INTERROGATING PSF'S RRM/RNA BINDING-MODE

Krystal Haislop

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2022

Supervisor of Dissertation

_____

Kristen W. Lynch

Benjamin Rush Professor and Chair, Department of Biochemistry and Biophysics

Graduate Group Chairperson

_____

Kim Sharp, Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Kushol Gupta, Research Assistant Professor of Biochemistry and Biophysics

Elizabeth Rhoades, Associate Professor of Chemistry

David W. Speicher, Professor, Wistar Institute

Matthew Weitzman, Professor of Pathology and Laboratory Medicine (Children's Hospital of Philadelphia)

I dedicate this thesis to my late father Michael Dean Haislop, who taught me to be kind, logical, persistent, and work hard for all I desire.

ACKNOWLEDGMENT

I would like to thank my advisor Kristen Lynch for her unending support throughout graduate school, scientifically and other. I remember to celebrate victories every step of the way because of you.

I would like to thank the National Institutes of Health for their financial support of my research.

I would like to thank my thesis committee for their guidance and advice.

I would like to thank members of the Biochemistry and Biophysics department, especially Kushol Gupta and Leland Mayne, for their technical assistance and guidance with this project.

I would like to thank the members of our lab, past and present, for being a family and home away from home. Thank you for celebrating the highs and making difficult times better. I'd like to thank Michael Mallory for helping scientifically, watering my plants when I was away, and helping me grow a 6-ft tall money tree in lab.

I would like to thank Sehe Han for being like a sister to me, supporting me in all endeavors, and inspiring me with your radical authenticity and fervor for life.

I would like to thank Adam Kessel for being my friend, physical therapist, and editor on retainer. Thank you for editing this thesis and helping my recovery from back injury and many concussions.

I would like to thank my classmates and friends for making graduate school and life a joyful and insightful experience. Thank you to my climbing community for being a space for my brain to rest and play, and for helping me get back on my feet after falling.

Finally, I would like to thank family for their support: my brother Michael, my sister Courtney, my sister Libby, my mother Chantal, and my late father Michael. Thank you, Mom, for consoling me after difficult days and giving me the strength push forward. I am deeply grateful.

ABSTRACT

INTERROGATING PTB-ASSOCIATED SPLICING FACTOR'S RNA RECOGNITION MOTIF/RNA BINDING-MODE

Krystal Haislop

Kristen W. Lynch

PTB-associated splicing factor (PSF) is a multifunctional nucleic acid binding protein vital for cell survival. PSF plays several roles throughout the cell, yet many aspects of PSF's mechanisms of interacting remain elusive. While protein and nucleic acid binding partners of PSF have been identified, binding consensus sequences reported in the literature vary. Previous studies in the Lynch Laboratory have found that PSF's second RNA recognition motif (RRM2) is necessary and sufficient for binding, but this binding can be occluded by interacting with cofactor protein TRAP150. Additionally, binding is dependent on phosphorylation of PSF in T-cells. Identifying the interaction interfaces of PSF/RNA and PSF/TRAP150 yields insight into the regulatory mechanism governing RNA-binding. First, I utilize biophysical techniques to overcome challenges characterizing PSF/RNA. The smallest RNA to bind with low/modest affinity is 65-nucleotides long, making the protein/RNA pair nonideal for crystallization or NMR experiments. PSF has long N- and C-terminal unstructured regions that aid in PSF's functional aggregation, creating another challenge for biophysical characterization. Here I use several biophysical techniques to characterize PSF's dynamics and binding interfaces. Mass spectrometry experiments indicate PSF contacts RNA using 30-60% of the accessible surface area. Interestingly, a charged hydrophobic loop region when mutated from "DDRGR" to "AAAAA" increases affinity for ESS-RNA 10-fold. A neighboring residue Lysine-413 also increases affinity for RNA when mutated to alanine. Arginine-474 and Threoinine-485 within the NOPS domain

lose their ability to bind RNA when mutated. It is not clear whether these residues affect local conformation of playing a direct role in binding. Additionally, I identify hundreds of PSF-dependent alternative polyadenylation changes and begin characterization of PSF/TRAP150 and phosphorylated PSF by HDX-MS. Together, these results indicate PSF interacts with RNA using a large portion of the surface accessible area to do so, and PSF uses a loop in RRM2 to aid in RNA-interaction.

TABLE OF CONTENTS

DEDICATION ..... ERROR! BOOKMARK NOT DEFINED.

ACKNOWLEDGMENT III

ABSTRACT IV

TABLE OF CONTENTS VI

LIST OF TABLES IX

LIST OF ILLUSTRATIONS X

CHAPTER 1: INTRODUCTION 1

WHAT IS AN RNA-BINDING PROTEIN? 1

PROCESSING OF MESSENGER RNA 3

STRUCTURE OF RNA-BINDING PROTEINS 9

REGULATION OF RNA-BINDING PROTEINS 14

POLYPYRIMIDINE TRACT-BINDING PROTEIN-ASSOCIATED SPLICING FACTOR 16

CURRENT INFORMATION ON PSF'S RNA-BINDING MODE 21

CHAPTER 2: BIOPHYSICAL PROBING OF PSF'S RNA-BINDING INTERFACE 24

INTRODUCTION 24

RESULTS 28

Crosslinking reveals an extended RNA-interacting interface 28

PSF structure, dynamics, and RNA interactions via Hydrogen-Deuterium Exchange Mass spectrometry 32

Comparative mapping of RNA interactions indicates similar regions contact RNA 37

RRM1/RNA interactions by Nuclear Magnetic Resonance 40

Hydrophobic surface of PSF via molecular simulation 40

DISCUSSION 46

CHAPTER 3: MUTATIONAL BINDING ANALYSIS 49

INTRODUCTION **49**

RESULTS **49**
Design of mutants 49
Threonine-485 to alanine mutation decreases affinity for RNA 52
Arginine-474 to alanine mutation decreases affinity for RNA 54
β2β3-loop gains affinity for RNA upon mutation to "AAAAA" 54
K413A increases affinity for ESS-RNA 59
Binding modality shared between multiple target RNAs 59

DISCUSSION **60**

CHAPTER 4: PSF'S ROLE IN ALTERNATIVE POLYADENYLATION IN T-CELLS **63**

INTRODUCTION **63**

RESULTS **64**
Determining PSF responsive alternative polyadenylation targets 64
Validating alternative polyadenylation events in Jurkat T cells by 3'-rapid amplification of cDNA ends (3'-RACE) 66

DISCUSSION **67**

CHAPTER 5: CONFORMATIONAL DYNAMICS OF PSF **69**

INTRODUCTION **69**

RESULTS **70**
Protein Purification of PSF Δ265 70
Conformation of PSF following *in vitro* phosphorylation by GSK3 70
PSF/TRAP150-PID dynamics measured by HDX-MS 72

DISCUSSION **73**

CHAPTER 6: CONCLUSIONS & FUTURE DIRECTIONS **74**

CHAPTER 7: MATERIALS AND METHODS **79**

Protein Purification **79**

RNA-Binding region Identification (RBR-ID) **79**

Hydrogen-Deuterium Exchange (HX-MS) **80**

Electromobility Shift Assay **81**

***In vitro* phosphorylation by GSK3** **82**

Differential Scanning Fluorimetry **82**

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS) **82**

Crystal screens **83**

Conservation mapping **84**

APPENDICES **85**

Appendix A: Thermal Melt of PSF by circular dichroism 85

Appendix B: Microscale Thermophoresis & Fluorescence Polarization 86

BIBLIOGRAPHY **89**

LIST OF TABLES

Table 2.1 *Additive RBR-ID scores across overlapping peptides in mouse embryonic stem cells* 31
Table 2.2 *Additive RBR-ID scores across overlapping peptides in vitro* 31
Table 2.3 *Summary of top protected peptides in presence of ESS-RNA in HDX-MS.* 34
Table 2.4 *Residues dewetting at a biasing potential of 2.2*…….. 45
Table 4.1 *Gene specific primers for 3'RACE validation.* 66

LIST OF ILLUSTRATIONS

Figure 1.1 Common RNA-binding proteins and domain structure.. 3
Figure 1.2 Splicing & regulation of alternative splicing. 5
Figure 1.3 Alternative Polyadenylation.. 9
Figure 1.4 Structure of RNA-binding proteins. 11
Figure 1.5 Structure of FUS and RBM45 RRMs bound to nucleic acid. 14
Figure 1.6 Drosophila behavior human splicing (DBHS) family of proteins. 17
Figure 1.7 MFold structure prediction of RNAs known to bind PSF 23
Figure 2.1 Structure of PSF homodimer (4wii). 25
Figure 2.2 Structure and binding-motifs of RNA Recognition Motifs.. 26
Figure 2.3 Structural details of PSF's RRMs.. 26
Figure 2.4 Mfold prediction of exonic splicing silencer (ESS) RNA 71
Figure 2.5 RNA-Binding Region Identification (RBR-ID) workflow & components. 29
Figure 2.6 RBR-ID peptide plots. 30
Figure 2.7 PSF's RNA binding by Hydrogen Deuterium Exchange Mass Spectrometry. 33
Figure 2.8 Hydrogen Deuterium Exchange Mass Spectrometry 35,36
Figure 2.9 Comparative mapping of PSF's RNA-binding indicate expansive binding interface…..38
Figure 2.10 Comparative mapping of PSF's NOPS domain 38
Figure 2.11 Comparative mapping of PSF's RRM1 39
Figure 2.12 Weak RRM1/ESS-RNA interactions via Nuclear Magnetic Resonance…….. 41
Figure 2.13 Mapped data of RRM1/ESS-RNA interactions by NMR…….. 42
Figure 2.14 Dewetted atoms in φ-ensemble simulation…….. 44
FIgure 2.15 Electrostatic potential & conservation of PSF……..…… 46
Figure 2.16 B-factors mapped to PSF homodimer……... 46
Figure 3.1 Mutational analysis PSF construct. 50
Figure 3.2 Binding analysis of mutations by electromobility shift assay (EMSA). 52
Figure 3.3 Protein stability by thermal melt 53
Figure 3.4 Size Exclusion Chromatography with Multi-Angle Light Scattering (SECMALS) 54
Figure 3.5 Hydrogen-Deuterium exchange data of β2β3-loop mutant 55-57
Figure 3.6 Heat map of HDX-MS WT vs. β2β3-loop mutant 58
Figure 3.7 K413A mutant increases binding, R474A decreases binding 59
Figure 3.8 Binding analysis of mutations with vault RNA (vtRNAs) 60
Figure 4.1 Alternative polyadenylation events using DaPars upon PSF knockdown.. 65
Figure 4.2 Alternative Polyadenylation site usage in Jurat T cells. 67
Figure 5.1 Purification of PSF Δ265 construct 70
Figure 5.2 Phosphorylation-induced dynamics measured by HDX-MS 70
Figure 5.3 Preliminary ^{1}H-^{2}D exchange data mapped onto the homodimer of PSF's exRRMs. ... 73

CHAPTER 1: Introduction

Cells in our body contain genetic information (genes) composed of **d**eoxy**r**ibonucleic **a**cid (DNA). DNA encodes recipes for proteins and RNA working together with the environment to create organisms, which develop, grow, and reproduce. Genes that are expressed from this DNA recipe dictate developmental timing, establish cell type differences, and change in response to an organism's environment to adapt. DNA expression begins with transcription of DNA to **r**ibo**n**ucleic **a**cid (RNA), creating a copy of the genetic message. Next, RNA is processed and after processing can be translated into protein. Proteins go on to fulfill many of the jobs within the cell. This concept of genetic expression is called the central dogma of biology and was first described over half a century ago (Crick, 1958). Since this initial description and over 70-years of advances in technology, our knowledge of gene expression and RNA processing has greatly expanded. Although much has been learned, there remains much to uncover regarding the mechanisms of interactions between protein and RNA that are integral to everyday cellular life.

WHAT IS AN RNA-BINDING PROTEIN?

RNA-binding proteins (RBPs) facilitate various processes in RNA biology. Once transcribed, RNA rapidly associates with RBPs that regulate its biogenesis, processing, localization, regulation, and metabolism (Geuens et al., 2016). Each step in the expression of DNA to protein requires RBPs to function. RNA polymerase II (RNA pol II) transcribes RNA from the DNA template. The spliceosome processes RNA into messenger RNA (mRNA), which can be translated to protein by the ribosome. RNA pol II, the spliceosome, and the ribosome are all cellular machines built from and regulated by RBPs. RNA-binding

proteins are integral to this machinery and often work in a combinatorial fashion to control cellular functions. While proper functioning of RNA-binding proteins establishes cell specificity, dysregulation of RBPs and RBP-networks are implicated in many maladies including neurological disease, muscular disorders, cancers, and other diseases (Conlon & Manley, 2017; Geuens et al., 2016). Studying variation in RNA-binding and regulation of RNA-binding proteins provides mechanistic insights into how cells function normally as well as in disease states. These insights help develop improved treatments of illness and disease.

Some examples of RBP families are serine-arginine rich (SR) proteins, heterogenous nuclear ribonucleoproteins (hnRNPs), as well as hnRNP-like proteins. SR proteins are characterized by containing both an RNA Recognition Motif (RRM) and an arginine/serine-rich sequence called an RS domain. The hnRNP family has many roles in RNA biogenesis, regulation, metabolism, and localization. The hnRNP family is diverse both in size and in domain structure and is named alphabetically from A to U. Members of this family contain at least one of the following four RBDs; RRMs being the most common, quasi-RNA Recognition Motifs (qRRMs), K-Homology domains (KH) and Arginine-Glycine-Glycine repeats (RGG). Finally, hnRNP-like proteins share similar characteristics as hnRNPs, however were not found in the initial discovery of this family (Dreyfuss et al., 1993). Like the hnRNP family, hnRNP-like proteins contain one or multiple RBDs, and participate in various parts of nucleic acid metabolism. Many of the RNA-binding proteins also contain Nuclear Localization Signals (NLS) for shuttling between the nucleus and cytoplasm, which is a key point of regulation.

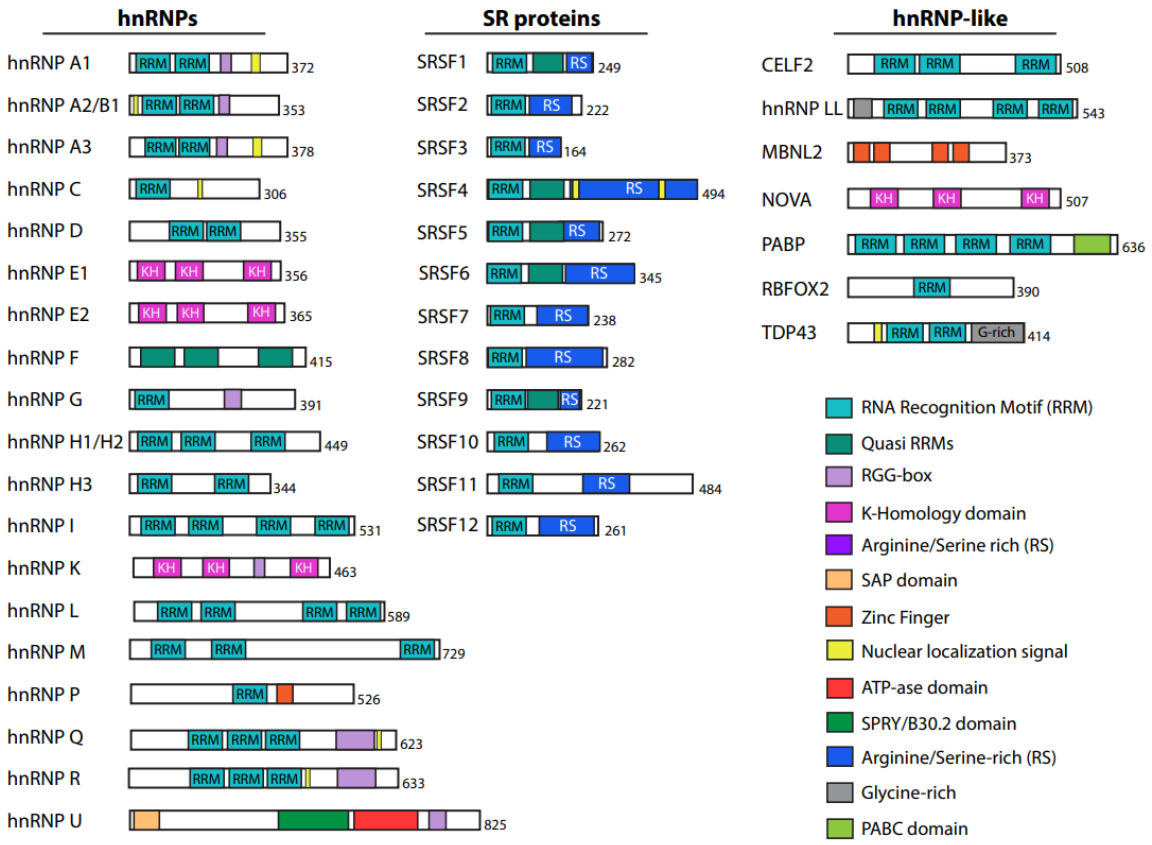


Figure 1.1 Common RNA-binding proteins and domain structure. The hnRNPs are a large family of RNA-binding proteins, which share general features but have distinct functions in cellular nucleic acid metabolism. The SR family of proteins all contain one RRM, a serine/arginine-rich region, and some have a quasi-RRM. hnRNP-like proteins are 45-60% similar to hnRNPs and have multiple roles in nucleic processing and metabolism.

PROCESSING OF MESSENGER RNA

During transcription of RNA, many steps of RNA processing occur simultaneously, or cotranscriptionally (Hirose & Manley, 1998, 2000; McNeil et al., 1998; Schroeder et al., 2000). Splicing is one cotranscriptional processing event, when the intronic sequences are removed from the precursor-mRNA (pre-mRNA) transcript followed by linkage (splicing) of exonic sequences. Genes have exons and introns. Exons become part of

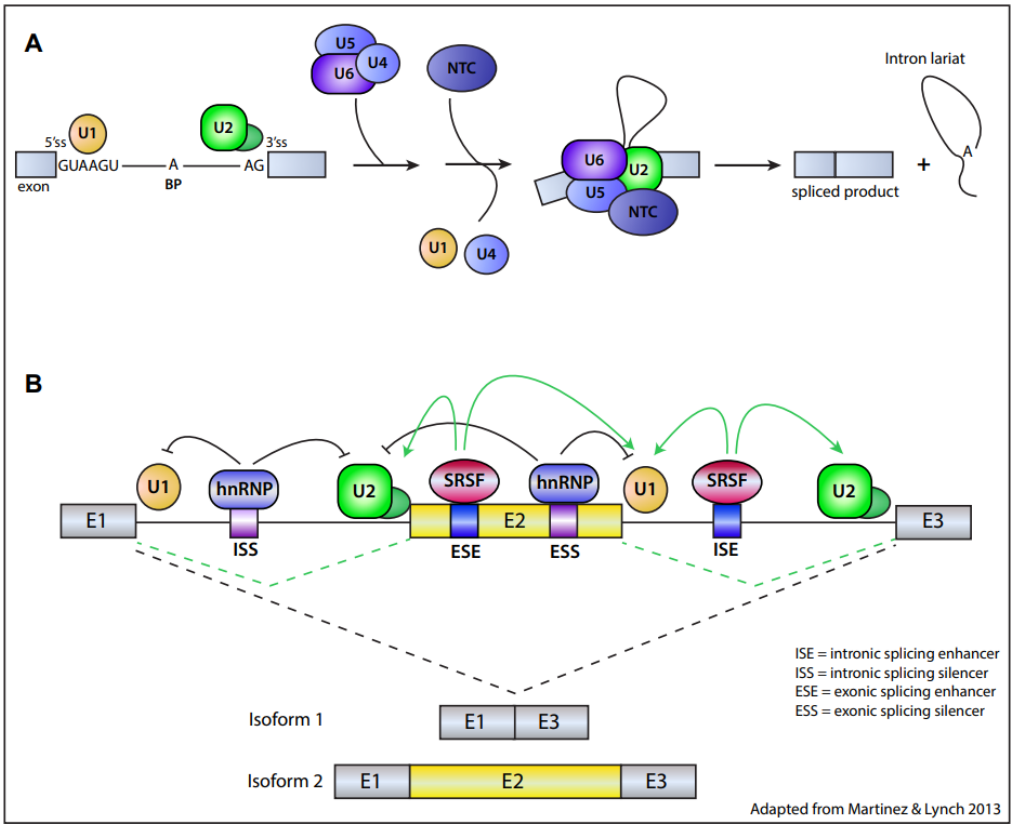
mature messenger RNA (mRNA) and can be translated into protein. Introns are intervening sequences removed during the RNA processing step called splicing, detailed further shortly. Another cotranscriptional event is the addition of a 7' methylguanosine cap to the 5' end pre-mRNA by phosphate linkage. The 5' cap protects RNA from degradation by ribonucleases and is recognized by proteins involved in assembling the ribosome onto mature mRNA to initiate translation in the cytoplasm. Lastly, 3' end processing occurs cotranscriptionally as well. A consensus sequence called the polyadenylation signal (PAS) within the mRNA transcript designates cleavage. Following cleavage, an enzyme called poly (A) polymerase (PAP) adds approximately 200 adenosines to create the 3' poly-A tail. Like the 5' cap, the poly-A tail protects the mRNA from degradation. Untranslated regions (UTRs) in the 5' and 3' ends of mRNA are scattered with binding sites for proteins involved in translation, stability, metabolism, and localization. The process of cleavage and polyadenylation is covered in detail later.

Splicing

Nearly all human transcripts undergo splicing during maturation from pre-mRNA to mRNA (Pan et al., 2008). A single pre-mRNA can be spliced to make multiple distinct protein isoforms. Splicing is an important part of expanding the complexity of the human proteome, as well as a key point of gene regulation (Blencowe, 2017). As mentioned, exons are portions of genes encoding protein and are interspersed with intervening sequences called introns. Splicing is the joining of exons and removal of intronic gene sequences. Splicing is a multi-step process catalyzing coordinated transesterification reactions to promote release of the 5' exon, and ligation to the 3' exon, which is catalyzed by a large molecular machine called the spliceosome (Fig. 1.2A) (Matera & Wang, 2014). The spliceosome is a 3-mDa complex made up of more than 100 proteins, including

five snRNPs (small nuclear ribonucleoproteins); U1, U2, U4 and U5. Progression through splicing requires assembly and disassembly of this machine along the way. Recognition of the 5' splice site by U1 and binding of U2 snRNP auxiliary factory (U2AF) to the 3' splice site characterizes "E (early) complex" formation. "A complex" involves the first ATP-dependent step, binding of the U2 snRNP to the branch point sequence at the 3' splice site. Next recruitment of the tri-snRNP complex, composed of U4-U6-U5, characterizes "B complex". Finally, through extensive

Figure 1.2 Splicing & regulation of alternative splicing. A) Splicing is a two-step reaction initiated by a nucleophilic attack by the branch point (BP) adenosine on the 5' splice site (ss), to



form an intron lariat intermediate. Next, the 5'ss attacks the 3' ss to splice exons together and remove the lariat. **B)** Alternative splicing is intricately regulated by RBPs that bind sequence elements in the transcript recruiting or blocking spliceosome assembly.

remodeling of both RNA and protein components of “B complex”, the active enzymatic splicesome, or “C complex,” can catalyze the splicing reaction.

The dynamic nature of the splicing reaction, whereby assembly and disassembly of an entire complex is required in a stepwise fashion, provides multiple opportunities for regulation. Alternative Splicing (AS) is a regulatory process generating distinct protein isoforms that can drastically impact cell physiology while adding complexity to the proteome (Blencowe, 2017) (Fig. 1.2B). Alternative splicing occurs in a cell type specific manner, in disease states such as cancers, as well as in responses to changes in the cellular environment, seen in the activation of T-cells.

Mis-splicing of RNA occurs in several diseases (Scotti & Swanson, 2016). For example, a single splice site mutation in dystrophin results in loss of function causing Duchenne muscular dystrophy, a devastating congenital disease characterized by muscle loss, scoliosis, intellectual disability, and short life expectancy, (Fletcher et al., 2013; Takeshima et al., 2010). Another disease caused by a single mis-splicing event is spinal muscular atrophy one of the major genetic disorders resulting in infant mortality (Singh et al., 2018). Survival of motor neuron (SMN) protein is critical for spliceosome assembly and is encoded by the SMN1 gene. A single point mutation causes mis-splicing resulting in the skipping of exon 7, creating a truncated isoform of SMN (Burghes & Beattie, 2009; Lorson et al., 1999; Monani et al., 1999). Understanding this mechanism enabled the development of an antisense oligonucleotide drug Spinraza, which targets splicing to force inclusion of exon 7 and results in statistically significant improvements in motor function and survival (Hua et al., 2007; Neil & Bisaccia, 2019). Additionally, mutations affecting alternative splicing of cell surface antigens occurs in some cancers, which makes some cancer cells resistant to immunotherapies (Black et al., 2018; Sotillo et al., 2015) These examples are relatively straight-forward, but many other cases are mechanistically more

complex or context-dependent, with changing outcomes based on a network of RNA-binding proteins rather than the direct action of one RBP or one splicing event (Ajith et al., 2016; Fu & Ares, 2014; Martinez et al., 2015; Melton et al., 2007). Much remains to be discovered about splicing in disease, and better understanding of mechanism can help inform treatment.

Understanding how cells utilize alternative splicing to respond to environmental cues provides insight into how dysregulation of splicing can affect/cause disease. Our lab has used the well-developed model of differential splicing of the CD45 pre-mRNA in resting versus stimulated T-cells to study the mechanisms of signal induced alternative splicing. The CD45 gene encodes a protein tyrosine phosphatase, a transmembrane protein on the cell surface that transduces extracellular signals into an appropriate molecular response. Several isoforms of this gene are expressed, to maintain balance in this pathway. A smaller isoform excluding exon 4 is upregulated following T-cell stimulation. Isoforms excluding exon 4 lack a domain required for integration in the cellular membrane, thus becomes cytoplasmic, and acts to dampen the immune response. Silent mutations disrupting the splicing of exon 4 increase propensity for autoimmune disease (Lynch & Weiss, 2001). Previous work identified major factors regulating CD45 pre-mRNA splicing in the context of T-cell activation, PSF, hnRNP L, hnRNP L-like, and hnRNP A1. PSF is the topic of this dissertation and described in detail later.

Cleavage and Polyadenylation

During RNA processing the pre-mRNA transcript is coated with RBPs from the 5' through 3' UTR, effecting RNA splicing, termination, and polyadenylation. Polyadenylation involves adding a uniform stretch of adenosines onto the 3’ end of protein-encoding RNA

transcripts. In metazoans, apart from replication-dependent histone transcripts, all mRNA maturation requires a large protein complex to endonucleolytically cleave the 3’ ends of nascent RNA and subsequently polyadenylate. The transcript contains sequences called polyadenylation signals (PAS). PAS are recognized by core machinery consisting of 4 multi-subunit protein complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), and mammalian cleavage factors I and II (CFIm and CFIIm) (Mandel et al., 2008). These signals are consensus sequences within the RNA transcript: the canonical hexamer A[A/U]UAAA is found 21 nucleotides upstream of cleavage and a U- or G/U- rich sequence is 10-30 nucleotides downstream of the cleavage site (Gruber et al., 2016). The downstream element (DSE) doesn’t have a consensus sequence, but in subsets of genes with degenerate hexamer sequences mutations in the DSE had a greater effect on 3’ end processing (Neve et al., 2017). The combinatorial nature of containing multiple sequence elements in 3’ end processing allows for fine-tuning and regulation.

Like the spliceosome, there are both core proteins of the poly(A) machinery, as well as regulatory proteins that bind machinery or cis elements in the 3’ UTRs of transcripts to affect PAS usage. Proteins binding 3' UTR of transcripts can regulate alternative polyadenylation (APA). Alternative 3' end of transcripts generated can control localization, translation, and stability of transcripts. Alternative polyadenylation occurs in over half of human transcripts, is tissue specific, found altered in disease context, and regulated in response to environmental cues (Tian & Manley, 2017). During T-cell stimulation APA is controlled through, CELF2 (Chatrikhi et al., 2019) and PSF (unpublished data), and likely other RBPs. Alternative APA site usage in the 3’ UTR creates transcripts with different 3’ ends to regulate stability and translation of transcripts (Fig. 1.3). For example, alternative 3’UTRs can include or exclude microRNA sites important for regulation (Sandberg et al.,

2008). Understanding how RBPs bind their target RNA sequence gives us mechanistic understanding of how cells control gene output to create proper cell function.

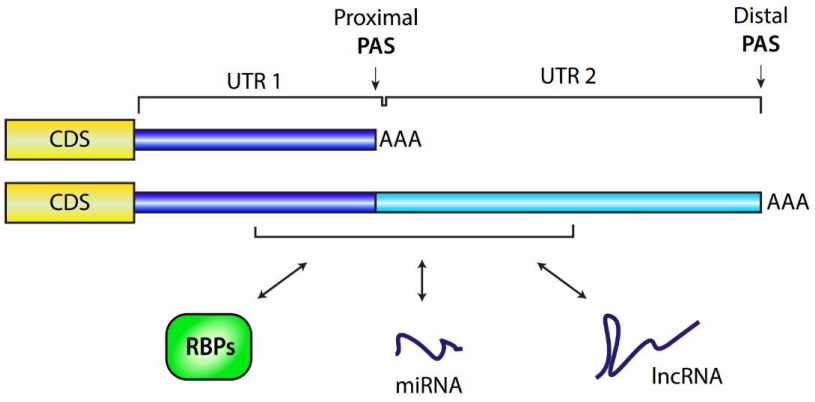


Figure 1.3 Alternative Polyadenylation. Two distinct mRNA isoforms with different 3' untranslated regions (3' UTRs) are created from utilizing different polyadenylation signals (PAS). Alternate 3' ends allow for differential regulation and metabolism of the mRNA isoforms by factors that bind the 3' UTR like RNA-binding proteins (RBPs), microRNAs (miRNAs), and long noncoding RNA (lncRNAs).

STRUCTURE OF RNA-BINDING PROTEINS

RNA-binding proteins accomplish diverse cellular tasks by binding both single-stranded and double-stranded RNA through one or multiple RNA-binding domains. Combination of domains and the modularity of domains, allows for dynamic regulation of binding a wide array of nucleic acid sequence and structural motifs. Several well-characterized RBDs have been studied for the last 30 years, namely zinc fingers (ZnF), K homology (KH) domains, double-stranded RNA-binding domain (dsRBD), RGG repeats, and the above-mentioned RRMs.

Zinc fingers were first identified as DNA-binding proteins in transcription factor IIIA (TFIIIA)(Brown et al., 1985; Diakun et al., 1986). Since then, ZnF containing proteins have

been found to complex with RNA, make protein-protein contacts and membranes interactions (Laity et al., 2001). Each “finger” is composed of an α-helix and β-sheet held together by the coordination of a zinc ion. Coordination of zinc is achieved using three cysteines and one histidine. The zinc finger domain is often found in tandem copies withing a single protein. For example, muscleblind-like 1 (MBNL1) contains four ZnF, which all have nearly identical folds. MBNL1 targets pre-mRNA containing the sequence YGCU(U/G)Y, and the two tandem ZnF domains (ZnF1/2 and ZnF3/4) were crystallized in the apo-state as well as ZnF3/4 bound to the sequence CGCUGU (Teplova & Patel, 2008). Both ZnF3/4 retain their conformation upon binding RNA (Teplova & Patel, 2008) (Fig. 1.4).

KH domains were first identified in heterogenous nuclear ribonucleoprotein K (hnRNP K) (Matunis et al., 1992). KH domains are ~70 amino acids comprised of three α -helices packed against a three-stranded antiparallel β-sheet surface. This family is divided into Type I and Type II KH domains based on N- and C-terminal extensions (Geuens et al., 2016). Initially, it was thought KH domain recognize and bind poly(C)-tracts, but new information uncovered that many proteins containing KH domains have affinity for A/C-rich in addition to G-rich sequences (Nicastro et al., 2015). For example, the KH3-domain of Nova-2 was crystallized bound to the standard A-form stem-loop RNA containing the sequence 5’-UCAY-3’ (Lewis et al., 2000) (Fig. 1.4).

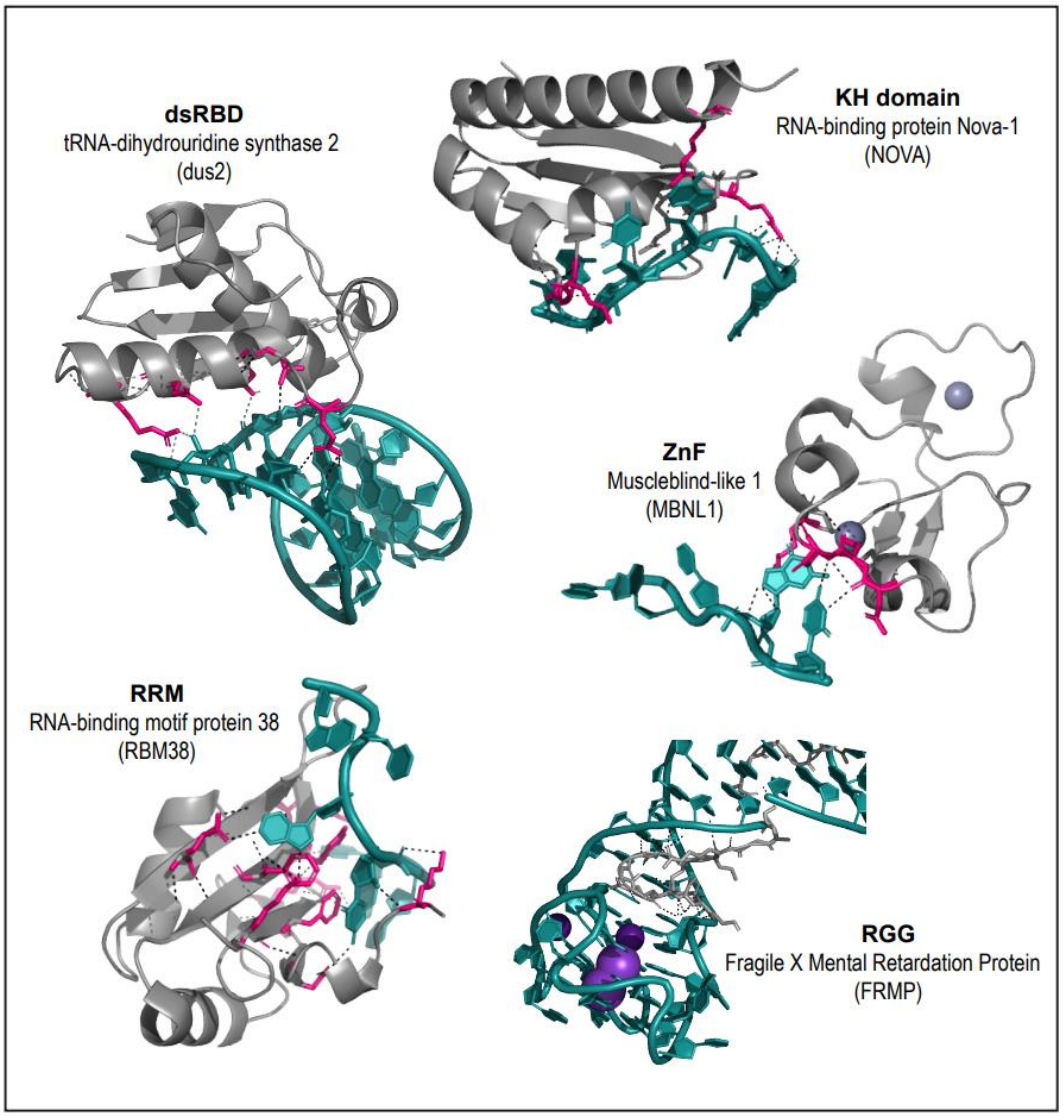


Figure 1.4 Structure of RNA-binding proteins. The third KH domain (KH3) of Nova-2 bound to a stem loop RNA uses an aliphatic αβ recognition surface that mediates binding to 5-UCAY-3. Muscleblind-like 1 contains four ZnF domain but shown here are ZnF3 bound to the sequence CGCUGU. The FRMP RGG motif fits into the duplex–quadruplex junction RNA through shape complementarity, cation–π interactions, and hydrogen bonding. Purple spheres represent K+ ions which stabilize the G-quadruplex structure. The RRM domain of RBM38 recognizes G(U/C/A)GUG through hydrogen bonds between the base atoms of RNA and main-chain or side-chain atoms, as well as through base stacking interactions with bases of RNA and two phenylanines along the β-sheet surface. Recognition of dsRNA by the dsRBD is achieved via three conserved regions: helix-1, helix-2, and the C-terminal part of the β1-β2 loop. PDB codes: 1EC6, 3D2S.5DEA,6JVX, and 5OC6 respectively(Bou-Nader et al., 2019; Lewis et al., 2000; Qian et al., 2020; Teplova & Patel, 2008; Vasilyev et al., 2015). Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

The double stranded RBD (dsRBD) is ~70 amino acid long with a versatile domain that folds into a compact αβββα structure and binds a variety of RNA structures. An example is the protein Staufen, which uses the dsRBD to interact with dsRNA through conserved loops, which insert into both the minor and major grooves, the intervening major groove of the RNA face, while making contacts with the phosphate backbone (Ryter & Schultz, 1998). The dsRBD has also been shown to bind RNA with complex structure, as in the case of dihydrouridine synthase 2 (dus2) which recognizes a tRNA ligand (Bou-Nader et al., 2019) (Fig. 1.4).

RGG domains are regions of arginine-glycine-rich (RGG) repeats, first recognized as a target for protein arginine methyltransferases (PRMTs) (Thandapani et al., 2013). The RGG domain is present in at least 31 different proteins (Masuzawa & Oyoshi, 2020), and makes many important interactions with G-quadruplex DNA and RNA. For example, the RGG domain of TLS/FUS is important for folding G-quadruplex telomere DNA (Takahama et al., 2013). Fragile X Mental Retardation Protein (FMRP) uses an arginine-glycine-rich (RGG) motif to interact with guanine (G)-quadruplex RNA, relying on interactions from a type I β-turn secondary element (Vasilyev et al., 2015). The RGG motif fits into a sharp turn at the duplex–quadruplex junction RNA through shape complementarity, cation–π interactions, and hydrogen bonding (Fig. 1.4).

Most importantly, RRM containing proteins comprise roughly 2% of human protein-coding, and often RRMs are found in multiple copies (44%, two-six) within a single protein (Bateman et al., 2002; Maris et al., 2005). RRMs are characterized by a ~90 amino acid long domain, which folds into a well conserved β1-α1-β2-β3-α2-β4 topology to form a four-stranded anti-parallel β-sheet packed with two α-helices. While this motif follows a canonical fold pattern, it is also structurally flexible and dynamic, which enables

recognition of a diverse range of RNA sequence and structure. Canonical RRMs contain conserved aromatic residues on the β-sheet surface that base stack with RNA. Base-stacking interactions create the main RNA-binding surface, with a 2-8-mer sequence recognition. For example, RBM38 uses two phenylalanines on its β-sheet surface to aid in binding to 5'UGUGUG-3' (Fig.1.4). RRMs are structurally and biochemically diverse with various modes of binding recognition (Cléry et al., 2008; Lunde et al., 2007). Some RRMs use loops to bind target RNA or contain N- or C- terminal extension of with a β-strand or an α-helix that can modulate binding (Maris et al., 2020). Only a few have been shown to bind stem-loop RNA, and the FUS-RRM interaction is an interesting example of this (Fig 1.6). As is typical for an RRM, the FUS RRM uses binding pockets on the β-sheet surface, but the orientation of the RNA bases is unique, forming a tight turn with mostly non-sequence-specific interactions, as it does not contain the canonical aromatic residues to base stack with RNA (Loughlin et al., 2019). Intriguingly, the positively charged loops α1/β2 and β2/β3 facilitate folding of their partner RNA (Basu et al., 2021). RRMs also bind ssDNA. RBM45 has two tandem RRMs that recognizes GAC sequence in either ssRNA or ssDNA as seen in the crystallization with 5'-CGACGGGACGC-3' ssDNA (Fig. 1.6). This interaction utilizes aromatic residues in the β-sheet as well as the interdomain linker between RRMs (X. Chen et al., 2021). Additionally, some RRMs mediate protein-protein interaction. For example, PTB contains four RRMs, and RRM3/RRm4 tightly associate through their α -helices (Vitali et al., 2006). Additionally, the U2AF homology motif (UHM) family of proteins use a negatively charged extended α -helix and a conserved R-X-F motif in the α2/β4 loop as a protein-protein interaction surface (Cléry et al., 2008). Studying the diverse array of RNA binding modes is important to better understand the mechanisms of RNA-protein interactions and RNA-binding proteins' contribution to disease.

Advances in RNA-protein interactome capturing and in structural studies of large ribonucleoprotein (RNP) complexes have uncovered complex RNA-protein interactions which do not require canonical RBDs (Castello et al., 2016; He et al., 2016; Wheeler et al., 2018). From these advances new RBPs were identified, expanding our understanding of how proteins recognize RNA while adding complexity to the modes and modularity of RNA-binding. Although many RNA-binding domains and protein-RNA interactions have been characterized, there is still much to be discovered given the complexity and breadth of interactions of RBPs.

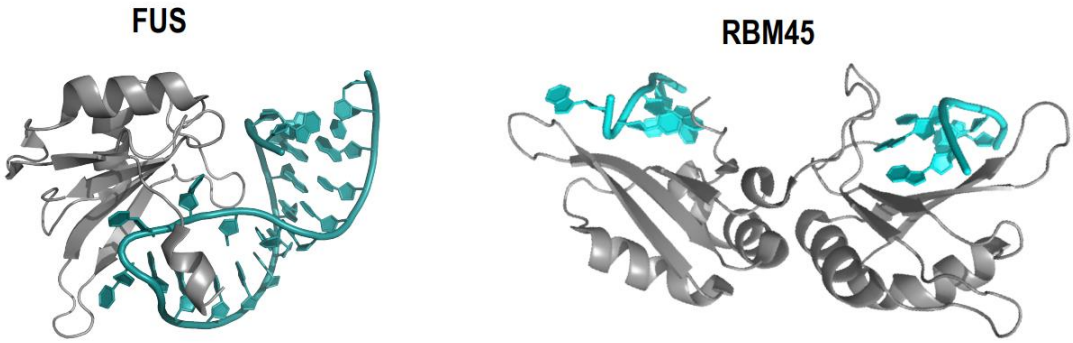


Figure 1.5 Structure of FUS and RBM45 RRMs bound to nucleic acid. FUS RRMs bind stem-loop RNA structure using two loop regions along with nonspecific interactions on the B-sheet surface. RBM45 binds ssDNA and RNA across its two tandem RRMs. RBM45 uses the typical aromatic residues on the β-sheet as well as interactions with the RRM linker. PDB codes: 6GBM and 7CZV (RBM45). Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

REGULATION OF RNA-BINDING PROTEINS

Regulation of nucleic acid binding proteins is complex and can involve altering protein levels, steric hinderance by mutually exclusive binding interactions (Yarosh et al., 2015), conformational changes to favor "open" and "closed" states (Duss et al., 2014; Kang et al., 2020), cooperative binding of multiple copies of proteins, post-translational

modifications to alter binding preference, as well as a combination of any of these mechanisms. Autoregulation of RBPs is widespread, as RBPs control their own synthesis, splicing, and translation (Müller-Mcnicoll et al., 2019).

Post-translational Modification

SR proteins are part of a large family of RNA-binding proteins that share a common motif, an arginine/serine-rich sequence called an RS domain, as well as an RRM domain. SR proteins SRSF1 and SRFS2 were discovered due to their role promoting U1 and U2 snRNP binding to the 5' and 3' splice sites respectively. Phosphorylation of SR proteins was discovered in 1990 (MB et al., 1990) and soon after the SRPK family and CLK family of kinases were shown to phosphorylate and control cellular localization of SR proteins (Colwill et al., 1996; Ghosh & Adams, 2011; Gui et al., 1994). For example phosphorylation mediated by SRPK1 (serine/arginine protein kinase-1) facilitates nuclear import of SR proteins (Kataoka et al., 1999; Yun & Fu, 2000).The phosphorylation state of SR proteins controls both localization and function (Zhou & Fu, 2013) Phosphorylation has also been show to facilitate a conformational switch in the RS domain of SRSF1, from a fully disordered state to a rigid arch-like structure (Xiang et al., 2013).

Protein Phosphatase 1 (PP1) dephosphorylates several splicing factors, including PSF, Non-POU domain-containing octamer-binding protein (NONO) and transformer2-beta1 (tra2-β1) (Liu et al., 2011; Novoyatleva et al., 2008). PP1 binds via conserved RVDF sequence within the RNA Recognition Motif. Dephosphorylation of tra2-β1 by PP1 alters its alternative RNA splicing activity, namely increased PP1 expression promoted the exclusion of tra2-β1-dependent exons (Novoyatleva et al., 2008). PSF and NONO dephosphorylated by PP1 also have altered splicing activity (Liu et al., 2011).

Many RBPs are autoregulated to play a role in their own homeostatic transcript or protein level. For example, increased expression of hnRNP L promotes the inclusion of a poison exon, and this splicing inclusion is mediated by hnRNP L binding to an upstream intronic splicing enhancer, ultimately leading to transcript degradation (Rossbach et al., 2009). CELF2 binds its own 3' UTR to regulate signal-induced alternative polyadenylation as well as intron retention (Chatrikhi et al., 2019). The splicing factor U2AF utilizes an intrinsically disordered region connecting it's two RRMs to reduce nonproductive binding to weak pyrimidine-tract RNA, proofreading to more accurately define 3' splice sites (Kang et al., 2020). Much remains to be learned about regulation of an RBPs and regulation of RBPs ability to bind nucleic acid.

POLYPYRIMIDINE TRACT-BINDING PROTEIN-ASSOCIATED SPLICING FACTOR

Polypyrimidine tract-binding protein-associated **S**plicing **F**actor, or PSF, is an essential and predominately nuclear protein with multiple roles in nucleic acid processing, binding to both RNA and DNA (Shav-Tal & Zipori, 2002). PSF is a member of the conserved DBHS (Drosophila behavior human splicing) family of proteins, and functions in cells as either a homodimer or a heterodimer with one of two additional protein family members, p54nrb and PSPC1. The DBHS family of proteins are multi-functional and highly conserved throughout vertebrates. All the DBHS proteins are found in paraspeckles, but only PSF and NONO are essential for paraspeckle formation. Paraspeckles are subnuclear structures localized to chromatin or DNA damage foci (Bond & Fox, 2009). Paraspeckles regulate certain genes by nuclear retention of RNA, and are dense structures built upon a long noncoding RNA NEAT-1 that acts as a protein scaffold to help formation.

PSF is a 76 kDa protein with N- and C-terminal unstructured regions. The middle portion contains the DBHS core consisting of two tandem RRMs, a homology domain known as NOPS (NONO/ParaSpeckle domain), an extended coiled-coil, and a nuclear localization signal (Fig. 1.6). The RRMs-coiled coil aid in dimerization, whereas the coiled-coil mediates oligomerization. Many different nuclear complexes containing PSF are involved in response to splicing regulation, DNA repair, translation, response to viral infection and regulating alternative polyadenylation. Mechanisms governing PSF's RNA-binding remain unclear.

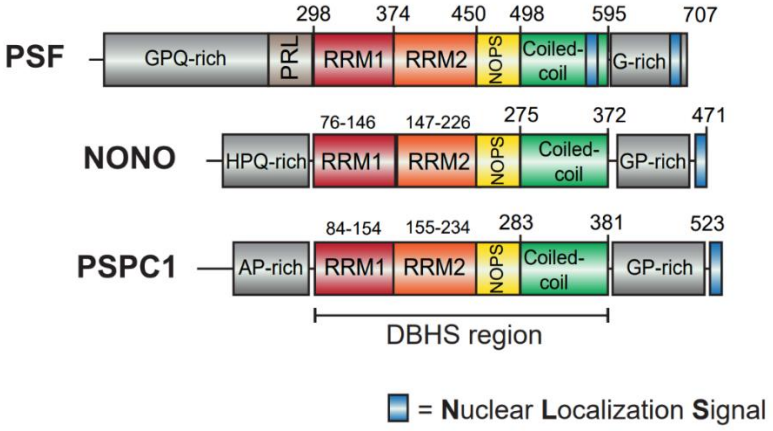


Figure 1.6 Drosophila behavior human splicing (DBHS) family of proteins. The DBHS family share a core region consisting of two tandem RRMs, a homology domain NOPS, and a unique extended coiled-coil region.

PSF in nucleic acid processing

PSF was first characterized by the Patton lab (Patton et al., 1993). Patton *et al.* discovered a protein complexed with PTB and necessary for pre-mRNA splicing. Immunodepletion of PSF blocked early spliceosome assembly. Other accounts of PSF regulating splicing have since been described. For example, PSF suppresses Tau exon

10 by stabilizing a stem-loop structure downstream of the exon. In the adult brain, exon 10 is regulated to control the ratio of isoforms Tau3R and Tau4R. Disrupting this balance is associated with tauopathies such as fronto-temporal lobar degeneration, Alzheimer's disease, and progressive supranuclear palsy (P. Ray et al., 2011). As mentioned, PSF also controls the balance of CD45 isoforms in T-cells to regulate the immune response. Additionally, PSF binds a purine-rich enhancer in exon 7 of SMN1/2 to promote inclusion (Cho et al., 2014).

PSF is also involved in 3' end processing. The snRNP-free U1A (SF-A) complex contains both PSF and p54nrb, a complex speculated to be a special adaptor between splicing and polyadenylation (Liang & Lutz, RNA 2006). Additionally, tethering of PSF to the 3' UTR of COX-2 (Cycloxygenase-2) containing a weak polyadenylation signal stimulated usage of this site (Lutz). PSF, NONO, and Matrin-3 complex to mediate nuclear retention of A-to-I edited RNA (Zhang & Carmichael, 2001). Paraspeckle formation is necessary for nuclear retention of edited RNA to occur in human embryonic stem cells (hESCs) and Hela cells, further highlighting PSF's role in nuclear retention of RNA (L. L. Chen & Carmichael, 2009). In our system of Jurkat T cells knockdown of PSF causes widespread alternative polyadenylation (unpublished data).

PSF also binds DNA. PSF aids in DNA repair by promoting DNA strand invasion during double stranded break repair (Akhmedov & Lopez, 2000), and found to stimulate duplex capture in homologous paring during meiosis (Morozumi et al., 2012). By interacting directly with Rad51D, a component of the DNA repair pathway, PSF mediates homology directed DNA repair and DNA damage response to various DNA damaging agents (Rajesh C., et. Al, NAR 2010). Phosphorylation also alters PSF's RNA- and DNA-binding ability and may be the cellular trigger responsible for directing PSF to DNA repair (Akhmedov & Lopez, 2000).

PSF also functions as a positive and negative regulator of transcription. PSF and p54nrb/NONO bind to the C-terminal domain of RNA polymerase II and recruit other splicing and polyadenylation factors (Emili et al., 2002). As a negative regulator, PSF recruits histone deacetylases (HDACs) to genes through the HDAC-associated protein SIN3A (Duong et al., 2011). PSF represses STAT6 mediated transcription of IgE by binding phosphorylated STAT6 and recruiting HDAC1 to the promoter (Dong et al., 2010).

PSF is predominantly nuclear, but cytoplasmic roles have been reported, namely in translation. In eukaryotic cells, translation of RNA requires cap binding complexes to initiate. However, viral RNAs acquired ways around this to hijack host cell machinery for production of viral proteins. One way is through sequences called internal ribosome entry sites (IRES). Some viral transcripts have IRES located in the 5' UTR that allow for cap-independent translation of viral RNA. During viral infection by the Coxsackievirus B3, PSF is translocated to the cytoplasm to act as an important IRES trans acting factor, positively mediating viral translation (Dave et al., 2017). PSF also interacts with the 5' UTR of cellular mRNAs encoding p53 and c-myc, as both transcripts have IRES (Cobbold et al., 2008; Sharathchandra et al., 2012).

Regulation of PSF

PSF plays many roles in cellular complexes, yet the regulation of PSF -- what determines when PSF is in one complex versus another remains poorly understood. PSF is post-translationally modified on several residues with differing effects to function. One way PSF is regulated is through phosphorylation. Phosphorylation by SRPK1 and DSK1 on serine 8 of PSF enhances binding to DNA, while inhibiting binding to the polypyrimidine tract of pre-mRNA in introns (Huang, et al., 2007). Additional studies show phosphorylation of PSF by protein kinase C stimulates DNA binding and D-loop formation

to promote homologous pairing in DNA repair while inhibiting RNA binding (Akhmedov & Lopez, 2000). N-terminal hyperphosphorylation of PSF occurs during apoptosis affects conformation, possibly explaining the shift in affinity for DNA/RNA (Shav-Tal et al., 2001). PSF and DBHS family member, p54nrb, are dephosphorylated by protein phosphatase 1, to alleviate known transcriptional corepressor activity of PSF/p54nrb on the androgen receptor. Dephosphorylation of PSF has been reported to regulate RNA splicing activity as well (Liu et al., 2011).

C-terminal phosphorylation of PSF increases binding to an accessory protein called TRAP150, which occludes RNA-binding to affect splicing of key factors in T-cells (Heyd & Lynch, 2011; Yarosh et al., 2015). In naïve T-cells, PSF is constitutively phosphorylated on a C-terminal threonine (T687), by Glycogen Synthase Kinase 3 (GSK3). Phosphorylation promotes binding of PSF to the 150-kDa component of Thyroid hormone Receptor-Associated Protein complex (TRAP150), and PSF/TRAP150 interactions are known to occlude PSF/RNA binding (Yarosh et al., 2015). Upon T-cell activation GSK3 is inactivated and newly translated PSF remains unphosphorylated. This PSF is not bound by TRAP150, and thus can bind the exonic splicing silencer motif (ESS) in CD45 RNA to alter splicing. What underlies the mutually exclusive binding of ESS and TRAP150 remains unclear. Limited proteolysis of T687 phosphorylated vs unphosphorylated PSF indicates a possible conformational change between the phospho-states, possibly explaining this selectivity (Yarosh, unpublished data). In HeLa cells, expression of RFP-tagged PSF mutants changed localization upon infection with CVB3: the phosphor-dead mutant T687A relocalized to the cytoplasm upon CVB3 infection, but the phosphomimetic T687D mutant remained in the nucleus, indicating phosphorylation of T687 may keep PSF

in the nucleus (Dave et al., 2017). In contrast, phosphorylation of PSF on tyrosine-293 by ALK (anaplastic lymphoma kinase) caused localization of PSF to the cytoplasm.

Finally, UV-crosslinking experiments in mammalian cells have found arginine methylation and arginine citrullination of PSF's N-terminal RGG box of affects mRNA binding (Snijders et al., 2015). Snijders et al. also discovered citrullination blocks in vitro methylation of PSF's RGG box, and that PRMT-1 (Protein arginine N-methyltransferase-1) catalyzes the addition of methyl groups onto the RGG Box. Methylation of arginine is important for regulating transcription, RNA splicing, and import/export of RNA/protein from the nucleus (Hwang et al., 2021). It's possible PTMs could play a role in regulating PSF's many functions.

CURRENT INFORMATION ON PSF'S RNA-BINDING MODE

Attempts at defining PSF's binding motif remains elusive, differing between techniques. For example, the earliest binding motif analysis of PSF was carried out by the Patton lab using an iterative selection technique called system evolution of ligands by exponential enrichment (SELEX). SELEX experiments yielded the binding consensus sequence of UGGAGGGAAC (Peng et al., 2002). Another *in vitro* method called RNAcompete derived a GUAGUGU motif (D. Ray et al., 2013). As part of the Encyclopedia of DNA Elements (ENCODE) project phase III, a large scale RBP-binding analysis resulted in multiple binding motifs with little overlap (Van Nostrand et al., 2020). *In vitro* binding experiments (RNA Bind-N-Seq) determined a 5-mer motif of UGUAA, whereas *in vivo* experiments yielded a CGCCG motif via enhanced cross-linking immunoprecipitation (Van Nostrand et al., 2020). PSF's proclivity for binding structured RNA, the length

dependency of RNA-binding, and the inability to define a consensus binding sequence indicate a binding modality dependent on RNA structure rather than sequence.

PSF binds structured RNA. For example, PSF suppresses Tau exon 10 by binding to and stabilizing a stem-loop structure downstream of the exon (P. Ray et al., 2011). Additionally, PSF binds a conserved stem in U5 snRNA, as well as interacts with a purine-rich enhancer sequence in exon 7 of SMN1/2 to promote exon 7 inclusion (Cho et al., 2014; Peng et al., 2002). In humans PSF also binds three of four vault RNAs (vtRNAs) which share ~84% sequence identity, range in size from 86-141 nucleotides in length, and contain conserved 5' and 3' stem-loops (J. Chen et al., 2018; van Zon et al., 2003). PSF binds a cloverleaf structure in the 5'UTR of the coxsackievirus B3 (CVB3) that contains an internal ribosome entry site (IRES) to enable efficient translation of viral RNA (Dave et al., 2017) (Fig. 1.7). The PSF/NONO heterodimer binds to primary microRNAs (pri-mRNAs) to aid in the processing and biogenesis of miRNAs and this process is scaffolded by NEAT (Jiang et al., 2017).

Finally, as for domains required for binding the RRM2 domain and part of the NOPS domain) are necessary and sufficient for binding, whereas RRM1 is dispensable (Yarosh et al., 2015). A domain containing just RRM1 and RRM2, with no amino acids N- or C-terminally had lower affinity for RNA compared to a construct containing RRM2/NOPS. RRM1 but not RRM2 contains the canonical aromatic residues typically involved in RNA-RRM interaction.

My thesis work aimed to elucidate mechanisms governing PSF's RNA-binding affinity and specificity. Mechanistic understanding of RBP's nucleic acid binding and binding regulation provide insight into the normal workings of cellular life, as well as into how RBPs can experience dysfunction to create disease, so treatment is possible. Given PSF's propensity for binding structured RNA, the lack of binding consensus sequence and

PSF's noncanonical RRM2 driving RNA-binding, I believe PSF's RRMs represent a unique and unusual mode of RNA-recognition and binding. Here I combine multiple biophysical and biochemical techniques to show PSF uses an expansive binding interface to bind RNA. Moreover, I find that increasing flexibility of the RRM2/NOPS domains by mutating a loop increases binding affinity. I show two residues located in the NOPS domain, which when mutated to alanine ablate RNA-binding ability.

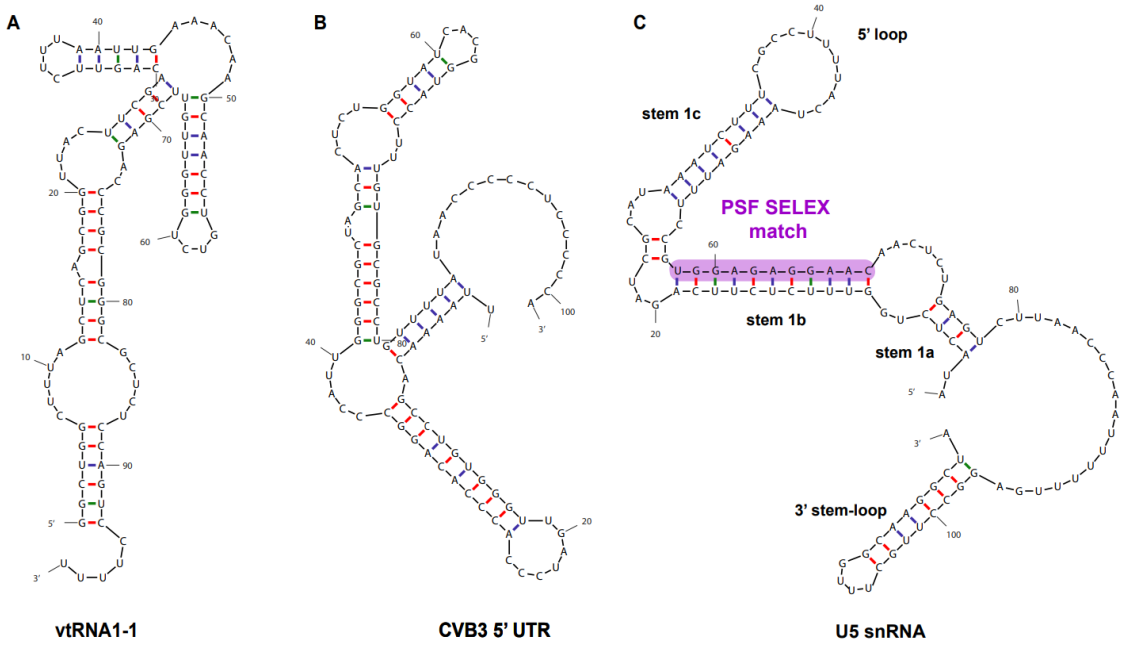


Figure 1.7 MFold structure prediction of RNA known to bind PSF. A) In humans, PSF binds three of four vault RNAs (vtRNAs), vtRNA1-1 shown. **B)** Structure prediction of the minimal truncation of coxsackievirus B3 5' UTR required to bind PSF. **C)** SELEX match for PSF/NONO located within the U5 snRNA

CHAPTER 2: Biophysical probing of PSF's RNA-binding interface

INTRODUCTION

In Chapter 2 I describe the characterization of PSF's interactions with RNA, as well as PSF's behavior in solution using several biophysical techniques. I present data from two different mass spectrometry approaches used to identify peptides involved in binding. Next, I outline preliminary data using Nuclear Magnetic Resonance to characterize PSF's RRM1 and RNA. Finally in this chapter I report data from molecular simulations describing PSF's hydrophobicity and electrostatics in solution in the apo, or unbound state.

As mentioned in chapter 1, PSF's core structured region consists of a proline-arginine rich linker (PRL), two RNA Recognition Motifs (RRMs), a homology domain known as NOPS (NONO/ParaSpeckle domain), an extended coiled-coil, and a nuclear localization signal (Fig. 2.1A). PSF contains N- and C-terminal unstructured regions, and functions in cells as a homodimer or a heterodimer with other DBHS proteins p54nrb/NONO or Pspc1. A partial crystal structure of PSF's homodimer was published by the Bond & Fox labs in 2015, which I utilize extensively in our research (PDB: 4WII) (Lee et al., 2015a) (Fig. 2.1). Monomers domain swap to aid in dimerization and result in an unconventional spatial arrangement of RRMs unique to the DBHS family (Fig. 2.1B).

RNA recognition motifs (RRMs) are the most abundant RNA-binding domain (Lunde et al., 2007). Abundance suggests an overall importance of RRMs in maintaining proper cell physiology. RRMs typically recognize single-stranded RNA or DNA using hydrophobic amino acids to base stack with RNA or to insert themselves between two sugar rings. The consensus sequences [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY], [ILV]-[FY]-[

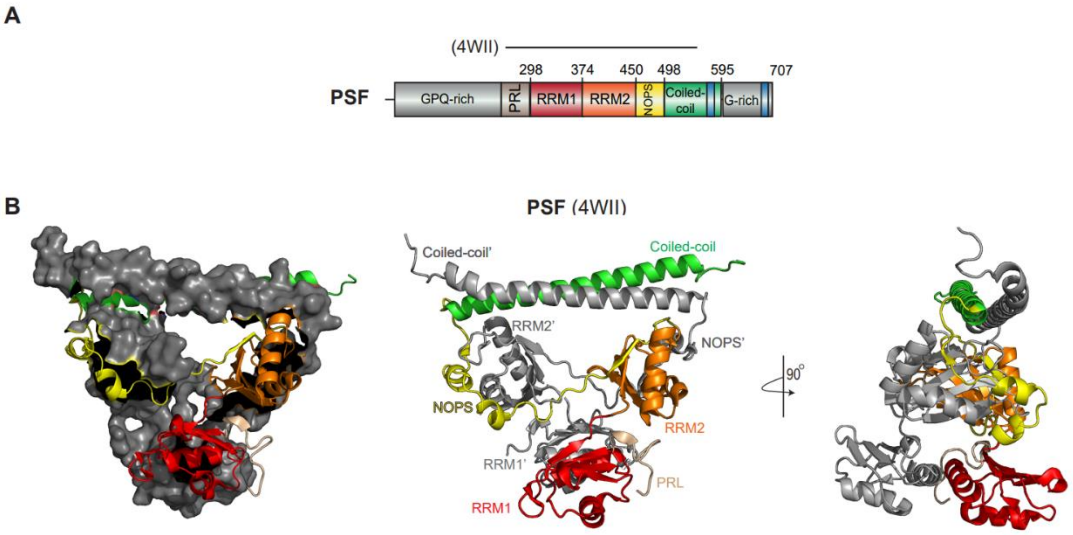


Figure 2.1 Structure of PSF homodimer (4wii). A) PSF domain structure with crystal structure 4wii spanning residues 276-535. **B)** Crystal structure showing one monomer in grey, first showing the surface model, followed by the more simplified ribbon diagram. Prime notation denotes one monomer from the other. Crystal structure solved by Lee et Al., NAR, 2015. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

ILV]-X-N-L, and [YW]-X-[DQ]-X are referred to as RNP1, RNP2, and RNP3. RNPs are in the β3-strand, β1-strand, and β2-strand respectively (Fig. 2.2) (Daubner et al., 2013). An example of canonical RNP mediated binding is in SRSF1 where residues in all three RNPs of RRM1 are used to base stack with “AACAAA” RNA (Fig 2.3D) (Cléry et al., 2021). Specificity requires multiple contacts with surrounding amino acids, and some RRMs have additional β-strands and α-helices which can aid in binding affinity and specificity.

Of PSFs two tandem RRMs, RRM1 contains the consensus RNPs. In contrast, RRM2 does not contain the canonical aromatic residues. Additionally, RRM2 is followed by the NOPS domain, which contains the fourth β-strand of RRM2’s β-sheet surface (Fig. 2.3B&C). The NOPS domain from one monomer packs in an extra helix against the other copy of RRM2 as well (Fig 2.3C). The PSF construct (366-484) containing RRM2

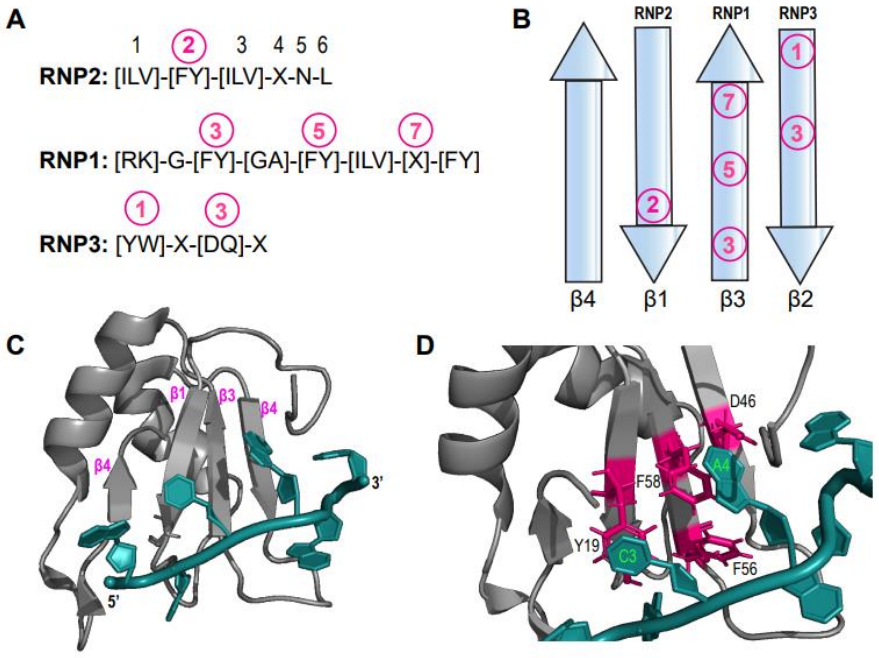


Figure 2.2 Structure and binding-motifs of RNA Recognition Motifs. A) Consensus binding motifs found in many RRMs. Numbers indicate amino acid position in the β-strand, **B)** Topology of β-strands creating a β-sheet binding surface for RNA. **C)** β-sheet binding surface of SRSF1 and RNA ligand AACAAA (Cléry, et al., 2021, PDB: 6HPJ). **D)** SRSF1 contains at least one residue in each RNP consensus to bind AACAA. Additional contacts are made by surround amino acids (not shown). Models made using PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

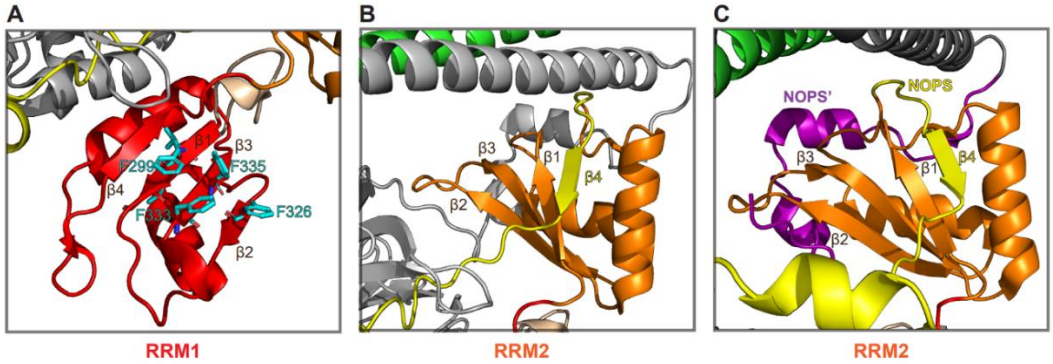


Figure 2.3 Structural details of PSF's RRMs. A) RRM1 contains canonical aromatic residues to base stack with RNA. **B)** RRM2 does not contain the canonical residues to base stack. Further, the β4-strand of RRM2 comes from the conserved NOPS domain. **C)** The NOPS domain from the other monomer domain swaps to pack two small helices with RRM2. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

and a portion of the NOPS domain is necessary and sufficient for binding RNA (Yarosh et al., 2015). Despite the canonical aromatic residues in RRM1, RRM1 is dispensable for binding (Yarosh et al., 2015).

Previously, we have shown PSF binds an exonic splicing silencer (ESS) sequence within exon 4 of CD45, which increasing skipping of exon 4 during T cell activation (Melton et al., 2007; Motta-Mena et al., 2011). The ESS sequence required for binding is 65 nucleotides (nts) long and is predicted to contain two short stem-loops by Mfold (Fig 2.4). Truncating ESS shorter on either end, or in the middle, ablates PSF's ability to bind. The ESS-RNA tends to dimerize, and molecular crowding (BSA titration) increases dimerization of the ESS RNA probe (unpublished data).

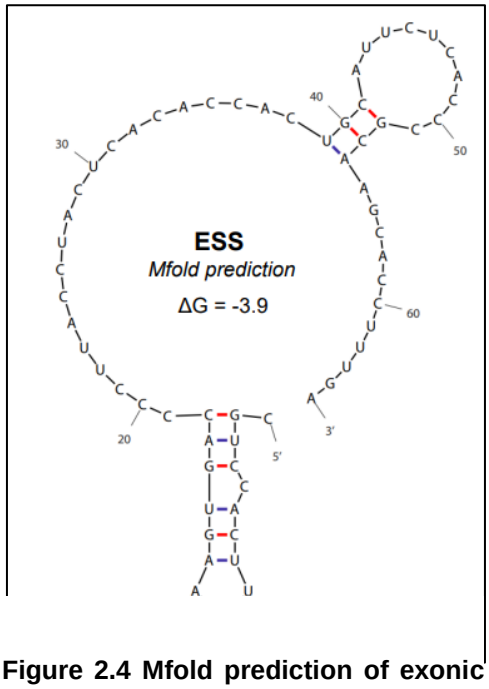


Figure 2.4 Mfold prediction of exonic splicing silencer (ESS) RNA.

PSF's preference for longer RNA makes it an unlikely candidate for crystallization as long flexible entities disrupt homogeneity. Screening attempts for crystallization were made with a 43- and 65-mer of ESS RNA but did not result in any co-crystals. Nuclear magnetic resonance is challenging on the PSF/RNA complex because of size. Instead, I combine multiple biophysical techniques to demonstrate PSF uses an expansive binding interface to bind RNA, indicating a modality of RNA-binding different from the canonical 2-8mer recognition.

RESULTS

Crosslinking reveals an extended RNA-interacting interface

Given evidence of an unusual RRM-RNA interface (RRM2 drives binding while lacking canonical RNPs), I first sought to determine regions of PSF that contact RNA using RNA-Binding Region Identification (RBR-ID). RBR-ID was developed at UPenn by Benjamin Garcia and Roberto Bonasio and utilizes a cross-linking mass spectrometry-based approach (Fig. 2.5A). Peptides cross-linked to RNA will shift in mass-to-charge ratio compared to the uncross-linked controls (Fig. 2.5B). A probabilistic value of RNA-binding can be calculated for each peptide, combining the extent of depletion and the p value (< 0.05 cutoff) according to the formula in (He et al., 2016).

I first examined published RBR-ID dataset (He et al., 2016), where *in vivo* cross-linking was conducted in mouse embryonic stem cells (mESCs). Human and mouse PSF share 99% sequence homology (Fig. 2.5D). Analysis of PSF peptides from this mouse study revealed several regions of PSF contact RNA, including regions outside of the RRMs (Fig 2.6A, Table 2.1). The region of PSF most likely to interact with RNA in was the NOPS domain, a domain extending from RRM2. The next most probable peptide to bind RNA is in RRM1 and has a phenylalanine corresponding to RNP 1. RNP 1 is one of the canonical binding sequences, is in β3-strand, and contains aromatic residues. Peptides from the extended coiled-coil were next and are close in sequence and in space to RRM2/NOPS. A single peptide from RRM2 spanning the β2-strand through loop 3 (β2β3-loop) was found (Table 2.1). While the in-cell experiments provide insight as to how PSF interacts with all nucleic acid in the cell in the context of other proteins, I was interested in how PSF alone

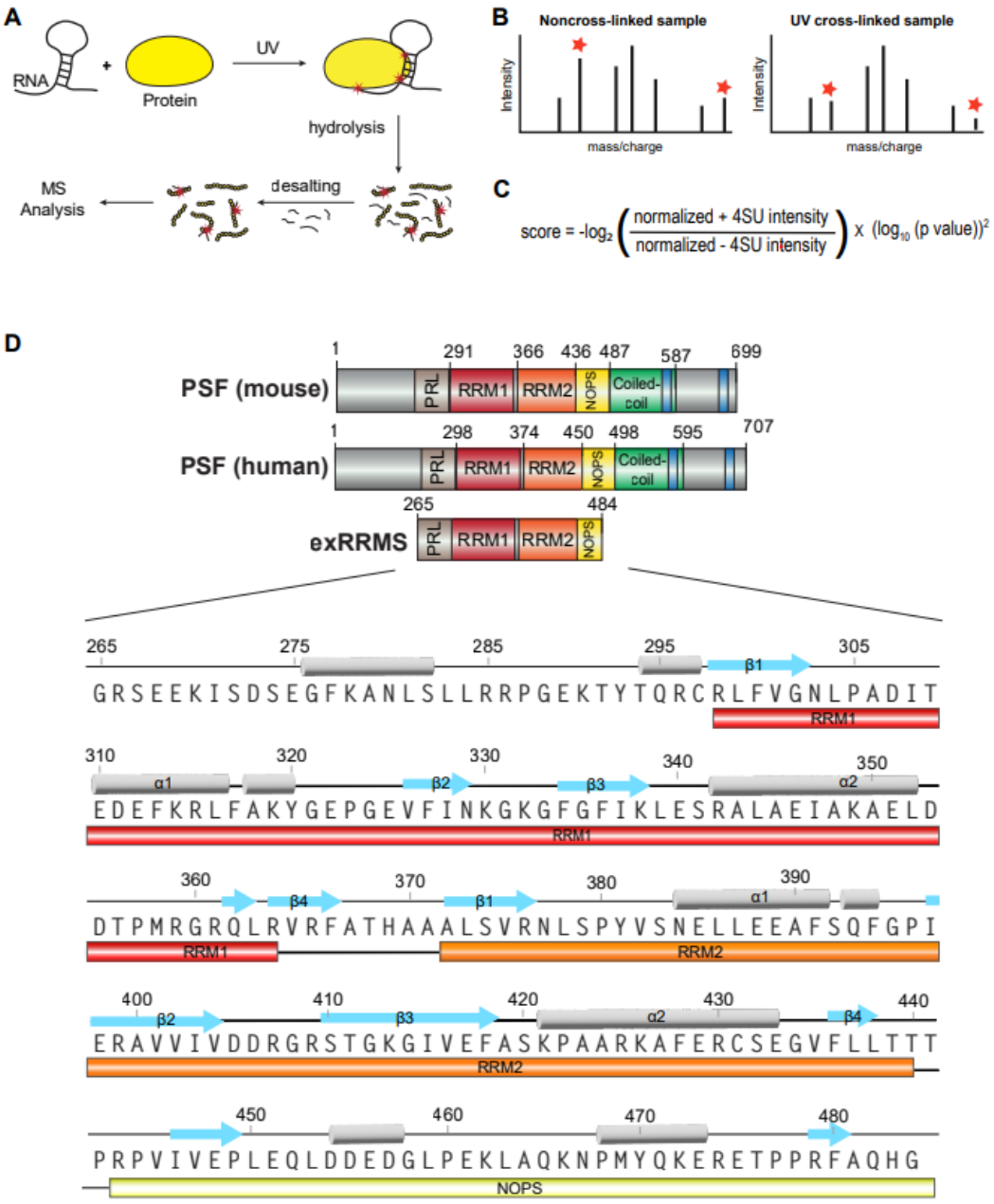


Figure 2.5 RNA-Binding Region Identification (RBR-ID) workflow & components. A) Workflow for RBR-ID consists of transcribing 4-thiouracil labeled RNA (either in cells or *in vitro*) followed by cross-linking at 312 nm UVB light, and normal mass spectrometry peptide preparation**. B)** Noncross-linked samples are compared to UV cross-linked samples and peptides crosslinked to RNA will change in mass-to-charge ratio resulting in a peptide drop out. **C)** RBR-ID score calculation **D)** Human and mouse PSF share 99% sequence similarity. The construct (human) used for *in vitro* RBR-ID experiments contains residues 265-284.

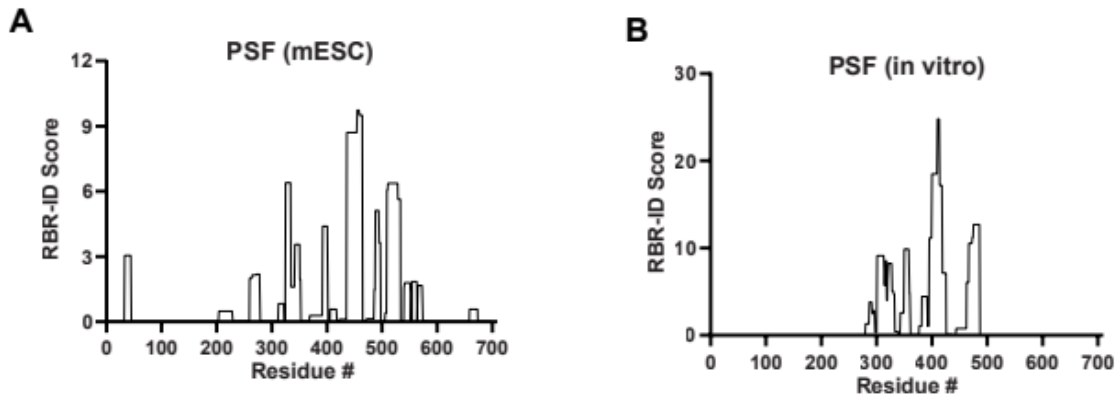


Figure 2.6 RBR-ID peptide plots. A) Peptides cross-linked to nucleic acid in mouse embryonic stem cells (mESC). **B)** Peptides cross-linked to ESS-RNA *in vitro* using recombinant PSF(exRRMs). RBR-ID score calculation shown in 2.5B.

own interacts RNA to understand PSF's mechanism of binding.

Our previous data demonstrates the CD45 ESS-RNA as a functionally relevant target of PSF, thus I specifically focused on the direct interaction of PSF with ESS-RNA. PSF's unstructured domains which aid in its functional aggregation as a paraspeckle component present a difficulty in purifying and working with soluble full-length protein. However, we've previously identified the minimal RNA-binding domain to contain RRM2 through the NOPS domain, and thus purified a truncated version of PSF for the biophysical characterizations termed extended RRMs PSF(exRRMs). The PSF(exRRMs) were incubated with 4-thiouracil labeled ESS-RNA, and then crosslinked and prepared as the other samples described in Figure 2.5A.

Given the domain structure for PSF in mouse and human are highly homologous, the data from the two experiments is easily comparable. Consistent with our previous work demonstrating the importance of RRM2 in the direct binding of PSF with target RNA, the peptides with highest probability of binding were in RRM2 and NOPS domains. Notably,

these include some of the same residues represented in the mESC data (Fig. 2.6B, Table 2.2) The coiled-coil domain was not included in the PSF(exRRMs) construct, but from past work we know it is dispensable for binding (Yarosh et al., 2015). In RRM1, peptides from β3-strand and β2- strand are indicated in binding corresponding to RNP2 and RNP3 respectively. The peptides involved in interacting with RNA represent roughly 30% of the accessible surface area of PSF(exRRMs) (Table 2.2, total ASA ~26316 Å)(Krissinel & Henrick, 2007). Together these data indicate an expansive RNA-binding interface. While several regions of PSF are interacting with RNA it does not necessarily mean they are required for binding and additional methods of testing are needed. I employed a complementary method to better understand PSF's interactions with RNA as well as PSF's dynamics in solution.

Table 2.1 Additive RNA Binding Region-Identification scores across overlapping peptides in mouse embryonic stem cells.

Residue	Sequence	RBR-ID Score	Domain
455-458	LAQK	9.7	NOPS
459-464	NPMYQK	9.5	NOPS
436-454	PVIVEPLEQLDDEDGLEPK	8.7	RRM2/NOPS
334-342	GFGIKLESR	6.4	RRM1
511-528	*LESEMEDAYHEHQANLLR*	6.4	Coiled-coil
509-510	DK	6.1	Coiled-coil
529-533	QDLMR	5.6	Coiled-coil
488-494	SLDEMEK	5.1	Coiled-coil
400-409	AVVIVDDRGR	4.4	RRM2
495-497	QQR	3.6	Coiled-coil

Table 2.2 Additive RNA Binding Region-Identification scores across overlapping peptides in vitro.

Residue	Sequence	RBR-ID Score	Domain	ASA
410-413	STGK	24.8	RRM2	435.6
400-409	AVVIVDDRGR	18.5	RRM2	1539.9
414-418	GIVEF	17.2	RRM2	17.8
475-484	ETPPRFAQHG	12.7	NOPS	2277.2
473-474	ER	11.2	NOPS	518.1
395-399	GPIER	11.1	RRM2	428.8
467-472	NPMYQK	10.5	NOPS	1221.0
301-313	VGNLPADITEDEF	9.1	RRM1	1087.3
316-317	LK	8.5	RRM1	83.0
321-327	GEPGEVF	8.2	RRM1	890.6

ASA = Accessible surface area (Å). PRL = Proline/Arginine linker. RRM1 = RNA Recognition Motif 1. RRM2 = RNA Recognition Motif 2. NOPS = NONO/Paraspeckle Domains.

PSF structure, dynamics, and RNA interactions via Hydrogen-Deuterium Exchange Mass spectrometry

As a complementary approach to *in vitro* RBR-ID, I used to Hydrogen-Deuterium Exchange Mass Spectrometry (HDX-MS), to study the impact of RNA-binding on PSF. HDX-MS measures solvent accessibility of the amide protein backbone (Fig. 2.7A). As hydrogens on the backbone exchange with deuterium in solvent, the mass to charge ratio of peptides changes, representing the change in mass unit from deuterium (Fig. 2.7B), with peptides that change most rapidly representing regions of greatest flexibility and/or accessibility in the protein. Regions that are stably structured in β-sheets and α -helices will exchange slowly, as do regions involved in intermolecular interactions, such as with ligand or, in this case, RNA. Analysis of HDX-MS data using the same PSF(exRRMs) as in the RBR-ID experiment above (Fig 2.5D), reveals peptides of

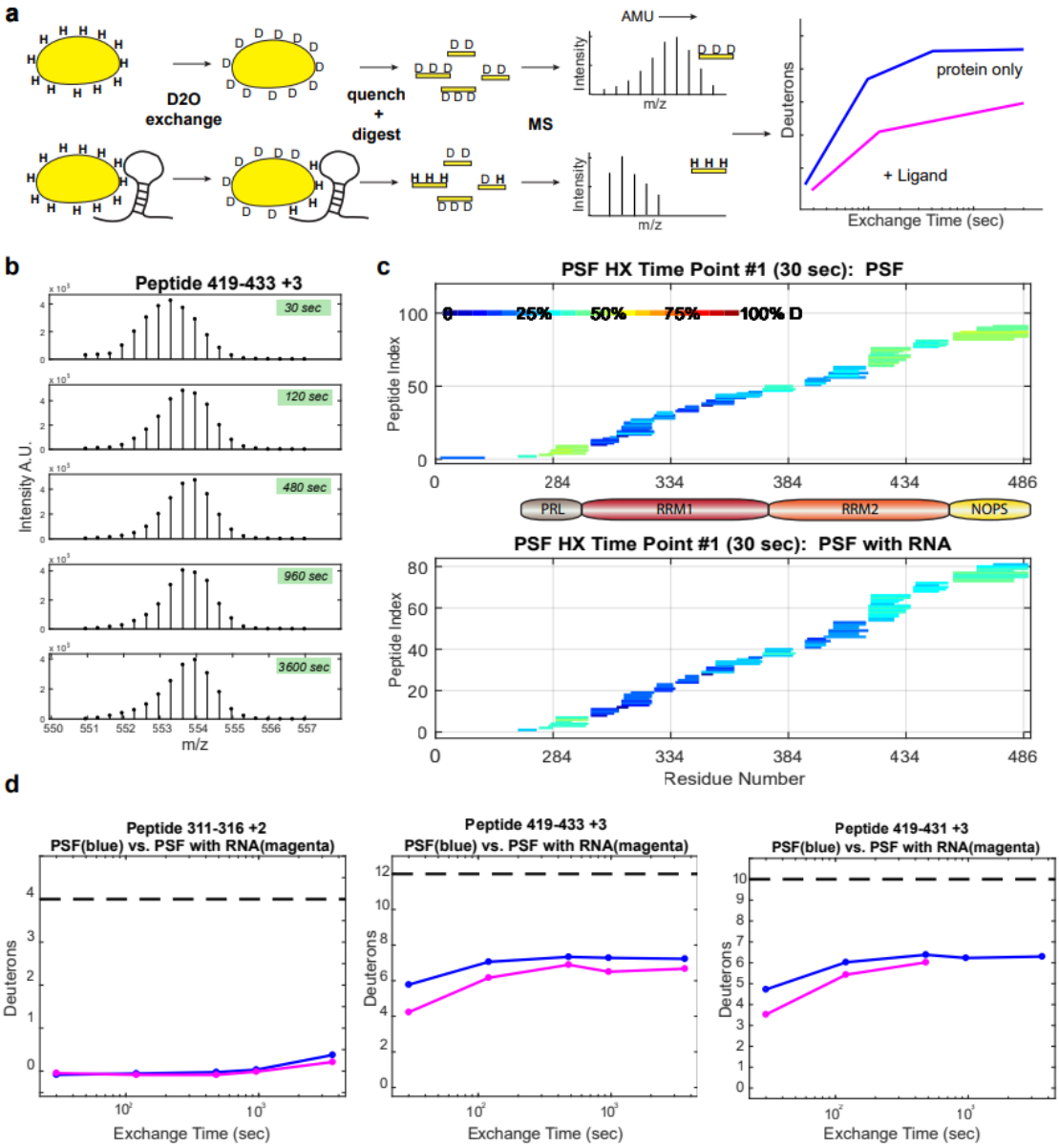


Figure 2.7 PSF's RNA binding by Hydrogen Deuterium Exchange Mass Spectrometry. a) Workflow of HDX-MS and expected output of peptide bound to RNA. **b)** Shift in mass to charge ratio of peptide upon uptake of deuterons. **c)** Deuteron exchange across the entire PSF(exRRMs) construct with and without RNA. **d)** Example peptides showing either no change upon RNA-binding or decrease uptake upon RNA-binding.

slower exchange (stably structured) largely correspond to residues confirmed by crystal structures as α-helices and β-sheets, as is to be expected. Conversely, peptides of loop

regions exchange more quickly (Fig. 2.7C). Exchange rates of RRM1 with solvent are slower than RRM2 indicating a more rigid structure than RRM2 (Fig 2.8). Flexibility of RRM may be required for RNA recognition.

At early timepoints, PSF samples incubated with ESS-RNA show differential exchange in the PRL, much of RRM2 and nearly all the NOPS domain (Fig. 2.7C). Differential exchange of peptides was seen in biological replicates, and importantly this difference is present in separate overlapping peptides, but not across the entire protein (Fig. 2.7D & Fig. 2.8). The most protected areas were the PRL, RRM2, and NOPS domain, with one peptide from RRM1 containing the canonical RNP1 (Fig. 2.7E). These peptides represent roughly 60% of the total accessible surface area (ASA) of PSFexRRMs (Table 2.3)(Krissinel & Henrick, 2007). Full deuterated samples were incubated for 1-hr at the melting point determined by circular dichroism (Appendix A).

Table 2.3 Summary of top protected peptides in presence of ESS-RNA in HDX-MS.

Residue	Sequence	% Protection	Domain	ASA
271-277	KISDSEG	14.6	PRL	*
285-292	LRRPGEKT	13.1	PRL	1440.5
418-431	ASKPAARKAFERC	13.1	RRM2	1635.3
403-414	IVDDRGRSTGKG	12.4	RRM2	1868.1
457-484	DGLPEKLAQKNPMYQKERETPP RFAQHG	10.6	NOPS	5884.8
328-335	INKGKGFG	10	RRM1	952.2
392-400	SQFGPIERA	8.6	RRM2	833.7
374-385	SVRNLSPYVSNE	7.4	RRM2	1372.2
437-452	LLTTTPRPVIVEPLEQ	4.9	RRM2/NOP S	2165.2

ASA = Accessible surface area (Å). PRL = Proline/Arginine linker. RRM1 = RNA Recognition Motif 1. RRM2 = RNA Recognition Motif 2. NOPS = NONO/Paraspeckle Domains.

* Residues not present in crystal structure to calculate ASA

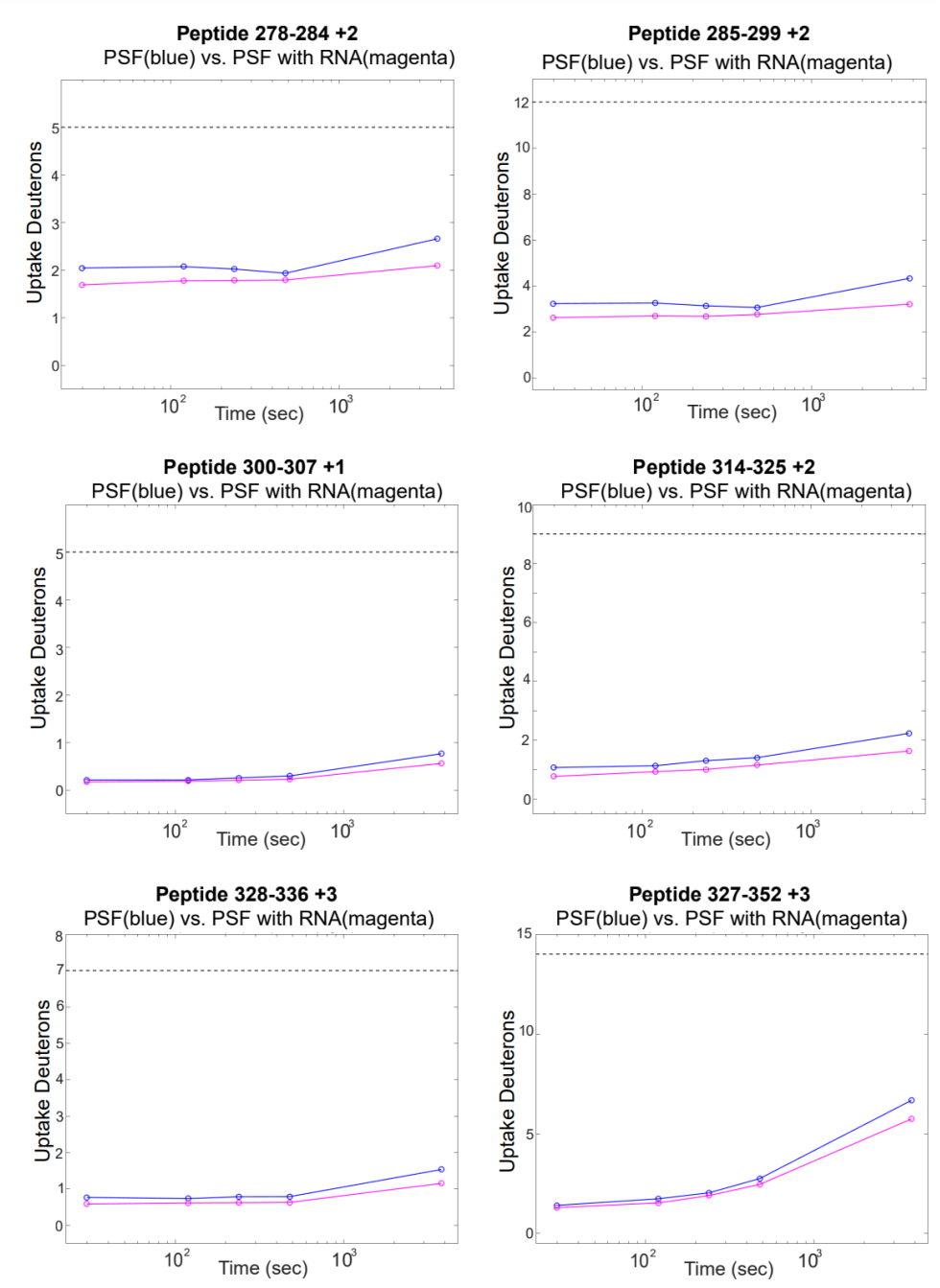
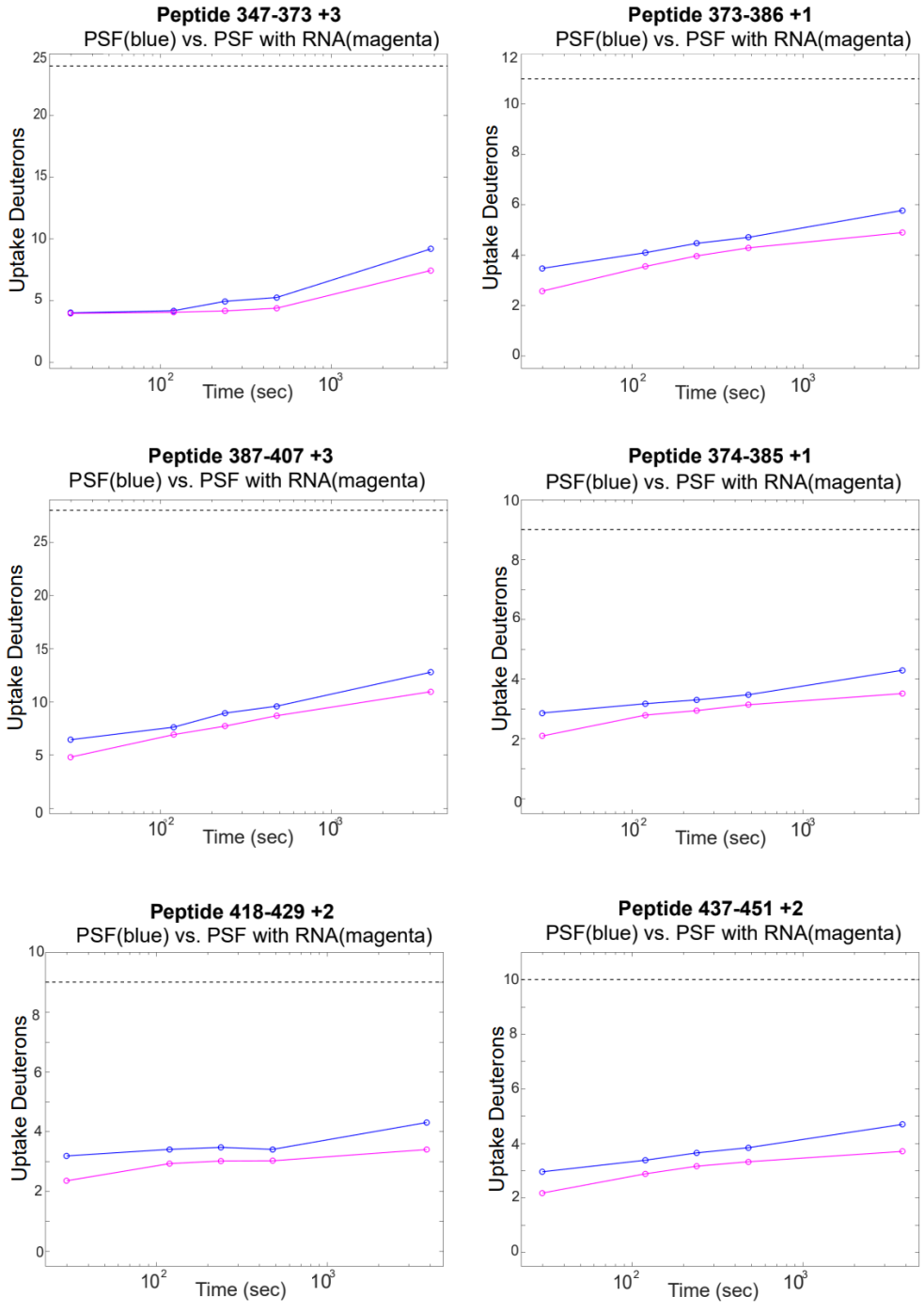


Figure 2.8 Hydrogen Deuterium Exchange Mass Spectrometry. Peptides exchange as uptake of deuterons overtime. PSF alone (blue) shown with PSF incubated with ESS-RNA (magenta). Dotted line indicates the theoretically calculated maximum exchange for each peptide. Peptides shown are a sampling of the whole data set. Peptides continue on the next page.

Figure 2.8 Hydrogen Deuterium Exchange Mass Spectrometry (cont'd).



Comparative mapping of RNA interactions indicates similar regions contact RNA

The two techniques used above tell a slightly different story about protein-RNA interactions. RBR-ID captures regions of protein near RNA, whereas HDX-MS measures regions on the protein that have an altered chemical environment in the presence of RNA (or another ligand like protein, small molecule, or DNA). To compare the RBR-ID data and HDX-MS data, I mapped my results onto the partial crystal structure of the PSF homodimer (PDB: 4WII, Lee et al., 2015), leaving out the coiled-coil domain as the constructs used do not contain this portion (Fig. 2.9A & B). Visualization of data shows regions interacting with RNA are concentrated in the RRM2/NOPS domain, as well as an expansive binding interface utilizing regions outside of the minimal binding domain. A few regions are of interest, namely the β2β3-loop and the NOPS domain (Fig. 2.9C & D). The loop is moderately structured making contacts between sidechains of the first aspartic acid and last arginine (Fig. 2.9D). Peptides in the NOPS domain also overlap in the two experiments (Fig. 2.10). *In vitro* RBR-ID data shows most of the domain interacting with RNA with spanning residues 467-484 the sequence “NPMTQKERETPPRFAQHG” (Table 2.1). As mentioned previously, the NOPS domain from one monomer packs against RRM2 of the other monomer and may either be directly involved in binding or modulate RNA’s access to RRM2. While RRM1 is dispensable for binding, it interacts with RNA in both experiments (Fig. 2.11). Contact points in RRM1 could aid in binding specificity. Peptides containing aromatic residues within RRM1’s canonical RNA-binding motifs are present in both experiments (Fig. 2.11 A&B). Phenylalanine-300 (F300) found in the β1-strand represents the canonical RNP2 binding motif. In the β3-strand sits RNP1, which includes F334 & F336 (Fig. 2.11B).

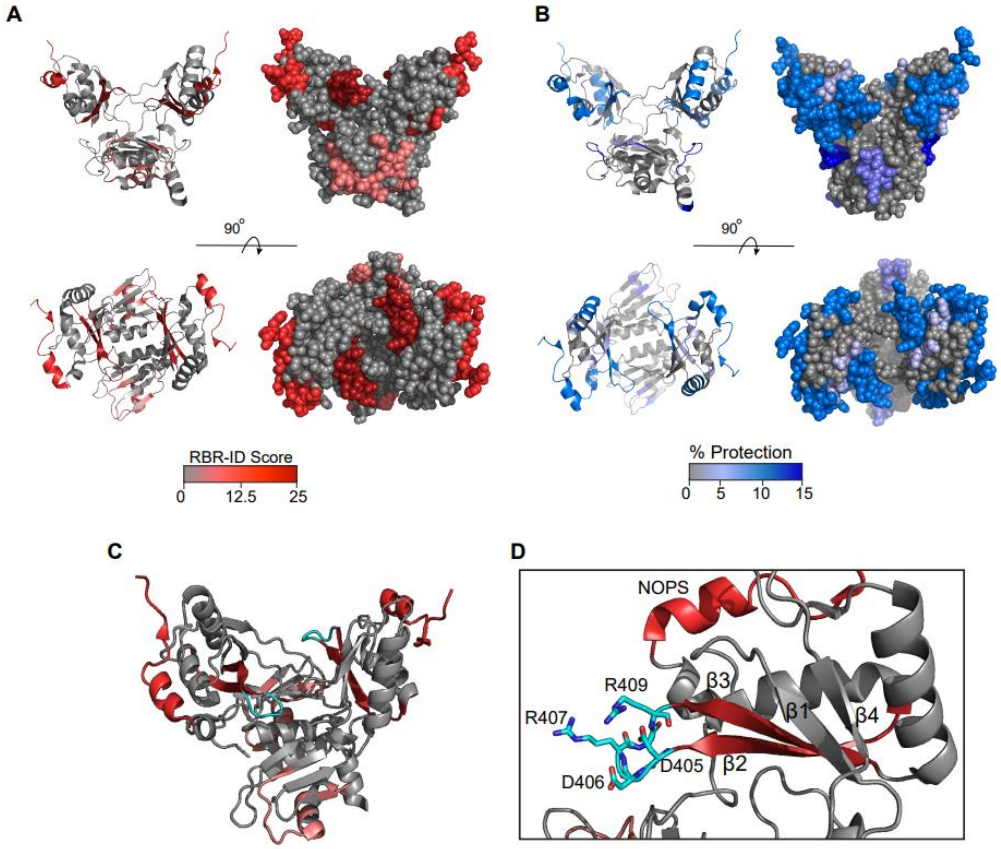
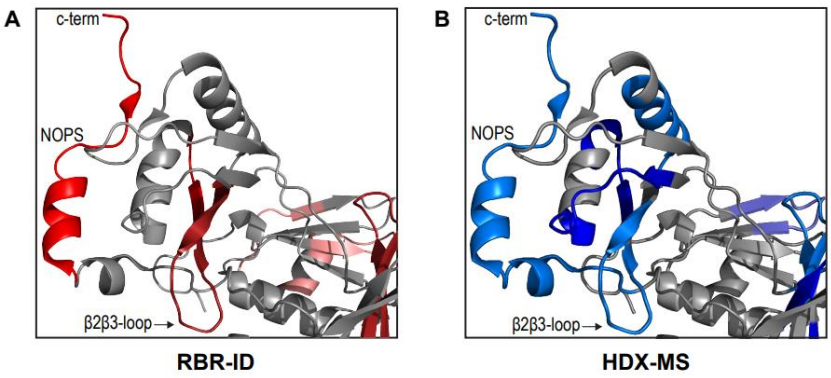


Figure 2.9 Comparative mapping of PSF's RNA-binding indicate expansive binding interface. A) RBR-ID data mapped from *in vitro* PSF/RNA cross-linking mapped onto partial crystal structure (PDB: 4wii). B) Exchange data from HDX-MS mapped onto same crystal structure. C) RBR-ID data mapped, cyan showing B2B3-loop. D) Zoom in of RRM2's β2β3-loop "DDRGR," and NOPS domain peptide consistently appearing in both experiments. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

Figure 2.10 Comparative mapping of PSF's NOPS domain. A) RBR-ID data mapped onto the crystal structure of PSF. B) Hydrogen-Deuterium exchange data mapped. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.



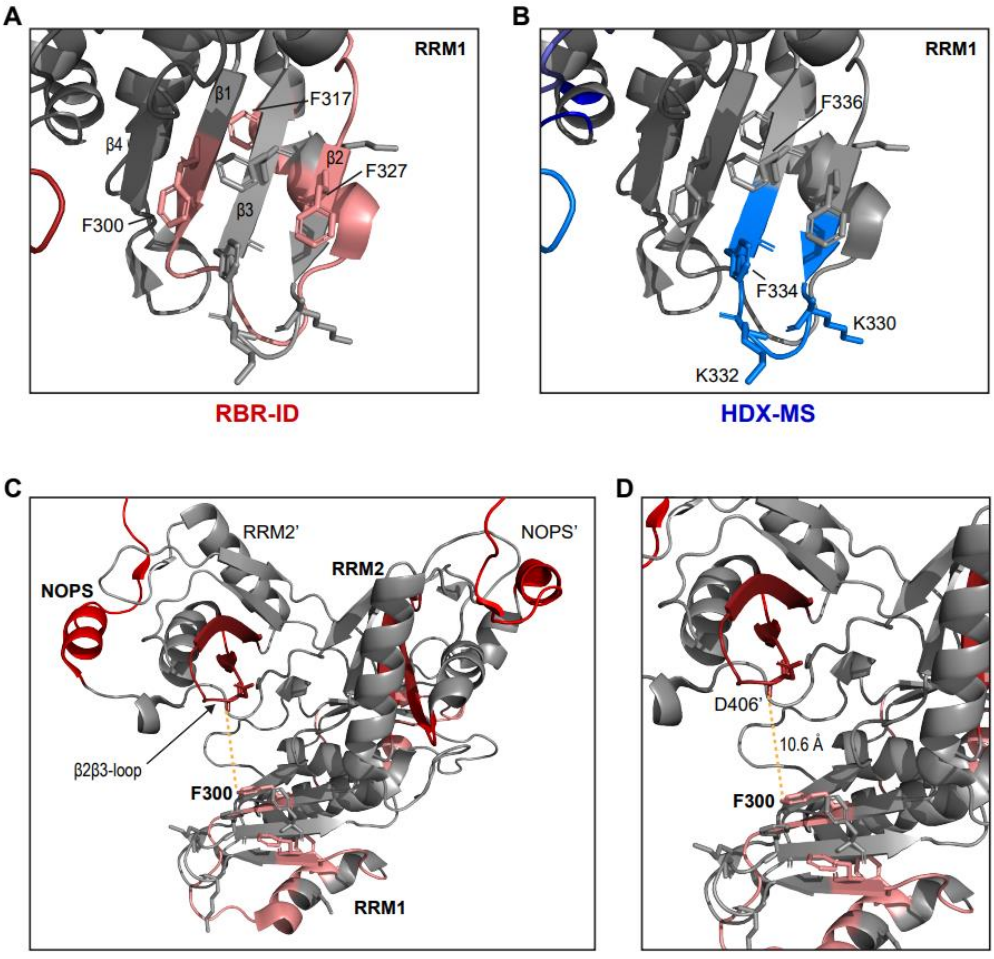


Figure 2.11 Comparative mapping of RRM1. A) Pictured are the aromatic residues making up the canonical RNPs in RRM1, which are shown to be interacting with RNA in RBR-ID. B) Phenylalanine-334 is in RRM1's β3-strand and represents the canonical RNP 1 motif. C) Phenylalanine-300 is part of RNP2 and spatially near the β2β3--loop. D) Aspartic Acid-406 begins the β2β3—loop in RRM2' and is 10.6 Angstroms (Å) away from F-300 in RRM1. Prime notation distinguishes one monomer from the other. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

Phenylalanine-300 of RRM1's β-sheet is only 10.7 Angstroms (Å) from Aspartic acid-406, which begins the β2β3-loop in RRM2' (Fig. 2.11 C&D). The striking degree of overlapping

regions in the biophysical mapping of RNA-interaction interface help to guide my mutational analysis described in the next chapter.

RRM1/RNA interactions by Nuclear Magnetic Resonance

As mentioned in Chapter 1, the size of PSF/ESS do not make it an ideal candidate NMR. However, in collaboration with the Sattler lab, we did attempt to collect data on smaller constructs/RNA. Because of the size limitations of NMR, both the protein construct and the RNA had to smaller, adding further complications. This reduces overall affinity and therefore reduces signal. Nishtha Galati collected Nuclear Magnetic Resonance (NMR) data on PSF(minRRM1) (residues 299-249) and the shorter 43-mer ESS-RNA. (Fig. 2.12). No strong interactions with the aromatic residues across the β-sheet surface were noted, but some weak changes across residues V326 K338, and R365 when ESS-RNA were present, as well as unspecific interactions between many copies of RRM1 and a single ESS-43mer (Fig. 2.12B&C). These residues are solvent accessible as analyzed by PISA using crystal structure 4wii. This data can be see mapped onto the crystal structure (PDB code: 4wii) in Figure 2.13.

Hydrophobic surface of PSF via molecular simulation

Hydrophobic and electrostatic interactions are important for protein-nucleic acid binding, so I set out to identify the hydrophilic and hydrophobic portions of PSF by molecular simulation. In collaboration with Nicholas Rego from Amish Patel's lab, we identified hydrophobic protein patches utilizing molecular simulation that systematically displace

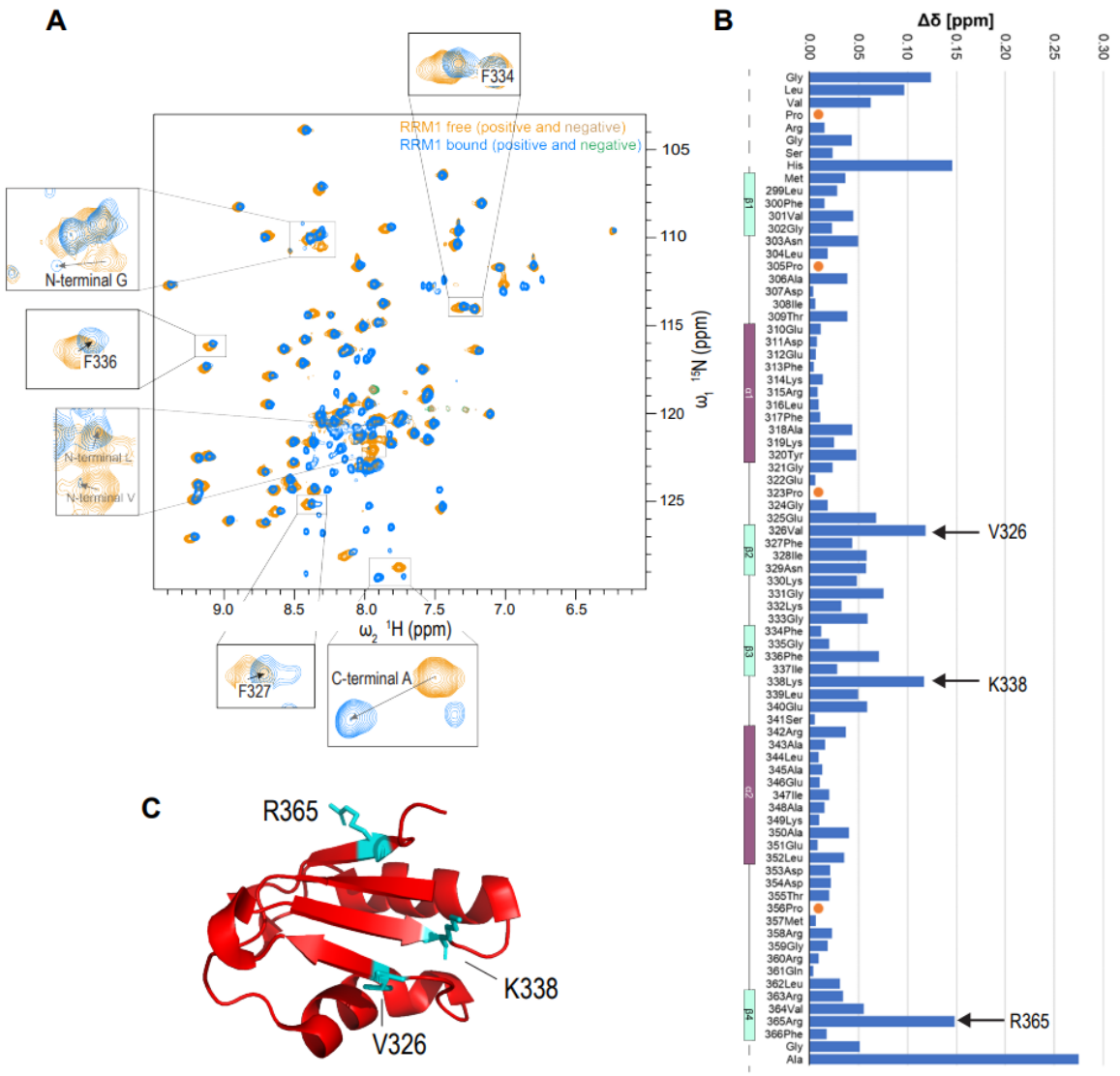


Figure 2.12 Weak RRM1/ESS-RNA interactions via Nuclear Magnetic Resonance. A) 2-dimensional NMR HNC2D ^{15}N–^{1}H HSQC spectra collected on RRM1 and RRM1/ESS 43-mer. B) Residues with individual ppm shifts between RRM1 spectra vs. RRM1/ESS 43-mer. C) Three residues showing weak interactions with RNA are shown in cyan: valine-326 (V326), lysine-338 (K338) arginine-365 (R365). Data collected and analyzed by Nishtha Galati. PDB code: 4wii. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

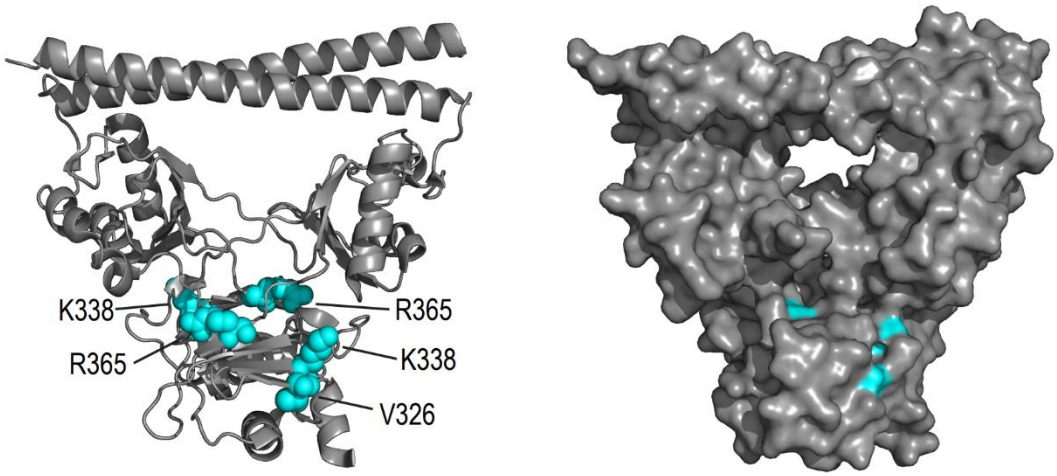


Figure 2.13 RRM1/ESS-RNA interactions by NMR mapped onto PSF homodimer (4wii). Shown are the residues in PSF(minRRM1) indicated to interact with RNA by NMR: V326, K338, and R365 were mapped to the crystal Structure solved by Lee et al., 2015. PDB code: 4wii. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

water molecules from the protein hydration shell (Rego, et al., 2021). Hydrophobic patches are not uniformly nonpolar but instead have polarity and charge at the nanoscale. To determine hydrophobic regions, we used molecular dynamic simulations where an unfavorable biasing potential (**φ)** was applied systematically, stripping protein hydration waters, revealing regions that have the weakest interactions with water (Fig. 2.14)(Rego et al., 2021). Removal of the protein hydration shell in response to applying biasing potential is referred to as “dewetting.” Dewetting is not uniform, and thus hydrophobic patches are revealed.

A list of the residues deemed hydrophobic by **φ** -ensemble simulations can be seen in Table 2.4. Chain A and B represent each monomer within the dimer. Residues along the coiled-coil domain dewetted. Given the coiled-coil domain facilitates oligomerization between PSF dimers, and hydrophobic regions are indicative of protein-protein interaction interfaces, it is unsurprising that a hydrophobic patch is present in this domain.

Another large hydrophobic patch revealed by biasing potential includes parts of PSF's RRM1, RRM2, and the NOPS domain. One of the residues in RRM1 that dewets, R365, was also chemically changed upon adding RNA in NMR (Table 2.4, Fig. 2.12 & 2.13). Portions of the NOPs domain that dewet also changed upon addition of RNA in HDX-MS experiments (Table 2.3 and 2.4). Despite the charged amino acid composition of RRM2's β2β3-loop (D-D-R-G-R), the dewetting algorithm indicates a more hydrophobic-like state. While this may sound counterintuitive, dewetting of polar molecules is common and occurred in all proteins analyzed using this simulation (Rego et al., 2021). When some non-polar regions dewet, the neighboring protein regions become more susceptible to dewetting, despite polarity(Rego et al., 2021). Predicting hydrophobicity must account for the solvent responses of all residues contextually as a part of the whole. Thus, hydrophobic patches on proteins are often heterogenous and poorly correlative with polarity.

Additionally, the β2β3-loop in RRM2 shows a positive electrostatic potential (Fig. 2.15A)(Pettersen et al., 2021). PSF is highly conserved, the loop is among the more conserved regions, and present in the peptides indicated to interact RNA in both RBR-ID and HDX-MS experiments (Fig. 2.15B, Table 2.2 & 2.3)(Ashkenazy et al., 2016; Pei & Grishin, 2001; Pettersen et al., 2021). Finally, the β2β3-loop loop has higher B-factors indicating more disorder in this region (Fig. 2.16).

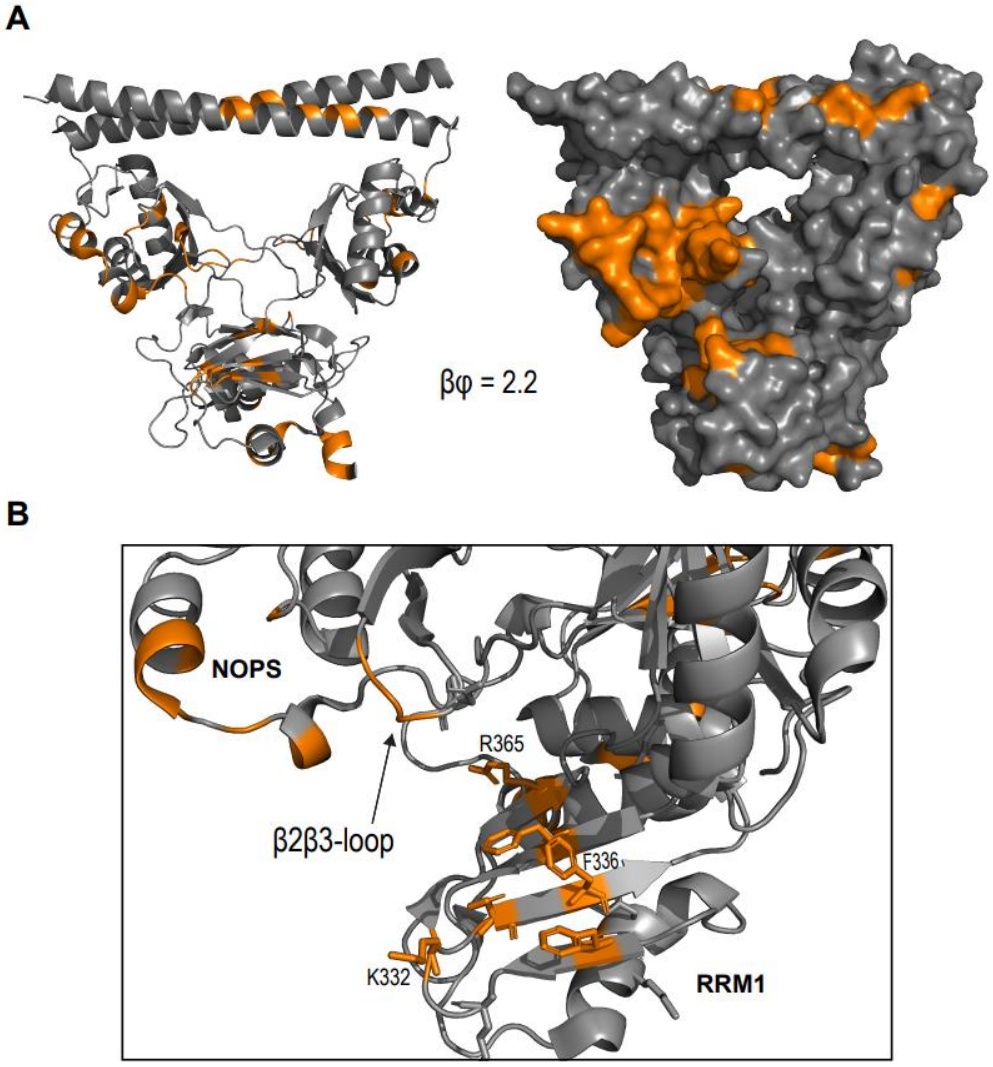


Figure 2.14 Dewetted atoms in φ-ensemble simulation. A) Cartoon ribbon diagram of PSF (L) with dewetted residues colored in orange. Surface diagram shown on right. Collective dewetting measured at the unfavorable biasing potential of βφ=2.2. **B)** Interestingly, not all residues in the hydrophobic patch are non-polar. Part of the β2β3-loop "RGR" dewets, thus exhibiting hydrophobic characteristics despite the polarity. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

Table 2.4 Residues dewetting at a biasing potential of 2.2.

Residue	Chain	Domain	Residue	Chain	Domain
GLY 277	A	PRL	PHE 300	B	RRM1
PHE 278	A	PRL	LEU 316	B	RRM1
ALA 280	A	PRL	PHE 327	B	RRM1
ASN 281	A	PRL	LYS 332	B	RRM1
LEU 282	A	PRL	PHE 334	B	RRM1
PHE 300	A	RRM1	PHE 336	B	RRM1
GLY 302	A	RRM1	ARG 363	B	RRM1
ASN 303	A	RRM1	PRO 380	B	RRM2
GLU 312	A	RRM1	TYR 381	B	RRM2
LEU 316	A	RRM1	ASP 405	B	RRM2
LYS 319	A	RRM1	ARG 407	B	RRM2
TYR 320	A	RRM1	GLY 408	B	RRM2
LYS 332	A	RRM1	ARG 409	B	RRM2
GLY 333	A	RRM1	SER 410	B	RRM2
PHE 334	A	RRM1	THR 411	B	RRM2
PHE 336	A	RRM1	GLY 412	B	RRM2
ARG 363	A	RRM1	PHE 428	B	RRM2
TYR 381	A	RRM2	LYS 462	B	NOPS
PHE 394	A	RRM2	LEU 463	B	NOPS
ARG 407	A	RRM2	GLN 465	B	NOPS
GLY 408	A	RRM2	ASN 467	B	NOPS
ARG 409	A	RRM2	PRO 468	B	NOPS
LEU 453	A	NOPS	MET 469	B	NOPS
GLY 458	A	NOPS	TYR 470	B	NOPS
LEU 459	A	NOPS	PHE 480	B	NOPS
PRO 460	A	NOPS	ARG 493	B	NOPS
LYS 462	A	NOPS	SER 496	B	NOPS
LEU 463	A	NOPS	LEU 497	B	NOPS
LYS 466	A	NOPS	MET 500	B	Coiled-coil
ASN 467	A	NOPS	GLN 504	B	Coiled-coil
PRO 468	A	NOPS	GLN 507	B	Coiled-coil
MET 469	A	NOPS	VAL 508	B	Coiled-coil
LYS 472	A	NOPS			
GLN 504	A	Coiled-coil			
GLN 507	A	Coiled-coil			
VAL 508	A	Coiled-coil			

PRL = Proline/Arginine linker. RRM1 = RNA Recognition Motif 1. RRM2 = RNA Recognition Motif 2. NOPS = NONO/Paraspeckle Domains.

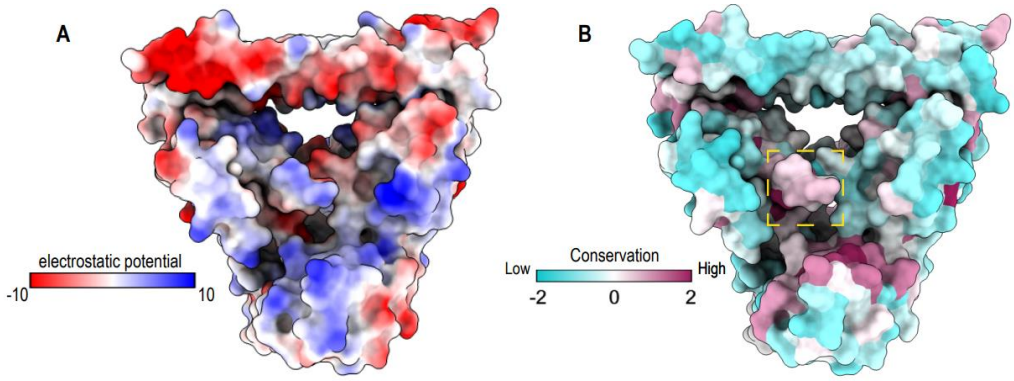


Figure 2.15 Electrostatic potential & conservation of PSF. A) Electrostatic potential of PSF mapped to the surface of crystal structure (PDB:4wii) **B)** Sequence conservation mapped onto PSF. Conservation analysis by AL2CO (Pei & Grishin, 2001). Yellow box indicates the conserved β2β3-loop.

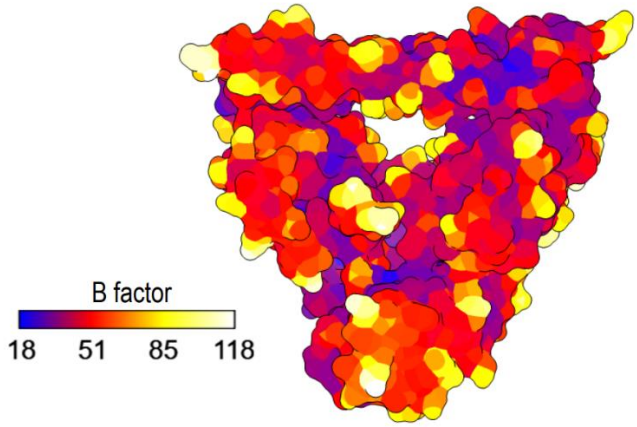


Figure 2.16 B-factors mapped to PSF homodimer (4wii). Model made using UCSF ChimeraX.

DISCUSSION

While RNA Recognition Motifs have a canonical fold and binding motifs, the structure and function of RRMs are diverse, with the ability to being ssRNA, ssDNA, as well as other proteins. PSF's RRM2 represents yet another flavor of diversity, interacting with RNA in a unique fashion with an expansive RNA-binding interface. Previously, we have shown RRM2 to be necessary and sufficient for binding to CD45 ESS-RNA. Which residues drive

PSF's RNA-binding was previously unknown. Here I show several regions of the protein interacting with RNA. Portions of RRM2 and NOPS are required for binding, and the well conserved β2β3-loop within RRM2 may be important for binding by several different methodologies. Peptides in the NOPS domain were present in both *in situ* and *in vitro* RBR-ID, the HDX-MS experiments, and molecular simulations determining hydrophobicity. Although RRM1 is dispensable for binding in *in vitro* binding assays (Yarosh et al., 2015), here I show RRM1 makes contacts which may aid in binding affinity or specificity via RBR-ID, HDX-MS, and NMR. Peptides bound to RNA could overrepresent transient interactions as UV-crosslinks occur based on proximity. Additionally, accounts of PSF binding DNA have been shown, and thus the in cell RBR-ID data may also include DNA-protein crosslinks. However, much of the binding surface seen in RBR-ID is corroborated by HDX-MS, which would not overrepresent transient interactions.

HDX-MS does come with other caveats, in that differential exchange merely indicates a change in chemical environment, and therefore could result from direct ligand binding or conformational change. If binding of RNA affects protein conformation some regions of the protein more/less susceptible to exchange with solvent. The high degree in overlap of these two techniques is instrumental to our analysis of this binding interface. Thus, complementary techniques suggest regions of PSF that may bind RNA.

Applying a biasing potential across the protein revealed hydrophobic patches on PSF. The β2β3-loop, R365 within RRM1, residues in the NOPS and coiled-coiled dewet at a biasing potential of φ=2.2. In general, hydrophobic regions are indicative of protein-protein interaction interfaces and might represent binding surface for regulation of binding. For example, this may represent a binding surface for PSF's own C-terminal end (C-term).

The phosphorylation state of PSF'S C-term regulates PSF's interaction with TRAP150. It could also be a binding surface for other proteins PSF interacts with. Additionally, these hydrophobic portions could be interacting with RNA via base stacking interactions, or through insertion of hydrophobic loops into the major and minor grooves RNA as shown in FUS's RRM in chapter 1.

In the next Chapter, I describe mutational analysis followed by binding assays to directly determine the relevance of individual residues in binding to RNA.

CHAPTER 3: Mutational Binding Analysis

INTRODUCTION

Several lines of evidence suggest that the RNA binding protein PSF interacts with RNA in a manner distinct from modes typically ascribed to RRM-containing proteins. In the previous chapter, I describe the use of multiple complementary biophysical techniques to identify regions of PSF that interact with RNA or are in a perturbed chemical environment when RNA is present. This information enabled us to make predictions about interaction between regions of PSF and RNA but does not confirm these residues drive binding. In this chapter, I use the structural information from Chapter 2 to guide mutational analysis of PSF. My results demonstrate that while several residues found within peptides from Chapter 2 have no effect on RNA-binding, residues T485 and R474 decrease binding when mutated. Creating these mutations did not change protein folding, stability, or dimerization. Mutating K413 in the β3-stand to alanine increased affinity for RNA. Mutating the entire β2β3-loop to alanine increased affinity for RNA 10-fold. It's possible that local or conformational changes induced by the increased flexibility of alanine mutation may regulate PSF's association with RNA.

RESULTS

Design of mutants

The biophysical characterization of PSF with RNA narrowed our focus on residues that may drive binding or interact directly with RNA. I created mutant protein constructs and conducted electromobility shift binding assays (EMSAs) to test the effects of mutation on RNA-binding directed by our biophysical data. A construct termed PSF(crystal-1) was

used for binding assays, which contains much of the coiled-coil domain, and the residues known to facilitate oligomerization (Fig. 3.1A).

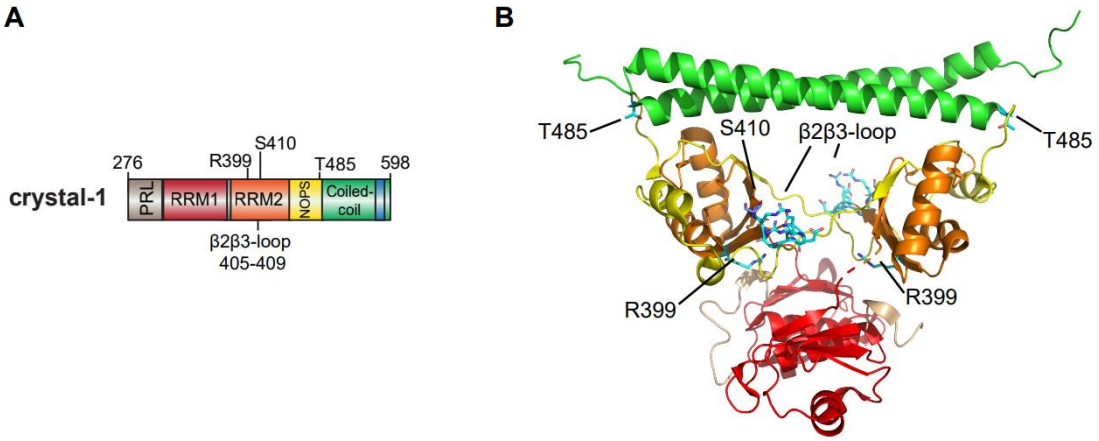


Figure 3.1 Mutational analysis PSF construct. A) PSF construct is named crystal-1 as it was the first crystal structure acquired (PDB: 4WIJ). B) Locations of mutations within the PSF homodimer are shown as sticks in cyan. Crystal structure solved by Lee et al, 2015. Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

A portion of RRM2 containing the β2-strand, the β2β3-loop, and the β3-strand are among the peptides most probable to bind RNA in the RBR-ID experiments (Fig. 2.6B & C). A similar overlapping peptide "IVDDRGRSTGKG" spanning residues 403-414 is seen in HDX-MS (Fig 2.7E & Table 2.3). We mutated several residues in and around this region, namely R399, S410, K413, and the entire β2β3-loop (Fig 3.1). Several factors about the β2β3-loop are intriguing. First, the recurrence of peptides from this area in the biophysical characterization indicate interaction with RNA. Next the β2β3-loop is a highly conserved evolutionarily as well as between DBHS protein family members (Fig. 2.15)(Ashkenazy et al., 2016; Pei & Grishin, 2001). Finally, despite polarity of most residues in the loop ("DDRGR"), these residues behave more like a hydrophobic patch within the protein

according to molecular simulations (Fig. 2.12). Thus, we mutated the β2β3-loop to all alanine “AAAAA.” In the center of the PSF homodimer, there is a gap in electron density spanning between two copies of RRM2. Lysine-413 (K413) is positioned inwardly towards the gap. This gap could accommodate an RNA strand and the positively charged residues clawing inwardly could help mediate the negatively charged RNA backbone. To test the potential impact of the positive charge on residue 413, I mutated K413 to alanine.

I also made mutations in the NOPS domain. Within the peptides interacting with RNA in the NOPS domains, we identified two threonine residues (T476 & T485) that are phosphorylation targets of Glycogen Synthase Kinase-(GSK3). GSK3 phosphorylates several residues in PSF; S33, T476, T485, and T687 (Shinde et al., 2017). Previously, we have demonstrated that phosphorylation of T687 on PSF’s C-terminal end promotes interactions with the protein TRAP150, which outcompetes RNA to bind PSF in a mutually exclusive fashion (Heyd & Lynch, 2010; Yarosh et al., 2015). We hypothesized that phosphorylation may drive conformational changes enabling TRAP150 to bind, and thereby preventing binding of RNA. N-terminal hyperphosphorylation of PSF has been shown to alter conformation and protein interactions (Shav-Tal et al., 2001). Therefore, we mutated T476 and T485 to test the effect on binding. Threonine-485 is also highly conserved and caps the beginning of the extended coiled-coil domain. Threonine is a helix breaker, so this placement may be vital to proper orientation of the coiled-coil domain(Imai & Mitaku, 2005). We also mutated several positively charged residues with the hypothesis that the residues might make electrostatic interactions with RNA’s negatively charged backbone either specifically or non-specifically. Several residues did not show significant effect on binding affinity: R399A, S410A, F436A, and T476A, are among these residues (Fig. 3.2A).

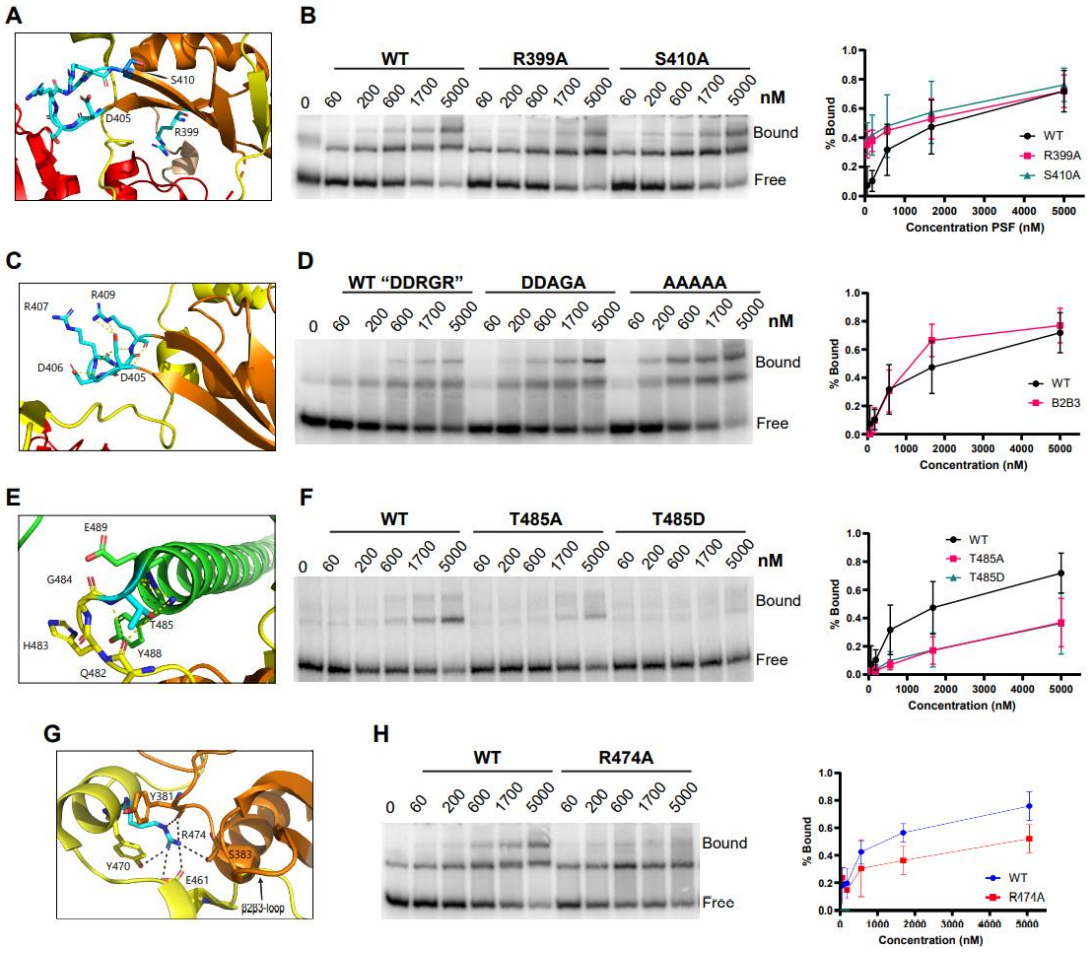


Figure 3.2 Binding analysis of mutations by electromobility shift assay (EMSA). A) Mutations R399A and S410A are in RRM2 and located in the β2-and β3-strands respectively. **B)** EMSAs gel, with protein concentrations starting at 5 μM and titrating down to 60 nM. Top species is the bound band while free probe is the bottom band. Quantification of n=3 EMSA on right. **C)** The β2β3-loop shown making charge contacts between loop residues. **D)** EMSA and quantification of β2β3-loop mutations**. E)** Shown is T485 located in the end of the NOPS domain, that caps the coiled-coil domain. **F)** EMSA of T485 mutations and quantification. **G)** R474 mutation located in the NOPS domain makes several contacts with surrounding residues. **H)** Ems of R474A and quantification (right). Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

Threonine-485 to alanine mutation decreases affinity for RNA

Mutating threonine-485 (T485) decreases affinity for RNA, which occurred whether we mutated to alanine or the phosphomimetic T485D (Fig. 3.2F). All mutants were stable and

folded in solution at working temperatures, albeit with a modest decrease in melting temperature compared to wildtype protein as seen by the left-shift in mutant melt curves (Fig. 3.3)(Baljinnyam et al., 2020). The melt curves are biphasic, which could represent dimer dissociation before protein unfolding. T485A shifts equilibrium back towards monomeric state, although the phosphomimetic T485D does not affect dimerization as measured by Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) (Fig 3.4) (Some et al., 2019). T485 is spatially close to histidine residues recently shown to coordinate zinc (Zn), and high levels of Zn lead to cytoplasmic aggregation of PSF (J. Huang et al., 2020). Mutating T485 may affect Zn coordination.

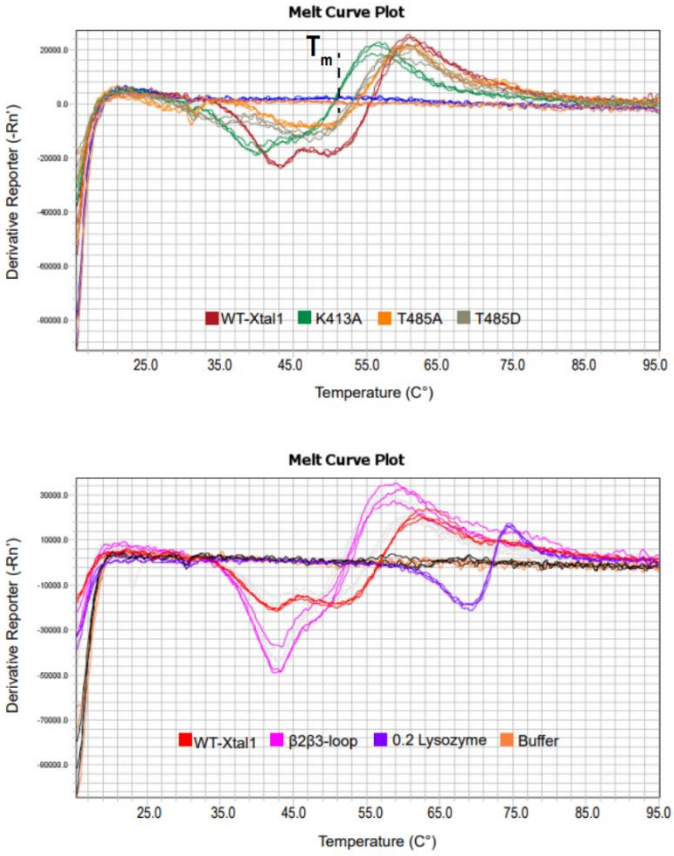


Figure 3.3 Protein stability by thermal melt. Mutant stability was tested using differential scanning fluorimetry (DSF). The melting temperature (T_m) is temperature at which 50% of protein in solution is unfolded, which corresponds to the inflection point on the curve. While some melt curves experience a shift, all mutants remain folded at binding condition temperature of 30°C. Lysozome is used as a positive control for the melt. Melting temperatures shown in table below.

Protein	Melting Temperature (°C)
WT	55.5
K413A	51
T485A	54
T485D	54
β2β3-loop	51
Lysozyme	73

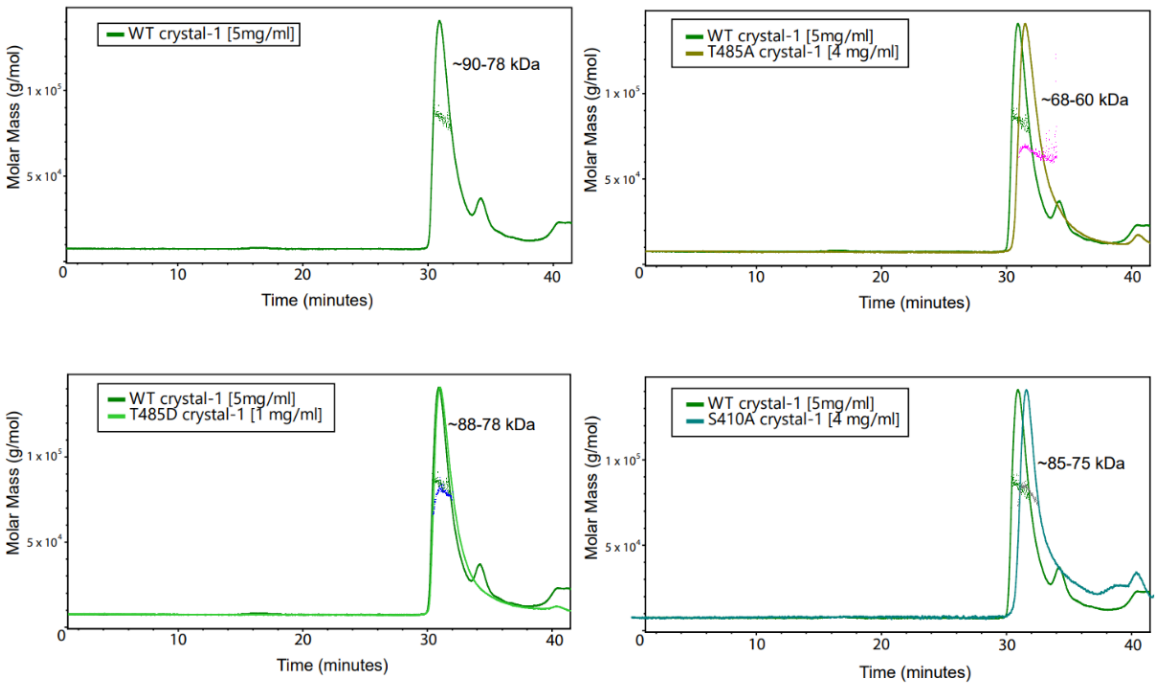


Figure 3.4 Size Exclusion Chromatography with Multi-Angle Light Scattering (SECMALS). Dimerization and overall shape/size of proteins were measured using SECMALS. Mutants T485A and S410A remain dimerized like wildtype. T485A shifts equilibrium towards monomeric state. Both S410A and T485A show a shift in overall shape/conformation, as indicated by the shift in retention time**.**

Arginine-474 to alanine mutation decreases affinity for RNA

Mutating a residue in the NOPS domain, Arginine-474 (R474), also resulted in a loss of affinity for RNA but more data is needed characterizing the stability and dimerization state of this mutant (Fig. 3.2H).

β2β3-loop gains affinity for RNA upon mutation to "AAAAA"

Mutating the β2β3-loop to all alanine increase PSF's affinity for ESS-RNA 10-fold (Fig. 3.2D). We made stepwise mutations to this loop, mutating either the basic or acidic

residues to alanine. Interestingly, the acidic to alanine mutation (AARGR) was not soluble and could not be purified. The β2β3-loop mutant is stable and folded in solution (Fig. 3.3). However, the β2β3-loop mutant may have some conformational changes that occur. Preliminary HDX-MS data shows the β2β3-loop mutant is folded but appears to be more dynamic in solution compared to wild-type protein (Fig. 3.5). RRM2 exchanges

Figure 3.5 Hydrogen-Deuterium exchange data of β2β3-loop mutant. Exchange data of PSF(WT) and PSF(β2β3) collected over time. Peptides are continued on the next two pages.

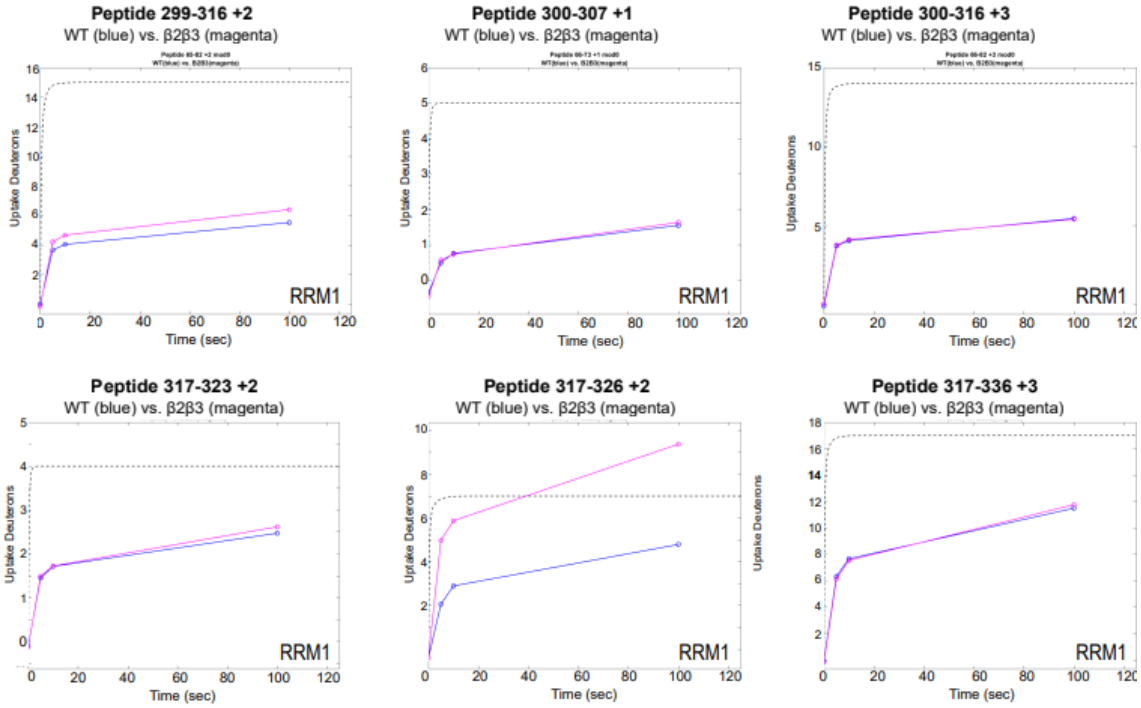


Figure 3.5 Hydrogen-Deuterium exchange data of β2β3-loop mutant. (cont'd)

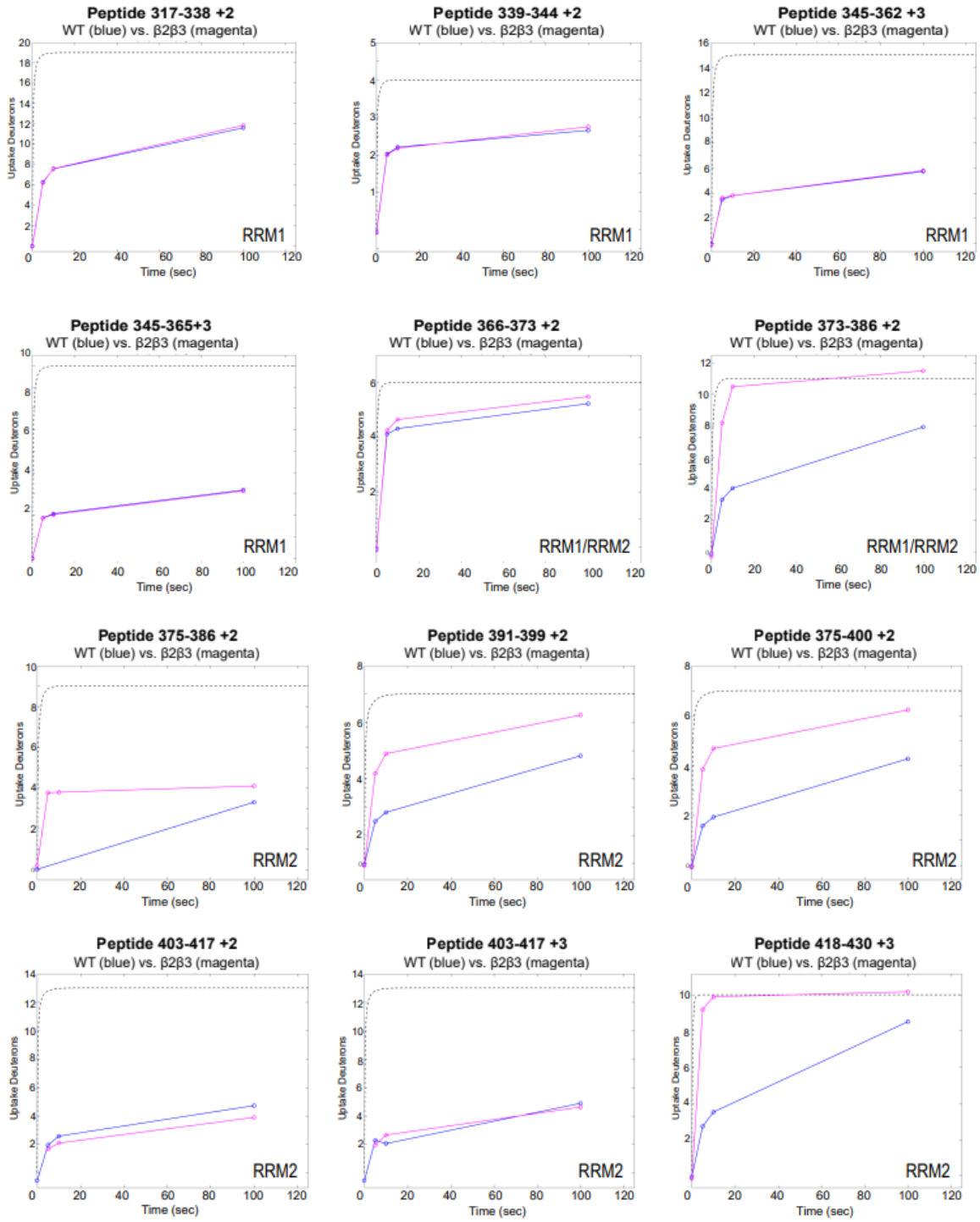
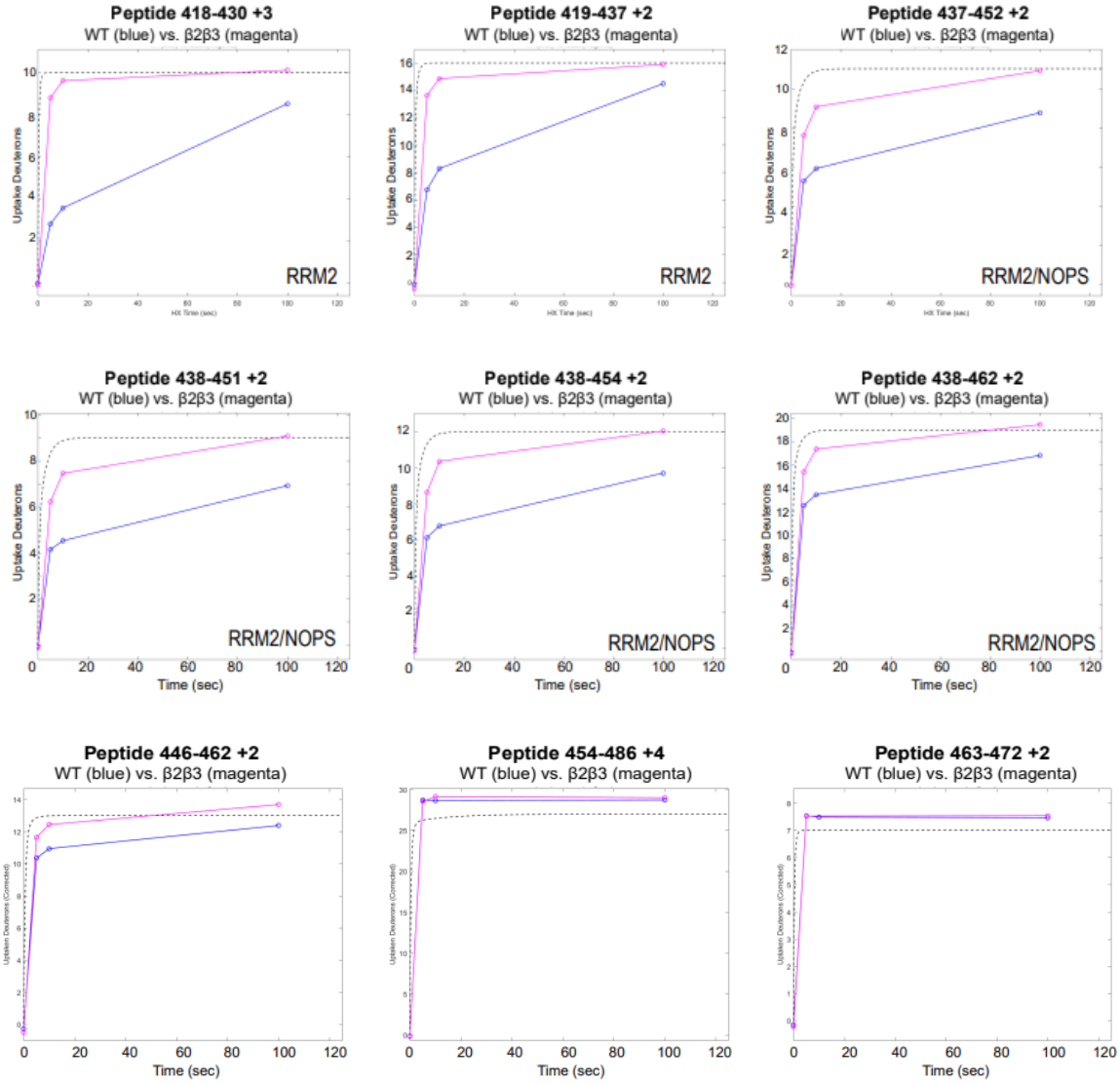


Figure 3.5 Hydrogen-Deuterium exchange data of β2β3-loop mutant. (cont'd)



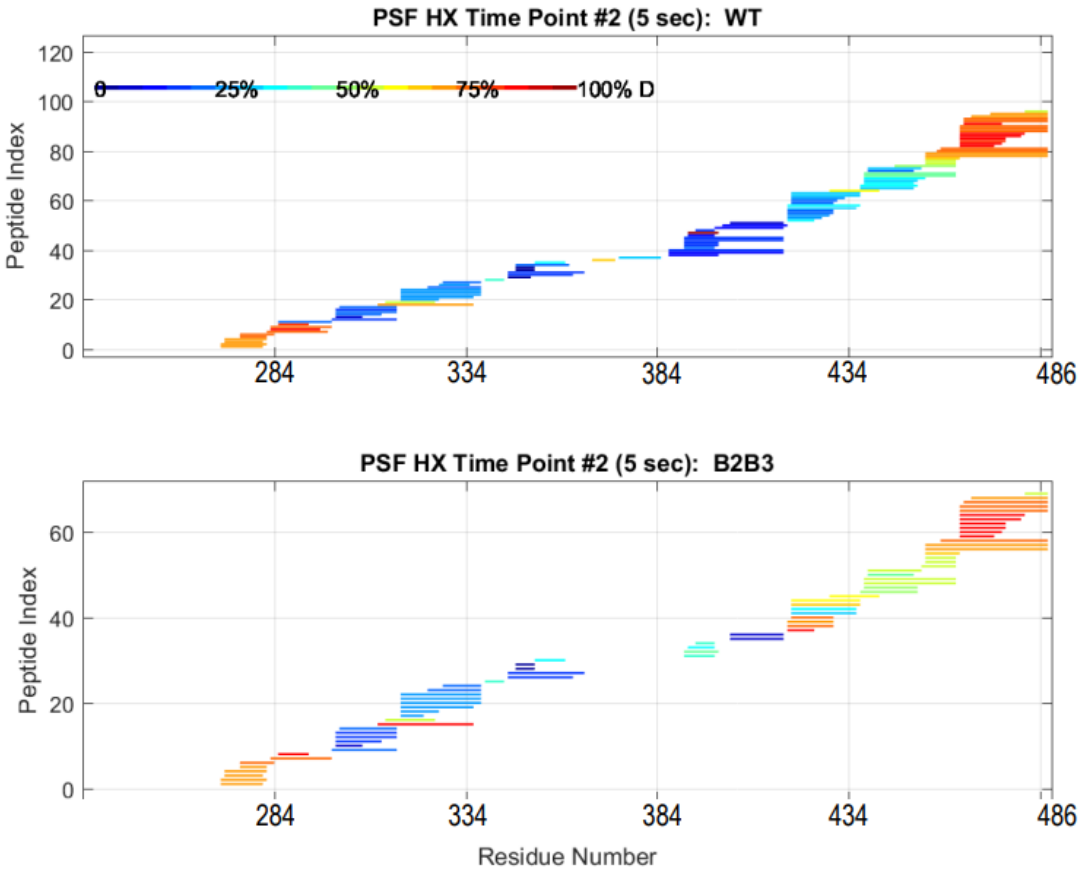


Figure 3.6 Heat map of HDX-MS WT vs. β2β3-loop mutant. Heat map shows percent deuteration of peptides at 5-seconds in D_2O. The NOPS and RRM2 domains of the β2β3-loop exchange more rapidly than WT. RRM1 exchange differences negligible. The β2β3-loop has low coverage and peptide count.

more rapidly in the β2β3-loop mutant, but differential exchange is not seen in RRM1. This may indicate the RRM2/NOPS region are more flexible and/or experience more breathing motions than wildtype protein. However, this experiment needs to be repeated to make further inferences about the loop-mutants dynamics as coverage and overall peptide count were poor (Fig. 3.6). Additionally, much of the NOPS domain exchanges too rapidly to see differences and slowing exchange to capture differences in more rapid changes would be ideal.

K413A increases affinity for ESS-RNA

Preliminary data suggest an increased affinity for ESS-RNA when mutating K413 to alanine (Fig. 3.7). However, this mutant was more likely to precipitate while purifying and resulted in low yield. K413 was present in several of the biophysical characterizations in Chapter 2 and would be of interest to further study biophysically if homogenous protein can be purified and remain soluble.

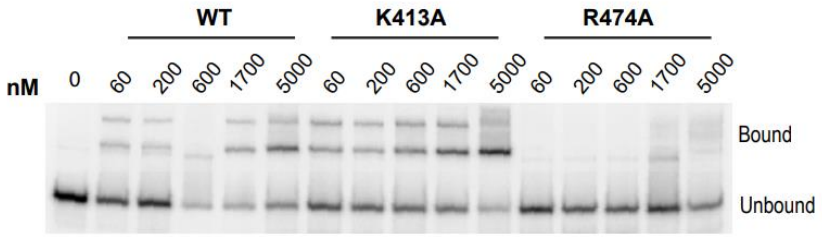


Figure 3.7 K413A mutant increases binding, R474A decreases binding. Electromobility shift assay with radiolabeled ESS-RNA and mutant recombinant protein. *The 600 nM WT lane is an outlier and likely due to pipetting error or sample evaporation.

Binding modality shared between multiple target RNA

To test if binding modality is shared for other RNA, I performed the same binding assays with some of PSF's other known target RNA vault RNAs (vtRNAs) 1-1 and 1-2. I observed similar effects on binding with vtRNA1-1 and vtRNA1-2 (Fig. 3.8). For example, mutations increasing affinity for ESS-RNA also increased affinity for vtRNAs. Therefore, PSF's mechanism of RNA-binding is likely shared between all three RNA.

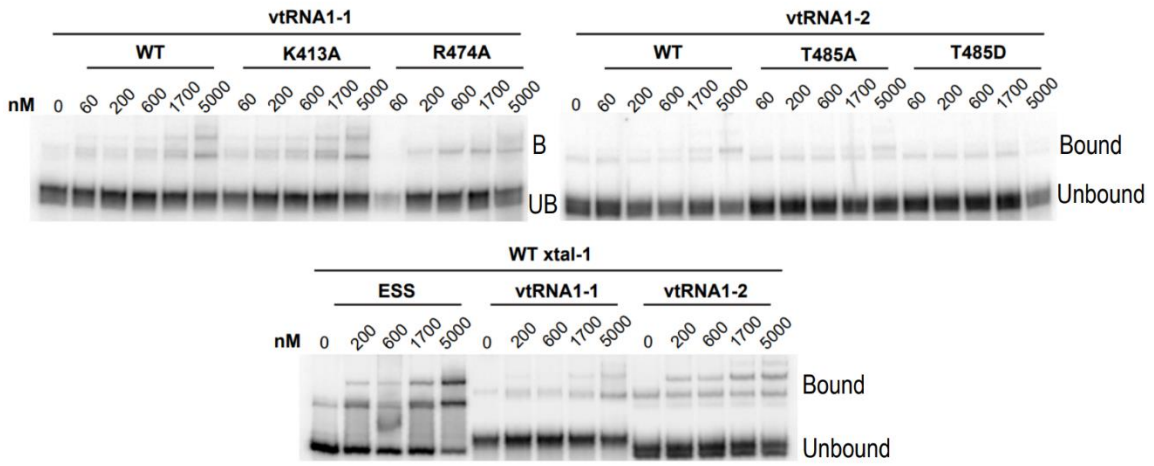


Figure 3.8 Binding analysis of mutations with vault RNAs (vtRNAs). Shown are two vault RNAs vtRNA1-1 and 1-2 in binding reactions with PSF. Electromobility shift assays show same effect on binding upon mutating K413A, R474A, T485A, and T485D

DISCUSSION

Here I show several residues occurring in the peptide data from Chapter 2 had no effect on RNA-binding, but a few did affect binding. Residues T485 and R474 lost affinity for RNA following mutation to alanine. Mutating β2β3-loop and K413 to alanine increased affinity for RNA. RNA-binding modality is shared among three different structured RNA.

Threonine-485 caps the extended coiled-coil domain N-terminally, and this coiled-coil domain mediates oligomerization (Lee et al., 2015). T485 may play a role in orienting this domain for RNA recognition or for proper conformation. Additionally, it was discovered that reversible aggregation of PSF is driven by the coordination of zinc ions. Coordination of zinc utilizes several residues spatially and sequentially near T485, namely H483, H528, and H530 (J. Huang et al., 2020). It is possible that phosphorylation of T485 induces local

changes that disrupt the orientation or charge of H483, hence disrupting coordination of zinc and subsequent aggregation.

Functional aggregation may play an important role in RNA-binding. In cells, PSF is one of the main components of paraspeckles, subnuclear membraneless nuclear organelles built around the non-coding RNA NEAT-1. RRM2 deletion disrupts paraspeckle formation, and given RRM2 is necessary and sufficient for binding, functional aggregation of PSF likely influences RNA-binding. Furthermore, phosphorylation of PSF may drive regulation in this fashion. PSF phosphorylation could cause conformational change to inhibit RNA binding, in turn destabilizing the oligomeric state to reduce functional aggregation and free up soluble PSF to participate in other complexes in the nucleus.

Argininie-474 was important for RNA-binding. This residue may be making direct contacts with RNA or perhaps R474 is vital for holding RRM2 in the right conformation to bind RNA. It is not clear, but R474A does make several contacts with neighboring residues, disruption of which could easily change structure in that area (Fig. 3.2G). Currently, it is unknown how this mutation affects stability and dimerization state of PSF.

The β2β3-loop is utilized to bind to stem-loop RNA by four different proteins: U1A, U170K, RBMY, and RBFOX1 (Y. Chen et al., 2016; Kondo et al., 2015; Oubridge et al., 1994; Skrisovska et al., 2007). PSF's β2β3-loop also appears important to recognize stem-loop/structured RNA. Mutating the β2β3-loop to all alanine and mutating K413A increased affinity of PSF for RNA. It is possible that the β2β3-loop acts as a gate to the gap in electron density between the two copies of RRM2 and mutating the charged loop to all alanine increases flexibility of the loop and surrounding β-strands. Additionally, increased dynamics may allow PSF to sample more conformations to find the bound state

more readily (Fig. 3.5 & 3.6). However, this data is preliminary and needs to be reproduced before inferring further.

As mentioned earlier, mutation of the loop to AARGR was not soluble. It unsurprising that disruption charge in area affected protein stability, as it is highly conserved. It is curious that the reciprocal Δbasic mutation (DDAGA) expressed and purified well despite alteration in charge. While the protein could not tolerate a positive charge in this region, it tolerates a negative charge well. Perhaps this region might also tolerate the negative charge of an RNA partner.

Finally, I show the effects on binding upon mutation were seen in three different RNA: ESS, vtRNA1-1, and vtRNA1-2. This indicates the modality of binding is shared between PSF and other RNA.

While there is still much to learn about PSF's RNA-binding ability, one thing is clear: PSF's RRM/RNA binding modality is unique and worth further study. In the next chapter I discuss PSF's role in alternative polyadenylation in T-cells.

CHAPTER 4: PSF's role in alternative polyadenylation in T-cells

INTRODUCTION

In eukaryotic cells, RNA is transcribed by one of three RNA polymerase, and further processed into the mature messenger RNA (mRNA). Protein coding mRNA is transcribed by RNA polymerase II (Pol II), cleaved at the 3' end, followed by addition of a 3' polyadenine (polyA) tail – a process referred to as cleavage and polyadenylation. Typically, polyA tails are 100-250 nucleotides long, but can be up to several kilobases in length. Alternative polyadenylation (APA) occurs when a gene containing multiple polyadenylation signal (PAS) sites gives rise to distinct mRNA transcripts with different 3' untranslated regions (UTRs). The 3'UTR contains sequences known as regulatory elements that play an important role in gene regulation. Sequences in the 3' UTRs can be bound by miRNAs, siRNAs, and RNA-Binding proteins (RBPs) to affect transcript translation, localization, stability, and nuclear export. Alternative polyadenylation is one mechanism the cell uses to aid in cell fate, response to stimuli, and when dysregulated can play a role in disease.

Here I present computational analysis of APA in three different cell types. Matthew Gazzara analyzed an RNA sequencing (RNA-seq) data-set from Jurkat T cells (JSL1) in the context of PSF knockdown using the DaPars algorithm (**D**ynamic **a**nalyses of **A**lternative **P**oly**A**denylation from **R**NA-**S**eq) (Gazzara et al., 2017; Xia et al., 2014). While I had data from stimulated and unstimulated cells, I focused on the stimulated data set. I was interested in PSF's contribution to alternative polyadenylation. In resting T cells, PSF is phosphorylated and bound by TRAP150, which occludes RNA-binding. In stimulated cells, PSF is hypophosphorylated and can bind RNA. In addition to this set of targets,

Matthew Gazzara conducted DaPars analysis on the publicly available ENCODE (www.encodeproject.org) data sets, both K562 cells and HepG2, in the context of PSF knockdown.

RESULTS

Determining PSF responsive alternative polyadenylation targets

Following DaPars analysis of the RNA-seq data, I filtered for changes in PAS usage with a difference of greater than 20%, a threshold based on recent work by Rakesh Chatrikhi looking at CELF2's regulation of polyadenylation in the same Jurkat T cell model system (Chatrikhi et al., 2019). In the stimulated data set, there were 371 events with greater than 20% change, 1245 events in K562 cells, and 591 events in HepG2 cells (Fig. 4.1). Next, I looked for overlapping events between the three cell types for potential targets to validate. I found 17 events shared between JSl1 and HepG2, 37 events between JSl1 and K562, and between the two encode data sets there were 138 shared events (Fig. 4.1). Four genes underwent APA changes following PSF knockdown in all cell lines: SUGP2, CCDC58, SIK2, PDCD6IP. Additionally, I added targets with large changes in PAS usage upon stimulation in the PSF KD condition. I looked for alternate PAS usage that created 3' UTR differences within 1-2 kb maximum. These targets are better suited for the validation technique 3'-RACE. I began my initial validation with 6 targets and designed primers in the terminal exons of the genes, seen in Table 4.1.

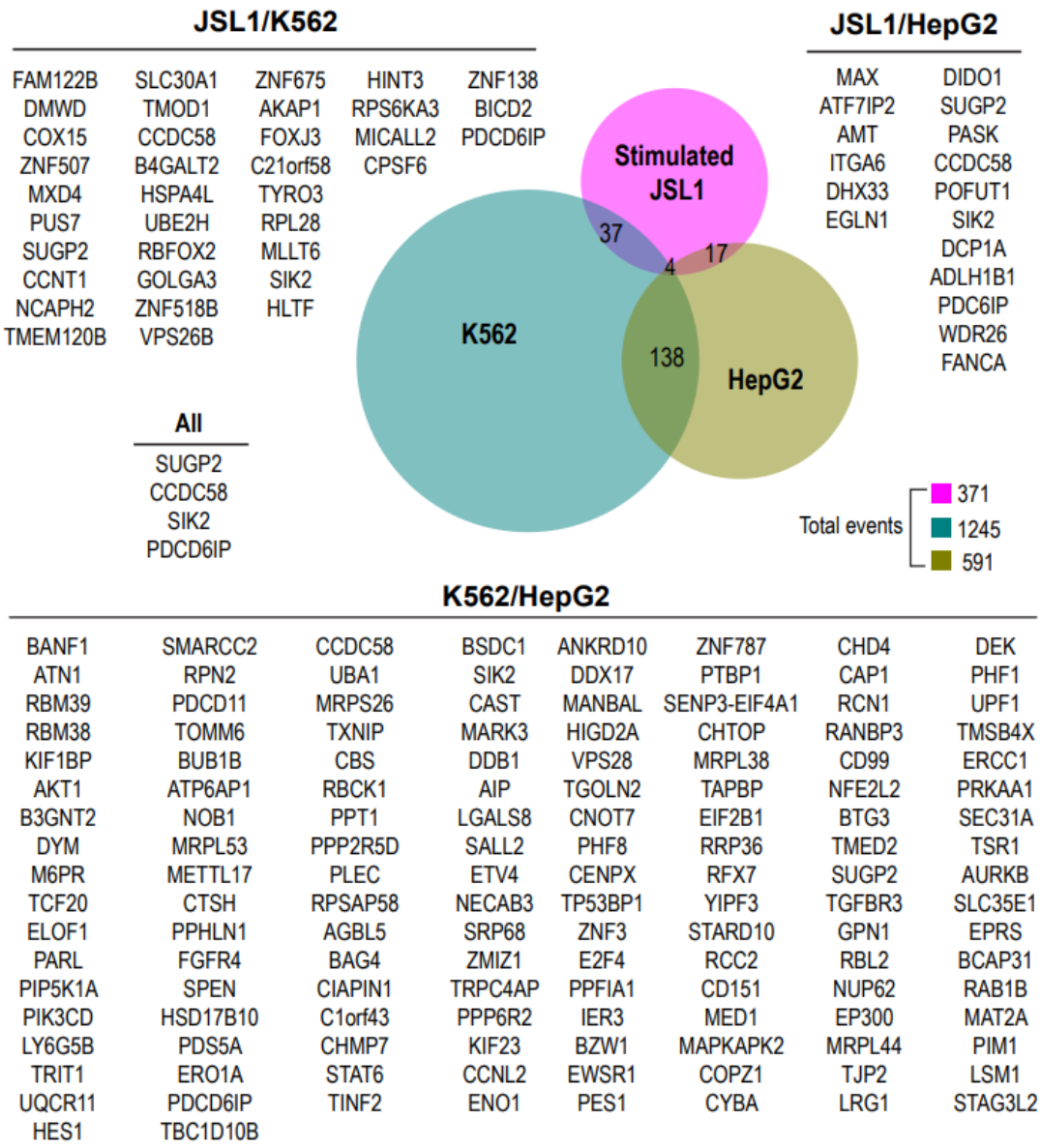


Figure 4.1 Alternative polyadenylation events using DaPars upon PSF knockdown. RNA-sequence data was analyzed from cell lines knocking down PSF with the DaPars algorithm. KD PSF Jurkat T cells (JSL1) were stimulated using PMA, and 371 APA changes occurred. Data from K562 & HepG2 cell lines come from ENCODE data, and we report 1245 and 591 APA changes respectively. Four genes experienced APA changes following PSF knockdown across all cell lines: SUGP2, CCDC58, SIK2, PDCD6IP. Genes with APA changes were overlapped in the three cell types, those gene names are shown above.

Table 4.1 Gene specific primers for 3'RACE validation.

Target Gene	Sequence	Primer name
ZNF814	5'-CACCAGAGTTCACACTGGAGAAAAGCC-3'	ZNF814 For 2 distal
ZNF814	5'-CAGGTGTGTCTCCCAAGAAGGCCC-3'	ZNF814 For 1 proximal
FAM18B/TVP23C	5'-TATTTCTCTCCTTGCGGCTTTGGGCC-3'	TVP23C for
RECQL5	5'-GCTTGGTGTAGTCAGGTTGTGTCCAGGC-	RECQL5 E9 forward
PASK	5'-GCAGCTTTCTTCAAGTCGCTCTTTAGCC	PASK 002 F RACE
TMOD1	5'-CGGAGTGGTGTCTAGTGTGTGGCGG-3'	TMOD1 F RACE
CCDC58	5'-AATGAACGCTGCCGAATTCACTTCAAGCC-	CCDC58 RACE FOR

Validating alternative polyadenylation events in Jurkat T cells by 3'-rapid amplification of cDNA ends (3'-RACE)

Reads were first visually confirmed on the genome browser to check where the DaPars algorithm called proximal and distal PAS sites. In many cases PAS designation is clear, as is shown in Figure 4.2A (note direction of gene transcription). Based on the RNA-seq reads, wildtype resting cells have a significant amount of distal PAS usage. When PSF is knocked down, distal usage decreases. In stimulated cells, WT cells use predominately the proximal PAS, and upon PSF KD distal PAS site usage increases (Fig. 4.1A). Next to validate these APA changes I used 3'-rapid amplification of cDNA ends (3'-RACE) with gene specific radiolabeled primers to visualize and resolve these APA changes (Fig 4.2B). RNA was generously provided by Michael Mallory, harvested from JSL1 cells using the four conditions: unstimulated, unstimulated + PSF knockdown (KD), stimulated, stimulated + KD. While TVP23C did not validate, RECQL5 did, and ZNF814 likely requires a different set of primers (Fig 4.2B).

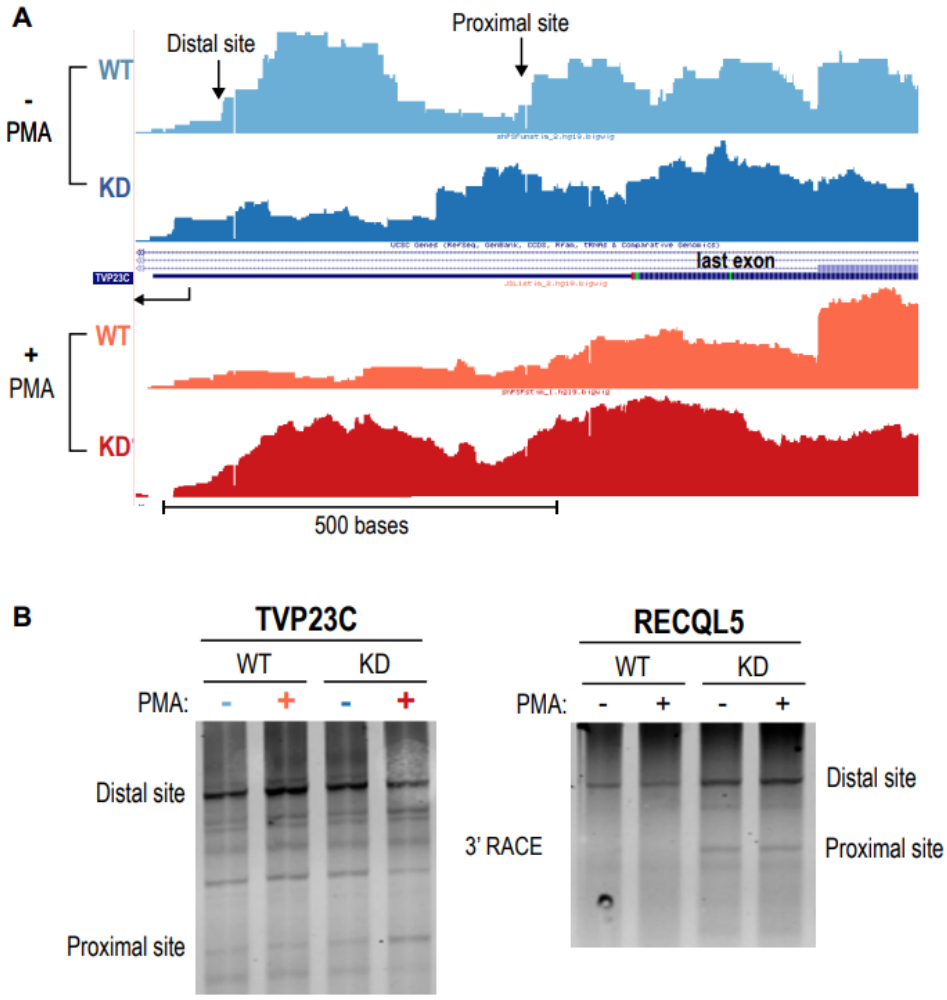


Figure 4.2 Alternative Polyadenylation site usage in Jurat T cells. A) RNA-sequencing read tracks from four conditions of TVP23C gene (Stimulated/unstimulated, WT/KD PSF). Black leftward pointing arrow indicate the direction of transcription. Distal and proximal sites indicated by DaPars are labeled. **B)** Polyacrylamide gel of 3'RACE products.

DISCUSSION

Alternative polyadenylation is another important step in expanding the proteome and regulating expression of the genome. PSF is a multifunctional RNA-binding protein with many roles in nucleic acid processing. Here I show that knocking down PSF alters polyadenylation in three separate cell types. This project has a clear path forward which should begin with working through validation of some of these targets, followed by conducting a motif analysis on the aggregate to see if any sequence patterns are established. Chatrikhi et al., 2019 serves as a great model for the future of this project, as he detailed CELF2-dependent APA changes in Jurkat T cells.

CHAPTER 5: Conformational dynamics of PSF

INTRODUCTION

In T cells, binding of PSF to the ESS sequence within exon 4 of CD45 is disrupted by a binding partner protein called TRAP150. In resting T cells, PSF is phosphorylated on C-terminal residue threonine-687 (T687) by GSK3, and in this phospho-state PSF favors binding to TRAP150. As mentioned in previous chapters, binding of PSF/TRAP150 and PSF/ESS-RNA is mutually exclusive. Preliminary data suggests phosphorylation of PSF's C-terminus regulates its interactions with TRAP150 through an allosteric mechanism as: (1) the C-terminus of PSF is not directly involved in the interaction with TRAP150, (2) we have observed differences in limited proteolysis between phosphorylated and dephosphorylated PSF and (3) pulldown data shows that PSFs C-terminus interacts with the exRRMs of PSF in *trans*. Therefore, we hypothesized phosphorylation induces a conformational change making PSF more suited to bind TRAP150 than RNA.

Building our understanding of the PSF/RNA interactions requires understanding the binding mechanisms of PSF/TRAP150, as well as understanding how phosphorylation effects structural conformation. To that end, using a longer construct PSFΔ265 I *in vitro* phosphorylate PSF, and look for dynamics/structural changes via HDX-MS. Additionally, I detail preliminary data looking at PSF(exRRMs) with the portion of TRAP150 known to interact with PSF called the PSF-interaction domain (PID).

Data for this section are inconclusive, but the preliminary data are promising. With further troubleshooting, future experiments detailing phosphorylation-induced conformational differences of PSF with HDX-MS are possible. The same is true for studying the interactions between PSF/TRAP with HDX-MS.

RESULTS

Protein Purification of PSF Δ265

For conformational studies I used a longer construct of PSF containing residues 266-707 (Fig. 5.1A). While difficult to purify and keep soluble, PSF**Δ**265 is still preferable to working with full length protein. Importantly, constructs containing the C-terminal have target residues for phosphorylation by GSK3. I purified the protein construct as described in Materials and Methods and Yarosh et. Al 2015. However, I observed a large portion of PSF**Δ**265 in the insoluble fraction (Fig. 5.1B, pellet). This is to be expected with a functionally aggregating protein, and either increasing salt concentrations or attempting purification from inclusion bodies would increase yields of the preparation.

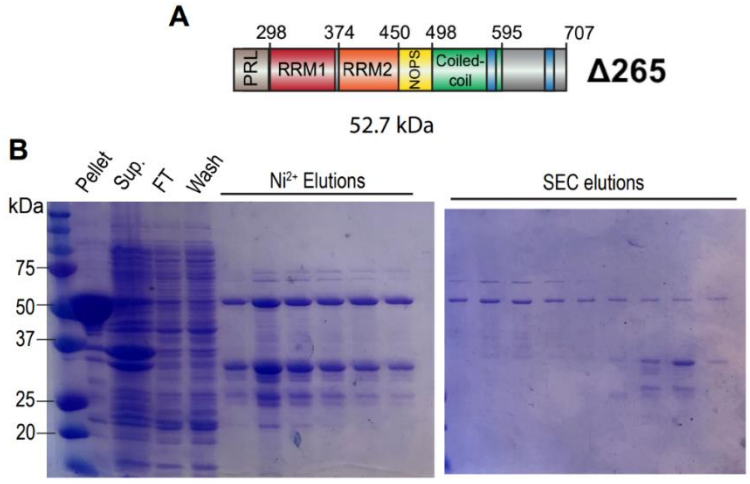


Figure 5.1 Purification of PSFΔ265 construct. A) PSF**Δ**265 contains residues 266-707 and an N-terminal 6His-tag. **B)** Coomasie stain of SDS-PAGE gel purifying **Δ**265 in a two-step purification. First affinity purification was done over a charged Ni2+ column, followed by gel filtration over a Superdex 200 column.

Conformation of PSF following *in vitro* phosphorylation by GSK3

The issue with this data set is the lack of peptides. Compared to the peptide index (y-axis) to other HDX data set, there is an ~66% reduction in peptides (Fig 5.2A & Fig. 2.7C). Part

of the problem stems from the low salt concentrations required for the in vitro phosphorylation reaction. PSF requires salt concentration of ~300 mM to remain soluble.

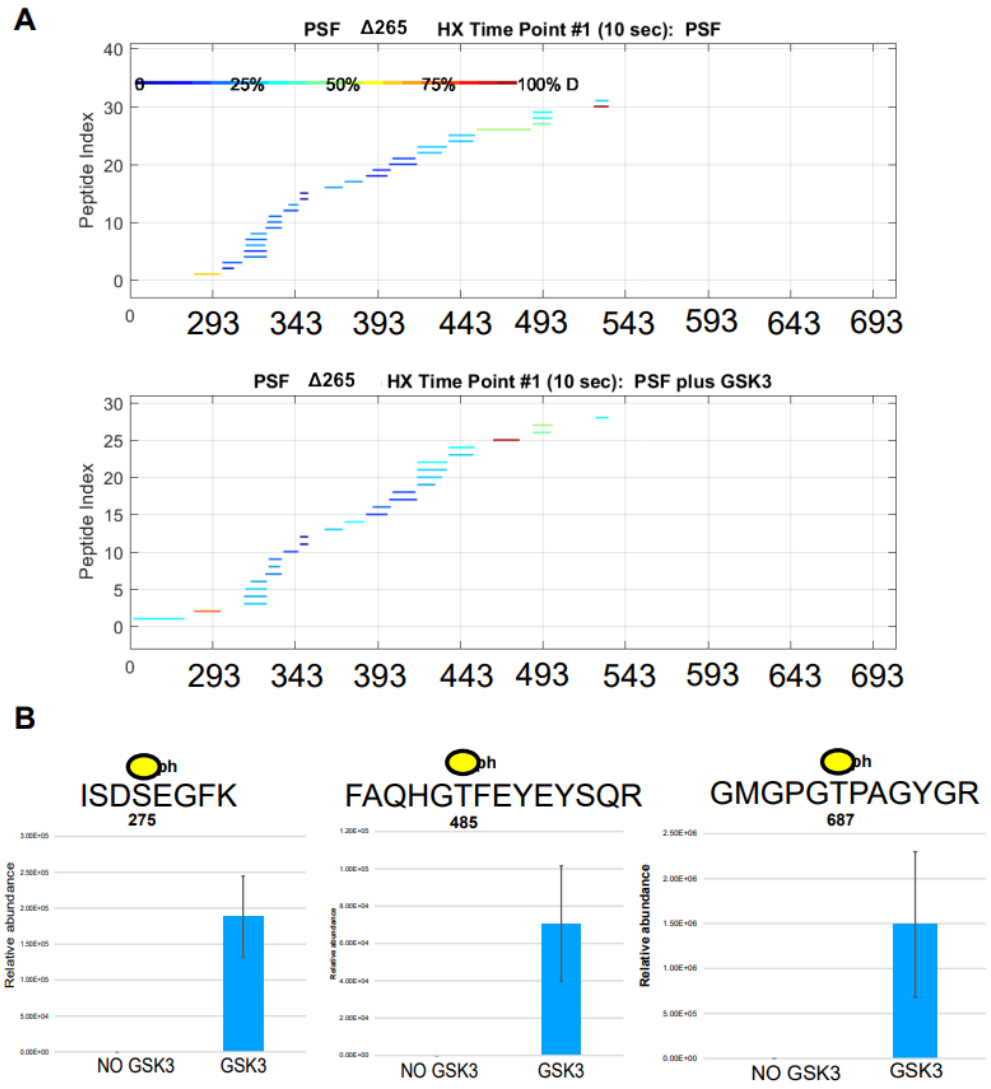


Figure 5.2 Phosphorylation-induced dynamics measured by HDX-MS. A) Peptides from H-D exchange of *in vitro* phosphorylated PSF B) Peptides found phosphorylated by GSK3 in vitro.

However, GSK3 is more active at low salt (100 mM) concentrations. It is possible some of the protein precipitated out solution, which could explain the low peptide signal in the

exchange data. Very few peptides covering the last 200 amino acids of the protein were observed (Fig. 5.2A). I confirmed the protein sequence could be easily digestible by pepsin, which I typically use for my HDX-MS experiments. The SEC elutions of the PSF(Δ265) construct were more heterogenous than other PSF constructs. While the dominant species is the desired size, larger and smaller species are noted (Fig. 5.1B). A significant amount of protein at a lower species (~30 kDa) can be seen. It's possible this is a truncated/early translation termination product. This is one explanation for why I am not seeing much coverage in the C-terminal end of the construct.

I simultaneously prepared samples from the *in vitro* phosphorylation reaction for mass spec analysis. This confirmed *in vitro* phosphorylation worked by the presence of three phosphorylated residues in the samples with GSK3: S235, T485, and T687 (Fig. 5.2B). The phosphorylation abundance mass spec processing and analysis was conducted by Mariel Mendoza at UPenn in Ben Garcia's laboratory.

PSF/TRAP150-PID dynamics measured by HDX-MS

In analyzing the dataset of PSF/TRAP150 interface I faced some issues. Due to low peptide coverage and high noise, the data is preliminary and must be repeated. In HX, to amplify signal we push equilibrium to the bound state by adding ligand in excess. However, when the ligand is another full protein, it reduces the number of peptides captured and analyzed from the protein of interest (PSF) because ligand peptides (TRAP150) are also coming across the detector, reducing the signal-to-noise ratio. While I mapped the results for visualization, these results need to be repeated (Fig. 5.3).

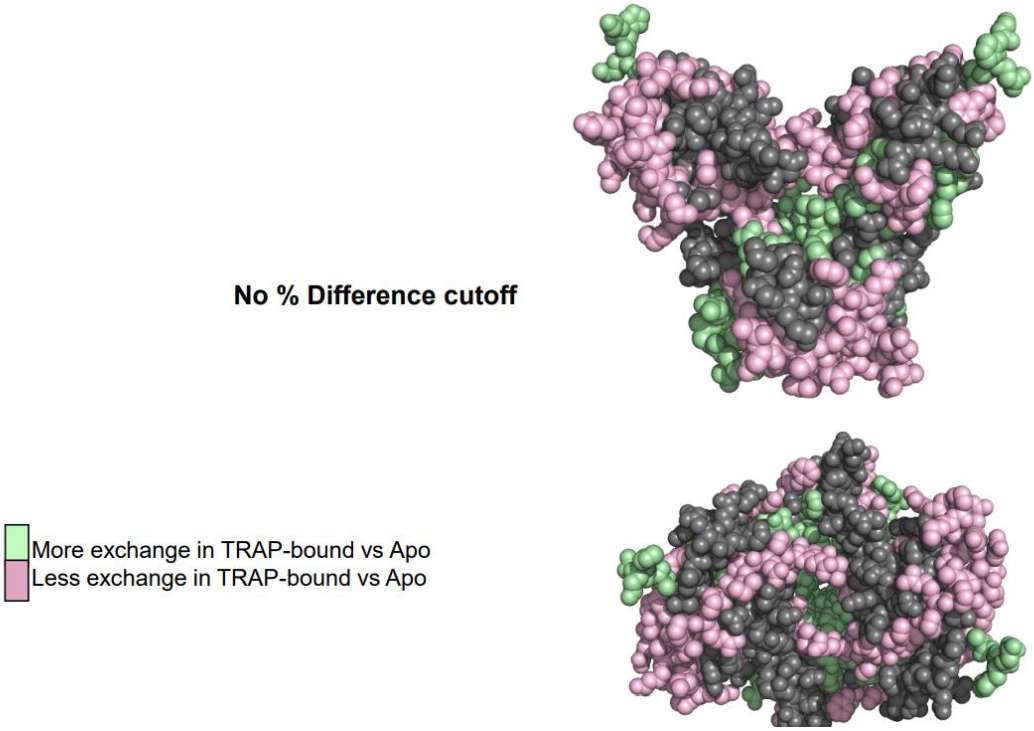


Figure 5.3 Preliminary ^{1}H-^{2}D exchange data mapped onto the homodimer of PSF's exRRMs. This maps the differential of PSF(crystal-1)exRRMs^{WT} alone vs PSF(crystal-1)exRRMs^{WT} bound to the TRAP150-PID Models made using the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.

DISCUSSION

Several lines of data indicate PSF likely undergoes conformational change dependent on phosphorylation. This conformational change likely shifts PSFs ability to bind certain binding partners and is a point of regulation. Understanding these conformations can inform the mechanism of mutual-exclusive binding with PSF/TRAP150 and PSF/ESS-RNA. Although data from this section are inconclusive, these experiments merit further considerations. Future experiments looking at phosphorylation-induced conformational differences of PSF by HDX-MS are possible. The same is true for studying the interactions between PSF/TRAP with HDX-MS.

CHAPTER 6: Conclusions & Future Directions

PSF is an interesting protein whose mechanism of binding and regulation have been difficult to pinpoint. PSF is an essential protein and knockout of PSF is lethal in embryo. RNA-binding analysis of PSF has produced a series of consensus motifs with little overlap or agreement between techniques. While I uncovered a great deal of information about PSF's structure, dynamics, and RNA-interaction surface, much remains unknown.

In Chapter 2 I characterize PSF biophysically data showing an expansive RNA-binding interface. I show where RNA crosslinks to PSF using a cross-linking mass spectrometry technique called RBR-ID. *In vitro* RBR-ID with recombinant PSF(exRRMs) and 4-SU-labeled ESS-RNA revealed peptides in the NOPS/RRM2 domain that interact with RNA. RRM1 interacted to a lesser degree, although RRM1 RNP containing peptides did cross-link. Next, I use a complementary approach to identify residues of PSF that experience chemical perturbations when RNA is present.

Recognizing the caveats to both RBR-ID and HDX-MS, I paired these techniques in tandem to confirm the data in each experiment. HDX-MS data share many overlapping peptides with the RBR-ID data. Importantly, exchange data does not necessarily indicate binding, but instead indicates a change in chemical environment that could also be explained by a conformational change to the protein structure upon binding, such as an induced fit model. For example, upon binding stem-loop RNA FUS shifts local conformation by using loops α1/β2 and β2/β3 to facilitate binding of its target RNA (Basu et al., 2021). If PSF binds stem-loop RNA in the same manner, this conformational change upon binding could explain some of the differential exchange data.

I present NMR data by our collaborator Nishtha Galati in Michael Sattler's lab on RRM1/RNA. Despite the size challenges that come with NMR, we wanted to try and get solution structures for the individual RRMs with RNA. We also used the 43-mer RNA to cut down on size. There were issues purifying protein constructs, but the data that was acquired with RRM1/ESS 43-mer showed weak interactions across RRM1's β-sheet surface, but not with the aromatic residues (Fig. 2.12). I would however expect the interactions to be weak as RRM1 is not the main driver of RNA-binding in PSF. This smaller RNA also loses significant affinity for PSF, and the interaction between PSF and the 65-mer isn't strong (PSF(crystal-1) construct/ESS-RNA 65-mer kD_{app} ~1.5 μM).

Finally, I further characterize PSF using molecular dynamics with the help of Nick Rego. Ensemble-simulations identify the areas of PSF's surface that most readily lose water or "dewet." As an unfavorable biasing potential is applied water molecules are stripped from the protein hydration shell, to provide information about PSFs hydrophobic/hydrophilic surfaces. Although the β2β3-loop is made from polar residues it behaves hydrophobically. The charge balance in this loop is vital for protein stability in solution, as was determined by the "DDRGR" to "AARGR" mutation that would not maintain solubility, despite several attempts of altering salt concentrations. Some areas determined to be hydrophobic patches by the ensemble simulations are to be expected, as is the case of RRM1's β-sheet surface, whose aromatic residues make a hydrophobic patch.

In Chapter 3, I outline a series of mutations designed, cloned, and purified to test the relevance of peptide data acquired in chapter two. Mutational design was aided by all

the biophysical data sets acquired and described in Chapter 2, homology, and inference based on comparison to other solved RRM/RNA models in the literature. Several other mutations have been made previously and can be seen in Chris Yarosh's thesis, my predecessor on this project. Those results also helped guide my design.

Minimal interacting domain data of PSF/ESS-RNA shows binding requires RRM2/partial NOPS domains. Thus, I focused my attention to residues in these domains. These domains were also among the most probable to bind RNA in my biophysical data. I made several mutations in and around the β2β3-loop, making mutations to residues in both strands, as well as sequential mutations to the loop itself. Some residues had no effect on binding whereas mutating the β2β3-loop or K413A increased affinity for RNA. It is possible that increase dynamics of the β2β3-loop mutant is aiding in binding by allowing the protein to more rapidly sample through conformations to bind the RNA target or possibly to mold with its RNA partner. Initial HDX-MS data show increased dynamics in RRM2/NOPS (Fig. 3.5). RBMY, U1A, U170K, and RBFOX1 all bind stem-loops [partially] by inserting their loops into RNA to help bind (Basu et al., 2021). PSF has been found to bind so many stem-loop /structured RNA, it seems like a potential mechanism for PSF to use its β2β3-loop to help recognize/bind/interact with RNA as well.

Additionally, I mutate two residues, R474A and T485A/T485D, in the NOPS domain decrease binding, suggesting a role for these residues in binding/interacting.

In Chapter 4, I briefly detail the beginnings of an alternative polyadenylation project with a large foundation of APA changes following PSF knockdown (Fig. 4.1). Matthew Gazzara performed the APA predicting algorithm DaPars on a previously obtained RNA-sequencing dataset, and we found hundreds of APA events occur upon PSF knockdown in JSL cells, as well as in the publicly available ENCODE data sets. I show one method of validating APA targets (3'-RACE) (Fig. 4.2). This project has would be an interesting future

direction for incoming lab members. Validating a series of the APA events discovered in using the DaPars algorithm using 3'-RACE is a clear path forward. While no clear gene ontology classes were indicated in the set of genes from APA data sets, there are several targets with large APA changes and potential for interesting biology.

Finally, in Chapter 5, I describe experiments which aim to inform on the PSF/TRAP150/ESS-RNA system and interplay. *In vitro* phosphorylation by GSK3 showed phosphorylation on S235, T485, and T687. Unfortunately, due to low peptide coverage, HDX-MS experiments don't yield much usable information and should be troubleshooted. HDX-MS experiments looking at PSF(exRRMs) bound to TRAP150(PID) should also be repeated due to low peptide coverage and high exchange noise. While data are preliminary, the experiments merit future efforts.

A way to push biophysical characterization of PSF forward is to focus on a few constructs purified from SF9 cells. PSF is post-translationally modified on many residues, not to mention on two vital threonine residues for the mechanisms I've previously described. Purifying a version of PSF post-translationally modified in cells may be more soluble and less prone to aggregation. This could reduce some of the noise that makes biophysical characterization difficult. It would be helpful for future studies regarding PSFs phosphorylation-induced conformational changes to have a construct which can be easily phosphorylated/dephosphorylated. Protein based HDX-MS experiments are relatively simple if the proteins are well behaved in solution. Thus, a new construct would provide data from protein in a more native state acquired relatively easily and reliably. Additionally, co-crystallization of PSFs with smaller RNA would yield the most resolute information. Screening a library of small RNA probes for binding in a high-thoughput fashion may yield a PSF/RNA pair smaller and more homogenous for co-crystallization. PSF is a vital,

unique, and elusive protein, whose further study presents an enticing challenge for a curious and ambitious mind.

CHAPTER 7: Materials and Methods

Protein Purification

PSF was expressed in Rosetta BL21 [DE3] (pLysS) cells and purified as previously described in detail (Yarosh et al., 2015). Briefly, protein constructs PSF(exRRMs) and PSF(crystal-1) containing residues 266-484 and 276-598 respectively were cloned into pET15 vectors containing an N-terminal His-tag. PSFexRRMs was expressed overnight (12-16 hrs) at 16°C. PSF(crystal-1) was expressed for 4 hours at 37°C, as it aggregates and precipitates at cooler temperatures. Cells were resuspended in His binding buffer with PMSF and were lysed using sonication. Lysate was treated with RNase A and RNase T1, and DNase before clarification by centrifugation. Proteins were first affinity purified using a Ni2+ column (Biorad, Cat. 1560131). After elution by 200 mM imidazole, the eluates were further separated by gel filtration (Superdex 200 10/300 increase GL, GE Healthcare) using buffer containing 300 mM NaCl, 20 mM Tris, pH 7.4, 0.2 mM EDTA, and 1 mM DTT. Fractions containing protein were pooled and concentrated using Millipore Amicon Ultra centrifugal filters (10 kDa molecular weight cutoff). Protein was used immediately or dialyzed against gel filtration buffer plus 20% glycerol, and frozen with liquid nitrogen before storing at -80°C.

RNA-Binding region Identification (RBR-ID)

RBR-ID Cells were pulsed with 500 mM 4-SU for 2-hr and crosslinked at 312 nm with 1 J/cm^2 UVB. Cells were lysed, nuclei were isolated and then lysed. The nuclei lysate was diluted in 50 mM ammonium bicarbonate and reduced with 5 mM dithiothreitol followed by alkylation of cysteines with 20mM iodoacetamide for 30 min. Trypsinization was performed

at a trypsin:sample ratio of 1:100 overnight at 37°C and blocked with 1% trifluoroacetic acid. Peptides were desalted, dried, and re- suspended in 0.1% formic acid prior to MS analysis. Crosslinked RNA was removed with Benzonase.

For *in vitro* RBR-ID of PSF and ESS-RNA, preformed protein-RNA complexes were cross-linked with 1 J/cm^2 UVB, treated with RNase A, and the protein digested with trypsin (as described above) or chymotrypsin at an enzyme:sample ratio of 1:20 overnight at 25°C. See (He et al., 2016) for detailed procedure.

Each peptide intensity was extracted from chromatograms and inter-run variability was achieved by normalizing for sum of all peptide intensities in each run. Depletion was calculated as the $\log_2$ ratio of mean intensity of peptides in +4SU/-4SU samples.

Hydrogen-Deuterium Exchange (HX-MS)

Solution-phase amide deuterium exchange was conducted by mixing protein (1-5 μM final) or protein/RNA complex with 16 μL of on-exchange buffer (20mM Tris, 300 mM NaCl, 0.2mM EDTA in D_2O, pD 7) for the indicated time at 25°C, yielding 80% D_2O concentration. For protein/RNA complex samples, 10-fold molar excess of ESS RNA was used: CGU CCA CUU UCA AGU GAC CCC UUA CCU ACU CAC ACC ACU GCA UUC UCA CCC GCA AGC ACC UUU GA (Dharmacon). Exchange was quenched by addition of 40 μL ice-cold quench buffer (0.8% formic acid, 1.5M Gd-HCl, 6 mM TCEP, 10% glycerol, pH 2.2, in dH2O), and frozen in liquid nitrogen until injection into LC/MS. All steps were precisely timed to prevent/minimize back exchange between samples. A fully deuterated sample

was used to mimic full exchange and help account for back-exchange (D->H) by incubating protein in deuterated buffer at 55°C for 12-24 hrs.

Each sample (50 mL) was melted on ice and injected into a HPLC set up housed at 4°C, using pre-chilled syringes. At a flow rate of 50 µL per minute (0.1% TFA), protein was digested flowing through an immobilized pepsin (Sigma) column. Pepsin was coupled to Poros 20 AL support (Applied Biosystems) and packed into column housings of 2 mm 2 cm (64 ml) (Upchurch). Peptide fragments were collected over a TARGA C8 Piccolo high-performance liquid chromatography column (1.0 x 5.0 mm, Higgins Analytical) and eluted through a C18 analytical column (Higgins Analytical) using a 15–100% buffer B gradient at 20µL per min (Buffer A: 0.1% formic acid; Buffer B: 0.1% formic acid, 99.9% acetonitrile). The effluent was electrosprayed into the mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific). Peptides from non-deuterated buffer were analyzed in tandem MS (LTQ Orbitrap XL, Thermo Fisher Scientific). Next, MS/MS data were analyzed using SEQUEST (Bioworks v3.3.1, Thermo Fisher Scientific) with a peptide tolerance of 8 ppm and a fragment tolerance of 0.1 AMU. Deuterated samples were analyzed using EX-MS2, a MATLAB-based program detailed in (Kan et al., 2019).

Electromobility Shift Assay

RNA probes were *in vitro* transcribed using radiolabeled (P^{-32}) α-CTP, gel-purified, and resuspended to 10^4 CPM/mL specific activity. Radiolabeled RNA and purified protein were incubated for 20 minutes at 30°C in binding buffer (1.3% polyvinyl alcohol, 25 ng/µL of yeast tRNA, 20 ng/µL of BSA, 3 mM MgCl2, 1 mM ATP, 20 mM phosphocreatine, 12 mM Tris pH 7.5, 0.1 mM EDTA, 12% glycerol, 120 mM NaCl, and 0.1 µL RNasin (Promega,

40 U/μL)). Next, heparin was added to a final concentration of 0.5 μg/μL and incubated for an additional 5 min at 30° and reactions were analyzed on a native gel (Acrylamide/Bis 29:1 BioRad) and visualized by autoradiography.

In vitro phosphorylation by GSK3

The *in vitro* kinase was incubated with 0.4 μg recombinant PSF at 30°C in buffer containing 800μm ATP for 60 minutes (100mM Tris pH 7.5, 5mM DTT, 10mM MgCl2) . Next, samples were either prepared for MS to check for phosphorylation (reduced, alkylated, trypsinized, and processed for MS) or set up directly into H-D exchange before quenching and processing samples for MS. PSF was purified from bacteria as described above. Recombinant GSK-3 was used at 50units/20μL reaction (New England Biolabs).

Differential Scanning Fluorimetry

Proteins (0.1 mg/mL final) were combined with Sypro orange (Thermofisher Sci, Cat S6650) in 20 mM Tris pH 7.4, 300 mM NaCl, and subjected to a melt curve from 25-95°C using a QuantStudio 6 Flex Real-Time PCR (Applied Biosystems).

Size-exclusion chromatography with multi-angle light scattering (SEC-MALS)

Molecular weights were determined by multi-angle light scatter coupled to refractive interferometric detection. A DAWN-HELEOS MALS detector with 16 detectors and a 658.9 nm laser beam, (Wyatt Technology, Santa Barbara, CA) and Optilab T-rEX

refractometer (Wyatt Technology) were used in-line with a size exclusion chromatography analytical column (Superdex 200 10/300 column, GE Healthcare). Injections of 20 µL (3-4 mg/mL) run over the system at a flow-rate of 0.5 mL/min in 20 mM Tris pH 7.4, 300 mM NaCl, 0.2 mM EDTA, and 1 mM DTT, at room temperature. System was calibrated with 10 mg/mL BSA Fraction V (Fisher Scientific). Data collection and SEC-MALS analysis were performed with ASTRA 6.1 software (Wyatt Technology).

Crystal screens

WT PSF(exRRMs) (19 mg/mL) protein and PSF(exRRMs)-β2β3 mutant protein (3 mg/mL) were combined with ESS 45D RNA (Dharmacon, sequence below) at 1:1.11 molar ratio and dialyzed overnight at 4°C against crystallization buffer (10 mM Tris HCl, pH 7.0, 150 mM NaCl, 1mM DTT, and 3 mM MgCl2). For each crystallization condition tested, 300 nL of protein/RNA solution were added to 300 nL of reservoir solution to create a 600 nL hanging drop over a 100 uL reservoir using a Mosquito Crystallization robot (TTP Labtech). Crystals were grown using vapor diffusion in Nunc 96 well plates (Thermo Fisher Scientific) at 21°C. Screens tested for WT: Crystal Screen 1&2 (Hampton Research), Crystal Screen Cryo (Hampton Research), Wizard 106 Classic 1&2 (Molecular Dimensions), Wizard Classic 3&4 (Molecular Dimensions), GV 1- N (custom screen sourced from the Van Duyne Lab at the University of Pennsylvania). Screens tested for β2β3: Crystal Screen 1&2 (Hampton Research), Wizard Classic 1&2 (Molecular Dimensions), and GV 1-N (custom screen sourced from the Van Duyne Lab at the University of Pennsylvania).

45D ESS RNA sequence:

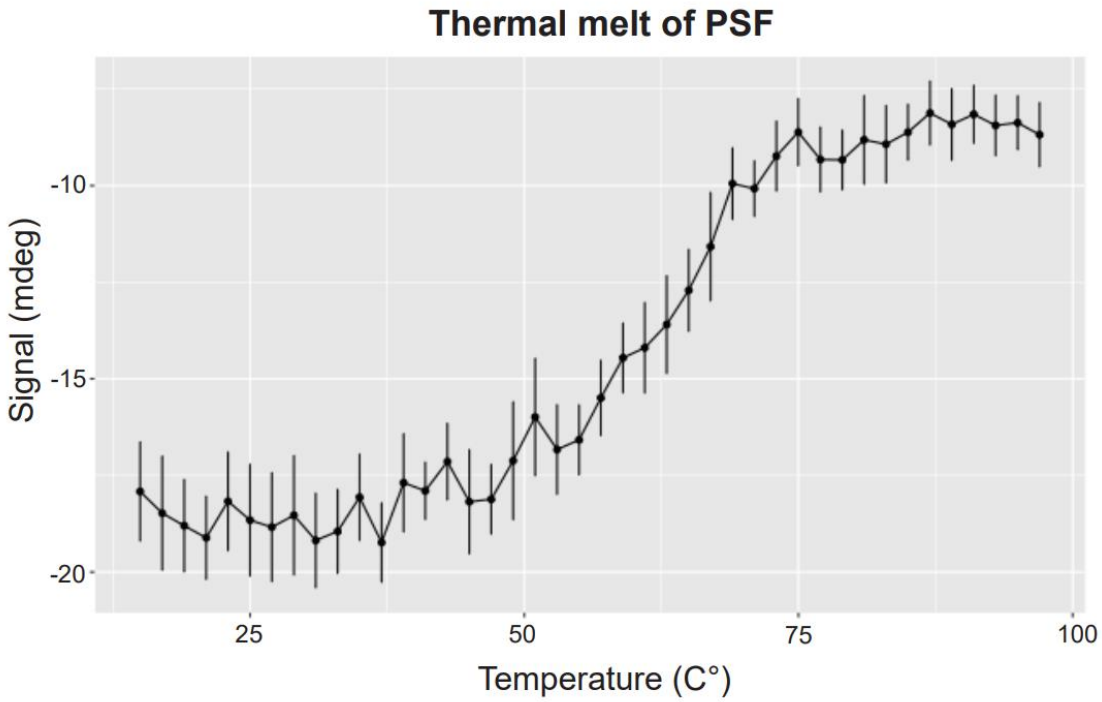
CCUUACCUACUCACACCACUGCAUUCUCACCCGCAAGCACCUU

Conservation mapping

Evolutionary conservation of amino acid positions was estimated using ConSurf (Ashkenazy et al., 2016). Alignments were generated with ClustalW, and conservation scores calculated using AL2CO and visualized with ChimeraX using default parameters of AL2CO (Pei & Grishin, 2001; Pettersen et al., 2021).

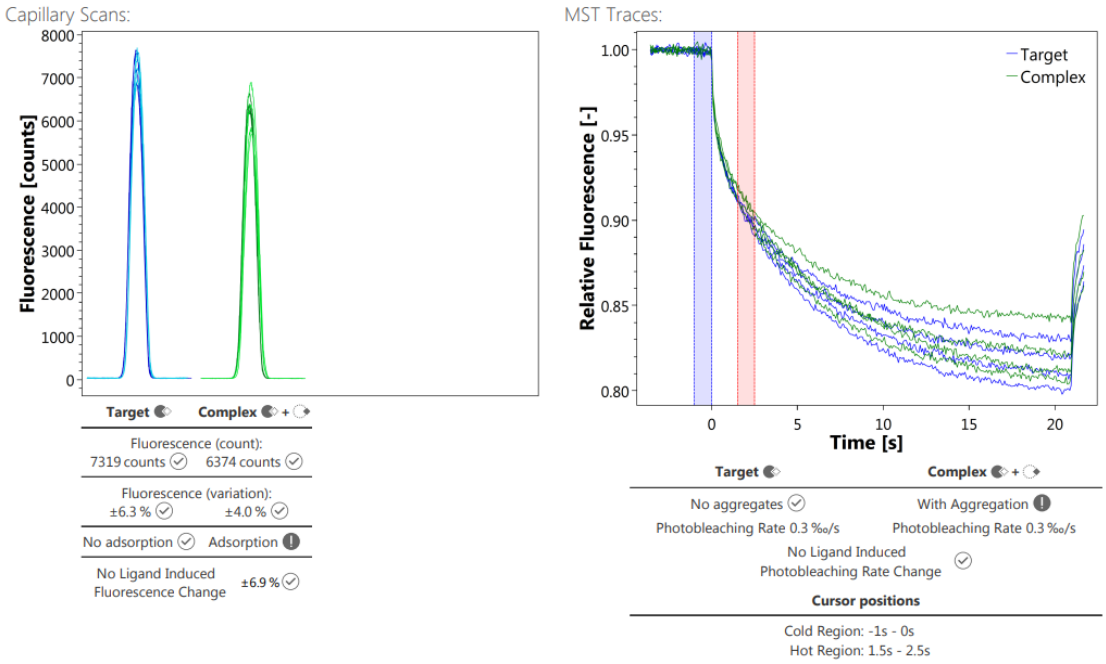
APPENDICES

Appendix A: Thermal Melt of PSF by circular dichroism



A thermal melt was performed with 25 μM PSF(exRRMS) in 50 mM phosphate buffer, 150 mM NaCl, pH 7.4 using an Aviv Biomedical model 410 circular dichroism spectrometer. From this data, a temperature of 55°C was used to create the fully deuterated control sample. Thermal melts determined by Differential Scanning Fluorimetry show a similar melting temperature profile.

Appendix B: Microscale Thermophoresis & Fluorescence Polarization



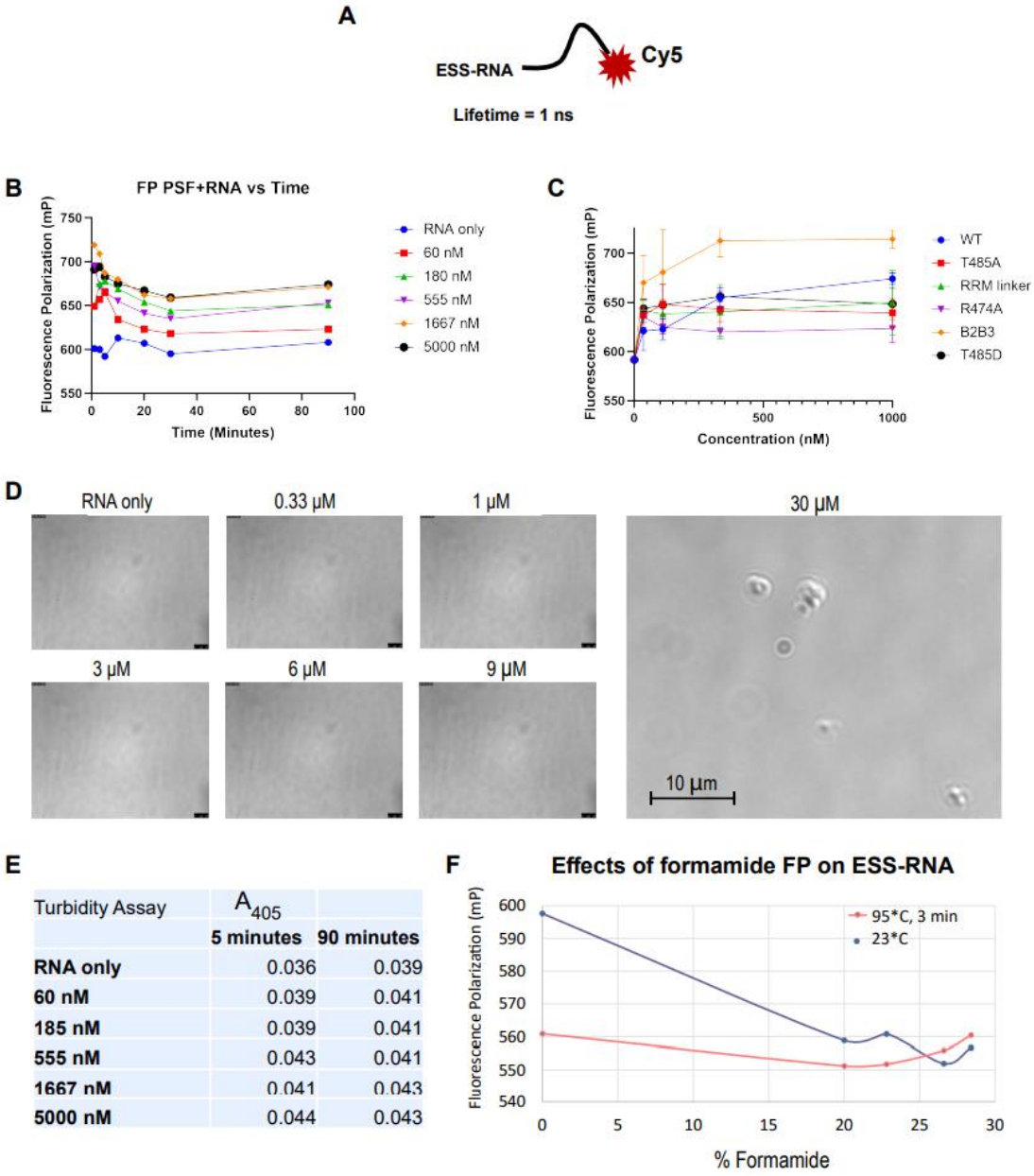
Appendix B1: Microscale Thermophoresis binding test of PSF and ESS-RNA. Capillary scan first checks that fluorescence in an optimal range with no adsorption. PSF/RNA complex failed this adsorption test. Next the MST traces check for aggregate formation. Unfortunately, PSF/RNA forms aggregates, making MST not a viable method for testing binding.

Several other higher through-put approaches to mutational binding analysis were attempted. However, the aggregation-prone nature of this protein/RNA duo proved ill-suited for other techniques. Microscale Thermophoresis (MST) is a fast and robust technique measuring picomolar dissociation constants with little sample consumption (Jerabek-Willemsen et al., 2011). MST is fluorescence-based and measures temperature-induced changes from the motion of molecules, providing information about ligand-binding through changes in size, charge, and hydration shell. While the technique worked well for a construct of PSF on its own, the addition of RNA resulted in aggregation and adsorption

which made data uninterpretable. Adsorption occurs when molecules stick to the glass capillaries used in the instrument. I saw adsorption despite using the best low-bind capillaries available. Aggregation and adsorption occurred in the RNA control samples. It is unclear whether the protein-RNA complex was also experiencing aggregation.

Another method for high-throughput binding analysis I tried was fluorescence polarization (FP). In this method, one component is fluorescently tagged, and the other is untagged. I used cy5-labeled ESS-RNA and unlabeled protein (Appendix B1). Extreme variation in the data was observed from day-to-day. Fluorescent quenching was observed overtime (Appendix B1). The binding mutants followed similar trends as the EMSAs, but the data was highly variable, and the results weren't reproducible (Appendix B2-C).

Due to the issues with aggregation in MST, I decided to look at samples using a differential interference contrast microscope to see if any liquid droplets/aggregates were forming. At the concentrations I had been using in the FP binding assays there did not appear to be any aggregates. However, at 30 μM of protein, aggregates form (Appendix B2-D). The plate reader also can check aggregation using absorbance at A_{405}, and I did not see aggregation in my samples (Appendix B2-E). Finally, I checked dimerization of the RNA-probe. Binding data via fluorescence polarization requires a size difference in the bound vs. unbound species. Smaller species rotate faster in solution compared to larger species, and the speed of fluorophore tumbling creates the fluorescence polarization. If the RNA probe was structured, it would be more compact, tumble faster, and have a smaller FP value. Upon boiling the structure there should be a slight increase in FP. If my RNA probe were predominately RNA dimer, upon boiling or adding formamide I should see a decrease in FP value – which is what I saw (Appendix B2-F). Unfortunately, FP was not a viable option for high-throughput binding assays.



Turbidity Assay	A_{405}	
	5 minutes	**90 minutes**
RNA only	0.036	0.039
60 nM	0.039	0.041
185 nM	0.039	0.041
555 nM	0.043	0.041
1667 nM	0.041	0.043
5000 nM	0.044	0.043

Appendix B2: Fluorescence Polarization with PSF/ESS-RNA. A) ESS-RNA was ordered synthesized with 3' Cy5 probe. B) Fluorescence polarization (FP) over time changes slightly, likely due to fluorophore quenching or aggregation.

BIBLIOGRAPHY

Ajith, S., Gazzara, M. R., Cole, B. S., Shankarling, G., Martinez, N. M., Mallory, M. J., & Lynch, K. W. (2016). *RNA Biology Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells*. https://doi.org/10.1080/15476286.2016.1176663

Akhmedov, A. T., & Lopez, B. S. (2000a). Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. In *Nucleic Acids Research* (Vol. 28, Issue 16).

Akhmedov, A. T., & Lopez, B. S. (2000b). Human 100-kDa homologous DNA-pairing protein is the splicing factor PSF and promotes DNA strand invasion. *Nucleic Acids Research*, *28*(16), 3022–3030. https://doi.org/10.1093/nar/28.16.3022

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*. https://doi.org/10.1093/nar/gkw408

Baljinnyam, B., Ronzetti, M., Yasgar, A., & Simeonov, A. (2020). Applications of Differential Scanning Fluorometry and Related Technologies in Characterization of Protein-Ligand Interactions. *Methods in Molecular Biology (Clifton, N.J.)*, *2089*, 47–68. https://doi.org/10.1007/978-1-0716-0163-1_4

Basu, S., Alagar, S., & Bahadur, R. P. (2021). Unusual RNA binding of FUS RRM studied by molecular dynamics simulation and enhanced sampling method. *Biophysical Journal*, *120*(9), 1765–1776. https://doi.org/10.1016/J.BPJ.2021.03.001

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S., Griffiths-Jones, Howe, K., Marshall, M., & Sonnhammer, E. (2002). The Pfam protein families database. *Nucleic Acids Research*, *30*(1), 276–280. https://doi.org/10.1093/NAR/30.1.276

Black, K. L., Naqvi, A. S., Asnani, M., Hayer, K. E., Yang, S. Y., Gillespie, E., Bagashev, A., Pillai, V., Tasian, S. K., Gazzara, M. R., Carroll, M., Taylor, D., Lynch, K. W., Barash, Y., & Thomas-Tikhonenko, A. (2018). Aberrant splicing in B-cell acute lymphoblastic leukemia. *Nucleic Acids Research*, *46*(21), 11357–11369. https://doi.org/10.1093/NAR/GKY946

Blencowe, B. J. (2017). The Relationship between Alternative Splicing and Proteomic Complexity. *Trends in Biochemical Sciences*, *42*(6), 407–408. https://doi.org/10.1016/J.TIBS.2017.04.001

Bond, C. S., & Fox, A. H. (2009). Paraspeckles: nuclear bodies built on long noncoding RNA. *The Journal of Cell Biology*, *186*(5), 637–644. https://doi.org/10.1083/JCB.200906113

Bou-Nader, C., Barraud, P., Pecqueur, L., Pérez, J., Velours, C., Shepard, W., Fontecave, M., Tisné, C., & Hamdane, D. (2019). Molecular basis for transfer RNA recognition by the

double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Research*, *47*(6), 3117–3126. https://doi.org/10.1093/NAR/GKY1302

Brown, R. S., Sander, C., & Argos, P. (1985). The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Letters*, *186*(2), 271–274. https://doi.org/10.1016/0014-5793(85)80723-7

Burghes, A. H. M., & Beattie, C. E. (2009). Splicing Missense mutation Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nature Publishing Group*, *10*. https://doi.org/10.1038/nrn2670

Castello, A., Fischer, B., Frese, C. K., Curk, T., Krijgsveld, J., Correspondence, M. W. H., Horos, R., Alleaume, A.-M., Foehr, S., & Hentze, M. W. (2016). Comprehensive Identification of RNA-Binding Domains in Human Cells Molecular Cell Resource Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell*, *63*, 696–710. https://doi.org/10.1016/j.molcel.2016.06.029

Chatrikhi, R., Mallory, M. J., Gazzara, M. R., Litterman, A. J., Ansel, K. M., & Lynch Correspondence, K. W. (2019). RNA Binding Protein CELF2 Regulates Signal-Induced Alternative Polyadenylation by Competing with Enhancers of the Polyadenylation Machinery. *CellReports*, *28*, 2795-2806.e3. https://doi.org/10.1016/j.celrep.2019.08.022

Chen, J., OuYang, H., An, X., & Liu, S. (2018). Vault RNAs partially induces drug resistance of human tumor cells MCF-7 by binding to the RNA/DNA-binding protein PSF and inducing oncogene GAGE6. *PloS One*, *13*(1), e0191325. https://doi.org/10.1371/journal.pone.0191325

Chen, L. L., & Carmichael, G. G. (2009). Altered Nuclear Retention of mRNAs Containing Inverted Repeats in Human Embryonic Stem Cells: Functional Role of a Nuclear Noncoding RNA. *Molecular Cell*, *35*(4), 467–478. https://doi.org/10.1016/j.molcel.2009.06.027

Chen, X., Yang, Z., Wang, W., Qian, K., Liu, M., Wang, J., & Wang, M. (2021). Structural basis for RNA recognition by the N-terminal tandem RRM domains of human RBM45. *Nucleic Acids Research*, *49*(5), 2946–2958. https://doi.org/10.1093/NAR/GKAB075

Chen, Y., Zubovic, L., Yang, F., Godin, K., Pavelitz, T., Castellanos, J., MacChi, P., & Varani, G. (2016). Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Research*, *44*(9), 4381–4395. https://doi.org/10.1093/NAR/GKW177

Cho, S., Moon, H., Loh, T. J., Oh, H. K., Williams, D. R., Liao, D. J., Zhou, J., Green, M. R., Zheng, X., & Shen, H. (2014). PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1839*(6), 517–525. https://doi.org/10.1016/j.bbagrm.2014.03.003

Cléry, A., Blatter, M., & Allain, F. H.-T. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, *18*(3), 290–298. https://doi.org/10.1016/J.SBI.2008.04.002

Cléry, A., Krepl, M., Nguyen, C. K. X., Moursy, A., Jorjani, H., Katsantoni, M., Okoniewski, M., Mittal, N., Zavolan, M., Sponer, J., Frédéric, &, & Allain, H.-T. (2021). Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20481-w

Cobbold, L. C., Spriggs, K. A., Haines, S. J., Dobbyn, H. C., Hayes, C., de Moor, C. H., Lilley, K. S., Bushell, M., & Willis, A. E. (2008). Identification of internal ribosome entry segment (IRES)-trans-acting factors for the Myc family of IRESs. *Molecular and Cellular Biology*, *28*(1), 40–49. https://doi.org/10.1128/MCB.01298-07

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., & Duncan, P. I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *The EMBO Journal*, *15*(2), 265–275. https://doi.org/10.1002/j.1460-2075.1996.tb00357.x

Conlon, E. G., & Manley, J. L. (2017). RNA-binding proteins in neurodegeneration: Mechanisms in aggregate. In *Genes and Development* (Vol. 31, Issue 15, pp. 1509–1528). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.304055.117

Crick, F. (1958). On protein synthesis. *Symp Soc Exp Biol.*, *12*, 138–163.

Daubner, G. M., Cléry, A., & Allain, F. H. T. (2013). RRM-RNA recognition: NMR or crystallography...and new findings. In *Current Opinion in Structural Biology* (Vol. 23, Issue 1, pp. 100–108). https://doi.org/10.1016/j.sbi.2012.11.006

Dave, P., George, B., Sharma, D. K., & Das, S. (2017). Polypyrimidine tract-binding protein (PTB) and PTB-associated splicing factor in CVB3 infection: An ITAF for an ITAF. *Nucleic Acids Research*, *45*(15), 9068–9084. https://doi.org/10.1093/nar/gkx519

Diakun, G. P., Fairall, L., & Klug, A. (1986). EXAFS study of the zinc-binding sites in the protein transcription factor IIIA. *Nature*, *324*(6098), 698–699. https://doi.org/10.1038/324698a0

Dong, L., Zhang, X., Fu, X., Zhang, X., Gao, X., Zhu, M., Wang, X., Yang, Z., Jensen, O. N., Saarikettu, J., Yao, Z., Silvennoinen, O., & Yang, J. (2010). *PTB-associated Splicing Factor (PSF) Functions as a Repressor of STAT6-mediated Ig Gene Transcription by Recruitment of HDAC1 **. https://doi.org/10.1074/jbc.M110.168377

Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., & Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annual Review of Biochemistry*, *62*, 289–321. https://doi.org/10.1146/annurev.bi.62.070193.001445

Duong, H. A., Robles, M. S., Knutti, D., & Weitz, C. J. (2011). A molecular mechanism for circadian clock negative feedback. *Science*, *332*(6036), 1436–1439. https://doi.org/10.1126/science.1196766

Duss, O., Michel, E., Diarra dit Konté, N., Schubert, M., & Allain, F. H.-T. (2014). Molecular basis for the wide range of affinity found in Csr/Rsm protein–RNA recognition. *Nucleic Acids Research*, *42*(8), 5332–5346. https://doi.org/10.1093/NAR/GKU141

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., Blencowe, B. J., & Ingles, C. J. (2002). Splicing and transcription-associated proteins PSF and p54nrb/NonO bind to the RNA polymerase II CTD. *RNA*, *8*(9), 1102–1111. https://doi.org/10.1017/S1355838202025037

Fletcher, S., Meloni, P. L., Johnsen, R. D., Wong, B. L., Muntoni, F., Wilton, S. D., & Stephen Wilton, C. D. (2013). Antisense suppression of donor splice site mutations in the dystrophin gene transcript. *Molecular Genetics & Genomic Medicine*, *1*(3), 162–173. https://doi.org/10.1002/mgg3.19

Fu, X. D., & Ares, M. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. In *Nature Reviews Genetics* (Vol. 15, Issue 10, pp. 689–701). Nature Publishing Group. https://doi.org/10.1038/nrg3778

Gazzara, M. R., Mallory, M. J., Roytenberg, R., Lindberg, J. P., Jha, A., Lynch, K. W., & Barash, Y. (2017). Ancient antagonism between CELF and RBFOX families tunes mRNA splicing outcomes. *Genome Research*, *27*(8), 1360–1370. https://doi.org/10.1101/gr.220517.117

Geuens, T., Bouhy, D., & Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. In *Human Genetics* (Vol. 135, Issue 8, pp. 851–867). Springer Verlag. https://doi.org/10.1007/s00439-016-1683-5

Ghosh, G., & Adams, J. A. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. *The FEBS Journal*, *278*(4), 587–597. https://doi.org/10.1111/j.1742-4658.2010.07992.x

Gruber, A. J., Schmidt, R., Gruber, A. R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., & Zavolan, M. (2016). A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Research*, *26*(8), 1145–1159. https://doi.org/10.1101/GR.202432.115

Gui, J.-F., Lane, W. S., & Fu, X.-D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature 1994 369:6482*, *369*(6482), 678–682. https://doi.org/10.1038/369678a0

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A., & Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular Cell*, *64*(2), 416–430. https://doi.org/10.1016/j.molcel.2016.09.034

Heyd, F., & Lynch, K. W. (2010). Phosphorylation-dependent regulation of PSF by GSK3 controls CD45 alternative splicing. *Molecular Cell*, *40*(1), 126–137. https://doi.org/10.1016/j.molcel.2010.09.013

Heyd, F., & Lynch, K. W. (2011). PSF controls expression of histone variants and cellular viability in thymocytes. *Biochemical and Biophysical Research Communications*, *414*(4), 743–749. https://doi.org/10.1016/j.bbrc.2011.09.149

Hirose, Y., & Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, *395*(6697), 93–96. https://doi.org/10.1038/25786

Hirose, Y., & Manley, J. L. (2000). *RNA polymerase II and the integration of nuclear events*.

Hua, Y., Vickers, T. A., Baker, B. F., Bennett, C. F., & Krainer, A. R. (2007). Enhancement of SMN2 Exon 7 Inclusion by Antisense Oligonucleotides Targeting the Exon. *PLoS Biology*, *5*(4), 729–744. https://doi.org/10.1371/JOURNAL.PBIO.0050073

Huang, C. J., Tang, Z., Lin, R. J., & Tucker, P. W. (2007). Phosphorylation by SR kinases regulates the binding of PTB-associated splicing factor (PSF) to the pre-mRNA polypyrimidine tract. *FEBS Letters*, *581*(2), 223–232. https://doi.org/10.1016/j.febslet.2006.12.015

Huang, J., Ringuet, M., Whitten, A. E., Caria, S., Lim, Y. W., Badhan, R., Anggono, V., & Lee, M. (2020). Structural basis of the zinc-induced cytoplasmic aggregation of the RNA-binding protein SFPQ. *Nucleic Acids Research*, *48*(6), 3356–3365. https://doi.org/10.1093/nar/gkaa076

Hwang, J. W., Cho, Y., Bae, G. U., Kim, S. N., & Kim, Y. K. (2021). Protein arginine methyltransferases: promising targets for cancer therapy. *Experimental & Molecular Medicine 2021 53:5*, *53*(5), 788–808. https://doi.org/10.1038/s12276-021-00613-y

Imai, K., & Mitaku, S. (2005). Mechanisms of secondary structure breakers in soluble proteins. *Biophysics*, *1*, 55. https://doi.org/10.2142/BIOPHYSICS.1.55

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011). Molecular Interaction Studies Using Microscale Thermophoresis. *Assay and Drug Development Technologies*, *9*(4), 342. https://doi.org/10.1089/ADT.2011.0380

Jiang, L., Shao, C., Wu, Q.-J., Chen, G., Zhou, J., Yang, B., Li, H., Gou, L.-T., Zhang, Y., Wang, Y., Yeo, G. W., Zhou, Y., & Fu, X.-D. (2017). *NEAT1 scaffolds RNA-binding proteins and the Microprocessor to globally enhance pri-miRNA processing*. https://doi.org/10.1038/nsmb.3455

Kan, Z. Y., Ye, X., Skinner, J. J., Mayne, L., & Englander, S. W. (2019). ExMS2: An Integrated Solution for Hydrogen-Deuterium Exchange Mass Spectrometry Data Analysis. *Analytical Chemistry*, *91*(11), 7474–7481. https://doi.org/10.1021/acs.analchem.9b01682

Kang, H.-S., Sánchez-Rico, C., Ebersberger, S., Sutandy, F. X. R., Busch, A., Welte, T., Stehle, R., Hipp, C., Schulz, L., Buchbender, A., Zarnack, K., König, J., & Sattler, M. (2020). An autoinhibitory intramolecular interaction proof-reads RNA recognition by the essential splicing factor U2AF2. *Proceedings of the National Academy of Sciences*, 201913483. https://doi.org/10.1073/pnas.1913483117

Kataoka, N., Bachorik, J. L., & Dreyfuss, G. (1999). Transportin-SR, a Nuclear Import Receptor for SR Proteins. *The Journal of Cell Biology*, *145*(6), 1145. https://doi.org/10.1083/JCB.145.6.1145

Kondo, Y., Oubridge, C., Van Roon, A. M. M., & Nagai, K. (2015). Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *ELife*, *2015*(4). https://doi.org/10.7554/ELIFE.04986.001

Krissinel, E., & Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, *372*(3), 774–797. https://doi.org/10.1016/J.JMB.2007.05.022

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*, *11*(1), 39–46. https://doi.org/10.1016/S0959-440X(00)00167-6

Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, B. S., Lu, Y., Iyer, K. S., Trewhella, J., Fox, A. H., & Bond, C. S. (2015a). The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv156

Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, B. S., Lu, Y., Iyer, K. S., Trewhella, J., Fox, A. H., & Bond, C. S. (2015b). The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv156

Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., & Burley, S. K. (2000). Sequence-Specific RNA Binding by a Nova KH Domain: Implications for Paraneoplastic Disease and the Fragile X Syndrome. *Cell*, *100*(3), 323–332. https://doi.org/10.1016/S0092-8674(00)80668-6

Liu, L., Xie, N., Rennie, P., Challis, J. R. G., Gleave, M., Lye, S. J., & Dong, X. (2011). Consensus PP1 Binding Motifs Regulate Transcriptional Corepression and Alternative RNA Splicing Activities of the Steroid Receptor Coregulators, p54nrb and PSF. *Molecular Endocrinology*, *25*(7), 1197–1210. https://doi.org/10.1210/me.2010-0517

Lorson, C. L., Hahnen, E., Androphy, E. J., & Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 6307. https://doi.org/10.1073/PNAS.96.11.6307

Loughlin, F. E., Lukavsky, P. J., Kazeeva, T., Reber, S., Hock, E. M., Colombo, M., Von Schroetter, C., Pauli, P., Cléry, A., Mühlemann, O., Polymenidou, M., Ruepp, M. D., & Allain, F. H. T. (2019). The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Molecular Cell*, *73*(3), 490-504.e6. https://doi.org/10.1016/J.MOLCEL.2018.11.012/ATTACHMENT/C703F259-90E0-475C-AE23-0FCD91940868/MMC1.PDF

Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. In *Nature Reviews Molecular Cell Biology* (Vol. 8, Issue 6, pp. 479–490). Nature Publishing Group. https://doi.org/10.1038/nrm2178

Lynch, K. W., & Weiss, A. (2001). *A CD45 Polymorphism Associated with Multiple Sclerosis Disrupts an Exonic Splicing Silencer**. https://doi.org/10.1074/jbc.M102175200

Mandel, C., Bai, Y., & Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences : CMLS*, *65*(7–8), 1099–1122. https://doi.org/10.1007/S00018-007-7474-3

Maris, C., Dominguez, C., & Allain, F. H. T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. In *FEBS Journal* (Vol. 272, Issue 9, pp. 2118–2131). https://doi.org/10.1111/j.1742-4658.2005.04653.x

Maris, C., Jayne, S., Damberger, F. F., Beusch, I., Dorn, G., Ravindranathan, S., Fréd, F. F., Frédéric, F., & Allain, H.-T. (2020). A transient-helix in the N-terminal RNA recognition motif of polypyrimidine tract binding protein senses RNA secondary structure. *Nucleic Acids Research*, *1*. https://doi.org/10.1093/nar/gkaa155

Martinez, N. M., Agosto, L., Qiu, J., Mallory, M. J., Gazzara, M. R., Barash, Y., Fu, X.-D., & Lynch, K. W. (2015). *Widespread JNK-dependent alternative splicing induces a positive feedback loop through CELF2-mediated regulation of MKK7 during T-cell activation*. https://doi.org/10.1101/gad.267245

Masuzawa, T., & Oyoshi, T. (2020). Roles of the RGG Domain and RNA Recognition Motif of Nucleolin in G-Quadruplex Stabilization. *ACS Applied Materials and Interfaces*, *2*, 6. https://doi.org/10.1021/ACSOMEGA.9B04221/SUPPL_FILE/AO9B04221_SI_001.PDF

Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology*, *15*(2), 108–121. https://doi.org/10.1038/nrm3742

Matunis, M. J., Michael, W. M., & Dreyfuss, G. (1992). Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Molecular and Cellular Biology*, *12*(1), 164–171. https://doi.org/10.1128/MCB.12.1.164-171.1992

MB, R., C, M., & JG, G. (1990). A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *The Journal of Cell Biology*, *111*(6 Pt 1), 2217–2223. https://doi.org/10.1083/JCB.111.6.2217

McNeil, J. B., Agah, H., & Bentley, D. (1998). Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. *Genes & Development*, *12*(16), 2510–2521. https://doi.org/10.1101/GAD.12.16.2510

Melton, A. A., Jackson, J., Wang, J., & Lynch, K. W. (2007). Combinatorial Control of Signal-Induced Exon Repression by hnRNP L and PSF †. *MOLECULAR AND CELLULAR BIOLOGY*, *27*(19), 6972–6984. https://doi.org/10.1128/MCB.00419-07

Monani, U. R., Lorson, C. L., Parsons, D. W., Prior, T. W., Androphy, E. J., Burghes, A. H. M., & McPherson, J. D. (1999). A Single Nucleotide Difference That Alters Splicing Patterns

Distinguishes the SMA Gene SMN1 From the Copy Gene SMN2. *Human Molecular Genetics*, *8*(7), 1177–1183. https://doi.org/10.1093/HMG/8.7.1177

Morozumi, Y., Ino, R., Takaku, M., Hosokawa, M., Chuma, S., & Kurumizaka, H. (2012). Human PSF concentrates DNA and stimulates duplex capture in DMC1-mediated homologous pairing. *Nucleic Acids Research*, *40*(7), 3031–3041. https://doi.org/10.1093/nar/gkr1229

Motta-Mena, L. B., Smith, S. A., Mallory, M. J., Jackson, J., Wang, J., & Lynch, K. W. (2011). *A Disease-associated Polymorphism Alters Splicing of the Human CD45 Phosphatase Gene by Disrupting Combinatorial Repression by Heterogeneous Nuclear Ribonucleoproteins (hnRNPs) * □ S*. https://doi.org/10.1074/jbc.M111.218727

Müller-Mcnicoll, M., Rossbach, O., Hui, J., & Medenbach, J. (2019). Auto-regulatory feedback by RNA-binding proteins. *Journal of Molecular Cell Biology*, *10*, 930–939. https://doi.org/10.1093/jmcb/mjz043

Neil, E. E., & Bisaccia, E. K. (2019). Nusinersen: A Novel Antisense Oligonucleotide for the Treatment of Spinal Muscular Atrophy. *The Journal of Pediatric Pharmacology and Therapeutics : JPPT*, *24*(3), 194. https://doi.org/10.5863/1551-6776-24.3.194

Neve, J., Patel, R., Wang, Z., Louey, A., & Furger, A. M. (2017). Cleavage and polyadenylation: Ending the message expands gene regulation. In *RNA Biology* (Vol. 14, Issue 7, pp. 865–890). Taylor and Francis Inc. https://doi.org/10.1080/15476286.2017.1306171

Nicastro, G., Taylor, I. A., & Ramos, A. (2015). KH-RNA interactions: Back in the groove. In *Current Opinion in Structural Biology* (Vol. 30, pp. 63–70). Elsevier Ltd. https://doi.org/10.1016/j.sbi.2015.01.002

Novoyatleva, T., Heinrich, B., Tang, Y., Benderska, N., Butchbach, M. E. R., Lorson, C. L., Lorson, M. A., Ben-Dov, C., Fehlbaum, P., Bracco, L., Burghes, A. H. M., Bollen, M., & Stamm, S. (2008). Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing. *Human Molecular Genetics*, *17*(1), 52–70. https://doi.org/10.1093/hmg/ddm284

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 1994 372:6505*, *372*(6505), 432–438. https://doi.org/10.1038/372432a0

Pan, Q., Shai, O., Lee, L., Frey, B., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. https://doi.org/10.1038/NG.259

Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., & Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes & Development*, *7*(3), 393–406.

Pei, J., & Grishin, N. v. (2001). AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics (Oxford, England)*, *17*(8), 700–712. https://doi.org/10.1093/BIOINFORMATICS/17.8.700

Peng, R., Dye, B. T., Pérez, I., Barnard, D. C., Thompson, A. B., & Patton, J. G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. *RNA*, *8*(10), 1334–1347. https://doi.org/10.1017/S1355838202022070

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science : A Publication of the Protein Society*, *30*(1), 70–82. https://doi.org/10.1002/PRO.3943

Qian, K., Li, M., Wang, J., Zhang, M., & Wang, M. (2020). Structural basis for mRNA recognition by human RBM38. *Biochemical Journal*, *477*(1), 161–172. https://doi.org/10.1042/BCJ20190652

Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecenas, D., … Hughes, T. R. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature 2013 499:7457*, *499*(7457), 172–177. https://doi.org/10.1038/nature12311

Ray, P., Kar, A., Fushimi, K., Havlioglu, N., Chen, X., & Wu, J. Y. (2011). PSF suppresses tau exon 10 inclusion by interacting with a stem-loop structure downstream of exon 10. *Journal of Molecular Neuroscience*, *45*(3), 453–466. https://doi.org/10.1007/s12031-011-9634-z

Rego, N. B., Xi, E., & Patel, A. J. (2021). Identifying hydrophobic protein patches to inform protein interaction interfaces. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(6). https://doi.org/10.1073/PNAS.2018234118/VIDEO-5

Rossbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., & Bindereif, A. (2009). Auto- and Cross-Regulation of the hnRNP L Proteins by Alternative Splicing. *Molecular and Cellular Biology*, *29*(6), 1442–1451. https://doi.org/10.1128/mcb.01689-08

Ryter, J. M., & Schultz, S. C. (1998). Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. In *The EMBO Journal* (Vol. 17, Issue 24).

Sandberg, R., Neilson, J., Sarma, A., Sharp, P., & Burge, C. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, *320*(5883), 1643–1647. https://doi.org/10.1126/science.1155390

Schroeder, S. C., Schwer, B., Shuman, S., & Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & Development*, *14*(19), 2435–2440. https://doi.org/10.1101/GAD.836300

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews. Genetics*, *17*(1), 19. https://doi.org/10.1038/NRG.2015.3

Sharathchandra, A., Lal, R., Khan, D., & Das, S. (2012). Annexin A2 and PSF proteins interact with p53 IRES and regulate translation of p53 mRNA. *Http://Dx.Doi.Org/10.4161/Rna.22707*, *9*(12), 1429–1439. https://doi.org/10.4161/RNA.22707

Shav-Tal, Y., Cohen, M., Lapter, S., Dye, B., Patton, J. G., Vandekerckhove, J., & Zipori, D. (2001). Nuclear Relocalization of the Pre-mRNA Splicing Factor PSF during Apoptosis Involves Hyperphosphorylation, Masking of Antigenic Epitopes, and Changes in Protein Interactions. *Molecular Biology of the Cell*, *12*(8), 2328–2340. https://doi.org/10.1091/mbc.12.8.2328

Shav-Tal, Y., & Zipori, D. (2002). PSF and p54 ^{nrb} /NonO - multi-functional nuclear proteins. *FEBS Letters*, *531*(2), 109–114. https://doi.org/10.1016/S0014-5793(02)03447-6

Shinde, M. Y., Sidoli, S., Kulej, K., Mallory, M. J., Radens, C. M., Reicherter, A. L., Myers, R. L., Barash, Y., Lynch, K. W., Garcia, B. A., & Klein, P. S. (2017). Phosphoproteomics reveals that glycogen synthase kinase-3 phosphorylates multiple splicing factors and is associated with alternative splicing. *The Journal of Biological Chemistry*, *292*(44), 18240–18255. https://doi.org/10.1074/jbc.M117.813527

Singh, G., Rife, B. D., Seufzer, B., Salemi, M., Rendahl, A., & Boris-Lawrie, K. (2018). Identification of conserved, primary sequence motifs that direct retrovirus RNA fate. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky369

Skrisovska, L., Bourgeois, C. F., Stefl, R., Grellscheid, S. N., Kister, L., Wenter, P., Elliott, D. J., Stevenin, J., & Allain, F. H. T. (2007). The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Reports*, *8*(4), 372–379. https://doi.org/10.1038/sj.embor.7400910

Snijders, A. P., Hautbergue, G. M., Bloom, A., Williamson, J. C., Minshull, T. C., Phillips, H. L., Mihaylov, S. R., Gjerde, D. T., Hornby, D. P., Wilson, S. A., Hurd, P. J., & Dickman, M. J. (2015). Arginine methylation and citrullination of splicing factor proline- and glutamine-rich (SFPQ/PSF) regulates its association with mRNA. *RNA (New York, N.Y.)*, *21*(3), 347–359. https://doi.org/10.1261/rna.045138.114

Some, D., Amartely, H., Tsadok, A., & Lebendiker, M. (2019). Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *Journal of Visualized Experiments : JoVE*, *2019*(148). https://doi.org/10.3791/59615

Sotillo, E., Barrett, D. M., Black, K. L., Bagashev, A., Oldridge, D., Wu, G., Sussman, R., Lanauze, C., Ruella, M., Gazzara, M. R., Martinez, N. M., Harrington, C. T., Chung, E. Y., Perazzelli, J., Hofmann, T. J., Maude, S. L., Raman, P., Barrera, A., Gill, S., … Thomas-Tikhonenko, A. (2015). Convergence of Acquired Mutations and Alternative Splicing of CD19 Enables Resistance to CART-19 Immunotherapy CANCER DISCOVERY | OF2. *Cancer Discovery*, *5*(12), 1282–1295. https://doi.org/10.1158/2159-8290.CD-15-1020

Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R., & Oyoshi, T. (2013). Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chemistry & Biology*, *20*(3), 341–350. https://doi.org/10.1016/J.CHEMBIOL.2013.02.013

Takeshima, Y., Yagi, M., Okizuka, Y., Awano, H., Zhang, Z., Yamauchi, Y., Nishio, H., & Matsuo, M. (2010). Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *Journal of Human Genetics*, *55*(6), 379–388. https://doi.org/10.1038/JHG.2010.49

Teplova, M., & Patel, D. J. (2008). Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nature Structural & Molecular Biology 2008 15:12*, *15*(12), 1343–1351. https://doi.org/10.1038/nsmb.1519

Thandapani, P., O'connor, T. R., Bailey, T. L., & Phane Richard, S. (2013). Molecular Cell Review Defining the RGG/RG Motif. *Molecular Cell*, *50*, 613–623. https://doi.org/10.1016/j.molcel.2013.05.021

Tian, B., & Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. In *Nature Reviews Molecular Cell Biology* (Vol. 18, Issue 1, pp. 18–30). Nature Publishing Group. https://doi.org/10.1038/nrm.2016.116

Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D., Olson, S., Sundararaman, B., Zhan, L., Bazile, C., Bouvrette, L. P. B., Bergalet, J., Duff, M. O., Garcia, K. E., Gelboin-Burkhart, C., … Yeo, G. W. (2020). A large-scale binding and functional map of human RNA-binding proteins. *Nature 2020 583:7818*, *583*(7818), 711–719. https://doi.org/10.1038/s41586-020-2077-3

van Zon, A., Mossink, M. H., Scheper, R. J., Sonneveld, P., & Wiemer, E. A. C. (2003). The vault complex. *Cellular and Molecular Life Sciences CMLS 2003 60:9*, *60*(9), 1828–1837. https://doi.org/10.1007/S00018-003-3030-Y

Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, J., Serganov, A., Designed Research, A. S., Performed Research, A. S., & Analyzed Data;, A. S. (2015). Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1515737112

Vitali, F., Henning, A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M., & Allain, F. H. T. (2006). Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO Journal*, *25*(1), 150–162. https://doi.org/10.1038/SJ.EMBOJ.7600911

Wheeler, E. C., Van Nostrand, E. L., & Yeo, G. W. (2018). Advances and challenges in the detection of transcriptome-wide protein–RNA interactions. *Wiley Interdisciplinary Reviews: RNA*, *9*(1). https://doi.org/10.1002/wrna.1436

Xia, Z., Donehower, L. A., Cooper, T. A., Neilson, J. R., Wheeler, D. A., Wagner, E. J., & Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nature Communications 2014 5:1*, *5*(1), 1–13. https://doi.org/10.1038/ncomms6274

Xiang, S., Gapsys, V., Kim, H. Y., Bessonov, S., Hsiao, H. H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., Lührmann, R., De Groot, B., & Zweckstetter, M. (2013). Phosphorylation Drives a Dynamic Switch in Serine/Arginine-Rich Proteins. *Structure*, *21*(12), 2162–2174. https://doi.org/10.1016/J.STR.2013.09.014

Yarosh, C. A., Tapescu, I., Thompson, M. G., Qiu, J., Mallory, M. J., Fu, X.-D., & Lynch, K. W. (2015). TRAP150 interacts with the RNA-binding domain of PSF and antagonizes splicing of numerous PSF-target genes in T cells. *Nucleic Acids Research*, *43*(18), 9006–9016. https://doi.org/10.1093/nar/gkv816

Yun, C. Y., & Fu, X.-D. (2000). Conserved Sr Protein Kinase Functions in Nuclear Import and Its Action Is Counteracted by Arginine Methylation in Saccharomyces cerevisiae. *Journal of Cell Biology*, *150*(4), 707–718. https://doi.org/10.1083/JCB.150.4.707

Zhang, Z., & Carmichael, G. G. (2001). The fate of dsRNA in the Nucleus: A p54nrb-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell*, *106*(4), 465–476. https://doi.org/10.1016/S0092-8674(01)00466-4

Zhou, Z., & Fu, X.-D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, *122*(3), 191–207. https://doi.org/10.1007/s00412-013-0407-z