prior to TEV cleavage, this number increase dramatically to close to 80% immediately upon seeing the formation of liquid-like droplets. About 90% of the protein were observed in polymeric structure after two-hour incubation, and NAI accessibilities of droplets after 18 hours of incubation reveal no difference to the ones deduced from hydrogel samples.



**Figure 2-19. Graphical Representation of Conversion of Soluble MBP:PTB:hnRNP LC Fusion Protein into Liquid-like Droplet State.**

The triple fusion linking maltose binding protein (MBP = blue circle), the RNA binding domain of pyrimidine track binding protein (PTB = green rectangle), and the low complexity domain of hnRNP2 (LC domain = wavy line) remains soluble and partially polymerized via the LC domain (red sheets) prior to TEV cleavage and exposure to synthetic RNA containing five PTB binding sites (yellow rectangle). Following TEV cleavage and exposure to RNA, MBP is left in solution and PTB:hnRNP LC domain fusion protein partitions into liquid-like droplet (grey shading) in a state of enhanced polymerization.

## **CHAPTER THREE**

### **Discussions and Conclusions**

#### **DISCUSSIONS**

Phase separation of protein solution was first described 70 years ago (Bungenberg de Jong, 1949). Bungenberg de Jong studied liquid-liquid phase separation extensively in a variety of colloidal systems, including aqueous protein solutions. By mixing gelatin solution with gum arabic, the system readily transitioned into a structure with colloid-rich oil phase containing both gelatin and gum arabic, surrounded by a tight skin of water molecules. The system has been extensively studied by physicists and scientists from food industry, and only until late 1980s, the phase separation phenomena drew attention of biochemists. Conditions and mechanism of phase separation of proteins were studied using crystallin and lysozyme (Broide et al., 1991; Muschol and Rosenberger, 1997; Tanaka et al., 1997; Taratuta et al., 1990; Tardieu et al., 1992; Thomson et al., 1987), mainly as a competing side-process to protein crystallization. It has been shown that phase transition of these highly soluble, globular proteins occurs in a temperature- and salt-dependent manner, under control of the weak forces acting in solution between macromolecules. Studies of the phase diagrams of these protein solutions suggested that van der Waals forces could be at the origin of the attractive interaction potentials, and other more specific effects might also contribute to the protein phase diagrams (Malfois et al., 1996).

Recent studies have focused on the role of liquid-liquid phase separation in eukaryotic cells. Hyman group first studied the dynamic properties of *C. elegans* P granules.

Having observed that germ line P granules around cell nuclei dripped down from nuclei and merged with each other, it has been concluded that P granules are liquid-like droplets (Brangwynne et al., 2009). The same group also studied nucleoli with similar live cell imaging approach, and the same liquid-like properties were reported (Brangwynne et al., 2011). Wang et al. studied the regulation of P granule dynamics by phosphorylation (Wang et al., 2014), and demonstrated that maternal-effect germline defective (MEG) -1 and MEG-3, the substrates of the kinase MBK-2/DYRK and the phosphatase PP2A regulates the dynamic properties of P granules. Phosphorylation of the MEGs promotes granule disassembly and dephosphorylation promotes granule assembly. It has also been shown that the P granules, even though behaving in a manner related to liquid-like droplets, are not homogenous oil droplets. GFP-tagged MEG-3 was shown by lattice light sheet microscopy to localize to a dynamic domain that surrounds and penetrates each granule. Furthermore, LC domain of RBM14 (RNA binding protein 14), an essential paraspeckle component, was found to form liquid-like droplets and hydrogels composed of amyloid-like fibers (Hennig et al., 2015). The LC domain of RBM14 was reported to be responsible for targeting these proteins to paraspeckles and was essential for the formation of paraspeckles. More recently, Elbaum-Garfinkle observed that LAF-1, a DDX3 RNA helicase found in P granules, phase separates into P granule-like droplets in vitro, and thus concluded that P granules are liquid-like droplets analogous to LAF-1 droplets (Elbaum-Garfinkle et al., 2015). Similar approaches have been adopted to study the nucleoli of *Xenopus laevis* oocytes with recombinant fibrillarin (Berry et al., 2015), mammalian cell nucleoli with recombinant nucleophosmin (Mitrea et al., 2016) and a reconstituted mitotic spindle matrix with

recombinant BuGZ and *Xenopus* egg extract (Jiang et al., 2015). Two studies investigated liquid-like droplets formed by LC domain of FUS protein. The first paper focused on the FUS droplets formed at DNA damage site, in a poly(ADP) ribose (PAR) dependent manner (Patel et al., 2015), while the second paper studied FUS droplets in neuronal cells (Murakami et al., 2015). Both studies on liquid-like droplets formed by FUS hypothesized that LC domain of FUS exists in a random coil state in liquid-like droplets and in cells, and aberrant formation of amyloid-like fibrils lacking proper regulatory factors is the cause of cytotoxicity. Patel et al. thereby proposed liquid-like droplet after prolonged incubation, which contains polymeric fibers, can be used as a model system to study molecular pathology of aging (Patel et al., 2015).

In considering the mechanism by which liquid-like droplets form, as compared with polymeric fibers or partitioning of LC containing proteins to pre-formed liquid-like droplets, I offer two contrasting perspectives. It is possible that the physical forces leading to the two states are entirely different. Recent studies of the LC domains of DDX4 protein and nucleoporin proteins, characterized by their abundance in phenylalanine residues, favor the utility of weak, non-specific chemical interactions to hold unstructured, random coil LC domains together (Nott et al., 2015; Patel et al., 2015). In presence of DDX4, charge-charge interaction between arginine and phenylalanine residues has been highlighted as a key chemical determinant for phase separation into liquid-like droplets.

My studies favor the contrasting idea that polymerization of LC domains into cross- $\beta$  structure is the driving force for liquid-like droplets partitioning. I reason that the sizes of polymers in hydrogels are much longer than those found in liquid-like droplets, and that the

size distribution and dynamics of LC domains in the latter setting may be a better representation of how LC domains function in living cells. The quantitative intensity of the NAI footprint in the various settings deployed in this study may be instructive in this regard. In hydrogel samples, the degree of NAI protection in ordered regions of the protein was roughly 3X that of denatured samples. In cells, the quantitative degree of protection was roughly 1.5X. The NAI footprint observed in freshly prepared liquid-like droplets yielded a quantitative degree of protection more closely matching to that of native hnRNPA2 as probed in isolated nuclei. Paradoxically, evidence of the existence of a low level of cross- $\beta$  structure was seen in samples of the MBP:PTB:hnRNPA2 LC domain triple fusion protein before TEV release of MBP, before exposure to the synthetic RNA containing iterative PTB binding sites, and before formation of liquid-like droplets. I interpret the observation of an NAI footprint in samples before the formation of liquid-like droplets to reflect the presence of fibril seeds in the soluble sample of MBP:PTB:hnRNPA2 LC. The fusion protein containing hnRNPA2 LC only phase separate when the RNA substrate was added and the MBP cleaved, so that fiber could grow long enough to be insoluble.

It has been suggested that that cross- $\beta$  polymerization of LC domains is not the driving force leading to the formation of liquid-like droplets, but that it may be of biologic utility during the maturation of liquid-like droplets and/or RNA granules (Lin et al., 2015; Molliex et al., 2015). Liquid-like droplets partitioning experiments (Figure 2-17) can be explained by extension of pre-exist polymeric fibers in matured droplets. Indeed, presence of fibrous seeds in the soluble form of MBP:PTB:hnRNPA2 LC may be enough to recruit eGFP:hnRNPA2 LC to liquid-like droplets, and liquid-like droplets partitioning might be

driven by different mechanism by which liquid-liquid biphasic separation is initially established.

During the preparation of hydrogel droplets, I have long observed that the concentrated protein solutions become cloudy prior to hydrogel formation. Direct dilution of a fully denatured His<sub>6</sub>-tagged LC domain of FUS from 8 M urea yielded similar cloudy solution, which was examined by light microscopy and was revealed to be composed of liquid-like droplets. I thereby exploited the liquid-like droplets formed by His<sub>6</sub>-tagged LC domain of FUS to study the mechanism of phase separation. Since the FUS LC protein in this experiment was directly diluted from a fully denatured stock, there should be no presence of pre-existing polymeric seeds, and the formation of liquid-like droplets cannot be extension of pre-existing polymers. Wild type or mutated variant of FUS LC domain (1-214) with one, two or three tyrosine residues mutated to serine (Figure 1-11A, mutants 1A – 3I) were subcloned into bacteria expression vector containing an N-terminal His<sub>6</sub>-tag. Recombinant His<sub>6</sub>-tagged FUS LC variants were overexpressed in BL21 cells, purified on Ni-NTA column in presence of 6 M guanidine HCl, and concentrated to 4 mM (75 mg/ml) stock solutions in 8 M urea. Concentrated stock solution of wild type or mutational variant of FUS LC were then diluted to 1.5, 2, 2.5 or 3 mg/ml in native buffer (25 mM HEPES 7.4, 100 mM KCl, 2% glycerol), and 60 µl of each sample was incubated in BSA coated 96-well plates at room temperature for 30 minutes before imaging (Figure 3-1A). Wild type FUS LC domain formed liquid-like droplets at concentrations as low as 1.5 mg/ml. Among mutated variants of FUS LC domain containing one tyrosine mutated to serine, 1A mutant formed liquid-like droplets at concentrations of 2.0 mg/ml or higher, while 1C mutant, which bound

hydrogel more weakly than 1A mutant, only formed liquid-like droplets at 2.5 mg/ml or higher. Mutated variants of FUS LC domain containing two tyrosine mutated to serine required higher concentration to form liquid-like droplets: the 2C mutant, which bound to hydrogel as well as wild type FUS LC, formed liquid-like droplets at 2.5 mg/ml; the 2F mutant, which exhibited hydrogel binding half that of the native protein, could not form liquid-like droplets until protein concentration was raised to 3 mg/ml; finally, no liquid-like droplets formation was detected even at the highest concentration tested (3 mg/ml) with the 2A or 2E mutants, which bind hydrogels only very weakly. None of the mutational variants with three tyrosine residues mutated to serine formed liquid-like droplets at any concentration tested.

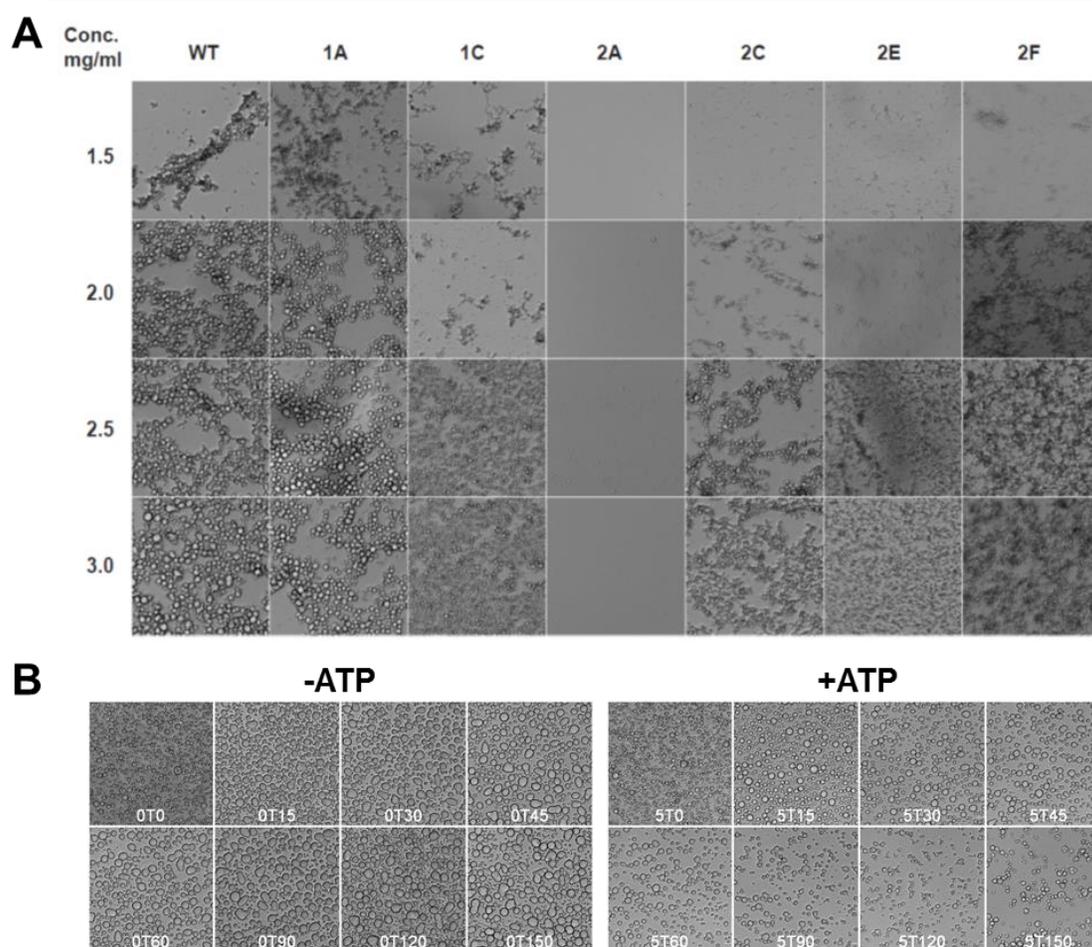
In summary, the abilities of liquid-like droplets formation of mutated variants of FUS LC were affected by two factors. Among mutated variants containing the same number of tyrosine-to-serine mutations, strong correlative relationship was observed between droplets formation and hydrogel binding. In addition, variants that bind hydrogel similarly but contain more tyrosine-to-serine mutations required higher concentration to form liquid-like droplets. I offer that the correlation between liquid-like droplets formation and hydrogel binding gave evidence that initial liquid-droplet formation required formation of cross- $\beta$  polymers, and the solubility of the protein may also affect this process.

LC domains are a cellular sink for post-translational modification, including phosphorylation, acetylation, methylation, glycosylation and PARPylation (Choudhary *et al.*, 2009; Lee, 2012; Zhang *et al.*, 2013). Knowing that phosphorylation of serine residues in the FUS LC domain regulates its polymer formation (Kato *et al.*, 2012), I reasoned that

phosphorylation may regulate liquid-like droplets formed by FUS LC domain. Supporting this idea, DNA-PK, the protein kinase that phosphorylates the LC domain of FUS, co-localized to FUS containing granules at DNA damage site generated by laser micro-irradiation (Altmeyer et al., 2015; Mastrocola et al., 2013; Uematsu et al., 2007). To study the role of phosphorylation in liquid-like droplet regulation, I reconstituted the phosphorylation of the FUS LC domain by DNA-PK with liquid-like droplets formed in vitro. His<sub>6</sub>-tagged FUS LC (wild type) was mixed with DNA-PK with or without ATP, in presence of salmon sperm DNA as activator, incubated in 96-well plate, and imaged every 15 minutes for the first hour, then every 30 minutes until 150 minutes (Figure 3-1B). In absence of ATP, liquid-like droplets were observed immediately after I mixed FUS LC with reaction buffer. The droplets merged with each other, and became large droplets. In presence of ATP, the size of liquid-like droplets peaked after 15 minutes of incubation, and shrank as a function of time. Having observed parallel effects of DNA-PK on hydrogel binding (Han et al., 2012) and the formation of liquid-like droplets, I propose that both phenomena are driving by cross- $\beta$  polymerization.

In summary, I made liquid-like droplets with PTB:hnRNPA2 LC and synthetic RNA substrate, tested liquid-like droplets partitioning by incubating the droplets with mutation variants of hnRNPA2 LC bearing single tyrosine or phenylalanine residues mutated to serine, and showed correlative relationship between hydrogel retention and liquid-like droplets partitioning (Figure 2-17). I studied the structural properties of these liquid-like droplet formed by PTB:hnRNPA2 LC by NAI footprinting and observed the same NAI footprint as obtained from hydrogels composed of mCherry:hnRNPA2 LC polymers (Figure 2-18). In

order to rule out the possibility that my results are an experimental artifact generated by basal level of polymeric seeds present in MBP:PTB:hnRNPA2 LC solution prior to TEV cleavage, I studied liquid-like droplets formed by dilution of fully denatured His<sub>6</sub>-tagged FUS LC. By diluting concentrated stocks of mutational variants of FUS LC out of 8 M urea, I observed correlation between hydrogel retention and the tendencies of forming liquid-like droplets. Moreover, the droplets formed by His<sub>6</sub>-tagged FUS LC can be regulated in an ATP dependent manner by DNA-PK, the same enzyme that regulates cross- $\beta$  polymerization of mCherry:FUS LC. Although these observations do not rule out the involvement of other chemical or physical forces in the formation of liquid-like droplets, I conclude that cross- $\beta$  interactions between LC domains are an important component of the forces facilitating both partitioning of LC domain containing protein into liquid-like droplets, and initial phase separation of LC sequences into liquid-like droplets.



**Figure 3-1. Liquid-like Droplets Formed by His<sub>6</sub>-tagged FUS LC**

- (A) Mutated variants of FUS LC were concentrated in 8 M urea, diluted to the indicated concentration with native buffer (25 mM HEPES pH 7.4, 100 mM KCl, 2% glycerol), and incubated for 30 minutes in 96-well plates coated with BSA. The liquid-like droplets at bottom of each well were imaged with phase contrast microscopy with use of a 20x lens.
- (B) Concentrated His<sub>6</sub>-tagged FUS LC wild type was diluted to DNA-PK reaction buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.2 mM EGTA, 0.1 mM EDTA, 1 mM DTT) with or without ATP, and subsequently mixed with DNA-PK and salmon sperm DNA. 60  $\mu$ l of each reaction mixtures were incubated in 96-well, and liquid-like droplets were imaged at the bottom of the wells.

## CONCLUSIONS

Eukaryotic cells contain a variety of subcellular puncta, which, unlike mitochondria, lysosomes, chloroplasts and peroxisomes, are not membrane bound. These non-membrane bound subcellular puncta include various RNA granules, such as stress granules, P-granules, processing bodies and neuronal granules. These various puncta may serve to increase the local concentration of RNA regulatory proteins and their RNA substrates, to sequester or specifically process certain RNAs, to facilitate RNA transportation or asymmetrically enrich certain regulatory factors to assist in determination of the germ lineage in fly or *C. elegans* embryos (Anderson and Kedersha, 2009). Similar non-membrane bound protein-RNA aggregates were described in the nuclei of cells, including nucleoli, nuclear speckles and para-speckles, PML bodies, Cajal bodies and transcription factories (Mao *et al.*, 2011). Fluorescence microscopy studies of RNA granules or nucleoli have shown liquid-like behavior of these puncta, including fission and fusion of granules and short recovery times of FRAP experiments, and led to the idea that the granule components exist in a liquid-like phase separated from bulk cytoplasm or nucleoplasm (Brangwynne *et al.*, 2009; Buchan and Parker, 2009; Marko, 2012).

Studies that may be pertinent to the biochemical forces leading to the organization of these cellular structures have begun to appear over the past several years. A potentially common conceptualization may tie two orthogonal approaches together. Rosen and colleagues have provided evidence that multivalent, polymeric structures form when proteins

containing repeated SRC homology 3 (SH3) domains are mixed with proteins containing repeated proline-rich motifs (PRMs). Upon heterotypic polymerization into dendritic assemblies, these proteins undergo phase separation into spherical, liquid-like droplets (Li *et al.*, 2012). The Rosen group later expanded their studies to various signaling system that exploit multivalent binding, including adhesion receptor Nephlin and its cytoplasmic partners, Nck and N-WASP (Banjade and Rosen, 2014; Banjade *et al.*, 2015), and T cell receptor signaling (Su *et al.*, 2016), and proposed that protein phase separation can create a distinct biochemical compartment that facilitates efficient signaling.

Parallel to the study of Rosen and colleagues, the McKnight laboratory has been studying the role of LC domains of RNA regulatory proteins containing repetitive sequence of variants of the G/S-Y-G/S motif. It has been shown that the G/S-Y-G/S motif containing LC domains are associated with a variety of RNA granules, and are required for RNA granule recruitment (Han *et al.*, 2012; Kato *et al.*, 2012). In our case, concentrated samples of these proteins have been observed to form hydrogels. Reasonably clear evidence has been gathered to support the conclusion that hydrogel formation equates to polymerization of the LC sequences. Studies of hydrogels have revealed X-ray diffraction patterns consistent with cross- $\beta$  structure, and electron microscopic evaluation of hydrogels has revealed homogeneous polymeric fibrils.

Of significant concern to us has been the question as to whether the polymeric structures being studied in test tube reactions are of biological relevance. Heretofore correlative mutagenesis provided evidence consistent with the idea that polymeric fibers are important for the LC domain function. Initially, FUS mutants were studied for their stress

granule recruitment in transiently transfected cells. Variants of FUS proteins containing tyrosine-to-serine mutations in its LC domain that abolish polymeric fiber formation were unable to be recruited to stress granules. Second, transcriptional activation has been studied with LC domains of the FET family RNA regulatory proteins: FUS, EWS and TAF15. All three of these paralogous proteins have amino terminal LC domains that can be translocated onto DNA binding domains as the causative event leading to human cancer. When fused to the DNA binding domain of GAL4, the LC domains of FET proteins function as transcriptional activation domains (Riggi *et al.*, 2007). Random mutations containing varying numbers of tyrosine residues changed to serine led to decreased activities in hydrogel retention (polymer extension). When transiently transfected to 293T cells, strong correlative relationship was observed between their polymerization capacity and capacity to activate transcription (Kwon *et al.*, 2013). Finally, hydrogels formed by the hnRNPA2 LC domain reversibly binds serine:arginine (SR) repeats that mimics the dynamics properties of nuclear speckles (Kwon *et al.*, 2014). SR-containing proteins in nuclear speckles are regulated by a family of cyclin-like kinase (CLK) enzymes that phosphorylate serine residues within SR domains (Aubol *et al.*, 2013; Colwill *et al.*, 1996; Duncan *et al.*, 1998; Lamond and Spector, 2003; Menegay *et al.*, 2000). Mutation of all serine residues within the SR domain of the SRSF2 protein to glycine resulted in a protein that still bound to hnRNPA2 LC hydrogels, but could not be liberated by the CLK enzymes and ATP. This same mutant, when expressed in living cells, led to the formation of nucleoli-proximal speckles that were resistant to phosphorylation-dependent release via over-expression of the CLK enzyme (Kwon *et al.*, 2014).

Here I have added a direct approach to inquire whether LC domains might function in cells via the same chemistries and structures leading to LC domain polymerization in test tubes. A footprinting method was developed using N-acetylimidazole (NAI). This chemical acetylates amino acid side chains in a manner influenced by protein structure and can be deployed as a reagent useful for both test tube biochemistry and the probing of native protein within freshly isolated nuclei (Figures 2-4, 2-5). I first validate the approach with GST enzyme as test protein, and showed that NAI generates a molecular footprint that negative correlates with solvent accessible surface deduced from crystal structure. I then showed that the method can be used to probe solvent accessible surface of endogenous PARP1 in intact cell nuclei. Following validation, I used this approach to probe NAI accessibility of recombinant polymers composed of hnRNPA2 LC domain. By constructing and studying 25 mutated variants of the LC domain of hnRNPA2, each containing single tyrosine or phenylalanine residues mutated to serine, I found mutants that significantly affect hydrogel binding, which mapped exclusively to the region of the hnRNPA2 LC domain protected from NAI acetylation in the polymeric state (Figure 2-11).

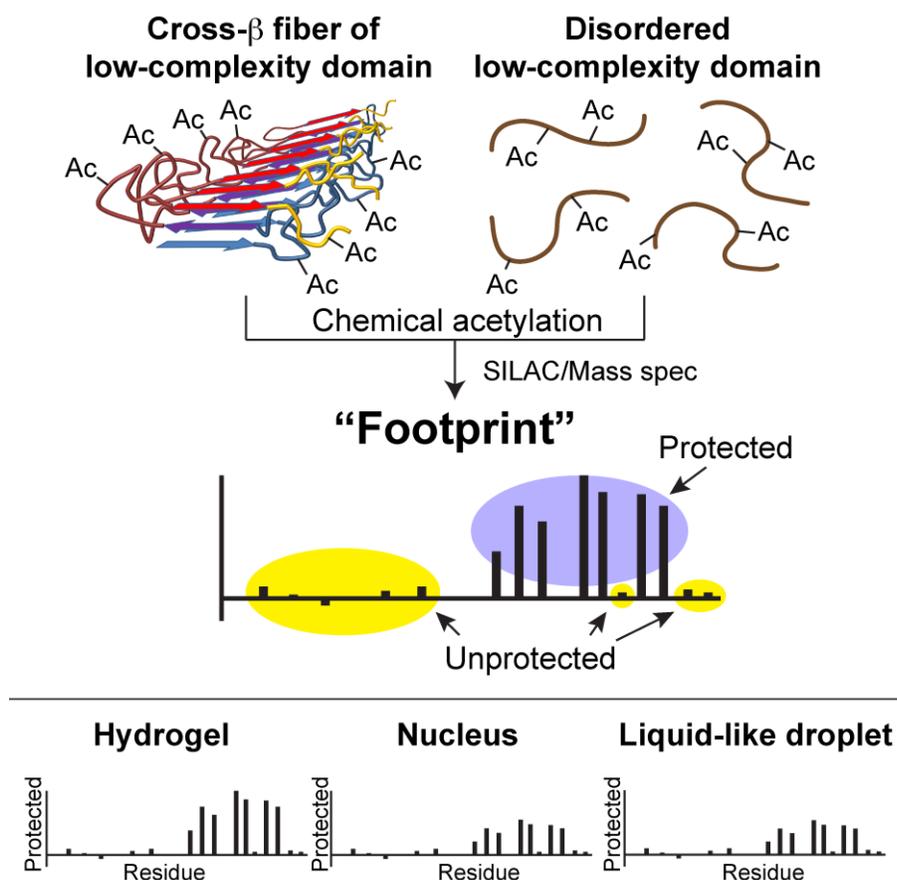
Using a similar methods for determining an NAI footprint for the nuclear form of the PARP enzyme, I probed the structure of native hnRNPA2 present in nuclei freshly prepared from 293T cells. These efforts demonstrated that the footprint of the recombinant amyloid-like fibers is highly related to the footprint observed in nuclei (Figures 2-13, 2-15). Qualitatively, the NAI footprint of 17 out of 18 residues from nuclear hnRNPA2 matched that from polymeric fibers formed by recombinant protein. The single exception to the footprint match was tyrosine residue 324. This residue was protected from NAI acetylation

in recombinant polymers of the LC domain of hnRNPA2, yet was sensitive to modification in the native, nuclear form of the protein. I thereby studied the enzymes associated with mCherry:hnRNPA2 fibers by hydrogel retention, and identified peptidyl-prolyl cis-trans isomerase (PPIA), an abundant protein with a cellular concentration of 10-30  $\mu$ M, that associates strongly with hydrogel droplets. Having noticed that tyrosine residue 324 is closely flanked by proline residues, I hypothesized that the isomeric state of the peptide bonds of these proline residues might affect NAI sensitivity. Indeed, when the LC domain of hnRNPA2 was co-expressed in bacterial cells with the PPIA enzyme, then purified, polymerized and exposed to NAI, the resulting footprint perfectly matched that of endogenous nuclear hnRNPA2. These observations are consistent with the conclusion that the LC domain of at least some proportion of hnRNPA2 in nuclei adopts a similar cross- $\beta$  structure as has been characterized with recombinant polymers.

In considering the virtues and properties of liquid-like droplets as compared with hydrogels, I conclude that cross- $\beta$  polymerization may be at the heart of formation of both hydrogels and liquid-like droplets. I adopted liquid-like droplet system developed by the Parker and Rosen groups (Lin et al., 2015), made droplets with synthetic RNA substrate and a fusion protein of PTB:hnRNPA2 LC, and tested previously mentioned tyrosine or phenylalanine to serine mutants of the LC domain of hnRNPA2 for liquid-like droplets partitioning. A strong correlative relationship was observed between hydrogel retention and liquid-like droplets partitioning (Figure 2-17). Mutants strongly impeded in hydrogel retention partitioned poorly into liquid-like droplets, mutants partially impeded in hydrogel retention were mildly impeded from entering liquid-like droplets, and mutants that bound

hydrogels as well as the wild type LC domain of hnRNPA2 partitioned effectively into liquid-like droplets. I deployed the NAI footprinting technique to probe liquid-like droplets and found the same footprint as was present in both polymers of the recombinant hnRNPA2 LC domain and mammalian cell nuclei (Figure 2-18). These observations favor the idea that the LC domain of hnRNPA2 has similar structures in hydrogels and in liquid-like droplets. To rule out the possibilities that the result of liquid-like droplets partitioning experiment might be artifact generated by short polymeric seeds in the solution of triple fusion protein MBP:PTB:hnRNPA2 LC prior to TEV cleavage, I complemented my studies on hnRNPA2 LC liquid-like droplets with phase separation experiments using fully denatured His<sub>6</sub>-tagged FUS LC as starting material. I purified wild type and mutated variants of His<sub>6</sub>-tagged FUS LC, concentrated them in 8 M urea, and diluted them directly in native buffer to check their tendencies to form liquid-like droplets. I compared mutants with one, two or three tyrosine residues mutated to serine, and observed strong correlative relationship between phase separation into liquid-like droplets and hydrogel retention.

In summary, I developed a chemical footprinting approach with NAI to probe solvent accessible surfaces of proteins in test tubes and cell nuclei, and found that the LC domain of hnRNPA2 LC adopts similar conformation in hydrogels composed of polymeric fibers, in liquid-like droplets, and in mammalian cell nuclei (Figure 3-2). In conclusion, I propose that the involvement of LC domains in the formation of non-membrane bound subcellular organelles, liquid-like droplets and hydrogels all rely on one in the same phenomenon – cross- $\beta$  polymerization.



**Figure 3-2. NAI Footprint Probes Structure of hnRNPA2 LC**

An NAI footprint method was developed to probe the solvent accessible surface of the LC domain of hnRNPA2. Similar footprints were observed from the hydrogels composed of recombinant fibers, the nuclei freshly prepared from mammalian cells, and the liquid-like droplets.

## CHAPTER FOUR

### Methodology

#### MATERIALS

N-acetylimidazole and DMSO (over molecular sieve) were purchased from Sigma-Aldrich (USA). An expression vector (pGEX2T) for recombinant GST was purchased from GE Life Science (USA). Methylamine vanadate (NanoVan) was purchased from Nanoprobe (USA). Centrifugal filters (Amicon Ultra, all sizes) were purchased from Millipore (USA). The parental vector for expression of the triple fusion protein of MBP:PTB:hnRNPA2 LC domain was provided by Dr. Michael Rosen of University of Texas Southwestern Medical School.

#### EXPERIMENTAL PROCEDURE

##### **Expression and Purification of mCherry or eGFP Fusion Proteins.**

His<sub>6</sub>:eGFP or His<sub>6</sub>:mCherry linked to the LC domain of hnRNPA2 (residues 181-341) and tyrosine-to-serine mutants of eGFP:hnRNPA2 LC was over-expressed in *E. coli* strain BL21(DE3) and purified with Ni-NTA resin (Qiagen, USA) as described previously (Kato et al., 2012). Briefly, one liter of bacteria cells expressing each fusion proteins were induced

with 0.4 mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) overnight at 16 °C, harvest, and resuspended in 30 ml of wash buffer (25 mM TRIS pH 8.0, 2 M guanidine hydrochloride, 500 mM NaCl, 15 mM imidazole pH 7.5, 5 mM 2-mercaptoethanol (BME)) with additional 1% Triton X-100 and protease inhibitor cocktail (Sigma-Aldrich, USA). The cells were then sonicated, and centrifuged at 16,000 rpm for one hour. The supernatant containing fusion proteins were loaded onto Ni-NTA tandem columns (2 ml resin x 2), washed with 20 column volume (CV) of wash buffer, and eluted in 15 ml of Ni elution buffer (25 mM TRIS pH 8.0, 300 mM imidazole, 2 M guanidine HCl, 5 mM BME). The elution was then concentrated, mixed with glycerol in 1:1 ration, and stored -20°C.

To purify mCherry:hnRNPA2 LC co-expressed with PPIA, human PPIA full-length cDNA was cloned into multi-cloning site 2 of pACYC-Duet1, co-transformed with mCherry:hnRNPA2 expression vectors into BL21 (DE3) cells, and mCherry:hnRNPA2 LC was purified by Ni-NTA column.

### **Expression and Purification of His<sub>6</sub>-tagged Proteins.**

cDNA sequences of FUS or hnRNPA2 LC domain were inserted into multi-cloning site of pHis-parallel vector and transformed to BL21 (DE3) bacteria cells for protein expression. The cultured bacteria was induced with 0.4 mM IPTG overnight at 16 °C, harvest, resuspended in GuHCl wash buffer (25 mM TRIS pH 8.0, 15 mM imidazole, 6 M guanidine hydrochloride, 5 mM BME), and lysed by sonication. After spun at 16,000 rpm for one hour, the supernatant was loaded onto Ni-NTA tandem columns (2 ml resin x 2),

washed with 20 CV of wash buffer, and eluted in 15 ml of GuHCl elution buffer (25 mM TRIS pH 8.0, 300 mM imidazole, 6 M guanidine HCl, 5 mM BME). The elution was then concentrated and stored at -20 °C.

### **Purification of Glutathione S-Transferase (GST).**

Vacant vector pGEX2T was transformed to BL21 (DE3) bacteria cells, and the bacteria culture (1 L) was induced at 22 °C overnight with 0.4 mM IPTG. The cells were harvested and lysed in 30 ml GST wash buffer (25 mM TRIS pH 7.4, 150 mM NaCl). After spinning at 16,000 rpm for one hour, the supernatant was loaded onto GS4B column (GE Life Science, USA), washed with 30 CV of wash buffer, and eluted in wash buffer with additional 5 mM glutathione (reduced).

### **Hydrogel-binding Assays.**

Hydrogel droplets of mCherry:LC domain of hnRNPA2 were prepared as described previously (Kato et al., 2012). The purified hnRNPA2 LC domain linked to mCherry was concentrated to 20 mg/ml and dialyzed against a gelation buffer containing 20 mM TRIS pH 7.5, 200 mM NaCl, 20 mM BME, 0.5 mM EDTA and 0.1 mM PMSF at room temperature for overnight. The dialyzed protein solution was sonicated briefly, and spun at 20,000x g for 2 minutes. The supernatant was concentrated to 50 mg/ml, and a 0.5 µl droplet of the protein was deposited onto a glass-bottomed microscope dish (MatTek, MA, USA). The dish was

sealed with Parafilm (Pechiney Plastic Packaging) and incubated overnight at room temperature for hydrogel formation.

To test hydrogel retention of mutation variants of eGFP:hnRNPA2 LC, glycerol stocks of the purified eGFP fusion test proteins were diluted to 1  $\mu$ M in 1 mL of gelation buffer. 1 ml of the eGFP test solution was pipetted into the hydrogel dish to soak the hydrogel droplets in the eGFP solution. After overnight incubation at 4°C, a horizontal section of the hydrogel droplet was scanned at both the mCherry and eGFP excitation wavelengths on a confocal microscope. eGFP signals at a boundary area of the hydrogel droplets were scanned by the program ImageJ (Schneider et al., 2012).

### **Heavy Protein Preparation.**

To prepare the stable-isotope labeled (heavy) hnRNPA2 LC, I followed published procedures (Baxa et al., 2007). Briefly, transformed cells were first cultured in 20 mL of LB medium overnight. The pre-culture was then inoculated in 500 mL of M9 minimum media supplemented with all 20 amino acids (no isotope. See Appendix A for medium recipe.). The culture was shaken at 37 °C until OD<sub>600</sub> reached 0.6. The cells were harvested and washed twice with cold PBS, and inoculated into 500 ml of pre-warmed M9 media supplemented with ring-[<sup>13</sup>C<sub>6</sub>]-tyrosine (labeled with <sup>13</sup>C on the six carbons of the phenyl ring, purchased from Cambridge Isotope Laboratories, Inc, USA) and all other 19 amino acids. The culture was shaken at 37 °C for 30 minutes, and protein expression was induced with 1 mM IPTG for 3 hours before harvest. The cells were then lysed in denaturing wash

buffer (25 mM TRIS pH 8.0, 15 mM imidazole, 5 mM BME, 6 M guanidine HCl). The protein was purified on Ni-NTA resin and eluted in denaturing wash buffer supplemented with 300 mM imidazole.

### **Acetylation of Recombinant Proteins.**

N-Acetylimidazole was prepared as 2 M stock solution in DMSO (Sigma-Aldrich, over molecular sieve), protected under argon gas, and stored at -20°C. To acetylate denatured (heavy) protein, the protein buffer was exchanged to denaturing reaction buffer (50 mM HEPES pH 7.4, 5 M guanidine thiocyanate) by centrifugal filtering units. The protein was first heated at 95 °C for 5 minutes for full denaturation. After cooling to room temperature, the heavy protein was diluted to 1 mg/ml, and acetylated with 30 mM NAI at room temperature for 15 minutes. The reaction was quenched with 0.8 M Tris-HCl (pH 8.8) at room temperature for 15 minutes. The native (light) protein (mCherry:hnRNPA2 LC fibers or GST, 1 mg/ml) was acetylated in native reaction buffer (50 mM HEPES pH 7.4, 150 mM NaCl) with 30 mM NAI at room temperature for 15 minutes, and quenched as described above. The native and denatured proteins were mixed at 1:1 ratio, neutralized by HCl (10 M stock solution, using 10% of the volume that the TRIS was added), and the buffer was exchanged to the denaturing reaction buffer. The mixed protein sample were then heated at 95 °C for 5 minutes for full denaturation.

### **Measuring Acetylation of Tyrosine Side Chains.**

Denatured bovine serum albumin (BSA) was acetylated in a variety of buffer conditions, and concentrated to 50  $\mu$ l with GuHCl buffer (50 mM HEPES pH 7.4, 6 M guanidine hydrochloride). A 20  $\mu$ l aliquot of the solution was diluted twice with the same GuHCl buffer, while the other 20  $\mu$ l of the solution was mixed 1:1 with 5 M hydroxylamine, which removes acetyl groups from tyrosine. UV absorbance of both solutions were measured at 278 nm and the difference was used to calculate the average number of tyrosine residues acetylated per molecule with the following equation, as described by Vallee group (Riordan et al., 1965b; Simpson et al., 1963):

$$N = (A_{HA} - A_{Ace}) / A_{HA} * \epsilon / \Delta\epsilon,$$

where  $A_{HA}$  is the UV absorbance of the hydroxylamine treated sample,  $A_{Ace}$  is the UV absorbance of the sample diluted with GuHCl buffer,  $\epsilon$  is the extinction coefficient of the protein ( $\epsilon = 47790$  for BSA), and  $\Delta\epsilon$  is the molar absorptivity difference between tyrosine and its O-acetyl derivative. Here I adopted the  $\Delta\epsilon = 1370$  derived from acetylated subtilisin (Myers and Glazer, 1971).

### **Mass Spectrometry Analysis and Data Processing.**

The buffer of heavy and light protein mixture was exchanged to urea buffer (25 mM HEPES pH 7.4, 8 M urea, 5 mM BME), and diluted four fold with digestion buffer (100 mM TRIS pH 8.0, 10 mM  $\text{CaCl}_2$ ). Sequencing grade chymotrypsin (Roche, USA) was added to a

final protease:protein ratio of 1:50 (w/w), and the digestion reaction was performed overnight at room temperature.

Digested peptides were desalted using SepPak C18 columns (Waters) according to manufacturer's instructions. The SILAC samples were analyzed by LC-MS/MS experiments on an LTQ Velos Pro Orbitrap mass spectrometer (Thermo, San Jose, CA) using the top twenty CID (collision-induced dissociation) method (Olsen et al., 2009). MS/MS spectra were searched against a composite database of the human IPI protein database (Version 3.60) and its reversed complement using the Sequest algorithm.

Search parameters allowed for chymotryptic peptides including zero to three miss-cleavage, with a static modification of 57.02146 Da on cysteine, and a dynamic modification of acetylation (42.01056 Da) on tyrosine, serine, threonine, lysine, arginine, glutamine or asparagine, oxidation (15.99491 Da) on methionine, stable isotope (6.02013 Da) on tyrosine, respectively. Search results were filtered to include <1% matches to the reverse data base by the linear discriminator function using parameters including Xcorr, dCN, missed cleavage, charge state (exclude 1+ peptides), mass accuracy, peptide length and fraction of ions matched to MS/MS spectra (Huttlin et al., 2010).

The abundances of acetylated peptides in the native and denatured samples were estimated by intensities of the monoisotopic ions from light and heavy labeled peptides, respectively, calculated with intensities averaged in the time window corresponding to the chromatographic peak extracted by heavy m/z. To calculate the NAI footprint, the ratio of the heavy and light peptides was used, and normalized by the heavy-to-light ratio of non-acetylated parental peptide.

### **Calculating Solvent Accessible Surface.**

Solvent accessible surface of protein side chains was calculated using the program “areaimol” of CCP4 (Collaborative Computational Project No. 4) (Winn et al., 2011). The solvent accessible surfaces of atoms forming covalent bonds with the acetyl groups (side chain nitrogen atoms on lysine, arginine or asparagine, and side chain oxygen atoms of tyrosine, serine or threonine) were used for correlative analysis.

### **293T Cell Culture and Nuclei Preparation.**

The tyrosine-negative DMEM high glucose was prepared according to Dulbecco and Freeman’s recipe, except that tyrosine was excluded (Dulbecco and Freeman, 1959). To make light and heavy medium, 100 mg/L of light tyrosine or [<sup>13</sup>C<sub>6</sub>]-tyrosine was added to the tyrosine negative medium, respectively (See Appendix B for DMEM recipe). These media were further supplemented with 10% dialyzed FBS (Sigma-Aldrich, USA), 2 mM L-glutamine and Penicillin/Streptomycin solution. For heavy tyrosine labeling, 293T cells were cultured in the heavy medium for eight generations with 1:5 splitting for each generation, and tested for heavy amino acids incorporation.

Intact nuclei were purified from 10<sup>8</sup> heavy or light cells, respectively. The cells were harvested by trypsin digestion, and washed twice with ice-cold hypotonic buffer (10 mM HEPES pH 7.4, 10 mM KCl, 2 mM MgCl<sub>2</sub>, 5 mM BME with protease inhibitors). The cells were then incubated on ice for 30 minutes, then grinded on ice with 25 strokes using a tissue

homogenizer. Nuclei were collected by spinning at 200x g at 4 °C for 10 minutes, and washed with the hypotonic buffer without BME.

### **Acetylation of 293T Nuclei.**

To prepare the native nuclei sample, intact nuclei from light cells were resuspended in 1 ml nuclei buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 2 mM MgCl<sub>2</sub> and protease inhibitor). For the denatured nuclei sample, the heavy nuclei were resuspended with 1 ml denaturing buffer (50 mM HEPES pH 7.4, 5 M guanidine thiocyanate and protease inhibitor), sonicated and heated at 95 °C for 5 minutes. Both heavy and light samples were acetylated with 30 mM NAI at room temperature for 15 minutes, quenched by 0.8 M TRIS pH 8.8, and mixed together. The mixture was then neutralized with HCl, sonicated, concentrated in denaturing buffer, and heated at 95 °C for 5 minutes.

To prepare samples for NAI footprinting of PARP1, the mixed heavy and light samples were concentrated in buffer with 8 M urea and diluted with nuclease buffer (20 mM HEPES 7.4, 2 mM MgCl<sub>2</sub>, 1% SDS). 2 µl of benzonase (Santa Cruz, USA) was added to the sample, which was concentrated 10-fold and diluted to the original volume with nuclease buffer. This procedure was repeated three times to remove DNA (Zhang et al., 2013), and the resulting protein solution was digested by chymotrypsin as described above.

To prepare samples for analyzing nuclear hnRNPA2, the sample heated in guanidine thiocyanate was filtered through 100 kDa cutoff centrifugal filtration units (Amicon Ultra),

then the flow-through was concentrated by 30 kDa Amicon Ultra in urea buffer (25 mM HEPES pH 7.4, 8 M urea, 5 mM BME) and digested by chymotrypsin as described above.

### **2D-HPLC Mass Spectrometry.**

Chymotryptic peptides from the nuclear SILAC sample were fractionated using an off-line two dimensional RP-RP-HPLC (basic reverse phase coupled with acidic reverse phase) protocol (Gilar et al., 2005). Briefly, lyophilized chymotryptic peptides were resuspended in 500  $\mu$ l buffer A (10 mM ammonium formate, pH 10) and injected onto an Agilent 300 Extended C18 column (2.1 mm  $\times$  100 mm, 3.5  $\mu$ m particle size). Gradient was developed over 60 minutes ranging from 5% to 35% buffer B (10 mM ammonium formate, pH 10, 90% acetonitrile) at a flow rate of 0.2 ml/min. Seventeen fractions were collected and lyophilized. Peptides were then desalted using SepPak C18 columns and lyophilized. In the second dimension, peptides were separated on a 75  $\mu$ m  $\times$  15 cm in-house packed RP column (Maccel 200-3-C18AQ, 3  $\mu$ m, 200  $\text{Å}$ ) using a gradient developed over 90 minutes ranging 0% to 37% buffer B (97% acetonitrile, 0.1% formic acid) at a flow rate of 300 nl/minute. Peptides were directly introduced into the mass spectrometer via a hand-pulled emitter.

## Identification of Hydrogel Binding Proteins from Heavy Cell Lysates

Hydrogels for proteomic experiments were prepared with mCherry:hnRNPA2 LC containing His<sub>6</sub>-tag at both end of the fusion protein. BL21 bacteria cell pellet expressing the fusion protein were resuspended in wash buffer (25 mM TRIS pH 8.0, 500 mM NaCl, 15 mM imidazole, 5 mM BME) with additional 1% Triton X-100 and protease inhibitor cocktail, sonicated and centrifuged at 16,000 rpm for one hour. The supernatant was then loaded onto Ni-NTA column, washed with 50 CV of wash buffer, and eluted with buffer containing 25 mM TRIS pH 8.0, 300 mM imidazole and 5 mM BME. The protein solution eluted was then diluted five folds with Buffer A (25 mM TRIS 8.0, 5 mM BME), loaded on to 5 ml x 2 HiTrap Q Sepharose HP column (GE Life Sciences), and purified over 200 ml gradient elution ranging from 0% to 50% buffer B (25 mM TRIS 8.0, 1 M NaCl, 5 mM BME). The purified fusion protein was supplemented with 2 M urea, concentrated to 20 mg/ml, and dialysed overnight at room temperature against gelation buffer (20 mM TRIS 7.5, 200 mM NaCl, 20 mM BME, 0.1 mM PMSF, 0.5 mM EDTA). To make hydrogel droplets, the dialyzed protein solution was sonicated briefly and spun at 20,000x g for two minutes. The supernatant was concentrated to 50 mg/ml, and plated as an array of 20 x 10 µl droplets on a 6-cm tissue culture plate. The plate was sealed and incubated for two days at room temperature before applying cell lysate.

To culture heavy amino acids labeled cells, heavy DMEM high glucose was prepared using SILAC media (Invitrogen, USA), supplemented with 100 mg/L of [<sup>13</sup>C<sub>6</sub>]-L-lysine, 100/L [<sup>13</sup>C<sub>6</sub>, <sup>15</sup>N<sub>4</sub>]-L-arginine (Cambridge Isotope Laboratories, Inc, USA), 100 mg/L light

L-proline, 10% dialyzed FBS (Sigma-Aldrich, USA), 2 mM L-glutamine and Penicillin/Streptomycin solution. To label cells with heavy amino acids, U2OS cells were cultured in the heavy medium for eight generations with 1:5 splitting for each generation, and tested for heavy amino acids incorporation. Heavy amino acid labeled cells were harvest by trypsin digestion, washed twice with cold PBS, and lysed is in cell lysis buffer (10 mM HEPES pH 7.4, 100 mM KCl, 1% Triton X-100, 2 mM MgCl<sub>2</sub>, 5 mM BME, with benzonase and protease inhibitor cocktail) by sonication, and diluted to 1 mg/ml.

A total of 200 µl mCherry:hnRNPA2 LC hydrogel was incubated with 5 ml of the heavy labeled cell lysate at 4 °C with shaking, washed three times with cell lysis buffer, and dissolved in GuHCl buffer (25 mM TRIS pH 8.0, 6 M guanidine hydrochloride, 10 mM imidazole, 5 mM BME), and loaded onto Ni-NTA columns to remove mCherry:hnRNPA2 LC. The resulting protein solution was concentrated, separated on SDS-PAGE gel, and subjected to mass spectrometry analysis.

#### **Purification of MBP:PTB:hnRNP LC Fusion Protein.**

BL21 (DE3) cells for co-expression of PPIA and His<sub>6</sub>-tagged MBP:PTB:hnRNPA2 LC were cultured in LB media to OD<sub>600</sub> of 0.8 at 37 °C and induced by 0.4 mM IPTG at 16°C overnight. The cells were harvested, resuspended and sonicated in lysis buffer (25 mM TRIS pH 8.0, 1.5 M NaCl, 10 mM imidazole, 10% glycerol, 5 mM BME). The bacteria lysate was cleared by centrifuging at 18,000 rpm for one hour, and loaded on a Ni-NTA column. Then the resin was washed with lysis buffer, and His<sub>6</sub>-tagged MBP:PTB:hnRNPA2 LC was eluted

with buffer containing 25 mM TRIS pH 8.0, 300 mM imidazole, 10 % glycerol and 5 mM BME. 5 nmoles of SNAP-Surface 549 (NEB) was added to the elution, stirred at 4°C overnight, and followed by purification with amylose resin (NEB, USA). The fusion protein was eluted in droplet buffer (25 mM HEPES pH 7.4, 100 mM NaCl, 10% glycerol) supplemented with 10 mM maltose, and then centrifuged at 100,000 x g for one hour to remove potential aggregates.

### **Liquid-like Droplet Formation by MBP:PTB:hnRNPA2 LC and NAI Footprinting.**

Liquid-like droplets were formed by mixing 5  $\mu$ M of MBP:PTB:hnRNPA2 LC with 1  $\mu$ M of RNA (5 copies of UCUCUAAAAA), followed by cleavage of MBP by TEV protease at room temperature. 30 mM NAI was added to the droplet solution at the indicated time points and quenched. Denatured His<sub>6</sub>-tagged hnRNPA2 LC (heavy) proteins were acetylated in the modified denaturing reaction buffer (50 mM HEPES 7.4, 5 M guanidine thiocyanate, 10% glycerol) with 30 mM NAI, quenched 0.8 M TRIS pH 8.8, and mixed with the native sample. Chymotrypsin digestion and mass spectrometry analysis were performed as described above.

### **eGFP:hnRNP A2 Recruitment by Liquid-like Droplet.**

Liquid-like droplets formed from MBP:PTB:hnRNPA2 LC and the synthetic RNA substrate were incubated at room temperature for one hour. 0.1  $\mu$ M of eGFP linked wild-

type or mutant hnRNPA2 LC proteins were added to the liquid-like droplet solution, and incubated on cover slides for 5 minutes at room temperature. The droplets were imaged by DeltaVision Elite fluorescent microscope using a 100x oil lens and eGFP signals inside and outside of liquid-like droplets were measured by the program Image J. The partition ratio of eGFP:hnRNPA2 protein was calculated by dividing the signal inside the droplet by the signal outside. This ratio was then normalized with the signal intensity of Alexa 549 that had been SNAP tagged onto the PTB:hnRNPA2 LC protein in the droplet.

### **Localization of PPIA to Stress Granules.**

U2OS cells were cultured on chamber slides overnight after seeding. Stress granules were induced by heat shock at 44 °C for 45 minutes. Cells were fixed with 4% paraformaldehyde for 10 minutes, permeabilized in PBST (PBS plus 0.2% Triton) for 20 minutes. After blocking with PBTA (3% BSA, 0.2% Triton in PBS) for 30 minutes, cells were incubated with 1:300 goat anti-TIA1 (Santa Cruz, USA) and 1:300 rabbit anti-PPIA (Abcam, USA) in PBTA overnight at 4°C. Primary antibodies were developed with 1:300 AlexaFluor 568 conjugated donkey anti-goat (Invitrogen, USA) and 1:500 AlexaFluor 488 conjugated donkey anti-rabbit (Invitrogen, USA) for 1 hour at room temperature. After washing with PBST, the cells were mounted with VECTASHIELD mounting medium with DAPI (Vector laboratories, USA), and imaged using a DeltaVision Elite microscope.

**Liquid-Like Droplets Formation by His<sub>6</sub>-tagged FUS LC.**

To prepare 96-well plate for liquid-like droplets experiments, 60  $\mu$ l of 5% BSA was incubated in each well of the plate (non-tissue culture treated) at room temperature for 10 minutes. The plate was then vacated and air-dried. Mutation variants of His<sub>6</sub>-tagged FUS LC proteins were diluted in urea buffer (25 mM TRIS 8.0, 8 M urea, 5 mM BME) and concentrated to 75 mg/ml. 1.5, 2, 2.5, or 3  $\mu$ l of the concentrated protein stocks were diluted to 60  $\mu$ l with droplet buffer (10 mM HEPES pH 7.4, 100 mM KCl, 2% glycerol), and incubated at room temperature for 30 minutes on the BSA coated plate. Droplet images were taken by ZOE Fluorescent Cell Imager (Bio-Rad, USA) with bright field from the bottom of each well.

Phosphorylation of His<sub>6</sub>-tagged FUS LC was performed with DNA-PK (Promega). His<sub>6</sub>-tagged FUS LC (wild type) was diluted with DNA-PK reaction buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl<sub>2</sub>, 0.2 mM EGTA, 0.1 mM EDTA, 1 mM DTT), with or without ATP, to 2.5 mg/ml, and subsequently mixed with 5 units of DNA-PK and 10  $\mu$ g/ml salmon sperm DNA. The mixtures were incubated in 96-well plate, and imaged every 15 minutes for the first hour, and then every 30 minutes until 150 minutes.

**Transmission Electron Microscopy of mCherry:hnRNPA2 Fibers.**

The recombinant mCherry:hnRNPA2 hydrogel was resuspended to a final concentration of 0.1 mg/ml by brief sonication in wash buffer (25 mM HEPES pH 7.4, 150

mM NaCl). Freshly glow-discharged (40 mA, 90 seconds) carbon-coated copper grids (carbon film on 400 square mesh copper) (Electron Microscopy Sciences, USA) were floated on the fiber suspension for 2 minutes for protein adsorption. Then the grids were washed in the wash buffer 3 times, and stained with 2% sodium phosphotungstate (pH 7.0) for 1 minute. TEM images were taken on a FEI Tecnai G2 Spirit Biotwin transmission electron microscope with Gatan 2Kx2K multiport readout post column CCD at 67,000 x magnification.

### **Immunoelectron Microscopy of SR Bound FUS Fibers.**

The recombinant eGFP:Flag:SR and mCherry:FUS proteins were mixed in 3:1 ratio, and incubated at room temperature for 15 minutes. Then the mixture, or FUS alone, was spun at 12,000x g for 3 minutes to collect fibers. The pelleted fibers were then resuspended in PBS-Citrate (PBS with 25 mM sodium citrate) to a final concentration of 0.1 mg/ml. Freshly glow-discharged (40 mA, 90 seconds) carbon-coated copper grids (carbon film on 400 square mesh copper, Electron Microscopy Sciences) were float on the fiber samples for two minutes for protein adsorption. The grids were then floated on 5 consecutive drops of wash buffer containing PBS supplemented with 10 mM glycine and 2.5% normal goat serum (Invitrogen, USA), and blocked for 15 minutes at room temperature (PBS supplemented with 10% normal goat serum, 1% BSA and 0.2% gelatin from cold water fish skin). For immunostaining, the grids were floated on primary antibody raised against flag-tag (mouse monoclonal antibody from Sigma Aldrich, 1:20 dilution) in 20  $\mu$ l of blocking

solution for 15 minutes at room temperature. The grids were then washed by floating on 5 consecutive drop of wash buffer, and incubated with 6 nm colloidal gold conjugated goat anti-mouse IgG + IgM (Jackson ImmunoResearch, 1:20 dilution in blocking solution) at room temperature for 15 minutes. Grids were washed five times in wash buffer, five times in water and negatively stained by 2% methylamine vanadate for 30 seconds. All images were taken by JEOL 1200EX transmission electron microscopes with Sis Morada 11 mpixel side mount CCD camera at a 100,000 x magnification.

## APPENDIX A Heavy Medium for Bacterial Cultures

To make 1 liter M9 medium, weighing:

11.3 grams M9 Minimal Salt,

0.0100 grams  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ,

2.0 grams glucose.

And Amino Acids (Light):

100 mg A	100 mg G	100 mg M	100 mg S
100 mg C	200 mg H	100 mg N	100 mg T
100 mg D	100 mg I	200 mg P	100 mg V
100 mg E	100 mg K	100 mg Q	200 mg W
200 mg F	100 mg L	100 mg R	200 mg Y

And adding the following components from stock solution:

Final Concentration	Stock	Volume
2mM $\text{MgSO}_4$	2M	1 ml
0.1 mM $\text{CaCl}_2$	1M	0.1 ml
0.0100 grams Thiamine	10mg/ml	1 ml

For heavy medium, use 100 mg heavy Tyrosine instead.

Filter through 0.22  $\mu$  filter to sterilize. Some precipitant may occur, but it will not affect the labeling.

## APPENDIX B

### Heavy DMEM Medium for Mammalian Cell Cultures

#### 10x AA (mg/1L)

Glycine	300	L-Arginine hydrochloride	840
L-Cystine 2HCl	63.0	L-Glutamine	5840
L-Histidine HCl-H <sub>2</sub> O	420	L-Isoleucine	1050
L-Leucine	1050	L-Lysine hydrochloride	1460
L-Methionine	300	L-Phenylalanine	660
L-Serine	420	L-Threonine	950
L-Tryptophan	160	L-Valine	940
*L-Tyr disodium dihydrate	1040	*Left out in the 10x, and added heavy in 1x	

#### 100x Vitamins (mg/L)

Adjust pH to dissolve Vitamins at 100x

Choline chloride	400	D-Calcium pantothenate	400
Folic Acid	400	Niacinamide	400
Pyridoxine hydrochloride	400	Riboflavin	40
Thiamine hydrochloride	400	i-Inositol	720

#### 10x Salt (g/L)

CaCl <sub>2</sub> (anhyd.)	2	Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	0.001
MgSO <sub>4</sub> (anhyd.)	1	KCl	4
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O	1.25		

#### Others (g/L)

add as powder in the 1x medium

NaHCO <sub>3</sub>	3.7	NaCl	6.4
D-Glucose	4.5	Phenol Red	0.015

After sterilized through 0.22 μ filter,

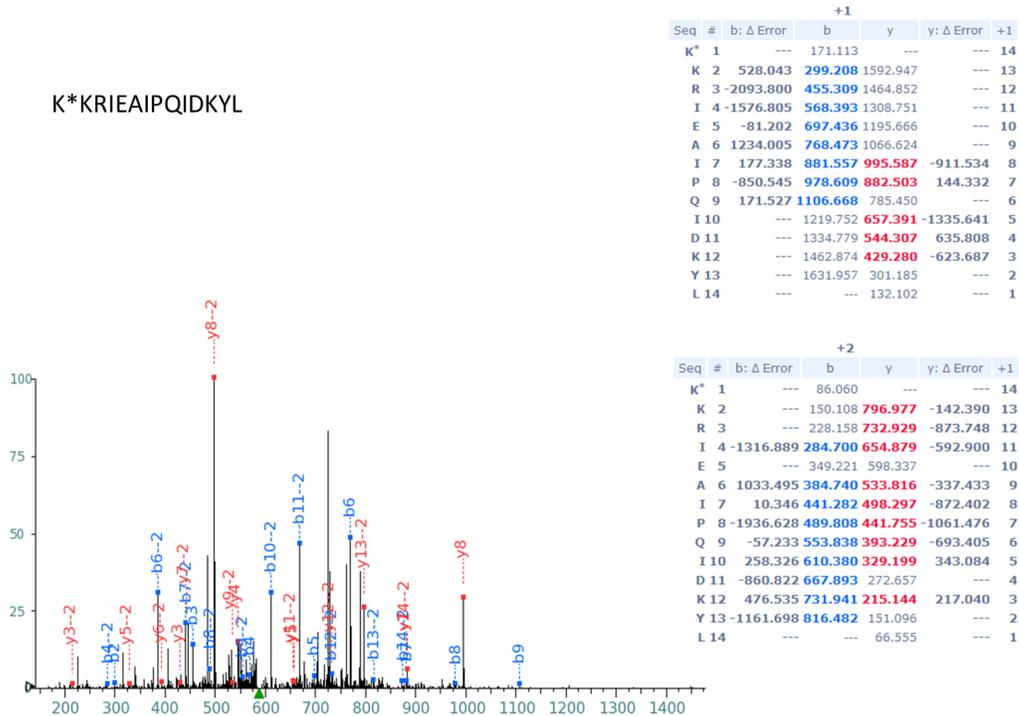
Add 10% dialyzed FBS, 10 ml 100x P/S, and 10 ml 100x L-Glutamine.

## APPENDIX C

### Mass Spectrum of Representative Peptides of GST Enzyme

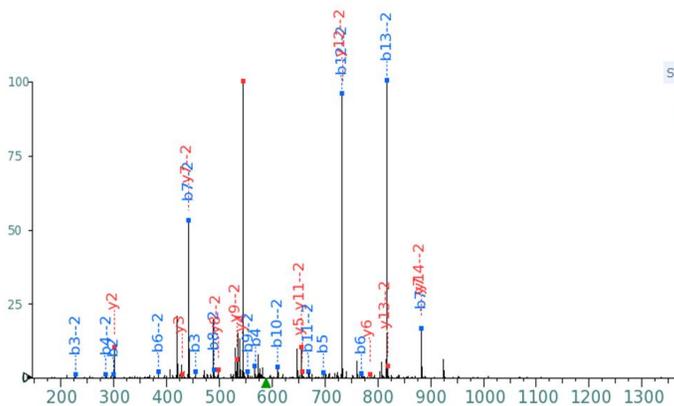
GST enzyme labeled with [ $^{13}\text{C}_6$ ]-Tyrosine were overexpressed in BL21 cells, denatured by 5 M guanidine thiocyanate, acetylated by 30 mM NAI at room temperature for 15 minutes, and digested with chymotrypsin in presence of 2 M urea. The resulting peptides were desalted and subjected to tandem mass spectrometry. Sequence-informative fragmentation ions are marked on the spectrum and summarized in the table, annotated in blue (b-ions) or red (y-ions). Identified peptides are shown with the acetylated residues marked by an asterisk.

(A)



(B)

KK\*RIEAIPQIDKYL

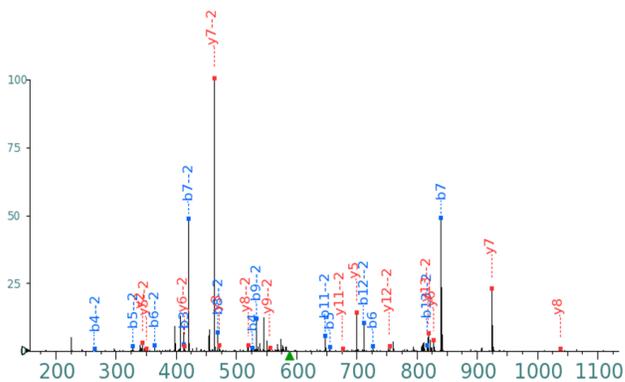


+1					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
K 1	---	129.102	---	---	14
K* 2	2856.225	<b>299.208</b>	1634.957	---	13
R 3	58.490	<b>455.309</b>	1464.852	---	12
I 4	-1570.126	<b>568.393</b>	1308.751	---	11
E 5	-98.883	<b>697.436</b>	1195.666	---	10
A 6	163.517	<b>768.473</b>	1066.624	---	9
I 7	-3.405	<b>881.557</b>	995.587	---	8
P 8	---	978.609	<b>882.503</b>	-1076.467	7
Q 9	---	1106.668	<b>785.450</b>	31.911	6
I 10	---	1219.752	<b>657.391</b>	-895.227	5
D 11	---	1334.779	<b>544.307</b>	1064.774	4
K 12	---	1462.874	<b>429.280</b>	387.453	3
Y 13	---	1631.957	<b>301.185</b>	-173.699	2
L 14	---	---	132.102	---	1

+2					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
K 1	---	65.055	---	---	14
K* 2	---	150.108	<b>817.982</b>	415.848	13
R 3	-4229.783	<b>228.158</b>	<b>732.929</b>	-827.451	12
I 4	-1684.371	<b>284.700</b>	<b>654.879</b>	1070.402	11
E 5	---	349.221	598.337	---	10
A 6	10.860	<b>384.740</b>	<b>533.816</b>	-1783.548	9
I 7	877.906	<b>441.282</b>	<b>498.297</b>	-1026.907	8
P 8	310.876	<b>489.808</b>	<b>441.755</b>	-192.987	7
Q 9	-206.598	<b>553.838</b>	393.229	---	6
I 10	-19.695	<b>610.380</b>	329.199	---	5
D 11	-108.364	<b>667.893</b>	272.657	---	4
K 12	522.769	<b>731.941</b>	215.144	---	3
Y 13	377.193	<b>816.482</b>	151.096	---	2
L 14	---	---	66.555	---	1

(C)

KKRIEAIPQIDKY\*L

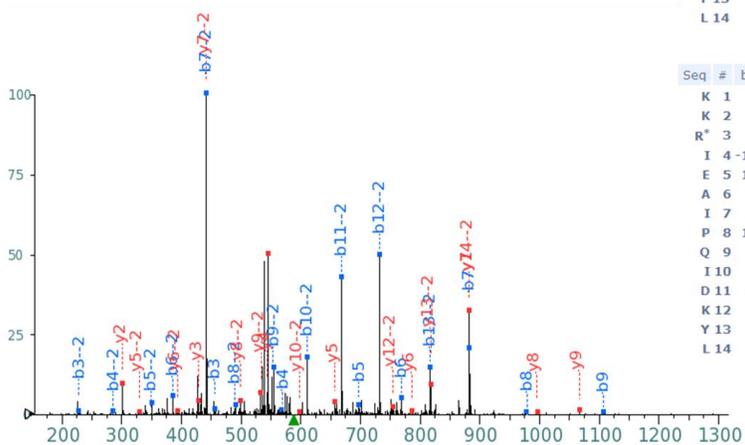


+1					
Seq #	b: Δ Error	b	y	Y: Δ Error	+1
K 1	---	129.102	---	---	14
K 2	---	257.197	1634.957	---	13
R 3	-2008.249	<b>413.298</b>	1506.862	---	12
I 4	-129.404	<b>526.382</b>	1350.761	---	11
E 5	417.323	<b>655.425</b>	1237.677	---	10
A 6	130.564	<b>726.462</b>	1108.634	---	9
I 7	29.873	<b>839.546</b>	<b>1037.597</b>	152.670	8
P 8	---	936.599	<b>924.513</b>	43.906	7
Q 9	---	1064.657	<b>827.460</b>	108.092	6
I 10	---	1177.742	<b>699.402</b>	27.079	5
D 11	---	1292.768	586.318	---	4
K 12	---	1420.863	<b>471.291</b>	460.060	3
Y* 13	---	1631.957	<b>343.196</b>	708.013	2
L 14	---	---	132.102	---	1

+2					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
K 1	---	65.055	---	---	14
K 2	---	129.102	<b>817.982</b>	972.685	13
R 3	---	207.153	<b>753.935</b>	647.579	12
I 4	2192.667	<b>263.695</b>	<b>675.884</b>	1107.768	11
E 5	-466.133	<b>328.216</b>	619.342	---	10
A 6	949.022	<b>363.735</b>	<b>554.821</b>	1185.834	9
I 7	389.541	<b>420.277</b>	<b>519.302</b>	850.367	8
P 8	1136.613	<b>468.803</b>	<b>462.760</b>	73.583	7
Q 9	-11.885	<b>532.832</b>	<b>414.234</b>	134.829	6
I 10	---	589.374	<b>350.205</b>	-1106.002	5
D 11	361.131	<b>646.888</b>	293.663	---	4
K 12	527.276	<b>710.935</b>	236.149	---	3
Y* 13	327.741	<b>816.482</b>	172.102	---	2
L 14	---	---	66.555	---	1

(D)

KKR\*IEAIPQIDKYL

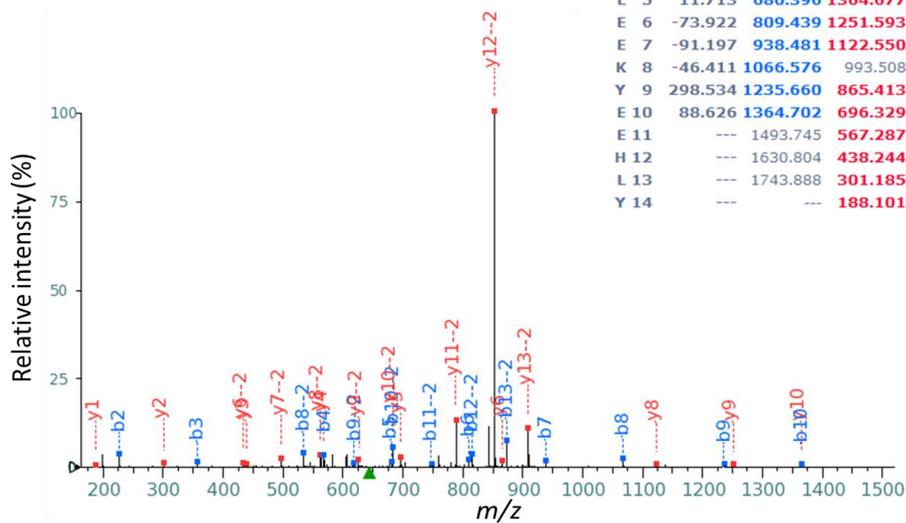


+1					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
K 1	---	129.102	---	---	14
K 2	---	257.197	1634.957	---	13
R* 3	49.241	<b>455.309</b>	1506.862	---	12
I 4	-800.739	<b>568.393</b>	1308.751	---	11
E 5	-130.395	<b>697.436</b>	1195.666	---	10
A 6	-136.777	<b>768.473</b>	<b>1066.624</b>	30.831	9
I 7	37.581	<b>881.557</b>	<b>995.587</b>	983.288	8
P 8	-662.904	<b>978.609</b>	<b>882.503</b>	72.973	7
Q 9	611.187	<b>1106.668</b>	<b>785.450</b>	249.973	6
I 10	---	1219.752	<b>657.391</b>	123.505	5
D 11	---	1334.779	<b>544.307</b>	1236.392	4
K 12	---	1462.874	<b>429.280</b>	20.920	3
Y 13	---	1631.957	<b>301.185</b>	75.585	2
L 14	---	---	132.102	---	1

+2					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
K 1	---	65.055	---	---	14
K 2	---	129.102	<b>817.982</b>	308.254	13
R* 3	154.884	<b>228.158</b>	<b>753.935</b>	680.079	12
I 4	-1575.539	<b>284.700</b>	654.879	---	11
E 5	1935.635	<b>349.221</b>	<b>598.337</b>	-357.372	10
A 6	106.826	<b>384.740</b>	<b>533.816</b>	-1423.605	9
I 7	865.894	<b>441.282</b>	<b>498.297</b>	-41.178	8
P 8	1492.415	<b>489.808</b>	<b>441.755</b>	-205.012	7
Q 9	579.744	<b>553.838</b>	<b>393.229</b>	1057.950	6
I 10	659.239	<b>610.380</b>	<b>329.199</b>	1078.373	5
D 11	446.153	<b>667.893</b>	272.657	---	4
K 12	744.800	<b>731.941</b>	215.144	---	3
Y 13	326.247	<b>816.482</b>	151.096	---	2
L 14	---	---	66.555	---	1

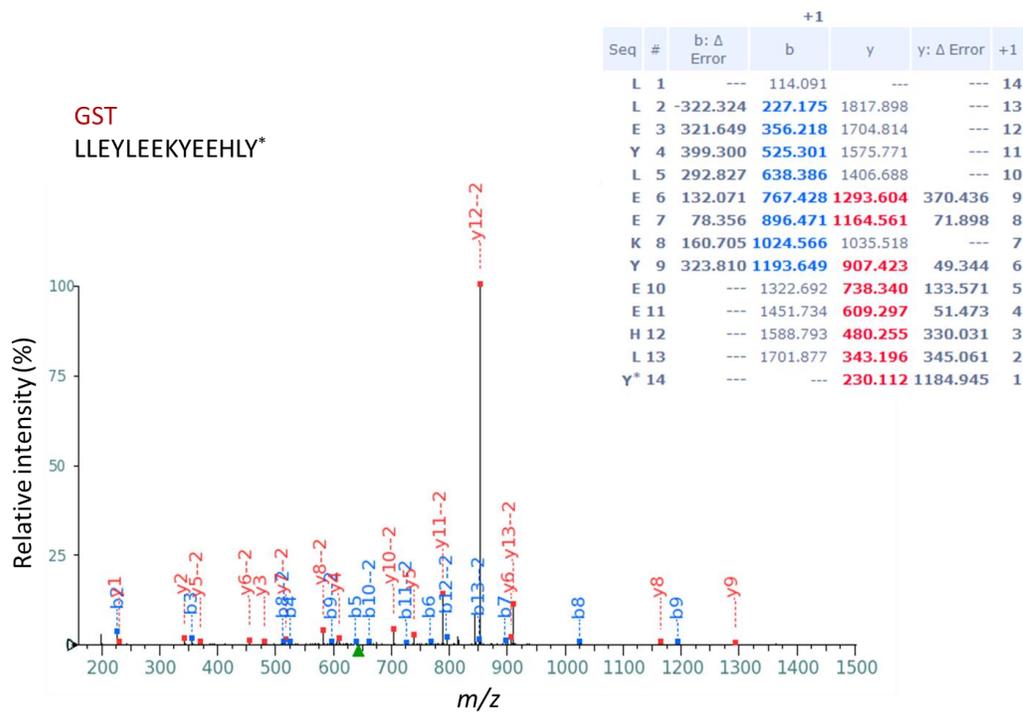
(E)

GST  
LLEY\*LEEKYEEHLY

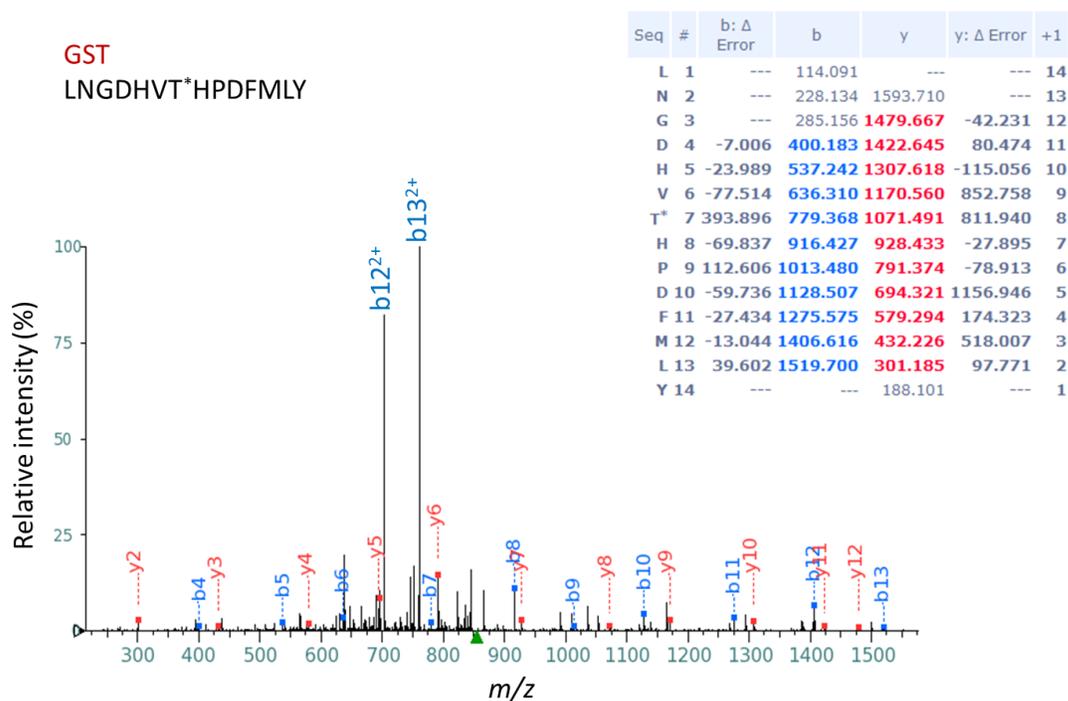


+1					
Seq #	b: Δ Error	b	y	y: Δ Error	+1
L 1	---	114.091	---	---	14
L 2	-585.860	<b>227.175</b>	1817.898	---	13
E 3	-547.337	<b>356.218</b>	1704.814	---	12
Y* 4	-140.756	<b>567.312</b>	1575.771	---	11
L 5	11.713	<b>680.396</b>	<b>1364.677</b>	107.054	10
E 6	-73.922	<b>809.439</b>	<b>1251.593</b>	236.809	9
E 7	-91.197	<b>938.481</b>	<b>1122.550</b>	-38.332	8
K 8	-46.411	<b>1066.576</b>	993.508	---	7
Y 9	298.534	<b>1235.660</b>	<b>865.413</b>	-83.459	6
E 10	88.626	<b>1364.702</b>	<b>696.329</b>	-70.030	5
E 11	---	1493.745	<b>567.287</b>	-96.416	4
H 12	---	1630.804	<b>438.244</b>	225.820	3
L 13	---	1743.888	<b>301.185</b>	-233.910	2
Y 14	---	---	<b>188.101</b>	-259.224	1

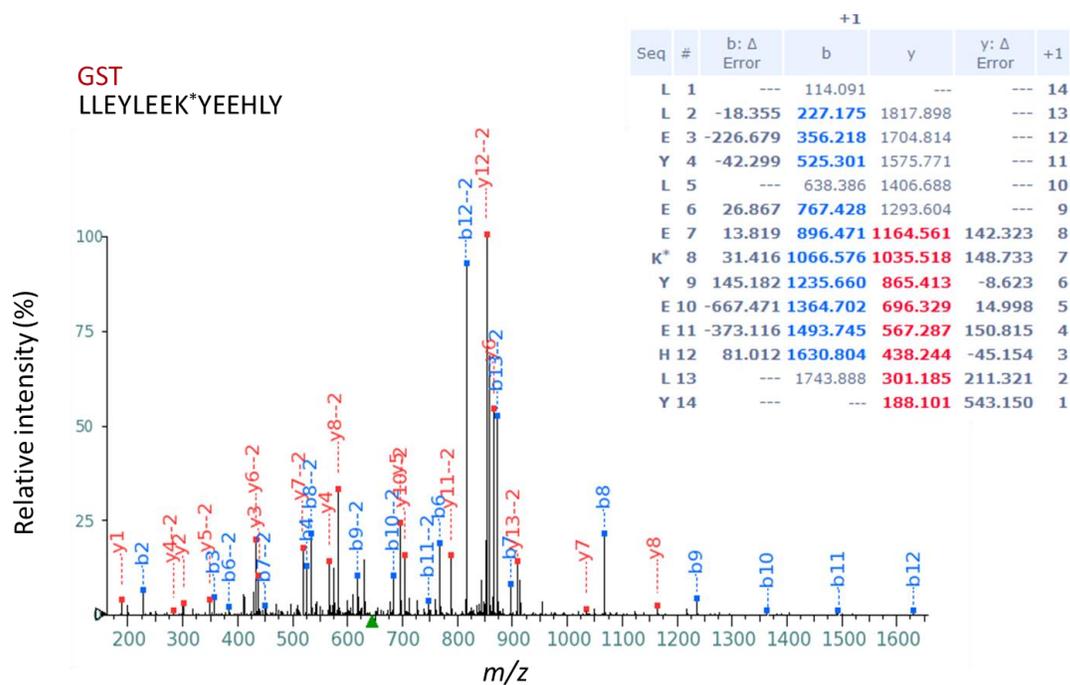
(F)



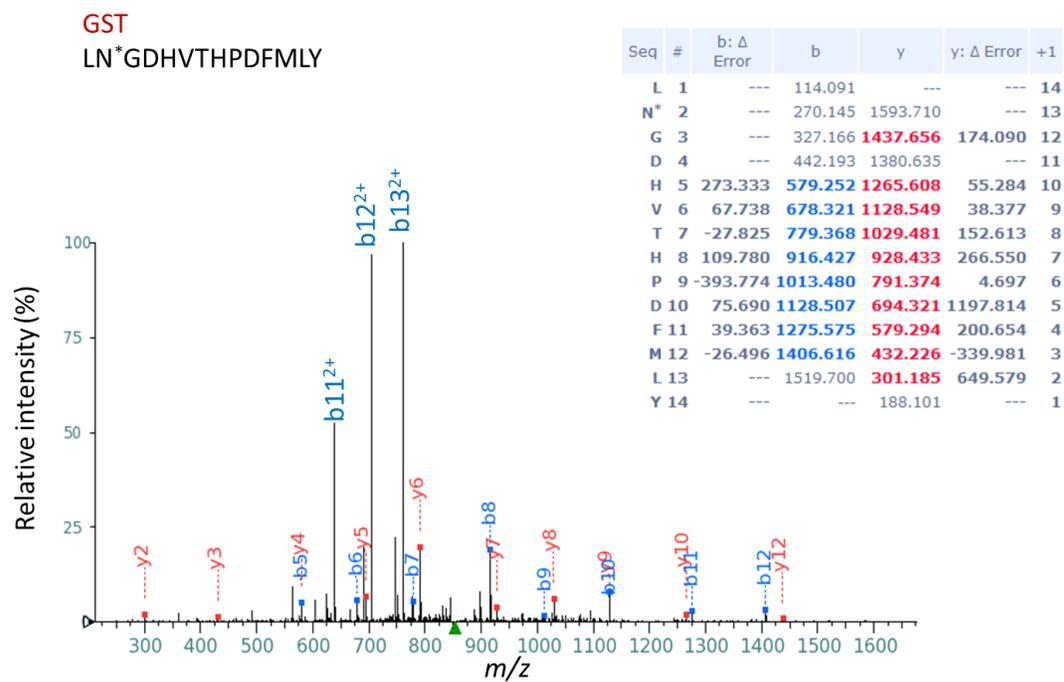
(G)



(H)



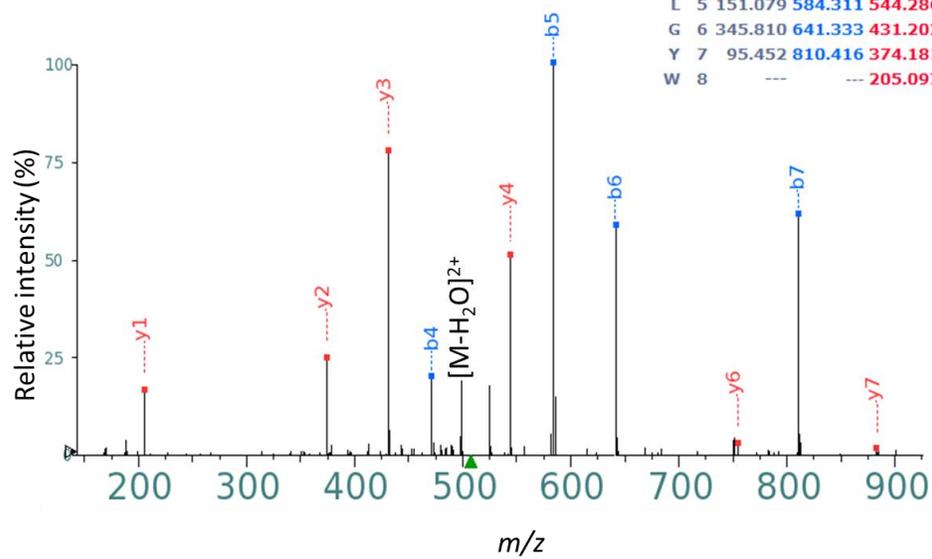
(I)



(J)

GST  
MS\*PILGYW

Seq	#	b: $\Delta$ Error	b	y	y: $\Delta$ Error	+1
M	1	---	132.048	---	---	8
S*	2	---	261.090	<b>883.466</b>	-755.664	7
P	3	---	358.143	<b>754.423</b>	212.010	6
I	4	184.519	<b>471.227</b>	657.370	---	5
L	5	151.079	<b>584.311</b>	<b>544.286</b>	231.162	4
G	6	345.810	<b>641.333</b>	<b>431.202</b>	179.334	3
Y	7	95.452	<b>810.416</b>	<b>374.181</b>	130.289	2
W	8	---	---	<b>205.097</b>	615.552	1

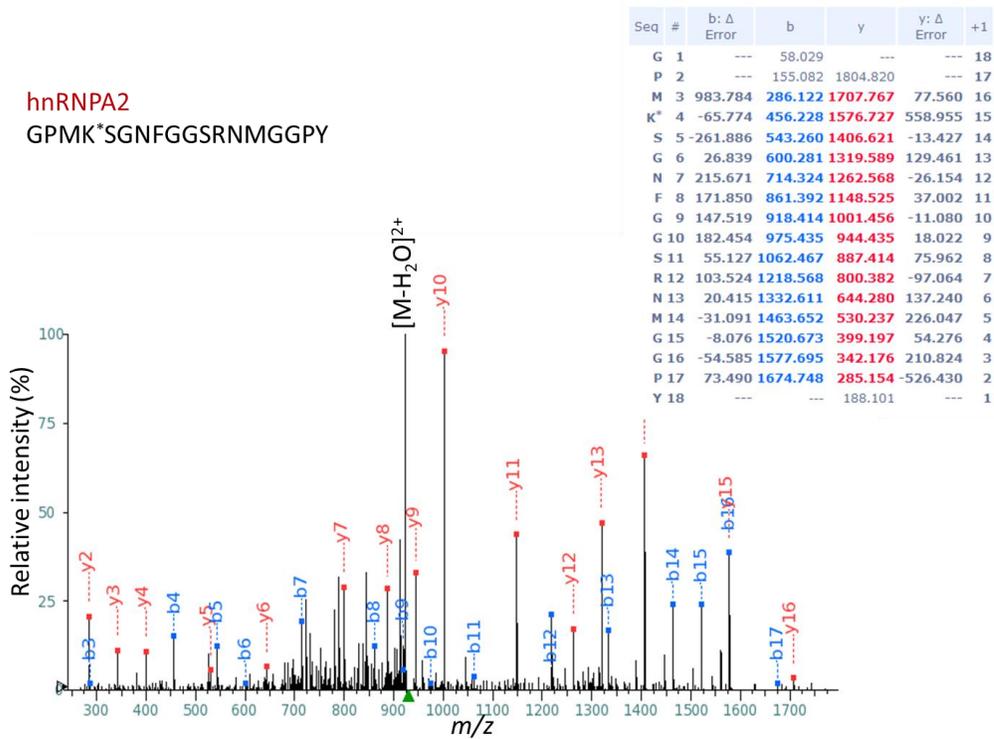


## APPENDIX D

### Mass Spectrum of Representative Peptides of hnRNPA2

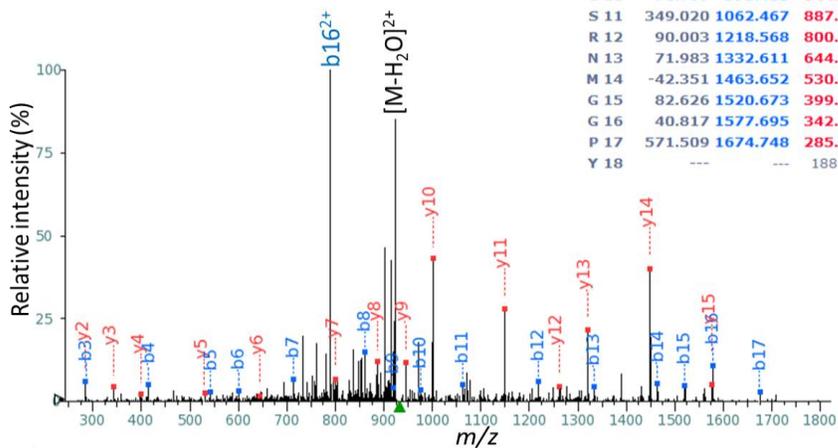
MS2 spectra of the representative hnRNPA2 peptides. Also shown is a fragment ion table (only +1 ions) showing the theoretical  $m/z$  of the b- and y-ions, and their corresponding mass error (in ppm). Assigned b- and y-ions are highlighted in blue and red, respectively. Identified peptides are shown with the acetylated residues marked by an asterisk.

(A)



(B)

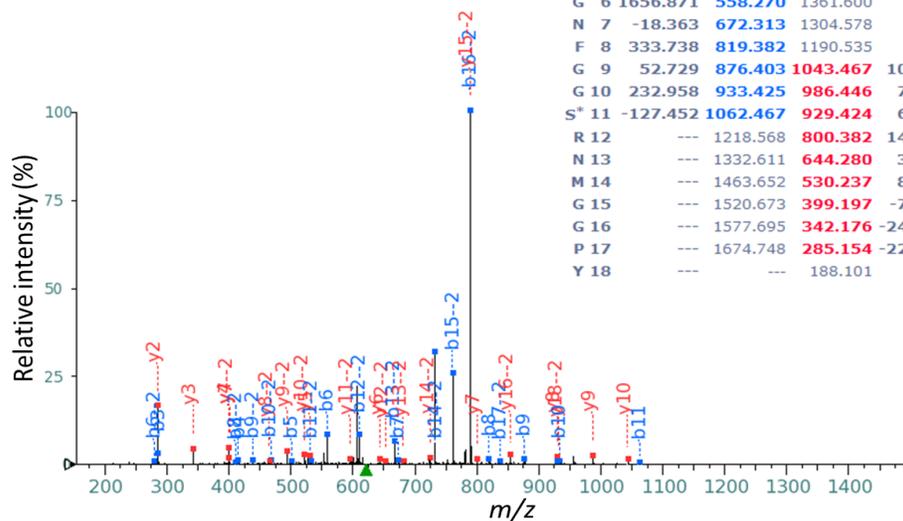
hnRNPA2  
GPMKS\*GNFGGSRNMGGPY



Seq	#	b: Δ Error	b	y	y: Δ Error	+1
G	1	---	58.029	---	---	18
P	2	---	155.082	1804.820	---	17
M	3	-3191.943	<b>286.122</b>	1707.767	---	16
K	4	-315.213	<b>414.217</b>	<b>1576.727</b>	-61.413	15
S*	5	-429.740	<b>543.260</b>	<b>1448.632</b>	18.601	14
G	6	-177.157	<b>600.281</b>	<b>1319.589</b>	-4.656	13
N	7	-119.155	<b>714.324</b>	<b>1262.568</b>	41.717	12
F	8	7.705	<b>861.392</b>	<b>1148.525</b>	54.325	11
G	9	899.093	<b>918.414</b>	<b>1001.456</b>	13.238	10
G	10	78.799	<b>975.435</b>	<b>944.435</b>	148.803	9
S	11	349.020	<b>1062.467</b>	<b>887.414</b>	113.646	8
R	12	90.003	<b>1218.568</b>	<b>800.382</b>	242.995	7
N	13	71.983	<b>1332.611</b>	<b>644.280</b>	334.856	6
M	14	-42.351	<b>1463.652</b>	<b>530.237</b>	224.552	5
G	15	82.626	<b>1520.673</b>	<b>399.197</b>	50.225	4
G	16	40.817	<b>1577.695</b>	<b>342.176</b>	-183.550	3
P	17	571.509	<b>1674.748</b>	<b>285.154</b>	201.766	2
Y	18	---	---	188.101	---	1

(C)

hnRNPA2  
GPMKSGNFGGS\*RNMGGPY



Seq	#	b: Δ Error	b	y	y: Δ Error	+1
G	1	---	58.029	---	---	18
P	2	---	155.082	1804.820	---	17
M	3	-358.217	<b>286.122</b>	1707.767	---	16
K	4	219.942	<b>414.217</b>	1576.727	---	15
S	5	77.403	<b>501.249</b>	1448.632	---	14
G	6	1656.871	<b>558.270</b>	1361.600	---	13
N	7	-18.363	<b>672.313</b>	1304.578	---	12
F	8	333.738	<b>819.382</b>	1190.535	---	11
G	9	52.729	<b>876.403</b>	<b>1043.467</b>	105.282	10
G	10	232.958	<b>933.425</b>	<b>986.446</b>	75.219	9
S*	11	-127.452	<b>1062.467</b>	<b>929.424</b>	64.250	8
R	12	---	1218.568	<b>800.382</b>	145.119	7
N	13	---	1332.611	<b>644.280</b>	36.460	6
M	14	---	1463.652	<b>530.237</b>	86.349	5
G	15	---	1520.673	<b>399.197</b>	-70.794	4
G	16	---	1577.695	<b>342.176</b>	-241.635	3
P	17	---	1674.748	<b>285.154</b>	-222.047	2
Y	18	---	---	188.101	---	1

## BIBLIOGRAPHY

- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A Systematic Survey Identifies Prions and Illuminates Sequence Features of Prionogenic Proteins. *Cell* 137, 146-158.
- Allen, B.L., and Taatjes, D.J. (2015). The Mediator Complex: A Central Integrator of Transcription. *Nat Rev Mol Cell Biol* 16, 155-166.
- Altmeyer, M., Neelsen, K.J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grofte, M., Rask, M.-B.D., Streicher, W., Jungmichel, S., Nielsen, M.L., *et al.* (2015). Liquid Demixing of Intrinsically Disordered Proteins Is Seeded by Poly(Adp-Ribose). *Nat Commun* 6.
- Anderson, P., and Kedersha, N. (2009). Rna Granules: Post-Transcriptional and Epigenetic Modulators of Gene Expression. *Nat Rev Mol Cell Biol* 10, 430-436.
- Andersson, M.K., Ståhlberg, A., Arvidsson, Y., Olofsson, A., Semb, H., Stenman, G., Nilsson, O., and Åman, P. (2008). The Multifunctional Fus, Ews and Taf15 Proto-Oncoproteins Show Cell Type-Specific Expression Patterns and Involvement in Cell Spreading and Stress Response. *BMC Cell Biology* 9, 1-17.
- Arvand, A., and Denny, C.T. (2001). Biology of Ews/Ets Fusions in Ewing's Family Tumors. *Oncogene* 20, 5747-5754.
- Astbury, W.T., Dickinson, S., and Bailey, K. (1935). The X-Ray Interpretation of Denaturation and the Structure of the Seed Globulins. *Biochemical Journal* 29, 2351-2360.2351.
- Aubol, B.E., Plocinik, R.M., Hagopian, J.C., Ma, C.-T., McGlone, M.L., Bandyopadhyay, R., Fu, X.-D., and Adams, J.A. (2013). Partitioning Rs Domain Phosphorylation in an Sr Protein through the Clk and SrpK Protein Kinases. *Journal of Molecular Biology* 425, 2894-2909.
- Back, J.W., de Jong, L., Muijsers, A.O., and de Koster, C.G. (2003). Chemical Cross-Linking and Mass Spectrometry for Protein Structural Modeling. *Journal of Molecular Biology* 331, 303-313.
- Banjade, S., and Rosen, M.K. (2014). Phase Transitions of Multivalent Proteins Can Promote Clustering of Membrane Receptors. *eLife* 3, e04123.
- Banjade, S., Wu, Q., Mittal, A., Peeples, W.B., Pappu, R.V., and Rosen, M.K. (2015). Conserved Interdomain Linker Promotes Phase Separation of the Multivalent

- Adaptor Protein Nck. *Proceedings of the National Academy of Sciences* *112*, E6426-E6435.
- Baxa, U., Wickner, R.B., Steven, A.C., Anderson, D.E., Marekov, L.N., Yau, W.M., and Tycko, R. (2007). Characterization of Beta-Sheet Structure in Ure2p1-89 Yeast Prion Fibrils by Solid-State Nuclear Magnetic Resonance. *Biochemistry* *46*, 13149-13162.
- Bentmann, E., Haass, C., and Dormann, D. (2013). Stress Granules in Neurodegeneration--Lessons Learnt from Tar DNA Binding Protein of 43 Kda and Fused in Sarcoma. *FEBS J* *280*, 4348-4370.
- Berry, J., Weber, S.C., Vaidya, N., Haataja, M., and Brangwynne, C.P. (2015). Rna Transcription Modulates Phase Transition-Driven Nuclear Body Assembly. *Proceedings of the National Academy of Sciences* *112*, E5237-E5245.
- Bertolotti, A., Bell, B., and Tora, L. (1999). The N-Terminal Domain of Human Tafii68 Displays Transactivation and Oncogenic Properties. *Oncogene* *18*, 8000-8010.
- Borggreffe, T., Davis, R., Erdjument-Bromage, H., Tempst, P., and Kornberg, R.D. (2002). A Complex of the Srb8, -9, -10, and -11 Transcriptional Regulatory Proteins from Yeast. *Journal of Biological Chemistry* *277*, 44202-44207.
- Bosco, D.A., Lemay, N., Ko, H.K., Zhou, H., Burke, C., Kwiatkowski, T.J., Sapp, P., McKenna-Yasek, D., Brown, R.H., and Hayward, L.J. (2010). Mutant Fus Proteins That Cause Amyotrophic Lateral Sclerosis Incorporate into Stress Granules. *Human Molecular Genetics* *19*, 4160-4175.
- Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* *324*, 1729-1732.
- Brangwynne, C.P., Mitchison, T.J., and Hyman, A.A. (2011). Active Liquid-Like Behavior of Nucleoli Determines Their Size and Shape in *Xenopus Laevis* Oocytes. *Proceedings of the National Academy of Sciences* *108*, 4334-4339.
- Broide, M.L., Berland, C.R., Pande, J., Ogun, O.O., and Benedek, G.B. (1991). Binary-Liquid Phase Separation of Lens Protein Solutions. *Proceedings of the National Academy of Sciences* *88*, 5660-5664.
- Buchan, J.R., and Parker, R. (2009). Eukaryotic Stress Granules: The Ins and Outs of Translation. *Molecular cell* *36*, 932-941.

- Bungenberg de Jong, H.G. (1949). Crystallisation- Coacervation- Flocculation. In *Colloid Science*, H.R. Kruyt, ed. (Amsterdam: Elsevier), pp. 232-258.
- Buratowski, S. (2009). Progression through the Rna Polymerase Ii Ctd Cycle. *Molecular cell* 36, 541-546.
- Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* 325, 834-840.
- Cioce, M., and Lamond, A.I. (2005). Cajal Bodies: A Long History of Discovery. *Annual Review of Cell and Developmental Biology* 21, 105-131.
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty Protein Kinase Phosphorylates Sr Splicing Factors and Regulates Their Intranuclear Distribution. *The EMBO Journal* 15, 265-275.
- Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A Unique Structure at the Carboxyl Terminus of the Largest Subunit of Eukaryotic Rna Polymerase Ii. *Proceedings of the National Academy of Sciences* 82, 7934-7938.
- Costantini, S., Colonna, G., and Facchiano, A.M. (2006). Amino Acid Propensities for Secondary Structures Are Influenced by the Protein Structural Class. *Biochemical and Biophysical Research Communications* 342, 441-451.
- Crozat, A., Aman, P., Mandahl, N., and Ron, D. (1993). Fusion of Chop to a Novel Rna-Binding Protein in Human Myxoid Liposarcoma. *Nature* 363, 640-644.
- Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a Glutamine/Asparagine-Rich Domain of Lsm4p Function in Processing Body Assembly in *Saccharomyces Cerevisiae*. *The Journal of Cell Biology* 179, 437-449.
- Dobson, C.M. (2003). Protein Folding and Misfolding. *Nature* 426, 884-890.
- Dulbecco, R., and Freeman, G. (1959). Plaque Production by the Polyoma Virus. *Virology* 8, 396-397.
- Duncan, P.I., Stojdl, D.F., Marius, R.M., Scheit, K.H., and Bell, J.C. (1998). The Clk2 and Clk3 Dual-Specificity Protein Kinases Regulate the Intranuclear Distribution of Sr Proteins and Influence Pre-Mrna Splicing. *Experimental Cell Research* 241, 300-308.
- Dundr, M., and Misteli, T. (2010). Biogenesis of Nuclear Bodies. *Cold Spring Harbor Perspectives in Biology* 2, a000711.

- Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C.C.-H., Eckmann, C.R., Myong, S., and Brangwynne, C.P. (2015). The Disordered P Granule Protein Laf-1 Drives Phase Separation into Droplets with Tunable Viscosity and Dynamics. *Proceedings of the National Academy of Sciences* *112*, 7189-7194.
- Emery, A.E.H. (2002). The Muscular Dystrophies. *The Lancet* *359*, 687-695.
- Frey, S., and Gorlich, D. (2007). A Saturated Fg-Repeat Hydrogel Can Reproduce the Permeability Properties of Nuclear Pore Complexes. *Cell* *130*, 512-523.
- Frey, S., Richter, R.P., and Gorlich, D. (2006). Fg-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. *Science* *314*, 815-817.
- Gallo, C.M., Wang, J.T., Motegi, F., and Seydoux, G. (2010). Cytoplasmic Partitioning of P Granule Components Is Not Required to Specify the Germline in *C. Elegans*. *Science* *330*, 1685-1689.
- Geddes, A.J., Parker, K.D., Atkins, E.D.T., and Beighton, E. (1968). "Cross-B" Conformation in Proteins. *Journal of Molecular Biology* *32*, 343-358.
- Ghamari, A., van de Corput, M.P.C., Thongjuea, S., van Cappellen, W.A., van IJcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B., and Grosveld, F.G. (2013). In Vivo Live Imaging of Rna Polymerase Ii Transcription Factories in Primary Cells. *Genes & Development* *27*, 767-777.
- Gilar, M., Olivova, P., Daly, A.E., and Gebler, J.C. (2005). Two-Dimensional Separation of Peptides Using Rp-Rp-Hplc System with Different Ph in First and Second Separation Dimensions. *J Sep Sci* *28*, 1694-1703.
- Guipaud, O., Guillonneau, F., Labas, V., Praseuth, D., Rossier, J., Lopez, B., and Bertrand, P. (2006). An in Vitro Enzymatic Assay Coupled to Proteomics Analysis Reveals a New DNA Processing Activity for Ewing Sarcoma and Taf(Ii)68 Proteins. *PROTEOMICS* *6*, 5962-5972.
- Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., *et al.* (2012). Cell-Free Formation of Rna Granules: Bound Rnas Identify Features and Components of Cellular Assemblies. *Cell* *149*, 768-779.
- Hennig, S., Kong, G., Mannen, T., Sadowska, A., Kobelke, S., Blythe, A., Knott, G.J., Iyer, K.S., Ho, D., Newcombe, E.A., *et al.* (2015). Prion-Like Domains in Rna Binding Proteins Are Essential for Building Subnuclear Paraspeckles. *The Journal of Cell Biology* *210*, 529-539.

- Hobbs, N.K., Bondareva, A.A., Barnett, S., Capecchi, M.R., and Schmidt, E.E. (2002). Removing the Vertebrate-Specific Tbp N Terminus Disrupts Placental B2m-Dependent Interactions with the Maternal Immune System. *Cell* 110, 43-54.
- Hope, I.A., Mahadevan, S., and Struhl, K. (1988). Structural and Functional Characterization of the Short Acidic Transcriptional Activation Region of Yeast Gcn4 Protein. *Nature* 333, 635-640.
- Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villen, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. *Cell* 143, 1174-1189.
- Jencks, W.P., and Carriuolo, J. (1959). Imidazole Catalysis: Ii. Acyl Transfer and the Reactions of Acetyl Imidazole with Water and Oxygen Anions. *Journal of Biological Chemistry* 234, 1272-1279.
- Jiang, H., Wang, S., Huang, Y., He, X., Cui, H., Zhu, X., and Zheng, Y. (2015). Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. *Cell* 163, 108-122.
- Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., *et al.* (2012). Cell-Free Formation of Rna Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* 149, 753-767.
- Katta, V., and Chait, B.T. (1991). Conformational Changes in Proteins Probed by Hydrogen-Exchange Electrospray-Ionization Mass Spectrometry. *Rapid Commun Mass Spectrom* 5, 214-217.
- Keshwani, Malik M., Hailey, Kendra L., Aubol, Brandon E., Fattet, L., McGlone, Maria L., Jennings, Patricia A., and Adams, Joseph A. (2015). Nuclear Protein Kinase Clk1 Uses a Non-Traditional Docking Mechanism to Select Physiological Substrates. *Biochemical Journal* 472, 329-338.
- 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.* (2013). Mutations in Prion-Like Domains in Hnrnpa2b1 and Hnrnpa1 Cause Multisystem Proteinopathy and Als. *Nature* 495, 467-473.
- Knowles, R.B., and Kosik, K.S. (1997). Neurotrophin-3 Signals Redistribute Rna in Neurons. *Proceedings of the National Academy of Sciences of the United States of America* 94, 14804-14808.
- Knuesel, M.T., Meyer, K.D., Donner, A.J., Espinosa, J.M., and Taatjes, D.J. (2009). The Human Cdk8 Subcomplex Is a Histone Kinase That Requires Med12 for Activity

- and Can Function Independently of Mediator. *Molecular and Cellular Biology* 29, 650-661.
- Koehl, P., and Levitt, M. (1999). Structure-Based Conformational Preferences of Amino Acids. *Proceedings of the National Academy of Sciences of the United States of America* 96, 12524-12529.
- Konermann, L., Pan, J., and Liu, Y.H. (2011). Hydrogen Exchange Mass Spectrometry for Studying Protein Structure and Dynamics. *Chem Soc Rev* 40, 1224-1234.
- Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J.L., and McKnight, S.L. (2013). Phosphorylation-Regulated Binding of Rna Polymerase Ii to Fibrous Polymers of Low-Complexity Domains. *Cell* 155, 1049-1060.
- Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S.L. (2014). Poly-Dipeptides Encoded by the C9orf72 Repeats Bind Nucleoli, Impede Rna Biogenesis, and Kill Cells. *Science* 345, 1139-1145.
- Lamond, A.I., and Spector, D.L. (2003). Nuclear Speckles: A Model for Nuclear Organelles. *Nat Rev Mol Cell Biol* 4, 605-612.
- Lauranzano, E., Pozzi, S., Pasetto, L., Stucchi, R., Massignan, T., Paoletta, K., Mombrini, M., Nardo, G., Lunetta, C., Corbo, M., *et al.* (2015). Peptidylprolyl Isomerase a Governs Tardbp Function and Assembly in Heterogeneous Nuclear Ribonucleoprotein Complexes. *Brain* 138, 974-991.
- Lee, E.K. (2012). Post-Translational Modifications of Rna-Binding Proteins and Their Roles in Rna Granules. *Curr Protein Pept Sci* 13, 331-336.
- Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. *Annual Review of Genetics* 34, 77-137.
- Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., *et al.* (2012). Phase Transitions in the Assembly of Multivalent Signalling Proteins. *Nature* 483, 336-340.
- Lin, Y., Protter, David S.W., Rosen, Michael K., and Parker, R. (2015). Formation and Maturation of Phase-Separated Liquid Droplets by Rna-Binding Proteins. *Molecular cell* 60, 208-219.
- Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006). Intrinsic Disorder in Transcription Factors. *Biochemistry* 45, 6873-6888.

- Loncle, N., Boube, M., Joulia, L., Boschiero, C., Werner, M., Cribbs, D.L., and Bourbon, H.M. (2007). Distinct Roles for Mediator Cdk8 Module Subunits in *Drosophila* Development. *The EMBO Journal* *26*, 1045-1054.
- Lorch, Y., and Kornberg, R.D. (2015). Chromatin-Remodeling and the Initiation of Transcription. *Quarterly Reviews of Biophysics* *48*, 465-470.
- Lunde, B.M., Moore, C., and Varani, G. (2007). Rna-Binding Proteins: Modular Design for Efficient Function. *Nat Rev Mol Cell Biol* *8*, 479-490.
- Ly, T., and Julian, R.R. (2006). Using Esi-MS to Probe Protein Structure by Site-Specific Noncovalent Attachment of 18-Crown-6. *Journal of the American Society for Mass Spectrometry* *17*, 1209-1215.
- Malfois, M., Bonneté, F., Belloni, L., and Tardieu, A. (1996). A Model of Attractive Interactions to Account for Fluid–Fluid Phase Separation of Protein Solutions. *The Journal of Chemical Physics* *105*, 3290-3300.
- Mao, Y.S., Zhang, B., and Spector, D.L. (2011). Biogenesis and Function of Nuclear Bodies. *Trends Genet* *27*, 295-306.
- Marko, J.F. (2012). The Liquid Drop Nature of Nucleoli. *Nucleus* *3*, 115-117.
- Mastrocola, A.S., Kim, S.H., Trinh, A.T., Rodenkirch, L.A., and Tibbetts, R.S. (2013). The Rna-Binding Protein Fused in Sarcoma (Fus) Functions Downstream of Poly(Adp-Ribose) Polymerase (Parp) in Response to DNA Damage. *Journal of Biological Chemistry* *288*, 24731-24741.
- May, W.A., Lessnick, S.L., Braun, B.S., Klemsz, M., Lewis, B.C., Lunsford, L.B., Hromas, R., and Denny, C.T. (1993). The Ewing's Sarcoma Ews/Fli-1 Fusion Gene Encodes a More Potent Transcriptional Activator and Is a More Powerful Transforming Gene Than Fli-1. *Molecular and Cellular Biology* *13*, 7393-7398.
- McKinley, M.P., Bolton, D.C., and Prusiner, S.B. (1983). A Protease-Resistant Protein Is a Structural Component of the Scrapie Prion. *Cell* *35*, 57-62.
- Meersman, F., and Dobson, C.M. (2006). Probing the Pressure–Temperature Stability of Amyloid Fibrils Provides New Insights into Their Molecular Properties. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* *1764*, 452-460.
- Mendoza, V.L., and Vachet, R.W. (2009). Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrometry Reviews* *28*, 785-815.

- Menegay, H.J., Myers, M.P., Moeslein, F.M., and Landreth, G.E. (2000). Biochemical Characterization and Localization of the Dual Specificity Kinase Clk1. *Journal of Cell Science* *113*, 3241-3253.
- Mitrea, D.M., Cika, J.A., Guy, C.S., Ban, D., Banerjee, P.R., Stanley, C.B., Nourse, A., Deniz, A.A., and Kriwacki, R.W. (2016). Nucleophosmin Integrates within the Nucleolus Via Multi-Modal Interactions with Proteins Displaying R-Rich Linear Motifs and Rrna. *eLife* *5*, e13571.
- Miyashiro, J., Woods, K.W., Park, C.H., Liu, X., Shi, Y., Johnson, E.F., Bouska, J.J., Olson, A.M., Luo, Y., Fry, E.H., *et al.* (2009). Synthesis and Sar of Novel Tricyclic Quinoxalinone Inhibitors of Poly(Adp-Ribose)Polymerase-1 (Parp-1). *Bioorg Med Chem Lett* *19*, 4050-4054.
- Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, Anderson P., Kim, Hong J., Mittag, T., and Taylor, J.P. (2015). Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* *163*, 123-133.
- Murakami, T., Qamar, S., Lin, Julie Q., Schierle, Gabriele S.K., Rees, E., Miyashita, A., Costa, Ana R., Dodd, Roger B., Chan, Fiona T.S., Michel, Claire H., *et al.* (2015). Als/Ftd Mutation-Induced Phase Transition of Fus Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs Rnp Granule Function. *Neuron* *88*, 678-690.
- Muschol, M., and Rosenberger, F. (1997). Liquid–Liquid Phase Separation in Supersaturated Lysozyme Solutions and Associated Precipitate Formation/Crystallization. *The Journal of Chemical Physics* *107*, 1953-1962.
- Myers, B., and Glazer, A.N. (1971). Spectroscopic Studies of the Exposure of Tyrosine Residues in Proteins with Special Reference to the Subtilisins. *Journal of Biological Chemistry* *246*, 412-419.
- Nott, Timothy J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T.D., Bazett-Jones, David P., Pawson, T., Forman-Kay, Julie D., *et al.* (2015). Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular cell* *57*, 936-947.
- Olsen, J.V., Schwartz, J.C., Griep-Raming, J., Nielsen, M.L., Damoc, E., Denisov, E., Lange, O., Remes, P., Taylor, D., Splendore, M., *et al.* (2009). A Dual Pressure Linear Ion Trap Orbitrap Instrument with Very High Sequencing Speed. *Mol Cell Proteomics* *8*, 2759-2769.

- Pabo, C.O., and Sauer, R.T. (1992). Transcription Factors: Structural Families and Principles of DNA Recognition. *Annu Rev Biochem* 61, 1053-1095.
- Parthasarathy, S., and Murthy, M.R. (1997). Analysis of Temperature Factor Distribution in High-Resolution Protein Structures. *Protein Science : A Publication of the Protein Society* 6, 2561-2567.
- Patel, A., Lee, Hyun O., Jawerth, L., Maharana, S., Jahnel, M., Hein, Marco Y., Stoykov, S., Mahamid, J., Saha, S., Franzmann, Titus M., *et al.* (2015). A Liquid-to-Solid Phase Transition of the Als Protein Fus Accelerated by Disease Mutation. *Cell* 162, 1066-1077.
- Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and Functions of the Rna Polymerase Ii Ctd. *Genes & Development* 20, 2922-2936.
- Rabbitts, T.H., Forster, A., Larson, R., and Nathan, P. (1993). Fusion of the Dominant Negative Transcription Regulator Chop with a Novel Gene Fus by Translocation T(12;16) in Malignant Liposarcoma. *Nat Genet* 4, 175-180.
- Radivojac, P., Obradovic, Z., Smith, D.K., Zhu, G., Vucetic, S., Brown, C.J., Lawson, J.D., and Dunker, A.K. (2004). Protein Flexibility and Intrinsic Disorder. *Protein Science* 13, 71-80.
- Riggi, N., Cironi, L., Suva, M.L., and Stamenkovic, I. (2007). Sarcomas: Genetics, Signalling, and Cellular Origins. Part 1: The Fellowship of Tet. *J Pathol* 213, 4-20.
- Riordan, J.F., and Vallee, B.L. (1972). [42] O-Acetyltyrosine. In *Methods in Enzymology* (Academic Press), pp. 500-506.
- Riordan, J.F., Wacker, W.E.C., and Vallee, B.L. (1965a). N-Acetylimidazole: A Reagent for Determination of "Free" Tyrosyl Residues of Proteins. *Biochemistry* 4, 1758-1765.
- Riordan, J.F., Wacker, W.E.C., and Vallee, B.L. (1965b). N-Acetylimidazole: A Reagent for Determination of "Free" Tyrosyl Residues of Proteins\*. *Biochemistry* 4, 1758-1765.
- Rook, M.S., Lu, M., and Kosik, K.S. (2000). Camkii $\alpha$  3' Untranslated Region-Directed Mrna Translocation in Living Neurons: Visualization by Gfp Linkage. *The Journal of Neuroscience* 20, 6385-6393.
- Rufer, A.C., Thiebach, L., Baer, K., Klein, H.W., and Hennig, M. (2005). X-Ray Structure of Glutathione S-Transferase from *Schistosoma Japonicum* in a New

- Crystal Form Reveals Flexibility of the Substrate-Binding Site. *Acta Crystallogr Sect F Struct Biol Cryst Commun* *61*, 263-265.
- Sánchez-García, I., and Rabbitts, T.H. (1994). Transcriptional Activation by Tal1 and Fus-Chop Proteins Expressed in Acute Malignancies as a Result of Chromosomal Abnormalities. *Proceedings of the National Academy of Sciences* *91*, 7869-7873.
- Schmidt, E.E., Bondareva, A.A., Radke, J.R., and Capecchi, M.R. (2003). Fundamental Cellular Processes Do Not Require Vertebrate-Specific Sequences within the Tata-Binding Protein. *Journal of Biological Chemistry* *278*, 6168-6174.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). Nih Image to Imagej: 25 Years of Image Analysis. *Nat Methods* *9*, 671-675.
- Shao, J., and Diamond, M.I. (2007). Polyglutamine Diseases: Emerging Concepts in Pathogenesis and Therapy. *Human Molecular Genetics* *16*, R115-R123.
- Sigler, P.B. (1988). Transcriptional Activation. Acid Blobs and Negative Noodles. *Nature* *333*, 210-212.
- Simpson, R.T., Riordan, J.F., and Vallee, B.L. (1963). Functional Tyrosyl Residues in the Active Center of Bovine Pancreatic Carboxypeptidase a\*. *Biochemistry* *2*, 616-622.
- Sinz, A. (2006). Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein-Protein Interactions. *Mass Spectrometry Reviews* *25*, 663-682.
- Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. *Cold Spring Harbor Perspectives in Biology* *3*.
- Strome, S., and Wood, W.B. (1983). Generation of Asymmetry and Segregation of Germ-Line Granules in Early C. *Elegans* Embryos. *Cell* *35*, 15-25.
- Su, X., Ditlev, J.A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D.S., Taunton, J., Rosen, M.K., and Vale, R.D. (2016). Phase Separation of Signaling Molecules Promotes T Cell Receptor Signal Transduction. *Science*.
- Sunde, M., and Blake, C. (1997). The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. In *Advances in Protein Chemistry*, D.S.E. Frederic M. Richards, and S.K. Peter, eds. (Academic Press), pp. 123-159.
- Tanaka, S., Yamamoto, M., Ito, K., Hayakawa, R., and Ataka, M. (1997). Relation between the Phase Separation and the Crystallization in Protein Solutions. *Physical Review E* *56*, R67-R69.

- Taratuta, V.G., Holschbach, A., Thurston, G.M., Blankschtein, D., and Benedek, G.B. (1990). Liquid-Liquid Phase Separation of Aqueous Lysozyme Solutions: Effects of Ph and Salt Identity. *The Journal of Physical Chemistry* *94*, 2140-2144.
- Tardieu, A., V  r  tout, F., Krop, B., and Slingsby, C. (1992). Protein Interactions in the Calf Eye Lens: Interactions between B-Crystallins Are Repulsive Whereas in  $\Gamma$ -Crystallins They Are Attractive. *European Biophysics Journal* *21*, 1-12.
- Taylor, J.P. (2015). Multisystem Proteinopathy: Intersecting Genetics in Muscle, Bone, and Brain Degeneration. *Neurology* *85*, 658-660.
- Thomson, J.A., Schurtenberger, P., Thurston, G.M., and Benedek, G.B. (1987). Binary Liquid Phase Separation and Critical Phenomena in a Protein/Water Solution. *Proceedings of the National Academy of Sciences* *84*, 7079-7083.
- Timasheff, S.N., and Gorbunoff, M.J. (1967). Conformation of Proteins. *Annual Review of Biochemistry* *36*, 13-54.
- Triezenberg, S.J. (1995). Structure and Function of Transcriptional Activation Domains. *Curr Opin Genet Dev* *5*, 190-196.
- Tullius, T.D., and Dombroski, B.A. (1986). Hydroxyl Radical "Footprinting": High-Resolution Information About DNA-Protein Contacts and Application to Lambda Repressor and Cro Protein. *Proceedings of the National Academy of Sciences* *83*, 5469-5473.
- Uematsu, N., Weterings, E., Yano, K.-i., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P.-O., van Gent, D.C., Chen, B.P.C., and Chen, D.J. (2007). Autophosphorylation of DNA-Pkcs Regulates Its Dynamics at DNA Double-Strand Breaks. *The Journal of Cell Biology* *177*, 219-229.
- van de Peppel, J., Kettelarij, N., van Bakel, H., Kockelkorn, T.T.J.P., van Leenen, D., and Holstege, F.C.P. (2005). Mediator Expression Profiling Epistasis Reveals a Signal Transduction Pathway with Antagonistic Submodules and Highly Specific Downstream Targets. *Molecular cell* *19*, 511-522.
- Vieira, N.M., Naslavsky, M.S., Licinio, L., Kok, F., Schlesinger, D., Vainzof, M., Sanchez, N., Kitajima, J.P., Gal, L., Cava  ana, N., *et al.* (2014). A Defect in the Rna-Processing Protein Hnrpd1 Causes Limb-Girdle Muscular Dystrophy 1g (Lgmd1g). *Human Molecular Genetics*.
- Voronina, E., Seydoux, G., Sassone-Corsi, P., and Nagamori, I. (2011). Rna Granules in Germ Cells. *Cold Spring Harbor Perspectives in Biology*.

- Wang, J.-C., Walker, A., Blackwell, T.K., and Yamamoto, K.R. (2004). The *Caenorhabditis Elegans* Ortholog of Trap240, Cetrp240/Let-19, Selectively Modulates Gene Expression and Is Essential for Embryogenesis. *Journal of Biological Chemistry* 279, 29270-29277.
- Wang, J.T., Smith, J., Chen, B.-C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B.G., Calidas, D., Betzig, E., and Seydoux, G. (2014). Regulation of Rna Granule Dynamics by Phosphorylation of Serine-Rich, Intrinsically Disordered Proteins in *C. Elegans*. *eLife* 3, e04591.
- Wang, L., and Chance, M.R. (2011). Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting. *Analytical Chemistry* 83, 7234-7241.
- Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., and Kimonis, V.E. (2004). Inclusion Body Myopathy Associated with Paget Disease of Bone and Frontotemporal Dementia Is Caused by Mutant Valosin-Containing Protein. *Nat Genet* 36, 377-381.
- Wehner, K.A., Schütz, S., and Sarnow, P. (2010). Ogfod1, a Novel Modulator of Eukaryotic Translation Initiation Factor 2 $\alpha$  Phosphorylation and the Cellular Response to Stress. *Molecular and Cellular Biology* 30, 2006-2016.
- Wickner, R.B. (1994). [Ure3] as an Altered Ure2 Protein: Evidence for a Prion Analog in *Saccharomyces Cerevisiae*. *Science* 264, 566-569.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the Ccp4 Suite and Current Developments. *Acta Crystallogr D Biol Crystallogr* 67, 235-242.
- Zhang, Y., Wang, J., Ding, M., and Yu, Y. (2013). Site-Specific Characterization of the Asp- and Glu-Adp-Ribosylated Proteome. *Nat Methods* 10, 981-984.
- Zoghbi, H.Y., and Orr, H.T. (2000). Glutamine Repeats and Neurodegeneration. *Annual Review of Neuroscience* 23, 217-247.
- Zydowsky, L.D., Etkorn, F.A., Chang, H.Y., Ferguson, S.B., Stolz, L.A., Ho, S.I., and Walsh, C.T. (1992). Active Site Mutants of Human Cyclophilin a Separate Peptidyl-Prolyl Isomerase Activity from Cyclosporin a Binding and Calcineurin Inhibition. *Protein Sci* 1, 1092-1099.