
Elucidating the structure-function relationship of the cell-type specific splicing regulator, RBPMS



Ruth Partridge

Supervisor: Professor C.W.J Smith

Department of Biochemistry, University of Cambridge

This thesis is submitted for the degree of *Doctor of Philosophy*

DECLARATION:

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

I would like to dedicate this thesis to my parents, for inspiring me to pursue scientific research.

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped and supported me throughout this period, without whom completion of this PhD research would not have been possible.

My Supervisor

Firstly, I would like to thank my supervisor Professor Chris Smith for his guidance and inspiration during my research, and for providing valuable feedback on any problems or questions I was facing. I am especially grateful for the friendly and collegial atmosphere he has encouraged within his lab group and the wider RNA lab, which made the research environment both enjoyable and productive.

My Fellow Lab Members

I wish to express my gratitude for the help and guidance provided by Doctor Yi Yang, as well as inspiration during my PhD project, particularly for his help with many of the trickier protein purification problems I faced. I would also like to thank Doctor Clare Gooding and Yuling Huang for their support and guidance, especially at the start of my PhD during the pandemic; and Doctor Aishwarya Jacob for providing me with valuable feedback throughout my research. Additionally, I would like to thank all the former and current RNA lab members for their support, guidance and friendship, including Michael Barnhart, Giselle Lee, Caia Duncan, Juan Mata, Alex Borodavka, Julia Acker, Xinyu Wang, Anna Git, Nancy Standart, and Charles Charalambos.

I would also like to thank Doctor Katherine Stott for her time, expertise and assistance with answering my questions about various biophysics techniques.

Industry Collaborators

I am very grateful for the opportunity given to me by AstraZeneca when they awarded me my studentship. The insight and support I have gained from this industrial partner has been extremely useful. In particular, I would like to thank my AstraZeneca supervisor Doctor Andrea Gohlke, who has been very helpful throughout my PhD, offering guidance and support. I would also like to thank the other members of the Biophysics team, Doctor Jamie Ware and Doctor Evelyn Barrey for their time and experimental expertise.

My Former Supervisors

I would like to thank my former supervisors, Doctor Alfredo Castello and his lab members for their support of me starting a PhD; and for Professor Shona Murphy, for her support and for introducing me

to the world of RNA. Additionally, I would like to thank my undergraduate tutors, Professor George Ratcliffe and Doctor Timothy Nott for their encouragement of me to pursue a PhD.

My Friends

I am extremely lucky to have met many wonderful friends over the years who have supported me throughout my studies. I am thankful to my friends that I met in Cambridge: Emma(s), Izzi, Lena, Alex, Alistair, Meg, Katie(s), Penny, and Anna; for our regular catchups, even when the only option was meeting outside in the cold and rain. For my friends from undergrad: Andrea, Tanya and Angel, for their support and laughter during the past eight years. For my friends from school: Kim, Rosie, Lucy, Cath, Alex, Hattie, Sorchia and Isa; whose continued kindness and friendship has always been able to reach me, even if they are as far away as South Korea (or Edinburgh). For these friends and many more, I am exceptionally grateful.

I would also like to thank my partner, Abhijit, for his unwavering love, unending kindness and encouragement during my PhD, and for always being there to provide support whenever I needed.

My Family

Last but certainly not least, I would like to thank my family for their support, who undoubtedly made the pursuit of my PhD possible. To my parents, for their constant support and love at every stage, and to my siblings, Emma and Matthew, for their friendship. Additionally, I would like to thank my grandparents for their encouragement and the motivation they have provided me. For them, I will always keep climbing my mountain.

ABSTRACT

The organization of RBPs into larger complexes is known to play a crucial role in regulating various aspects of RNA metabolism, including splicing. The RNA binding protein, RNA binding protein with multiple splicing (RBPMS), regulates the alternative splicing programme in differentiated vascular smooth muscle cells (VSMCs). RBPMS has an N-terminal RNA recognition motif (RRM) that mediates both RNA binding and dimerization, allowing binding to pairs of trinucleotide CAC motifs. The C-terminal tail is predicted to be unstructured, and such regions of other RNA-binding proteins are often important for higher-order structures including condensates. There are two main isoforms expressed in differentiated VSMCs, RBPMS-A and RBPMS-B, and they have unique extreme C-terminal tails. RBPMS-A is the more active isoform, both activating and repressing splicing events, whereas RBPMS-B has only been observed to activate splicing. Previously, recombinant full-length RBPMS-A was shown to be sufficient in the repression of *Tpm1* exon 3, a model differentiated VSMC splicing event. Additionally, deletion of the C-terminal twenty amino acids of RBPMS-A caused a loss of higher order assembly and a reduction of splicing repressor activity. In this thesis, I describe detailed mutational analyses of the RBPMS-A C-terminal tail which revealed two clusters of aromatic residues were necessary for the activity of transfected RBPMS-A, whereas basic residues within the same region appear less important for activity. To allow analysis of the biophysical properties of RBPMS-A associated with different modes of splicing regulation, model splicing substrates were developed for *in vitro* RBPMS-mediated repression and activation assays. Wild-type RBPMS-A and basic mutant were fully active in both *in vitro* splicing assays, whereas the aromatic mutants were impaired for activity. Mass photometry and glutaraldehyde crosslinking have shown that these residues are also important for the ability of RBPMS-A to undergo higher order assembly. Remarkably, mutation of the basic residues also diminished higher-order assembly, despite these mutants still retaining full splicing activity. The C-terminus of RBPMS-B, however, was found to be antagonistic to splicing activity, with mutation of this region leading to full splicing activity to the level of RBPMS-A. Peptide studies of this region revealed that it was able to form very stable fibril-like structures, but the full-length protein forms large spherical structures. The different oligomerisation behaviours and activities of these isoforms and their mutants found in this study gives a greater insight into the mechanism of splicing regulation by RBPMS and the control of the splicing programme in smooth muscle cells.

TABLE OF CONTENTS

DECLARATION:.....	v
ACKNOWLEDGEMENTS	vii
ABSTRACT.....	x
TABLE OF CONTENTS	xii
TABLE OF FIGURES & TABLES.....	xvi
ABBREVIATIONS	xx
INTRODUCTION:.....	1
1.1 TRANSCRIPTIONAL CONTROL	1
1.2 CO- AND POST-TRANSCRIPTIONAL mRNA PROCESSING	4
1.21 5' AND 3' END PROCESSING	4
1.22 SPLICING.....	5
1.23 THE COUPLING OF RNA PROCESSING AND TRANSCRIPTION.....	9
1.24 POST-TRANSCRIPTIONAL REGULATION	9
1.3 ALTERNATIVE SPLICING	12
1.31 CONSEQUENCES OF ALTERNATIVE SPLICING	12
1.32 REGULATION OF ALTERNATIVE SPLICING	14
1.33 CELL-TYPE SPECIFICITY AND ALTERNATIVE SPLICING	20
1.4 ALTERNATIVE SPLICING IN VASCULAR SMOOTH MUSCLE CELLS.....	21
1.41 VASCULAR SMOOTH MUSCLE CELLS HAVE PHENOTYPIC PLASTICITY	21
1.42 ALTERNATIVE SPLICING AND THE REGULATION OF VSMC PLASTICITY	22
1.5 RBPMS: A MASTER SPLICING REGULATOR IN VSMCs.....	25
1.51 THE ROLE OF RBPMS IN VSMCs	25
1.52 NON-VSMC FUNCTIONS OF RBPMS	26

1.53 THE STRUCTURE – FUNCTION RELATIONSHIP OF RBPMS	28
1.6 ROLE OF HIGHER ORDER ASSEMBLY IN RNA BIOLOGY.....	33
1.61 AMINO ACID SEQUENCE DRIVES HIGHER ORDER ASSEMBLY	33
1.62 ROLE OF RBP HIGHER ORDER ASSEMBLY/BIOCONDENSATES IN SPLICING	36
1.7 REMAINING QUESTIONS TO BE ADDRESSED.....	39
MATERIAL AND METHODS	42
2.1: CLONING	42
2.11: DNA CONSTRUCTS.....	43
2.2 MAMMALIAN CELL CULTURE	50
2.3 RECOMBINANT PROTEIN PRODUCTION.....	54
2.31 PROTEIN PURIFICATION – RBPMS-A.....	54
2.32 PROTEIN PURIFICATION – RBPMS-B.....	56
2.4 <i>IN VITRO</i> ASSAYS	58
2.41 RNA AND SPLICING ASSAYS.....	58
2.42 <i>IN VITRO</i> BIOPHYSICAL ASSAYS	59
2.43 INTERACTION STUDIES.....	62
RESULTS.....	65
3 RBPMS CAN ACTIVATE EXON INCLUSION <i>IN VITRO</i>	65
3.1 CANDIDATE SELECTION	65
3.2 <i>PTPRF</i> : AN IDEAL TARGET FOR THE <i>IN VITRO</i> SPLICING ACTIVATION ASSAY	73
3.3 RBPMS-DEPENDENT ACTIVATION OF <i>PTPRF</i> REQUIRES RBPMS AND MBNL BINDING SITES..	83
3.4 CONCLUDING REMARKS.....	90
3.5 FUTURE PLANS	92
4 SELF-ASSEMBLY OF RBPMS IS NOT REQUIRED FOR FUNCTION	94
4.1 BACKGROUND	94
4.2 C-TERMINAL TAIL SUBSTITUTIONS IDENTIFIED IMPORTANT RESIDUES FOR SPLICING ACTIVITY	97
4.3 <i>IN VITRO</i> SPLICING ASSAYS SUCCESSFULLY REPLICATE TRANSFECTION STUDIES	103

4.4	THE C-TERMINAL MUTATIONS DIMINISH RBPMS CROSSLINKING TO RNA	112
4.5	HIGHER ORDER ASSEMBLY AND ACTIVITY OF RBPMS-A CAN BE UNCOUPLED.....	118
4.6	AROMATIC RESIDUES MAY BE REQUIRED FOR HETEROTYPIC INTERACTIONS	137
4.7	CONCLUDING REMARKS.....	142
4.8	FUTURE PLANS	146
5	EXTREME HIGHER ORDER ASSEMBLY OF RBPMS-B PREVENTS FULL ACTIVITY.....	148
5.1	BACKGROUND	148
5.2	RBPMS ISOFORMS HAVE DIFFERING ACTIVITY IN PAC1 CELLS.....	149
5.3	ACTIVITY OF RBPMS-B CAN BE INCREASED VIA MUTATION OF ITS HSA REGION	152
5.4	AROMATIC RESIDUES OF RBPMS-B C-TERMINUS ARE IMPORTANT FOR SPLICING FUNCTION	161
5.5	HSA PEPTIDE HAS FIBRIL-LIKE PROPERTIES	165
5.6	FULL LENGTH PROTEIN FORMS LARGE SPHERICAL STRUCTURES	175
5.7	CONCLUDING REMARKS.....	185
5.8	FUTURE PLANS	187
	GENERAL DISCUSSION.....	190
6.1	STRUCTURE VERSUS FUNCTION OF RBPMS	190
6.11	RELATING OLIGOMERISATION TO FUNCTION	191
6.12	TYPES OF INTERACTIONS REQUIRED FOR RBPMS OLIGOMERISATION.....	198
6.13	ALTERNATIVE SPLICING OF IDR REGIONS OFTEN MEDIATES CHANGES IN ASSEMBLY BEHAVIOUR OF RBPS.....	201
6.2	MECHANISMS OF SPLICING ACTIVATION VERSUS REPRESSION.....	203
6.21	SUCCESSFUL PRODUCTION OF AN <i>IN VITRO</i> SPLICING ACTIVATION ASSAY.....	203
6.22	POTENTIAL MECHANISMS OF SPLICING ACTIVATION	204
6.3	POTENTIAL ROLES OF RBPMS-A VERSUS RBPMS-B.....	212
6.4	CONCLUSION	214
6.5	FUTURE DIRECTIONS.....	218
	REFERENCES.....	221

TABLE OF FIGURES & TABLES

Figure 1.1 Modes of gene regulation	3
Figure 1.2 Mechanism of Splicing	6
Figure 1.3 Spliceosome assembly	7
Figure 1.4 Post-translational modifications alter the activity of proteins	11
Figure 1.5 RBPs affect alternative splicing outcomes	13
Figure 1.6 VSMCs exhibit phenotypic plasticity	22
Figure 1.7 RBPMS is a master splicing regulator of differentiated VSMCs	26
Figure 1.8 Oligomerisation of RBPMS requires its extreme C-terminus	30
Figure 1.9 RBPMS-A and RBPMS2 have a high degree of conservation	32
Figure 1.10 Multivalent interactions via IDRs drives assembly and phase separation	35
Table 2.1 Primers for cloning DNA constructs	45
Table 2.2 Minigene mutagenesis primers	47
Table 2.3: Protein construct mutagenesis primers	50
Table 2.4: RT-PCR primers used to visualise splicing events of minigene/in vitro targets or endogenous targets and siRNA and qPCR primers to assess knockdown of RBPMS isoforms	54
Figure 2.1 Schematic of the purification of recombinant RBPMS	57
Table 2.5: Primary antibodies used in immunoblot and immunofluorescence studies	63
Figure 3.1 Minigene construct sizes	66
Figure 3.2 Transfection experimental setup	68
Figure 3.3 Minigenes show differential response to RBPMS-A and RBPMS-B	70
Figure 3.4 Myocd PCR bands	71
Figure 3.5 Minigene construct conservation and deletions	73
Figure 3.6 Minigenes still respond to RBPMS-A after the first round of intronic deletions	74
Figure 3.7 PTPRF was found to be the ideal the candidate for in vitro studies	75
Figure 3.8 Purification of recombinant RBPMS-A	77
Figure 3.9 RBPMS activates PTPRF exon 23 in vitro	78
Figure 3.10 RBPMS titration shows some reciprocal change with PTPRF	79
Figure 3.11 “Cold” in vitro splicing assay	80
Figure 3.12 Recombinant RBPMS-A regulates splicing in vitro	82
Figure 3.13 Deletion of the first and second CAC clusters abolishes RBPMS-A dependent splicing activation of PTPRF exon 23	84
Figure 3.14 The CACs surrounding the regulated exon of Tpm1 are necessary for RBPMS-A activity	86

Figure 3.15 Conservation of Tpm1 CAC clusters	87
Figure 3.16 MBNL binding site mutations near CAC clusters in PTPRF affect RBPMS-A activity	89
.....	95
Figure 4.1 Deletions of the extreme C-terminus have minimal effects	96
Figure 4.2 Substitution mutants of the C-terminus.....	98
Figure 4.3 Aromatic residues of the C-terminus are important for activity	99
Figure 4.4 Reversal of charge depletes activity of RBPMS-A	101
Figure 4.5 Phosphomimetic mutant loses splicing activity	102
Figure 4.6 Recombinant untagged mutant RBPMS-A protein production	104
.....	105
Figure 4.7 Recombinant STREP tagged mutant RBPMS-A protein production.....	105
Figure 4.8 Aromatic RBPMS-A mutants lose splicing activity in vitro.	107
Figure 4.9 Introduction of negative charges in RBPMS-A reduces splicing activity in vitro.....	109
Figure 4.10 Titration of the basic and phosphomutants of RBPMS-A confirms that phosphomimetic mutant loses splicing activity	111
Figure 4.11 Wild-type RBPMS-A crosslinks RNA in nuclear extract.....	113
Figure 4.12 All RBPMS-A mutants have reduced UV crosslinking relative to wild-type	115
Figure 4.13 RKR/A and TT/A mutants show minimal UV crosslinking in nuclear extract relative to wild-type RBPMS-A	117
Figure 4.14 Recombinant RBPMS-A can phase separate in the presence of crowding reagents	120
Figure 4.15 Recombinant RBPMS-A droplet formation is dependent on concentration of protein and PVA.....	122
Figure 4.16 Aromatic residues of RBPMS-A C-terminus are important for higher-order assembly..	124
.....	124
Figure 4.17 Basic residues of RBPMS-A C-terminus are important for higher-order assembly.....	126
Figure 4.18 Higher concentration confirms basic and aromatic residues of RBPMS-A are important for higher-order assembly.....	128
Figure 4.19 Glutaraldehyde crosslinking can specifically capture the higher-order species of RBPMS	129
.....	129
Figure 4.20 Crosslinking of RBPMS-A confirm higher-order assembly behaviour is dependent on aromatic and basic residues.....	133
Figure 4.21 RNA decreases higher-order assembly of RBPMS-A.....	135
Figure 4.22 RBPMS-A WWW/A mutant is unaffected by addition of RNA.....	137

Figure 4.23 Pilot pull-down experiments confirm that RBPMS-A can interact with MBNL and RBFOX2	139
Figure 4.24 Preliminary pull-down data suggests the RBPMS-A aromatic mutants lose interaction with RBFOX2 and RBM14	141
Table 4.1 Summary of the results from the RBPMS-A mutants tested in this results chapter	142
Figure 5.1 RBPMS isoforms are produced by an alternative splicing event	149
Figure 5.2 Individual RBPMS isoforms can be knocked down in pilot experiment	151
Figure 5.3 Deletion of aggregation prone residues causes an increase in splicing activity for RBPMS-B with endogenous targets	154
Figure 5.4 Mutants of RBPMS-B have differential effects with the minigene activation targets	156
Figure 5.5 Phosphatase study confirms upper band in RBPMS western blot is due to phosphorylation	157
Figure 5.6 AlphaFold predicts that the HSA region of RBPMS-B forms antiparallel beta sheets.....	159
Figure 5.7 Proline substitutions in the HSA region enhance the splicing activity of RBPMS-B	161
Figure 5.8 Aromatic residues in RBPMS-B C-terminus were mutated to alanine	162
Table 5.1 Splicing activity and predicted aggregation propensity of RBPMS-B and its mutants.....	165
Figure 5.9 Aromatic residues aid splicing activity of RBPMS-B.....	165
Figure 5.10 Circular dichroism confirms that the HSA region of RBPMS-B can form highly stable anti-parallel beta sheets	167
Figure 5.11 Negative stain electron microscopy images show HSA peptide of RBPMS-B forms large structures.....	169
Figure 5.12 Fluorescence plate reader studies shows ThT binds specifically to fibrils	170
Figure 5.13 Fluorescence plate reader studies show ThT binds to the HSA peptide	171
Figure 5.14 Guanidine HCl can denature HSA fibrils at high concentration	173
Figure 5.15 Amyloid peptide and HSA form similar structures with AlphaFold.....	174
Figure 5.16 Recombinant RBPMS-B sediments into heaviest fractions of 10-60% glycerol gradient	176
Figure 5.17 Recombinant RBPMS-B can be successfully purified	178
Figure 5.18 Recombinant RBPMS-B is mainly insoluble after protein induction	179
Figure 5.19 42°C protein induction incubation temperature leads to more soluble RBPMS-B being produced.....	180
Figure 5.20 Negative stain EM images show RBPMS-B forms large circular structures in vitro	182
Figure 5.21 Recombinant RBPMS-B P2 mutant protein can be purified in the same manner as RBPMS-A	184

Figure 6.1 Splicing repression by RBPMS.....	193
Figure 6.2 Assemblies of RBPMS-A and RBPMS-B.....	198
Figure 6.3 PTPRF has a weak polypyrimidine tract downstream from exon 23.....	207
Figure 6.4 Potential activation model with PTPRF and RBPMS	209
Figure 6.5 Summary of the findings about the structure-function relationship of RBPMS.....	217

ABBREVIATIONS

ΔPSI = PSI change

3'SS = 3' splice site

5'SS = 5' splice site

A = Adenine

A = Alanine

ACTN1 = Actinin1

Ad2 = Adenovirus 2

ALS = Amyotrophic lateral sclerosis

AS = Alternative splicing

ASE = Alternative splicing event

ASO = Antisense oligonucleotide

bp = Base pairs

BSA = Bovine Serum Albumin

C = Cysteine

C = Cytosine

Cald1 = Caldesmon1

cDNA = complementary DNA

CE = Cassette exon

CLIP = UV crosslinking and immunoprecipitation

CTD = Carboxy terminal domain

CVD = cardiovascular disease

DMEM = Dulbecco's Modified Eagle medium

DNA = Deoxyribonucleic acid

DTT = Dithiothreitol

E = Glutamate

ECM = Extracellular matrix

EDTA = Ethylenediaminetetraacetic acid

EMSA = Electrophoretic mobility shift assay

ESE = Exonic splicing enhancer

ESS = Exonic splicing silencer

F = Phenylalanine

FBS = Fetal bovine serum

Flnb = Filamin B

FRAP = Fluorescence recovery after photobleaching

FT = Flow-through

FUS = Fused in Sarcoma

G = Glycine

G = Guanine

H = Histidine

HEK293T = Human embryonic kidney 293 T antigen cells

Hermes = Heart RRM Expressed Sequence

hnRNP = Heterogeneous nuclear ribonucleoproteins

HSA = Hot spot for aggregation

IPTG = Isopropyl β -D-1-thiogalactopyranoside

IR = Intron retention

ISE = Intronic splicing enhancer

ISS = Intronic splicing silencer

K = Lysine

KCl = Potassium chloride

KD = Knockdown

kDa = Kilodalton

L = Leucine

LARKS = Low-complexity amyloid-like kinked segments

LASR = Large assembly of splicing regulators

LC = Low complexity

LLPS = Liquid-liquid phase separation

MBNL = Muscleblind

MgCl₂ = Magnesium Chloride

MP = Mass Photometry

mRNA = messenger RNA

Myocd = Myocardin

N = Asparagine

NM = non-muscle

NMD = Nonsense-mediated decay

ns = not significant

nt = Nucleotides

P = Proline

PAC1 = Rat pulmonary artery cell line

PAR-CLIP = Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation

PBS = Phosphate-buffered saline

PCR = Polymerase chain reaction

PEG = Polyethylene glycol

Pol II = RNA polymerase II

Poly(A) = Poly adenosine

PPFIA1 = PTPRF Interacting Protein Alpha 1

PPFIBP1 = PPFIA Binding Protein 1

PPI = Protein protein interaction

PSI = Percentage spliced in

PTBP1 = Polypyrimidine tract binding protein

PTC = Premature termination codon

PTM = Post-translational modification

PTPRF = Protein Tyrosine Phosphatase Receptor type F

PVA = Polyvinyl alcohol

P = P-value

QKI = Quaking protein

R = Arginine

RBP = RNA binding protein

RBPMS = RNA binding protein with multiple splicing

RNA = Ribonucleic acid

RNAseq = RNA sequencing

RRM = RNA recognition motif

RT = Reverse transcriptase

RT-PCR = Reverse transcription PCR

S = Serine

SDS = Sodium dodecyl sulphate

SD = Standard Deviation

SEC = Size-exclusion chromatography

SF = Splicing factor

SF1 = Splicing factor protein

SM = Smooth muscle

SMC = Smooth muscle cell

snRNP = Small nuclear ribonuclearproteins

SR protein = Serine Arginine rich protein

T = Threonine

T = Thymine

TF = Transcription factor

TPM1 = Tropomyosin 1

U = Uracil

UV = Ultraviolet light

VSMC = Vascular smooth muscle cells

W = Tryptophan

WT = Wild-type

Y = Tyrosine

INTRODUCTION:

1.1. TRANSCRIPTIONAL CONTROL

Gene expression involves the flow of genetic information from DNA to RNA, and finally to protein. The DNA stores the genetic information necessary to synthesise proteins and functional RNA products. DNA is transcribed into messenger RNA (mRNA) during transcription, and this mRNA acts as the intermediary between DNA and protein synthesis (Jacob & Monod, 1961; Nirenberg & Matthaei, 1961). In eukaryotes, messenger RNA gets shuttled from the nucleus, where DNA is stored, to the cytoplasm where protein synthesis occurs, during translation. Gene expression can be regulated pre-transcription, during transcription, after transcription, during translation, and post translation (Fig.1.1). These many layers of regulation allow for exquisite control over gene expression, which can be fine-tuned depending on development, tissue-type, cell-type and particular cellular conditions.

Transcriptional control is dependent on *cis* sequences within the DNA and on *trans* transcription factors (TFs) which bind to the DNA. Promoter sequences within the DNA near the transcription initiation site “promote” transcription via the binding of the transcriptional machinery (Fig.1.1). Other *cis* elements in the DNA can activate or repress transcription via the binding of *trans*-acting transcription factors (Dyner, 1989). The activating *cis* and *trans* elements work together to achieve successful transcription, and this involves the recruitment of the transcriptional machinery and RNA polymerase II (pol II) which catalyses transcription. In addition, the DNA inside eukaryotic cells is not freely accessible in the nucleus. Instead, it exists as chromatin, which is a mixture of DNA and protein, namely histone proteins. This chromatin is modified via the addition of post translational modifications (PTMs) such as acetylation, methylation and phosphorylation, and these markers alter the compaction state of chromatin. Certain markers such as H3K27 tri-methylation and H3K9 methylation are associated with inactive chromatin, which is more compact, making the DNA inaccessible to transcription machinery. Active chromatin markers such as H3K4 methylation and H3K27 acetylation are associated with relaxed chromatin which is readily accessible to the transcription machinery. Thus, the interplay between the transcription machinery and chromatin modifiers is another level in which gene expression is regulated (Berger, 2002; Cosma, 2002; Narlikar *et al*, 2002).

Sequences that are located far away from the start site of transcription have also been found to be important for the regulation of transcription. These sequences, called enhancers, are also bound by transcription factors and aid transcription initiation (Dyner, 1989; Spitz & Furlong, 2012; Panigrahi & O’Malley, 2021). These enhancer sequences are non-coding and act independently from the distance

and orientation of their target gene, and multiple enhancers can be found together to form a super-enhancer. Enhancers, and in particular super-enhancers, have been found to be important in regulating transcription in a cell-type specific manner during development. Here, cell-type specific transcription factors and chromatin modifying enzymes bind to the enhancers and activate transcription of cell-type specific proteins, including RNA processing factors, TFs and signalling proteins. Transcriptional control represents the first layer of gene regulation, and requires coordination from pol II, transcription factors bound at the promoter/promoter proximal sites and enhancers, as well as enzymes which modify the chromatin to make it accessible for these proteins. This allows for controlled gene expression which can be turned on or off with specific cellular signals, but this control can be further fine-tuned at the level of co-transcriptional and post-transcriptional processing (Fig.1.1).

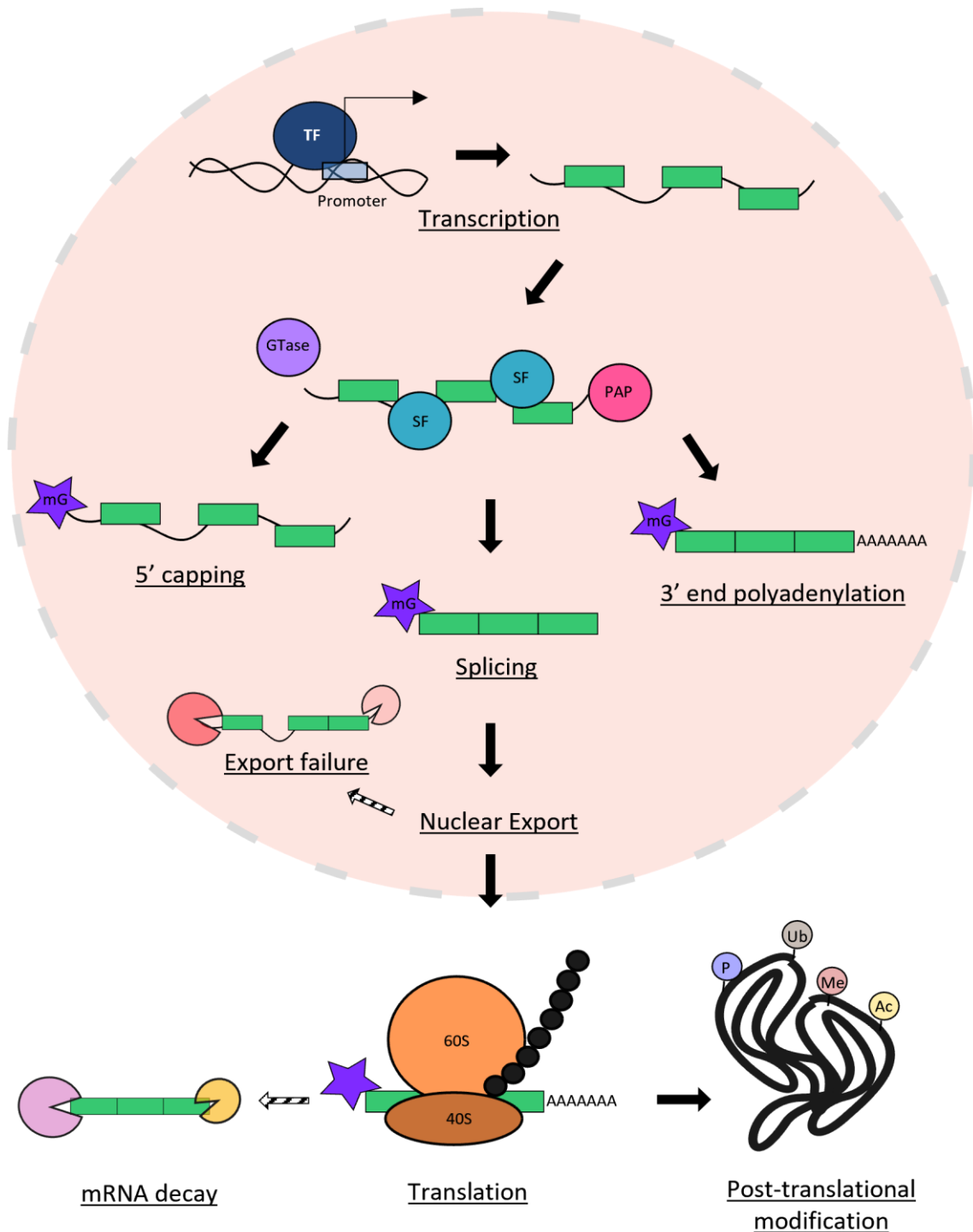


Figure 1.1 Modes of gene regulation: Schematic of the processes of gene regulation from transcription of DNA to RNA, to post-translational modification of proteins. Genes are transcribed from DNA to RNA, and co-transcriptional processing factors bind to the RNA to produce mature mRNA. These proteins include the guanylyltransferase (GTase) which adds the guanine at the 5' end, splicing factors (SF) which recruit the spliceosome complex to remove introns, and the poly adenylation polymerase (PAP) which adds the 3' polyA tail. Mature mRNA is then exported from the nucleus to the cytoplasm where it is translated by the 80S ribosome (made up of the 60S and 40S subunits) into protein, which can then undergo various post translation modifications: phosphorylation (P), ubiquitination (Ub), methylation (Me) and acetylation (Ac) amongst others. Improper processing can prevent export leading to mRNA decay by exonucleases. Improper signals within the mRNA itself can also lead to mRNA decay after export. Pathways leading to RNA turnover are indicated by dashed arrows. TF = transcription factor.

1.2 CO- AND POST-TRANSCRIPTIONAL mRNA PROCESSING

As the DNA is transcribed to pre-messenger RNA, it starts to be processed to get it ready for translation into protein. This processing occurs during and after transcription and involves the addition of a 5' cap and 3' poly adenosine (poly(A)) tail, and the removal of non-coding regions of the RNA via splicing. These non-coding regions of the RNA are known as introns and must be removed to produce a mature mRNA, which only contains the coding regions of the gene, known as exons. This processing is necessary to produce a stable mRNA that will not be targeted by exonucleases and properly exported, and will be recognised by the translation machinery, thus allowing successful production of protein (Fig.1.1). As with transcription, these processes are tightly controlled, and each step allows for more fine-tuning of gene regulation.

1.21 5' AND 3' END PROCESSING

5' end capping

The 5' end of the mRNA must be capped with a Guanosine Monophosphate (GMP) moiety to ensure stability of the transcript, as well as allowing recognition from the translation initiation factor, eIF-4E (Proudfoot *et al*, 2002). This capping occurs early on in transcription, after approximately 20 – 30 nucleotides have been transcribed, and therefore occurs in close proximity to the largest subunit of pol II, the C-terminal domain (CTD) (Rasmussen & Lis, 1993). Briefly, the capping process involves the removal of the γ phosphate of the 5' end triphosphate of the RNA by RNA triphosphatase, followed by the addition of GMP by the capping enzyme guanylyltransferase via a 5' – 5' linkage, forming GpppN structure. The cap is then methylated by a methyltransferase forming the m7GpppN structure (reviewed in Bentley, 2014). The CTD of pol II can directly interact with the methyltransferase and the guanylyltransferase, which it activates. The capping enzyme are recruited to genes at the promotor-proximal pause step of transcription, and the capping enzyme interacts with the transcription elongation factor DSIF. This elongation factor is thought to cause pausing to allow for the checking of the cap to ensure it has occurred properly (Mandal *et al*, 2004).

3' end processing

The 3' end of the pre-mRNA undergoes a cleavage and a subsequent addition of a poly(A) tail which is signalled by a AAUAAA sequence in the 3' end of the pre-mRNA, located upstream of the cleavage site. The AAUAAA signals the CPSF protein to bind, and another protein complex, CstF binds downstream at a GU-rich region, and together with Cleavage Factors I and II, they cleave the mRNA predominantly at a CA dinucleotide approximately 10 – 30 nucleotides downstream of the AAUAAA sequence

(Proudfoot *et al*, 2002; Bentley, 2014). Poly(A) polymerase (PAP) adds the poly(A) tail whilst the poly(A) binding protein (PABP) binds the poly(A) tail as it emerges, regulating its length by enhancing the processivity of PAP. The poly(A) tail protects the mRNA from exonucleases, and along with the 5' cap, it is recognised by export factors to allow for nuclear export into the cytoplasm where PABP can interact with translation initiation machinery (Fig.1.1). Polyadenylation has been linked to termination of transcription and to the removal of pol II from the DNA template (Whitelaw & Proudfoot, 1986; Logan *et al*, 1987; Nojima *et al*, 2013).

1.22 SPLICING

The splicing reaction

Another critical step in the maturation of pre-mRNA is splicing, which involves removing the non-coding intronic regions of the RNA and joining the protein coding sequences, the exons (Figures 1.1 & 1.2) (Sharp, 1987). The splicing reaction itself involves two transesterification reactions. The first of which involves the 2'OH of the branch point adenosine (bA) attacking the phosphate group at the 5' splice site (5'SS), resulting in a cleavage of the RNA at the 5'SS, and the intron forms a "lariat" intermediate structure via a 2' – 5' phosphodiester bond. This leaves a 3'OH at the upstream exon which then attacks the phosphate group at the 3' splice site (3'SS) in the second transesterification reaction, joining together the two exons and releasing the intronic lariat.

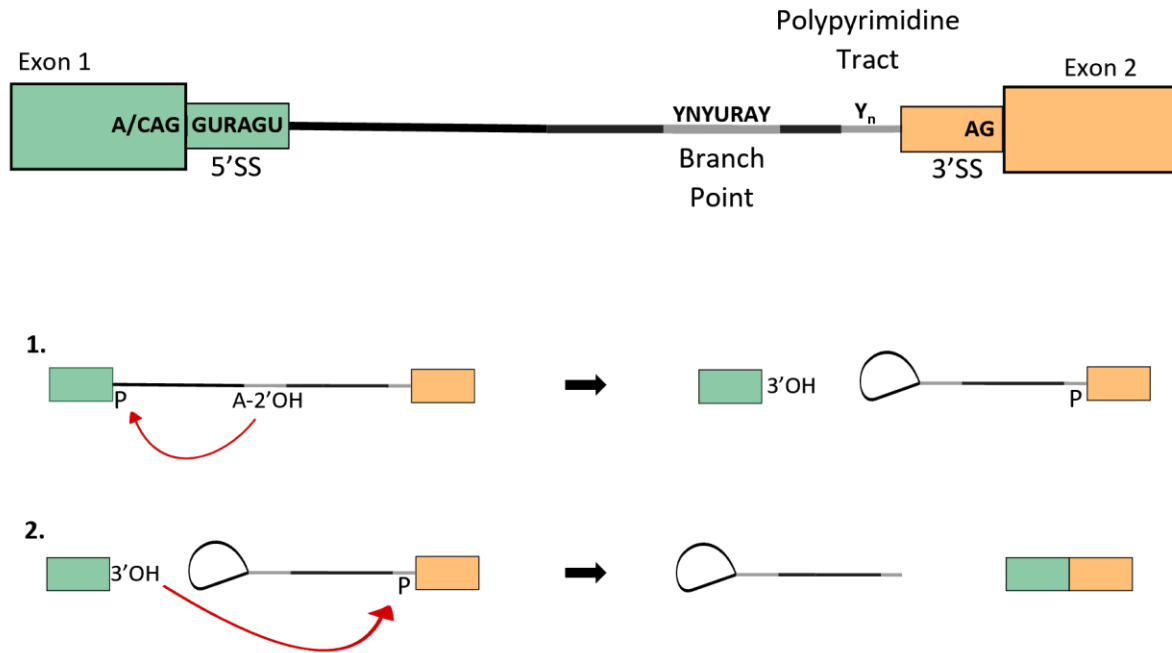
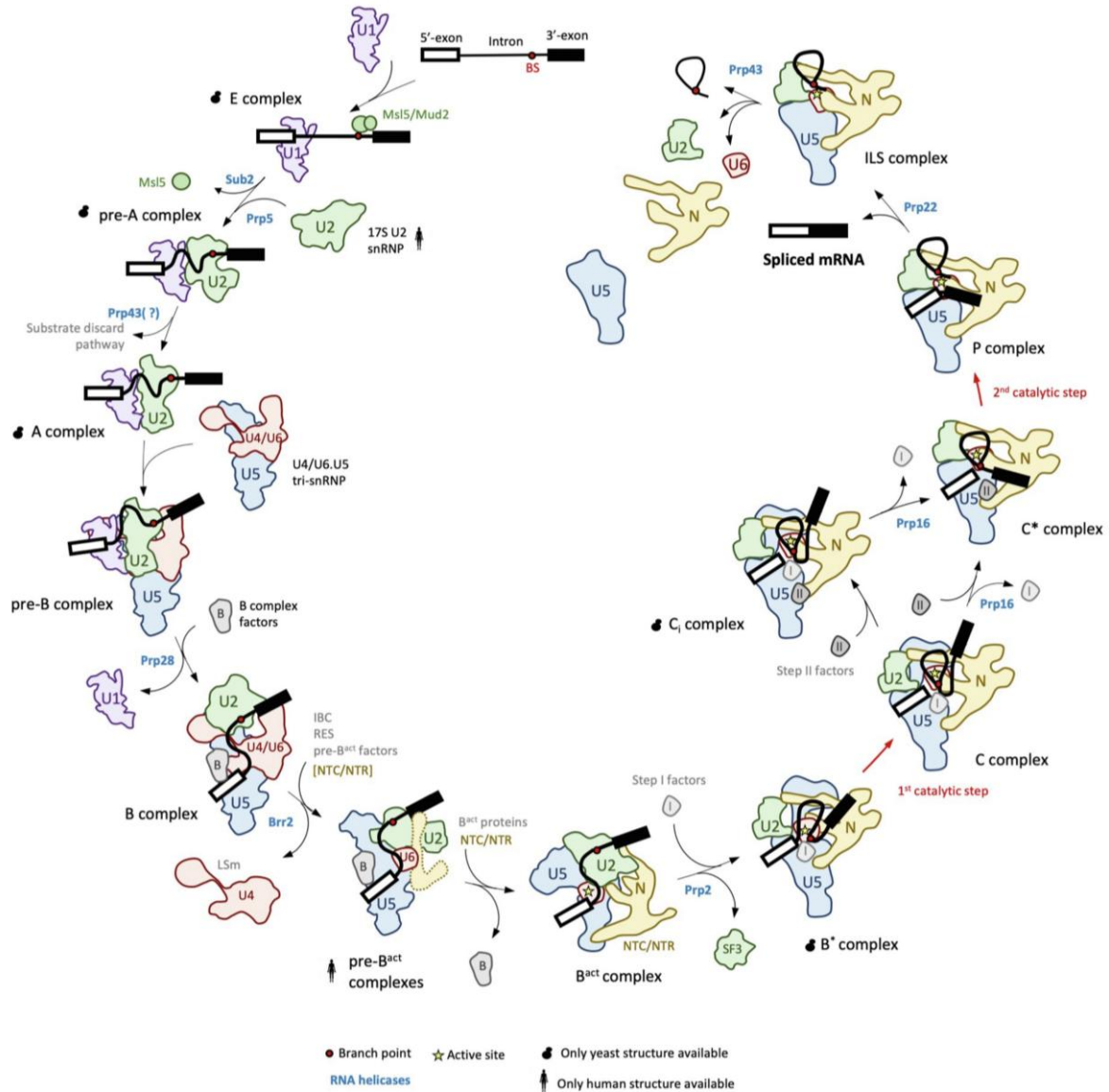


Figure 1.2 Mechanism of Splicing: Splicing reaction schematic showing the cis-elements involved in mRNA splicing and the two trans-esterification reactions resulting in an intron being spliced out as a lariat structure and the two exons being brought together. 5'SS = 5' splice site, 3'SS = 3' splice site. P = phosphate.

The splicing reaction is carried out by a complex of proteins and RNA known as the spliceosome. There are five small nuclear RNAs (snRNAs): U1, U2, U4, U5 and U6 that associate with proteins to form snRNPs (small nuclear ribonucleoproteins) which play an important role in splicing. The spliceosome undergoes various structural and compositional changes throughout the splicing reaction, forming various different complexes (Fig.1.3) (Will & Luhrmann, 2011; Hartmuth *et al*, 2012; Tholen & Galej, 2022). The early (E) complex forms when the 5'SS splice site is recognised by U1 snRNP via U1 snRNA:5'SS base pairing and the U2 auxiliary factor (U2AF) binds to the polypyrimidine tract. This protein lies between the branchpoint adenosine and the 3'SS and helps to position U2 snRNP at the branchpoint and dislodges the branch point binding protein (SF1). The A complex forms when the U2 snRNA base pairs with branch point sequence, with U2 snRNP forming non-specific RNA:protein contacts, causing the bA to bulge out ready to attack the 5'SS phosphate group. The remodelling to form the A complex requires ATP, as does the remodelling to form the next complex, the pre-B complex. This complex forms when the tri-snRNP complex (U4, U6 and U5 snRNPs) joins the complex. Major changes to the arrangements of RNA:RNA and RNA:protein contacts occur, and U1 is released, forming the B complex. U4 snRNP is then lost during the transition from the B to the B^{act} complex. This frees U6 snRNA to fold and associate with U2 snRNA, forming the catalytic core of the spliceosome. This complex gets catalytic activation via Prp2, a DEAH-box RNA helicase, to form the B* complex which

catalyses the first transesterification step, and then the C complex catalyses the last step. This subsequently forms the P complex; the spliceosome dissociates and the snRNPs are remodelled so that they can catalyse further rounds of splicing. All steps of spliceosome assembly after E complex formation require ATP hydrolysis.



Current Opinion in Structural Biology

Figure 1.3 Spliceosome assembly: Schematic of simplified spliceosome assembly with snRNPs and some trans-acting factors. Schematic is largely based off yeast (*Saccharomyces cerevisiae*) apart from 17S U2 snRNP and pre-B^{act}. Taken from Tholen & Galej, 2022.

The exon and intron definition models of splicing

The splicing reaction depicted in Figure 1.2 shows the intron definition model of splicing, whereby the splicing machinery is thought to form across the intron which is to be excised (De Conti *et al*, 2013). This is what is primarily observed in yeast, and in higher eukaryotes when introns are shorter, and exons are longer. In cases where introns are longer and exons are shorter, it was proposed instead that the spliceosome first defined and spread along exons, in the exon definition model (Robberson *et al*, 1990; Berget, 1995). The evidence for this mode of recognition came from the observation that mutation of the 5'SS of an exon resulted in this exon being skipped, demonstrating that the spliceosome requires exon recognition for constitutive splicing to occur. If spliceosome assembly occurred through intron definition, this mutation would have resulted instead in intron retention. The 5'SS and 3'SS *cis* elements are important for recruiting the spliceosome (U1 snRNP and U2AF respectively), but other *cis* elements are also required to ensure further exon definition by allowing the splicing machinery to recognise “true” exons over “false” exons (Blencowe, 2000; Wagner & Garcia-Blanco, 2001; Cartegni *et al*, 2002; Ladd & Cooper, 2002). These sequences are known as splicing enhancers and they recruit *trans* acting factors such as the SR family of splicing factors, which can stabilise the exon-defined spliceosome complex. During the splicing reaction, it is thought that the exon-defined complex switches to an intron-defined complex, as the chemical reaction of splicing occurs across the intron. This switching is not fully understood, but it has recently been suggested to occur at the B complex stage, and not involve U1 – U2 interaction, as previously thought (Will & Luhrmann, 2011; Zhang *et al*, 2024). Instead, it is thought that extensive rearrangements of the exon defined B complex occur to generate the catalytically activated B complex. This study demonstrated that the spliceosome commits to a particular 5'SS at the pre-B to B complex assembly stage, when the 5'SS is bound by the U6 snRNA after U1 is dislodged. Throughout the formation of the spliceosome and the splicing reaction, there are multiple fidelity steps that occur that involve the snRNAs and associated proteins (Tholen & Galej, 2022). These steps include the rejection of improper substrates by the U2 snRNP branch-point stem loop unwinding, and the blocking of SF3B1 binding to the U2 snRNP branch helix due to the branch-point adenosine not bulging out properly. This is thought to prevent SF3B1 from transitioning from an open to a closed conformation, and therefore progression of the spliceosome to a fully formed A complex does not occur. These checkpoints allow for splicing to occur in the proper manner and prevents incorrect splicing sequences from being used, thus preventing aberrant splicing.

1.23 THE COUPLING OF RNA PROCESSING AND TRANSCRIPTION

The various mRNA processing steps influence each other and they themselves are also influenced by transcription. This has been demonstrated to be due to members of the mRNA processing machineries interacting with pol II as it carries out transcription, in particular with its C-terminal domain (Proudfoot *et al*, 2002; Kwek *et al*, 2002; Anvar *et al*, 2018). The CTD of pol II can be post-translationally modified, with its heptad repeat of YSPTSPS undergoing phosphorylation (Y/S/T), *cis-trans* isomerisation (P) and glycosylation (T/S) (reviewed in Zaborowska *et al*, 2016). These modifications have been found to be required for various RNA processing and splicing factors, with capping enzymes recognising Serine 5 phosphorylation and poly(A) factors binding to phosphorylated Serine 2. Additionally, serine 5 phosphorylation has been linked to the recruitment of splicing factor SRSF1. Furthermore, the polypyrimidine tract binding protein U2AF65 has been found to directly interact with the phosphorylated CTD of pol II and consequently enhances splicing (David *et al*, 2011). These CTD dependent interactions with the RNA processing machineries suggests that these processes occur co-transcriptionally.

The U1-70K protein was observed to interact with pol II subunits RPB2 and RBP12 which positions the 5'SS close to the exit channel of pol II (Zhang *et al*, 2021). This is thought to lead to the looping out of the intron as the RNA is transcribed which may facilitate the identification of the 3'SS and assembly of the spliceosome. Spliceosomes have been observed to be associated with chromatin which again suggests they are acting co-transcriptionally, and is suggested to aid in guiding the splicing machinery to exons (Kfir *et al*, 2015). Splicing has also been shown, however, to occur post-transcriptionally at nuclear speckles which are condensates within cells made up primarily of RNA and processing and splicing factors. Proportionally, it is thought that splicing is ~80% co-transcriptional and ~20% post-transcriptional (Girard *et al*, 2012).

1.24 POST-TRANSCRIPTIONAL REGULATION

RNA turnover

The above RNA processing steps all have various checkpoints to ensure that the pre-mRNA is being correctly processed to become a mature mRNA that can be translated into protein. If these checkpoints detect defects in the pre-mRNA, for example improper capping or intron retention, then the RNA can be degraded in a process called RNA turnover. Additionally, signals within exons can lead to RNA turnover, for example premature stop codons. These signals can be recognised by the translation

machinery and lead to the RNA being degraded by the nonsense-mediated decay (NMD) pathway (Houseley & Tollervey, 2009).

RNA turnover is important for maintaining high fidelity transcripts and is another way in which gene expression is controlled. The various pathways of RNA turnover can respond to environmental signals and is important for cellular homeostasis. There are multiple pathways for RNA turnover in cells (Fig.1.1), which can occur in both the nucleus and cytoplasm and there is often redundancy between these pathways which allows RNA degradation to be a highly efficient process (Houseley & Tollervey, 2009; Horvathova *et al*, 2017; Westbrook *et al*, 2023). The RNA turnover process is itself highly regulated by various factors including proteins which bind to RNAs and influence their stability, micro RNAs (miRNAs) can also bind and influence the stability of RNAs, and cellular signals can activate or inhibit specific degradation pathways. This allows cells to fine-tune gene expression both as transcripts are being transcribed and processed, and after the mRNA has been processed and exported from the cytoplasm.

Post-translational modifications

Once mRNA has been transcribed, processed and exported to the cytoplasm it is ready to be translated into protein. As with transcription, translation is also a highly regulated process, although the principal regulation occurs at the initiation stage, so as to conserve resources and energy within the cell and to allow rapid gene expression control (as reviewed in Jackson *et al*, 2010). This regulation occurs as the ribosome is recruited to the mRNA and begins protein synthesis. Once proteins have been synthesised, their activity and lifetime can be further modulated by the addition of post-translation modifications (PTMs) (Fig.1.4). PTMs are often used as a reversible way to modulate the activity of a protein, for example by changing binding sites and altering protein-protein interaction (PPI) networks. Individual PTMs can cause such alterations in activity/binding, but additionally, multiple PTMs can be added which can further alter activity of a protein, for example by changing which proteins recognise it. During the cell-cycle, many proteins will be phosphorylated by cyclin-dependent kinases, one such protein is the Retinoblastoma protein (Rb), which normally inhibits E2F transcription factors to prevent the cell progressing for the G1 to S phase. Hyperphosphorylation leads to inactivation of Rb and thus allowing cell-cycle progression (Mittnacht & Weinberg, 1991). This process is tightly regulated but can become dysregulated during cancer, leading to uncontrolled proliferation (Roesch *et al*, 2005). As previously discussed, the phosphorylation of the CTD of pol II can affect RNA processing as well as transcription. These modifications are often added by cyclin-dependent kinases such as CDK9 and CDK7, which allows the linking of transcription with cell-cycle stage, as well as various other cellular signals (Bregman *et al*, 2000). Various other proteins can be modified post-translationally with a variety

of moieties, for example, the histone proteins in chromatin are known to undergo various types of modification including methylation, acetylation and phosphorylation. These markers can alter the state of the chromatin and either allow or inhibit the recruitment of the transcription machinery, and is also known to alter splicing due to different histone “marks” (PTMs) at exons and introns (Hunter, 2007; Schwartz *et al*, 2009). Another type of post-translational modification that proteins can undergo is arginine methylation, which is known to affect RNA processing and RNA binding domains are prone to arginine methylation (Blanc, 2016; Larsen *et al*, 2016).

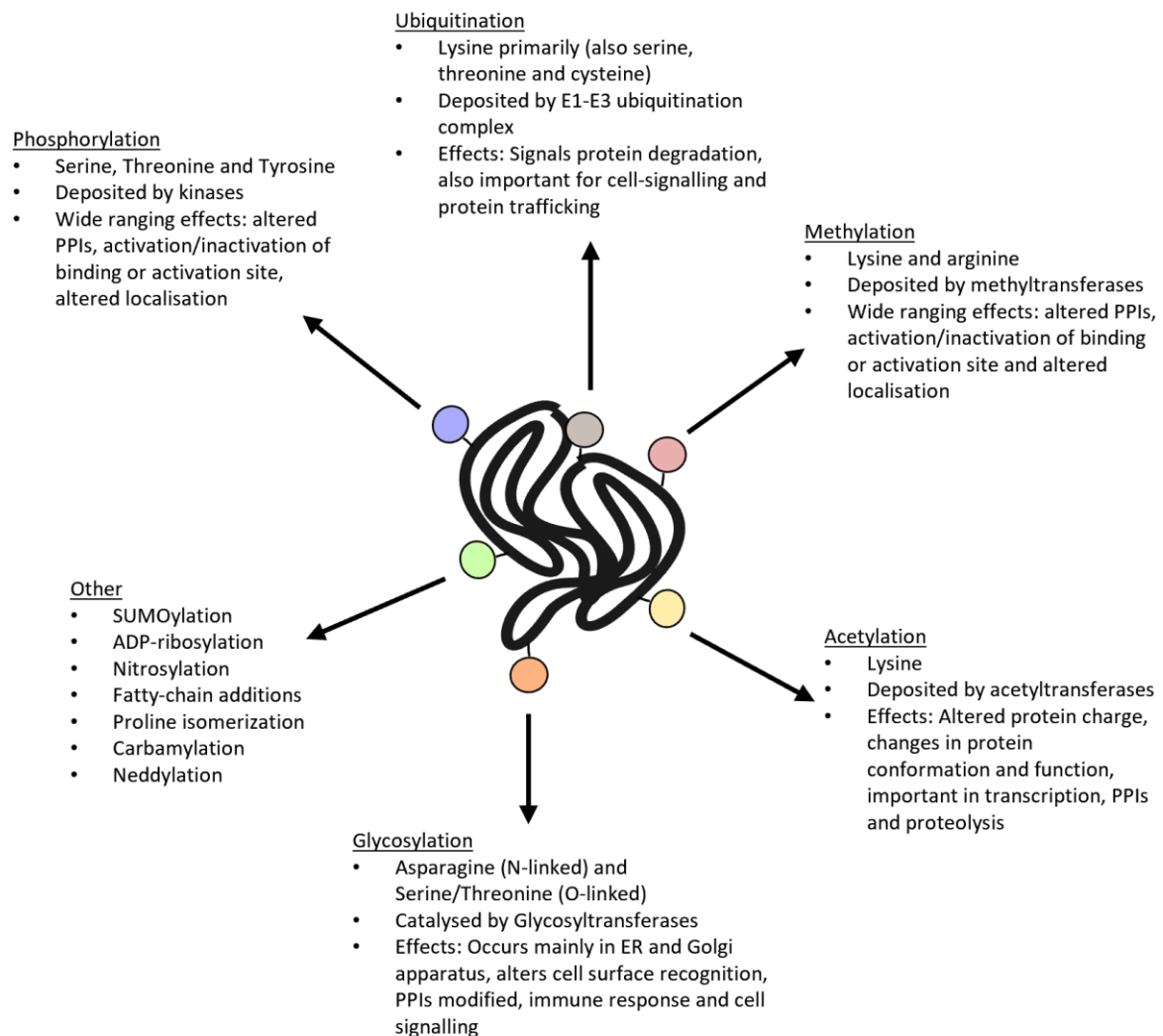


Figure 1.4 Post-translational modifications alter the activity of proteins: Schematic of various types of PTMs, how they occur and some of the effects they have on proteins. PPI = protein-protein interaction, ER = endoplasmic reticulum. Information from (Zhong *et al*, 2023)

1.3 ALTERNATIVE SPLICING

1.31 CONSEQUENCES OF ALTERNATIVE SPLICING

The previous description of splicing (the removal of introns and the joining of exons) is known as constitutive splicing and occurs due to the *cis* elements shown in Figure 1.2, which are recognised by the core splicing machinery. In higher eukaryotes, these splice sites have insufficient information to unambiguously define an authentic splice site from a false one (Wang & Burge, 2008). Other *cis*-acting sequences within a pre-mRNA molecule influence splicing via the binding of *trans*-acting splicing factors, which allow the splicing machinery to distinguish true exons over pseudo exons. (Graveley *et al*, 1998; Zhu *et al*, 2001; Zhang & Chasin, 2004). These *trans* factors can promote splicing, as just described, or they can inhibit the splicing machinery and cause it to miss exons or splice sites. This differential activity of splicing factors and other RBPs can lead to a process known as alternative splicing (AS) (Fig.1.5). There are various types of alternative splicing (Fig.1.5), which can have different outcomes. Alternative exon skipping/inclusion leads to the loss or gain of an exon, which may introduce different functional domains, active sites or binding regions thus altering activity. Alternative splice sites can change the length of an exon which can also alter binding sites or perhaps PTM sites, or in some circumstances it alters the open reading frame which leads to a truncated protein if the new reading frame contains an earlier stop codon. Splicing of the first and last exons is known to alter transcription start site and polyadenylation respectively. Additionally, AS can silence certain genes, which can be through intron retention where the intron is not spliced out, or it can be through the inclusion of an exon (through cassette exon splicing or alternative splice site usage) that has a premature termination codon (PTC) (Barash *et al*, 2010; Baralle & Giudice, 2017). Such exons are known as “poison” exons and if translated, would form a truncated version of the protein. These intron-retained and poison-exon transcripts are generally removed via the nonsense-mediated decay (NMD) RNA turnover pathway first, thus silencing these genes. In some cases, transcripts containing retained introns are still exported and translated, and such introns are known as exitrons (Shamnas *et al*, 2024). These exitrons have characteristics of both exons and introns and can confer different activities to proteins.

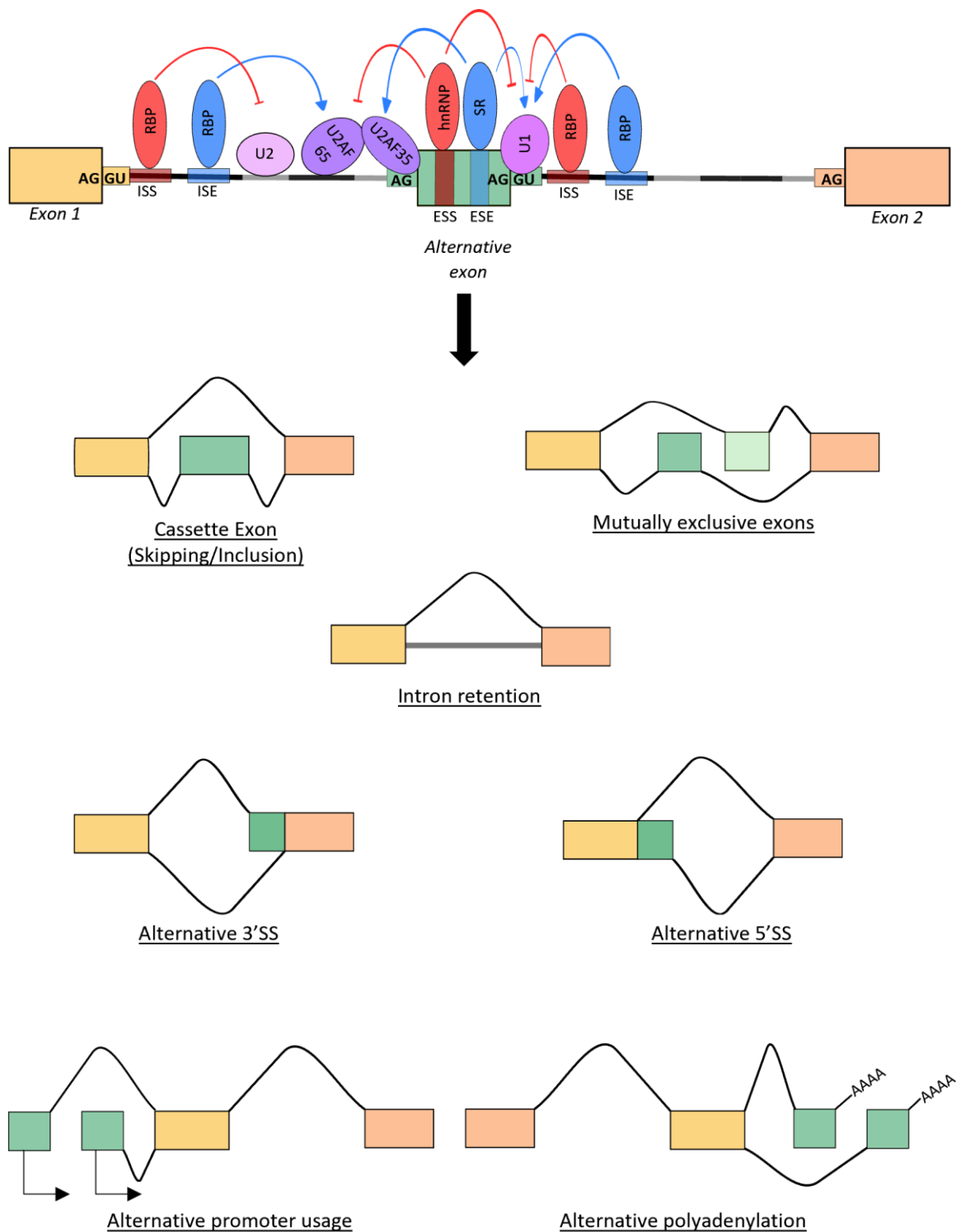


Figure 1.5 RBPs affect alternative splicing outcomes: Schematic of RBPs binding to pre-mRNA with differential effects on alternative splicing. The regulatory trans factors bind to their respective cis elements and either inhibit (red) or activate (blue) splicing via interaction with the core splicing machinery (pink/purple) leading to various alternative splicing outcomes (lower panel). ISS = Intronic Splicing Silencer, ISE = Intronic Splicing Enhancer, ESS = Exonic Splicing Silencer, ESE = Exonic Splicing Enhancer, RBP = RNA-binding protein, SR = serine/arginine rich splicing factor, 3'SS = 3' splice site, 5'SS = 5' splice site. Adapted from Park et al, 2018; Liu et al, 2022.

Alternative splicing results in multiple transcripts being produced from a single gene, and was first hinted at with the discovery of RNA splicing, when it was observed that the exons of the transcript from the Ad2 gene from Adenovirus were not contiguous in the mRNA sequence (Berget *et al*, 1977; Chow *et al*, 1977; Gilbert, 1978). Further examples of multiple transcripts from a single gene, such as the immunoglobulin μ gene giving rise to two mRNA transcripts μ_m and μ_s , demonstrated that this phenomenon was widely occurring (Early *et al*, 1980). Alternative splicing has since been shown to be almost ubiquitous in mammalian cells, and is thought to occur in ~95% of multi-exon genes in humans where it has been implicated in modifying PPI networks via the introduction of different binding domains, or altering PTM sites leading to different protein activities (Pan *et al*, 2008; Wang & Burge, 2008; Buljan *et al*, 2012; Ellis *et al*, 2012). Thus, alternative splicing can greatly increase the diversity of the proteome without requiring a large number of genes and may explain why higher eukaryotes do not have huge genomes compared to lower eukaryotes, such as yeast. The importance and use of AS has been found to vary between higher eukaryotes. A study comparing AS in *A. thaliana* versus animals (*C. elegans*, *D. melanogaster*, and *H. sapiens*) identified that plants generally use AS to regulate gene expression in response to stresses, whilst AS in animals appears to mainly be used for tissue-specific proteome production (Martín *et al*, 2021).

The idea that AS is functional is contested, as although many AS events (ASEs) occur, exactly how many of these ASEs lead to functional changes and how much is just “noise” is not known (Tress *et al*, 2017; Ule & Blencowe, 2019). It has been suggested that much of this “noise” of many seemingly neutral ASEs, may poise organisms for rapid evolutionary changes, allowing organisms to adapt quickly due to environmental changes (Modrek & Lee, 2003; Wright *et al*, 2022). Furthermore, it is widely accepted that conserved ASEs specific to certain cell-types largely have functional outcomes. For example, in neuronal differentiation, AS of disabled homolog 1 (DAB1) is regulated by neuro-oncological ventral antigen 2 (NOVA2). Here, NOVA2 represses the inclusion of exons 7b and 7c of DAB1 to produce a shorter isoform, which promotes proper brain development and neuronal migration, whereas if NOVA2 is knocked out, the longer isoform is instead produced and leads to neuronal migration defects (Yano *et al*, 2010; Baralle & Giudice, 2017). AS has been observed to be important for cell differentiation programmes of many different tissues, often affecting different groups of genes from transcriptional control, thus showing that these two mechanisms of gene regulation are complementary (Wang *et al*, 2008; Baralle & Giudice, 2017; Wright *et al*, 2022).

1.32 REGULATION OF ALTERNATIVE SPLICING

Various *cis* elements of RNA are known to affect alternative splicing by which *trans* acting factors they attract and where they are located. These sequences can be in the intronic regions of pre-mRNA:

intronic splicing silencers (ISS) which act to inhibit splicing, and intronic splicing enhancers (ISEs) which activate splicing (Wang & Burge, 2008). Similar sequences also exist in exons, exonic splicing silencers (ESSs) and exonic splicing enhancers (ESEs) (Fig.1.5). Splicing factors known as the SR family proteins are often recruited to ESEs and bind via their RNA recognition motifs (RRMs) and then attract and interact with spliceosomal proteins with their serine-arginine (SR) rich domains, generally assisting in the formation of the early (E) spliceosome. Splicing silencing sequences are instead predominantly bound to heterogeneous nuclear ribonucleoproteins (hnRNPs) which act to inhibit the formation of functional spliceosomes, for example by blocking the binding of U1 and U2 snRNPs. These SR proteins and hnRNPs can also compete for binding at similar locations, and the outcome of these competitions can depend on the location of a particular RNA sequences they are binding, the specificity of each protein, and on the relative abundance of the proteins (Zhu *et al*, 2001). The positive splicing sequences within exon and introns (ESEs and ISEs) are thought to help in exon/intron boundary recognition, as the 5'SS, 3'SS, branchpoint sequence and polypyrimidine tract were found to be insufficient for proper exon recognition (Mardon *et al*, 1987; Krainer *et al*, 1990; De Conti *et al*, 2013). Therefore, the antagonism between the different regulators of splicing can lead to differential recognition of an exon and thus determines the fate of a particular alternative exon.

Recognition of splice sites

Alternative exons are generally shorter than constitutive exons, thus they are favoured by the exon-definition model for splicing. Indeed, some exons are too short for conventional splicing to occur. These exons are known as microexons and are particularly prevalent in the neuronal network (Irimia *et al*, 2014). These microexons are 3 – 27 nucleotides (nt) in length, and their inclusion requires the binding of SRRM4 to an ISE where it antagonises the repressive function of the PTBP1 protein, and its enhancer of microexon domain (eMIC) interacts with SF1 and U2AF to recruit U2 snRNP to activate formation of the spliceosome (Ule & Blencowe, 2019). Many splicing factors act at the early stages of spliceosome assembly, with many SR proteins and hnRNPs affecting splice site recognition. SR proteins can recruit U1 snRNP to the 5'SS and promote 3'SS binding by interacting with the U2AF complex and U2 snRNP, generally by binding to ESEs (Fig.1.5) (Chen & Manley, 2009). Such interactions are mediated by their arginine-serine (RS) domains, which can be phosphorylated/dephosphorylated to alter their activity. These SR proteins can form large complexes with other splicing enhancer proteins such as TRA2, SRRM1 and SRRM2 which can further enhance splice site recognition. In addition, splicing activator proteins can bind at intronic regions to promote splicing, for example SAM68 binds to upstream intronic regions in a phosphorylation dependent manner, and recruits U2AF to the 3'SS (Tisserant *et al*, 2008). Inhibition of splice site recognition can act to sterically block access of snRNPs and pro-

splicing factors to their *cis* elements, for example hnRNP A1 binds to ISSs upstream of exon 3 in *Tat* of HIV to prevent U2 snRNP from binding (Tange *et al*, 2001).

Later stage regulation of the spliceosome

Alternative splicing regulation can also occur after the initial splice site recognition, when the exon definition spliceosome complex switches to an intron definition complex, even after U1 and U2 snRNPs have bound their splice site in the E-complex. The RBP hnRNP L has been shown to inhibit splicing by binding to an ESS of exon 4 of *CD45* pre-mRNA and preventing the cross-intron pairing of U1 and U2 snRNPs, leading to a stalled A-like spliceosome complex (House & Lynch, 2006). The splicing regulator RBM5 has also been shown to affect spliceosome assembly, by binding to exonic sequences in exon 6 of *CD95* pre-mRNA and inhibiting the association of the U4/U6-U5 tri-snRNP thus preventing spliceosome assembly progression (Bonnal *et al*, 2008). These examples show how splicing regulation can occur at various stages of spliceosome assembly, not just at splice site selection. Furthermore, the recent structural study of the cross-exon pre-B spliceosome complex demonstrates how selection of 5'SS by the pre-B complexes can lead to alternative splicing depending on which 5'SS the complex interacts with (Zhang *et al*, 2024). Canonical splicing would occur if the 5'SS interacted with was directly upstream from U2-triSNP complex. Competition between different U1-bound 5'SS to interact with the U2-triSNP complex can lead to alternative splice site usage or exon skipping, or even back splicing where a downstream splice site “donor” is linked to an upstream splice “acceptor” site.

Position-dependent activity of splicing regulators

Some proteins can act at further distances from an alternative exon, by multimerising along the RNA and thus blocking the spliceosome from being able to recognise an exon and assemble, as seen with some PTBP1 repressed exons (Spellman & Smith, 2006). PTBP1 has also been observed to bind at distal sites either side of alternative exon N1 of the *Src* pre-mRNA and promote exon skipping by forming a hyper-stable non-productive U1 snRNP:5'SS complex, meaning spliceosome assembly does not progress (Sharma *et al*, 2011; Wongpalee *et al*, 2016). RBPs can loop out the alternative exon by binding to sites upstream and downstream of the exon and forming protein-protein interactions, as has been observed for hnRNP A1 and its own pre-mRNA at exon 7 (Blanchette & Chabot, 1999). Here, binding to the splice sites by U1 snRNP or U2AF is not prevented, instead formation of the spliceosome across the looped out exon is thought to be sterically hindered and thus the exon is skipped (Chen & Manley, 2009). The location of the *cis* elements for splicing regulators can affect the mechanism by which they alter splicing, and these different mechanisms have been observed even for the same protein, for example with PTBP1 and hnRNP A1 (Ule & Blencowe, 2019). RBPs can act as both splicing activators and repressors depending on where on the RNA they bind. PTBP1 has been shown to have

this position-dependent activity. If its binding sites are located only downstream of an alternative exon, the exon is activated and spliced in (Llorian *et al*, 2010). This is thought to be due to PTBP1 promoting productive binding of U1 snRNP to the 5'SS of the alternative exon (Hamid & Makeyev, 2017). When PTBP1 is acting in a repressive manner, it was found to bind across the exon (Llorian *et al*, 2010). This is thought to prevent the splicing machinery from being able to recognise the exon, for example by blocking the binding of U2AF65 to the polypyrimidine tract, and thus the exon is skipped. Additionally, with the aforementioned c-Src exon N1, PTBP1 binds both upstream and downstream of the alternative exon and promotes the hyper-stable U1 snRNP-5'SS interaction and prevents further spliceosome assembly, thus leading to exon skipping (Hamid & Makeyev, 2017). The characteristics that distinguish PTBP1 activated exons from repressed ones appear to be the presence of PTBP1 binding mainly in the downstream intron of the activated exon, and the presence of a weak 5'SS. This suggests that PTBP1 recruitment to a weak 5'SS leads to productive binding and proper formation of the spliceosome, whereas U1 snRNP recruitment by PTBP1 to strong 5'SS leads instead to the hyper-stable complex which is non-productive. A study looking at positional dependent effects on alternative splicing used MS2 tagging to position SR proteins upstream of the 5'SS of an alternative exon or downstream (Erkelenz *et al*, 2013). The authors identified that if SR proteins were located upstream of the 5'SS they promoted inclusion and if they were downstream, they promoted exclusion. They also identified that with MS2 tagged hnRNPs at the same sites, the opposite splicing patterns were observed, i.e. an upstream location led to skipping and downstream led to inclusion, as was seen with PTBP1. This suggests that this position-dependent activity of AS regulators may be a common mechanism, and such activity has been identified for various RBPs such as NOVA, RBFOX, and the MBNL proteins (Ule & Blencowe, 2019). The picture can get more complex if the RBP is part of a larger complex which can bind to different motifs and have different effects on splicing, as is the case for the RBPs RBFOX1/2, which are part of the large assembly of splicing regulators (LASR) complex (Damianov *et al*, 2016; Jiang *et al*, 2021). The splicing activity of RBFOX has been seen to be position dependent, and due to it binding to a larger complex of proteins, it does not necessarily need to bind to its canonical GCAUG sites, or even directly to the RNA itself, to affect splicing (Zhou *et al*, 2021; Jacob *et al*, 2022). This demonstrates that the *cis* elements in the RNA work in combination with different *trans* factors to alter AS, and even if the canonical binding sites for a particular RBP are not present, they can still affect splicing by interacting with other co-regulator or antagonistic proteins. Additionally, the local concentration of certain RBPs will impact their distribution along their target RNAs and thus protein concentration can alter alternative splicing. The concentration of specific RBPs can change from cell-type to cell-type and in cellular extracts, leading to different alternative splicing patterns (Tange *et al*, 2001; Tao *et al*, 2024).

RNA structure can alter alternative splicing

As well as direct interaction between negatively regulating proteins and positively regulating proteins and the spliceosome, RNA structure can regulate alternative splicing. In some circumstances, RNA secondary structures can promote splicing by recruiting SR proteins, and in others it can prevent proper binding of core splicing factors (Blanchette & Chabot, 1999; Buratti *et al*, 2004). For example, a stem loop RNA secondary structure led to the sequestering of the alternative exon 6B of chicken β -tropomyosin, which led to its exclusion (Libri *et al*, 1992). RNA:RNA interactions can lead to specific exon inclusion/exclusion, as has been observed with the *Dscam* pre-mRNA in *D. melanogaster*, where its exon 6 cluster has 48 mutually exclusive exons. RNA:RNA pairing between a conserved sequence downstream of exon 5 and a sequence known as the selector sequence, which is found upstream of each exon 6, leads to the inclusion of only one of these variants (Olson *et al*, 2007; Chen & Manley, 2009). All other selector sequences are unable to bind due to the presence of the hrp36 protein which binds at the other selector sequences. Additionally, RNA secondary structures have been associated with the formation of circular RNAs (circRNA) through back-splicing. The proteins NF90 and NF110 bound at *Alu* repeat elements are thought to promote circRNA biogenesis by stabilising intronic RNA secondary structures to juxtapose the circRNA exon (Kristensen *et al*, 2019). These proteins have been found to be important in forming viral circRNAs upon viral infection which aids in immune signalling via the binding of RIG-I.

Co-transcriptional regulation of AS

As discussed previously, splicing is known to occur co- and post-transcriptionally, hence when splicing is coupled to transcription, alternative splicing outcomes can be affected by transcription. The rate of transcription elongation has been shown to affect some alternative splicing events (Mata *et al*, 2003; Nilsen & Graveley, 2010; Bentley, 2014). When an alternative exon has weak splice sites (non-canonical), if pol II is traversing the pre-mRNA quickly, the spliceosome may not have time to recognise and stably assemble upon these weaker SS. This results in the alternative exon being “missed” by the spliceosome and it is therefore skipped. This means that during fast transcription elongation, stronger splice sites are opted for by the spliceosome. During slow transcription elongation, the spliceosome has more time to recognise and assemble upon weaker splice sites which often co-exist with alternative exons. This means that these exons have a greater chance of being spliced in, as is the case for the Fibronectin exon EDI which has been shown to be included with slow pol II elongation rates (Mata *et al*, 2003). It can therefore be expected that some splicing factors may act to alter pol II elongation rates. This is what is observed for SR family protein SRSF2, which has been shown to interact with transcription elongation factor pTEFb (CDK9) to positively affect transcription elongation (Lin *et*

al, 2008; Ji *et al*, 2013). Additionally, different chromatin marks via histone PTMs can affect transcription rates and thus also affect alternative splicing, with areas of repressed chromatin leading to reduced pol II elongation rates and thus affecting alternative splicing outcomes (Schwartz *et al*, 2009; Agirre *et al*, 2021). Exons where slow elongation rates promote inclusion are known as class I exons, whereas those that are more likely to be included due to fast elongation are called class II exons (Marasco *et al*, 2023). An example of a class II exon is the *SMN2* exon 7, where transcription elongation was promoted by using antisense oligos (ASOs) against HDACs, and this led to the increased inclusion of exon 7. Thus, transcription rates and chromatin signatures are important in the regulation of alternative splicing.

Dysregulation of AS in disease

The dysregulation of AS networks has been found to be a key cause of disease and disease progression (Tao *et al*, 2024). Defects in splicing can be caused by mutated *cis* elements that lead to aberrant splicing, for example by disrupting proper 5'SS binding by U1 snRNP, or to the ESE/ESSs that splicing factors bind (Cartegni *et al*, 2002; Wang & Burge, 2008). Dysregulation can also be due to changes in the RBPs that modulate AS, for example by mutations that disrupt protein function, their expression, and their stability (protein or RNA). Spinal muscular atrophy (SMA) is caused by a loss of function of the *SMN1* gene whose expression normally produces the SMN protein which is required for assembly of snRNP complexes (Cartegni & Krainer, 2002; Kashima *et al*, 2007). An almost identical copy of this gene, *SMN2*, is opted for instead, which also produces the SMN protein, but this protein is truncated and less stable due to an exclusion event at exon 7. This was found to be due to C->T substitution between *SMN1* and *SMN2* in exon 7, which creates an ESS sequence for hnRNP A1 protein and prevents the binding of SRSF1, causing this exon to be skipped. Krainer *et al* have since created a drug called Sprinraza29 that blocks binding by hnRNP A7 to an ISS downstream of exon 7, promoting exon 7 inclusion and thus a functional protein is produced (Shorrock *et al*, 2018). RBPs themselves can cause or contribute to disease due to mutations or changes in their expression, and this has been found to be the case in many forms of cancer, where there is significant upregulation in the amount of AS events (ASEs) (Wang *et al*, 2022). RBPs that may normally act as pro-differentiation factors may be downregulated, meaning the events they normally control become dysregulated leading to uncontrolled proliferation and metastasis, as is the case for RBFOX2 and pancreatic cancer development (Maurin *et al*, 2023). The key role that RBPs and AS have in the pathogenesis of disease highlights how important these processes are for normal development and homeostasis.

1.33 CELL-TYPE SPECIFICITY AND ALTERNATIVE SPLICING

Given that alternative splicing decisions are affected by the competition between positive and negative splicing regulators, changes in the levels of such proteins can lead to changes in AS. This can give rise to tissue or cell-specific AS networks (Pan *et al*, 2008; Wang *et al*, 2008). AS can remodel PPI networks and alter PTMs between different tissue types, allowing a single gene to meet different requirements depending on cell needs (Ellis *et al*, 2012). Cell-type specific AS networks have been found to be important for both development of different tissue types, and in their homeostasis (Baralle & Giudice, 2017). Splicing events that led to “switch-like” changes between different cell types were correlated with higher levels of sequence conservation. This suggests that the regulation of cell-specific AS may be achieved by specific splicing regulators via the binding of a set of common *cis*-acting elements. Factors regulating alternative splicing such as transcription and RNA processing, translation rates and protein turnover, influence tissue specific AS networks. Expression of certain splicing regulators will be upregulated or downregulated in different cell-types or during different stages of development (Baralle & Giudice, 2017). These may be tissue-specific splicing factors, or core splicing factors. Specific disruption of the genes encoding core splicing proteins SF2 and SC35 in the heart demonstrated that they were playing an important role in tissue development (Ding *et al*, 2004; Xu *et al*, 2005). Additionally, mutations in core splicing proteins PRPF31, PRPF8, PRPF3 and RP9 which form part of the U4/U6-U5 tri-snRNP complex cause autosomal dominant retinitis pigmentosa (RP) (Cooper *et al*, 2009). This demonstrates how tissue specific changes can occur through modulation of the core splicing machinery.

Alternative splicing itself can alter AS networks by altering the splicing activity of other splicing factors, via changing the protein, or RNA binding domains, or PTM sites, and can also alter the levels of splicing factors via AS NMD (Barash *et al*, 2010; Jangi & Sharp, 2014). These AS networks are important for remodelling cells to allow them to perform specific functions and can be controlled by specific RBPs whose expression is upregulated with environmental or developmental cues. This is the case in cardiac cells during the transition from fibroblasts to myofibroblasts during wound healing, where the protein MBNL1 acts as a pro-differentiation factor by regulating AS (Bugg *et al*, 2022; Acharya *et al*, 2024). MBNL1 regulates differentiation specific transcripts and transcripts that remodel the cytoskeleton to aid in the remodelling of fibroblasts to myofibroblasts, targeting proteins such as the serum response factor (SRF) and Calcineurin A β , which play important roles in cardiac wound healing. MBNL1 has also been found to be important in cardiomyocyte development and homeostasis, where it is also thought to be playing a role in AS regulation, and its dysregulation has been linked to disease progression of myotonic dystrophy I (DM1) (Philips *et al*, 1998; Lee *et al*, 2022). AS-NMD pathways are often observed

in autoregulatory loops to negatively or positively regulate the expression of RBPs in a tissue specific manner, and has been observed for many proteins, including the SR proteins (García-Moreno & Romão, 2020; Ding *et al*, 2022). In the brain, proteins PTBP1, PTBP2 and SRRM4 interact and regulate each other as their levels change during brain development. In early stages of development, PTBP1 levels are high and it negatively regulates PTBP2 by causing the skipping of exon 10 which results in the introduction of a PTC, leading to AS-NMD of PTBP2 mRNA (Boutz *et al*, 2007; Spellman *et al*, 2007). During neuron differentiation, PTBP1 is downregulated and PTBP2 exon 10 is activated by SRRM4, and PTBP2 levels increase which aids in neuronal development and maintenance (Raj *et al*, 2014). The protein RBFOX2 has also been shown to play an important role in AS-NMD of various other RBPs, including PTBP2, meaning this protein has far-reaching effects on different AS networks depending on its expression (Jangi *et al*, 2014; Jangi & Sharp, 2014).

1.4 ALTERNATIVE SPLICING IN VASCULAR SMOOTH MUSCLE CELLS

1.41 VASCULAR SMOOTH MUSCLE CELLS HAVE PHENOTYPIC PLASTICITY

Vascular smooth muscle cells (VSMCs) line the walls of blood vessels and have an elongated spindle shape which aids in their principal function, contraction to maintain the blood pressure and structural stability of blood vessels (Kao & Carsten, 1997; Frismantiene *et al*, 2018). These VSMCs exhibit cellular plasticity, whereby mature and differentiated cells undergo phenotypic switching, transiently becoming de-differentiated and more proliferative (Fig.1.6) (Owens, 1995; Owens *et al*, 2004). This phenotypic plasticity is dependent on environmental cues such as damage to the wall of the blood vessel, and the cells de-differentiate to promote the renewing and repairing of tissue. The de-differentiated cells lose their spindle shape and the ability to contract, and instead become more motile, produce more extracellular matrix (ECM), and secrete migratory and proliferation factors. The phenotypic switching of VSMCs involves many changes, such as reduced calcium signalling, cytoskeletal changes and alteration of the focal adhesion machineries. As the VSMCs de-differentiate, the levels of SMC-specific markers, such as SM- α -actin, SM-myosin heavy chain and *h*-caldesmon, decrease (Frismantiene *et al*, 2018). Phenotypic switching occurs at low levels in adult tissue, and if an injury occurs and VSMCs undergo de-differentiation, they switch back to their differentiated state once the repair has been achieved. Consistent and unregulated VSMC phenotypic switching, however, has been linked to various vascular diseases. Specifically, the phenotypic transition of VSMCs has been linked to disease-progression of atherosclerosis and restenosis, both of which result in the narrowing of blood vessels and can lead to myocardial infarction (Shankman *et al*, 2015; Frismantiene *et al*, 2018). Given that cardiovascular disease (CVD) poses the biggest mortality risk to adults in both Europe and

the United States, it is important to identify the regulatory mechanisms that are behind the phenotypic plasticity of VSMCs (Timmis *et al*, 2024; Woodruff *et al*, 2024).

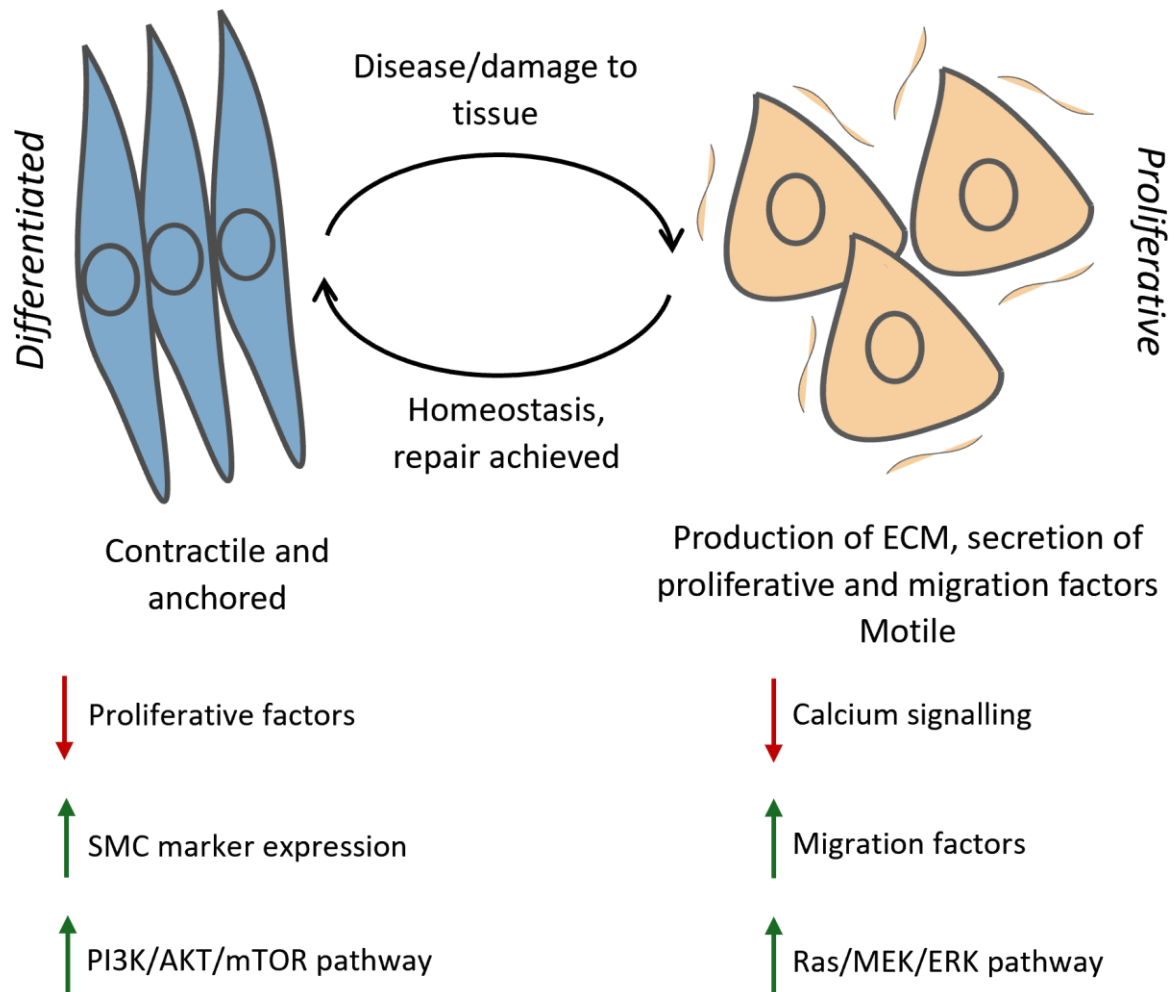


Figure 1.6 VSMCs exhibit phenotypic plasticity: Schematic of de-differentiation of vascular smooth muscle cells and associated changes. Red arrows represent downregulation, green arrows represent upregulation. ECM = extracellular matrix. Adapted from Nakagaki-Silva *et al*, 2019.

1.42 ALTERNATIVE SPLICING AND THE REGULATION OF VSMC PLASTICITY

Signalling pathways and transcriptional regulation of VSMC plasticity

The phenotypic switching of VSMCs is highly regulated, as is the maintenance of the steady state of VSMCs to ensure that the contractile cells perform their function properly. Pro-differentiation signalling is thought to occur through various signalling pathways including via the cytokine TGF β and the PI3K/AKT/mTOR pathways (Frismantiene *et al*, 2018). De-differentiation signalling is thought to occur either through the Ras/MEK/ERK pathway and is triggered by the growth factor PDGF; or through β -catenin signalling. These signalling cascades act through a variety of ways from silencing/activating

transcription through to altering translation and PTMs. One of the most well-characterised modes of regulation of phenotypic switching of VSMCs is at the transcription level. The changes in transcription are activated by specific transcriptional activators, such as Myocardin (MYOCD) and the Serum response factor (SRF) (Allahverdian *et al*, 2018). SRF is widely expressed in many cell-types and its role modulating VSMC plasticity relies on MYOCD. SRF and MYOCD form a complex and bind to CArG boxes near the promoters of target genes, to promote VSMC differentiation and thus the contractile phenotype is produced. MYOCD is antagonised by KLF4, which competes for binding at CArG sites, where it represses the expression of differentiation genes (Miano, 2015; Frismantiene *et al*, 2018). Epigenetic regulation also plays a role in phenotypic switching through programmed remodelling of chromatin, specifically at the CArG elements of SMC differentiation markers where the histone proteins H3 and H4 are acetylated by histone acetyltransferases (HATs) to allow binding of SRF-MYOCD complexes and promote differentiation (McDonald & Owens, 2007; Allahverdian *et al*, 2018). These histones get deacetylated during PDGF-mediated de-differentiation via KLF4 recruitment of histone deacetylases (HDACs). This limits accessibility to CArG boxes for MYOCD-SRF, whereas KLF4 and associated proteins bind upstream to G/C elements and repress expression of SMC differentiation marker genes.

AS and its regulation of VSMCs

Alternative splicing is another mechanism by which VSMC maintenance and phenotypic switching is regulated, and has long been viewed as an important mode of control of VSMC differentiation/de-differentiation, although these processes are not as well-characterised as transcriptional regulation (Owens, 1995; Owens *et al*, 2004; Brozovich *et al*, 2016). Many of the SMC differentiation markers have been found to be alternatively spliced, including the myosin heavy and light chain proteins, caldesmon, vinculin, α -actinin, and tropomyosin (Nabeshima *et al*, 1987; Wieczorek *et al*, 1988; Lenz *et al*, 1989; Owens, 1995). The AS of these genes can either produce a non-muscle (NM) variant or a smooth-muscle (SM) variant depending on cell-type, where the SM isoforms are selected for in mature differentiated VSMCs. These SM variants are thought to promote the normal contractile function of VSMCs and thus when they are lost through altered AS during de-differentiation, the cells lose these functions. The Myosin phosphatase target 1 (*MYPT1*) is alternatively spliced to produce four variants which have different Ca^{2+} sensitivities, allowing VSMCs to have a range of contractile behaviours. It has a leucine-zipper (LZ) domain which is created by the exclusion of alternative exon 23, whereas inclusion of this exon under the regulation of TRA2 β causes a loss of this LZ domain (Zhang & Fisher, 2007; Shukla & Fisher, 2008). High expression of TRA2 β and hence the isoform lacking the LZ domain has been linked to hypertension, thus demonstrating AS may prove to be a useful therapeutic target for the treatment of CVD. The SMC transcription factor MYOCD itself undergoes AS to produce a SM specific isoform,

where exon 2a contains a PTC which is very close to the start codon, thus another start codon must be utilised instead and a N-terminally truncated protein is produced (Creemers *et al*, 2006; Miano, 2015). If this exon is skipped, then a cardiac specific isoform is produced which binds to an interaction partner MEF2 and activates cardiac specific genes.

The above examples indicate the importance of AS for the phenotypic switching of VSMCs and demonstrate that this process cannot rely on transcription alone. In fact, the changes in AS events and transcription upon de-differentiation were found to affect different sets of genes (Llorian *et al*, 2016). The ASEs were primarily within cytoskeletal and focal adhesion genes enabling the VSMCs to change from a contractile state to a more motile one. This study identified that cassette exon (CE) ASEs were the most common event changed between differentiation and de-differentiation, but the overall number of CE events did not change between each state, whilst intron retention (IR) events were highly skewed towards differentiated cells. The IR events, as well as the inclusion of poison exons were found to be common in snRNP proteins in the differentiated state, as well as in transcriptional regulators. These splicing events are non-productive and result in NMD, showing that this mode of gene down-regulation is a frequent way in which gene expression can be modulated beyond transcriptional control in differentiated VSMCs. The cassette exons that were included in the differentiated state were found to be shorter than typical CEs and had weak splice sites. Additionally, they were found to have PTBP1 binding sites in the upstream intron, which has previously been shown to be how PTBP1 represses exons, which was confirmed with PTBP1 knockdown (Spellman *et al*, 2007; Llorian *et al*, 2010, 2016). Canonical and non-canonical binding sites for Quaking (QKI) were found surrounding both the differentiated and proliferative upregulated exons, with their position suggesting QKI may be playing a role in both SM exon activation and repression. QKI was previously thought to promote de-differentiation as it can repress exon 2a of *MYOCD*, and thus its role in VSMC phenotypic plasticity remains unclear (Veer *et al*, 2013). Additionally, the role of PTBP1 in VSMC homeostasis or phenotypic switching is confounding, as for some events it was seen to repress the SM exon, and in others its action results in its inclusion. Two mutually exclusive events show this opposite action of PTPBP1, *Actn1* and *Tpm1* (Southby *et al*, 1999; Gooding *et al*, 2013). The SM exon of *Actn1*, which produces a non-functional Ca²⁺ binding EF hand to stabilise α -actinin, is repressed by PTPBP1 in favour of the NM exon being spliced in. The opposite is true for *Tpm1*, where its NM exon is instead repressed by PTPBP1. In addition, the knockdown of PTBP1 and PTBP2 was not found to be sufficient for switching splicing patterns to differentiated levels. These results with PTBP1 and QKI demonstrate that there may be other proteins involved in the differentiated AS network in VSMCs which prevents the repression of SM variants.

1.5 RBPMS: A MASTER SPLICING REGULATOR IN VSMCs

1.51 THE ROLE OF RBPMS IN VSMCs

RNA binding protein RBPMS was identified in the lab as being a master regulator of alternative splicing in VSMCs (Nakagaki-Silva *et al*, 2019). Unlike the other RBPs identified in the previous VSMC de-differentiation study, it was found to act specifically in differentiated VSMCs as it was downregulated upon de-differentiation (Fig.1.7) (Llorian *et al*, 2016; Nakagaki-Silva *et al*, 2019). In differentiated cells, RBPMS acts to promote ASEs that are important for VSMC contractility, such as within cytoskeletal genes and focal adhesion genes. The postulated role of master splicing regulators in different tissues was proposed by Jangi and Sharp who state that such proteins would act to either control cell differentiation/de-differentiation, or as regulators of tissue homeostasis of the differentiated state (Jangi & Sharp, 2014). The authors reason that if the master regulator was removed, differentiation or de-differentiation would not properly occur. The master regulator would target genes that are important for cellular function, as well as transcriptional regulators and splicing regulators thus creating a large network of gene expression control. Jangi and Sharp also suggested that such master regulators can be identified if they are under the transcriptional control of a super-enhancer. RBPMS was found to be under the control of a super-enhancer and was highly expressed in differentiated VSMCs (Nakagaki-Silva *et al*, 2019). ASE changes upon RBPMS knockdown in differentiated VSMC model cell line PAC1 cells was compared with the ASEs from PAC1 cell de-differentiation, and it was found that RBPMS was fully responsible for 20% of the AS changes between the differentiated and proliferative state. These targets included other splicing regulators such as *Mbnl1* and *Rbfox2*; transcriptional regulators including the key SMC TF *Myocd*, and SMC markers *Tpm1*, *Actn1*, *Vcl* (Vinculin) and *Cald1* (Caldesmon 1). Therefore, RBPMS was said to be acting as a master splicing regulator in VSMCs, where it is thought to be important in maintaining the differentiated state of these cells. A recent study from the lab further demonstrated this by identifying that the overexpression of RBPMS promoted the contractile phenotype of VSMCs derived from human embryonic stem cells (hESCs) (Jacob *et al*, 2022). As before, RBPMS was found to potently drive smooth muscle specific AS programme via modulation of the actin-cytoskeleton and focal adhesion. Further RBPMS overexpression in these hESC-derived VSMCs led to reduced cell motility and proliferation, thus they behaved similarly to mature and differentiated VSMCs. RBPMS was also found in this study to functionally interact with another splicing regulator, RBFOX2, and together they mediate SM-specific AS at a subset of targets. Moreover, their binding sites were found to overlap surrounding and within target exons. As with many other RBPs that are able to regulate both activation of exon inclusion and repression, RBPMS was suggested to have a positional dependence on its splicing regulation (Nakagaki-

Silva *et al*, 2019; Jacob *et al*, 2022). Suggested RBPMS binding sites of repressed exons occurred upstream and within the alternative exon, whilst RBPMS binding sites associated with activated exons mainly occurred downstream of the alternative exon.

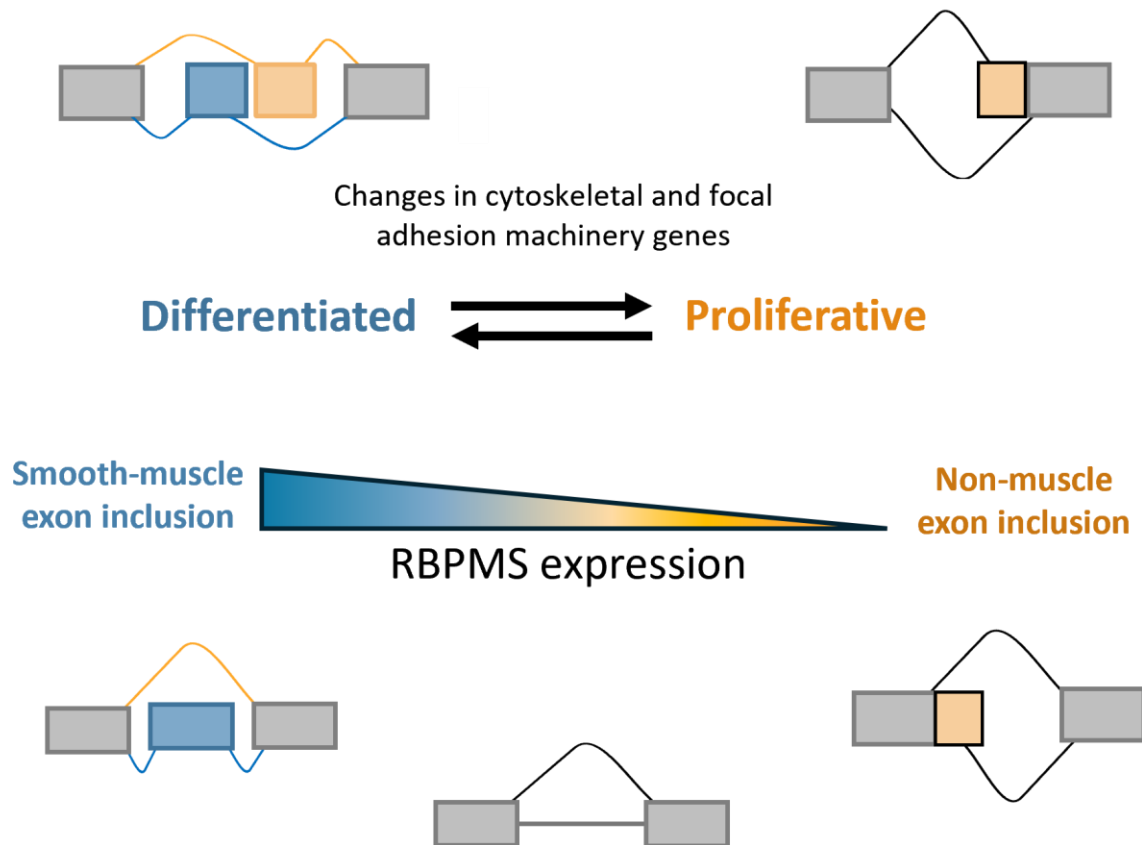


Figure 1.7 RBPMS is a master splicing regulator of differentiated VSMCs: Alternative splicing changes were found to affect a different set of genes than transcription upon de-differentiation, and RBPMS was identified as the master regulator of alternative splicing in the differentiated state. RBPMS is downregulated upon de-differentiation and is responsible for 20% of the ASEs changed.

1.52 NON-VSMC FUNCTIONS OF RBPMS

Prior to its role in VSMCs, RBPMS had not been observed to be involved in splicing, although it was found to undergo alternative splicing itself to produce multiple mRNA and protein isoforms (Shimamoto *et al*, 1996). A similar protein with 76% sequence similarity to the human RBPMS was also identified in *X. laevis* in a cDNA screen, which was named Hermes and has since been identified as a homologue of RBPMS, also known as RBPMS2 (Gerber *et al*, 1999; Akerberg *et al*, 2019).

The discovery of RBPMS2 (Hermes) in *X. laevis* identified that RBPMS2 was predominantly expressed in the embryonic heart, but the role it was playing was unknown (Gerber *et al*, 1999). The first known

function of RBPMS was in transcriptional regulation, where it interacts with the Smad family of transcription factors to enhance transcription, and was also found to bind to RNA *in vivo* (Sun *et al*, 2006). Since this discovery, the RBPMS family proteins have been found to play a variety of roles in regulation gene expression from AS to translation (reviewed in Akerberg *et al*, 2019). RBPMS2 has been found to promote the maturation of oocytes in *X. laevis* via 3'UTR binding to *nanos* RNA, which sequesters it leading to translational repression (Song *et al*, 2007; Agüero *et al*, 2016). RBPMS2 has also been found to be important in another type of SMCs in gastrointestinal tissue, where it is thought to promote de-differentiation via translational regulation (Notarnicola *et al*, 2012; Sagnol *et al*, 2014). Both RBPMS and RBPMS2 are highly expressed in cardiomyocytes and are important for heart development, with deficient expression of either resulting in improper cardiac function and impaired alternative splicing of a number of targets including *RBFOX2* and *Pdlim* (Gerber *et al*, 1999; Akerberg *et al*, 2022; Bartsch *et al*, 2023; Gan *et al*, 2022, 2024). RBPMS was also shown to be important in translational regulation of mRNA in human embryonic stem cells (hESCs) where it was found to play a role in cardiomyocyte determination (Bartsch *et al*, 2023). Gan *et al* recently showed the importance of RBPMS in the proper function of adult heart tissue in mice and promoted cardiomyocyte contraction via the AS mediated modulation of sarcomeric genes (Gan *et al*, 2024). Furthermore, RBPMS was found to interact with a number RBPs found to be important in heart development, including *MBNL1/2*, *RBFOX2*, *RBM2*, and the RBPMS homologue *BPMS2*.

RBPMS and disease

As well as the role of RBPMS in VSMCs and cardiac cells, it has been implicated in disease, primarily in cancer. RBPMS has been observed to act as a tumour suppressor, which was first observed for breast cancer (Fu *et al*, 2015). Here, two isoforms of RBPMS, RBPMS-A and RBPMS-B (named in this study as RBPMS1-A and RBPMS1-C) were found to inhibit breast cancer cell growth by regulating cFos and Smad mediated AP-1 signalling, where AP-1 normally acts to progress tumour growth but is inhibited by RBPMS. RBPMS has since been found to play similar tumour suppression roles in ovarian and bladder cancer, and also increases sensitivity to cancer drugs Cisplatin and Gefitinib (Rabelo-Fernández *et al*, 2022a, 2022b; Yao *et al*, 2023; Gong *et al*, 2024). RBPMS was identified as playing a general role in tumour suppression through its association with NOTCH2 mRNA, by promoting its de-adenylation, thus inhibiting cancer cell proliferation (Yu *et al*, 2023). Additionally, decreased RBPMS expression was found to be linked to the pathogenesis of osteoarthritis, where it is regulated by TGF- β and IL-1 β (Shanmugaapriya *et al*, 2016). From this finding, it was deemed likely that RBPMS plays an important role in the maintenance of normal cartilage. The importance for RBPMS in heart development and function also identified that loss of *Rbpms* led to splicing defects and major contractile issues which caused dilated cardiomyopathy (DCM), and was lethal (Akerberg *et al*, 2022; Gan *et al*, 2022, 2024; Shi,

2024). Furthermore, *Rbpms*^{-/-} mice showed major failure in their SMC development. Given the importance of RBPMS in maintaining regular function across a variety of tissues, particularly in cardiac cells where it is regulating the AS network, it is likely that RBPMS is playing a similarly important role in VSMCs. Hence, it is important to elucidate the mechanisms by which RBPMS controls AS and differentiation to understand how this may be dysregulated in disease.

1.53 THE STRUCTURE – FUNCTION RELATIONSHIP OF RBPMS

RBPMS contains a single RNA recognition motif (RRM), located at its N-terminus (Fig.1.8), followed by an intrinsically disordered and proline rich C-terminus. Two major isoforms of RBPMS were discovered in the differentiated PAC1 cells, RBPMS-A and RBPMS-B, which differ only in their extreme C-terminal tails, after residue 177 (Fig.1.8) (Nakagaki-Silva *et al*, 2019). The RRM of RBPMS was found to bind to a pair of tandem CAC repeats separated by 1 – 12 nucleotides (CAC_{n1-12}CAC) initially through PAR-CLIP studies and then subsequently through crystallographic studies (Farazi *et al*, 2014; Sagnol *et al*, 2014; Teplova *et al*, 2016; Soufari & MacKereth, 2017). These crystallographic studies demonstrated that RBPMS bound to its RNA as a homodimer, which was mediated by the RRM domain (Fig.1.8). The homodimerization interface involves the residues N34, K36, E39 and N87, and the interaction with RNA is mainly mediated by the C-terminal loop of the RRM, particularly with K100, and the β -strand residues, E97 and F27. These structural studies used a C-terminally truncated form of RBPMS, without its C-terminal tail, as this region is predicted to be an intrinsically disordered region (Fig.1.8) and produced heterogenous structures not amenable to crystallography (Farazi *et al*, 2014; Teplova *et al*, 2016). The authors noted that although this truncated RBPMS was mainly dimeric, they did observe oligomers likely to be tetrameric via gel filtration. They hypothesised that based on the closeness of several tandem CAC pairs in the target RNA, RBPMS likely undergoes oligomerisation. Two studies carried out recently in the laboratory, showed that RBPMS can indeed undergo oligomerisation *in vitro*, forming oligomers of ~27 subunits (Barnhart *et al*, 2022; Yang *et al*, 2023). Yang *et al* showed that this oligomerisation behaviour requires the final twenty amino acids of RBPMS-A, as a C-terminal deletion mutant (Δ C20) was only able to form dimer/trimers (Fig.1.8). This mutant also lost splicing repression function with RBPMS target *Tpm1* exon 3, which suggested a link between this self-oligomerisation behaviour and the function of RBPMS-A. A proteomic study of RBPMS-A in HeLa nuclear extract revealed that RBPMS-A binds to many RBPs known to be involved in splicing, from core splicing proteins such as SF1, SRSF1 and SF3B, and other alternative splicing regulators such as RBFOX2 and MBNL1. This study showed many similar interacting partners to the proteomic study in cardiomyocytes by Gan *et al* (Gan *et al*, 2024). The Δ C20 mutant, however, did not interact with most of these proteins, showing that it had an impaired PPI network. This demonstrated that the final twenty amino acids of

BPMS-A are also necessary for mediating heterotypic interactions. BPMS-A has also been observed to form liquid-liquid phase separation (LLPS) droplets *in vitro*, and this droplet formation was found to be inhibited by 1,6-hexanediol and high salt, suggesting that LLPS of BPMS-A involves weak hydrophobic interactions and electrostatic interactions (Peacey, Part III dissertation 2020). BPMS has previously been observed to enter stress granules in the cytoplasm of cells during oxidative stress, suggesting it has some propensity for entering phase separated foci in cells (Farazi *et al*, 2014; Teplova *et al*, 2016). Additionally, higher order assembly of BPMS-A has been suggested to be regulated via PTM of its IDR (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019; Barnhart *et al*, 2022). BPMS-A has been shown to be phosphorylated via ERK on residues T113 and T118 (Fig.1.8) and phosphomimetic studies suggested that this phosphorylation decreased the activity of BPMS for both activation and repression of exon inclusion. It was also suggested to decrease self-oligomerisation and RNA binding (Barnhart *et al*, 2022). This further implies that the higher order assembly of BPMS-A and its splicing activity are linked. This study also gave an insight into how BPMS activity might be regulated via PTMs, with ERK activation being triggered by PDGF which is associated with VSMC de-differentiation and cellular stress.

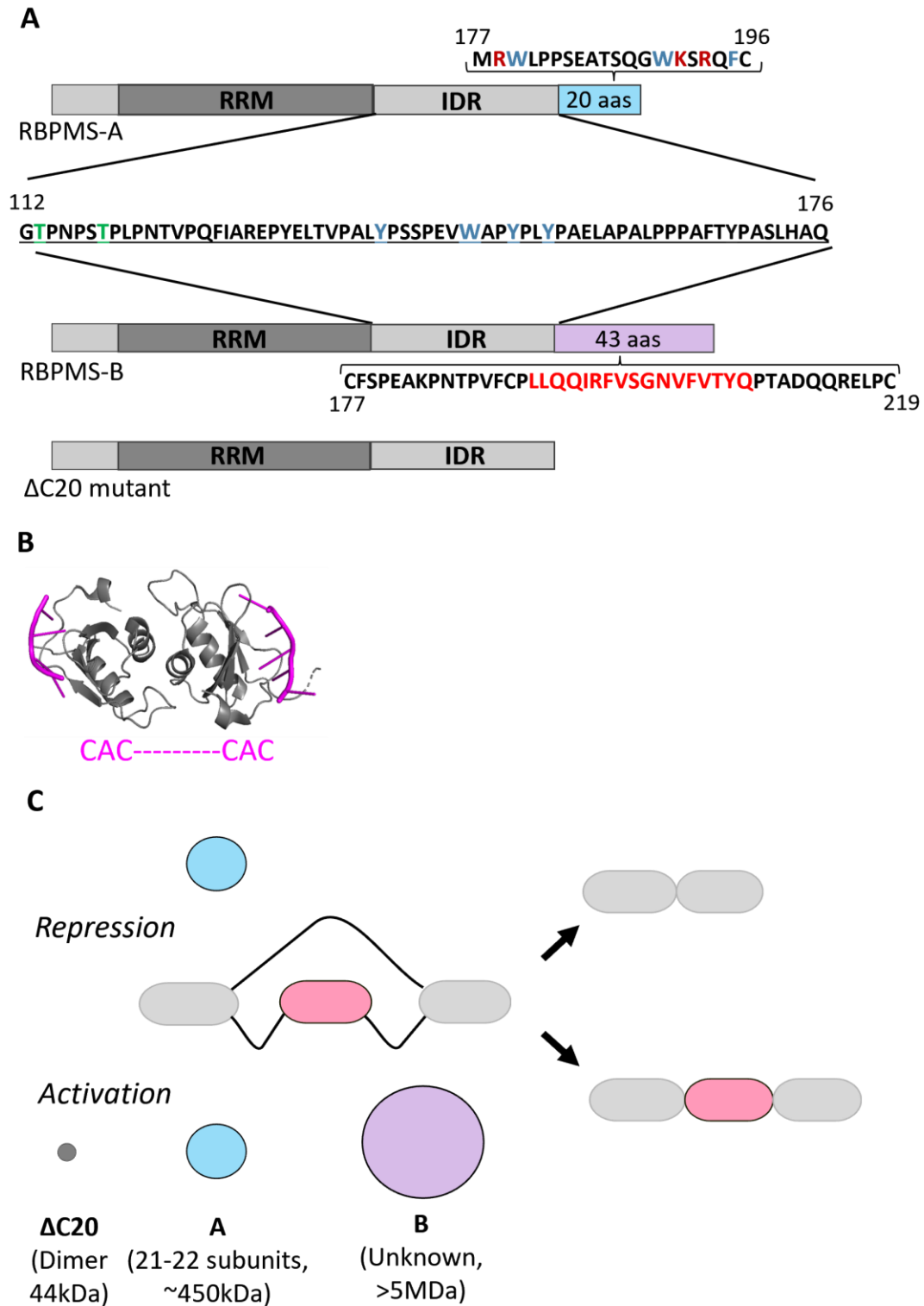


Figure 1.8 Oligomerisation of RBPMS requires its extreme C-terminus: **A:** Domain structure of RBPMS-A, RBPMS-B and ΔC20 deletion mutant. Aromatic (blue), basic (dark red) residues are highlighted along with threonines 113 & 118 (green), and the hot spot for aggregation region of RBPMS-B (red). **B:** Crystal structure of RNA recognition motif (RRM) of RBPMS as a dimer and bound to tandem CAC repeat (magenta), crystal structure from Teplova et al, 2016. Modelled in pymol. **C:** Schematic of splicing activity of RBPMS isoforms and ΔC20 mutant with exon activation or repression, approximate or suspected sizes of each protein are indicated by size of the circles, from in vitro data (Yang et al, 2023; Yang, unpublished).

As well as RBPMS-A and RBPMS-B, the homologue protein RBPMS2 was also found to be expressed in PAC1 cells. This protein, however, was found to be mainly expressed as a single isoform. RBPMS-A shares greater sequence conservation with RBPMS2 than RBPMS-B does at its C-terminus (~70% versus ~60%) (Fig.1.9). Additionally, transfected RBPMS2 was found to have very similar activity to RBPMS-A across various targets, whereas RBPMS-B had weaker splicing activity, particularly with repression targets. This would suggest that RBPMS-A may be the more ancestral form of RBPMS (Fig.1.9B). The Δ C20 mutant, although having impaired ability to repress targets exons compared to RBPMS-A, still had higher activity than RBPMS-B (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019). This finding suggested that the unique C-terminus of RBPMS-B acts antagonistically towards the splicing repression function of RBPMS-B. Fusion of the RBPMS-B C-terminus onto RBPMS-A caused a loss of full repressive activity even greater than the loss of the C-terminus Δ C20 mutant, whereas fusion of the RBPMS-A C-terminal tail onto RBPMS-B allowed RBPMS-B/A to gain repressive activity. These results further showed that the C-terminus of RBPMS-B antagonises splicing activity. Analysis of the RBPMS-B C-terminus with an online aggregation tool (AGGRESCAN) found a stretch of 17 amino acids that was predicted to be a “hot spot for aggregation” (HSA) (Fig.1.8A) (Conchillo-Solé *et al*, 2007; Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019). When this HSA region was deleted, the activity of RBPMS-B increased to levels similar to A, i.e. it could now act as a repressor. Initial attempts to purify the recombinant RBPMS-B protein showed that it was mainly insoluble and formed complexes of >7MDa (Fig.1.8), suggesting that this isoform of RBPMS may have even more extreme higher order assembly behaviour than RBPMS-A. How this potential for extreme higher order assembly is linked to its activity has not been elucidated.

A

```

RBPM2 Xenopus laevis MSGIKSDTEH-----NNNNIEEEVRTL FVSGLPIDIKPRELYLLFRPFKGYEGS 49
RBPM2 Mus musculus MSNLKPDVEHC-----TGAGTGSPL EEEVRTL FVSGLPVDIKPRELYLLFRPFKGYEGS 54
RBPM2 Homo sapiens MSNLKPDGEHGGSTGTGSGAGSGGAL EEEVRTL FVSGLPVDIKPRELYLLFRPFKGYEGS 60
RBPM2A Mus musculus -----MNGGGKAEKENTPSEANLQEEEVRTL FVSGLPDIKPRELYLLFRPFKGYEGS 53
RBPM2A Homo sapiens -----MNGGGKAEKENTPSEANLQEEEVRTL FVSGLPDIKPRELYLLFRPFKGYEGS 53
RBPM2B Mus musculus -----MNGGGKAEKENTPSEANLQEEEVRTL FVSGLPDIKPRELYLLFRPFKGYEGS 53
RBPM2B Homo sapiens -----MNGGGKAEKENTPSEANLQEEEVRTL FVSGLPDIKPRELYLLFRPFKGYEGS 53
                                     : . *****:*****

RBPM2 Xenopus laevis LIKLTSPQVGFVTFDNRAGAEAAKNALNGIRFDPENPQTLRLLEFAKANTKMAKNKLMAT 109
RBPM2 Mus musculus LIKLTSPQVGFVIFDSRAGAEAAKNALNGIRFDPENPQTLRLLEFAKANTKMAKSKLIAT 114
RBPM2 Homo sapiens LIKLTARQPVGFVIFDSRAGAEAAKNALNGIRFDPENPQTLRLLEFAKANTKMAKSKLMAT 120
RBPM2A Mus musculus LIKLTSPQVGFVFSFDSRSEAEAAKNALNGIRFDPEIPQTLRLLEFAKANTKMAKNKLVGT 113
RBPM2A Homo sapiens LIKLTSPQVGFVFSFDSRSEAEAAKNALNGIRFDPEIPQTLRLLEFAKANTKMAKNKLVGT 113
RBPM2B Mus musculus LIKLTSPQVGFVFSFDSRSEAEAAKNALNGIRFDPEIPQTLRLLEFAKANTKMAKNKLVGT 113
RBPM2B Homo sapiens LIKLTSPQVGFVFSFDSRSEAEAAKNALNGIRFDPEIPQTLRLLEFAKANTKMAKNKLVGT 113
*****:***** **.*: ***** ***** *****:*****

RBPM2 Xenopus laevis PNPTNFHPALGAHFIARDPYDFTGAALIPASPEAWAPYPLYTAE LAPAIPH-AAFTYPAA 168
RBPM2 Mus musculus PNPTSVHPALGAHFIARDPYDLMTALIPASPEAWAPYPLYTTELTPAISH-TTFTYPAA 173
RBPM2 Homo sapiens PNPSNVHPALGAHFIARDPYDLMTALIPASPEAWAPYPLYTTELTPAISH-AAFTYPTA 179
RBPM2A Mus musculus PNPSTPLPNTVPQFIAREPYELTVPALYPSSPEVWAPYPLYPAELAPALPPPAAFTYP-- 171
RBPM2A Homo sapiens PNPSTPLPNTVPQFIAREPYELTVPALYPSSPEVWAPYPLYPAELAPALPPP-AAFTYP-- 170
RBPM2B Mus musculus PNPSTPLPNTVPQFIAREPYELTVPALYPSSPEVWAPYPLYPAELAPALPPPAAFTYP-- 171
RBPM2B Homo sapiens PNPSTPLPNTVPQFIAREPYELTVPALYPSSPEVWAPYPLYPAELAPALPPP-AAFTYP-- 170
***:. * :***:***: ** *:***.***** :*:***: :****

RBPM2 Xenopus laevis A-----AAALHAQMRWYPPSEATQQGWKSRQFC----- 196
RBPM2 Mus musculus TAAAAAATLHAQVRWYPSSDTTQQGWKYRQFC----- 206
RBPM2 Homo sapiens TAA---AAALHAQVRWYPSSDTTQQGWKYRQFC----- 209
RBPM2A Mus musculus -----ASLHAQMRWIPPSEATSQGWKSRQFC----- 197
RBPM2A Homo sapiens -----ASLHAQMRWLPPSEATSQGWKSRQFC----- 196
RBPM2B Mus musculus -----ASLHAQCFSPSEAK-----PNTPVFCPLLQQIRFVSGNVFVTYQPTADQQREL 218
RBPM2B Homo sapiens -----ASLHAQCFSPSEAK-----PNTPVFCPLLQQIRFVSGNVFVTYQPTADQQREL 217
               *:**** . : **

RBPM2 Xenopus laevis -- 196
RBPM2 Mus musculus -- 206
RBPM2 Homo sapiens -- 209
RBPM2A Mus musculus -- 197
RBPM2A Homo sapiens -- 196
RBPM2B Mus musculus PC 220
RBPM2B Homo sapiens PC 219

```

B

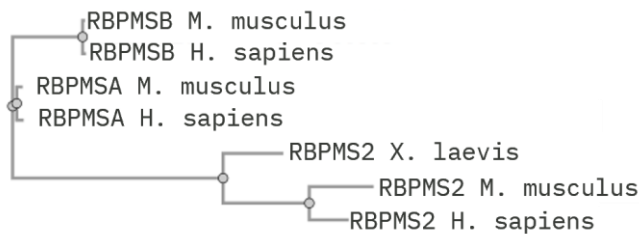


Figure 1.9 RBPM2-A and RBPM2 have a high degree of conservation: **A:** Multi sequence alignment of RBPM2, RBPM2-A and RBPM2-B from *X. laevis* (Hermes), *M. musculus*, and *H. sapiens* generated with Clustal Omega (1.2.4), and colours are from clustal 2 where red represents small and hydrophobic side chains, blue for acidic, magenta for basic, green for hydroxyl, sulfhydryl, and amine side chains. The (*) indicates fully conserved residues, (:) indicates residues of strongly similar properties and (.) indicates residues of weakly similar properties. **B:** Phylogenetic tree of RBPM2, RBPM2-A and RBPM2-B from *H. sapiens*, *X. laevis* and *M. musculus*.

1.6 ROLE OF HIGHER ORDER ASSEMBLY IN RNA BIOLOGY

Many proteins do not perform their functions as a single protein, and instead require assembly into larger protein complexes of multiple copies of the same protein and/or other proteins, for example RBFOX2 functions as part of the LASR complex of proteins (Damianov *et al*, 2016). An extreme form of higher-order assembly is when proteins undergo phase separation, most commonly liquid-liquid phase separation (LLPS), or hydrogel formation. LLPS is the de-mixing of liquids into dilute and dense phase, for example favouring macromolecule-macromolecule interactions over macromolecule-water interactions. At concentrations greater than the critical saturation concentration (C_{sat}), the dense phase condenses into spherical liquid droplets within the dilute phase (Alberti *et al*, 2019). Hydrogels form when these liquid droplets undergo a solid phase transition, and even more extreme assembly can lead to large protein aggregates forming in the cell (Lin *et al*, 2015; Patel *et al*, 2015; Guan *et al*, 2023; Ottoz *et al*, 2023). Large membraneless organelles known as biocondensates have been found within cells, whose formation is driven by phase separation. Many intracellular biocondensates have been identified, from the initial discovery of the nucleolus and Cajal bodies as non-membrane bound organelles (later identified as biocondensates), to stress granules and nuclear speckles, all of which have different protein/RNA/DNA compositions and perform distinct functions in different circumstances (Montgomery, 1898; Cajal, 1903; Collier & Schlesinger, 1986; Collier *et al*, 1988; Banani *et al*, 2017; Sabari *et al*, 2020; Lafontaine *et al*, 2021).

1.61 AMINO ACID SEQUENCE DRIVES HIGHER ORDER ASSEMBLY

Amino acid sequence composition of IDRs determines their behaviour

The amino acid composition of proteins is important for their ability to undergo higher order assembly and phase separation, in particular regions of low-complexity (LC) or intrinsically disordered regions (IDRs) have been found to drive phase separation (Fig.1.9) (Mao *et al*, 2013; Soranno *et al*, 2014; Boija *et al*, 2018; Bratek-skicki *et al*, 2020; Mittag & Pappu, 2022). These regions do not fold into domains, often lack secondary structure and are often proline rich, allowing these regions to remain extended. Some IDRs may be rich in glycine residues which makes them highly flexible. These properties allow them to form multi-valent interactions with other proteins and molecules such as DNA and RNA. In addition to proline and glycine, polar residues such as serine/threonine residues, and glutamine/asparagine residues have been shown to be enriched in some IDRs (Sun *et al*, 2011; Mao *et al*, 2013; Banani *et al*, 2017; Boija *et al*, 2018; Alberti *et al*, 2019). These polar residues have been shown to in some circumstances to increase solubility of proteins in LLPS droplets, for example in the low-complexity structures known as low-complexity aromatic-rich kinked segments (LARKS). In other

cases they have been found to promote the transition to a solid state, as seen in Fused in Sarcoma (FUS) mutants (Hughes *et al*, 2018; Wang *et al*, 2018). Other IDRs and low-complexity regions of proteins might be enriched in aromatic residues or charged residues (Ying *et al*, 2017; Boija *et al*, 2018). The charged residues of IDRs may be forming electrostatic interactions with oppositely charged residues of other subunits or proteins to mediate higher order assembly and phase separation. This would mean that the higher-order assembly would be sensitive to ionic changes and pH changes which affect charge. As well as the classic electrostatic charges between a positive and negative side chain, positively charged residues (lysine and arginine) can engage in cation – π interactions (Fig.1.9) (Gallivan & Dougherty, 1999; Gromiha *et al*, 2004; Banani *et al*, 2017). The π refers to the δ negative (δ -ve) charge from an aromatic side chain (tryptophan, tyrosine, or phenylalanine). Aromatic side chains can also interact with each other, forming π – π interactions, where one side chain is δ -ve and the other is δ +ve. These types of interactions have been shown to be important for higher order oligomerisation and phase separation of proteins. For example in the IDR of DDX4, phenylalanine and arginine residues were found to form cation – π interactions with DDX4 subunits and other proteins, and the aromatic rings of the phenylalanine were also suggested to be forming π – π interactions (Nott *et al*, 2015).

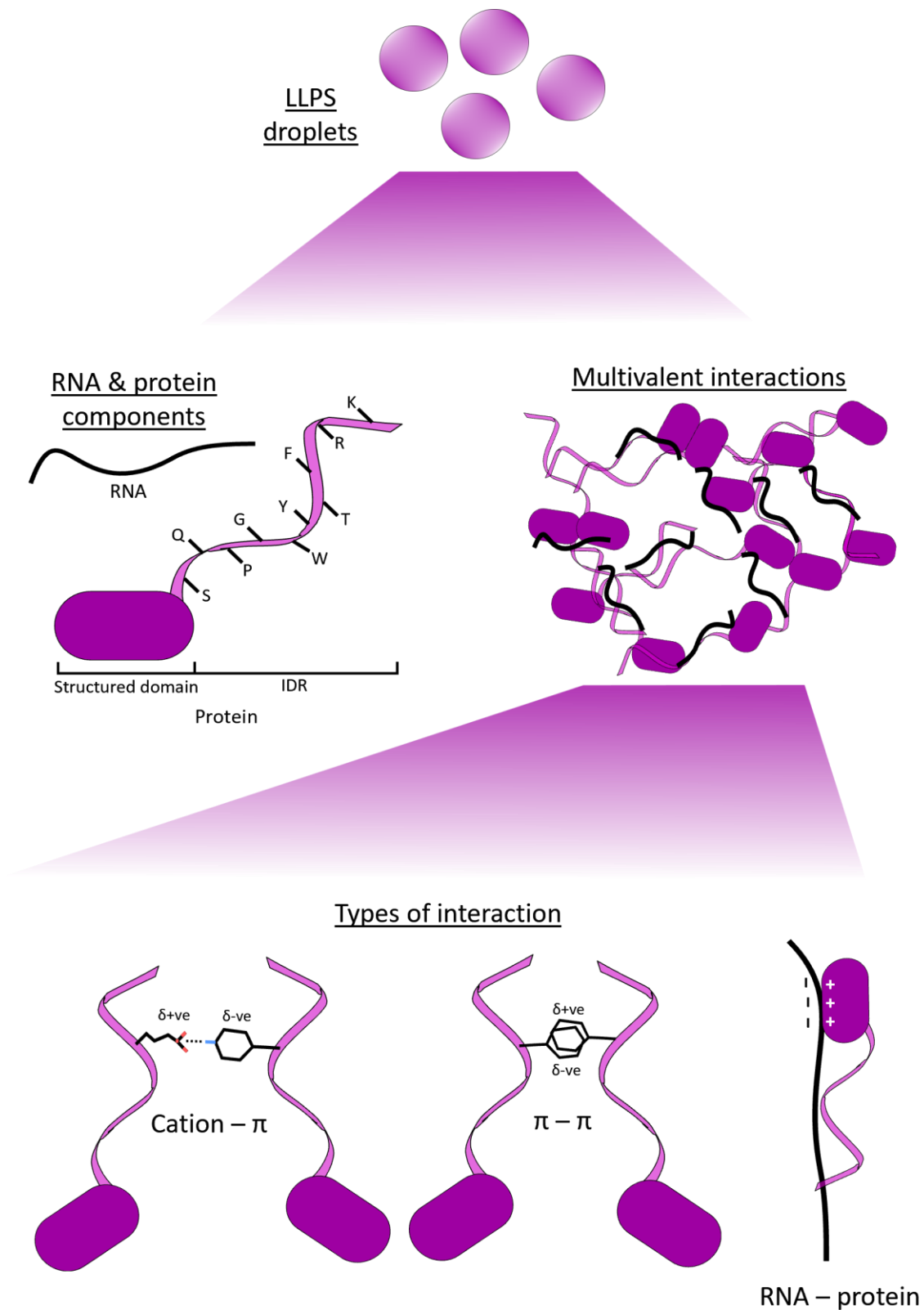


Figure 1.10 Multivalent interactions via IDRs drives assembly and phase separation: Schematic demonstrating the types of interactions that lead to higher order assembly and formation of liquid-liquid phase separated (LLPS) droplets formed from proteins and RNA. Typical domain structure of an RBP is shown with some of the amino acids known to be prevalent in IDRs shown. Lower panel demonstrates some of the multivalent interactions that can occur in higher order assembly of protein and RNA.

Post-translational modifications can affect phase separation

As the composition of amino acids is important for oligomerisation and phase separation due to the properties of their side chains, it therefore follows that post-translational modifications will also affect phase separation. Phosphorylation sites are common within IDRs and alter phase separation due to the introduction of negative charge (Iakoucheva *et al*, 2004; Banani *et al*, 2017). In the case of the serine/arginine rich splicing factors (SRSFs), phosphorylation was found to promote phase separation and results in their recruitment to sites of transcription. Phospho-tyrosine residues in Nephrin were found to promote the formation of higher order oligomers of Nephrin and Nck through interaction with SH2 domains (Misteli *et al*, 1998; Li *et al*, 2012). In contrast, phosphorylation of the low-complexity region of FUS was found to reduce phase separation and reduced pathogenic aggregation (Monahan *et al*, 2017). Arginine methylation has also been found to affect phase separation in a negative manner with a number of proteins, in particular with FUS where it is thought to regulate the strength of cation – π interactions, and may be important in preventing FUS-aggregation (Hofweber *et al*, 2018; Qamar *et al*, 2018).

1.62 ROLE OF RBP HIGHER ORDER ASSEMBLY/BIOCONDENSATES IN SPLICING

RNA can affect higher order assembly

The domain structure of RBPs lends itself to higher order interactions, and in many cases these proteins can undergo LLPS and other phase transitions. This is because many RBPs have IDR/LC regions, as well as modular structured domains that can bind RNA (Ule & Blencowe, 2019). The structured domains are thought to form specific interactions with other proteins or with RNA, and have been given the monicker “stickers” (Mittag & Pappu, 2022). Certain clusters of residues within IDRs can also act as “stickers”, particularly those residues forming electrostatic interactions, cation – π or π – π interactions. This allows for assemblies and interactions with other proteins, allowing assemblies to perform certain functions, such as splicing. The rest of the IDR, or “linker/spacer” regions are thought to be important for increased solubility and thus play an important role in the type and extent of assembly and phase separation observed. The interplay between IDRs and structured domains in how proteins perform their function has also been demonstrated with transcription factors (Gromiha *et al*, 2004; Boija *et al*, 2018). TFs have been shown to form large complexes and, in some cases, small, localised phase separated bubbles (for example at super-enhancers), mediated through multi-valent interactions with their IDRs and binding with target DNA through their structured DNA binding domains. RBPs display very similar behaviour with RNA to aid in co- and post-transcriptional processes (Lin *et al*, 2015). In both cases, RNA and DNA are known to affect higher order assembly and phase separation, and the

presence of RNA has been shown to be important for driving the assembly of various biocondensates, such as stress granules via the interaction with their RNA binding protein partners. These RNA-protein interactions can also be multivalent with structured domains likely binding specifically and IDRs non-specifically to promote phase separation and higher affinity binding to RNA (Li *et al*, 2012; Lin *et al*, 2015). Most RNA binding domains of RBPs recognise only a few nucleotides of RNA and thus by themselves these interactions are often not highly specific (Ule & Blencowe, 2019). Most RBPs therefore often bind to several motifs through multiple RNA-binding domains to allow for a higher affinity to their target RNAs (Lunde *et al*, 2007; Tauber *et al*, 2020). This is demonstrated well by the Pumilio (Puf) family of RBPs which recognise RNA via eight PUM repeats, which individually only recognise single RNA nucleotides, but in combination can bind RNA to a highly specific degree (Lunde *et al*, 2007). These interactions between RBPs and RNA can allow RNA to act as a scaffold for proteins to bind, which can interact with each other and RNA, thus driving phase separation. For example, the G3BP1/2 proteins have two types of RNA binding domains, an RRM and an arginine/glycine rich RGG domain, and this multi-domain protein has been found to be important in the formation of stress granules (Tauber *et al*, 2020). Additionally, when a further RNA binding domain was added to this protein, the liquid-like properties of this protein in stress granules were altered (Yang *et al*, 2020). PTBP1 is also a multidomain RBP, with four RRM domains and thus is able to form many interactions with multivalent RNA molecules that have multiple PTBP1 binding motifs, leading to condensation (Li *et al*, 2012). Even proteins with a single RRM can form RNPs which condense if they are able to oligomerise through their IDRs, and thus form many multivalent interactions upon the RNA, and with each other (Banani *et al*, 2017; Ule & Blencowe, 2019). The effect on assembly and phase separation has been observed to be concentration dependent. Low levels of RNA have been observed to not have much of an effect, whilst at levels higher than the critical level for phase separation, RNA can disrupt assembly and phase separation through the increased electrostatic repulsion brought about by the negative charge on the RNA (Dutagaci *et al*, 2021).

Phase separation/assembly can affect splicing

Higher order assemblies of proteins and RNA allow for close proximity of RBPs and their targets, which can aid in mRNA processing, such as constitutive and alternative splicing at nuclear speckles. These condensates contain many splicing proteins, including an enrichment of SR proteins (Sabari *et al*, 2020). It has been demonstrated that these proteins get phosphorylated, and this diverts them from nuclear speckles to sites of active transcription (Misteli *et al*, 1998). The serine/arginine repetitive matrix 2 protein (SRRM2) and the SON protein have been identified as important scaffold proteins for the formation of nuclear speckles (Giudice & Jiang, 2024). SRRM2 drives condensation through its IDRs, which have also been found to be important for regulating the majority of its splicing targets (Xu

et al, 2022). Nuclear speckles have been linked to post-transcriptional splicing, but the observation by some that activated genes can localise to nuclear speckles and have enhanced gene transcription, co-transcriptional splicing and processing, suggests that nuclear speckles may be a hub for actively transcribing genes to be quickly processed (Liao & Regev, 2021; Giudice & Jiang, 2024). Specifically, it was proposed that splicing was occurring at the interface of nuclear speckles and the nucleoplasm, with SR-bound exons being found inside the speckles and hnRNP-bound introns in the nucleoplasm (Liao & Regev, 2021; Paul *et al*, 2024). Nuclear speckles and other condensates/large assemblies of RBPs can influence AS decisions by increasing the local concentrations of splicing regulators. This allows multiple copies of the same protein (i.e. PTBP1 binding across exons) or proteins binding their co-regulators, as is the case for RBFOX interacting with the LASR complex components via its tyrosine rich C-terminus (Llorian *et al*, 2010; Wongpalee *et al*, 2016; Ying *et al*, 2017).

IDRs are often alternatively spliced

IDRs are often alternative spliced themselves, demonstrating AS can regulate the higher order assembly and condensation behaviour of RBPs. It is theorised that the alternative splicing of these IDRs is one way in which higher organisms have achieved complexity in remodelling PPI networks (Gueroussov *et al*, 2017b). The protein hnRNP DL has three isoforms which can have two IDRs at each terminus, an arginine rich N-terminal and a tyrosine rich C-terminal one, only the tyrosine rich IDR, or neither. These proteins have different states of assembly, with the none-IDR isoform not being able to undergo phase separation, the double IDR isoform undergoing LLPS and the third isoform with only the tyrosine rich region undergoing amyloid like aggregation (Giudice & Jiang, 2024). The amyloid fibril forming isoform is actually the predominant one, despite many instances of amyloid fibres being linked to pathogenesis, although mutations that accelerate aggregation have been linked to limb-girdle muscular dystrophy pathogenesis (Akagi *et al*, 2000; Press *et al*, 2019).

Phase separation behaviour and disease

The splicing factors TARD43 (TDP43) and FUS have been found to undergo phase separation, which can transition to aggregation in amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD) and other disorders (Lagier-Tourenne & Cleveland, 2009; Sun *et al*, 2011; Patel *et al*, 2015). Disease-associated mutations of FUS and TDP-43 were found to be in regions of the proteins that were important for PPIs and promoted aggregation (Lagier-Tourenne & Cleveland, 2009). The normal role of these proteins in neurons is to regulate alternative splicing, and this function is lost when the proteins are stuck in the pathogenic aggregates. Both proteins were found to undergo phase separation under physiological conditions and this behaviour was required for the proper function of these proteins (Sun *et al*, 2011; Wang *et al*, 2018; Hallegger *et al*, 2021). FUS is inherently more

aggregation prone than TDP43, with LLPS droplets formed by FUS via arginine and tyrosine residues of its prion-like domain undergoing a solid phase transition with time (Sun *et al*, 2011; Patel *et al*, 2015; Wang *et al*, 2018). ALS mutations were found to accelerate this solid phase transition and lead to aggregation. TDP43 phase separation is required for assembly upon a subset of RNA splicing targets, where it formed “binding-region condensates” driven by homo-oligomerisation via its C-terminus (Hallegger *et al*, 2021). This mechanism is one way in which TDP43 autoregulates its expression, and ALS mutations were found to perturb the formation of these binding-region condensates and hence prevent autoregulation, leading to excess TDP43 and aberrant splicing. The splicing factor RBFOX1 was found to also form amyloid like fibril aggregates but in this case, these structures were found linked to its splicing activation function (Ying *et al*, 2017). Wild-type RBFOX1 was found to phase separate and form fibrous structures in a hydrogel-like phase *in vitro* via its tyrosine rich CTD, which was required for self-aggregation and interaction with the LASR complex. This behaviour was required for splicing activation but not repression, as when tyrosine residues were mutated, the protein could not form hydrogels and did not activate target exons but was still able to repress other targets. These RBPs demonstrate the importance of LLPS and how this behaviour can lead to formation of hydrogel fibril-like structures in some circumstances and highlights the necessity for regulation of these behaviours through AS and PTMs to prevent pathological aggregation.

1.7 REMAINING QUESTIONS TO BE ADDRESSED

How important is the structure of RBPMS-A for its ability to regulate splicing?

The previous studies on RBPMS-A have demonstrated that it can assemble into higher order oligomers and this behaviour may be linked to its ability to repress splicing. The residues that are required for this activity or oligomerisation, and whether the same residues are required for both, is not known. Work in the lab has primarily focussed on RBPMS-A and its repression of splicing, specifically with the candidate gene *Tpm1*. The RBPMS-A full-length protein was purified and was found to be fully active in HeLa nuclear extract, promoting the VSMC splicing pattern of *Tpm1*. Without RBPMS-A added, *Tpm1* exon 3 (the NM exon) was fully included, and upon RBPMS-A addition, this exon was fully repressed. Additionally, this *Tpm1* construct has been produced *in vitro* and was used to identify the complex that RBPMS-A forms on the transcript to understand its mechanism of repression (Yang *et al*, 2023). RBPMS-A was also found to be involved in the activation of splicing of genes important for the differentiated state of VSMCs, but the importance of the C-terminus of RBPMS-A on splicing activation is still unknown (Nakagaki-Silva *et al*, 2019). Thus, an activation target of RBPMS-A that can be used for *in vitro* study is required to fully elucidate the relationship between splicing activation and repression, and the ability of RBPMS-A to undergo higher order assembly. The production of active full-length

protein that is able to undergo higher order assembly and even LLPS *in vitro* presents a unique opportunity to explore the relationship between self-oligomerisation and the biological function of the same protein.

Based on this previous work on RBPMS-A splicing regulation and oligomerisation, the following questions remain to be addressed: can an activation target for RBPMS-A be produced *in vitro*? What *cis* elements are required for activation by RBPMS-A? Which residues of the C-terminus are required for splicing activation/repression? Are the residues required the same for activation/repression? Do these residues also affect higher order assembly and PPIs? Can a mutant be produced that uncouples these behaviours? Do these mutants have reduced RNA binding? Does RNA influence higher order assembly of RBPMS-A?

Chapters 3 and 4 aim to address these questions, with chapter 3 focusing on identifying and creating an *in vitro* activation assay for RBPMS-A to complement the existing *Tpm1* repression assay. Chapter 4 investigates the importance of C-terminal residues for activity of RBPMS-A and how this relates to its higher order assembly and interactions with other proteins.

How do the biophysical behaviours and activities of RBPMS-A and RBPMS-B compare?

RBPMS-A and RBPMS-B are both highly expressed at the RNA level in differentiated VSMCs but have different apparent splicing activity in transfection assays due to their different C-terminal tails. The C-terminus of RBPMS-B appears to antagonise its splicing activity, with a seventeen amino acid region being predicted to be a “hot spot for aggregation” which led to increased activity upon deletion. Further elucidation on the different splicing activities of these two isoforms and how their oligomerisation compares, may shed light on how different degrees of oligomerisation and the nature of the assemblies the proteins form can alter protein function.

The questions that remain are: how does this HSA region antagonise intrinsic activity of RBPMS-B? Can mutants be produced which rescue the activity of RBPMS-B and behave more similarly to RBPMS-A? What are the features of the rest of the RBPMS-B C-terminus that allows repressive activity when the HSA region is mutated? How does the activity of RBPMS-A and RBPMS-B compare with activation targets? How does this region affect higher order assembly of RBPMS-B? Does RBPMS-B form similar higher order oligomers to RBPMS-A?

The final results chapter, chapter 5 aims to address these questions, focusing on the unique C-terminus of RBPMS-B and investigating which residues impact activity and potential higher order assembly of RBPMS-B, and how this compares with RBPMS-A. The findings from these experiments and from the other results chapters are discussed in chapter 6.

MATERIAL AND METHODS

2.1: CLONING

Transformation and growth of bacterial cultures

Competent *Escherichia coli* (*E. coli*) cells (TG-1 of DH5 α) were pre-made in the laboratory using CaCl₂. The heat-shock method was followed for transformation of the bacteria. 10 μ l of ligation reaction mix or 100 ng plasmid DNA was mixed with 100 μ l of competent cells and incubated on ice for 30 min. The mix was heat-shocked at 42°C for 30 sec and then incubated on ice for 5 min. The cells were then spread onto LB-agar plates supplemented with ampicillin (50 μ g/ml final concentration of ampicillin) and grown overnight at 37°C. Colonies were picked and grown up in either 2 ml or 200 ml liquid LB broth with 1:500 ampicillin (final concentration 50 μ g/ml) respectively for miniprep and maxiprep purification.

DNA preparation

Plasmid DNA was prepared following either the QIAGEN miniprep kit protocol for small scale production or QIAGEN maxiprep kit protocol for large scale production. The bacterial cultures in liquid LB 1:500 ampicillin broth were grown at 37°C, shaking at 280 rpm for either >6 hr for miniprep, or overnight for the maxiprep. Pelleting, lysis and DNA purification steps were completed according to the QIAGEN QIAprep Spin Miniprep or Maxiprep kits. The final samples were eluted in 50 μ l water for miniprep and 200 μ l for maxiprep. The DNA concentration was measured using a NanoDrop spectrometer (Thermo Fisher Scientific).

DNA ligation

Ligations were carried out with 0.5 μ l T4 DNA ligase (NEB) to a final volume of 10 μ l and incubated at room temperature for 10 min, or 20 min for blunt end ligation. The ratios of plasmid DNA to insert used followed the T4 DNA ligase manufacturers guidelines (NEB). The ligation mixes were used to transform TG1 cells as described previously.

Extraction of HEK293T genomic DNA

HEK293T cells resuspended in Phosphate Buffered Saline (PBS, Thermo Fisher Scientific) were centrifuged twice at 500 x g for 5 min at room temperature, washing the pellet in PBS after each spin. The pellet was resuspended in 930 μ l 1x Proteinase K Buffer and 20 μ l of Proteinase K (Thermo Fisher Scientific, 20 μ g/ml final concentration) and incubated for 1 hr at 37°C. Phenol extraction was

performed with an equal volume of phenol/chloroform/isoamyl alcohol (MP Biomedicals) and centrifuging at 18 000 x g for 1 min, before the top layer was added to a new tube. 1/10X volume of 3 M sodium acetate pH 5 and 2.5X volume of ice-cold 70% ethanol were added to precipitate the DNA overnight at -20°C. The sample was centrifuged twice at 18 000 x g for 10 min and the pellet was washed in 1 ml of room temperature 70% ethanol before the pellet was airdried for 15 min at room temperature. The pellet was resuspended in 50 µl water and RNase-treated by adding 6.5 µl of 10X RNase buffer and 8.5 µl of RNase A/T (Thermo Fisher Scientific) and incubating for 1 hr at 37°C. Phenol extraction and ethanol precipitation was performed once more as described previously before the pellet was finally resuspended in 100 µl water. DNA concentration was measured via NanoDrop.

Mutagenesis

Mutagenesis was performed using the QuikChange II XL kit (Agilent) or using Gibson Assembly with the HiFi DNA Assembly kit (NEB) when large or numerous regions were mutated. For the QuikChange II XL, PCR was performed with Pfu Turbo polymerase (Agilent). Samples were incubated with DpnI (NEB) for 1 hr at 37°C to degrade parental methylated DNA template. For Gibson Assembly, DNA fragments were incubated with HiFi DNA Assembly Master Mix (NEB) at 50°C for 30 – 60 min. In both cases, plasmids were transformed and purified as described above. For QuikChange II XL mutagenesis, primers were designed using the Agilent QuikChange Primer Design software. For Gibson Assembly, primers/fragments were designed with 20 – 30 nucleotide overlap between fragments (Tables 2.2 & 2.3). All mutagenesis and cloning was verified by sequencing.

2.11: DNA CONSTRUCTS

Minigene constructs

Minigene reporter constructs were obtained to investigate the splicing activation activity of the RBPMS proteins. These were either three-exon constructs (as described in Gooding *et al*, 2013; Gromak *et al*, 2003) or exon-trap constructs, in both cases the plasmid backbone was from an SV40 vector (Sigma Aldrich). The three-exon minigenes contained the exon reported to be regulated by RBPMS and the two flanking introns and exons. The exon-trap constructs only contain the target exon and partial flanking intronic regions and are fused to GFP exons or Tpm1 exons 1 and 4. The target exons used for the minigene reporters were all previously identified in the lab as being activated by RBPMS (Nakagaki-Silva *et al*, 2019; Jacob *et al*, 2022). The minigene construct DNA fragments were obtained by PCR amplification using Phusion HotStart Polymerase (Thermo Fisher Scientific) with genomic DNA extracted from HEK293T cells (described above). The PCR fragments and plasmids were digested with either NotI HF and MluI (NEB), or NotI HF and Tth111I (NEB) for three-exon constructs and exon-trap

constructs respectively. They were gel extracted from 0.5 µg/ml ethidium bromide 0.8% agarose gels following the QIAGEN QIAquick spin gel extraction kit. The DNA fragments and plasmids were then ligated and transformed and purified as described above.

For the *in vitro* minigene constructs, deletions were made via PCR with Phusion Hotstart polymerase as above, amplifying with primers in opposite directions, removing the desired segment of DNA. T7 promoter sequences were added (Table 2.1) before the upstream constitutive exon, and consensus 5'SS sequences were added after the downstream constitutive exon to improve splicing efficiency.

Protein expression constructs

RBPMs-A, RBPMs-B and their mutants were cloned into pCI 3xFLAG mammalian expression vector (Promega) as described in (Barnhart *et al*, 2022). RBPMs proteins were cloned in using *EcoRI* and *BamHI* sites. The 3xFLAG epitope tag was placed at the N-terminus of RBPMs, as the C-terminus was previously found to be necessary for activity (Yang *et al*, 2023).

For recombinant protein expression, RBPMs without the 3xFLAG were cloned into a modified pET15b vector (as described in Yang *et al*, 2023), with either a N-terminal 6xHis-TEV tag or a N-terminal 6xHis STREP-II tag.

DNA Construct Sequences and Primers

NAME/SAMPLE	AMINO ACID SEQUENCE
3xFlag sequence	DYKDHDGDYKDHDIDYKDDDDK
6xHis-TEV sequence	HHHHHHSSGENLYFQG
STREPII-6xHis sequence	WSHPQFEKGGGSHHHHHH
NAME/SAMPLE	OLIGO (5' → 3')
<i>PTPRF</i> minigene	ATAAGAGCGGCCGCTCCTCTAAGGATGAGCAGTCGATCG GAACTGACGCGTACTCCTGGGAGAACTTGAGGCCATC
<i>PPFIBP1</i> minigene	ATAAGAGCGGCCGCACTGGCATCTTTAAAAGAACAACCTAGA GAACTGACGCGTCTTCATTTGCTGCCATCAAGG
<i>HSPG2</i> minigene	ATAAGAGCGGCCGCTGGATGGCTGGGTTTTTGTGG GAACTGACGCGTTTGAGCTCATCAGACATGTCCCTGC
<i>RBPMs</i> minigene	ATAAGAGCGGCCGCTATGAGGGTTCTCTTATAAAGCTCACATC GACGCGTAACTGCAAAGCATTCTTTGCAGCCTCTG

<i>FLNB</i> minigene	ATAAGAGCGGCCGCAGGTCCTGGAATCGCCTCC GAACTGACGCGTGTGATTTCTCCTTTCTGACAGC
<i>CALD1</i> minigene	CCGGAATTCGAGATCGAGACCATCCTGGC CCGGAATTCGCTACAAATCACTTCACCAGTCTC
<i>Myocd</i> minigene	CCGGAATTCGGAGCAAGCTCAGAATG CCGGAATTCGCAGAGAGATGTTAGGG
<i>BPMS-A</i> transfection construct	CCGGAATTCAACGGCGGCGGCA CGCGGATCCTTAGCAGAACTGCCGGGACTTC
<i>BPMS-B</i> transfection construct	CCGGAATTCAACGGCGGCGGCA GCGGGATCCTTAACATGGGAGCTCCCTCTGTTG
pET15b vector backbone	GGGTTGCCTTACTGGTTAGCAGAATG CATTCTGCTAACCAGTAAGGCAACCC
<i>BPMS-A</i> recombinant RRM fragment	CGAGAAGGAGAACACCCCGA TCGGGGTGTCTCCTTCTCG
<i>BPMS-A</i> recombinant C-tail fragment	AGTCCCGGCAGTTCTGCTAA TTAGCAGAACTGCCGGGACT
<i>BPMS-A</i> WWF/A recombinant C-tail fragment	CGAAGTCCCGGCAGGCTGCTAA TTAGCAGGCCTGCCGGGACTTCG
<i>BPMS-A</i> RKR/A recombinant C-tail fragment	CTGGGCGTCCGCGCAGTTCTG CAGAACTGCGCGGACGCCAG
<i>BPMS-B</i> C-terminus specific fragment	GTTTCTCTCCAGAGGCAAAGCC GGCTTTGCCTCTGGAGAGAAAC
<i>BPMS-B</i> C-tail fragment	CAACCCACTGCAGACCAACAG CTGTTGGTCTGCAGTGGGTTG
FLAG vector backbone	CGTGAGTTTTCGTTCCACTGAGCG CGCTCAGTGGAAACGAAACTCACG
Addition of T7 promoter	GGCCTAATACGACTCACTATAGG GGCCCCTATAGTGAGTCGTATTA
5'SS insertion	CGCGGTAAGTA CGCGTACTTAC

Table 2.1 Primers for cloning DNA constructs

Mutagenesis minigene primers

NAME/SAMPLE	OLIGO (5' → 3')
<i>PTPRF</i> deletion 1	AAGGAGGCAATGAGCAGGGCACAG GGCCCAGGAACCTTCAGTGGCGAC
<i>PTPRF</i> deletion 2	CTTGAGAGACCTGCCAGGC GAGGAAGCAAGGTGGGCAG
<i>PPFIBP1</i> deletion 1	CTGGACAGGAGATAGACCCTTTCAG CCTGGCTGGATTCTGCTGCTCTTTC
<i>PPFIBP1</i> deletion 2	CTGAGCTCAAGCAGTCC CCGGTTATACTGTAGAG
<i>RBPM5</i> deletion	CAAATAGCTGGGACTACAGGC GTGCTAATTCCTAATACATCTCAC
<i>CALD1</i> deletion 1	CCGGAATTCCTGCAAGACTGCATTGAAAGC CCGGAATTCGCTACAAATCACTTCACCACTCTC
<i>CALD1</i> deletion 2	CCGGAATTCGAGATCGAGACCATCCTGGC CCGGAATTCCTCAATATCATCTTCAGTTCACC
<i>FLNB</i> deletion 1	GCTGCAGAGATGCCTTGTG CTCTACAAGGTCATGACTGCTTAAC
<i>FLNB</i> deletion 2	GTAGAATTCAGCAGCTC CACGGGAAAGTGAAAAGTG
<i>PTPRF</i> CAC deletion 1	CGCCCTCTGCTCTCTCCAC GGGGCACTAGGCCCTGC
<i>PTPRF</i> CAC deletion 2	ATACATGCGTTGGGGCTTTCTC ACCATGGAGCACTTGGAAGT
<i>PTPRF</i> CAC deletion 3	TAATGCTGCCTGTGTATGCCCTAC GCAGAGAAAGCCCCAACGC
<i>PTPRF</i> CAC→CCC cluster 1a fragment	CTGTTTTGGGGGCGGGGGGCTGAGGGGGGGGCAC GTGCCCCCTCAGCCCCCGCCCCAAAACAG
<i>PTPRF</i> CAC→CCC cluster 1b fragment	GGACAGAGGGCGGGGTTGGGCCATGAGCATGG CCATGCTCATGGCCCAACCCCGCCCTCTGTCC
<i>PTPRF</i> CAC→CCC cluster 1c fragment	TTGTGCCATGAGCATGGGAGCCTGTTTTGGGG CCCCAAAACAGGCTCCCATGCTCATGGCACAA

<i>PTPRF</i> CAC→CCC cluster 2 fragment	CCCAACGCATGTATGGGCACATGGGAA CATGGGGGACCATGGAGCACTT AAGTGCTCCATGGTCCCCCATGTTCCCAT GTGCCATACATGCGTTGGG
<i>PTPRF</i> CAC→CCC cluster 3 fragment	GCGTTGGGGCTTTCTCTGCCCCCTGCTCAAG CCTCCCACTAATGCTGCCTGTGTATGCCCTAC GTAGGGCATAACACAGGCAGCATTAGTGGGAG GCTTGAGCAGGGGGGCAGAGAAAGCCCCAACGC
<i>PTPRF</i> MBNL mutation 1 fragment	TGCCAGGCAGGGCCTAGTCGCCCCACTCAGCACC CCCCACCAAAACAGGCTCCACATCGTCATGGCAC AACACCCCCCTCTGTCTCTCCAC GTGGGAGAGGACAGAGGGGGGTGTTGTGCCATG ACGATGTGGAGCCTGTTTTGGTGGGGGGGTGCTG AGTGGGGCGACTAGGCCCTGCCTGGCA
<i>PTPRF</i> MBNL mutation 2 fragment	CCCTCTGTCCTCTCCACCCTCCCCATCCCTGTCCC CGCATGTCGTGCTGTCTCCATCCCACCAGTTCCAAG TCGTCCATGGTCACACATGTTAC GTGAACATGTGTGACCATGGACGACTTGGAAGTGG TGGGATGGAGACACGACGACATGCGGGGACAGGG ATGGGGGAGGGTGGGAGAGGACAGAGGG
<i>PTPRF</i> MBNL mutation 3 fragment	CATGCGTTGGGGCTTTCTCTCGCACACTCGTCAAGC CTCACACTAATGCTGCCTGTGTATGCCCTAC GTAGGGCATAACAGGCAGCATTAGTGTGAGGCTT GACGAGTGTGCGAGAGAAAGCCCCAACGCATG

Table 2.2 *Minigene mutagenesis primers*

Mutagenesis RBPMS primers

NAME/SAMPLE	OLIGO (5' → 3')
RBPMS-A HAQMRW deletion	CTCCCTCCCTCCGAGGCTAC CAGTGAAGCGGGATAGGTGAAAG
RBPMS-A LPPSEAT deletion	TCCCAGGGCTGGAAGTCC CCAGCGCATCTGGGCATGCAG
RBPMS-A QGWKSQF deletion	TGCTAAGGATCCTCTGATCCTC GGAAGTAGCCTCGGAGGGAG
RBPMS-A HAQ deletion	ATCCCGCTTCACTGATGCGCTGGCTCCC GGGAGCCAGCGCATCAGTGAAGCGGGAT
RBPMS-A MRW deletion	ACTGCATGCCAGCTCCCTCCCTCCG CGGAGGGAGGGAGCTGGGCATGCAGT

RBPMs-A SRQF deletion	CAGAGGATCCTTAGCACTTCCAGCCCTGGGAA TTCCCAGGGCTGGAAGTGCTAAGGATCCTCTG
RBPMs-A QGWK deletion	CCGAGGCTACTTCTCCCGGCAGTTCTG CAGAACTGCCGGGAGGAAGTAGCCTCGG
RBPMs-A H174A	CAGCGCATCTGGGCAGCCAGTGAAGCGGGATA TATCCCGCTTCACTGGCTGCCAGATGCGCTG
RBPMs-A W179A	GGAGGGAGGGAGCGCGCATCTGGGCA TGCCCAGATGCGCGCTCCCTCCCTCC
RBPMs-A W190A	ACTGCCGGGACTTCGCGCCCTGGGAAGTAG CTACTTCCCAGGGCGCGAAGTCCCAGGCAGT
RBPMs-A F195A	GAGGATCCTTAGCAGGCCTGCCGGGACTTCGC GCGAAGTCCCAGGCAGGCCTGCTAAGGATCCTC
RBPMs-A R178A	GGGAGGGAGCCAGGCCATCTGGGCAGCC GGCTGCCAGATGGCCTGGCTCCCTCCC
RBPMs-A K191A	AGAACTGCCGGGACGCCAGCCCTGGGAAG CTTCCCAGGGCTGGGCGTCCCGGCAGTTCT
RBPMs-A R193A	TTAGCAGAACTGCGCGGACGCCAGCCC GGGCTGGGCGTCCGCGCAGTTCTGCTAA
RBPMs-A Q176A	GGAGCCAGCGCATCGCGCATGCAGTGAAG CTTCACTGCATGCCGCGATGCGCTGGCTCC
RBPMs-A Q188A	GGGACTTCCAGCCCGCGGAAGTAGCCTCGG CCGAGGCTACTTCCGCGGGCTGGAAGTCCC
RBPMs-A Q194A	GATCCTTAGCAGAACGCCCGGACTTCCAGCC GGCTGGAAGTCCCGGGCGTTCTGCTAAGGATC
RBPMs-A R178E	GGAGGGAGGGAGCCACTCCATCTGGGCATGCAG CTGCATGCCAGATGGAGTGGCTCCCTCCCTCC
RBPMs-A K191E	CTGCCGGGACTCCCAGCCCTGGG CCCAGGGCTGGGAGTCCCGGCAG
RBPMs-A R193E	CCTTAGCAGAACTGCTCGGACTCCCAGCCCTG CAGGGCTGGGAGTCCGAGCAGTTCTGCTAAGG
RBPMs-A P160G	GAGGAGGAAGAGCACCCGCTAACTCCGCTG CAGCGGAGTTAGCGGGTGCTCTTCCTCCTC

BPMS-A P164-166G	GGTGAAAGCGGCACCACCACCAAGAGCACC CGCTAACTCCGCTGG CCAGCGGAGTTAGCGGGTGCTCTTGGTGGTG GTGCCGCTTTCACC
BPMS-A P170G	TGCAGTGAAGCGCCATAGGTGAAAGCGGCACC GGTGCCGCTTTCACCTATGGCGCTTCACTGCA
BPMS-B QQ deletion	CCTGTCTTTTGTCCACTTCTTATTAGATTGTCTCTGGGAAT ATTCCCAGAGACAAATCTAATAAGAAGTGACAAAAGACAGG
BPMS-B VFVY deletion	TCTGCAGTGGGTTGATTCCCAGAGACAAATCTAATTTGC GCAAATTAGATTGTCTCTGGGAATCAACCCACTGCAGA
BPMS-B L192-3A, I196A	CAAACACATTCCCAGAGACAAATCTAGCTTGCTGAGCAG CTGGACAAAAGACAGGTGTGTTGGGCT AGCCCAACACACCTGTCTTTTGTCCAGCTGCTCAGCAAG CTAGATTTGTCTCTGGGAATGTGTTTG
BPMS-B F198A, V199A	TGACAAACACATTCCCAGAGGCAGCTCTAGCTTGCTGAGCAGCTG CAGCTGCTCAGCAAGCTAGAGCTGCCTCTGGGAATGTGTTTGCA
BPMS-B V203A, F204A, V205A, Y207A	TGTTGGTCTGCAGTGGGTTGAGCCGTGGCAGCCGCATTCCCAGA GGCAGCTCTAG CTAGAGCTGCCTCTGGGAATGCGGCTGCCACGGCTCAACCCACT GCAGACCAACA
BPMS-B I196P	CACATTCCCAGAGACAAATCTAGGTTGCTGAAGAAGTGACAAAAGA TCTTTTGTCCACTTCTTCAGCAACCTAGATTGTCTCTGGGAATGTG
BPMS-B V205P	GCAGTGGGTTGATACGTGGGAAACACATTCCCAGAGACAA TTGTCTCTGGGAATGTGTTTCCACGTATCAACCCACTGC
BPMS-B V203P	GGGTTGATACGTGGGAAACGGATTCCCAGAGACAAATCTA TAGATTGTCTCTGGGAATCCGTTTCCACGTATCAACCC
BPMS-B C177S	GCCTCTGGAGAGAAAGACTGGGCATGCAGTG CACTGCATGCCAGTCTTTCTCTCCAGAGGC
BPMS-B C190S	CTGAAGAAGTGAGAAAAGACAGGTGTGTTGGGCT AGCCCAACACACCTGTCTTTTCTCCACTTCTTCAG
BPMS-B F178/F189A fragment	CTATCCCGCTTCACTGCATGCCAGTGTGCCTCTCCA GAGGCAAAGCCCAACACACCTGTCGCTGTCCACTTCTT AAGAAGTGACAAAGCGACAGGTGTGTTGGGCTTTCCT CTGGAGAGGCACACTGGGCATGCAGTGAAGCGGGATAG

RBPMs-B wild-type F198/F204 fragment	CACCTGTCGCTTGTCACCTTCTTCAGCAAATTAGAGCTGT CTCTGGGAATGTGGCTGTACGTATCAACCCACTGCAGACCAACAG CTGTTGGTCTGCAGTGGGTTGATACGTGACAGCCACATT CCCAGAGACAGCTCTAATTTGCTGAAGAAGTGGACAAGCGACAGGTG
RBPMs-B P2 F198/F204 fragment	CACCTGTCGCTTGTCACCTTCTTCAGCAACCTAGAGCTG TCTCTGGGAATGTGGCTCCCACGTATCAACCCACTGCAGACCAACAG CTGTTGGTCTGCAGTGGGTTGATACGTGGGAGCCACAT TCCCAGAGACAGCTCTAGGTTGCTGAAGAAGTGGACAAGCGACAGGTG
RBPMs-B wild-type F198/F204/Y207 fragment	CACCTGTCGCTTGTCACCTTCTTCAGCAAATTAGAGCTG TCTCTGGGAATGTGGCTGTACGGCTCAACCCACTGCAGACCAACAG CTGTTGGTCTGCAGTGGGTTGAGCCGTGACAGCCACAT TCCCAGAGACAGCTCTAATTTGCTGAAGAAGTGGACAAGCGACAGGTG
RBPMs-B P2 F198/F204/Y207 fragment	CACCTGTCGCTTGTCACCTTCTTCAGCAACCTAGAGCTG TCTCTGGGAATGTGGCTCCCACGGCTCAACCCACTGCAGACCAACAG CTGTTGGTCTGCAGTGGGTTGAGCCGTGGGAGCCACAT TCCCAGAGACAGCTCTAGGTTGCTGAAGAAGTGGACAAGCGACAGGTG

Table 2.3: Protein construct mutagenesis primers

2.2 MAMMALIAN CELL CULTURE

General cell culture

HEK293T and PAC1 cells were grown in 10 cm dishes in 15 ml Dulbecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific) with 10% Fetal Bovine Serum (FBS, Thermo Fisher Scientific) and all incubations were at 37°C and 5% CO₂. The cells were split by washing twice with 7 ml PBS followed by incubation with 1 ml trypsin for 3 min, before 14 ml of DMEM media 10% FBS was added to quench the trypsin. 1 ml of this mix was added to a new plate with 14 ml of fresh DMEM media 10% FBS to split cells 1:15.

Transfections

Transient transfections were carried out in HEK293T cells via lipofection (lipofectamine 2000, Invitrogen) with mammalian reporter minigene constructs and RBPMs-A, RBPMs-B or their mutants to assess splicing efficiency on the target exons (as described in Nakagaki-Silva *et al*, 2019; Barnhart *et al*, 2022). Briefly, HEK293T cells were seeded at a concentration of 2x10⁵ cells/ml per well of a 6-well plate and DMEM media 10% FBS was added to each well to give a total volume of 1.5 ml per well. The cells were then incubated for >8 hr. 2 µl of lipofectamine per well and 2 µg of chosen DNA plasmid(s) per well was used. Three repeats were performed for each transfection for RNA extraction plus one

repeat for protein extraction. 480 µl of DNA/Optimem/Lipofectamine mix was added to the appropriate wells followed by incubation for 48 hr.

RNA Extraction and RT-PCR

RNA extraction was performed as described in (Nakagaki-Silva *et al*, 2019; Barnhart *et al*, 2022) with TRI reagent (Sigma Aldrich) and according to manufacturer's instructions, followed by ethanol precipitation. The pellets were then Turbo DNase (Ambion) treated and the RNA concentration was measured via NanoDrop. Reverse transcription (RT) with SuperScriptII (Thermo Fisher Scientific) reverse transcriptase was performed to produce cDNA with oligo dT (0.5 mg/ml) from 1 µg of total RNA. Non-RT reactions were prepared in the same way except without the reverse transcriptase. 1 µl of the RT mix was added to a PCR reaction with JumpStart Taq DNA polymerase (Sigma Aldrich). The PCR products were then separated via the QIAxcel Advanced System (QIAGEN) to visualise and quantify the splicing products. The percentage of spliced in (PSI) was calculated with QIAxcel ScreenGel software (QIAGEN).

Protein isolation and western blotting

Protein isolation was performed by suspending cells with Laemmli 2xSDS sample buffer (4% SDS, 10% glycerol, 125 mM Tris-HCl, 5 mg of Bromophenol blue, and 5% of β-mercaptoethanol) and freezing in -80°C. Samples were boiled for 5 min at 90°C and then loaded onto a 15% acrylamide SDS-page gel and transferred to an Immobilon-FL PVDF membrane (Millipore). Color Prestained Protein Standard Broad Range (NEB) was used as a molecular weight standard in all western blots. Proteins were detected with primary antibodies (Table 2.5) and secondary antibodies were IgG antibodies conjugated to horseradish (Strattech Scientifica). Detection was performed via luminol/iodophenol chemiluminescence and imaged on a Bio-Rad ChemiDoc. Analysis was performed using Image Lab software, v6.1.

siRNA mediated knockdown

Knockdown was performed with RBPMS isoform specific siRNAs, pan-RBPMS siRNA, control siRNA or mock treated (Table 2.4). Knockdown was performed in the rat VSMC model cell-line PAC1 which were kept in their differentiated state (as described in Llorian *et al*, 2016). siRNA was transfected at 60 – 90 pmol via lipofection (RNAiMAX, Invitrogen) with dilution in Opti-MEM (Thermo Fisher Scientific) and incubated at RT for 20 min. Cells were seeded at 2.5×10^5 cells/ml in 6-well plates and the siRNA-RNAiMAX mix was added to the wells. A second hit was performed after 24 hr in the same manner, and cells were harvested 48 hr after the second hit. RNA extraction was performed as described above and splicing differences were observed via RT-PCR. Differences in RNA levels were measured using

qPCR, where the cDNA from the reverse-transcription reaction was diluted 1:4 and 2 µl was added to a qPCR 2X Sybergreen mix (Thermo Fisher Scientific) with appropriate primers (Table 2.4). qPCR was performed with Rotor-Gene Q thermocycler and the data was analysed with the Rotor-Gene Q software (QIAGEN). Proteins were extracted by pelleting cells via centrifugation at 8 000 rpm for 1 min and washing twice in PBS. Cells were resuspended in 50 µl RIPA buffer (150 mM NaCl, 1% Triton X-100, 0.1% SDS, and 50 mM Tris pH 8, 1:25 miniEDTA free protease inhibitor cocktail (Roche)), and 10 U of Benzonase (Sigma Aldrich) was added and incubated for 30 min on ice. Cells were centrifuged for 16 000 x g for 15 min at 4°C and supernatants were collected. Total protein concentration was calculated using BCA analysis (Thermo Fisher Scientific). Laemmli 1X SDS sample buffer was added, and the proteins were treated as in the western blotting section above.

HeLa nuclear extraction

HeLa nuclei were extracted as described in Hartmuth *et al*, 2012 and Yang *et al*, 2023. Briefly, the cells were prepared as follows: HeLa S3 cells were suspended in 5-1 T-flask in SMEM media (Thermo Fisher Scientific) with 10% FBS (Sigma Aldrich). They were grown with constant agitation at 80 rpm at 37°C. Harvesting was done at a concentration at 5×10^5 cells/ml via centrifugation at 2 000 rpm for 10 min at 4°C. Pellets were resuspended twice with ice-cold PBS (Thermo Fisher Scientific) 40X cell volume. The extraction protocol was performed according to the S10 protocol (described in Hartmuth *et al*, 2012).

Phosphatase treatment

Whole-cell extracts of HEK293T cells overexpressing RBPMS (transfection section above) were harvested in gentle lysis buffer (150mM NaCl, 1% Triton X-100, 0.1% SDS, 50mM Tris pH 8). Extracts (3 µL) were incubated in NEB Buffer Pack for Protein MetalloPhosphatases (PMP; 50mM HEPES, 100mM NaCl, 2 mM DTT, 0.01% Brij 35, pH 7.5), 10 mM MnCl₂, and 4 U of lambda protein phosphatase for 30 min at 30°C. Protein was then resolved on a 15-20% SDS-PAGE followed by western blotting.

RT-PCR and siRNA primers

NAME/SAMPLE	OLIGO (5' → 3')
RT-PCR for endogenous <i>FLNB</i> (HEK293T)	ATCGCCTCCACTGTGAAAAC AATCCCAGGCCGTTTCATGTC
RT-PCR for endogenous <i>TPM1</i> (HEK293T)	GAGCTGGCAGAGAAAAAG AGGAGGACATCGCGGCCA TCCAACCTCTCCTCAACCAG

RT-PCR for endogenous <i>ACTN1</i> (HEK293T)	CAGATCCTGACCCGGGATG CATGACTTGGTCTGCTGTATCTGT
RT-PCR for endogenous <i>PPFIBP1</i> (HEK293T)	GGTTTGCAAGATGAAAGGAGAAGG GCCATCAAGGACTCCACAGC
siRNA RBPMS-A (Source: Sigma Aldrich)	GGCTGGAAGTCCCGGCAGT
siRNA RBPMS-B (Source: Sigma Aldrich)	GAATGTGTTTGTCACGTAT
siRNA <i>pan</i> RBPMS (Ref: Nakagaki-Silva <i>et al</i> , 2019)	GGCGGCAAAGCCGAGAAGGCCGAACA
C2 siRNA (Source: Dharmacon)	AAGGTCCGGCTCCCCCAAATG
RT-PCR primers for endogenous <i>Actn1</i> (PAC1)	CAGATCTTGACCCGGGATG GCCTGGAACGTCACTACCC
RT-PCR primers for endogenous <i>Ptprf</i> (PAC1)	CGTCAAAGGATGAGCAGTCAATC CCCATCATTGGCTTTGAGAC
RT-PCR primers for endogenous <i>Ppfibp1</i> (PAC1)	GACGAAAGGAGAAGGGGTCTG CCATCAGAGACTCCACTGCC
RT-PCR primers for endogenous <i>Myocd</i> (PAC1)	GTTTTCAATTCCATCCCCAAC AGTTCATCCATCTGCTGACTCC
RT-PCR primers for endogenous <i>Ppfia1</i> (PAC1)	GCCAGTTGCAAGAACGTCTG TTGCCATGTCTCTCCTCAGC
qPCR primers for RBPMS-A	AGCTCACAGTACCTGCACTT CTTCCAGCCCTGGGAAGTAG
qPCR primers for RBPMS-B	TTCTCTCCAGAGGCAAAGCC GAGCTCCCTCTGTTGGTCT
qPCR primers for <i>pan Rbpms</i>	CTCACATCTAAACAGCCCGTAG CGTTCTTTGCGGCCTCTG
qPCR primers for housekeeping gene <i>Rpl-32</i>	GCCCAAGATCGTCAAAAAGAGG ATCAGGATCTGGCCCTTGAATC
RT-PCR Transfection primers for three-exon minigenes	TTGCAAAAAGACCATGGGTACC GGCATGCAAGCTTACGCG
RT-PCR Transfection primers for exon-trap minigenes and <i>ACTN1</i> and <i>Tpm1</i> minigenes	TTTTGGAGGCCTAGGCTTTTGCAAAAAG ACTCACTGCGTTCCAGGCAATGCT
RT primer for <i>PTPRF</i> and <i>FLNB</i> <i>in vitro</i> constructs	TTACGCGTACTTAC
RT primer for <i>Tpm1</i> <i>in vitro</i> construct	CCAGACGCTCTGAGCGCG
RT-PCR primers for <i>PTPRF</i> <i>in vitro</i> construct	GCCGCGTCTCTAAGGATGAG

	CTTACCGCGTTACTCCTGGG
RT-PCR primers for <i>FLNB</i> <i>in vitro</i> construct	CCGCAGGTCCTGGAATC CTTACCGCGTGTGATTCTCC
RT-PCR primers for <i>Tpm1</i> <i>in vitro</i> construct	CAGATGCTGAAGCTCGACAA TGCTACGTCAGCTTCAGCTG

Table 2.4: RT-PCR primers used to visualise splicing events of minigene/*in vitro* targets or endogenous targets and siRNA and qPCR primers to assess knockdown of RBPMS isoforms

2.3 RECOMBINANT PROTEIN PRODUCTION

Bacterial growth and protein expression

Competent BL21 cells were transformed with pET15b plasmids (Sigma Aldrich) containing RBPMS, they were heat shocked at 42°C for 30 seconds, and incubated on ice for 5 min. Transformed bacteria were plated onto ampicillin LB-agar plates and incubated overnight at 37°C. Colonies were picked and grown overnight in 10 ml LB with 50 µg/ml ampicillin, shaken at 280 rpm at 37°C. The overnight cultures were then added to 2 L flasks and topped up to 500 ml LB with 50 µg/ml ampicillin. These flasks were shaken at 250 rpm for 1.5 – 2 hr at 37°C until the OD (optical density) had reached 0.6 – 0.8, indicating log-phase growth. 0.2 mM of IPTG (isopropyl b-D-1- thiogalactopyranoside, Melford) was then added to induce protein expression, and the cells were shaken at 250 rpm at 37°C for 2 hr. For RBPMS-B, after addition of IPTG, cells were shaken at either 19°C for >5 hr, or 37°C or 42°C for 2 hr. The cells were then centrifuged at 7 000 x g for 10 min at 4°C, and the pellets were moved to 50 ml falcon tubes and stored at -20°C.

Recombinant TEV protease (rTEV) for tag removal was expressed in a similar manner except once IPTG had been added, cells were incubated at 19°C for 5 hr before centrifugation.

2.31 PROTEIN PURIFICATION – RBPMS-A

Cell lysis

The cell pellets were resuspended in HisA buffer (50 mM Tris, 40 mM Imidazole, 500 mM KCl, 10% glycerol, pH 8.5) and 1 M LiCl₂, 0.5 mM DTT and 1 miniEDTA free protease inhibitor tablet (Roche) (final volume 15 ml). Cells were sonicated at a rate 10s on, 10s off at amplitude 15 µm with Sonicator Soniprep 150 (MSE).

HisTrap purification

For RBPMS-A and its mutants, the cell lysate was then centrifuged at 45 000 x g and supernatant was passed through a 0.45 µm filter. The cell lysate was then loaded onto a HisTrap HP column (Cytiva) (Figure 2.1) on an AKTA PURE system (Cytiva), using buffers HisB (50 mM Tris, 500 mM Imidazole, 500 mM KCl, 10% glycerol, 0.5 mM DTT, pH 8.5) and HisA2 (50 mM Tris, 2 M NaCl, and 10% glycerol, pH 8.5). Peak fractions were confirmed with SDS-PAGE gel electrophoresis and Coomassie blue staining.

TEV production and tag removal

Recombinant His-tagged TEV was purified as above, with a HisTrap Excel column (Cytiva), and peak fractions were buffer exchanged to Phosphate-Buffered Saline (PBS, Thermo Fisher Scientific).

1 ml of 1 M Tris (not pH adjusted) was added to 7 ml peak fractions of RBPMS preparations to disrupt potential higher-order species. Proteins were then buffer exchanged to QA buffer (25 mM CAPS, 50 mM KCl, 10% glycerol, 0.5 mM DTT, pH 10) using either PD-10 columns (Cytiva) on the bench for small volumes (≤ 5 ml) or on a HiPrep Desalting 26/10 column (Cytiva) on the AKTA PURE system. For proteins that required tag removal, 1 mg/ml rTEV was added at 1:20 and incubated at 30°C for 2 hr, followed by incubation overnight at 4°C.

MonoQ column purification and SEC

rTEV cleaved proteins and tagged proteins in QA buffer were loaded onto a MonoQ column (Cytiva) with buffer QB (25 mM CAPS, 50 mM KCl, 1 M NaCl, 10% glycerol, 0.5 mM DTT, pH 10). Clean peak fractions were confirmed via gel electrophoresis and Coomassie Blue stained. Samples were concentrated with Amicon Ultra filter (MWCO 10kDa) (Millipore).

Buffer exchange

Larger volume samples were buffer exchanged with HiPrep desalting column 26/10 column (Cytiva), and 2.5 – 5 ml volumes via PD-10 columns (Cytiva). Sample volumes of ≤ 130 µl were buffer exchanged with Zeba Spin (Thermo Fisher Scientific) columns. Proteins with volumes > 500 µl were concentrated with 10 kDa MWCO Amicon Ultra membrane filters (Millipore). Proteins at ≤ 500 µl were concentrated using Vivaspin 500 columns (Sartorius).

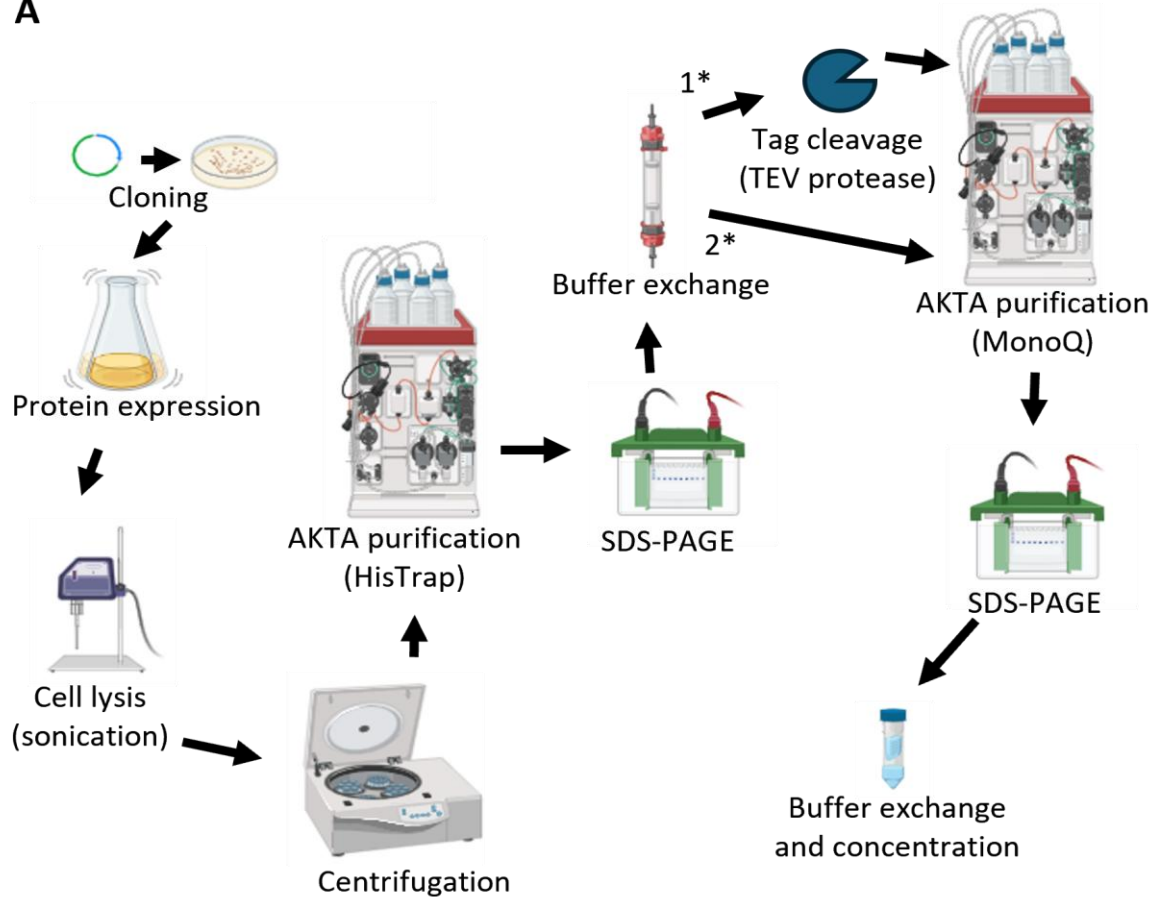
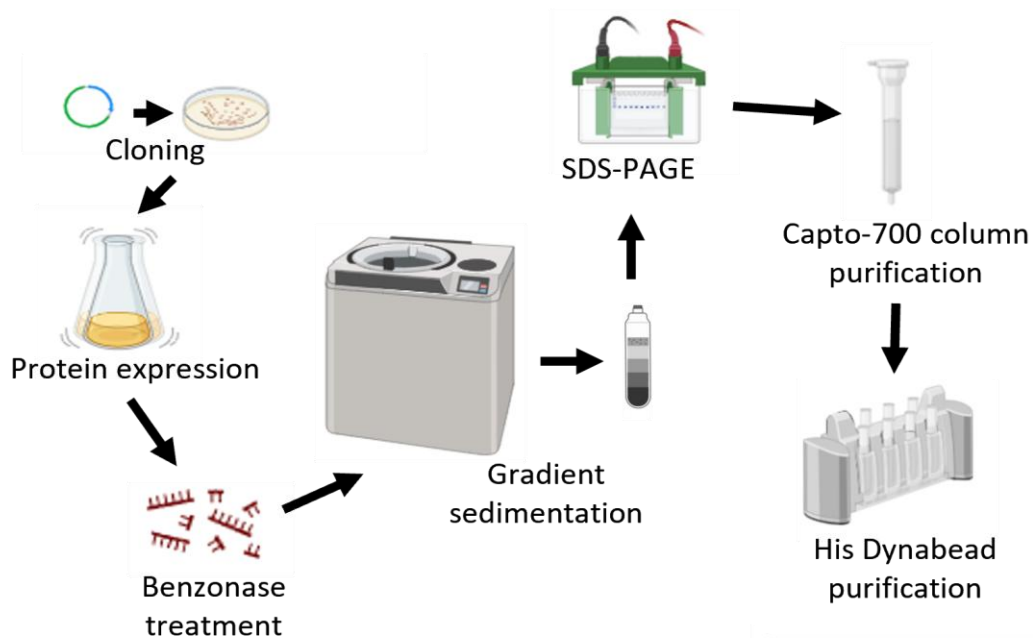
2.32 PROTEIN PURIFICATION – RBPMS-B

Sedimentation

Cell pellets were resuspended in HisA buffer and 1 miniEDTA free protease inhibitor tablet (Roche) (final volume 15 ml). Cell lysates were treated with 10 U/ml of Benzonase (Milliore) for 30 min on ice, followed by centrifugation at 18 000 x g for 15 min (Figure 2.1). 500 µl of supernatant was then loaded on 10 – 60% glycerol gradient (100 mM KCl, 20 mM HEPES pH 7.9, 0.5 mM DTT) set up in 14 mm × 95 mm tubes (Beckman Coulter). Samples were then subjected to ultracentrifugation in an SW40 Ti rotor (Beckman Coulter) at 32 000 rpm for 15.5 hr at 4°C. Gradients were fractionated into 24 × 0.5 ml fractions. E. coli ribosome subunits (30S and 50S) were used as size markers and identified by UV absorption at 260 nm, EDTA at 2.5 mM was added to keep them separated. Fractions were analysed by SDS-PAGE gel electrophoresis and Coomassie Blue stain.

Sample clean-up

Fractions containing RBPMS-B were added to a HiTrap Capto 700 (Cytiva) with low salt buffer (100 mM KCl, 20 mM HEPES pH 7.9 and 0.5 mM DTT). The flow-through (FT) was collected and concentrated with 10 kDa MWCO Amicon Ultra filters (Millipore). The concentrated sample was then added to Dynabeads His-Tag (Invitrogen) and the isolation protocol was followed. Briefly, the concentrated FT samples were buffer exchanged to HisA buffer that did not contain DTT, and 2 mg of Dynabeads was added to 50 µl of the buffer exchange sample and rotated at RT for 5 min. Samples were added to a magnetic rack and washed with HisA. This wash step was repeated four times, before the beads were eluted with HisB buffer and rotated for 10 min at RT. Purity of sample was assessed by SDS-PAGE gel electrophoresis and Coomassie Blue staining.

A**B**

Created in BioRender.com

Figure 2.1 Schematic of the purification of recombinant RBPMS. **A:** Purification of wild-type RBPMS-A and its mutants, and the RBPMS-B P2 mutant. 1* pathway for the TEV cleavable His (6xHis-TEV) tagged proteins, and 2* for the STREP-6xHis tagged proteins. **B:** Purification of wild-type STREP-6xHis tagged RBPMS-B.

2.4 *IN VITRO* ASSAYS

2.41 RNA AND SPLICING ASSAYS

In Vitro transcription

DNA plasmids were linearised with XhoI (*Tpm1*) or Hind III (*PTPRF*, *RBPM5*, and *FLNB*) and then DNA was extracted by adding equal volume phenol/chloroform/isoamyl alcohol as described in gDNA extraction from HEK293T cells, except after ethanol precipitation, samples were spun for 10 min at 18 000 x g, then washed with 1ml 70% room temperature ethanol and samples were spun for 5 min at 18 000 x g. Ethanol was removed and pellets were airdried for 15 min and resuspended with 10 µl water. 1 µl of 1mg/ml linearised DNA was added to a transcription mix. For radiolabelled *in vitro* splicing samples, this was 10 µl mix: 1X Transcription buffer (Thermo Fisher Scientific), 1 µl α[32P]-UTP/CTP (Revvity), 0.4 µl RiboLock (Thermo Fisher Scientific), 0.5 mM Cap Analog (NEB), 1 µl rNTP mix (5 mM GTP, 5 mM ATP, 2/5 mM CTP, 2/5 mM UTP (Thermo Fisher Scientific), depending on which radiolabel used) and 1 µl T7 RNA polymerase (Invitrogen). For radiolabelled UV crosslinking samples, the transcription mix was the same as above, except 2 µl of α[32P]-CTP and T7 polymerase was added. For non-radiolabelled splicing transcripts, the mix was 50 µl: 1X Transcription Buffer, 10 µl rNTP mix (10 mM ATP, 10 mM UTP, 10 mM CTP, 2.5 mM GTP), 1.6 mM CAP Analog, 1 µl RiboLock and 1.5 µl of T7 polymerase.

For radiolabelled transcripts, transcription reactions were performed at 37°C for 2 hr. Samples were then phenol extracted with equal volumes of water-saturated phenol (MP Biomedicals) and spinning down for 1 min at 18 000 x g. The top layer was added to a G50 column (Cytiva) and spun down at 2800 rpm for 2 min. RNA concentration was calculated by calculating the counts per minute (cpm) of the transcript before and after phenol extraction via LS 6500 scintillation counter (Beckman Coulter).

For non-radiolabelled transcripts, reactions were performed at 37°C for 2 hr, then 1 µl Turbo DNase (Thermo Fisher Scientific) was added and incubated for 20 min at 37°C. RNA was purified using the Monarch RNA Cleanup kits (NEB) and concentration was determined via Nanodrop.

In Vitro splicing

In vitro splicing assays were performed as described in (Krainer *et al*, 1984). Briefly, RNA transcripts at 12.5 ng/µl (radiolabelled) or 2.5 nM (non-radiolabelled) were added to 10 µl splicing mix containing 30% HeLa nuclear extract, 0.4 µl RiboLock, 2.6% Polyvinyl Alcohol (PVA, Sigma Aldrich), 2 mM ATP (Thermo Fisher Scientific) 2 mM MgCl₂, 20 mM Creatine Phosphate (Boehringer Mannheim GmbH)

and buffer D (20 mM HEPES-KOH pH 7.9, 0.1 M KCl, 1.5 mM MgCl₂, 0.2 mM EDTA (pH 8), 10% (v/v) glycerol). Recombinant RBPMS was added at the desired concentration in Splicing Buffer (SPB: 25 mM HEPES pH 7.9, 60 mM KCl, 0.5 mM DTT). Samples were incubated at 30°C for desired times (0 – 3 hr), and the reaction was stopped by adding proteinase K (Thermo Fisher Scientific, final concentration 20 µg/ml) and incubating for 30 min at 37°C.

Samples were phenol extracted with equal volumes of water-saturated phenol (MP Biomedicals) twice and precipitated with 100% ethanol at -20°C overnight. Samples were centrifuged at 16 000 x g for 15 min and pellets were airdried for 10 min. RNA pellets were resuspended in 10 µl water. For radiolabelled samples, 5 µl of formamide dye was added and samples were boiled at 90°C for 4 min. Samples were then loaded onto a 4% UREA-PAGE gel. Gels were run in 1X TBE for 1.5 hr and fixed in 10% methanol and 10% acetic acid. Gels were dried and transferred to a Phosphor screen and imaged using a Typhoon FLA 9000 (Cytiva). For non-radiolabelled samples, specific RT primers were added (Table 2.4) and an RT and RT-PCR reaction were performed as described above. Visualisation was performed with a Qiaxcel (QIAGEN).

For lariat de-branching samples, RNA pellets were resuspended in buffer D and 30% of a splicing-poor HeLa nuclear extract, and incubated at 30°C for 30 min, before the reaction was stopped as described above and RNA was extracted again.

RNA crosslinking

The RNA crosslinking reactions were set up in 96-well untreated polystyrene plates (Greiner) with Binding Buffer (10 mM HEPES pH 7.9, 3 mM MgCl₂, 5% Glycerol and 1 mM DTT), 40 mM KCl, 75 µg/ml tRNA, 2 nM radiolabelled RNA, 20% nuclear extract or buffer E (20 mM HEPES pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, and 0.5 mM DTT), and 0 – 2 µM of recombinant RBPMS. Binding reactions were performed at 30°C for 30 min, and then UV crosslinked with 2x 960 mJ at 254 nm using a Spectrolinker XL-1500 UV Crosslinker. Non-crosslinked RNA was then degraded using RNase (1 mg/ml of RNase A and 1 U/ml of RNase T1 (Thermo Fisher Scientific)) for 10 min at 37°C. Loading dye was added, and samples were added to 15% acrylamide SDS-page gels. Coomassie Blue stain was used to fix protein, and then gels were dried and transferred to phosphoscreens. Imaging was performed as described above for the splicing gels.

2.42 *IN VITRO* BIOPHYSICAL ASSAYS

For all biophysical assays, unless stated otherwise, recombinant RBPMS was buffer exchanged to low salt buffer (100 mM KCl, 20 mM HEPES pH 7.9, 0.5 mM DTT).

Phase separation studies

Buffer exchange on purified recombinant RBPMS-A was carried out using a PD-10 desalting column (Cytiva), into splicing buffer. Nanodrop was used to determine protein concentration. Phase separation of the purified protein was promoted by the addition of crowding reagents at various concentrations: polyvinyl alcohol (PVA), polyethylene glycol (PEG) 3350 or PEG 7000-9000 (Sigma Aldrich). Purified recombinant RBPMS-A with fluorophore Alexa-647 attached to the sole cysteine at the C- terminus (described below), was added to allow visualisation of the phase separated droplets via microscopy using an Oxford Nanolmager. Measurements were taken at $\lambda = 647$ nm. Turbidity measurements were taken using the nanodrop at $\lambda = 340$ nm, $\epsilon = 1.176$ (abs 0.1% (=1 g/l)).

Labelling proteins

Proteins were fluorescently labelled with Alexa Fluor™ 647 C2 Maleimide (Invitrogen). Briefly, proteins were buffer exchanged in labelling buffer (100 mM KCl, 20 mM HEPES pH 7.3, 0.5 mM TCEP) using a Zeba spin desalting column (7K MWCO; Thermo Fisher Scientific). Alexa Fluor 647 C2 maleimide was added at 10X molar excess and incubated for 2 hr at RT or 4°C overnight in the dark. The reaction was quenched by addition of β -mercaptoethanol (Sigma Aldrich) in molar excess (1 μ l of 14 M) and incubated at RT for 5 min. pH was corrected by addition of 25 mM CAPS pH 10 (final concentration). The fluorescently labelled protein was then buffer exchanged to desired buffer using an Amicon Ultra centrifugal filter (10K MWCO; Thermo Fisher Scientific) until the filtrate ran clear (protein was washed and concentrated 3/4 times).

The amount of free fluorophore in the mixture was estimated by SDS-PAGE analysis, running labelled and unlabelled protein together and imaging on a Bio Rad ChemiDoc. Activity of labelled protein was compared to unlabelled protein in *in vitro* splicing assays.

Mass photometry

Mass photometry uses the interference between light that is scattered as it comes into contact with a glass coverslip and the light reflected back at the coverslip to identify the mass of protein molecules (Young *et al*, 2018). This is because the interference is directly proportional to the mass of single protein molecules (or other macromolecules) on the glass surface. The limitations of mass photometry are in the lower molecular weight limitation (the smallest mass that can be detected by mass photometry is 42 kDa), and the low concentration requirements (≤ 150 nM of protein only). Mass Photometer TwoMP (Refeyn) was used to assess higher-order assembly of RBPMS. RBPMS-A and its mutants were buffer exchanged to MP buffer (150 mM KCl, 20 mM HEPES pH 7.9, and 0.5 mM DTT), and rapidly diluted to 100 – 150 nM and recording of the binding events were collected for 60s using

the Acquire MP software (Refeyn). Each experiment was performed in triplicate and data was merged to create histograms with the Discover MP software (Refeyn). Samples and buffer were centrifuged briefly before each experiment to avoid aggregates and particulates.

Glutaraldehyde crosslinking

Glutaraldehyde is used to crosslink proteins in order to stabilise structures/higher order assemblies by reacting with the primary amine groups of lysine residues, but the reaction must be quenched to prevent over crosslinking resulting in protein aggregation (Kim *et al*, 2024). Glutaraldehyde crosslinking was used to stabilise oligomers of RBPMS-A before diluting for mass photometry experiments. Glutaraldehyde (Sigma Aldrich) at 0.0015% was added to 10 – 20 μ M RBPMS-A wild-type or mutants, and samples were incubated for 30 min on ice. Crosslinking was quenched by addition of 200 mM Glycine pH 6 and incubation for 5 min on ice. For SDS-PAGE analysis, samples were run on 4-12% Bis-Tris NuPAGE gels (Invitrogen) and Coomassie Blue or silver stained to show higher-order assembly. For mass photometry analysis, samples were diluted to 100 – 150 nM as previous.

RNA and glutaraldehyde crosslinking analysis

For RNA-protein crosslinking studies, short segments of *in vitro* transcribed RNA were used which were 42 nucleotides (CAC₂) and 120 nucleotides (3xCAC₂) long, and either contained a single pair of CACs separated by 10 nucleotides of spacer RNA, or three pairs of CACs separated by 10 – 16 nucleotides of spacer RNA. RNA (0 nM – 1800 nM) was mixed with 10 μ M RBPMS and incubated at 30°C for 1 hr. Glutaraldehyde (Sigma Aldrich) crosslinking was performed as above. After quenching, 5 U/ml Benzonase (Millipore) was added to the samples for 15 min at 30°C before centrifugation at 18 000 x g for 5 min. The supernatant was then taken for mass photometry analysis or ran on a 4 – 12% Bis-Tris gradient NuPAGE gel (Invitrogen).

Circular Dichroism

Circular dichroism (CD) utilises the differential absorption of circularly polarised by chiral molecules such as proteins, and is often used to elucidate protein secondary structure (Rogers *et al*, 2019). The different secondary structures of proteins have different arrangements of amino acids and thus have altered absorption of circularly polarised light in the far UV spectrum (λ = 180 – 260 nm). CD is limited by buffer restraints as certain ions absorb UV differently at the wavelengths required by CD, and depends on the purity and concentration of the protein leading to relatively low resolution of secondary structure (Greenfield, 2007). The peptide of the HSA region of RBPMS-B (residues 192 – 208) was produced by Genscript and dissolved with DMSO (Sigma Aldrich). The peptide was dialysed into CD buffer (32.5% Acetonitrile, 10 mM phosphate buffer and 50 mM NaF) overnight at 4°C with

Slide-A-Lyzer™ MINI Dialysis Devices, 2 kDa MWCO (Thermo Fisher Scientific). Circular dichroism was performed with the Aviv 410 CD instrument and a pathlength of 0.1 cm and stain-free QS quartz 450 µl cuvettes were used. The peptide was diluted to 0.1 mg/ml and measurements were taken at 25°C, 90°C and at 25°C again post 90°C heating. Absorption was measured at $\lambda = 185 - 250$ nm. Three measurements were taken at each temperature.

Plate reader studies

Fluorescence studies with the HSA peptide and control proteins (Carbonic Anhydrase I (Sigma Aldrich), rat islet amylin polypeptide (Cambridge Bioscience), and the amyloid β 40 peptide (Sigma Aldrich)) were performed with the PHERAstar FSX plate reader (BMG LabTech). 0 – 20 µM Thioflavin T (ThT, Sigma Aldrich) was added to 0 – 10 µM of protein in 1X PBS. Samples were prepared at 4°C on 384 well microplates (Corning-3575) and sealed with LightCycler 480 sealing foil (Roche). Plates were centrifuged briefly before starting the assay. ThT was excited at $\lambda = 440$ nm and emission at $\lambda = 480$ nm was measured every 15 min for 18 hr, at 37°C.

For assays involving Guanidine HCl (Sigma Aldrich), the HSA peptide was dialysed overnight into PBS buffer containing 0 – 6 M Guanidine at 4°C. Samples were then prepared as described above and measurements at $\lambda = 480$ nm were taken every 15 min for 7 hr at 37°C.

Dynamic Light Scattering

The Zetasizer Nano S DLS machine was used to measure the dynamic light scattering of the HSA peptide in CD buffer at 0.1 mg/ml in 70 µl Cuvettes (Sigma Aldrich). Three repeats were performed.

Negative stain electron microscopy

Transmission electron microscopy with the HSA peptide or full-length RBPMS-B was performed with Tecnai G2 80-200keV transmission electron microscope. Copper grids were used with a carbon film and Uranyl Acetate was used to stain samples. HSA peptide was in CD buffer at 0.1 mg/ml, and RBPMS-B was in low salt buffer at 0.5 mg/ml.

2.43 INTERACTION STUDIES

Pull-down studies

Pull-down experiments were performed as described in (Yang *et al*, 2023). Briefly, MagStrep “type3” XT beads (IBA Lifesciences) were initially washed three times with bead wash buffer (20 mM Tris pH 7.5, 200 mM KCl, 1 mM EDTA) followed by washing three times with G.A buffer (20 mM HEPES pH 7.9, 50mM KCl, 2.2 mM MgCl₂, 200 µM EDTA, 5% glycerol). The beads were blocked with 1.5% BSA (Acros

Organics) and 7.2 mg/ml tRNA from yeast (Roche) and 0.08% NP-40 Alternative (CALBIOCHEM) overnight at 4°C. 6xHis STREP-II tagged RBPMS-A and its mutants were buffer exchanged to low salt buffer and added at 2 µM final concentration to a reaction containing 60% HeLa nuclear extract, 2.6% PVA, and 2.2 mM MgCl₂ and incubated for 15 min at 30°C. The pull-down mix was added to the pre-blocked beads and incubated for 1 hr at 4°C. Beads were washed 6x with G150 buffer (20 mM HEPES pH 7.9, 150 mM KCl, 250 µM EDTA, 2.2 mM MgCl₂, 0.05% NP-40, 10% glycerol) before being eluted with biotin buffer (100 mM Tris, 150 mM NaCl, 1 mM EDTA, 50 mM Biotin, pH 8). Samples were probed by SDS-PAGE, silver stain and western blot analysis.

ANTIBODY	HOST	SOURCE	DILUTION
FLAG	<i>Mouse (Monoclonal)</i>	<i>Sigma-Aldrich, F1804</i>	<i>1:2000</i>
MBNL1	<i>Mouse (Monoclonal)</i>	<i>Sigma-Aldrich, M3320</i>	<i>1:2000</i>
RBFOX2	<i>Rabbit (Polyclonal)</i>	<i>Bethyl, A300-864A</i>	<i>1:2000</i>
RBM14	<i>Rabbit (Polyclonal)</i>	<i>Abcam, ab70636</i>	<i>1:2000</i>
RBPMS	<i>Rabbit (Polyclonal)</i>	<i>Sigma-Aldrich HPA056999 ATLAS</i>	<i>1:500</i>
TUBULIN	<i>rat (Monoclonal)</i>	<i>Abcam, ab6160</i>	<i>1:10000</i>

Table 2.5: Primary antibodies used in immunoblot and immunofluorescence studies

RESULTS

3 RBPMS CAN ACTIVATE EXON INCLUSION *IN VITRO*

3.1 CANDIDATE SELECTION

Identification of target exons that are activated by RBPMS

The mechanism of splicing repression by RBPMS has previously been investigated using the well-characterised vascular smooth muscle cell transcript, tropomyosin 1 (*Tpm1*) (Mullen *et al*, 1991; Gooding & Smith, 2008; Gooding *et al*, 2013; Nakagaki-Silva *et al*, 2019; Barnhart *et al*, 2022; Yang *et al*, 2023). In differentiated VSMCs, tropomyosin exon 3 is repressed, and its alternative mutually exclusive partner, exon 2, is spliced in. This splicing pattern was efficiently reconstituted in cell-free *in vitro* conditions upon addition of recombinant RBPMS-A, showing RBPMS-A represses exon 3 and that it can be captured in an *in vitro* splicing assay. The use of an *in vitro* splicing assay allows for a detailed analysis of the relationship between splicing repression activity and higher-order assembly of RBPMS.

A similar target was required to investigate splicing activation by RBPMS and allow a comparison of the mechanisms of splicing activation and repression regulation. A selection of exons were identified as being included in RBPMS overexpression data or skipped in RBPMS knockdown data, as well as being included in differentiated versus skipped in the proliferative model rat smooth muscle cell line, PAC1 cells (Nakagaki-Silva *et al*, 2019). Additionally, these target exons were found to be included in rat aorta tissue cells and in human embryonic stem-VSMC cells with RBPMS overexpression (Jacob *et al*, 2022). These candidate exons were in the following genes: *Ptprf*, *Ppfibp1*, *Cald1*, *Myocd*, *Flnb*, *Hspg2*, and *Rbpms* itself.

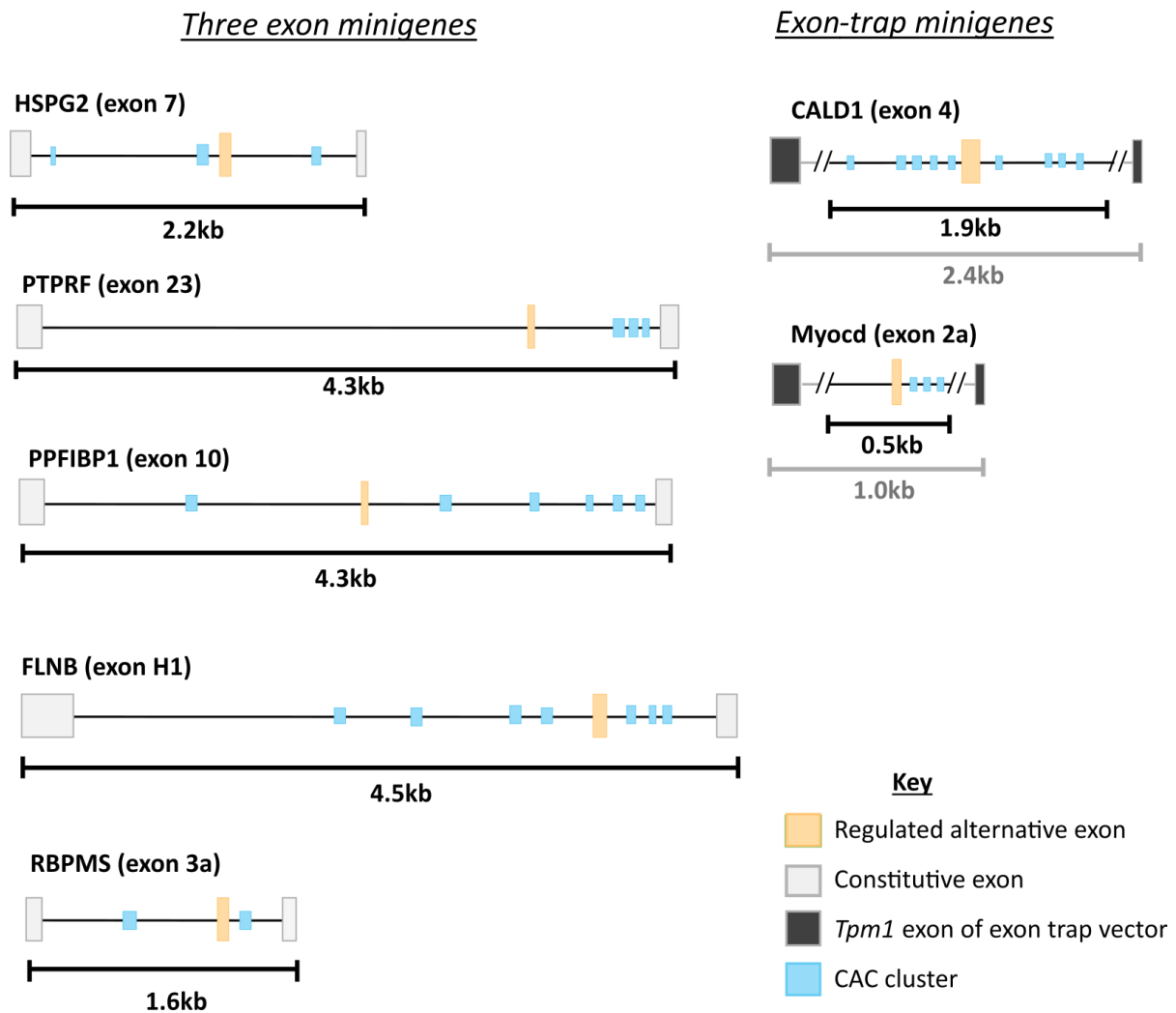


Figure 3.1 Minigene construct sizes: Schematics show the length of each minigenes and the different construct types. The regulated exon numbers are indicated above each minigene and are coloured orange in the schematics. The CAC sites where RBPMS may bind are indicated in blue.

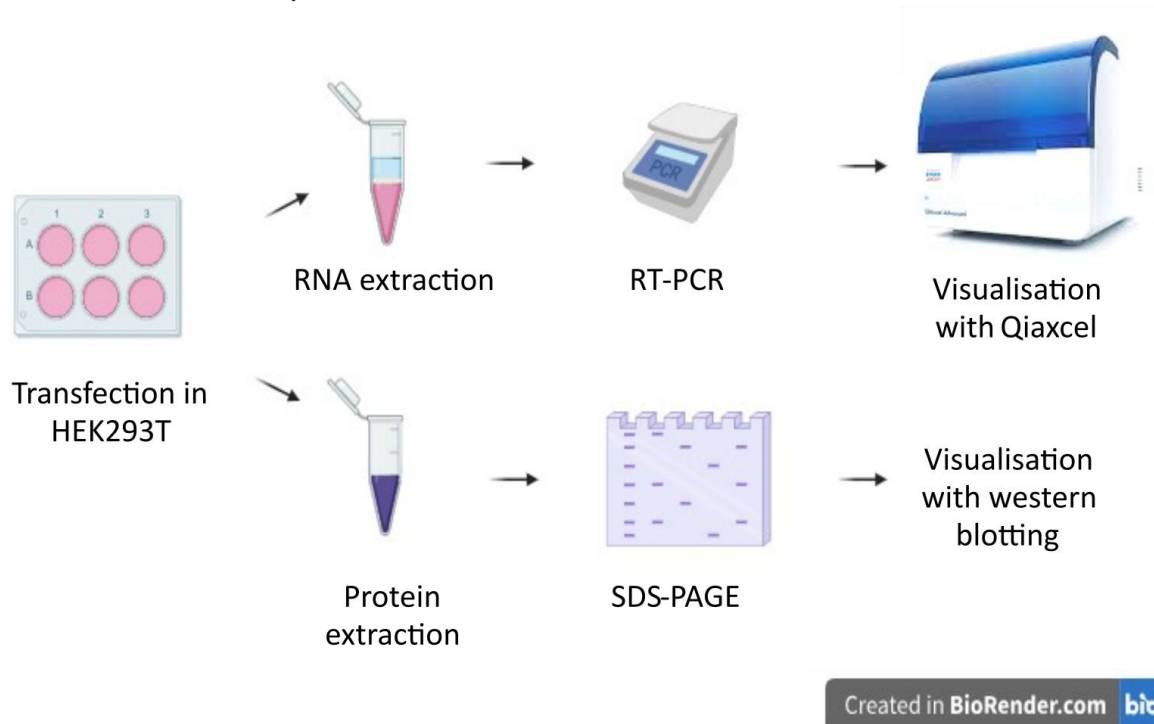
PTPRF is a receptor tyrosine phosphatase and its exon 23 (e23) is a cassette exon which was found to be included in RBPMS overexpression data (Nakagaki-Silva *et al*, 2019). PPFIBP1 is an interacting protein of PTPRF and its exon 10 (e10) is also a cassette exon. Both PTPRF and PPFIBP1 are thought to be important for the regulatory signalling of the focal adhesion complexes of VSMCs, although the exact role the inclusion of their regulated exons have is unknown. CALD1 is an integrin adhesome component and is important for the contractility of VSMCs (Horton *et al*, 2015; Llorian *et al*, 2016; Nakagaki-Silva *et al*, 2019). The alternative splicing event (ASE) of CALD1 involves cassette exon 4 (e4) and an alternative 5'SS on its third exon. This splicing event produces the heavy caldesmon isoform (h-CALD1), a marker of SMC differentiation. As previously discussed, MYOCD is a master transcriptional regulator in SMCs and cardiac cells. Its exon 2a (e2a) was found to be included in the overexpression and aorta datasets, and contains a premature stop codon which causes the use of a downstream Met

residue in exon 4, producing an N-terminally truncated SMC specific isoform (Creemers *et al*, 2006; Miano, 2015; Frismantiene *et al*, 2018; Nakagaki-Silva *et al*, 2019). Exon 2a skipping produces a cardiac specific isoform that retains a domain that interacts with co-activator MEF2, allowing myocardin to regulate the gene expression of cardiac specific genes. In the truncated SMC isoform, this domain is lost, thus the expression of cardiac genes does not occur. FLNB is an actin-binding protein, with the RBPMS regulated exon being a cassette exon, exon 30/H1 (eH1). This exon contains a hinge domain which leads to nuclear localisation of FLNB and is known to antagonise the FOXC1 transcription factor (Li *et al*, 2018; Nakagaki-Silva *et al*, 2019). HSPG2 or Perlecan, is a basement heparan sulfate proteoglycan and is an important component of the extracellular matrix and inhibitor of VSMC proliferation (Lundquist *et al*, 1996; Nakagaki-Silva *et al*, 2019). Its exon 7 (e7) is included in the RBPMS overexpression datasets. The *Rbpms* gene itself was found to contain a candidate exon, exon 3a (e3a). This is a potential poison exon that is highly conserved, with stop codons in all three frames, suggesting possible negative autoregulatory activity. The introduction of poison exons or intron retention leading to nonsense-mediated decay is a common mechanism for regulation of gene expression of RNA binding proteins, as well as a more frequent event in differentiated versus proliferative VSMCs (Jangi *et al*, 2014; Llorian *et al*, 2016; Baralle & Giudice, 2017).

RBPMS regulates these exons in transfection studies

To test whether RBPMS can regulate these exons in minigene transfection experiments, the candidate exons were cloned from HEK293T cell genomic DNA into three exon constructs (*PTPRF* e23, *PPFIBP1* e10, *HSPG2* e7, *RBPMS* e3a, and *FLNB* eH1) or cloned between two heterologous constitutive exons in exon-trap minigene constructs (*CALD1* e4), depending on length of intronic regions between regulated and constitutive exons (see methods section 2.1) (Figure 3.1). *Myocd* was previously used in transfection assays in RBPMS-A studies in the lab and had been cloned from PAC1 genomic DNA, and this exon 2a was cloned into the exon-trap minigene construct. These constructs were then co-transfected into HEK293T cells with RBPMS-A or RBPMS-B 3xFlag constructs (see methods section 2.1). Splicing patterns were observed by extracting RNA forty-eight hours post-transfection and performing RT-PCR with primers specific to the vector constructs (Figure 3.2). Splicing of endogenous targets present in HEK293T cells were also analysed using primers specific to the endogenous regions of these transcripts.

A Transfection experiment



B Qiaxcel visualisation

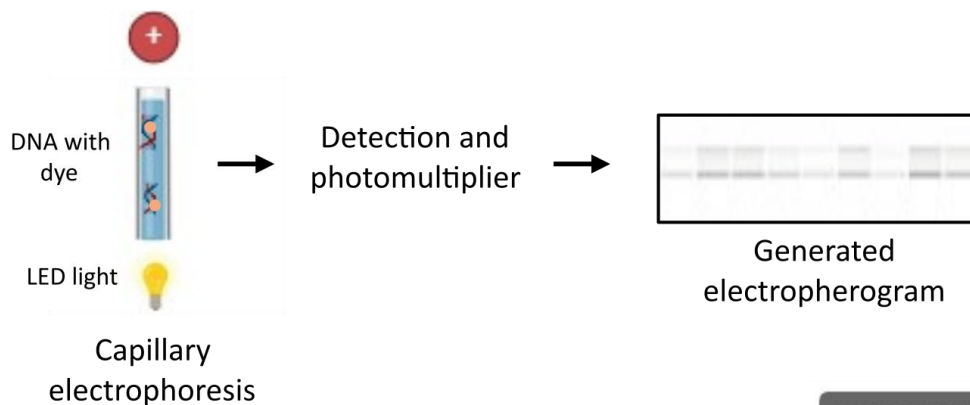


Figure 3.2 Transfection experimental setup: **A:** Schematic demonstrating the co-transfection protocol. **B:** Schematic showing process of data visualisation using the Qiaxcel (QIAGEN). Figure made using biorender.com

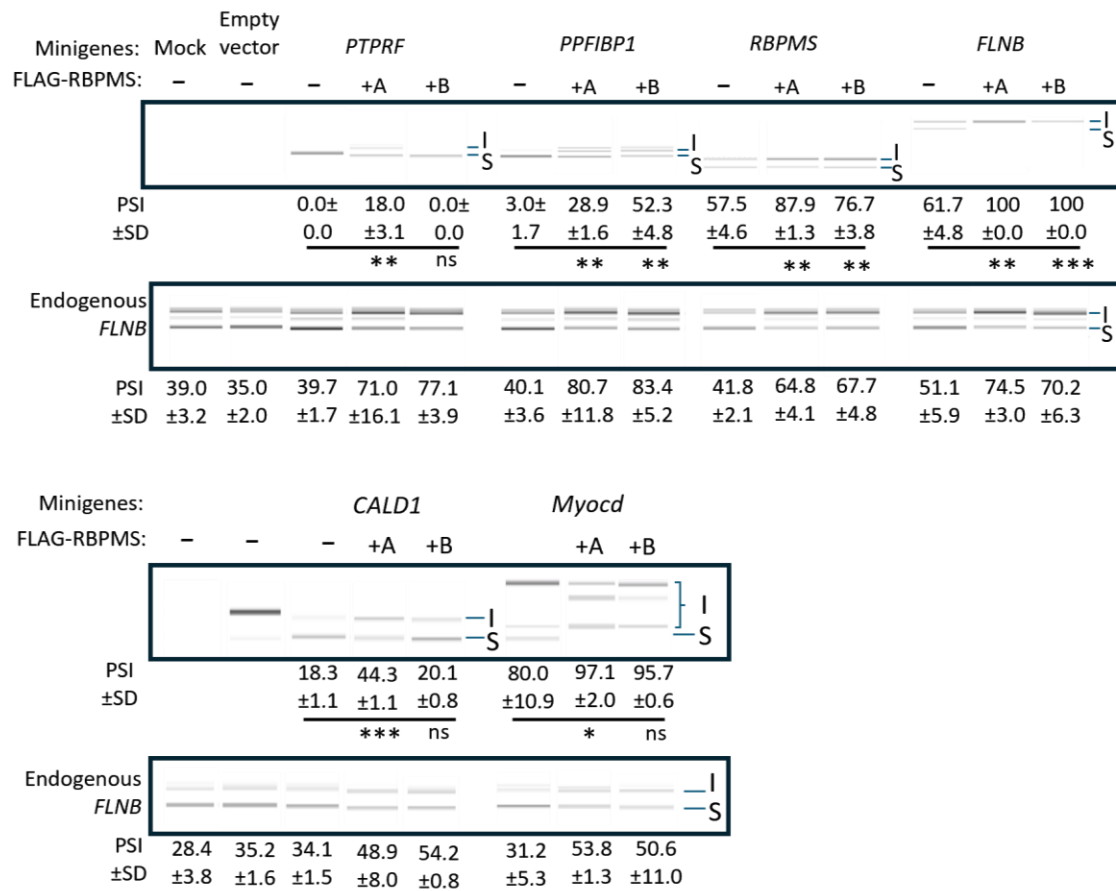
Figure 3.3 shows the splicing outcomes of the co-transfection assays of the minigenes with both isoforms of RBPMS, and this is summarised in Table 3.1. *HSPG2* exon 7 showed no response with RBPMS in co-transfection assays and thus was not pursued further (data not shown). Endogenous *FLNB*

splicing of exon H1 was analysed in addition to the minigene constructs to confirm that RBPMS was active and able to regulate splicing events with this transfection setup. The regulated exons in *PPFIBP1*, *FLNB*, and *BPMS* showed increased inclusion in response to both RBPMS-A and RBPMS-B, and *PTPRF*, *CALD1* and *Myocd* only responded to RBPMS-A. All targets showed significant increases in inclusion of their regulated exons upon addition of RBPMS-A, with the greatest change in percentage spliced in (Δ PSI) being for the *FLNB* minigene, at ~40% (Fig.3.3). Thus, they will likely all be useful targets to continue investigating RBPMS-A splicing activity.

Minigene	Coordinates of genomic regions	Activation by RBPMS?
<i>CALD1</i>	hg38: chr7:134,934,768-134,936,643	Only activated by RBPMS-A
<i>FLNB</i>	hg38: chr3:58,138,283-58,142,750	Activated by both isoforms
<i>HSPG2</i>	hg38: chr1:21,887,934-21,890,141	No response in co-transfection assays (data not shown).
<i>Myocd</i>	rn7: chr10:49,886,536-49,887,083	Activated by RBPMS-A only
<i>PPFIBP1</i>	hg38: chr12:27,656,616-27,660,943	Activated by both isoforms
<i>PTPRF</i>	hg38: chr1:43,609,382-43,613,731	Activated only by RBPMS-A
<i>BPMS</i>	hg38: chr8:30,477,797-30,479,379	Activated by both isoforms

Table 3.1: Location of minigene genomic regions and a summary of their response to RBPMS in co-transfection assays.

A



B

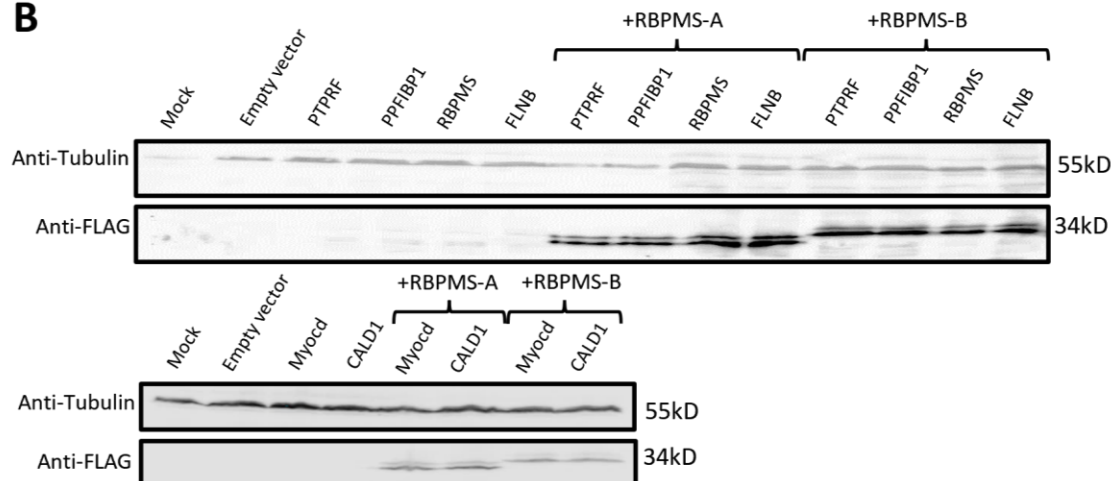


Figure 3.3 Minigenes show differential response to RBPMS-A and RBPMS-B: Results from the co-transfection of RBPMS-A or RBPMS-B and the minigene targets. **A:** Qiaxcel data from RT-PCR showing splicing outcomes of the minigenes when RBPMS-A or RBPMS-B is added. **B:** Western blot showing protein expression. I = Included, S = Skipped, PSI = percentage spliced in. Statistical tests were student *t* tests comparing no RBPMS samples with the samples where RBPMS-A or RBPMS-B had been co-transfected for each minigene: * *p* < 0.05, ** *p* < 0.01 and *** *p* < 0.001.

There are upper bands present for the splicing results of *PTPRF* and *PPFIBP1* which are larger than the expected size of the inclusion band (Fig.3.3). Sequencing suggested that this was likely a heteroduplex product formed of both the splicing skipping and inclusion products, a common artefact of PCR between sequences of similar identity including splicing products (Eckhart *et al*, 1999; Thompson *et al*, 2002). Slow-cooling and reducing cycle number diminished the amount of heteroduplex present, but they were not able to be eliminated completely. Adding single strand endonuclease may have proven more effective in removing them (Thompson *et al*, 2002). As these heteroduplexes contained inclusion and skipped products, they were counted as both inclusion and skipped product in the calculation of PSI (see Section 2.2). In the splicing results of the *Myocd* construct, there are two upper bands that are much larger than the inclusion or skipped products. These bands were found to be a cryptic exon inclusion/partially spliced product, and a product where the downstream intron after the regulated exon was included (Fig.3.4). The cryptic exon product is likely forming due to a cryptic splice site present within the vector used for this construct, mutating this cryptic site would likely solve this issue. The intron retention product has not been observed for endogenous *Myocd* and is likely an artefact of the exon-trap vector as the downstream intron is a hybrid of the *Myocd* original intron and that of the vector, originating from *Tpm1*. As there were other targets that showed greater Δ PSI than *Myocd*, this construct was not pursued further.

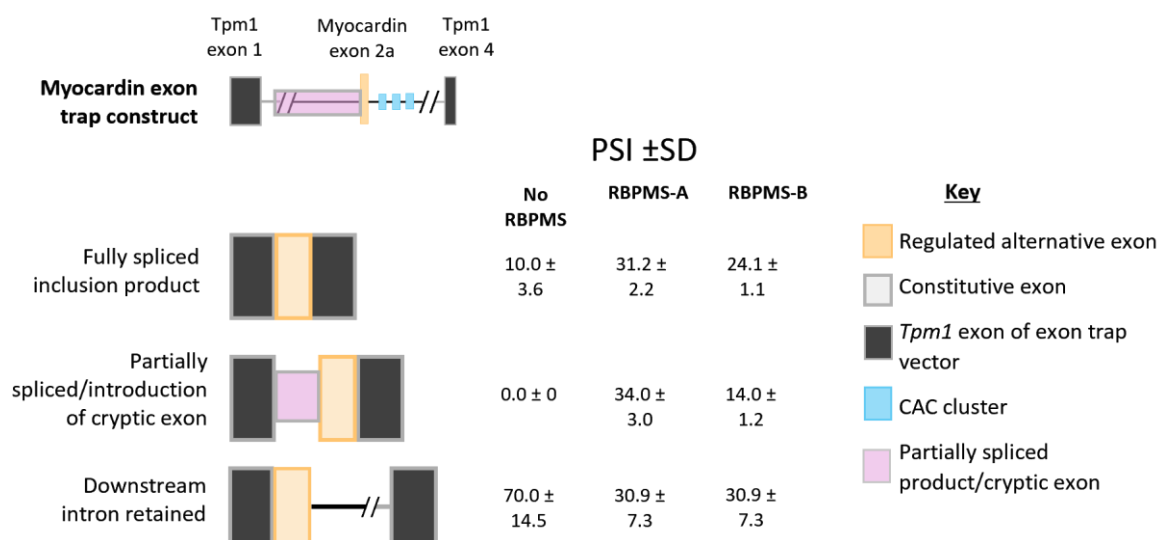


Figure 3.4 Myocd PCR bands: Sequencing results identified the other bands present in the RT-PCR results from Mycod minigene. Top schematic shows full-length construct and location of potential cryptic exon. The Percentage Spliced In (PSI) of each band is shown before and after addition of each isoform.

In four of the six candidates (*PTPRF e23*, *BPMS e3a*, *CALD1 e4* and *Mycod e2a*), BPMS-A activated the exons to a greater extent than BPMS-B, with the *FLNB* and *PPFIBP1* constructs being the exceptions. In previous studies, BPMS-A was found to be the more active isoform in regards to it being able to both repress and activate splicing events, whereas BPMS-B is only able to activate splicing events (Nakagaki-Silva *et al*, 2019). The results in Figure 3.3 show that in some cases BPMS-B is also less effective as an activator than BPMS-A, however, in the case of *PPFIBP1* BPMS-B addition led to a greater level of inclusion than with BPMS-A. The extent of the difference in activation activity of the BPMS isoforms would ideally be further investigated with a greater selection of targets.

3.2 PTPRF: AN IDEAL TARGET FOR THE *IN VITRO* SPLICING ACTIVATION ASSAY

Preparation of constructs for in vitro splicing assays

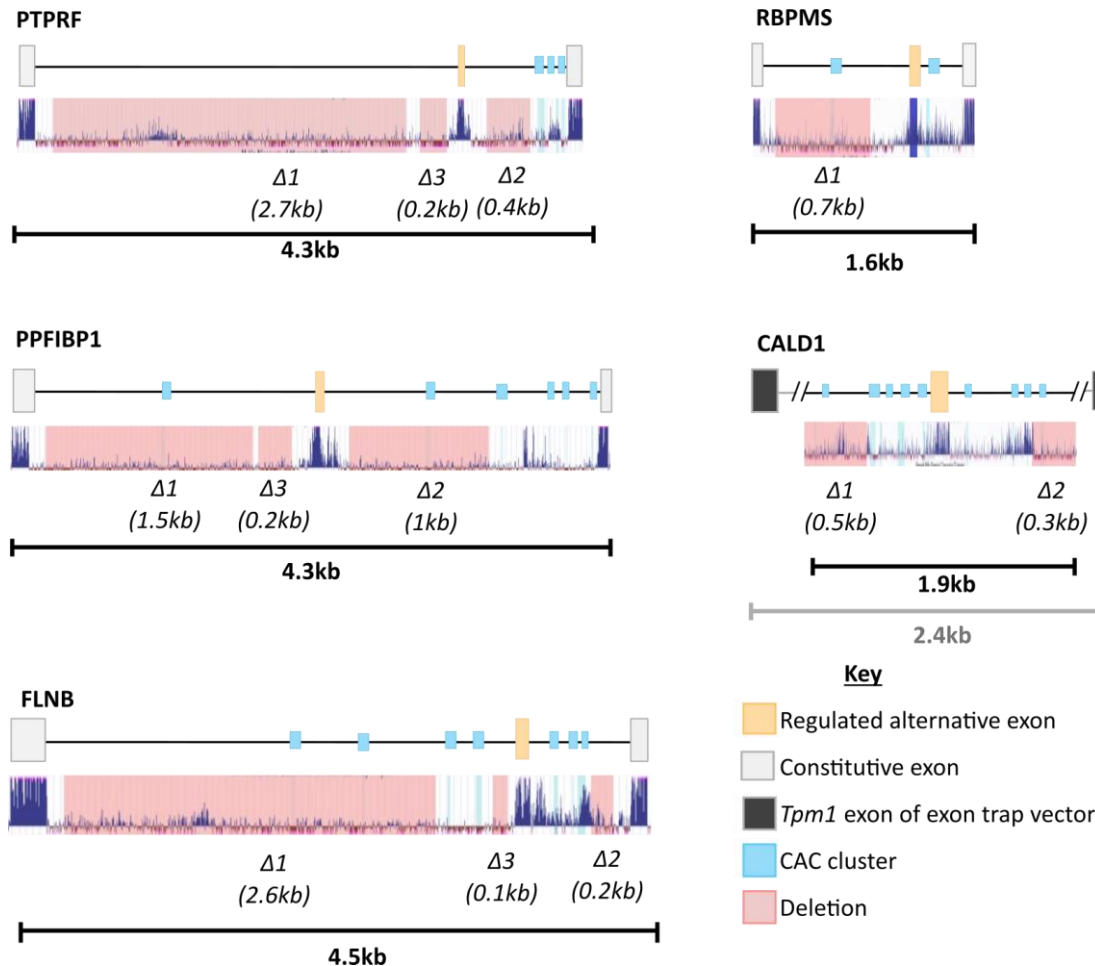


Figure 3.5 Minigene construct conservation and deletions: Schematics show the length of each minigene, their conservation, CAC sites and where deletions were made. The conservation diagrams were produced using the UCSC browser, displaying 100 vertebrates Basewise Conservation by PhyloP. The CAC clusters refer to where there were potential RBPMS binding sites, formed from (CAC_{n1-12}CAC).

The candidate exon constructs that were shown to be responsive to RBPMS in transfection studies needed to be optimised to allow for efficient *in vitro* transcription and splicing reactions (Krainer *et al*, 1984). This meant reducing the sizes of the constructs, ideally to ~1kb or below. All the candidate exons are conserved in mammals (Fig.3.5), and therefore it is likely that the important splicing regulatory elements will be conserved. Initially, large deletions were made in less conserved areas that did not contain RBPMS binding sites (CAC_{n1-12}CAC). These deletion constructs were then tested in co-transfection assays with 3xFLAG RBPMS-A to test whether basal splicing had changed and whether they still responded to RBPMS (Fig.3.6). Further deletion constructs were made (Fig.3.7) and tested in the same way.

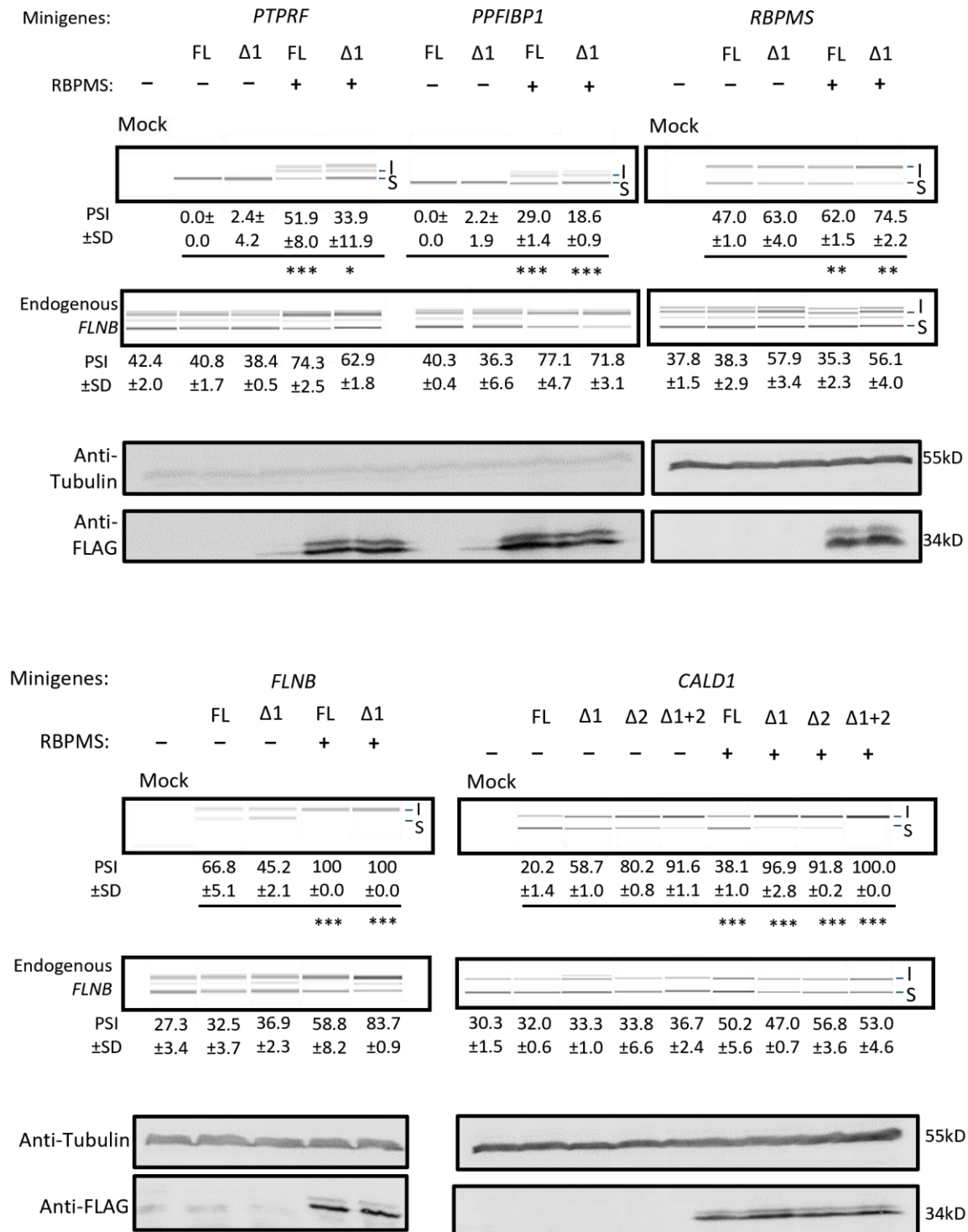


Figure 3.6 Minigenes still respond to RBPMS-A after the first round of intronic deletions: Results from minigene deletion co-transfection with RBPMS-A alongside full-length constructs to assess if they can still be efficiently activated by RBPMS. I = Included, S = Skipped, PSI = percentage spliced in. Statistical tests were student t tests comparing -RBPMS and +RBPMS samples: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

The basal level of activation of *PTPRF* e23 remained the same as the full-length construct after the initial ~2.7kb deletion (Fig.3.6), but its response to RBPMS-A was dampened. Endogenous *FLNB* was also lower for this sample suggesting that the differences in activation may not be due to the deletion but due to lower RBPMS-A activity overall. For *PPFIBP1* e10, however, there does appear to be a reduced response with RBPMS-A after the deletion, as here endogenous *FLNB* activity is similar between the full-length and deletion samples. The deletion construct of *RBPMS* had a higher level of basal inclusion than the full-length construct but the Δ PSI with and without RBPMS-A present was similar for both constructs. The *FLNB* deletion construct had a reduced basal level of inclusion, but exon H1 was fully included in the presence of RBPMS-A, leading to a greater Δ PSI than the full-length construct. *CALD1* deletions all showed an increase in basal level of inclusion, but Δ PSI after the addition of RBPMS-A were all significant, with Δ 1 having the greatest Δ PSI. These deletions left the construct still relatively long, and further deletions were difficult to obtain due to the high frequency of conserved CAC clusters. Therefore, this construct was not pursued further.

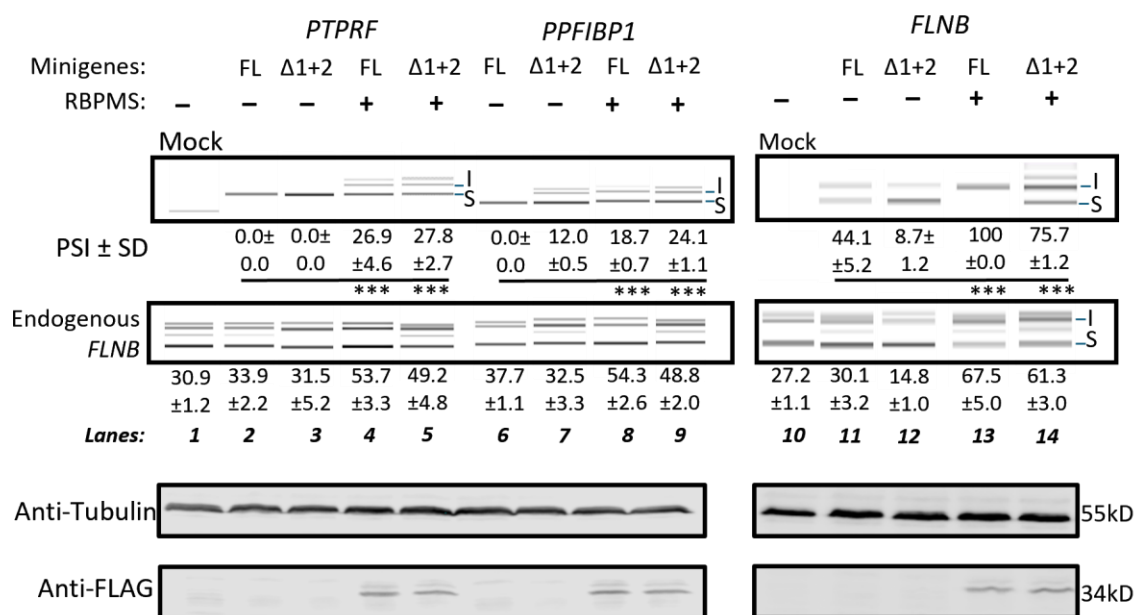


Figure 3.7 *PTPRF* was found to be the ideal the candidate for in vitro studies: Results from second deletions of the intronic regions of candidate minigenes followed by co-transfection with RBPMS-A. I = Included, S= Skipped, PSI = percentage spliced in. Statistical tests were student t tests: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

The double *FLNB* deletion construct ($\Delta 1+2$) also had reduced basal splicing but a significant activation of the regulated exon with RBPMS-A (Fig.3.7), however, subsequent deletions lost any basal inclusion and as well as regulation by RBPMS-A (not shown). The second deletion brought *FLNB* down to 1.7kb which may not be efficiently transcribed and spliced *in vitro* and so initially this target was not pursued. The double *PPFIBP1* deletion construct ($\Delta 1+2$) had a reduced Δ PSI compared to full-length and subsequent deletions led to no splicing inclusion and further deletions lost splicing activation (not shown). *PTPRF* was able to be reduced to 1.2kb with a second deletion and this double deletion construct retained the same level of splicing activation by RBPMS-A as the full-length construct (Fig.3.7). A third deletion was made of *PTPRF* (Fig.3.5), but later it was found to reduce the efficiency of splicing *in vitro* compared to the second deletion mutant (data not shown). The first deletion of *RBPM5* also brought it down to an appropriate size for *in vitro* transcription (0.9kb) and splicing, however it was not efficiently spliced *in vitro* (not shown). Based on these transfection experiments, *PTPRF* was selected as the most suitable candidate for *in vitro* studies.

Testing the in vitro constructs in splicing assays

In order to test the activation constructs *in vitro*, recombinant RBPMS-A was purified. N-terminally 6xHis tagged RBPMS-A was expressed in *E. coli* and purified initially with a HisTrap column (GE Healthcare) (Fig.3.8A). The His-tag was then cleaved with TEV protease (made recombinantly, see section 2.3), and the recombinant RBPMS-A was purified with a Mono Q column (GE Healthcare), and clean fractions were selected (Fig.3.8B, lanes 3 – 9).

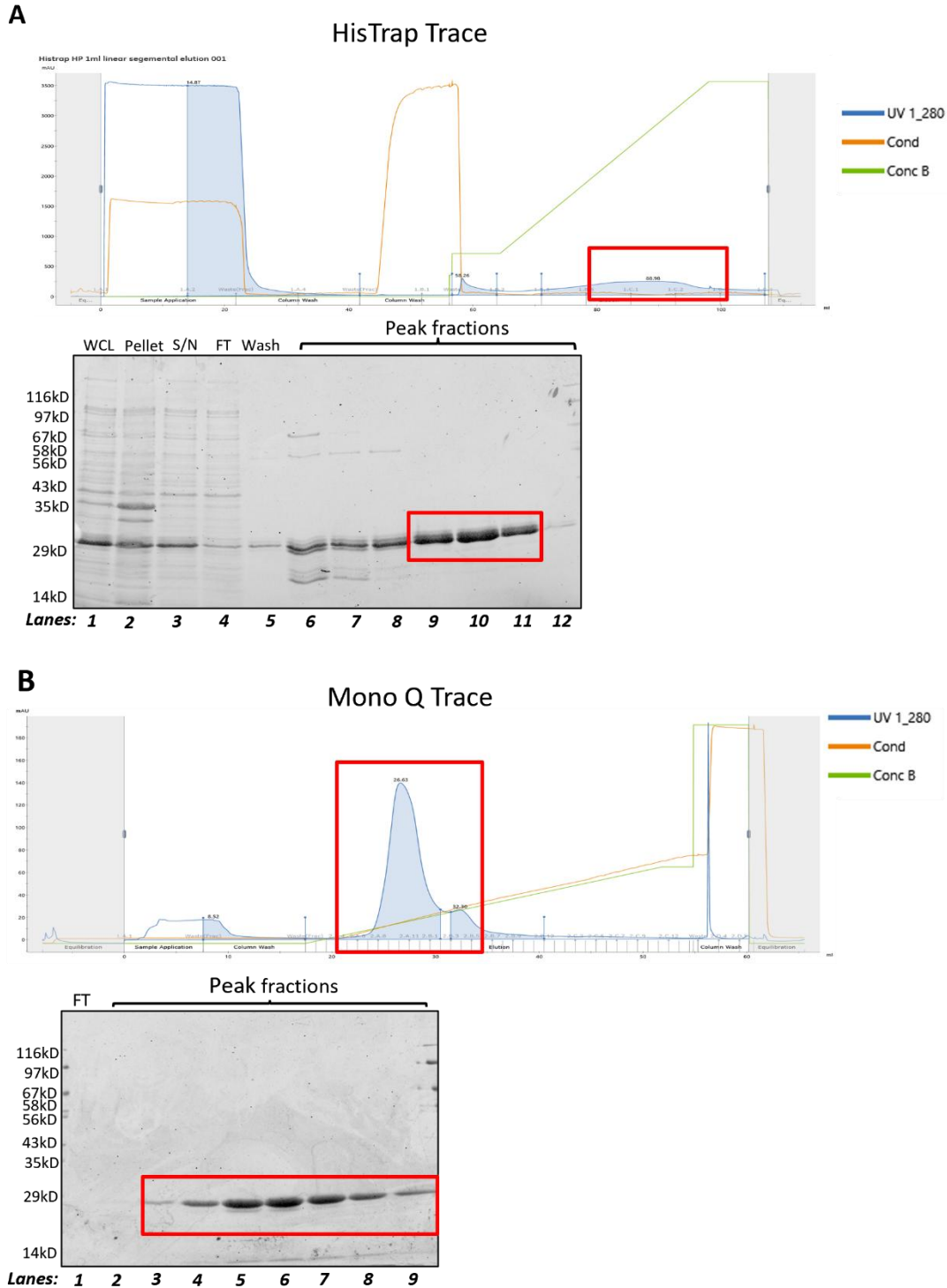


Figure 3.8 Purification of recombinant RBPMS-A: **A:** HisTrap column purification of recombinant His-RBPMS-A. Top = UV/Vis spectra traces produced by the UNICORN AKTA software. Bottom = Coomassie stained SDS-PAGE gels showing RBPMS-A purification after HisTrap. **B:** Mono Q column purification of recombinant RBPMS-A. Top = UV/Vis spectra traces produced by the UNICORN AKTA software. Bottom = Coomassie stained SDS-PAGE gels showing RBPMS-A purification after Mono Q purification. The red box indicates the clean fractions that were taken for future experimental work. WCL = whole cell lysate before purification, S/N = supernatant before HisTrap purification, FT = flow-through during column purification, Cond = conductivity (salt concentration), Conc B refers to concentration of either HisB buffer or QB (see section 2.3), UV 1_280 refers to UV absorbance at $\lambda = 280\text{nm}$.

Initial *in vitro* splicing assays involved radiolabelling mRNA by transcription with $\alpha[32\text{P}]\text{-UTP}$, performing the splicing reaction at 30°C in HeLa nuclear extract and running the splicing products out on a urea-PAGE sequencing gel (Section 2.4) (Krainer *et al*, 1984; Yang *et al*, 2023). T7 RNA polymerase promoter sequences were added at the start of the transfection constructs for *in vitro* transcription, and a more conserved 5'SS (GTAAGT) was added to the downstream constitutive exon for *PTPRF*. This 5'SS is stronger and allows for more efficient splicing, as the initial attempt with the original 5'SS yielded very few splicing products (data not shown). A time-course splicing assay for *PTPRF* was carried out, with 0 hr to 3 hr incubation with HeLa nuclear extract, with and without $2\ \mu\text{M}$ recombinant RBPMS-A added (Fig.3.9). Despite adding the stronger 5'SS, splicing of *PTPRF* was still relatively inefficient, although the final splicing products (both inclusion and skipping) can be seen on the gel (lanes 1 – 4). The fully spliced inclusion product increases with time and with the addition of RBPMS-A (lanes 5 – 8), and intermediate products disappear. This suggests that *PTPRF* can be spliced *in vitro*, and that RBPMS-A increases inclusion.

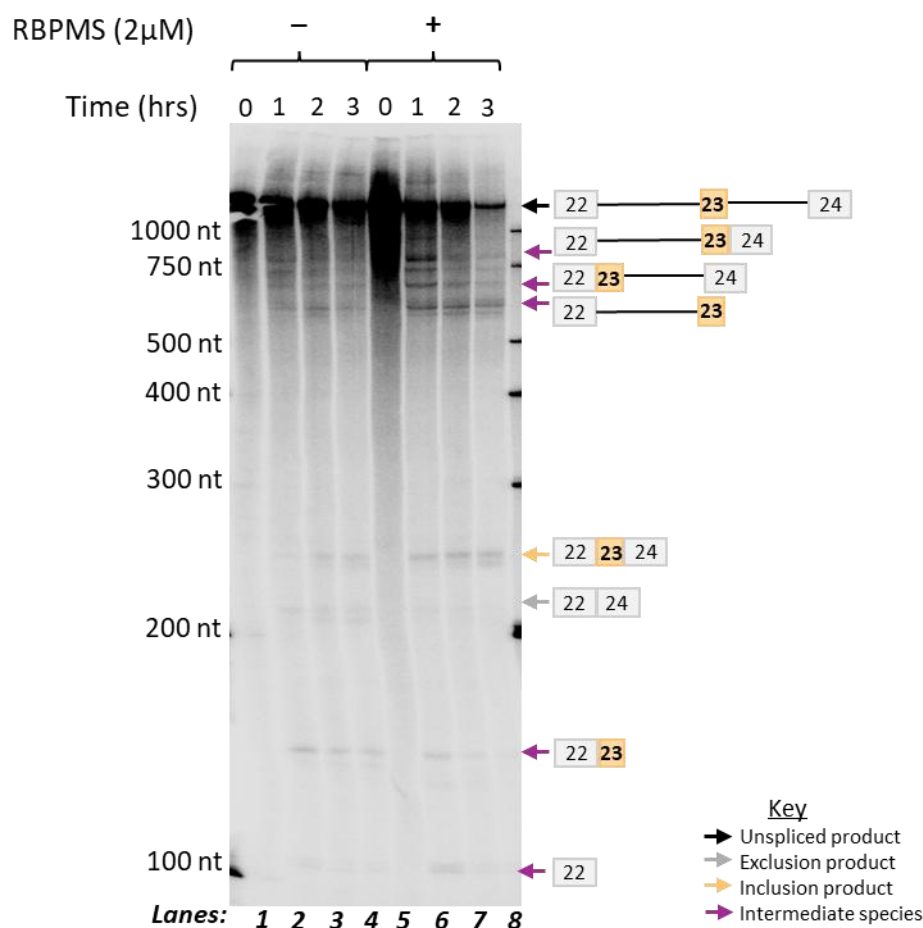


Figure 3.9 RBPMS activates PTPRF exon 23 in vitro: Radiograph showing the splicing species of PTPRF from a time-course assay from 0 – 3 hr, with or without $2\ \mu\text{M}$ of RBPMS-A. nt = nucleotides.

Titrations of RBPMS-A up to 2 μ M were performed in splicing reactions carried out for 3 hr (Fig.3.10A). The intermediate products show changes as RBPMS-A concentration increases, which further indicates that *PTPRF* splicing is activated by RBPMS-A, but the low amount of final splicing products suggest that overall splicing of *PTPRF* is not very efficient (Fig.3.10A, lane 1 versus 4). Lariat products and intermediates move at different speeds compared to linear RNA, making it difficult to discern which bands relate to which potential lariat species (Padgett *et al*, 1984). To elucidate these structures, debranching experiments were performed to linearise the lariat structures after the splicing reactions had occurred. These products were run in parallel with the unbranched products to identify the lariat species (Fig.3.10B). From these debranching experiments, as well seeing an increase in the fully spliced inclusion product when RBPMS-A is added, the upstream intron inclusion lariat species also appears upon addition of RBPMS-A. Additionally, the exclusion lariat species recedes when RBPMS-A is added in lane 3 in figure 3.10B. This species is a fully spliced skipped product, and so as *PTPRF* exon 23 gets activated by RBPMS, the amount of this species diminishes. The appearance and disappearance of these species respectively, indicates RBPMS-A is activating inclusion of exon 23 of *PTPRF* *in vitro*.

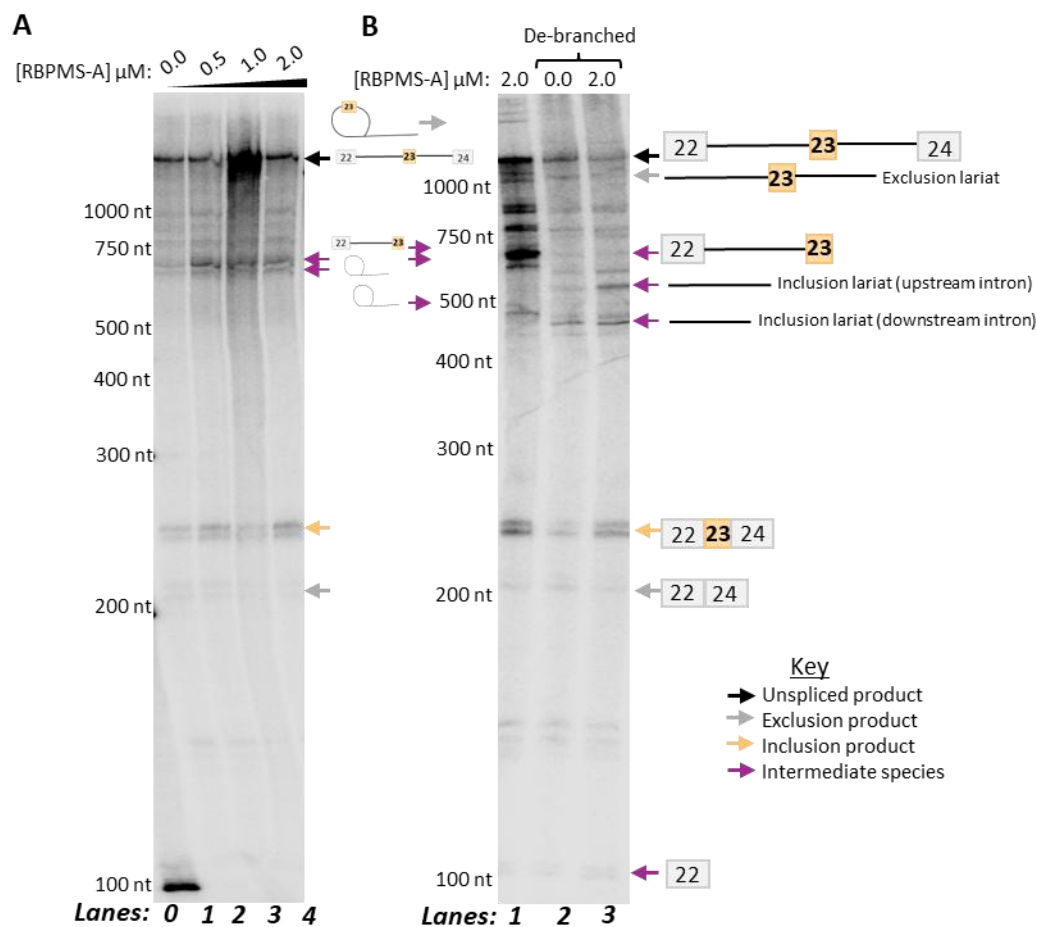


Figure 3.10 RBPMS titration shows some reciprocal change with PTPRF: A: Radiograph displaying the splicing species of *PTPRF* with titration of RBPMS-A. **B:** Radiograph with and without a de-branching reaction to elucidate the lariat species which normally run with irregular speeds due to their structure.

Compared to our previous model of repressed *Tpm1* exon 3, which is spliced efficiently *in vitro* and can respond to RBPMS-A with reciprocal changes in exon inclusion and skipping products, the *PTPRF* transcript was much less efficiently spliced (Yang *et al*, 2023). As an alternative analysis of *in vitro* splicing, RNA constructs were transcribed without radiolabelling, and the splicing was instead analysed by RT-PCR and Qiaxcel (Figure 3.11). This is a more indirect method for measuring splicing, but the amplification step during RT-PCR allows for more sensitive detection of splicing products from inefficiently spliced constructs.

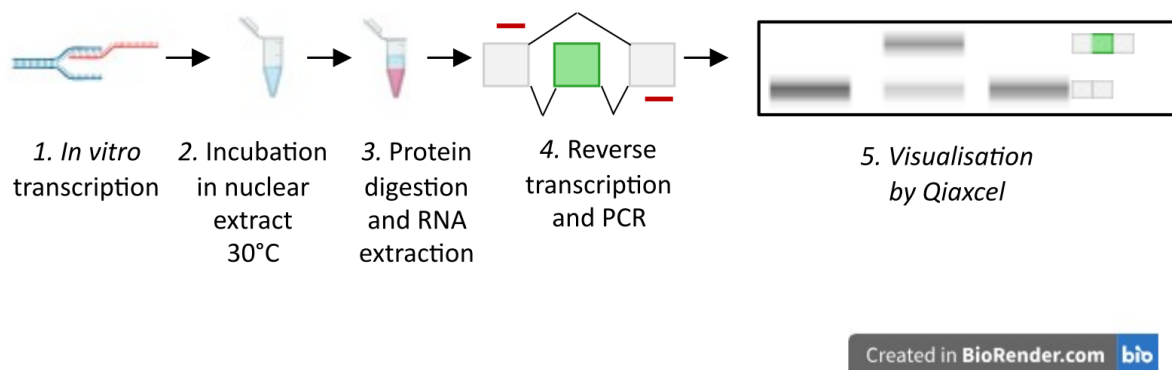


Figure 3.11 “Cold” *in vitro* splicing assay: Schematic showing the *in vitro* splicing assay method with RT-PCR, with splicing outcomes shown on the Qiaxcel automated electrophoretogram. Red lines on the splicing construct represent the primers used during RT-PCR. Made using biorender.com.

Time-course assays with the RT-PCR method were performed (Fig.3.12), for activation constructs *PTPRF* and *FLNB*, as well as the repression target *Tpm1*, which has been used extensively for *in vitro* studies (Mullen *et al*, 1991; Barnhart *et al*, 2022; Yang *et al*, 2023). *FLNB* was reduced to 1.7kb by the second deletion (Fig.3.7), which although still long, showed a relatively large change in PSI upon co-transfection of RBPMS-A. Therefore, it was included in the *in vitro* splicing assays after *RBPMS* proved unsuccessful. The *Tpm1* construct only contains the non-muscle specific exon 3 that RBPMS-A represses, between constitutive exons 1 and 4.

For each substrate, at the 0 min timepoint, there were no RT-PCR products seen (Fig.3.12). This shows that there is not any mispriming from unspliced RNA, as well as proving the primers are specific only for the constructs used and not for any endogenous RNA. The inclusion products of the activated constructs increased as the incubation time increased. For *PTPRF*, the amount of inclusion product continues to increase at every timepoint after RBPMS-A is added. This is likely due to splicing inclusion requiring two separate splicing events to the formation of the fully included product. *FLNB* shows a similar trend, although splicing inclusion is only seen after 120 min. The Δ PSI for the *PTPRF* construct

was greater between the -RBPMS and +RBPMS lanes than with *FLNB*. Therefore, *PTPRF* was the target selected for future activation experiments, and 3 hr was the chosen incubation time for future experiments. The band at the bottom of the *FLNB* lanes was identified by sequencing to be a template shift which gave a partial skipped product and thus was counted towards exon skipping. RBPMS-A is known to repress exon 3 of *Tpm1*, and this is what is observed in the time-course assay for this construct, where the inclusion product of 1-3-4 is diminished as RBPMS-A is added (Nakagaki-Silva *et al*, 2019; Yang *et al*, 2023). The splicing of this construct in this assay does not appear to be efficient as there is an intermediate product on the gel image, as this construct still has exon 3 it was counted as an inclusion product. Overall, this construct responded well to addition of RBPMS-A, giving a Δ PSI of ~50% after only 60 min, and hence will be a useful target to study splicing repression by RBPMS-A in future *in vitro* experiments.

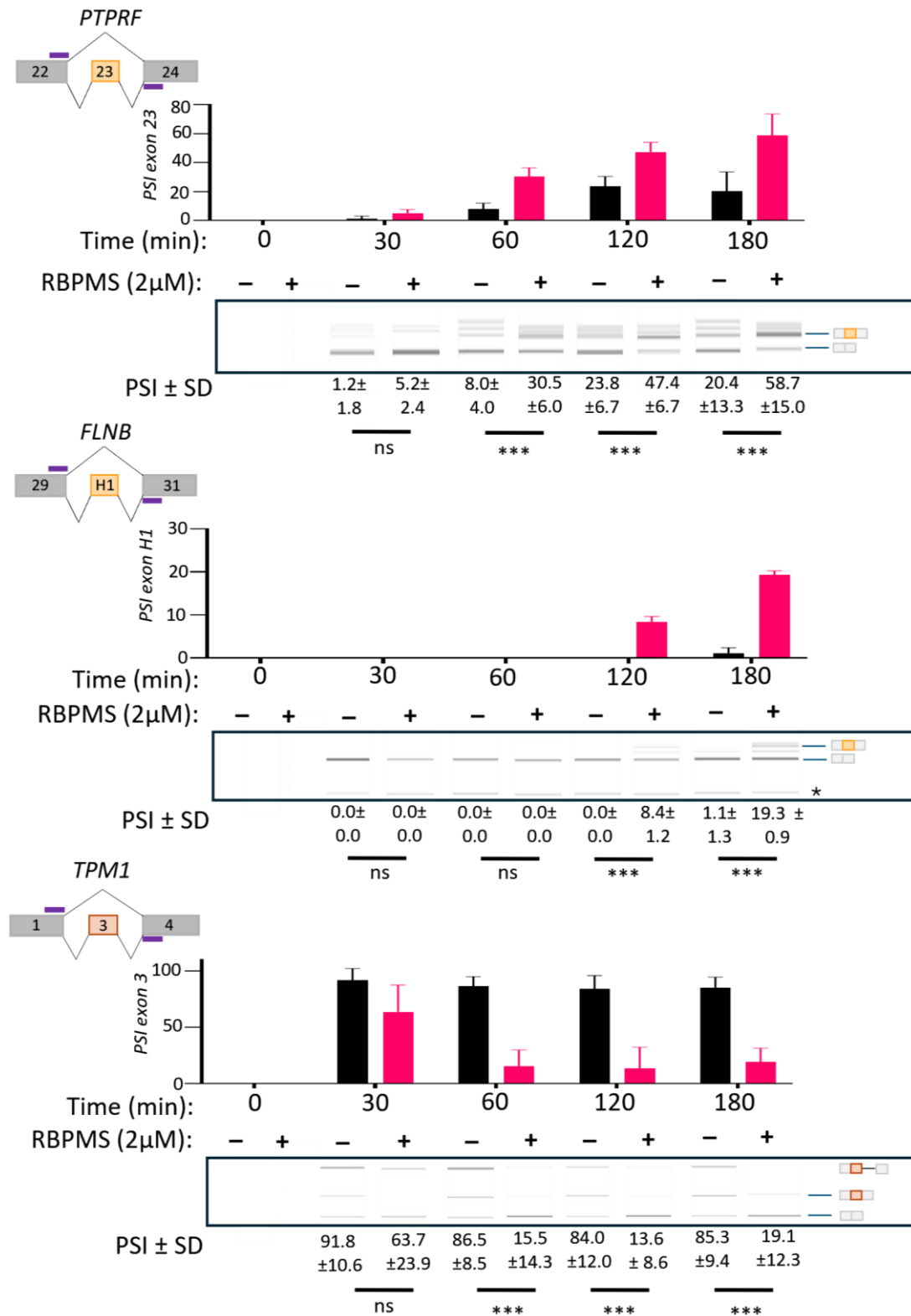


Figure 3.12 Recombinant RBPMS-A regulates splicing in vitro: In vitro splicing assay results using RT-PCR and Qiaxcel. Time-course assays +/-RBPMS with each in vitro construct. The band at the top of Tpm1 was an intermediate product and counted towards inclusion. * template shift formed by partial sections of exons 29 and 31 of FLNB, confirmed by sequencing and counted towards skipping. Statistical tests were either student t test (PTPRF) or Welch test (FLNB and Tpm1) (depending on whether they passed equal standard deviation assumption, comparing +/- RBPMS results at each time-point: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$). I = Included, S= Skipped, PSI = percentage spliced in.

3.3 RBPMS-A DEPENDENT ACTIVATION OF *PTPRF* REQUIRES RBPMS AND MBNL

BINDING SITES

Mutation of RBPMS binding sites leads to a loss of RBPMS-dependent activation

Alternative splicing is regulated by a combination of *trans*-acting factors (RBPs) and *cis* elements (their binding sites) (Figure 1.5). RBPMS binding sites (CAC_{n1-12}CAC) are position dependent for skipped versus included exons. The sites are mainly upstream and within the regulated exon for skipped exons, and for included exons, the sites are often located downstream of the regulated exon (Nakagaki-Silva *et al*, 2019; Jacob *et al*, 2022). The CAC clusters adjacent to RBPMS regulated exon 3 of *Tpm1* were found to be crucial for RBPMS-dependent repression (Nakagaki-Silva *et al*, 2019; Yang *et al*, 2023). *PTPRF* contains three clusters of CACs downstream of regulated exon 23 which are conserved (Fig.3.5 & 3.13A). To investigate the importance of these clusters of RBPMS activity, these regions were initially deleted either alone or in combination with other clusters.

Splicing analysis of these cluster deletions (Fig.3.13B) revealed that none of the clusters in isolation were solely responsible for RBPMS-dependent activation (p-values for the Δ PSI of all single cluster deletions upon RBPMS-A addition were <0.001), but when both clusters one and two were deleted together, or all three were deleted, splicing activation was unaffected by addition of RBPMS-A, and exon 23 inclusion remained at basal level (lanes 9 & 10, p-values = 0.055 and 0.016 respectively). This indicates that RBPMS-A activation of *PTPRF* e23 is mediated by CAC clusters 1 and 2, most likely by direct binding at these locations. Cluster three appears to be redundant for RBPMS-A activity, as deletion of all three clusters did not affect activity more than just deletion clusters one and two together. Deletion of cluster 2 by itself had less of an effect on activity than the other clusters upon addition of RBPMS (lane 6), but when in combination with other clusters its effect was greater when the other clusters, 1 and 3, were deleted together (lanes 10, 12 & 14). This suggests that RBPMS bound at cluster 2 may interact with RBPMS molecules bound at cluster 3, and particularly at cluster 1. Additionally, it could be due to interactions with other co-regulators of RBPMS bound at or between these two clusters.

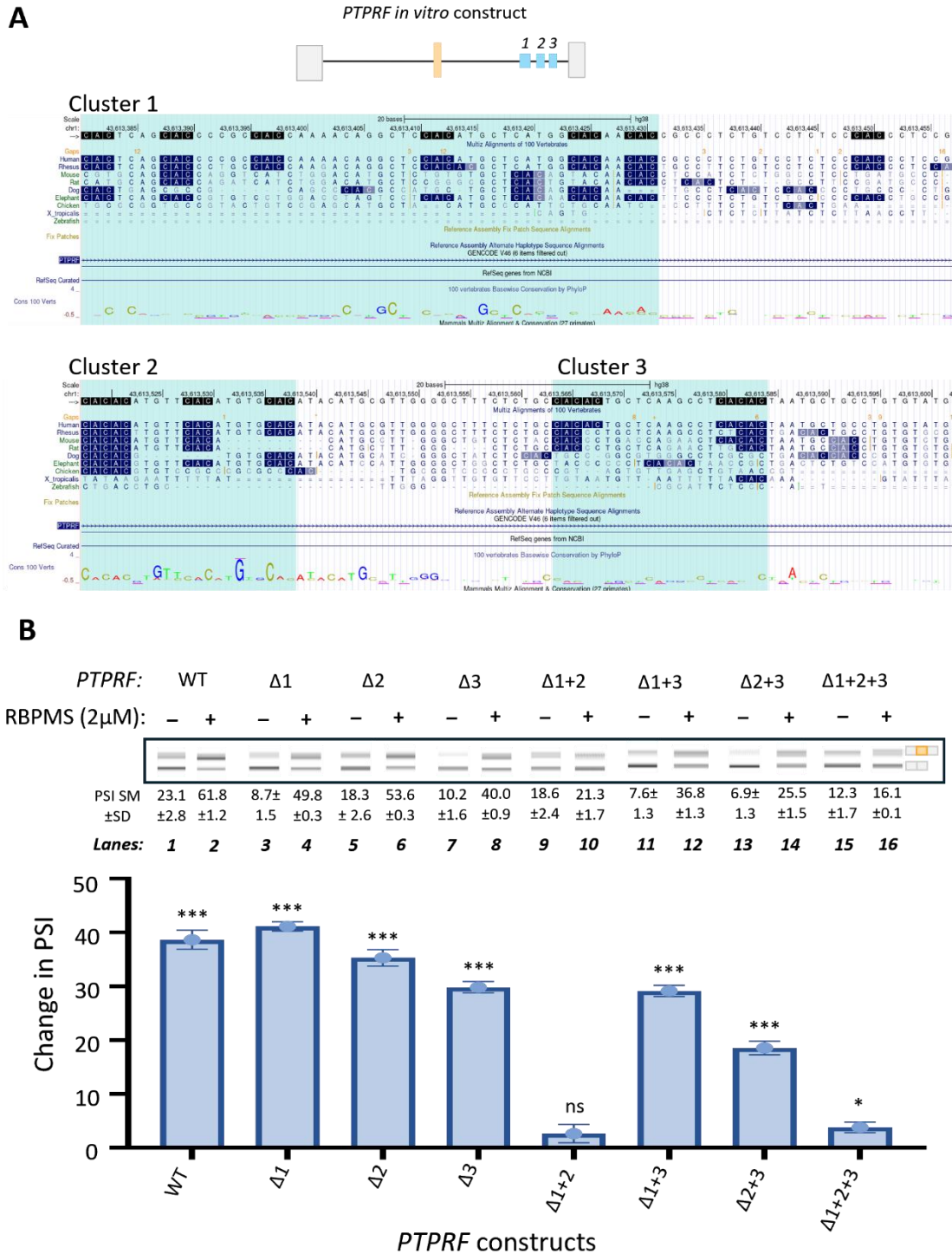


Figure 3.13 Deletion of the first and second CAC clusters abolishes RBPMs-A dependent splicing activation of PTPRF exon 23: **A:** Schematic of in vitro PTPRF construct with location of CAC clusters, and screenshots from the UCSC browser displaying the conservation of the CAC clusters in the downstream intron of the regulated exon (exon 23) of PTPRF. The conservation diagrams were produced using the UCSC browser, displaying 100 vertebrates Basewise Conservation by PhyloP. The CAC clusters are highlighted in light blue, with CACs themselves in dark blue. **B:** Qiaxcel data of in vitro splicing assay with deletion of the CAC clusters of PTPRF $\pm$ 1 μ M RBPMs-A. The graph shows the change in PSI for each construct when RBPMs-A is added at 1 μ M. Statistical tests were student t tests comparing samples with and without RBPMs-A added: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. WT = wild-type, Δ = deletion, PSI = percentage spliced in.

The deletion of the clusters alone led to an increase in the inclusion product when no RBPMS-A had been added (Fig.3.13B, lanes 3, 5 & 7). This may be due to loss of other regulatory sequences that would normally affect the splicing of *PTPRF*. To ensure that the effects observed were due to loss of RBPMS binding sites (CAC₂) and not due to deleting other possible regulatory sequences, the CACs within the clusters found to be important for activity (clusters 1 and 2) were mutated to CCC. RBPMS-A was titrated between 0 – 2000 nM with either the wild-type *PTPRF* construct, or the mutated CAC construct (Fig.3.14A). The mutant construct splicing activation remained the same as the basal level at all concentrations of RBPMS-A. This confirms that CAC clusters 1 and 2 in the downstream intron of *PTPRF* are indeed responsible for RBPMS-A activity and are likely to be where RBPMS-A is binding. In addition, a titration experiment with *Tpm1* with CACs upstream and downstream of the regulated exon mutated was performed (Figure 3.14B). Here, the CAC mutant does have reduced RBPMS-dependent splicing repression activity but does not reduce to basal level. This may be due to other CACs that RBPMS-A may bind to being present, for example within exon 3 itself (Fig.3.15). In future experiments, these sites should be included in a CAC mutant made to investigate whether full loss of RBPMS-dependent splicing repression of *Tpm1* can be achieved with *cis* element mutations.

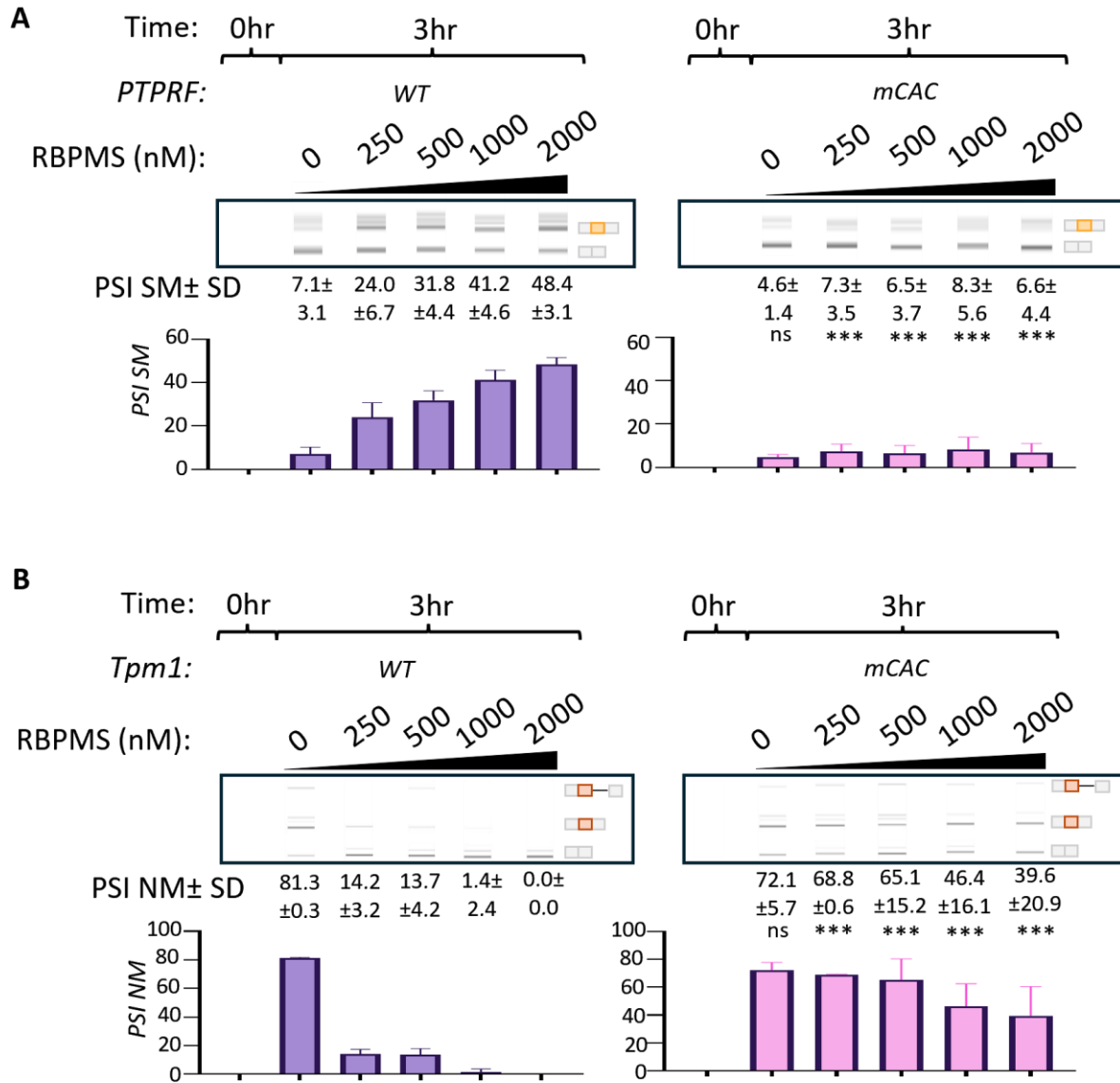
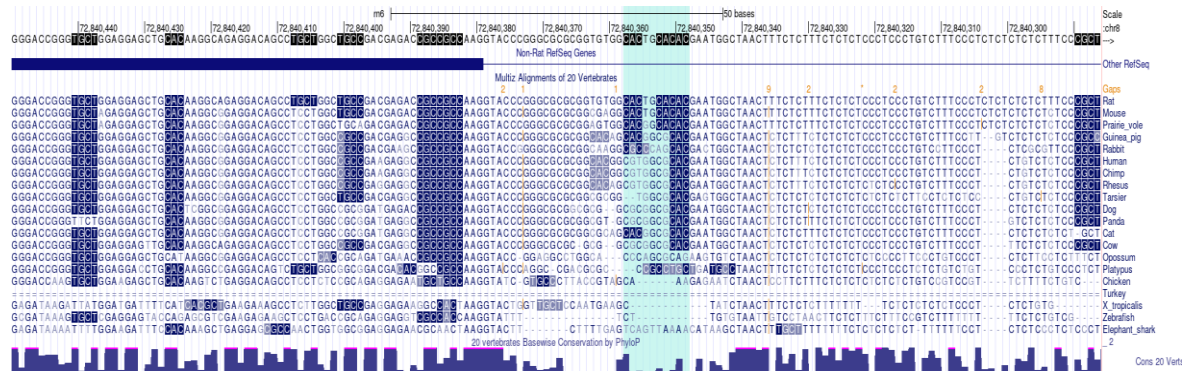
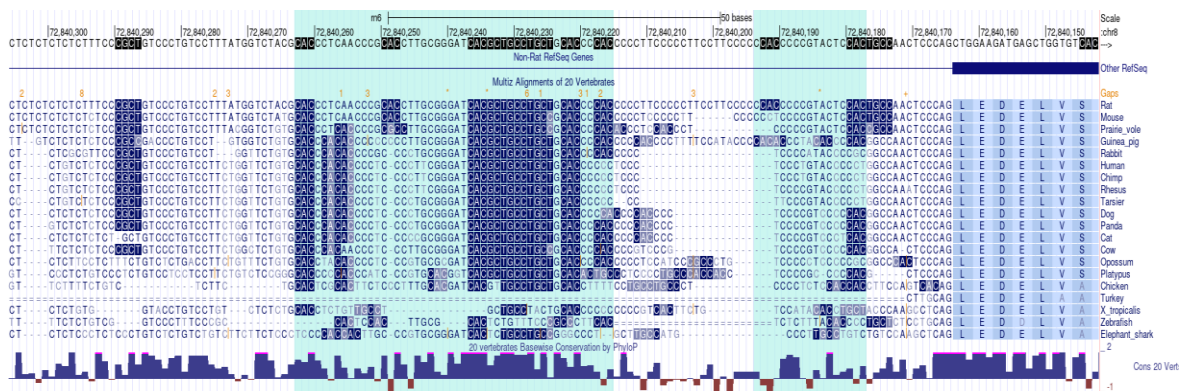


Figure 3.14 The CACs surrounding the regulated exon of *Tpm1* are necessary for RBPMs-A activity: In vitro splicing assay results with RBPMs-A titration with either wild-type construct, or with mutated CAC sites. **A:** PTPRF results. **B:** *Tpm1* results. WT = wild-type transcript, mCAC = transcript with mutated CAC sites, SM = smooth muscle, NM = non-muscle, PSI = percentage spliced in. Statistical test were student t tests comparing each sample with and without RBPMs-A added: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

Upstream intron CAC cluster 1 (and exon 2)



Upstream intron CAC clusters 2 and 3 and beginning of exon 3



Downstream CAC clusters and end of exon 3

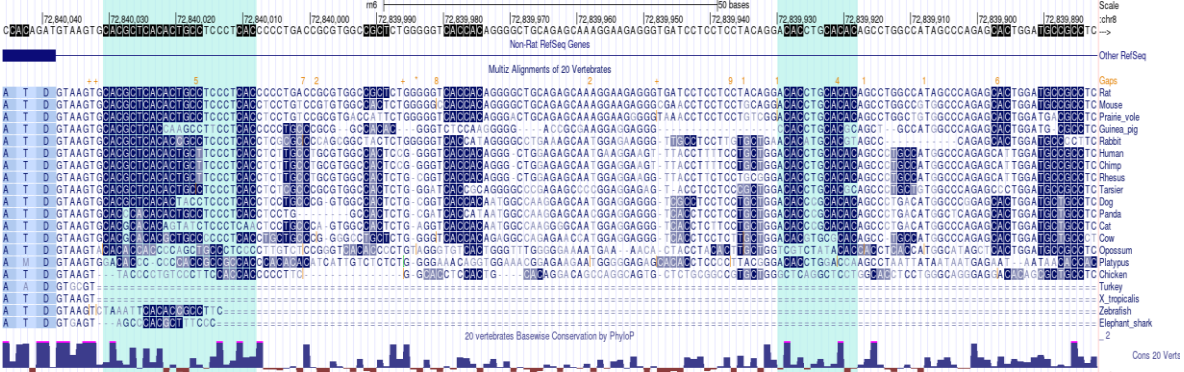
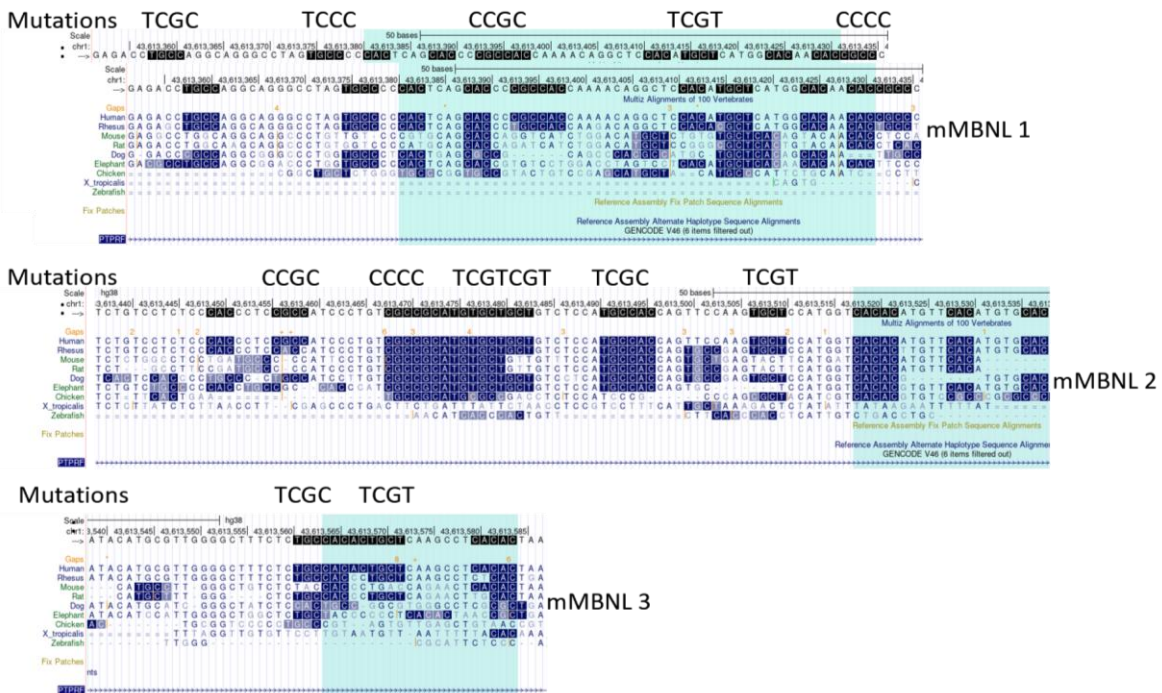


Figure 3.15 Conservation of Tpm1 CAC clusters: Screenshots from the UCSC browser displaying the conservation of the CAC clusters (20 vertebrates Basewise Conservation by PhyloP) in the downstream intron of the regulated exon (exon 3) of Tpm1. The clusters are highlighted in light blue, with CACs themselves in dark blue. The MBNL sites (YGCY) have also been highlighted in dark blue.

MBNL binding sites are also important for RBPMS splicing activity

In addition to the CAC sites towards the end of the downstream intron of *PTPRF*, many YGCY motifs were also observed, and the RBPs Muscle-blind 1/2 (MBNL1/2) are known to bind to YGCY sites and affect alternative splicing (Goers *et al*, 2010; Gooding *et al*, 2013). RBPMS-A has been found to interact with MBNL1 and together they co-regulate a number of both activation and repression ASEs (Yang *et al*, 2023). These sites in the *PTPRF* construct were therefore mutated to either YCGY, or YCCY (in some cases, reversing the order G and C recreated another YGCY motif) to see the effect on basal and RBPMS-A dependent *in vitro* splicing (Fig.3.16). Mutating the MBNL sites around CAC clusters 1 and 2 had the greatest effect of RBPMS-dependent activation of exon 23 (Figure 3.16B, lanes 13 & 14). This suggests that MBNL does influence inclusion of e23, likely by binding at the sites tested here. Interestingly, mutating sites 1 and 2 has a greater effect than mutating all the sites. MBNL site 3 mutation alone increases levels of inclusion relative to wild-type *PTPRF*, this may suggest that the mutations introduced into this cluster aid RBPMS-dependent splicing. This cluster is close to the polypyrimidine tract and branch point, and mutations here may lead to stronger binding of a splicing enhancer factor that may interact with RBPMS in some way to lead to greater inclusion of e23. None of the MBNL site mutations abrogated splicing activation of e23 by RBPMS as much as the CAC mutations. This suggests that MBNL may play a more minor role than RBPMS in the activation of exon 23 in this context.



B

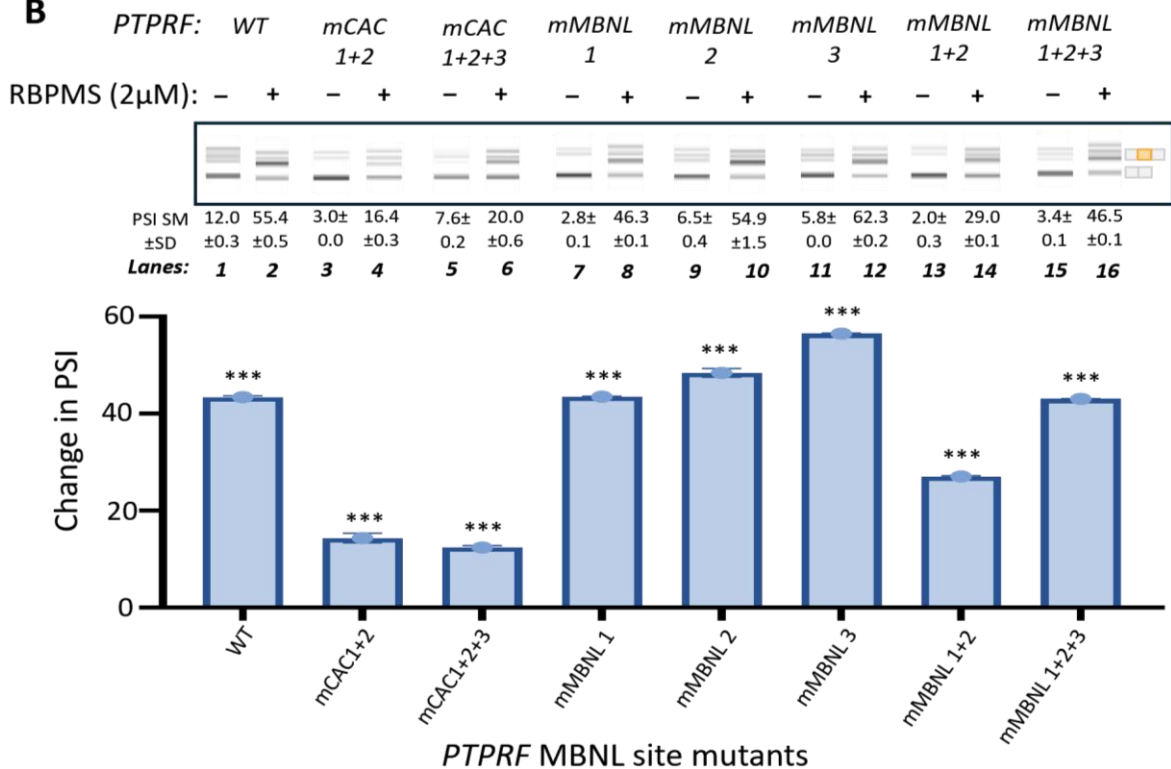


Figure 3.16 MBNL binding site mutations near CAC clusters in PTPRF affect RBPMS-A activity: **A:** Screenshots from the UCSC browser displaying the conservation of the CAC clusters (100 vertebrates Basewise Conservation by PhyloP) and coincidence of MBNL binding sites (YGCY) in the downstream intron of PTPRF exon 23. The CAC clusters 1, 2 and 3 are highlighted in light blue, with CACs and YGCYs in dark blue, as well as a single RBFOX2 site (GCATG) between CAC clusters 1 and 2. Above the images are the mutations introduced to the YGCY sites. **B:** Qiaxcel data of in vitro splicing assay with mutation of the YGCY sites around the CAC clusters of PTPRF. WT = wild-type transcript, mMBNL = transcript with mutated MBNL sites, PSI = percentage spliced in. Statistical test was a student t test comparing samples with and without RBPMS-A added: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.

3.4 CONCLUDING REMARKS

To gain a full understanding of the mechanism of action of the splicing regulator RBPMS, assays to measure repression and activation activity were required. The repression activity of RBPMS-A has been tested with exon 3 of *Tpm1* (Nakagaki-Silva *et al*, 2019; Barnhart *et al*, 2022; Yang *et al*, 2023). In differentiated VSMCs, RBPMS-A represses exon 3, and exon 2 gets spliced in. For the *in vitro* construct, exon 2 had been removed but the regulatory sequences for RBPMS and the splicing machinery were maintained. This allowed the construct to be small enough to be efficiently transcribed and spliced *in vitro* whilst still allowing repression by RBPMS-A. In the *in vitro* studies with this construct, addition of RBPMS-A to the HeLa nuclear extract allowed recapitulation of the differentiated VSMC specific splicing pattern, skipping of exon 3 (Fig.3.14) (Yang *et al*, 2023). As RBPMS activates splicing events as well as repressing them, it was necessary to find a similar target which RBPMS can activate.

An ideal target for activation would need to respond to recombinant RBPMS in HeLa nuclear extract whilst being short enough to be efficiently transcribed and spliced. Co-transfection studies with target minigenes and their subsequent deletion mutants found that *PTPRF*, *RBPMS*, and *FLNB* were promising candidates (Figures 3.3, 3.6 & 3.7). Radiolabelled *RBPMS*, however, was not efficiently spliced in HeLa nuclear extract. *FLNB* was efficiently spliced, but the Δ PSI upon addition of recombinant RBPMS-A was relatively small compared to *PTPRF*. Therefore, *PTPRF* was ultimately chosen to be the activation target to continue elucidating the mechanism of splicing regulation by RBPMS (Chapter 4). Additionally, *PTPRF* may be an interesting target for studying the differences between RBPMS-A and RBPMS-B, as only RBPMS-A activated exon 23 in the co-transfection studies (Fig.3.4). *FLNB* exon H1 was activated equally by each isoform, so may also prove to be an interesting target to observe *in vitro*. Additionally, *PPFIBP1* is likely a good option for studying RBPMS-A versus RBPMS-B, as RBPMS-B had greater activity than RBPMS-A with this candidate. *PPFIBP1* was not able to be reduced to an acceptable size during this study whilst retaining responsiveness to RBPMS, but deletions in different regions of the minigene may prove to be less disruptive to splicing regulation.

RBPMS was shown to bind tandem CAC repeats of RNA separated by 1-12 nucleotides (Farazi *et al*, 2014; Teplova *et al*, 2016). The CACs surrounding the regulated repressed exon of *Tpm1* were shown to be required for the activity of RBPMS, and when mutated they did not show a differentiated VSMC splicing pattern (Yang *et al*, 2023). For activated exons, CAC clusters were often found to be downstream of the regulated exon in differentiated VSMCs (Nakagaki-Silva *et al*, 2019). Furthermore, the CACs in the downstream intron of the regulated exon of *Flnb* (exon H1) were found to be necessary for RBPMS-dependent activation. The downstream intronic CAC clusters were therefore not deleted

from the activation minigene targets when preparing the constructs *in vitro*. The CAC clusters of *PTPRF* were investigated (Figures 3.13 & 3.14A) to see if RBPMS-dependent activation of exon 23 is reliant on these sites. Deletion and mutation experiments identified that the CAC clusters 1 and 2 were most important for RBPMS-dependent activation, and cluster 3 appeared to be redundant. The *Tpm1* CAC mutant was still affected by the addition of RBPMS-A, as exon 3 was still partially repressed at higher RBPMS-A concentrations. This construct also had a noisy basal level of splicing, but the RBPMS-dependent repression of this mutant may also be due to RBPMS-A binding at other sites. For both substrates, response to RBPMS-A was able to be diminished significantly by introducing these CAC site mutations, showing that the activity of RBPMS-A on these targets is specific.

There are a few single CACs near to MBNL sites in the *Tpm1* construct, therefore it is possible that RBPMS-A may not be directly binding to its primary target sequence and is in fact recruited by other proteins, such as MBNL, as was suggested in previous splicing and crosslinking studies with an MBNL site mutant of *Tpm1* (Yang *et al*, 2023). MBNL was also found to directly interact with RBPMS and affect splicing of several RBPMS targets in the same direction upon knockdown in PAC1 cells. MBNL is endogenously expressed in HeLa nuclear extract and so may be regulating the splicing of the *in vitro* constructs. The *PTPRF* MBNL mutant with mutations around the CAC clusters 1 and 2, also showed reduced splicing activation compared to the wild-type construct (Fig.3.16). Hence, these *cis* elements are important for RBPMS activity, and this may be due to endogenous MBNL acting in a similar manner to RBPMS in activating *PTPRF*, as demonstrated by siRNA mediated knockdown experiments of MBNL and RBPMS in PAC1 cells (Yang *et al*, 2023, Huang *et al*, unpublished). These mutants had many repetitive cytosines introduced due to these mutations and thus there was a possibility that these may be introducing new binding sites for other RBPs that bind C-rich sequences and affect splicing, for example hnRNP K (Expert-Bezancon *et al*, 2002; Chkheidze & Liebhaber, 2003; Bampton *et al*, 2021). If this was occurring, the basal level of splicing would likely be different between this mutant and the wild-type of CAC mutant constructs, however, this was not observed.

One limitation of the current *PTPRF in vitro* construct is it is not efficiently spliced, hence the need for the indirect RT-PCR method as a readout for splicing changes. This inefficient splicing means that it is unlikely to be an effective platform for activation complex assembly. Repression complex assembly on the *Tpm1* construct was achieved, but this was because it was much more efficiently spliced *in vitro* (Yang *et al*, 2023). *PTPRF*, however, is still an extremely useful construct as it can be used to uncover the mechanism of action for RBPMS-dependent splicing activation versus repression by comparison with the *Tpm1* construct. Both constructs responded to the addition of RBPMS-A *in vitro* and this response was proven to be specific with the introduction of CAC site mutations, with these *cis* mutants

being able to be used as good negative controls for future experiments testing splicing regulation by RBPMS.

3.5 FUTURE PLANS

To further understand splicing activation and splicing repression by RBPMS, follow up experiments would be required:

- Investigation of the relationship between the splicing activity of RBPMS, and its previously observed higher order assembly behaviour. These *in vitro* splicing assays allow for the direct comparison of activity with biophysical behaviour of recombinant proteins.
- UV crosslinking and immunoprecipitation (CLIP) experiments should be performed to gain a better understanding of where RBPMS binds on these targets and compare this to known binding sites of co-regulator proteins.
- Production of a more efficient *in vitro* activation construct to elucidate the splicing activation complex formed by RBPMS.
- Further development of *PPFIBP1* as an *in vitro* construct to better understand the splicing regulation by RBPMS-B versus RBPMS-A.

4 SELF-ASSEMBLY OF RBPMS IS NOT REQUIRED FOR FUNCTION

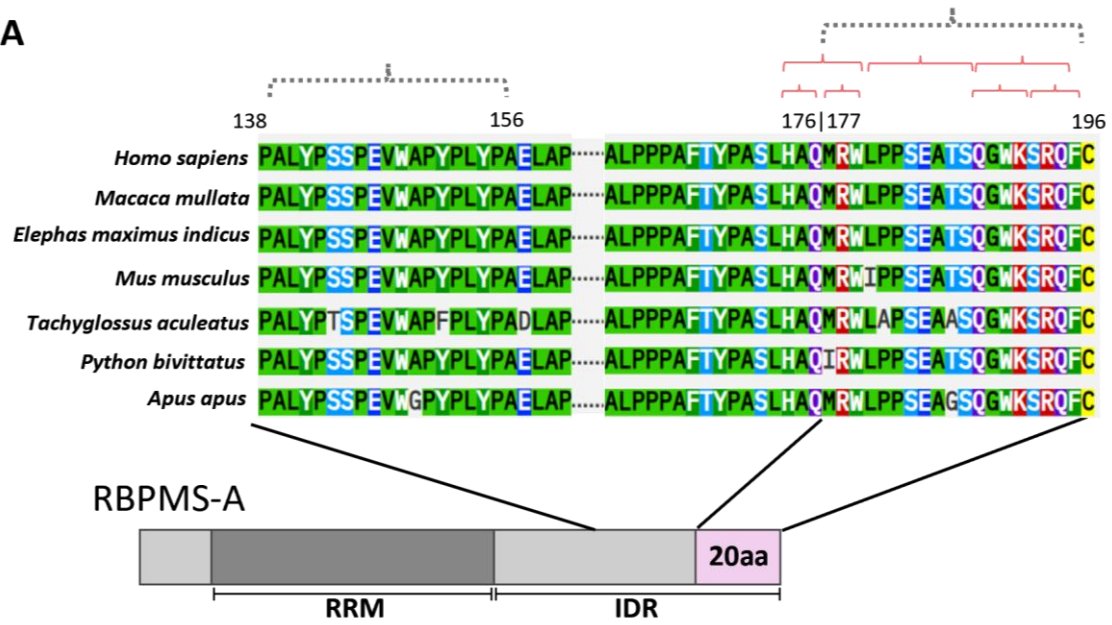
4.1 BACKGROUND

BPMS regulation of splicing is via the C-terminus

Many RBPs have intrinsically disordered regions (IDRs) as well as structured RNA-binding domains. These IDRs have been found to be important for the function of these proteins, for example they can mediate homotypic interactions to form large complexes, in extreme cases mediating phase separation behaviour (Boija *et al*, 2018; Sabari *et al*, 2020). Additionally, these regions are often important for forming heterotypic interactions with co-regulator proteins, as is the case for the splicing regulator protein RBFOX and its LASR complex (Ying *et al*, 2017). These unstructured regions can also aid in RNA binding, and can be subject to post-translational modification such as phosphorylation and arginine methylation, which can in turn alter activity and higher-order assembly (Woodsmith *et al*, 2018; Barnhart *et al*, 2022; Malki *et al*, 2022). BPMS contains a structured RNA binding domain, a RNA recognition domain (RRM) that dimerises and binds RNA (Figures 1.8 & 4.1A), and an intrinsically disordered proline rich C-terminus (Teplova *et al*, 2016). The extreme C-terminus of BPMS-A was found to be necessary for splicing function in a previous study in the laboratory (Yang *et al*, 2023). Here, the final twenty amino acids were deleted to create the mutant, $\Delta C20$. This mutant was unable to repress exon 3 of *Tpm1* in transfection and *in vitro* studies. This region of BPMS-A was also found to be necessary for self-assembly of BPMS-A and for interacting with various co-regulator proteins, in particular MBNL1 and RBFOX2.

In addition to the final twenty amino acids of BPMS-A being found to be important for its activity, further C-terminal deletions were produced previously to identify if other regions of the C-terminus are important for activity (Barnhart, PhD Thesis, University of Cambridge, 2022). Here, deletions within the IDR before the C-terminus were also found to abrogate the splicing activity of BPMS-A, specifically, deletions of residues between 139-156 (Fig.4.1A), which identified a tyrosine rich region surrounding tryptophan 148. This mutant significantly decreased both splicing repression (with *ACTN1* and *TPM1*) and activation (with *FLNB* and *MPRIIP*). This suggests that residues within this region are necessary for the splicing regulation activity of BPMS-A. These regions were investigated in the following experiments to identify which residues of the C-terminus were important for activity, and how they affected the biophysical behaviour to elucidate how the structure and function of BPMS-A are linked.

A



B

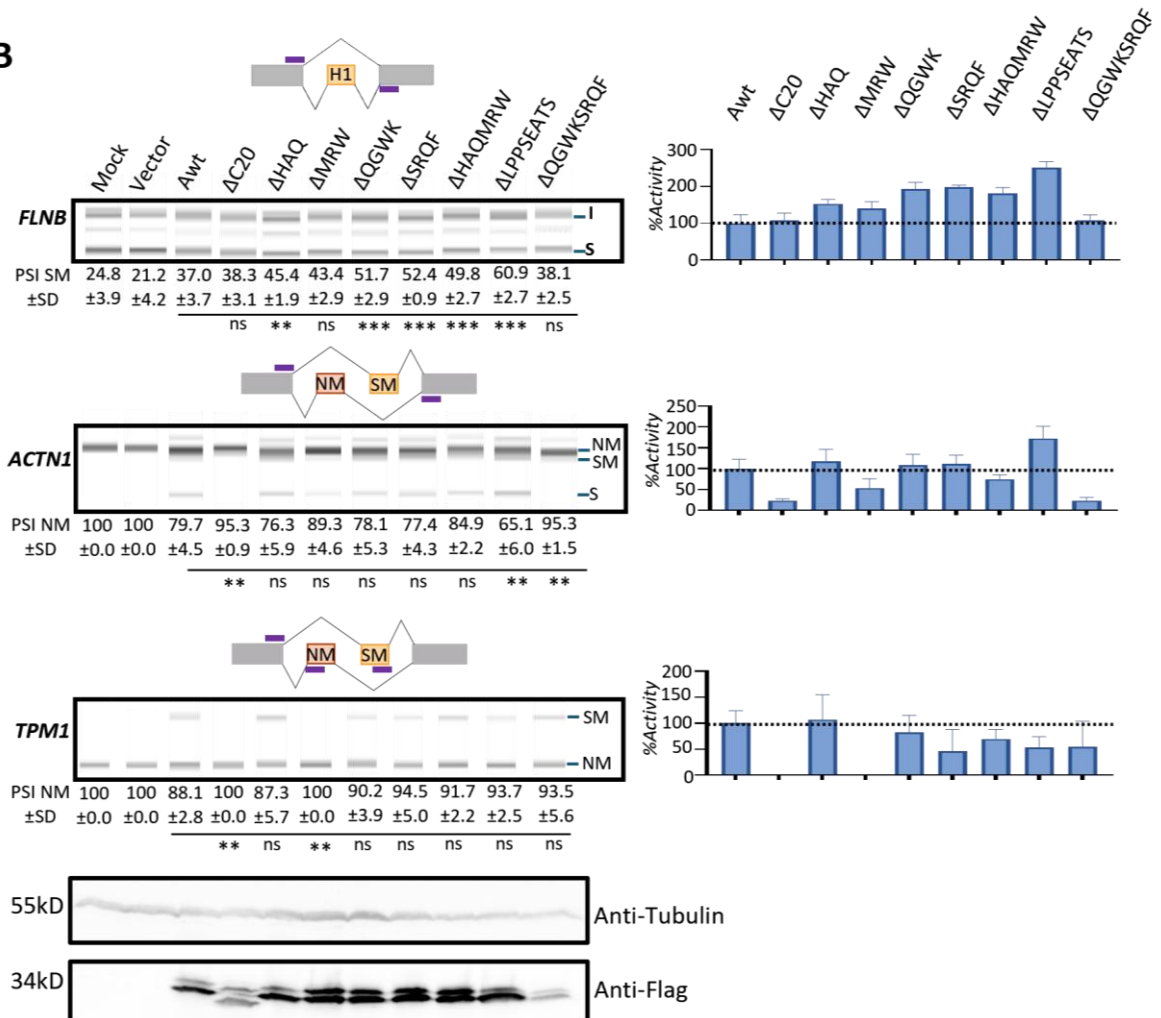


Figure legend on next page.

Figure 4.1 Deletions of the extreme C-terminus have minimal effects: **A:** Domain structure of RBPMS-A and the location of the deletions that were introduced are shown with red brackets (made in this study) or grey dashed brackets (made previously), with conservation across various vertebrate species of the latter section of the IDR. **B:** Qiaxcel RT-PCR data and western blot results of deletion mutants with endogenous targets in HEK293T. %Activity refers to the percentage activity of RBPMS-A wild-type. Purple lines on the schematics of the endogenous targets represent the location of the primers used in the RT-PCR. Statistical test was a one-way ANOVA and Dunnett's multiple comparison test, comparing activity of wild-type to the mutants, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe/Levene test used to check variance, QQ plot for distribution. RRM = RNA Recognition Motif, IDR = intrinsically disordered region, aa = amino acids, PSI = percentage spliced in, SM = smooth muscle, NM = non-muscle, I = inclusion, S = skipped.

RBPMS deletion mutations show important functional residues are not clustered together

To further understand the importance of the final C-terminal twenty amino acids of RBPMS-A, smaller deletions within this region were made (Fig.4.1A). The deletions were made based on conserved and non-conserved regions of the C-terminus. The activities of these mutants were tested in transfection assays in HEK293T and the splicing of endogenous RBPMS targets were analysed (Fig.4.1B).

These deletion mutants had mixed effects on the splicing activity of RBPMS-A. The deletion mutants Δ HAQ, Δ QGWK, Δ SRQF and Δ LPPSEATS had increased activity for splicing activation as they led to greater inclusion of the *FLNB* H1 exon (p-values = 0.008 for Δ HAQ and <0.001 for the other mutants), whereas Δ MRW, Δ C20, and Δ QGWKSRQF did not have significantly different activity to wild-type (p-values = 0.051, 0.992, and 0.997 respectively). For the repression targets, many of the deletions had no effect on activity. For *ACTN1*, Δ LPPSEATS was also found to increase activity compared to wild-type, leading to greater repression of the NM exon in favour of the SM exon (p-value = 0.004), whereas the Δ QGWKSRQF and Δ C20 mutants had significantly reduced activity (p-values = 0.002 for both mutants). For *TPM1*, none of the deletions increased activity but Δ C20 and the Δ MRW mutant lost repression activity (p-values = 0.007 for both mutants). The western blot results (Fig.4.1B) showed that the Δ QGWKSRQF and Δ C20 mutants had reduced expression and so the differences seen in activity cannot be conclusively said to be because of the deletions alone. The other deletion mutants apart from Δ HAQ had increased protein expression according to the western blot results and so the increase in activity seen for the activation and repression targets may be due to this rather than the deletions introduced. Given the lack of fresh insight by the preceding experiment, the next approach chosen was mutagenesis analysis of particular types of amino acid within the C-terminus.

4.2 C-TERMINAL TAIL SUBSTITUTIONS IDENTIFIED IMPORTANT RESIDUES FOR

SPLICING ACTIVITY

Higher-order assembly and protein-protein interactions are often mediated by cation – π and π – π interactions (Section 1.61) (Wang *et al*, 2018; Sabari *et al*, 2020). Basic side chains of lysine and arginine contribute the cation charge for these interactions, and aromatic side chains from phenylalanine, tyrosine and tryptophan form the π interactions. The regions found by deletion analysis to be important for RBPMS splicing activity contained aromatic residues and basic residues (Fig.4.2), therefore these residues were targeted on the assumption they may be forming cation – π and π – π interactions. The tyrosines in the 141-154 region were mutated to alanine to make a triple tyrosine mutant (Y141/151/154A, YYY/A), and was produced by Michael Barnhart (Barnhart, PhD Thesis, University of Cambridge, 2022). There are two tryptophans in the C-terminus and one phenylalanine, these were mutated to alanine (W179/190/F195A, WWF/A), and all three tryptophans across these two regions were also included in another mutant (W148/179/190A, WWW/A) and this latter mutant was produced by Michael Barnhart (Barnhart, PhD Thesis, University of Cambridge, 2022). There are three basic residues in the extreme C-terminus of RBPMS-A, and these were mutated to alanine as a neutral residue that does not take part in cation – π or π – π interactions (R178/K191/R193A, RKR/A). Histidine (H174) was also mutated alongside the basic or aromatic residues of the final twenty amino acids to create the additional mutants, HWWF/A and HRKR/A. This is because, depending on cellular environment (pH), histidine can be involved in cation – π interactions, as it can behave as a cation when protonated, or as an aromatic side chain when neutral (Liao *et al*, 2013). Histidine can also therefore be involved in π – π interactions when at neutral pH. The glutamine residues of the C-terminus were highly conserved, and such residues have been found to be important in other RBPs for shaping the properties of higher-order assemblies, mainly reducing liquid-like fluidity of phase separated droplets (Wang *et al*, 2018). Therefore, these residues were also mutated to alanine (Q176/188/194A, QQQ/A).

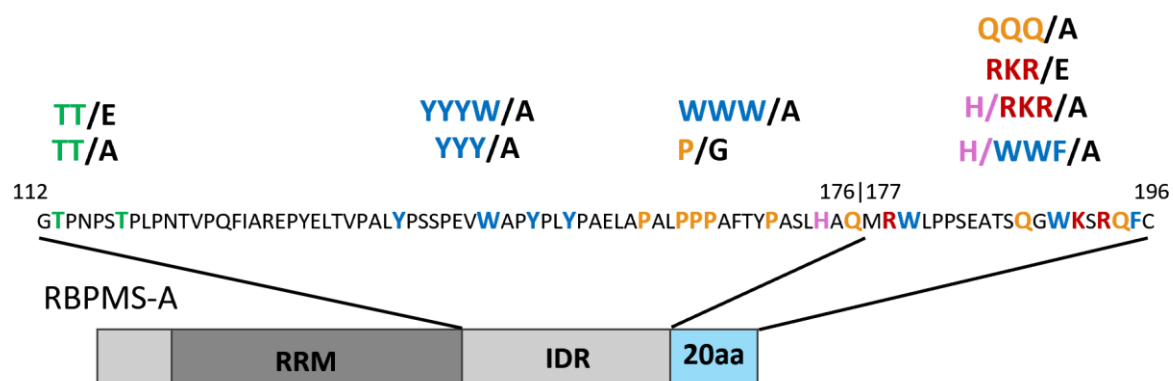


Figure 4.2 Substitution mutants of the C-terminus: Schematic of domain structure of RBPMS-A with residues that have been mutated shown in different colours (aromatics in blue, basic residues in red, phosphorylation sites in green, histidine residues in pink, and proline and glutamines in orange).

Aromatic residues of the C-terminus are required for the splicing activity of RBPMS

The basic, aromatic and glutamine mutants were tested for their effect on the splicing activity of RBPMS-A in transfection studies in HEK293T cells (Figure 4.3). The western blot shows that all the substitution mutants have similar expression to wild-type protein, allowing fair comparison of their activities. The minigene target *FLNB* was used to test activation behaviour, and endogenous *ACTN1* and *TPM1* were used to test repression. The activation of *FLNB* H1 exon was significantly reduced by all the aromatic mutants relative to the wild-type protein, suggesting that these residues play an important role in splicing activation by RBPMS-A (p-values were <0.001). The basic and glutamine mutants had no significant effect on activation; meaning these residues are unlikely to be important for activation of *FLNB* exon H1 by RBPMS-A (p-values were >0.999). The WWW/A mutant appears to be the most potent of the aromatic mutants, reducing activity significantly more than the other mutants. Similar effects were observed on the repression targets *ACTN1* and *TPM1*; the aromatic mutants lost activity, but basic mutations had no effect. This suggests that the RBPMS-A aromatic residues are important for both splicing activation and repression. The glutamine mutant had a slightly greater activity for repression than wild-type RBPMS-A. This indicates that the glutamine residues may moderate the activity of RBPMS-A in some way, perhaps through preventing more liquid-like phase separation behaviour, although the increase in activity relative to wild-type is only modest. The addition of the histidine to the WWF and RKR mutants (HWWF/A and HRKR/A) did not have an effect, suggesting that this residue is not important for activity.

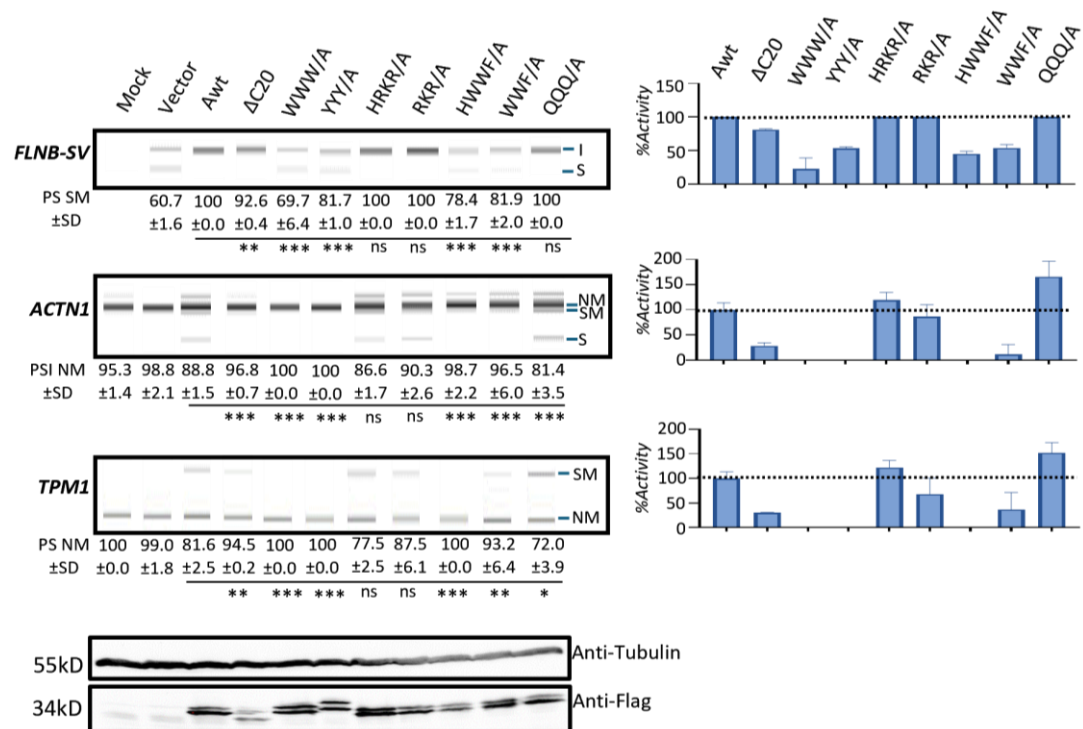


Figure 4.3 Aromatic residues of the C-terminus are important for activity: Qiagel RT-PCR data and western blot results from the substitution mutant transfection into HEK293T, with endogenous targets ACTN1 and TPM1, and minigene FLNB-SV used to measure differences in splicing activity. %Activity refers to the percentage activity of RBPMS-A wild-type. Statistical test was a one-way ANOVA followed by a Dunnett's multiple comparison test, comparing activity of wild-type to the mutants, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test for equal variance and QQ plot for normality. PSI = percentage spliced in, SV = refers to the minigene vector, SM = smooth muscle, NM = non-muscle, S = skipping, I = inclusion.

A second set of substitution mutants further tested the roles of the aromatic and basic residues, as well as a cluster of prolines that separate the two aromatic clusters (Fig.4.2). To further investigate the positively charged residues in the C-terminus, a reversal of charge mutant (RKR/E) was produced. This was to see if the addition of negative charge alters activity, as loss of positive charge through alanine substitution did not have an effect. Another aromatic mutant was made, with the tryptophan (W148) in the cluster of tyrosines at position 140-154 being mutated alongside the tyrosines (YYYW/A). This mutant was made as the triple tryptophan mutant (WWW/A), which included this residue, had reduced activity upon FLNB splicing relative to the triple tyrosine mutant (YYY/A). This YYYW/A mutant was produced by Michael Barnhart (Barnhart, PhD Thesis, University of Cambridge, 2022). Additionally, the stretch of amino acids between the two clusters found to be important for activity is proline rich. Proline residues are often important in providing rigidity and preventing flexibility due its constrained N – C α bond, and they are often found in IDRs (MacArthur & Thornton, 1991; Mao *et al*, 2013; Mittag & Pappu, 2022). Based on the assumption that the proline linker in this case may be

important for keeping these two sections of the C-terminus rigidly separated, the prolines here were mutated to glycine which allow much more flexibility (P160/164/165/166/170G, PPPPP/G).

These mutants were tested in transfection assays with the minigene *FLNB* for an activation target, and the minigene *Actn1* or endogenous *TPM1* as repression targets (Fig.4.4). The YYYW/A substitutions had a greater effect in activity than the tyrosine and tryptophan mutants (YYY/A and WWW/A) with *FLNB* (p-values = <0.001 for all aromatic mutants). With *Actn1* NM exon repression, the WWW/A and YYYW/A mutations again affected activity more than the YYY/A and WWF/A mutations, whilst for *TPM1*, all the aromatic mutants lost activity (p-values = <0.001 with *ACTN1* for all mutants and 0.004 with *TPM1*). When W148 was mutated singularly in previous studies in the laboratory it did not affect splicing activity (Barnhart, PhD Thesis, University of Cambridge, 2022). Therefore, it is likely that this residue is required in conjunction with other aromatic residues, eg the tyrosines surrounding it or with the other tryptophans. The PPPPP/G mutant had mixed effects on activity, having no effect on *FLNB* and *TPM1* (p-values = >0.99), but with *Actn1*, repression activity was increased (p-value = <0.001). This suggests that the importance of this proline linker is not universal for all splicing targets of RBPMS. A wider range of targets would be needed to investigate how important this linker is to RBPMS-A splicing activity. In these assays, the RKR/A mutant had slightly reduced activity for *Actn1* NM repression (p-value = 0.023), but with *TPM1* and *FLNB* the activity was the same as wild-type (p-values = >0.99). The RKR/E mutation had much more pronounced defects with activation and repression with each target, particularly with *TPM1*, whereas RKR/A had 100% activity whilst RKR/E had no activity (p-values = <0.001 with *FLNB* and *ACTN1*, 0.004 with *TPM1*). This suggests that the positive charges are not themselves responsible for the activity of RBPMS-A, but introducing negative charge does prevent RBPMS-A from being fully active.

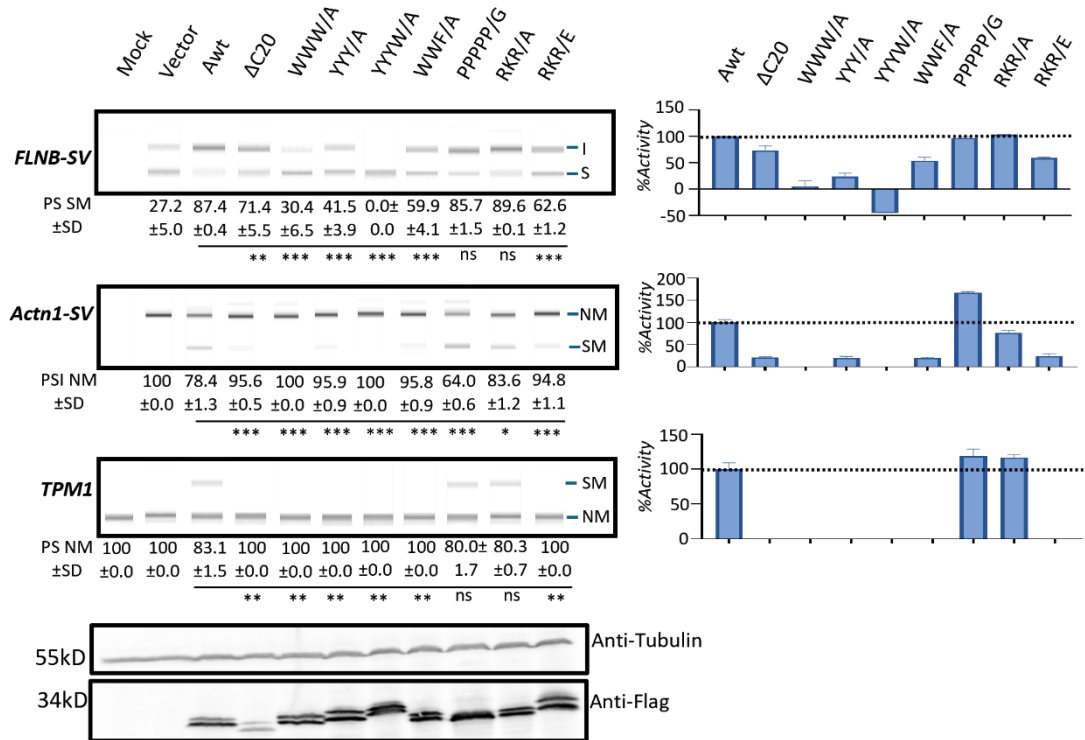


Figure 4.4 Reversal of charge depletes activity of RBPMS-A: Qiaxcel and western blot results from co-transfection of RBPMS substitution mutants and minigenes *FLNB* and *Actn1*, endogenous *TPM1* was also used to assess splicing ability. %Activity refers to the percentage activity of RBPMS-A wild-type. Statistical test was a one-way ANOVA followed by a Dunnett's multiple comparison test, comparing activity of wild-type to the mutants, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test was used to test for equal variance and QQ plot for normal distribution. For *TPM1*, a Kruskal Wallance test was performed (no normal distribution). PSI = percentage spliced in, SV refers to the minigene vector used, SM = smooth muscle, NM = non-muscle, S = skipping, I = inclusion.

Previous studies on the post-translational modifications of RBPMS revealed that two threonine residues just after the RRM (T113 and T118) can be phosphorylated by ERK2, and this phosphorylation negatively regulates the activity of RBPMS (Barnhart *et al*, 2022). This work involved using two mutants of these threonine, either mutating the threonines to glutamate to mimic the phosphorylation, or to alanine as a phospho-null mutant (Fig.4.2). These mutants were tested alongside the aromatic and basic mutants previously discussed. These mutants were tested in HEK293T cell transfections with the aromatic and basic mutants with activation targets, minigene *PTPRF* and endogenous *FLNB*, and endogenous repression targets, *TPM1* and *ACTN1* (Figure 4.5). The *PTPRF* minigene was used to assess how these mutants may behave in the *in vitro* splicing activation assay established in the previous chapter (Chapter 3). Overall, the aromatic and basic mutants behaved similarly with each target as they have previously (Figures 4.3 and 4.4), and they showed similar effects with *PTPRF* as they did with *FLNB* (p -values = < 0.001). The phospho-null mutant (TT/A) does not abrogate activation activity and had only a small decrease in activity with *TPM1*, and increased activity for *ACTN1*, confirming that

these threonines are not required for the splicing activity of RBPMS (p-value with *PTPRF* = 0.120, with *FLNB* = 0.996, *ACTN1* = 0.004, *TPM1* = 0.045). The phosphomimetic mutant (TT/E) had reduced activity for activation and repression, except with the endogenous *FLNB* target (p-value with *FLNB* = 0.139, and with *PTPRF*, *ACTN1* and *TPM1* = <0.001). This confirms that phosphorylation of these threonines likely negatively affects the activity of RBPMS-A. This may also explain why the TT/A mutant had higher activity than wild-type RBPMS-A with *ACTN1*, if there is endogenous phosphorylation of the wild-type protein this would likely limit its activity more than the TT/A mutant, although it does not appear to have a great effect as it is not observed with the other targets.

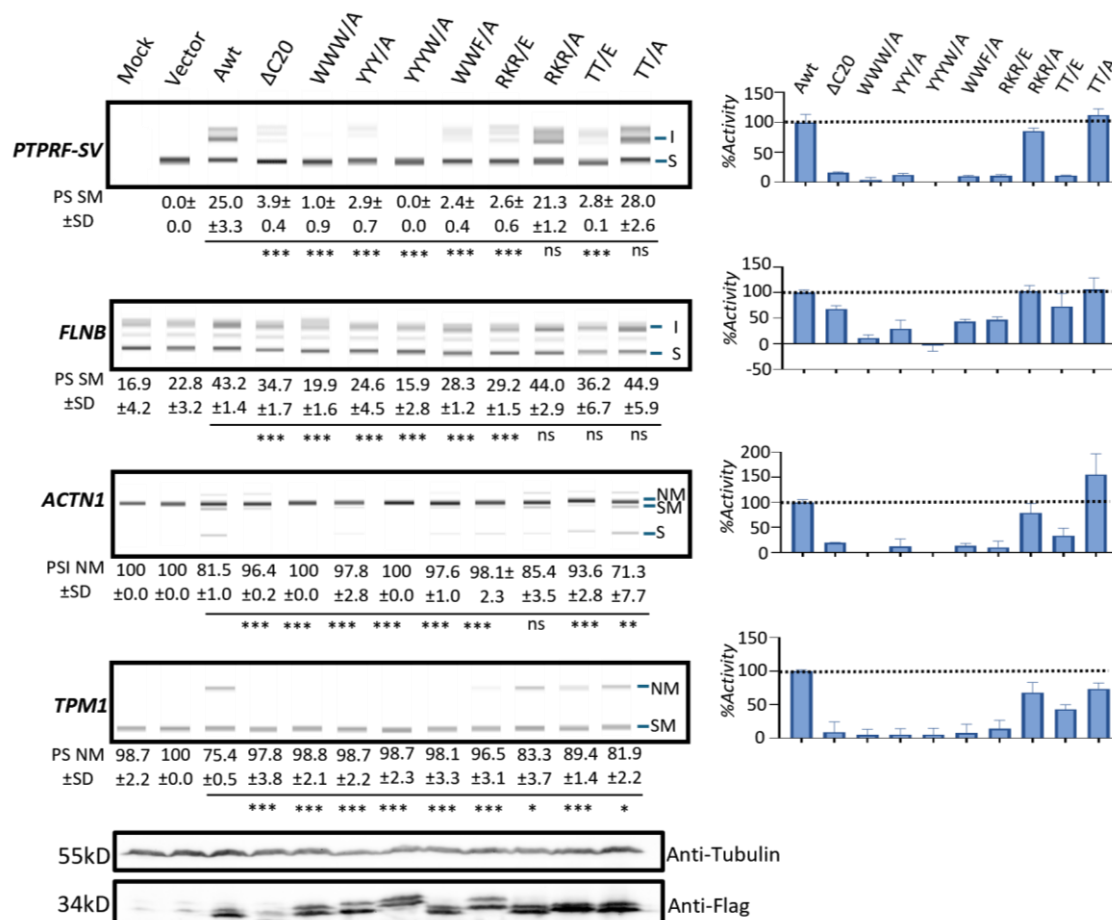


Figure 4.5 Phosphomimetic mutant loses splicing activity: Qiaxcel and western blot data of RBPMS-A substitution mutant transfection. Minigene PTPRF and endogenous FLNB, ACTN1, and TPM1 were used to assess splicing activity. %Activity refers to the percentage activity of RBPMS-A wild-type. Statistical test was a one-way ANOVA followed by a Dunnett's multiple comparison test, comparing activity of wild-type to the mutants, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test was performed for equal variance and QQ plot for normality. PSI = percentage spliced in, SM = smooth muscle, NM = non-muscle, I = inclusion, S = skipped.

4.3 *IN VITRO* SPLICING ASSAYS SUCCESSFULLY REPLICATE TRANSFECTION STUDIES

From the substitution mutants tested in the above section, it appears that the aromatic residues are required for both splicing activation and repression activity of RBPMS-A. There are also two sets of mutants where mutating the residues to alanine had no effect on activity, but introduction of a negative charge caused a significant reduction in activity (RKR/A & E, and TT/A & E). In the case of the TT/E mutant, this negative charge mimics the phosphorylation of these residues and has previously been shown to reduce activity (Barnhart *et al*, 2022). The basic residues themselves appear to not be necessary for activity, but switching to negative charge had a deleterious effect. To delve into the structure-function relationship of the splicing activity of RBPMS-A, the role that these residues play in activity and assembly needed to be further dissected. To investigate this, *in vitro* systems, such as the splicing assays introduced in chapter 3, were employed. Use of recombinant proteins and *in vitro* transcribed RNA allows precise control over the concentrations of different proteins and RNAs. This ensures that any observed changes in activity are directly attributable to the introduced mutations, rather than differences in expression levels or protein turnover. Moreover, *in vitro* systems allow for the direct comparison of activity and biophysical properties of the mutants, therefore allowing for the elucidation of the structure-function relationship of RBPMS-A.

Production of recombinant RBPMS

The mutant proteins were made recombinantly as described in sections 2.3 and 3.2. The proteins produced were either untagged or had an N-terminal STREP-6xHis tag. The TT/A and TT/E mutants were purified with a similar protocol as the wild-type protein as they formed similar size complexes during the HisTrap purification and Mono Q purification. The aromatic and basic mutants had a slightly adjusted protocol where the fractionation occurred via smaller steps (2 ml versus 7 ml for HisTrap, and 0.5 ml versus 2 ml for Mono Q) to allow for greater separation of each fraction, since they were observed to form smaller complexes than wild-type protein during their purification. This means that if the larger fraction volumes were used, the sample would be fully eluted in the first fraction, and there would be no way to separate “clean” fractions from fractions containing contaminating proteins for these mutants. The tagged proteins were used for pull-down experiments and for when an additional detection tag for antibodies was required (section 4.6). Figures 4.6 & 4.7 shows the fractions selected after the final stage of purification.

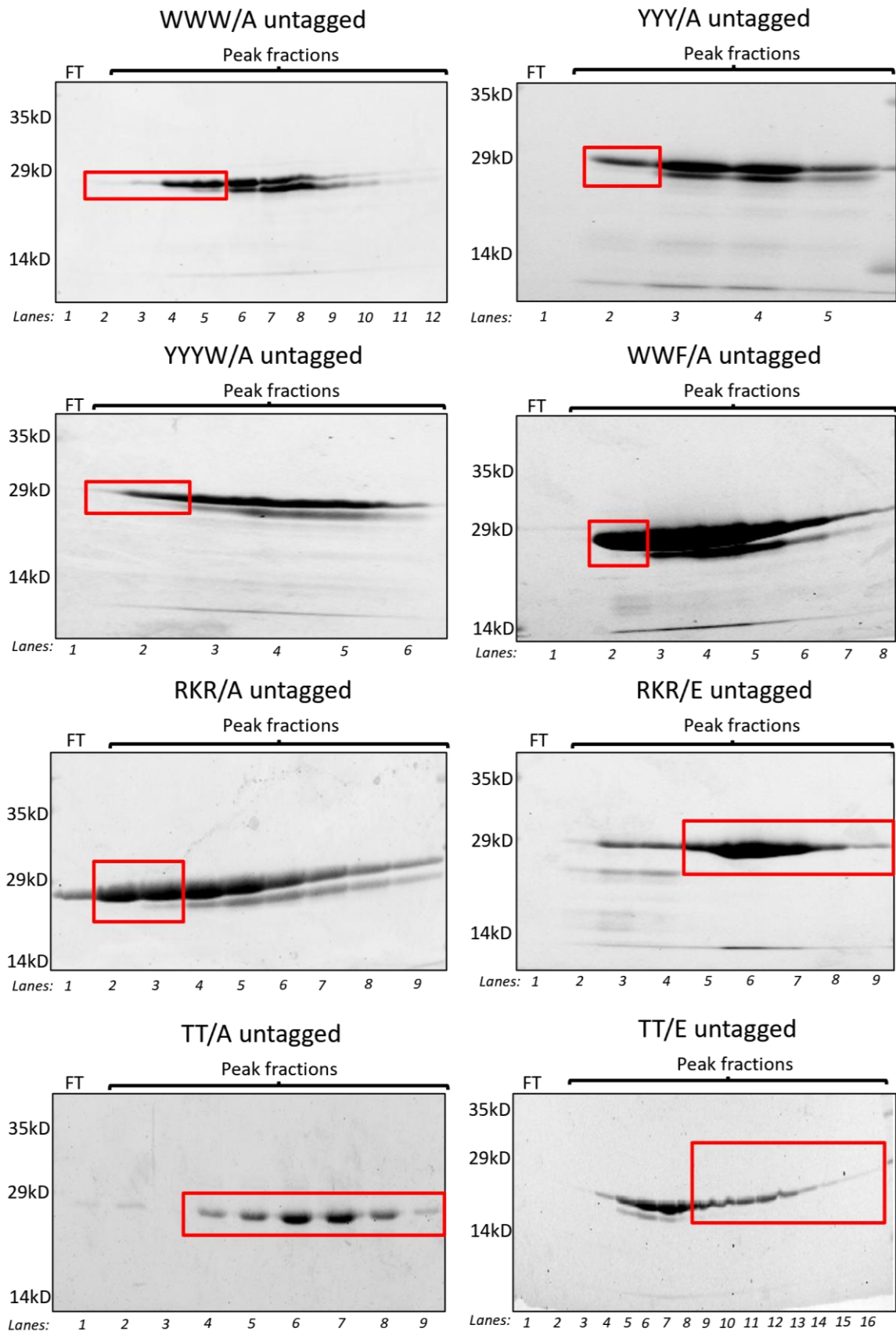


Figure 4.6 Recombinant untagged mutant RBPMS-A protein production: Coomassie Blue stained SDS-PAGE gels showing which fractions were chosen of recombinant untagged RBPMS-A mutants post HisTrap, post TEV cleavage and post MonoQ. FT = flow-through.

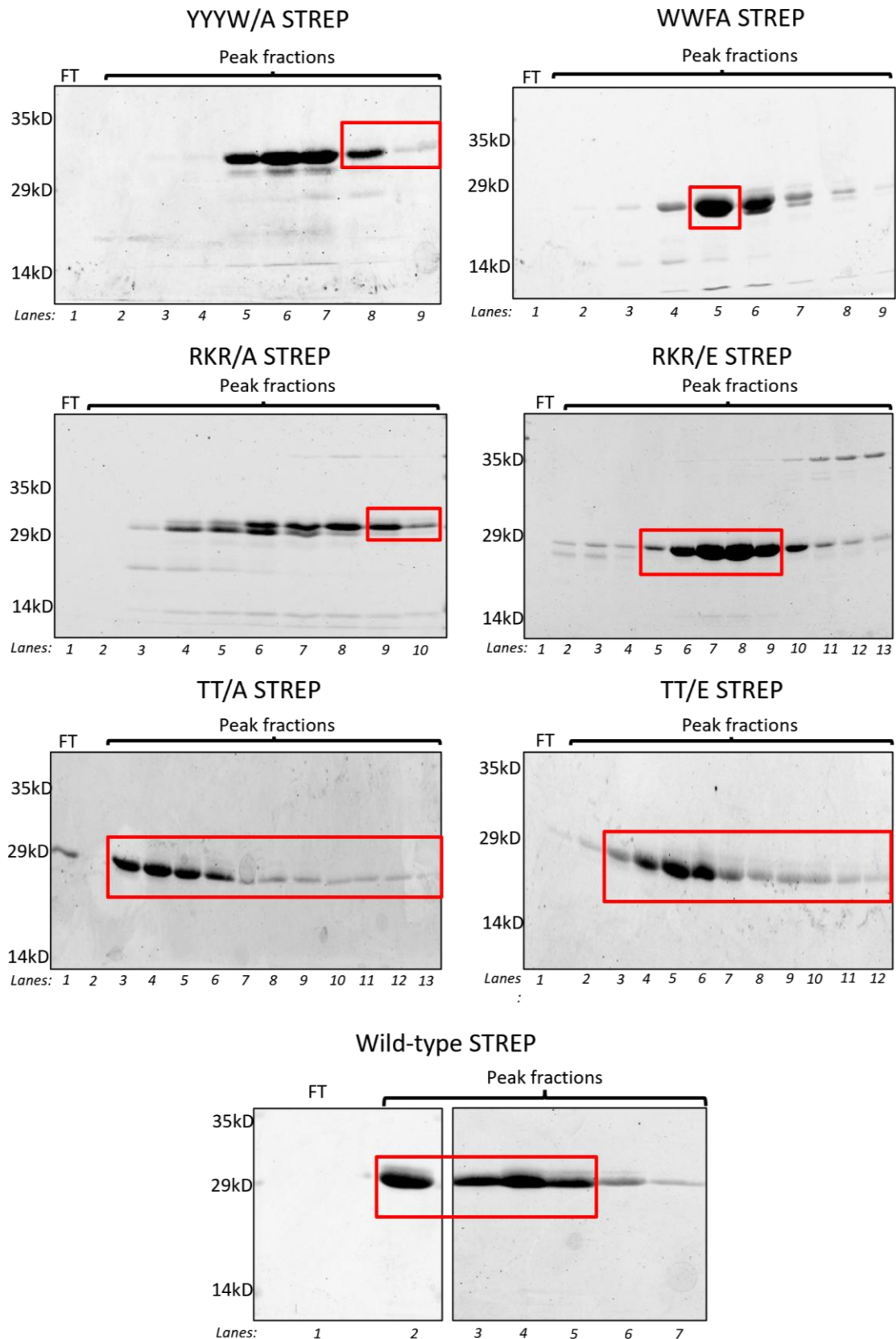


Figure 4.7 Recombinant STREP tagged mutant RBPMS-A protein production: Coomassie Blue stained SDS-PAGE gels showing which fractions were chosen of recombinant 6xHis STREP II tagged RBPMS-A mutants post HisTrap and post MonoQ. FT = flow-through.

In vitro splicing assays confirm that the aromatic residues are required for the splicing activity of RBPMS-A

The splicing activities of the recombinant proteins were tested with the *in vitro* constructs *PTPRF* and *Tpm1* (Fig.4.8). The wild-type constructs and the mutated CAC constructs were used with each protein. Each protein was added to a final concentration of 1 μ M, with Figure 4.8A demonstrating concentration was the same between the recombinant proteins. The aromatic mutants lost activity for both splicing activation and for repression (Fig.4.8B). The amount of inclusion for *PTPRF* exon 23 and *Tpm1* exon 3 with the aromatic substitution mutants stayed approximately at basal levels (no statistical significance, all p-values when compared to wild-type RBPMS-A = <0.001 for all mutants). These results agree well with the transfection data, indicating that these assays are an appropriate test of RBPMS-A splicing activity and confirm that the aromatic residues are essential for both splicing activation and repression by RBPMS-A.

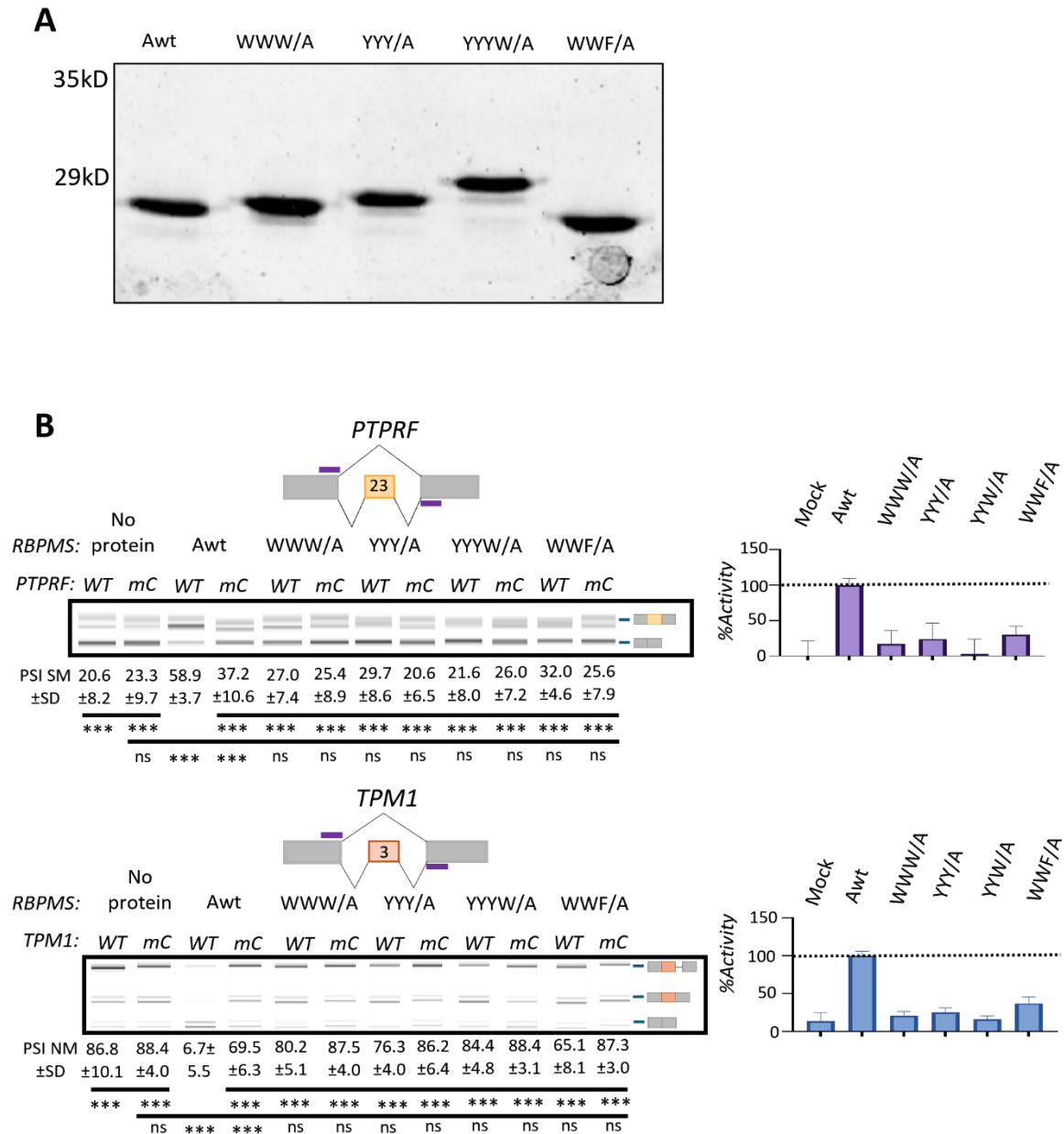


Figure 4.8 Aromatic RBPMs-A mutants lose splicing activity in vitro: **A:** Image of Coomassie Blue stained SDS-PAGE gel showing recombinant proteins used in the in vitro splicing assay. **B:** Qiaxcel data from in vitro splicing assays with activation target PTPRF WT/mutated CAC mutant or repression target Tpm1 WT/mutated CAC mutant with 1 μ M recombinant RBPMs-A wild-type and aromatic mutant proteins. %Activity refers to the percentage activity of RBPMs-A wild-type with the wild-type transcript, and only compares mutant activity with wild-type transcript. Statistical test was a two-way ANOVA followed by a Tukey's multiple comparison test to compare activity of RBPMs-A wild-type and WT transcript with every other sample (top line), or mock and WT transcript with every other sample (bottom line). * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test to show equal variance, QQ plot for normal distribution. PSI = Percentage Spliced In, SM = Smooth Muscle, NM = Non-Muscle, WT = wild-type transcript, mC = mutated CAC/CCC transcript, Awt = RBPMs-A wild-type.

The basic mutants and phospho-mutants were next tested (Fig.4.9). As in the transfection studies, the RKR/A and TT/A mutants had no effect on activity for activation and repression (p-values of RKR/A and TT/A vs wild-type = 0.183 and = 0.426 respectively for *PTPRF*, 0.385 and 0.225 for *TPM1*). The RKR/E and TT/E mutants both had reduced activity compared to wild-type with *PTPRF* and *Tpm1* (p-values of RKR/E and TT/E with *PTPRF* and *TPM1* = <0.001). With *PTPRF*, the RKR/E mutant had a Δ PSI of ~10% relative to basal levels, whilst TT/E had a Δ PSI of ~5%. For the repression assay, RKR/E lost activity relative to wild-type and inclusion of exon 3 remained at basal levels. The TT/E mutant had reduced activity compared to WT RBPMS-A, but exon 3 skipping was significantly higher than the basal level (Δ PSI = ~27%). This suggests that the mutation of the threonines to phosphate-mimicking glutamate residues dampens activity of RBPMS-A but does not cause a complete loss like the mutation of the aromatic residues does. These studies confirmed that out of all the residues tested, the aromatic residues are the most important for the splicing activation and repression activity of RBPMS-A.

To explore whether there were differences at lower concentrations between wild-type and the null mutants (RKR/A and TT/A) and TT/E which had some residual activity, titration assays were performed (Figure 4.10). RKR/A and WT RBPMS-A proteins showed similar activity with *PTPRF* and *TPM1*, except for a slight increase in activity compared to wild-type at 250 nM with *PTPRF* (p-value = 0.004). TT/A and wild-type proteins had no difference in activity at any concentration tested for activation (*PTPRF*) and repression (*Tpm1*). This confirms that the substitution of these residues does not greatly affect the activity of RBPMS-A within the concentration range tested. The TT/E mutant protein had reduced activity at each concentration for the activation target *PTPRF*. With *Tpm1*, at lower concentration of TT/E and wild-type proteins there was no increase in exon 3 skipping above basal levels, but at higher concentrations (≥ 500 nM) the wild-type repressed exon 3 above basal levels and there was a significant difference between TT/E and wild-type proteins (p-values at 500 nM = 0.01, 1000 nM = <0.001, and 2000 nM = 0.008). The TT/E mutant had partial activity compared to basal levels for all concentrations of protein with *PTPRF* (p-values at 250nM = 0.006, 500 nM = <0.001, 1000 nM = <0.001, and 2000 nM = 0.015), and only at the highest concentration (1000 nM) for repression of *Tpm1*. The results for the TT/E and TT/A proteins only contained two replicates, thus more repeats are required for robust statistical analysis to confirm these observations. The activity of wild-type RBPMS-A with *Tpm1* at the highest concentration was reduced in the experiments with TT/A and TT/E proteins compared to the RKR/A experiment, which is likely due to the experiments being performed at different times with different aliquots of nuclear extract. These experiments, however, were still able to successfully confirm the trend that wild-type and TT/A RBPMS-A have similar activity, whilst the TT/E mutant had reduced activity.

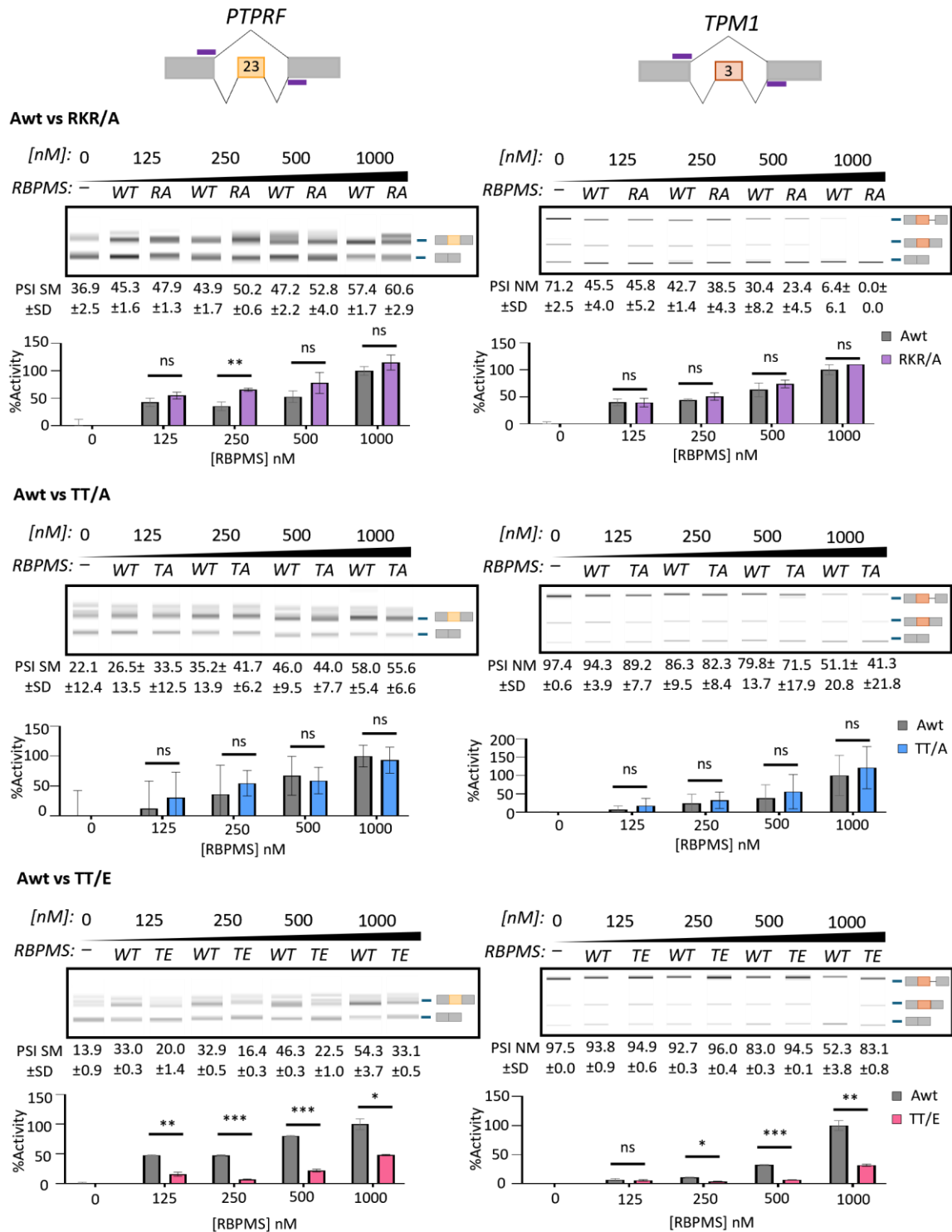


Figure 4.10 Titration of the basic and phosphomutants of RBPMS-A confirms that phosphomimetic mutant loses splicing activity: Qiagxcel data from *in vitro* splicing assays with PTPRF and Tpm1 and titration of RKR/A, TT/A and TT/E mutants compared and wild-type. TT/A and TT/E titration experiments only have two biological repeats, RKR/A titration had three. %Activity is relative to RBPMS-A wild-type at the highest concentration. Statistical tests were student tests comparing activity at each concentration: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. PSI = Percentage Spliced In, SM = Smooth Muscle, NM = Non-muscle, WT = RBPMS-A Wild-type, RA = RKR/A mutant. TA = TT/A mutant, TE = TT/E mutant.

4.4 THE C-TERMINAL MUTATIONS DIMINISH RBPMS CROSSLINKING TO RNA

RBPMs-A crosslinks to RNA

To assess how important the residues of the C-terminus are for interacting with RNA, UV crosslinking experiments with radiolabelled RNA were performed. The RNA substrate was *Tpm1* exon 3 (TM3) with flanking intronic regions containing RBPMS binding sites (or mutated binding sites) and regulatory elements (Fig.4.11A). To observe if the UV crosslinking experiment can detect differences in crosslinking between the wild-type and the mutant proteins, a pilot experiment was performed with titrations of wild-type RBPMS-A and the YYYW/A mutant. The RNA binding reaction occurred in either 20% HeLa nuclear extract or buffer. The use of nuclear extract recreates the conditions used for *in vitro* splicing and means that RBPMS will be competing for binding to the transcript with other antagonistic or co-regulatory nuclear proteins. In the buffer only lanes of the radiograph (Fig.4.11B, lanes 12 – 17), RBPMS-A wild-type and YYYW/A can both crosslink the wild-type TM3 transcript. In nuclear extract (lanes 2 – 7), only RBPMS-A wild-type is observed to crosslink to the transcript, with YYYW/A showing no crosslinking at any concentration. There is also no interaction with the CCC mutant by the wild-type or YYYW/A mutant protein (lanes 9 & 10), showing that the crosslinking to WT TM3 is specific in a competitive environment. The lack of crosslinking by YYYW/A in nuclear extract may clarify why this mutant loses splicing activity, as lack of UV crosslinking suggests this mutant may not be able to bind RNA efficiently in nuclear extract. The $\Delta C20$ mutant was also observed to not crosslink RNA in splicing conditions (Yang *et al*, 2023). This was hypothesised to be due to a loss of interaction with other proteins that may aid RBPMS recruitment to the RNA.

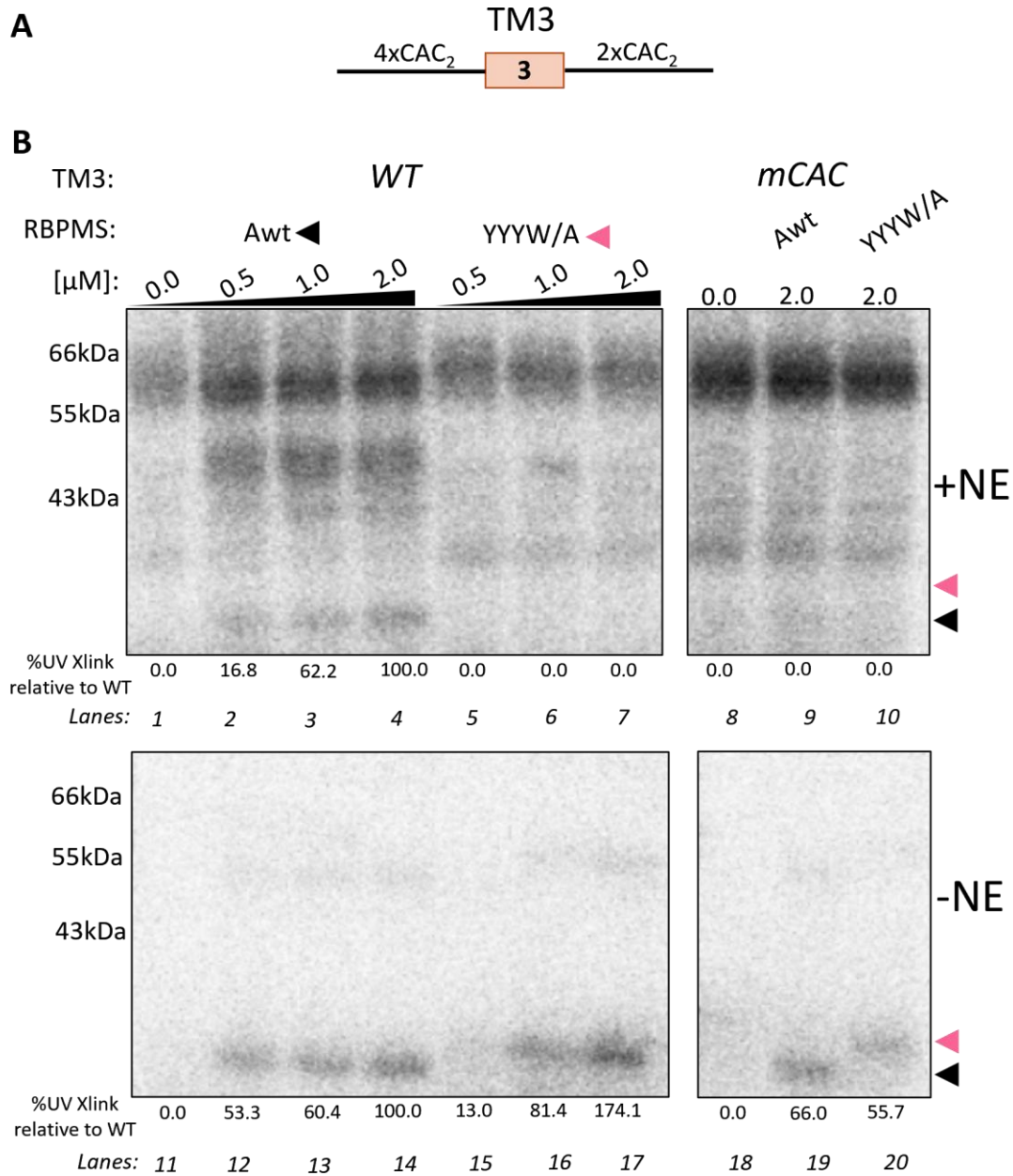


Figure 4.11 Wild-type RBPMS-A crosslinks RNA in nuclear extract: **A:** Schematic of the transcript used in the UV crosslinking experiment. This transcript is 586 nt and has exon 3 of the Tpm1 construct and its flanking intronic regions, containing all the regulatory splicing sequences and likely RBPMS-A binding sites (CAC₂). **B:** Radiograph of UV crosslinking experiment with 2 nM of radio-labelled TM3 wild-type (WT) or mutated CAC TM3 (mCAC), and a titration recombinant RBPMS-A wild-type or YYYW/A mutant. RNA binding reaction was with and without the addition of 20% HeLa nuclear extract. Expected mobilities for wild-type RBPMS-A and the YYYW/A mutant proteins are shown with black and pink arrows respectively. %UV Xlink refers to the percentage of UV crosslinking relative to wild-type at the highest concentration with wild-type TM3 in $\pm$ nuclear extract conditions. One repeat was performed. Xlink = Crosslink, NE = nuclear extract, Awt = RBPMS-A wild-type.

RBPM5-A mutants have reduced crosslinking

All the mutants were tested for UV crosslinking to the wild-type TM3 transcript at 1 μ M final concentration, in 20% nuclear extract or buffer (Fig.4.12). The buffer only lanes (11 – 20, lower panel) show that all the mutants except RKR/E retained some RNA crosslinking ability, but they all had significantly reduced levels relative to the wild-type protein (p-values = <0.001). The RKR/E mutant has completely lost any RNA crosslinking ability, whereas the RKR/A mutant still retains ~30% of wild-type RNA crosslinking. In nuclear extract, only wild-type RBPM5-A is seen to crosslink the substrate. There are very faint bands for RKR/A and TT/A that were detected by the ImageQuant software, but they were not easily observable by eye. For the other mutants, there was no signal detected above background (p-values relative to wild-type RBPM5-A = <0.001). These two mutants were the only ones that were fully active in the *in vitro* splicing assays and so it was expected that these should still be able to retain crosslinking ability in a competitive environment. It was therefore surprising that their crosslinking ability was significantly reduced relative to wild-type, indicating that these mutants have reduced RNA binding ability compared to wild-type.

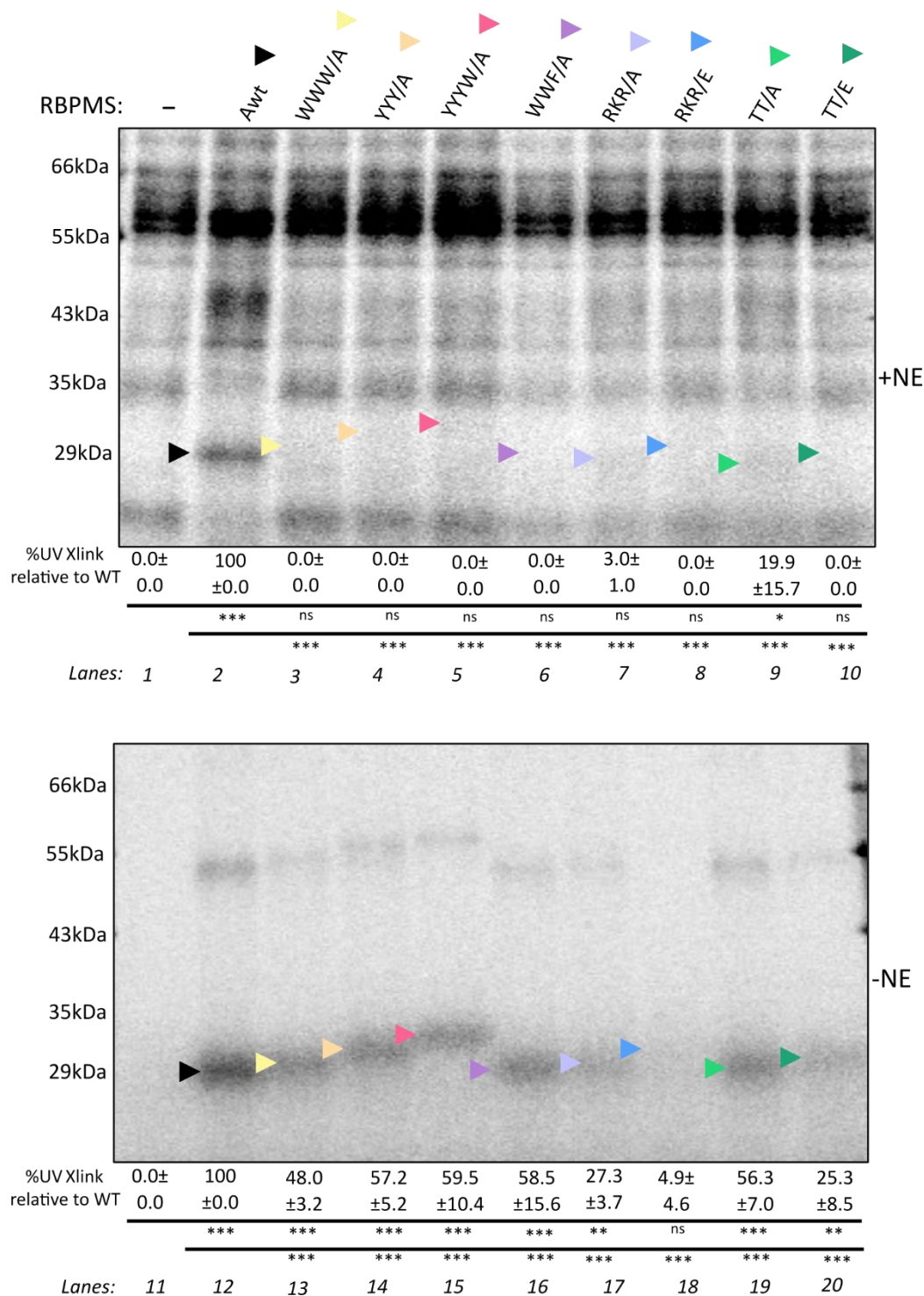


Figure 4.12 All RBPMS-A mutants have reduced UV crosslinking relative to wild-type: Radiograph of UV crosslinking experiment with 2 nM of radio-labelled TM3 wild-type (WT) and the addition of 1 μ M of recombinant RBPMS-A wild-type and mutant proteins. RNA binding reaction was performed $\pm$ 20% HeLa nuclear extract. Expected mobility of each protein is shown by a coloured arrow. %UV Xlink refers to the percentage of UV crosslinking relative to wild-type in $\pm$ nuclear extract conditions. Three repeats were performed. Statistical test was a one-way ANOVA followed by a Dunnett's multiple comparison test, comparing crosslinked RNA in the no protein lane (lanes 1 and 11) relative to addition of RBPMS (top line), or comparing wild-type (lane 2 and 12) and its mutants crosslinking abilities (bottom line), * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test for equal variance and QQ plot for normality. Xlink = Crosslink, NE = nuclear extract, Awt = RBPMS-A wild-type.

Titration of TT/A and RKR/A show they retain low levels of UV crosslinking ability in nuclear extract

To further investigate the UV crosslinking capabilities of these two mutants (RKR/A and TT/A), titration experiments were performed alongside wild-type protein (Fig.4.13). Between 0.5 – 1 μ M of protein, TT/A and RKR/A mutants did not show any significant crosslinking in nuclear extract (lanes 3 – 7). At 2 μ M, there was low level crosslinking for both mutants above background, but significantly less than wild-type (lanes 8 – 10, p-values = <0.001 for both RKR/A and TT/A mutants). In buffer, the amount of RNA crosslinked by TT/A and wild-type was not significantly different, except at 0.5 μ M (p-value = 0.01) (lanes 16 & 18). RKR/A UV crosslinking to RNA in buffer was reduced relative to wild-type at each concentration, being 50% of that of wild-type at 2 μ M (lanes 22 & 23).

Altogether, these results show that the basic residues and aromatic residues are aiding in UV crosslinking ability and therefore potentially in RNA binding ability. The threonine residues do not impact RNA crosslinking in buffer, as the TT/A and wild-type proteins had similar efficiency (Fig.4.23), but mutation to alanine does affect crosslinking in nuclear extract. The conditions, amount of protein and amount of RNA, were all similar to the *in vitro* splicing assays which suggests that this is the context in which RKR/A and TT/A are able to regulate splicing. The fact that RKR/A and TT/A are still active and have such low UV crosslinking may suggest that in fact splicing regulation by RBPMS-A only requires minimal RNA binding.

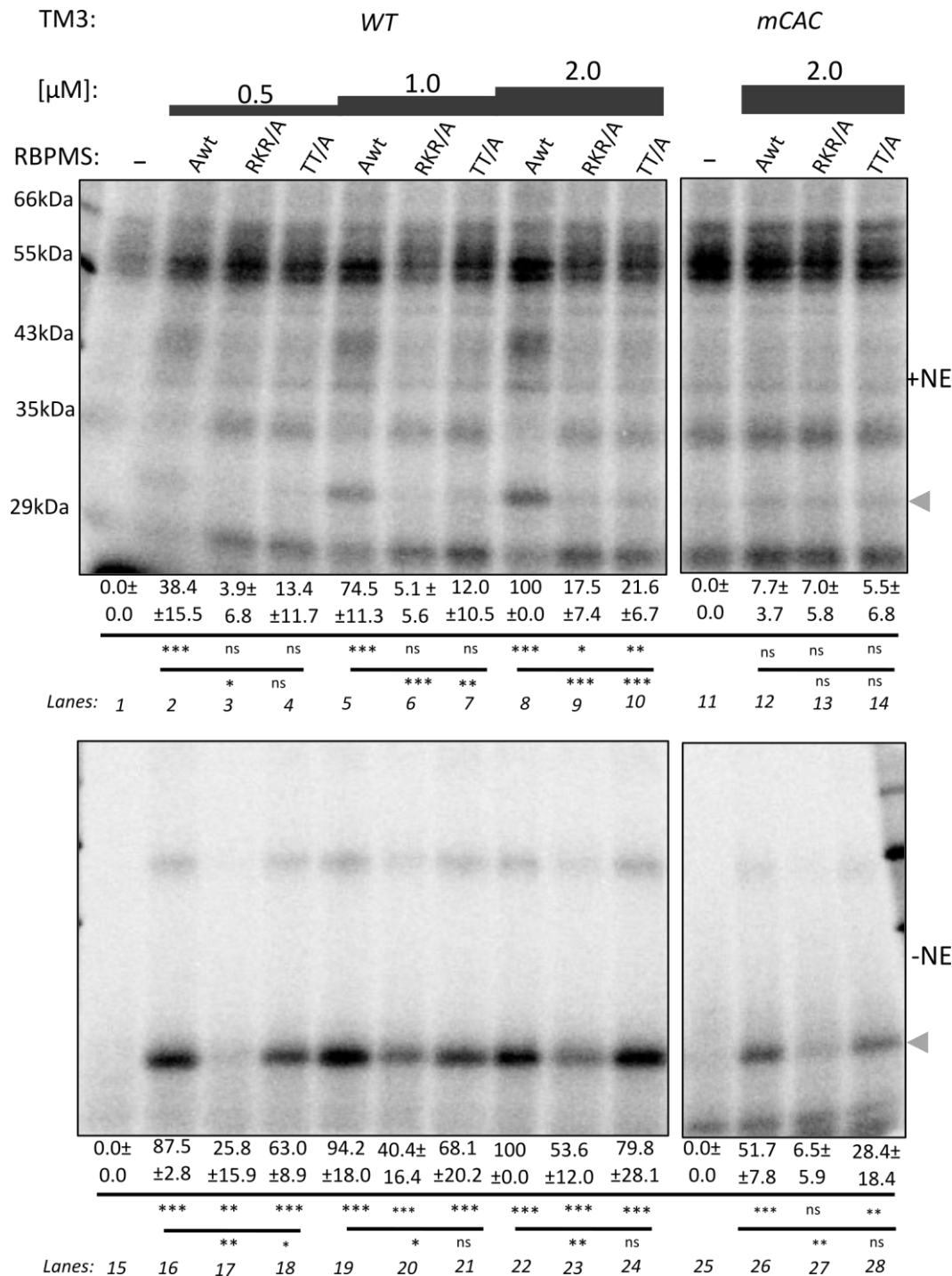


Figure 4.13 RKR/A and TT/A mutants show minimal UV crosslinking in nuclear extract relative to wild-type RBPMS-A: Radiograph of UV crosslinking experiment with 2 nM of radio-labelled TM3 wild-type (WT) or mutated CAC TM3 (mCAC), and a titration recombinant RBPMS-A wild-type, RKR/A or TT/A mutants. The RNA binding reaction was performed $\pm$ 20% HeLa nuclear extract. Expected mobility of each protein is shown by a grey arrow. %UV Xlink refers to the percentage of UV crosslinking relative to wild-type at the highest concentration with wild-type TM3 in $\pm$ nuclear extract conditions. Three repeats were performed. Statistical tests were a one-way ANOVA followed by a Dunnett's multiple comparison test comparing UV crosslinking of each concentration of protein to the no protein lane (lanes 1, 11, 15, & 26) or student t tests comparing UV crosslinking of the mutants to wild-type protein at each concentration. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test for equal variance and QQ plot for normality. Xlink = Crosslink, NE = nuclear extract, Awt = RBPMS-A wild-type.

4.5 HIGHER ORDER ASSEMBLY AND ACTIVITY OF RBPMS-A CAN BE UNCOUPLED

The C-terminus of RBPMS-A was found to be necessary for activity and for higher-order assembly (Section 1.53) (Yang *et al*, 2023). Here, analytical ultracentrifugation (AUC) studies showed that RBPMS-A is able to form multiple higher-order species, with assemblies up to 27 subunits being observed. These species were dependent on protein concentration, with lower concentrations reducing the number of higher-order species seen, and on salt concentration. At lower salt (60 mM KCl), there was a larger range of sizes formed, but at high salt (150 mM KCl) there was a greater proportion of species forming higher order assemblies, suggesting high salt favours higher-order assembly. This indicates that electrostatic interactions may be influencing the higher-order assembly of RBPMS-A. The deletion mutant $\Delta C20$ did not undergo higher-order assembly at any concentration, with a trimer being the largest species being observed. The final twenty amino acids of RBPMS-A are therefore important for both higher-order assembly and its activity, which suggests there is a relationship between the two. Furthermore, the phospho-null mutant and wild-type RBPMS-A were shown to have greater levels of higher-order assembly than the phospho-mimetic mutant using AUC (Barnhart *et al*, 2022). This implies that phosphorylation of RBPMS-A at these residues leads to a decrease in both activity and higher-order assembly. This again suggests that higher-order assembly of RBPMS-A and its ability to regulate splicing are correlated.

RBPMS-A can undergo LLPS

Many RBPs have been observed to undergo extreme higher-order assembly *in vitro* and in cells, in some cases forming condensates, either on their own or with other proteins and RNA. These proteins are often undergoing liquid-liquid phase separation (LLPS). The intrinsically disordered regions and low complexity domains of RBPs have often been found to drive this phase separation due to their propensity for multivalent interactions with other proteins and RNA (Lin *et al*, 2015; Ying *et al*, 2017; Ule & Blencowe, 2019). This phase separation behaviour of these proteins can aid or regulate their activity by putting them in close proximity with their target RNAs and co-regulatory or antagonistic proteins.

To investigate if recombinant RBPMS-A can phase separate, recombinant RBPMS-A was buffer exchanged to a low salt buffer (100 mM KCl, 20 mM HEPES pH 7.9) (Section 2.42). Turbidity measurements at $\lambda = 340$ nm are often used to measure protein aggregation, and thus it was used to assess whether recombinant protein was aggregating and therefore undergoing LLPS (Shmueli *et al*, 2017). There was no increase in signal at 340 nm for recombinant RBPMS-A wild-type on its own, suggesting it was not undergoing phase separation. There are several studies that suggest that some

proteins need the addition of crowding reagents to phase separate, to help recreate the crowded cellular environment (Uversky, 2009; Soranno *et al*, 2014; Bratek-skicki *et al*, 2020). Hence, several crowding reagents were tested with RBPMS-A to see how they affected phase separation. Protein was added at 20 μ M along with a small amount of C-terminally fluorescently labelled recombinant RBPMS-A (0.5 μ M final concentration). Crowding reagents polyethylene glycol (PEG) or polyvinyl alcohol (PVA), which is used in the *in vitro* splicing assays as a crowding reagent, were then added. RBPMS-A has a single cysteine at its C-terminus, and this was found to not affect activity (Figure 4.14A). Therefore, maleimide Alexa-647 dye was able to be conjugated to the C-terminal cysteine to fluorescently label this protein. Figure 4.14B shows that recombinant wild-type RBPMS-A can form droplets in the presence of crowding reagents.

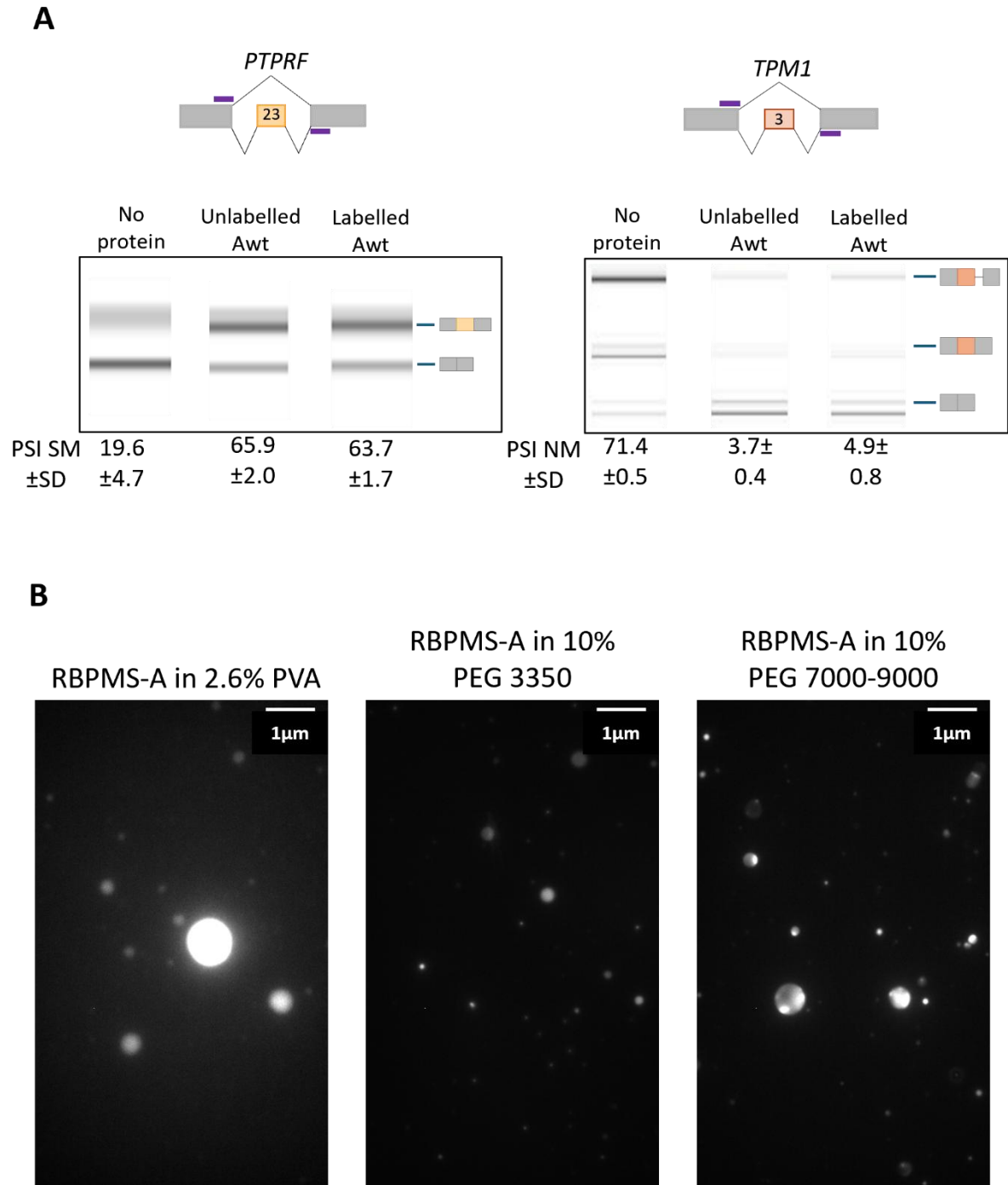


Figure 4.14 Recombinant RBPMs-A can phase separate in the presence of crowding reagents: A: Qiaxcel results showing unlabelled wild-type RBPMs-A (Awt) has the same splicing activity as RBPMs-A labelled with fluorophore Alexa 647. **B:** Microscopy images of recombinant RBPMs-A with different crowding reagents. PEG = Polyethylene glycol, PVA = polyvinyl alcohol. Measurements taken at $\lambda = 647$ nm.

To further investigate the effect of crowding reagents on the phase separation behaviour of RBPMS-A, different concentrations of protein and PVA were tested via microscopy (Fig.4.15). PVA was chosen as the crowding reagent in this study as it produced the largest droplets (Fig.4.14B). Additionally, PVA at 2.6% is used to promote efficient splicing *in vitro* (Krainer *et al*, 1984). A high concentration of recombinant RBPMS-A was found to not phase separate by itself (Fig.4.15), but if $\geq 1.3\%$ PVA is added, then droplets start to appear, at concentrations as low as $0.5\ \mu\text{M}$, although the droplets are very faint. $22\ \mu\text{M}$ of protein and 2.6% of PVA gives rise to the largest droplets. Liquid-liquid phase separation has been reported to be affected by both protein concentration and crowding reagents; therefore it is likely that the droplets seen in Figures 4.14 & 4.15 are LLPS droplets. More testing, such as fluorescence recovery after photobleaching (FRAP) would be necessary to confirm this. RBPMS-A has not been observed to undergo LLPS in cells, although it can enter stress granules during oxidative stress (Farazi *et al*, 2014). Hence, it is likely that phase separation of RBPMS-A at a large and observable scale in cells is not physiological, and the phase separation seen here can instead be used as an indicator of the propensity of RBPMS-A to form higher-order assemblies, with LLPS droplets being an extreme example of this behaviour.

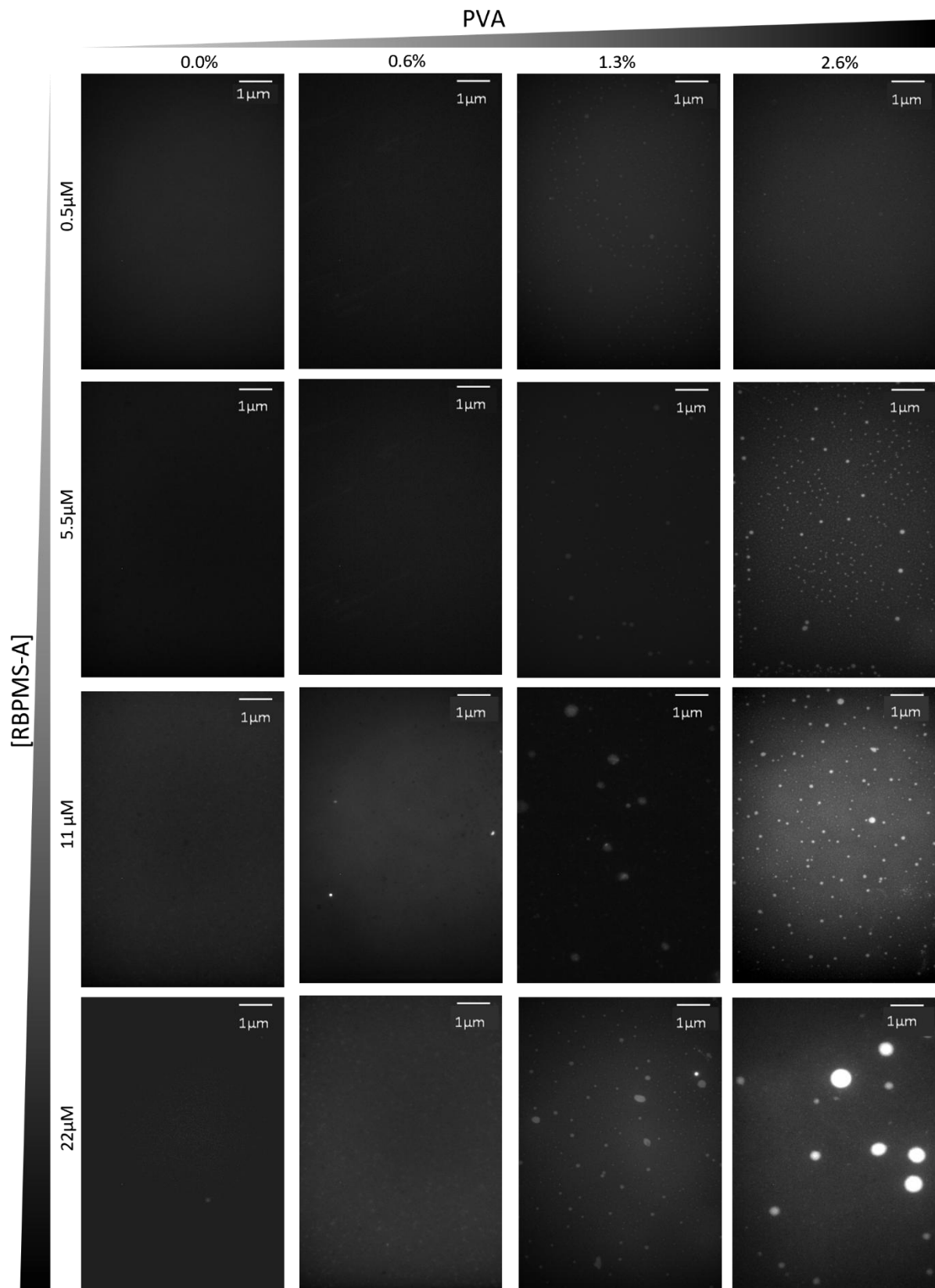


Figure 4.15 Recombinant RBPMS-A droplet formation is dependent on concentration of protein and PVA: Microscopy images of recombinant RBPMS-A. Increasing concentration of RBPMS-A and PVA were used. Measurements were taken at $\lambda = 647$ nm.

Mass photometry reveals higher-order assembly of RBPMS-A mutants is abrogated relative to wild-type

To uncover how important the aromatic, basic and phosphorylated threonine residues are for the higher order assembly of RBPMS-A, mass photometry experiments were performed. Mass photometry measures the interference between the light that gets scattered as it encounters a glass coverslip whilst being subjected to a beam of light, and the light reflected back at the coverslip (Young *et al*, 2018). This interference is directly proportional to the mass of single protein molecules (or other macromolecules) on the glass surface. The smallest mass that can be detected by mass photometry is 42 kDa, which is too large to capture RBPMS-A monomers (22 kDa) but is still able to capture dimer species and above. The AUC data of RBPMS-A and Δ C20 proteins did not show any monomers forming, hence mass photometry is an ideal method to use for the study of RBPMS oligomerisation (Yang *et al*, 2023). This method allows for measurements of mass in solution and can show molecular heterogeneity, as it measures the mass of each molecule that touches the glass coverslip. This is an important advantage for measuring the mass of RBPMS-A assemblies, as they have previously been observed to be highly dynamic (Yang *et al*, 2023).

Wild-type RBPMS-A was observed to form a dynamic range of oligomers, from dimers to oligomers with >20 subunits (Fig.4.16). This agrees with previous AUC data, demonstrating that mass photometry is an appropriate tool for observing RBPMS self-assembly. The aromatic mutants all have diminished higher-order assembly significantly, down to approximately dimer/trimer. This implicates these aromatic residues in forming important interactions necessary for the self-assembly of RBPMS-A and may link to the loss of activity of these mutants.

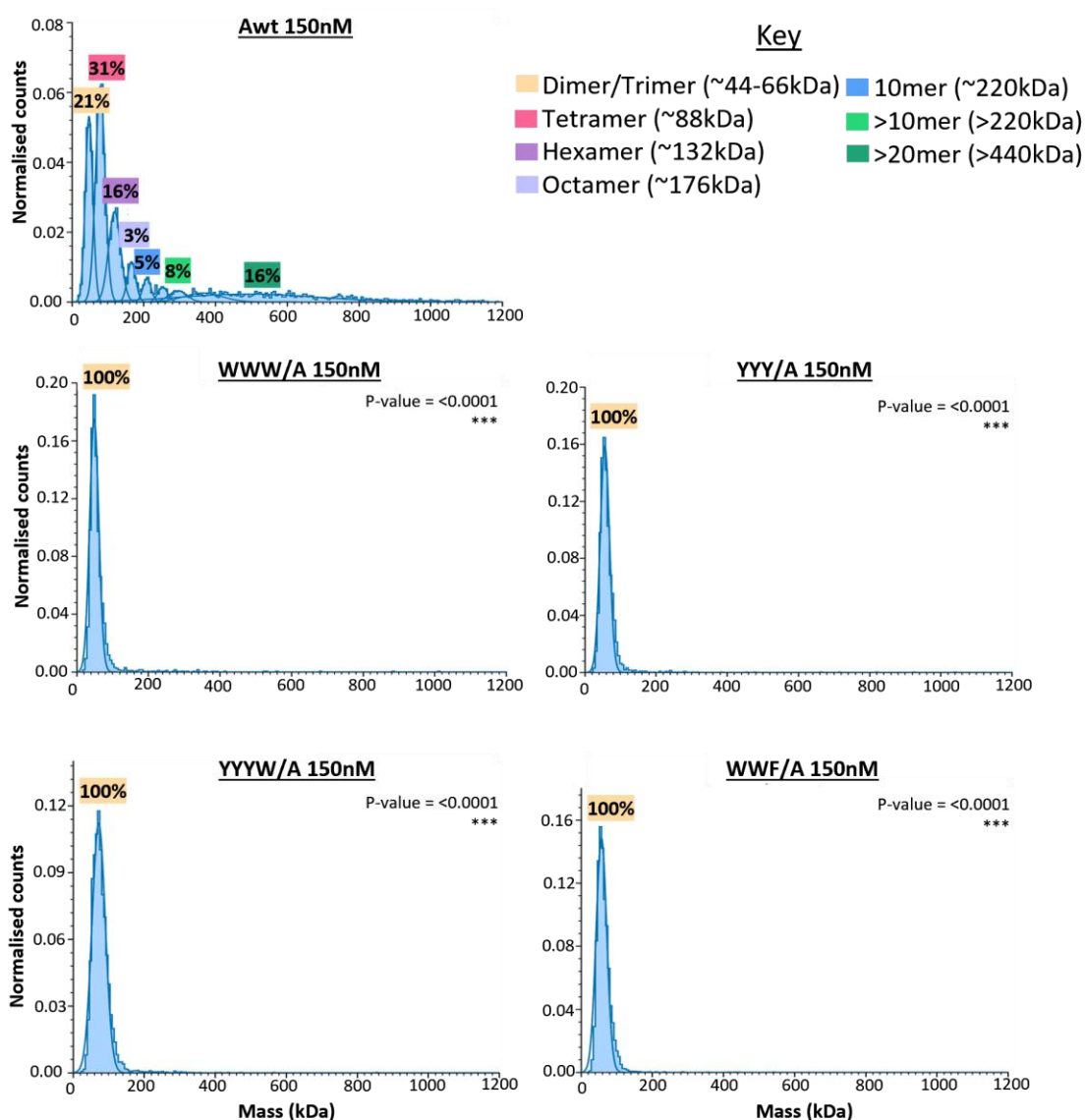


Figure 4.16 Aromatic residues of RBPMS-A C-terminus are important for higher-order assembly: Mass Photometry results with recombinant RBPMS-A and its aromatic mutants at 150 nM concentration. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed. Statistical test was a non-parametric two-sample Kolmogorov-Smirnov test examining cumulative distribution functions between the overall distribution of oligomers formed by the mutants relative to the wild-type (p value indicated in the top right-hand corner of each graph).

The phospho and basic mutants also had reduced higher order assembly compared to wild-type (Fig.4.17). The RKR/E mutant had reduced higher-order assembly, being 96% dimer/trimer, with very smaller peaks for hexamer and octamer. Surprisingly, the RKR/A mutant also had reduced higher order assembly and mainly formed dimer/tetramer species (52% and 41% respectively), as well as a small amount of hexamer and octamer, despite this mutant having full activity in the *in vitro* splicing assays and transfections. The RKR/A mutant result suggests that higher-order assembly and activity of RBPMS-A can be uncoupled, even though for the $\Delta C20$ mutant these properties appeared to be closely related (Yang *et al*, 2023). Statistically, the cumulative distribution functions (CDF) of the oligomeric states of the RKR/A and RKR/E mutants showed no difference (Kolmogorov-Smirnov test, p-value = 0.9993), but both profiles were significantly different to wild-type (Fig.4.17). This result may simply indicate tetrameric RBPMS-A is the oligomer that is required for activity, as the presence of this species is what separates this mutant from the inactive aromatic and reverse charge mutants. Wild-type RBPMS-A showed no significant difference in its CDF relative to the TT/A mutant, but did with TT/E, although the p-value was high (0.0281, *). Additionally, the phospho mutants both had very similar higher-order assemblies to each other (p-value = 0.9993), with both able to form some species >10 although only TT/A formed species above 20mers. The TT/A mutant >20mer peak showed a lower frequency compared to wild-type, thus a further Kolmogorov-Smirnov statistical test was performed to compare the peaks only for >20mer of the wild-type and TT/A proteins. This showed there was only a slight significant difference between them (16% 20mer for wild-type, 3% for TT/A giving a p-value of 0.0310, *). The result of the TT/E mutant retaining a degree of higher order assembly was unexpected as it reduced both activation and repression activity of RBPMS. This mutant is a further example of higher-order assembly and activity being uncoupled.

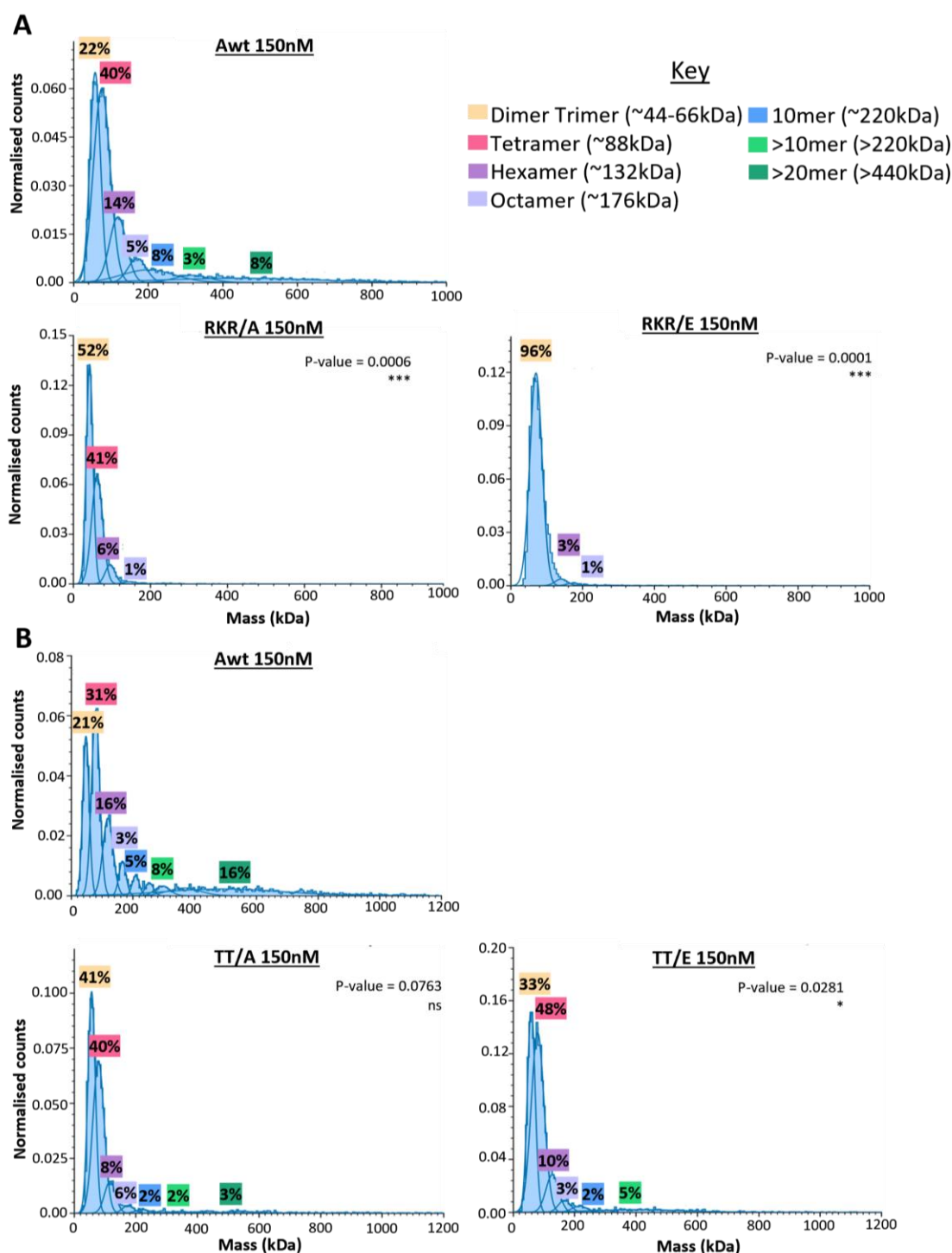


Figure 4.17 Basic residues of RBPMS-A C-terminus are important for higher-order assembly: A: Mass Photometry results with recombinant RBPMS-A and its basic mutants at 150 nM concentration. **B:** Mass photometry results for RBPMS-A wild-type and phospho-mutants. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed. Statistical test was a Kolmogorov-Smirnov test examining cumulative distribution functions between the overall distribution of oligomers formed by the mutants relative to the wild-type (indicated in the top right-hand corner of each graph) or comparing cumulative distribution functions of individual oligomeric states of the mutants or wild-type (indicated in the text).

Previous studies of higher-order assembly of RBPMS-A used μM range (Barnhart *et al*, 2022; Yang *et al*, 2023). Mass photometry is limited by concentration, with the allowed concentration range being between 100 pM – 150 nM approximately, which is below the concentration used in the splicing assays (1 – 2 μM). Additionally, the splicing assays are performed with other proteins from the HeLa nuclear extract present, and with crowding reagents. Therefore, it is possible that with measurements taken at 150 nM, the mutants do not show higher-order assembly behaviour, but this may be rescued at higher concentrations of protein. In particular, this may be the case with the differences between the active/inactive RKR/A versus RKR/E and TT/A versus TT/E proteins. With measurements taken above 150 nM, the signal to noise ratio is high and it makes it difficult to accurately discern the molecular weight of individual species. The mass photometry profiles of wild-type RBPMS-A and the mutants were tested at 1000 nM (Fig.4.18). Here, the noisy signal is noticeable by the large negative peaks, this is due to the glass coverslip becoming too overcrowded with protein, so other protein molecules are not able to bind to the glass surface itself. This means these particles will encounter other protein molecules instead of the glass slide causing the bound protein to detach from the glass, leading to a void binding (unbinding) event. Despite this noise, there is a clear difference between wild-type and the mutant proteins. Wild-type RBPMS-A forms many higher order species, whereas the sizes of the complexes formed by the other mutants are greatly reduced, except for TT/A. The RBPMS-A mutant protein TT/A is still able to form some higher order species (>20mer), although the amount seen is reduced relative to wild-type RBPMS-A.

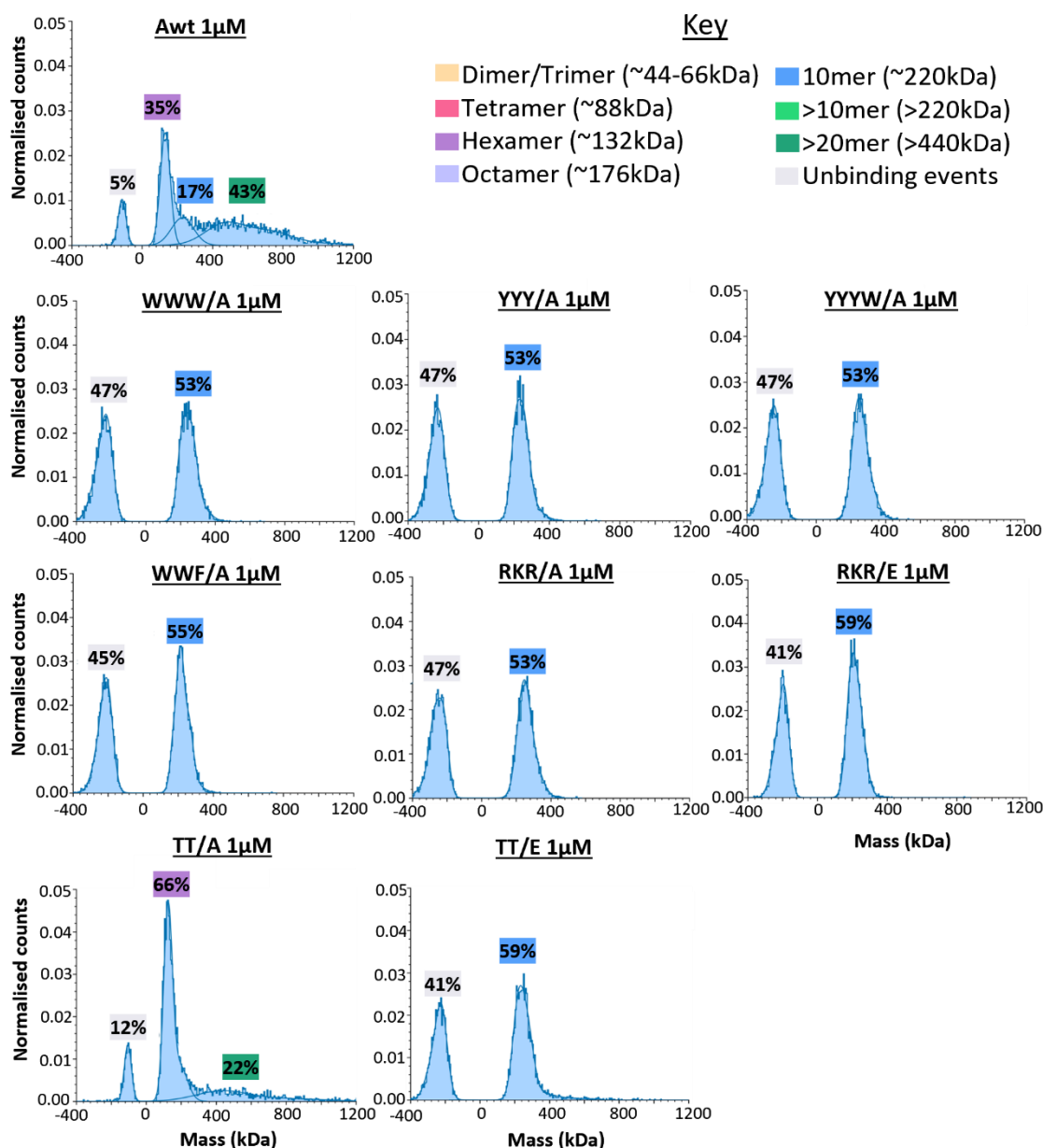


Figure 4.18 Higher concentration confirms basic and aromatic residues of RBPMS-A are important for higher-order assembly: Mass Photometry results with recombinant RBPMS-A and its aromatic mutants at 1 μ M concentration. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed.

To understand the importance of the substituted residues on oligomerisation at higher concentration without the issue of high signal to noise ratios, glutaraldehyde crosslinking was used to stabilise oligomers, before rapid dilution to appropriate concentrations for mass photometry. To assess the amount of glutaraldehyde to use, a pilot experiment was performed using wild-type RBPMS-A and the

$\Delta C20$ mutant proteins with different amounts of glutaraldehyde. The $\Delta C20$ mutant does not form higher-order species, and thus was used as a negative control. After crosslinking, samples were run on SDS-PAGE 4-12% Bis-Tris gradient gels and Coomassie Blue stained (Fig.4.19). For wild-type RBPMS-A samples, the higher order assemblies were able to be captured at each glutaraldehyde concentration. The lowest amount of glutaraldehyde (0.0015%) was decided as the most appropriate, since it still captures the higher-order assemblies of the wild-type sample but not the $\Delta C20$ mutant (lane 2 versus lane 6), indicating it is not capturing non-specific crosslinking events. Lane 1 shows some bands which are likely as a result spillage from the sample in lane 2, but this does not obscure the outcome of this pilot, which is that 0.0015% glutaraldehyde is a suitable concentration for future experiments.

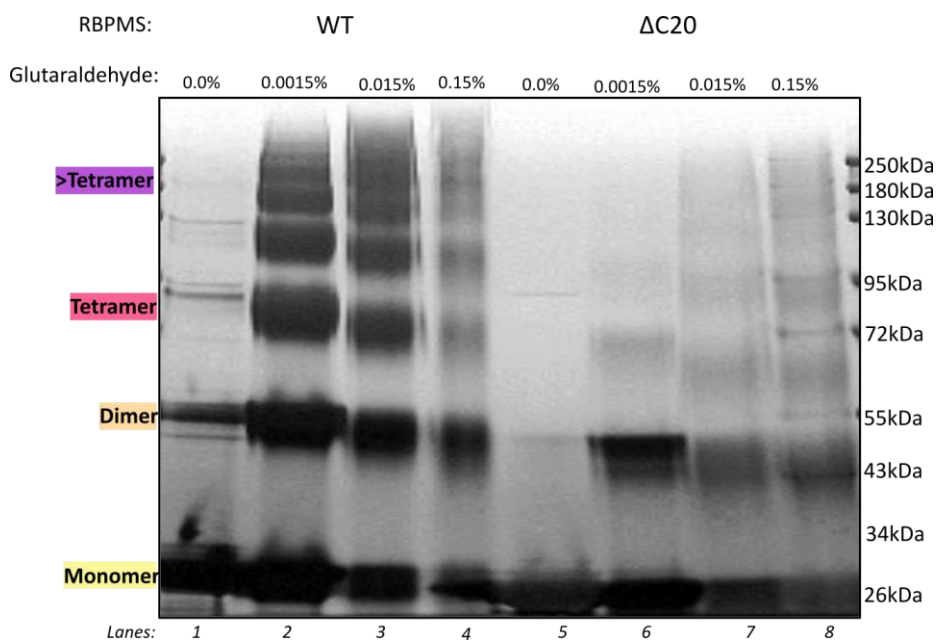


Figure 4.19 Glutaraldehyde crosslinking can specifically capture the higher-order species of RBPMS: Coomassie stained 4-12% Bis-Tris gel showing oligomeric species of RBPMS-A wild-type vs $\Delta C20$ mutant with different concentrations of glutaraldehyde. The bands correlating with the expected sizes of oligomeric species of RBPMS-A are shown on the left side of the gel.

The recombinant RBPMS mutants were crosslinked at 20 μ M, and oligomeric state was detected with SDS-PAGE (Fig.4.20A) and mass photometry (MP) (Fig.4.20B). The aromatic and basic mutants had reduced higher-order assembly relative to wild-type, although there are still some higher-order species present (8-10% >tetramer species, versus 40% for wild-type) (Fig.4.20A). The basic mutants again had reduced higher-order assembly relative to wild-type, with the RKR/A mutant appearing similar to the aromatic mutants, and RKR/E having even less higher order assembly than RKR/A (14% >tetramer for RKR/A and 6% for RKR/E). The phospho-mutants have retained more higher-order assembly, with the TT/A mutant now appearing more similar to wild-type (~40% >tetramer with WT, versus ~30% for

TT/A). Quantitation of individual species above tetramer was difficult from the gel due to the smear effect that the higher-order assemblies produced. To be able to quantify how much of each mutant was in a particular state, mass photometry was used (Figure 4.20B). Here, the samples were rapidly diluted to 100 nM just before the MP experiment was performed. After crosslinking, wild-type RBPMS-A had a higher proportion of subunits in higher oligomeric states than in the previous experiment with 150 nM protein (70% vs 16% >20mer; p-value = 0.0009 ***, Figures 4.16 & 4.17). This confirms that wild-type RBPMS-A higher order assembly is dynamic and dependent on protein concentration, which was shown in the phase separation assays (Fig.4.15) and in previous studies (Yang *et al*, 2023). The aromatic and basic mutants now have some higher order species, mainly dimer, tetramer and hexamer, but they still do not form the more extreme higher order assemblies that the wild-type protein does (>20mer). This confirms that the basic and aromatic residues are important for forming homotypic interactions necessary for RBPMS-A self-assembly into higher-order oligomers. The lack of higher-order assembly of the RKR/A mutant compared to the wild-type protein once again points towards uncoupling of higher order oligomerisation and splicing activity. Both basic mutants lack a distinct peak for the dimer/trimer species, suggesting there may be more tetrameric species than dimer/trimers for these mutants, although this was not indicated by the SDS-PAGE gel quantitation. The SDS-PAGE gel showed the RKR/E mutant protein had reduced assembly compared to RKR/A, which the mass photometry also shows (Fig.4.20). However, the RKR/E mutant on the gel had reduced assembly compared to the aromatic mutants also, which was not seen in the mass photometry experiment. This may be due to the RKR/E mutant having a higher proportion of protein in a monomeric state compared to the other proteins, which cannot be detected by mass photometry. If these monomeric species are removed from the gel quantitation, the proportion of the RKR/E mutant in dimer/tetramer/>tetramer states appears similar to the aromatic mutants but is still reduced relative to the RKR/A mutant (data not shown). Given that the RKR/A mutant formed more molecular weight species (five different species were observed) compared to RKR/E, where the discover MP software was only able to call three distinct peaks, it appears that the RKR/A mutant can form more heterogeneous higher-order structures >tetramer, compared to RKR/E. RKR/A also formed more octameric species than RKR/E (18% versus 5%, p-value = 0.026, *), and less hexamers (26% versus 43%, p-value = 0.01, *), and was able to some >10mer species (5%, p-value = <0.0001, ***), which the RKR/E and aromatic mutants did not. In the lower concentration studies, the RKR/A mutant was also able to form more oligomers than the RKR/E, with four versus three distinct species for RKR/A and RKR/E respectively (Fig.4.17A). This broader range of oligomers that the RKR/A mutant can form may be linked to its retention of splicing activity relative to the RKR/E and aromatic mutants. Glutaraldehyde crosslinks via lysine residues, thus there was some concern that the RKR mutants would be less efficiently crosslinked than the other mutant proteins

(Kim *et al*, 2024). This is unlikely to be the case here, however, as RBPMS-A contains twelve lysine residues altogether, thus mutating one is unlikely to influence glutaraldehyde crosslinking efficiency. Furthermore, Figure 4.17A showed the RKR mutants both had reduced oligomerisation compared to the wild-type protein when not crosslinked, suggesting the crosslinked observations for these mutants are true. The phospho-mutants once again showed similar profiles to each other, the main difference being that the TT/A mutant had small amount of >20mer species present (8%) whereas TT/E did not, although this was still much reduced compared to wild-type (~70% >20mer, p-value = <0.0001, ***). Higher order assembly >tetramer was also much more reduced for TT/A relative to wild-type (40% >tetramer with TT/A versus 80% >tetramer for wild-type), whereas these proportions were much more similar on the SDS-PAGE gel. SDS-PAGE is much harsher than mass photometry, and thus any protein that is not crosslinked will be reduced to monomer. This means that the amount of protein in each state captured on the gel may not represent the amount of protein actually in the various oligomeric states, whereas mass photometry does capture this. These differences between the SDS-PAGE and mass photometry techniques likely explains any differences observed in the amount of protein captured in each oligomeric state.

Overall, these higher-order assembly results have shown that the aromatic and basic residues of the C-terminus are necessary for the full oligomerisation behaviour of RBPMS-A. This suggests that these residues may be forming cation – π or π – π interactions with other RBPMS-A subunits and that these interactions are necessary for the more extreme higher-order assembly behaviour of RBPMS-A. The threonines 113 and 118 also appear to be important for extreme higher-order assembly, as substituting these residues to alanine or glutamate affected oligomerisation. The RKR and TT mutants have shown that higher-order assembly of RBPMS-A and its splicing activity can be uncoupled, as RKR/A and TT/A are fully active, at all concentrations tested in Figure 4.10, but both have reduced oligomerisation compared to the wild-type protein at low and high concentration (Figures 4.17 & 4.20).

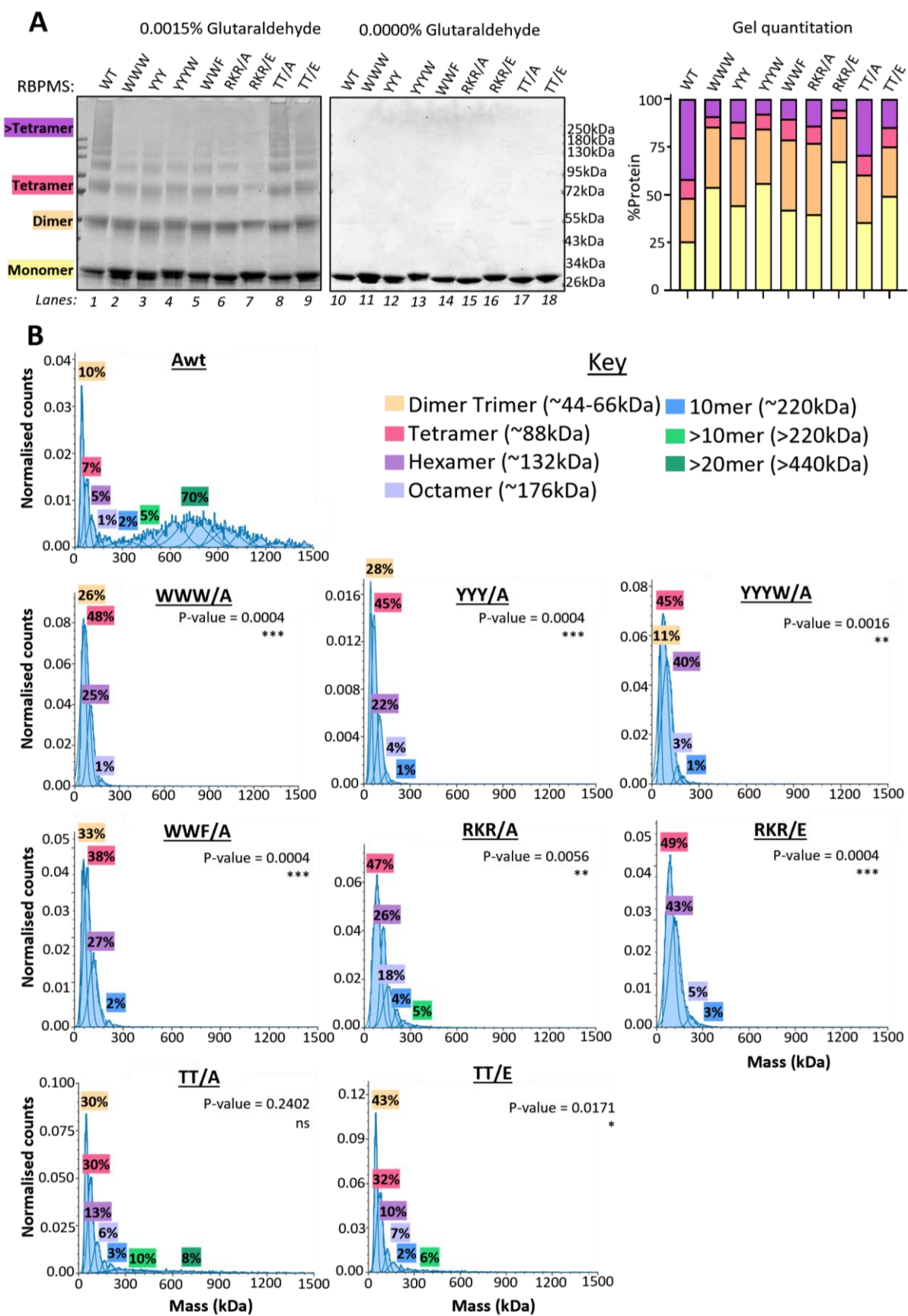


Figure legend on next page.

Figure 4.20 Crosslinking of RBPMS-A confirm higher-order assembly behaviour is dependent on aromatic and basic residues: **A:** Coomassie Blue stained 4-12% Bis-Tris gels showing oligomeric species of wild-type RBPMS-A vs its mutants with and without glutaraldehyde. Bands correlating with the expected sizes of the oligomeric species of RBPMS-A are shown on the left side of the gel, and quantitation of these areas of each lane is shown on the graph on the right. **B:** Mass Photometry results with recombinant RBPMS-A and its mutants at 150 nM concentration, post glutaraldehyde crosslinking. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed. Statistical test was a Kolmogorov-Smirnov test examining cumulative distribution functions between the overall distribution of oligomers formed by the mutants relative to the wild-type (indicated in the top right-hand corner of each graph) or comparing cumulative distribution functions of individual oligomeric states of the mutants or wild-type (indicated in the text).

RNA affects higher-order assembly of RBPMS-A

RNA is known to influence the phase separation and higher-order assembly behaviour of RBPs, and is often found in biological condensates, such as stress granules and nuclear speckles (Buchan & Parker, 2009; Lin *et al*, 2015; Banani *et al*, 2017; Bratek-skicki *et al*, 2020). To understand how RNA may affect higher-order assembly of RBPMS-A, RNA containing a single tandem repeat of CAC₂ or a CCC₂ mutant (42 nucleotides total length), or three tandem CAC₂/CCC₂ pairs (120 nucleotides total length) was titrated against fixed concentration of wild-type RBPMS-A or the WWW/A mutant at 10 µM. The single tandem repeat CAC RNAs are able to mediate simple binding of RBPMS dimers, whereas the triplet tandem repeat CAC RNAs can mediate multivalent cooperative binding of RBPMS (Yang *et al*, 2023).

These RNA-protein complexes were crosslinked with glutaraldehyde, then quenched with glycine, and Benzonase was added so only the protein would be observed by mass photometry. Wild-type protein with each RNA type was tested (Fig.4.21) and with the single tandem CAC RNA, higher order assembly of >20mer species is reduced from 49% to 29% with 150 nM RNA (p-value = 0.105, ns), which was significantly further reduced by increasing RNA concentration. Interestingly, for the CCC mutant, there is an emphatic and significant loss of higher-order species (11% at 150 nM, p-value = <0.0001, ***), although this does not seem to be much affected by increasing RNA concentration above 150 nM. The amount of dimer species increases with the CAC₂ and CCC₂ RNA relative to the 0 nM RNA sample, but this was not found to be statistically significant. The lack of sequence specificity suggests that the RNA effect is not through engagement of the sequence-specific binding site on the RRM. Indeed, the lower potency of the WT CAC₂ RNA could be due its binding to the RRM leading to lower availability for the disruptive effect on oligomerisation.

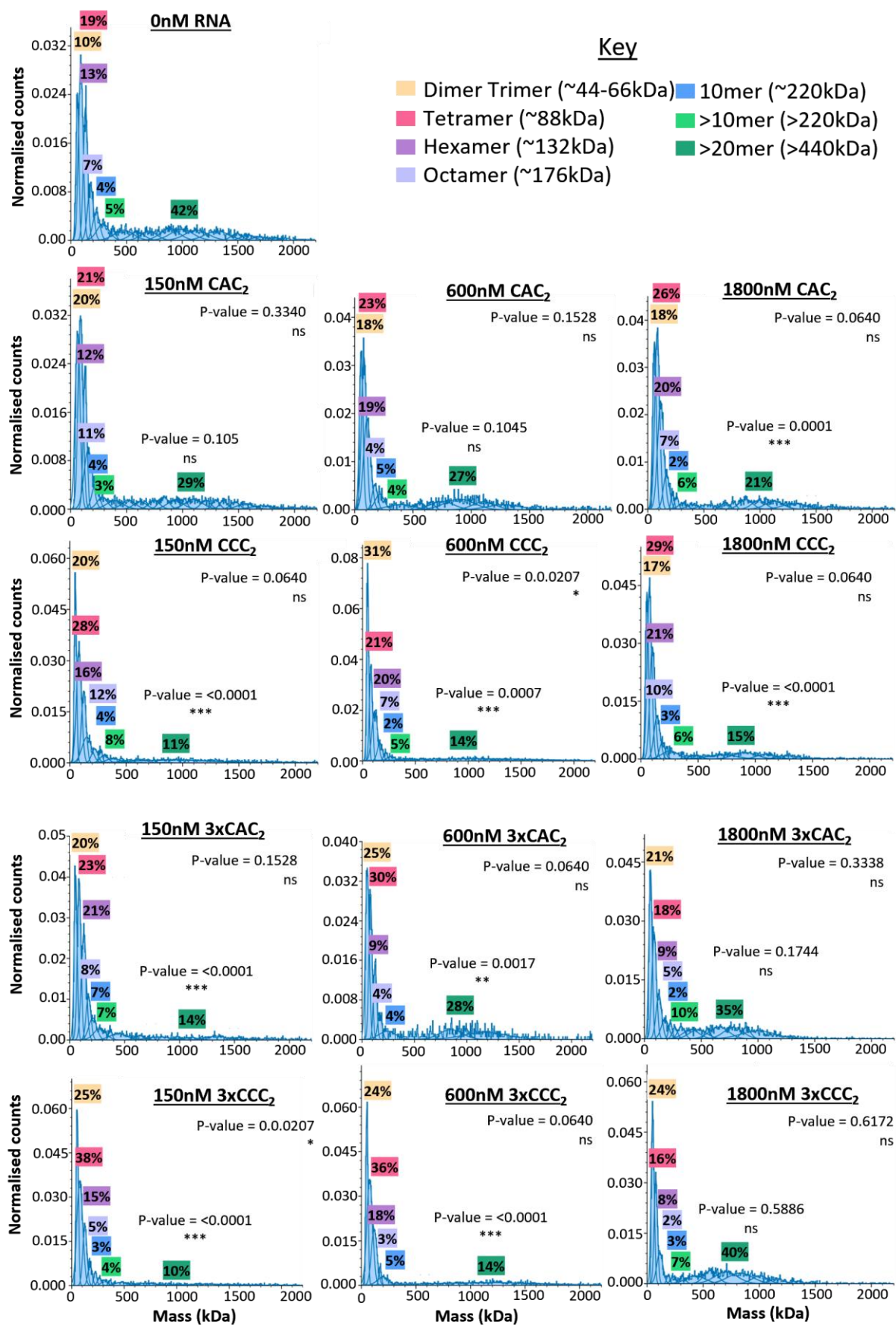


Figure legend on next page.

Figure 4.21 RNA decreases higher-order assembly of RBPMS-A: Mass Photometry results with recombinant wild-type RBPMS-A post addition of RNA at various concentration, either a 42 nt sequence containing a single CAC₂/CCC₂ or a 120 nt sequence containing 3xCAC₂/CCC₂. Samples were crosslinked with 0.0015% glutaraldehyde to stabilize oligomeric species, and Benzonase treated to ensure a clean sample for Mass Photometry. Protein samples were diluted to 150 nM before Mass Photometry was performed. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed. Statistical test was a Kolmogorov-Smirnov test examining cumulative distribution functions between the overall distribution of oligomers formed by each sample to the 0nM RNA sample (indicated in the top right-hand corner of each graph) or comparing cumulative distribution functions the >20mer species of each sample to the 0nM RNA sample (indicated above >20mer peaks).

For the 3xCAC₂ RNA, a biphasic response was observed. The initial significant reduction of >20mer species at 150 nM 3xCAC₂ (42% to 14%, p-value = <0.0001, ***) was followed by an increase at 600 nM (28%, **) and 1800 nM (35%, ns), showing a similar profile to the 0nM RNA sample. A similar biphasic trend was also seen with the 3xCCC₂ RNA. Overall, these results show that the presence of RNA does affect higher order assembly of RBPMS-A, and that length of RNA can alter this, but sequence specificity appears to be less important.

The WWW/A mutant protein was also tested with each RNA type (Fig.4.22). With the single CAC₂/CCC₂ RNAs, there is a decrease in the amount of octameric species forming compared with the 0 nM RNA sample (2-7% compared to 14%), and an increase in dimer species for the CCC₂ RNA above 150 nM concentration. The longer RNAs (3xCAC₂/CCC₂), also reduced the amount of octamer present. None of these changes were found to be statistically significant, and the overall profiles of each sample with the 0nM RNA sample were not statistically different. Compared to wild-type RBPMS-A, the effect of RNA on this mutant appears to be minimal, and higher order assembly is not induced with or without RNA.

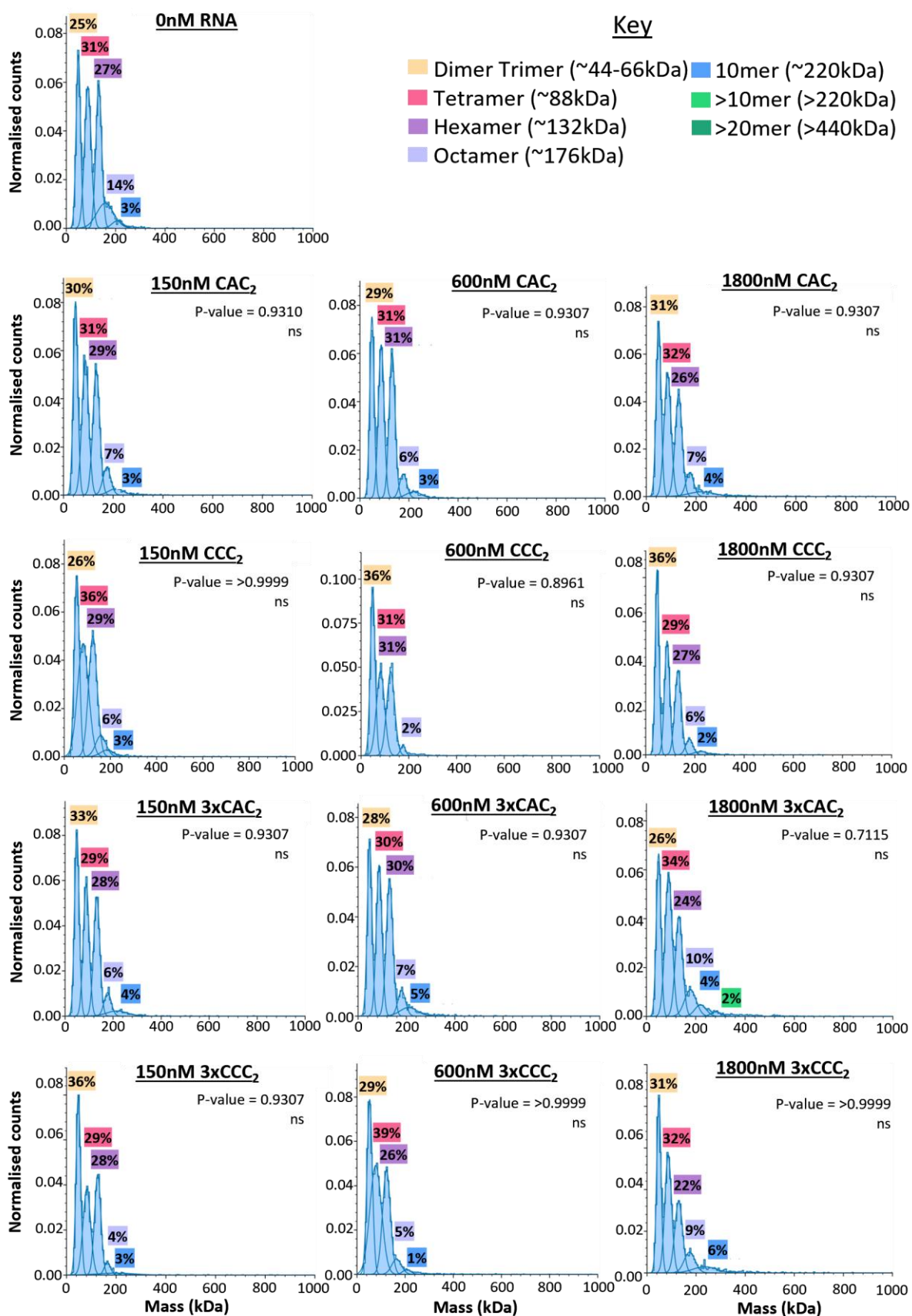


Figure legend on next page.

Figure 4.22 RBPMS-A WWW/A mutant is unaffected by addition of RNA: Mass Photometry results with recombinant RBPMS-A WWW/A mutant post addition of RNA at various concentration, either a 42 nt sequence containing a single CAC₂/CCC₂ or a 120 nt sequence containing 3xCAC₂/CCC₂. Samples were crosslinked with 0.0015% glutaraldehyde to stabilize oligomeric species, and Benzonase treated to ensure a clean sample for Mass Photometry. Protein samples were diluted to 150 nM before Mass Photometry was performed. Histograms for each protein were plotted using the Discovery MP software and represent three technical repeats. Mass (in kDa) is shown along the X axis and normalized counts are shown along the Y axis. The normalization process divides the counts in each peak by the total number of events. Each peak correlates to an oligomer species of RBPMS which have been colour coded, and the percentage of each oligomer type present is displayed. Statistical test was a Kolmogorov-Smirnov test examining cumulative distribution functions between the overall distribution of oligomers formed by each sample to the 0nM RNA sample (indicated in the top right-hand corner of each graph).

4.6 AROMATIC RESIDUES MAY BE REQUIRED FOR HETEROTYPIC INTERACTIONS

Background

The focus of the higher-order assembly experiments discussed above were on the self-assembly of RBPMS-A. Heterotypic interactions between different proteins, however, are also often important for RBPs to perform their function. They may bind antagonistic factors that prevent them from executing a specific function, or they may require binding to co-regulator proteins to have full activity, for example, by binding to core splicing factors and other co-regulator proteins. RBPMS-A has been shown to interact with various splicing factors, including SR proteins, hnRNPs, members of the core spliceosome, and other regulatory proteins, including RBFOX2 and MBNL1 (Yang *et al*, 2023). Knockdown of RBPMS, RBFOX2 and MBNL shows that these proteins regulated many of the same splicing targets in a similar manner. The Δ C20 mutant lost most of these key interactions, suggesting that the residues within this extreme C-terminal region may be important for these heterotypic interactions. Therefore, pull-down analyses of RBPMS-A and its mutants were attempted in HeLa nuclear extract to observe if there were any differences between the mutants and wild-type protein in their protein-protein interactions (PPIs).

RBPMS-A can be specifically pulled down in nuclear extract

Initially, a pilot experiment was performed with and without wild-type RBPMS-A to see if any of the known interactors of RBPMS-A could be captured. N-terminally STREP 6xHis tagged RBPMS-A or buffer was added to 60% nuclear extract and a binding reaction was performed. This nuclear extract mix was then added to STREP-tag beads which were washed, and RBPMS-A and its bound proteins were eluted using biotin. Efficiency of the wash and elution steps were measured by silver stain analysis of the wash

sample, the amount of sample left on the beads after the elution step, and the eluate itself (Fig.4.23A). The sample still on the beads after the biotin elution was identified by boiling the beads to elute what was left and analysing them on a SDS-PAGE gel (Fig.4.23A, lanes 4 & 5). These lanes show that there is some protein left on the beads, and there is a band that is the right size for RBPMS-A, showing that the elution was not 100% efficient. Lanes 6 and 7 show the eluate of the $\pm$ RBPMS-A samples. The highest signal is from RBPMS-A itself in the +RBPMS-A lane (lane 6), and there are other bands present that likely show proteins specifically interacting with RBPMS-A and not the beads as most are not present in the -RBPMS-A control (lane 7). Western blot analysis was used to identify if any previously reported interactors of RBPMS-A were present in this pull-down (Fig.4.23B). RBFOX2 and MBNL were seen to be specifically pulled down with RBPMS-A in this experiment. RBPMS-A was also identified in the pull-down by use of the anti-STREP antibody, confirming that elution was successful. The pulldown of RBFOX2 and MBNL1 was RBPMS dependent, as they are not seen in the -RBPMS-A lanes (lanes 2 versus 3).

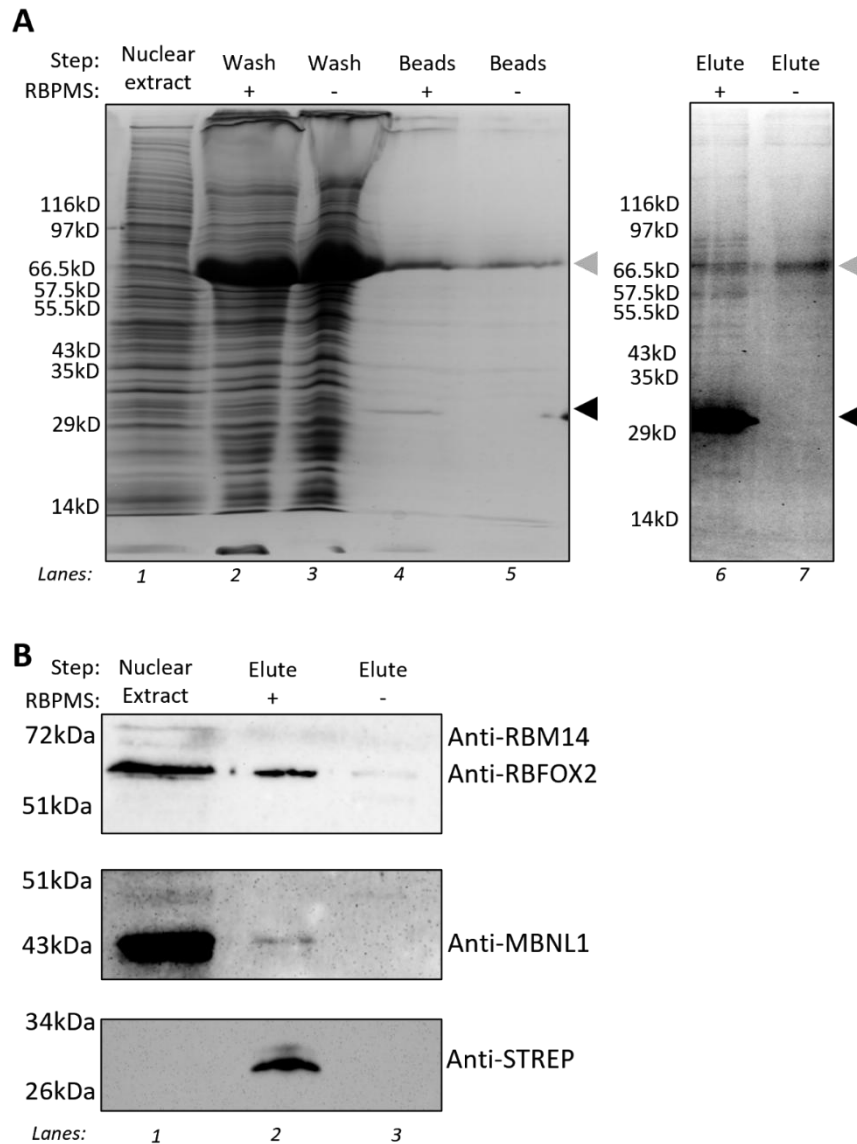


Figure 4.23 Pilot pull-down experiments confirm that RBPMS-A can interact with MBNL and RBFOX2:
A: Silver stain of pilot pull down. Gel on the left-hand side shows the unbound proteins from the wash step (wash), and the proteins that were still on the beads post biotin elution (beads). The gel on the right-hand side shows the proteins found in the eluate. The expected mobility of RBPMS-A and BSA (used in the bead blocking step) are shown by a black or grey arrow respectively. **B:** The eluate with and without RBPMS-A were probed for RBFOX2 and MBNL in a western blot analysis. RBPMS itself was probed for using an anti-STREP antibody.

Pull-down analysis with RBPMS mutants in nuclear extract suggests aromatic residues are required for heterotypic interactions

A pull-down experiment with two aromatic mutants, the basic mutants and the phosphomimetic mutant was performed in nuclear extract, as described previously. Silver stain analysis was carried out to confirm efficiency of each step of the protocol as before (Fig.4.24A). The boiled beads sample (lanes 10 – 16) show that there is still RBPMS bound to the beads, showing the elution was not 100% successful, however, in the eluate (lanes 17 – 23) the RBPMS signal was strong, suggesting most of the protein was eluted. The no protein control (lane 17) has very few bands compared to the other samples, showing that pull-down was clean and that most of the non-specific interactions have likely been lost during the wash steps. Additionally, there appears to be differences in the number of bands present and their location between the different mutants and wild-type, indicating that the pull-down is capturing differences in the PPIs of these proteins. A proteomic LC-MS/MS experiment was performed but it did not show the presence of many known RBPMS interactors, or many other RBPs at all (data not shown). This suggests the pull down may not be efficient, which may have been due to true interactions being lost in the wash steps. Further work is needed to improve the pull-down technique during the wash and elution steps to get a full picture of the interactomes of RBPMS-A wild-type and its mutants. Western blot analysis was performed on the elutes with a small number of targets to identify if any interactors could be observed (Fig.4.24B). Here, RBM14 can now be seen to be interacting with wild-type RBPMS-A, as well as the RKR mutants, particularly to RKR/E. RBM14 was not seen to interact with wild-type RBPMS-A in the pilot experiment (Fig.4.23B lane 2), showing inconsistencies between the two experiments. Different batches of nuclear extract were used in each experiment, with the nuclear extract in the pilot showing potentially less RBM14 than in the experiment with the mutants (Fig.4.23B lane 1 & Fig.4.24B lane 1). Additionally, MBNL was seen to interact with wild-type RBPMS-A in the pilot experiment (Fig.4.23B), but there was no signal with either wild-type RBPMS-A or its mutants in this experiment (data not shown). RBFOX2 pulled down by wild-type RBPMS-A in this experiment, as well as RKR/A and TT/E (Fig.4.24B). This suggests that these two mutants retain their interactions with this protein. RKR/A retains interactions with both RBM14 and RBFOX2, although the amount of RBM14 seen is significantly less than is seen for wild-type or RKR/E. Future experiments where more interacting proteins can be identified will reveal if the RKR/A mutant consistently retains interactions whilst the other splicing incompetent mutants lose interactions, revealing if there is a link between ability to form heterotypic interactions and splicing activity. The aromatic mutants (WWF/A and YYYW/A) in Figure 4.24B showed a loss of interaction with

RBFOX2 and RBM14 relative to wild-type, suggesting that these residues are important for forming protein-protein interactions.

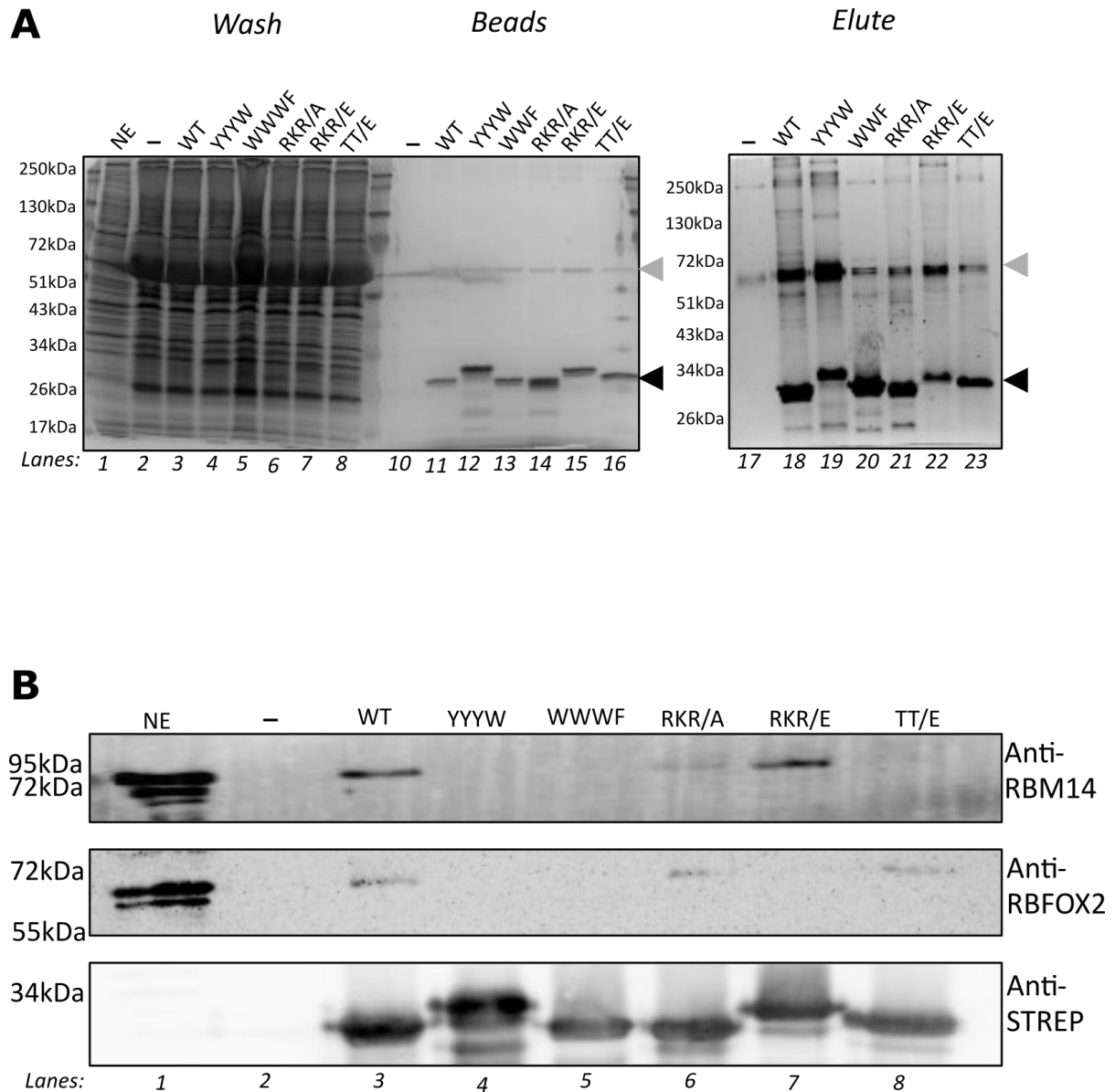


Figure 4.24 Preliminary pull-down data suggests the RBPMS-A aromatic mutants lose interaction with RBFOX2 and RBM14: **A:** Silver stain of pull down with wild-type RBPMS-A and mutants. Gel on the left-hand side shows the unbound proteins from the wash step (wash), and the proteins that were still on the beads post biotin elution (beads). The gel on the right-hand side shows the proteins found in the eluate. The expected mobility of RBPMS-A and BSA (used in the blocking step) are shown by a black or grey arrow respectively. **B:** The eluates from each protein sample were probed for RBFOX2 and RBM14 in a western blot analysis. RBPMS itself was probed for using an anti-STREP antibody. NE = nuclear extract.

4.7 CONCLUDING REMARKS

Protein	Splicing activity	RNA crosslinking ability (in buffer/Nuclear extract)	Higher order oligomerisation potential	Protein-protein interactions
Wild-type	++++	++++/++++	++++	++++
WWW/A mutant	-	+++/-	+	NT
YYY/A mutant	-	+++/-	+	NT
YYYW/A mutant	-	+++/-	+	-
WWF/A mutant	-	+++/-	+	-
RKR/A mutant	++++	++/+	++	+
RKR/E mutant	-	-/-	+	+
TT/A mutant	++++	+++/+	+++	NT
TT/E mutant	+	++/-	++	+

Table 4.1 Summary of the results from the RBPMS-A mutants tested in this results chapter: Each RBPMS-A mutant is compared to wild-type RBPMS-A in each category tested in this chapter. ++++ indicates the highest activity seen, and – equals no activity. NT = not tested.

RNA binding proteins are often able to undergo oligomerisation to perform certain functions. How these proteins oligomerise, the size of the assemblies they form, and their PPI networks depend on their amino acid composition. Intrinsically disordered regions and low complexity domains have frequently been found to be important for oligomerisation due to the residues being more likely to encounter residues from other subunits and proteins as they are not buried in the core of a protein domain. Instead, these regions can extend, allowing the residues to form multi-valent interactions with protein and RNA targets alike. The intrinsically disordered C-terminal tail of RBPMS-A was found to be required for splicing regulatory activity as well as for forming homo- and hetero-typic interactions. This discovery was based on a deletion of the final twenty amino acids of RBPMS-A, and thus further investigation was required to uncover which residues in this region were important and why (Yang *et al*, 2023). Identifying whether residues that were necessary for activity also impact the biophysical behaviours of RBPMS-A would help to confirm if the ability of RBPMS-A to undergo higher-order assembly and its splicing activity are always linked.

Substitution mutants were made to uncover which residues within the IDR of RBPMS-A are important for activity (Figures 4.2 – 4.5, and 4.8 – 4.10). Aromatic and basic residues were chosen as such residues

are able to form cation – π and π – π interactions, known to aid higher-order assembly of other proteins (Sabari *et al*, 2020). Both the overexpression data (Figures 4.2 – 4.5) and *in vitro* splicing assays (4.8 – 4.10), identified that the aromatic residues of the IDR are required for the splicing activity of RBPMS-A, but the basic residues of the extreme C-terminus (RKR) are not. Other residues previously reported to be important for oligomerisation (histidines, glutamines and prolines) were also mutated but these had mixed effects on activity. Given that the aromatic residues are necessary for activity and that activity and higher-order assembly has previously been linked, π – π interactions are likely important for activity and assembly of RBPMS-A; whilst basic residues and therefore cation – π interactions are not.

With the success of the *in vitro* splicing assays replicating the results that were seen in the co-transfection experiments, the biophysical behaviour of the mutants could be directly compared to their activity. The full-length protein was shown here to exhibit higher order assembly behaviour and can even be pushed to undergo LLPS in certain conditions (Figures 4.14 & 4.15). It was expected that the mutants whose activity had been lost or reduced (WWW/A, YYY/A, YYYW/A, WWF/A, RKR/E and TT/E), would also lose the ability to form the higher order oligomers that the wild-type protein has, as had been observed for the Δ C20 mutant. Moreover, the mutants that did not lose activity (RKR/A and TT/A) would be expected to retain the higher-order assembly behaviour of the wild-type protein. As anticipated, the aromatic mutants had reduced higher order assembly relative to the wild-type protein at all concentrations tested (Figures 4.16, 4.18 & 4.20). This demonstrates that the aromatic residues are required for both activity and higher order oligomerisation and implies that these two abilities of RBPMS-A are connected. It is also consistent with π – π interactions being important for oligomerisation and driving this process. The basic mutants, however, had very similar oligomerisation profiles to each other at each concentration tested, and they both had impaired ability to form higher order oligomers compared to the wild-type protein (Figures 4.17, 4.18 & 4.20). This was surprising, as the RKR/A mutant had full activity, and the RKR/E mutant had no activity in the *in vitro* assays. These results challenge the hypothesis that the splicing activity and higher-order assembly of RBPMS are linked, and that higher order assembly is required for activity. The RKR/A mutant protein uncouples higher-order assembly and activity and suggests that cation – π interactions may be required for oligomerisation, but not activity. At each concentration, although it was not observed to form higher order oligomers of >20mers like wild-type, the RKR/A mutant did form a wider variety of species and higher-molecular weight species than its inactive counterpart, RKR/E. This may mean that the ability for RBPMS-A to regulate splicing does not necessarily require the more extreme higher-order species seen (>20mer) but does require more diversity in the oligomers that can form, which the aromatic and RKR/E mutants lack the ability to do.

Both phospho-mutants had reduced numbers of extreme higher order assemblies (>20mer) relative to the wild-type protein, although they both formed greater oligomers than the aromatic and basic mutants. The SDS-PAGE gel results from the crosslinked proteins, however, did show TT/A had a higher proportion of >tetrameric species than the other mutants and was more similar to the wild-type protein (Fig.4.20). Additionally, TT/A has previously been reported to have similar oligomerisation behaviour to the wild-type protein (Barnhart *et al*, 2022). This study used analytical ultracentrifugation (AUC) to test higher-order assembly which is impacted by heterogeneous samples. It may be that mass photometry is more sensitive and can detect the differences between the wild-type and TT/A proteins. To resolve this discrepancy, AUC experiments can be repeated with TT/A and wild-type proteins. The TT/E mutant did not have any splicing activity, thus its retention of some higher order assembly was unexpected. Previous NMR data indicated that the TT/E region of RBPMS-A directly interacts with the RNA binding surface of the RRM, thus affecting RNA binding. In the case of this mutant, it is therefore likely that the lack of activity is not solely due to differences in higher-order assembly and is instead due to altered RNA binding.

These mass photometry experiments only observed oligomerisation of the protein by itself, whereas in cells, RBPs interact with a variety of different proteins and can form hetero-oligomers, and they also interact with RNA which can influence their higher-order assembly and phase separation behaviour. Figure 4.21 demonstrates that the higher order assembly of RBPMS-A wild-type is reduced by the addition of RNA, and that this response does not appear to be sequence-specific. In fact, both the CCC₂ and 3xCCC₂ mutants had a greater impact on higher-order assembly than their CAC counterparts, particularly in the 150 – 600 nM concentration range. The triple tryptophan (WWW/A) mutant was mainly unaffected by RNA, likely because WWW/A mutations have already reduced oligomerisation beyond that which addition of RNA can (Fig.4.22). The UV crosslinking data (Figures 4.11 – 4.13) shows that the wild-type protein can crosslink to RNA in both nuclear extract (a competitive environment) and in buffer only, which likely means that wild-type RBPMS-A can bind to the substrate RNA in both environments. Wild-type protein was also seen to interact non-specifically with the CCC mutant (Figures 4.11 lane 19 & 4.13 lane 26) when nuclear extract is not present. Given that we have seen RBPMS-A higher order assembly be affected by non-specific RNA, it is not surprising that in buffer alone RBPMS-A is likely interacting or binding in some way to the CCC mutant. The RKR/A mutant had reduced overall crosslinking compared to the wild-type protein but did show a small amount of potential interaction in nuclear extract (Fig.4.13 lane 9) and showed no significant interaction with the mutant transcript (Fig.4.13 lane 27). This may suggest that the RKR residues of the RBPMS-A C-terminus are responsible for non-specific interactions with RNA, which is why overall crosslinking of RKR/A is reduced compared to wild-type even in buffer. These RKR residues interacting with the non-

specific RNA in the mass photometry experiments may be why there is reduced higher order assembly upon addition of RNA. The fact that the RKR/A mutant is fully active despite having lower order oligomers and reduced potential RNA binding (Figures 4.20 & 4.13) suggests that for full activity, the extreme self-assembly observed for wild-type RBPMS-A may not be required and that non-specific RNA interactions are also not needed. This can also be explained by wild-type RBPMS-A not crosslinking to the non-specific RNA in the competitive environment of the nuclear extract. Future experiments would need to be carried out to confirm these observations.

The above discussion has focussed on the homotypic interactions between RBPMS-A subunits, and its interaction with RNA. RBPMS-A, as with many other RBPs, interacts with other proteins, which likely help it perform its function. An in-depth analysis of the importance of heterotypic interactions, which residues are necessary for this, and how this impacts the formation of hetero oligomers was not achieved in this study. Despite this, RBPMS-A was successfully pulled down, and shown to be interacting with other RBPs (Figure 4.23B & 4.24B). Furthermore, Figure 4.24B suggests that the aromatic mutants YYYW/A and WWF/A lost interaction with RBFOX2 and RBM14, whilst wild-type and the basic mutant(s) retained partial ability to interact. The WWW/A and YYY/A mutant proteins were tested against RBPMS-A in a pull-down experiment in nuclear extract by a Part III student in the lab (Lee, Part III dissertation, 2022). This experiment found that these aromatic mutations caused a reduction or loss of interaction with several RBPMS-A interactors (MATRIN3, RBM14, RBM47, RBFOX2, RBM4, and ESRP2). The YYY/A mutant pull-down had reduced amounts of all interactors compared to the wild-type protein except MATRIN3, and the WWW/A mutant lost all interactions (Table 4.1). This further implicates these aromatic residues in forming heterotypic interactions and suggests that the tryptophan residues are more important in these interactions than the tryosines, although no such distinction was possible between the YYYW/A and WWF/A mutant proteins with the limited number of proteins tested in Figure 4.24B. The aromatic mutants therefore may be important in forming $\pi - \pi$ interactions or potentially providing the π side of cation – π interactions for interacting with other proteins and with other RBPMS subunits. The basic residues on the other hand, are likely forming cation – π interactions, which appear to be important for the self-assembly but not activity. Additionally, these residues may be forming non-specific interactions with RNA, but loss of these residues does not fully prevent binding to target RNA or other to other proteins, which allows the RKR/A mutant to be fully active. The TT/A mutant has reduced higher-order assembly and UV crosslinking relative to wild-type but given that it still has full activity, the lost RNA binding and oligomerisation appears to not be necessary for splicing function. The negatively charged mutants RKR/E and TT/E had reduced higher order assembly compared to the wild-type protein, and slightly reduced relative to their alanine substitution counterparts. They still showed some interaction with

the other proteins (RKR/E and RBM14, and TT/E and RBFOX2) but they likely lost their activity due to their lack of crosslinking to RNA in nuclear extract, and even in buffer alone for the RKR/E mutant. This demonstrates that the negative charge introduced in these regions likely prevents proper RNA binding in a competitive environment. This loss may be due to reduced oligomerisation of RBPMS itself or a lack of interaction with key co-regulator proteins, or as a result of reduced RNA binding, which the UV crosslinking showed for RKR/E (Fig.4.12, lanes 8 & 18). Electrophoretic mobility shift assays (EMSAs) were previously performed with the TT/E mutant, and these results confirmed that TT/E does not bind to target RNA as effectively as wild-type or TT/A proteins (Barnhart *et al*, 2022). Therefore, lack of RNA binding is likely to be at least partially attributable to the loss of activity for the TT/E and RKR/E mutant proteins.

4.8 FUTURE PLANS

To further understand the role the residues in the IDR of RBPMS-A are playing, and how this relates to its structure-function relationship, follow up experiments are required:

- Full pull-down and proteomic analysis of RBPMS-A and its mutants to identify if the basic and aromatic residues are also required for heterotypic interactions, as well investigating how phosphorylation of RBPMS may alter PPIs via the TT/E mutant.
- Further analysis of RNA binding abilities of RBPMS-A and its mutants through electrophoretic mobility shift assays
- Mass photometry experiments with the RKR mutants and RNA to observe if oligomerisation of these mutants is influenced by RNA and if it is sequence specific
- Single molecule studies to identify the splicing/protein complexes that RBPMS-A and its mutants form on an RNA target
- Higher order assembly studies of RBPMS-A and its mutants in nuclear extract, eg sedimentation studies $\pm$ nuclear extract, to see if their higher order assemblies can be influenced by interacting (or lack of interaction) with other proteins.
- The Click-chemistry approach (Kolb *et al*, 2001) could be used to detect where along the C-terminus homo-typic and hetero-typic interactions are occurring, thus identifying/confirming the important residues for the higher order assembly of RBPMS

5 EXTREME HIGHER ORDER ASSEMBLY OF RBPMS-B PREVENTS FULL ACTIVITY

5.1 BACKGROUND

Two major isoforms of RBPMS were detected at the RNA level in rat PAC1 cells, RBPMS-A and RBPMS-B (Section 1.53) (Nakagaki-Silva *et al*, 2019). Additionally, western blot data from PAC1 cells detected a protein doublet assumed to be RBPMS-A and RBPMS-B, although this had not been formally proven. RBPMS-A has a unique 20 amino acids C-terminal tail which occurs from the skipping of exon 7, and inclusion of exon 8 (Fig.5.1). RBPMS-B has a unique 43 amino acid C-terminal tail which arises from inclusion of exon 7 which is an alternative 3' end exon (Fig.5.1). Previously, transfected RBPMS-B was found to have lower splicing activity relative to RBPMS-A in HEK293T and PAC1 cells with minigene targets (Nakagaki-Silva *et al*, 2019). RBPMS-B overexpression had virtually no effect on the repression of the NM exons of *Tpm1* or *Actn1*, and lower activity with the activation target *Flnb*, although increased levels of inclusion of the H1 exon relative to basal levels were observed. Given that the two isoforms have shared sequence identity up to residue 176, the unique C-termini of each isoform must cause the differences in activity observed. The unique C-terminal tail of RBPMS-B was found through chimeric studies (Section 1.53) to be antagonising the activity of RBPMS-B. Exactly how the activities of RBPMS-A and RBPMS-B differed across various endogenous or minigene targets, and whether this was due to differences in protein expression or biophysical behaviour caused by these distinct C-terminal tails remained to be investigated.

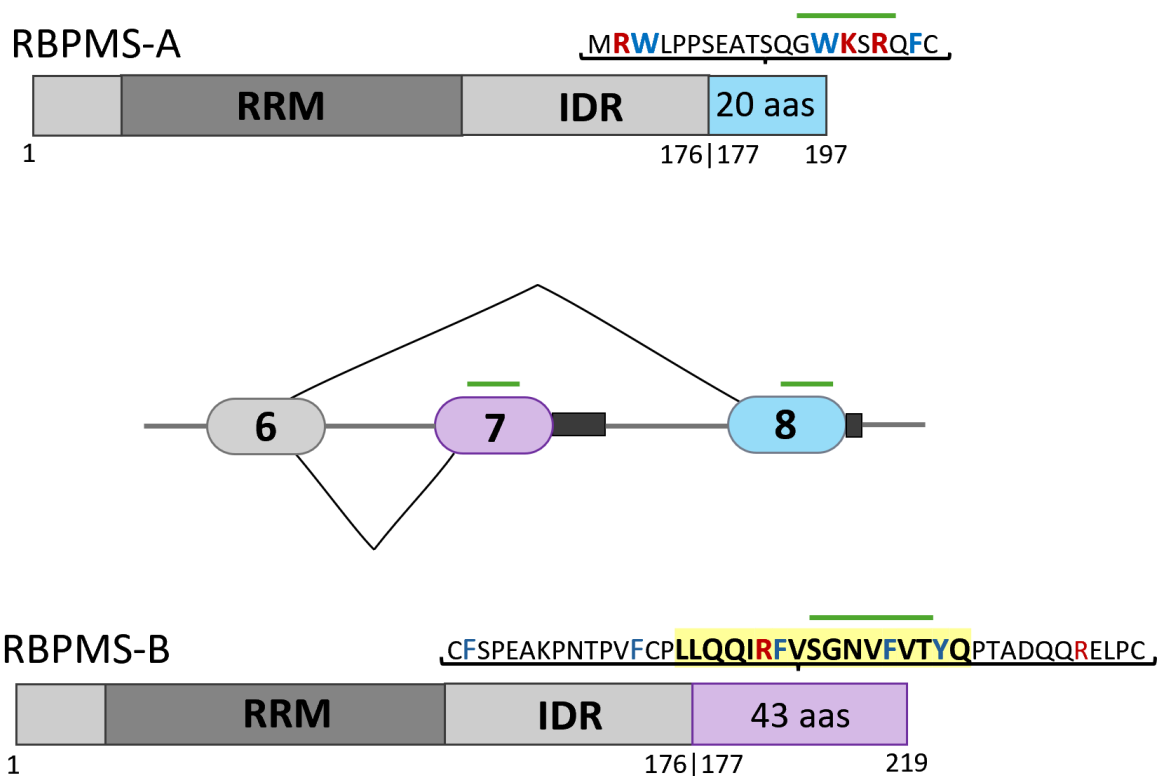


Figure 5.1 RBPMS isoforms are produced by an alternative splicing event: Schematic showing the domain structure of the main RBPMS isoforms found in VSMCs and the splicing event that produces them, with exon 7 being spliced in for RBPMS-B and exon 8 for RBPMS-A. Green lines represent where isoform specific siRNAs were designed to. The Hot Spot for Aggregation (HSA) region of RBPMS-B is highlighted in yellow and is bold. The aromatic residues of each unique C-terminus are in blue, and the basic residues are in red.

5.2 RBPMS ISOFORMS HAVE DIFFERING SPLICING ACTIVITY IN PAC1 CELLS

Pilot knockdown data demonstrates PAC1 cell RBPMS-A has higher activity than RBPMS-B

Previous depletion experiments have knocked down both isoforms at once with *pan* RBPMS small interfering RNA (siRNA), and whilst this led to important discoveries about how RBPMS affects splicing in PAC1 cells, it did not reveal what impact each isoform was having (Nakagaki-Silva *et al*, 2019; Yang *et al*, 2023). To investigate the effect that each isoform had on splicing in PAC1 cells, a pilot individual isoform knockdown experiment was performed (Fig.5.2).

For each isoform, siRNAs were designed to the specific exons for their unique C-terminal tails (Fig.5.1 & Table 2.4). These siRNAs were transfected into differentiated PAC1 cells along with control siRNA which was non-specific (C2 siRNA) and *pan* RBPMS siRNA which was specific to the first exon of RBPMS for knockdown of both isoforms. RNA and protein were extracted after knockdown, and western blot

with a *pan* anti-RBPMS antibody was used to assess protein levels (Fig.5.2A). The western blot data shows that there are similar levels of each protein in the control (C2) cells, and the *pan* RBPMS siRNA decreased levels of both isoforms at the protein level, showing that knockdown of total RBPMS was successful (lanes 1 & 2). RBPMS-A and RBPMS-B siRNA lanes (3 & 4) indicate that the pilot experiment was successful at the protein level as each isoform was able to be specifically depleted. Additionally, the isoform specific siRNA confirms that the doublet for RBPMS previously observed does indeed correspond to RBPMS-A and RBPMS-B. When RBPMS-A siRNA was used, there appears to be increased expression of RBPMS-B, but more repeats are needed to see if this response consistently occurs. A protein sample extracted from proliferative PAC1 cells previously in the laboratory by Clare Gooding was run alongside the other samples to demonstrate the reduced expression of each isoform in the proliferative state. The amount of both isoforms in proliferative cells is much reduced, as expected from previous experiments which showed *Rbpms* expression decreases upon de-differentiation (Nakagaki-Silva *et al*, 2019).

Splicing changes of several targets were observed via RT-PCR with primers to endogenous targets in the PAC1 cells (Fig.5.2B). RBPMS-A knockdown caused a greater reduction in the amount of skipping seen with *Actn1*, as well as reduced inclusion of the regulated exons of *Ptprf* and *Ppfia1* compared with RBPMS-B knockdown (p-values with *Actn1* = 0.001, 0.023 with *Ptprf*, and 0.008 with *Ppfia1*). With *Myocd* and *Ppfibp1*, RBPMS-B knockdown reduced the amount of inclusion of their regulated exons even further than RBPMS-A knockdown (p-values with *Myocd* and *Ppfibp1* = <0.001). In the case of these two splicing events, knockdown of either isoform did not reduce levels of inclusion to the same extent as *pan* RBPMS knockdown. This suggests that both isoforms contribute to the splicing regulation of these targets. This was not observed for *Actn1*, *Ptprf* and *Ppfia1*, where knockdown of RBPMS-A alone reduced skipping or inclusion (respectively) to the same level as full RBPMS knockdown. This demonstrates that for these exons, RBPMS-A is the main regulator, although knockdown of RBPMS-B still had an effect.

Overall, from the targets investigated, it appears that both isoforms of RBPMS regulate alternative splicing events in PAC1 cells and that different targets have different susceptibilities to each isoform. A greater number of targets would need to be investigated, as well as more experimental replicates, to observe if there is a trend in which targets respond more to knockdown of which isoform.

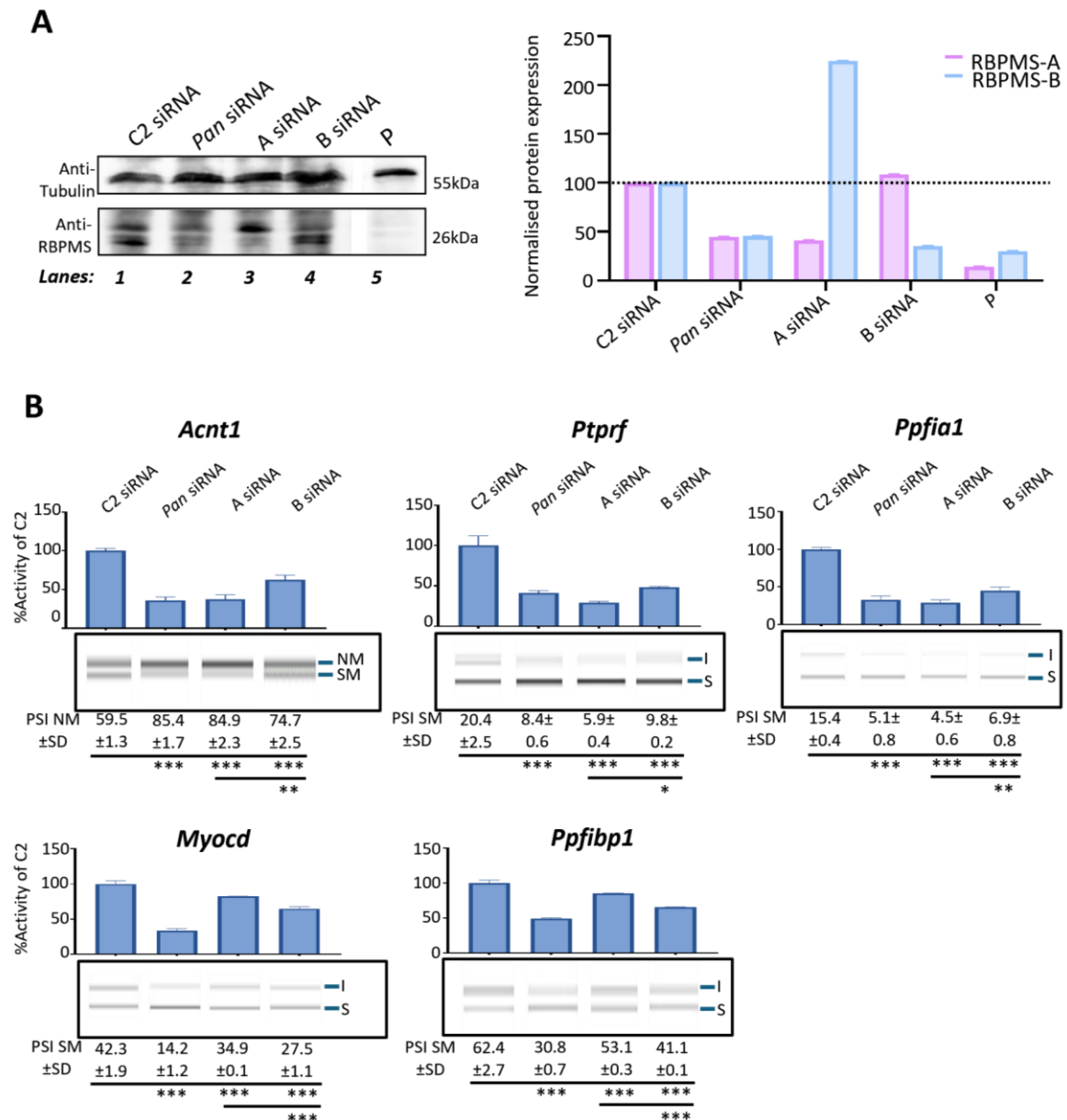


Figure 5.2 Individual RBPMS isoforms can be knocked down in pilot experiment: **A:** Western blot results showing protein expression in each knockdown sample and quantitation of each RBPMS isoform, normalized to Tubulin (loading control) and to levels in control PAC1 (C2) cells. Samples from untreated proliferative (P) cells were included in the western blot to show expression in the de-differentiated state. **B:** Qiaxcel results showing the splicing changes that occur with each siRNA type used with various endogenous targets. Statistical test was a one-way ANOVA followed by Tukey's multiple comparison test, comparing activity of RBPMS in C2 siRNA sample with other samples, (top row) or activity of RBPMS in RBPMS-A siRNA and RBPMS-B siRNA samples. * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test was used to check variance, QQ plot for normal distribution. PSI = percentage spliced in, SM = smooth muscle, NM = non-muscle, I = inclusion, S = skipped.

5.3 ACTIVITY OF RBPMS-B CAN BE INCREASED VIA MUTATION OF ITS HSA REGION

Aggregation reducing deletion mutation increases splicing activity of RBPMS-B

Previously, the C-terminus of RBPMS-B was investigated for its potential to form aggregates (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019). Using an online tool called AGGRESCAN, a seventeen amino acid sequence within the C-terminus was identified as being a “Hot Spot for Aggregation” (HSA) (Conchillo-Solé *et al*, 2007). Upon deletion of this HSA region, RBPMS-B was now able to repress exon 3 of the *Tpm1* minigene to the same extent as RBPMS-A. This suggested that this region of the C-terminus is antagonising the activity of RBPMS-B and that the rest of the C-terminus presumably confers similar activity to the A isoform. Whether this region is indeed aggregation prone and if this is linked to the diminished splicing activity of RBPMS-B remained to be elucidated.

In the present study, AGGRESCAN was again used to further test the C-terminus of RBPMS-B to identify if there were any smaller deletions to the HSA region that could be made which affected its aggregation propensity score. Deletion of two glutamines (Q194 & Q195, Δ QQ) at the start of the sequence was found to lead to a higher predicted aggregation propensity than wild-type, and deletion of the amino acids at the end of the region (V203 – Y207, Δ VFVTY), was predicted to have a reduced aggregation propensity (Figure 5.3A & B). In addition to the smaller HSA deletion mutants, a mutant with two cysteine residues of the C-terminus of RBPMS-B mutated to serine (C177/C190S, CC/S) was also included. This mutant was produced to see if it would be possible to add a single fluorophore at the C-terminal C219 of RBPMS-B, as had been done for RBPMS-A. In the case of RBPMS-A, mutating the single cysteine at the end of RBPMS-A had no effect on activity (Fig.4.14A) and thus a fluorophore was able to be added. It was hoped that two of the cysteine residues in RBPMS-B could be substituted without affecting activity. This would allow the other cysteine to be used for addition of a single fluorophore for future experiments that may require fluorescently labelled proteins (such as phase separation microscopy experiments). AGGRESCAN, however, predicted that these substitutions decreased aggregation propensity, thus this mutant was included in this study (Figure 5.3A & B). Subsequently, these deletion mutations and the cysteine substitution mutant were cloned into transfection constructs and tested in co-transfection assays in HEK293T with minigene and endogenous targets alongside wild-type RBPMS-A and RBPMS-B for comparison (Figures 5.3 & 5.4).

All mutant proteins expressed to similar levels (Fig.5.3C, lower panel) allowing direct comparison of their activities. With endogenous *FLNB*, all the mutants except the glutamine mutant (Δ QQ) and the cystine mutant (CC/S) showed similar levels of increased inclusion of exon H1 compared to wild-type RBPMS-B (p-value for the Δ C20 = 0.873, CC/s = 0.005, Δ VFVTY = 0.114, and Δ QQ = 0.006). Wild-type

BPMS-A and BPMS-B also had similar activity here, suggesting that for splicing activation, the BPMS-B C-terminus is not antagonistic (p-value = >0.999). The cysteine and Δ VFVTY mutants had significantly higher activities than either wild-type isoform, whilst the Δ QQ mutant had reduced activity relative to both isoforms. With the repression target *Actn1*, wild-type BPMS-B overexpression did not lead to any increased levels of NM skipping (p-value relative to mock = 0.995), but wild-type BPMS-A overexpression did (p-value relative to mock = <0.001). The Δ QQ mutant also had no activity (p-value relative to mock = >0.999), but the Δ VFVTY deletion mutant and the CC/S mutant both had increased activity compared to wild-type BPMS-B (p-values = <0.001 and 0.044 respectively, relative to wild-type BPMS-B). The level of repression for the Δ VFVTY mutant was now similar to BPMS-A. The AGGRESCAN aggregation propensity prediction results in combination with the experimental data suggest a link between aggregation propensity and activity. The pro-aggregation mutant Δ QQ has reduced activity compared to wild-type BPMS-B for both activation and repression, whereas the mutants with reduced predicted aggregation propensity (Δ VFVTY and CC/S) have higher activity.

A

177 192 208 219
CFSPEAKPNTVPV**C**LL**QQ**IRFVSGN**V**FV**TY**QPTADQQRELPC

B

Full length protein scores	Awt	Bwt	ΔHSA	ΔVFVTY	ΔQQ	C177S+C190S (CC/S)
Normalized aa aggregation propensity Sequence Sum for 100 residues:	-6.9	-4.4	-8.2	-7.3	-3.3	-5.1

C

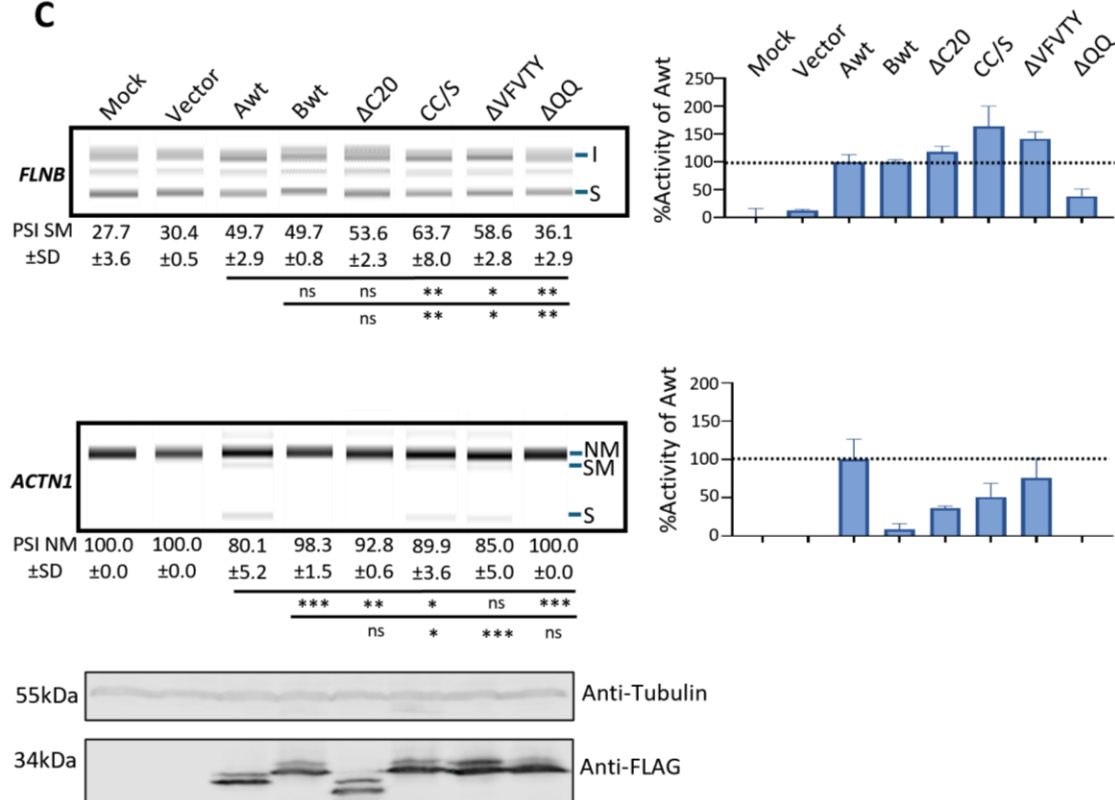


Figure 5.3 Deletion of aggregation prone residues causes an increase in splicing activity for RBPMS-B with endogenous targets: **A:** Final 43 amino acids of RBPMS-B with residues that have been mutated in bold. The HSA region is highlighted, and residues that have been deleted have been underlined. **B:** Table of mutants from RBPMS-B designed to be more/less aggregation prone than the wild-type protein. Values for aggregation propensities closer to 0 are predicted to be more likely to aggregate. Values from AGGRESCAN tool. **C:** Qiagel results from the transfection of deletion mutants of RBPMS-B HSA region and the effect they have on splicing activation (FLNB) and (ACTN1). Statistical test was a one-way ANOVA followed by Tukey's multiple comparison test, comparing activity of each isoform to the mutants (top row = RBPMS-A, bottom row = RBPMS-B), * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test used to check variance, QQ plot for normal distribution. PSI = percentage spliced in, SM = smooth muscle, NM = non-muscle, I = inclusion, S = skipped, Awt = RBPMS-A wild-type, Bwt = RBPMS-B wild-type, aa = amino acid.

The mutants in this co-transfection assay were also tested with the activation minigene targets produced in chapter 3 (Figure 5.4). With *FLNB*, both RBPMS-A and RBPMS-B overexpression led to full inclusion of exon H1, and all the mutants were also fully active (p-value relative to vector control = <0.001 for all proteins). This was true even for the Δ QQ mutant, which previously had lower activity with the endogenous target. Activation of *RPMS* exon 3a was reduced by the QQ mutation (p-value vs wild-type RBPMS-B = 0.005), but the Δ VFVTY and Cysteine mutant behaved similarly to wild-type RBPMS-B (p-values = 0.942 and >0.999, respectively). The Δ C20 mutant had lower activity than wild-type RBPMS-A but showed no difference with RBPMS-B (p-values = 0.010 and 0.517 respectively), and neither isoform showed a significant difference in activity (p-value = 0.270). With *PPFIBP1*, as shown previously (Fig.3.3), RBPMS-B overexpression led to higher levels of inclusion of exon 10 than with RBPMS-A (A vs B p-value = <0.001). All mutants had reduced levels of inclusion compared to wild-type RBPMS-B except the CC/S mutant which had slightly higher activity (p-values = <0.001 for all mutants except CC/S = 0.041 and Δ QQ = 0.121). *PTPRF* was also tested but this minigene did not respond to any of the mutants or to RBPMS-B wild-type (data not shown). Overall, it appears that for most targets, the predicted aggregation propensity of RBPMS-B and splicing activity appear to be linked, as decreasing this predicted propensity via mutations (Δ VFVTY and CC/S) causes an increase in splicing activity compared to wild-type or the Δ QQ proteins which had higher predicted aggregation propensity.

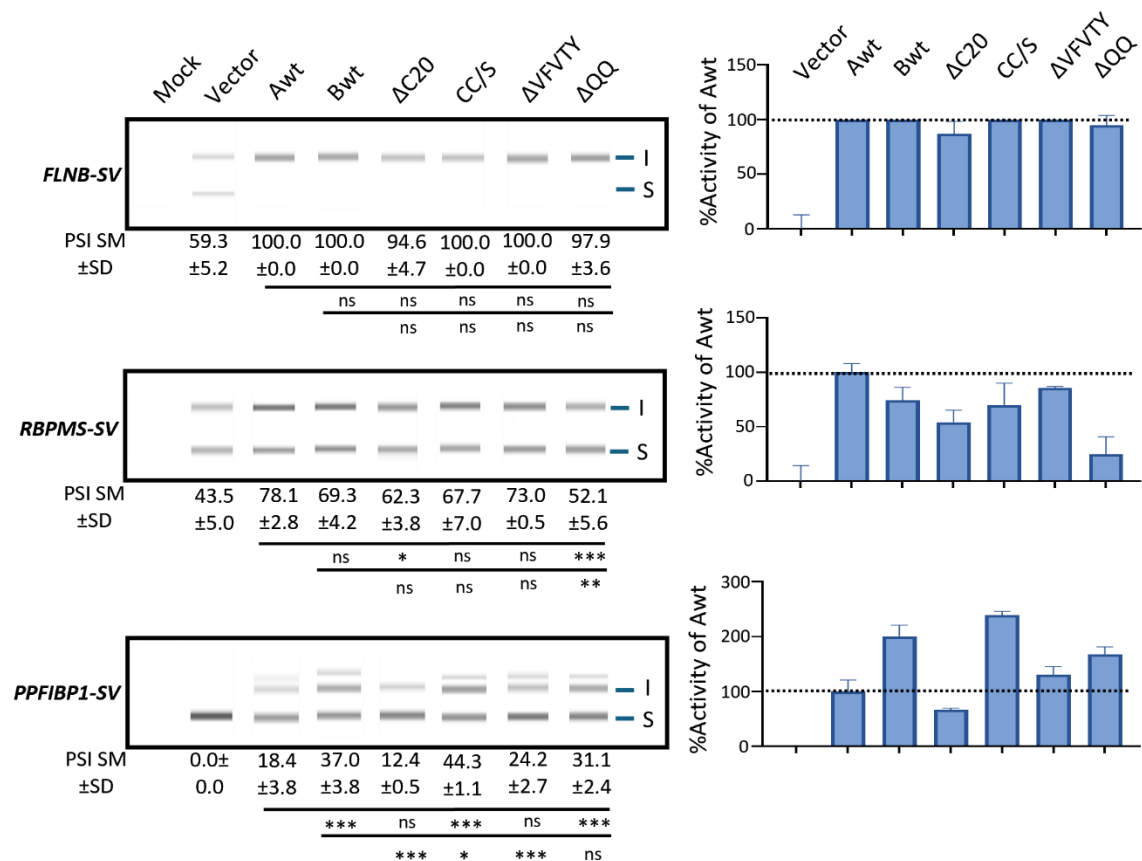


Figure 5.4 Mutants of RBPMS-B have differential effects with the minigene activation targets: Qiaxcel results showing the results from the co-transfection of deletion mutations of the RBPMS-B HSA region and the effect they have on splicing activation with minigene constructs (SV signifies the backbone vector of the minigenes). Statistical test was a one-way ANOVA followed by Tukey's multiple comparison test, comparing activity of each isoform to the mutants (top row = RBPMS-A, bottom row = RBPMS-B), * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test used to check variance, QQ plot for normal distribution. PSI = percentage spliced in, I = inclusion, S = skipped, Awt = RBPMS-A wild-type, Bwt = RBPMS-B wild-type.

Phosphorylation of RBPMS-B confirmed via phosphatase treatment

The double band seen for RBPMS-A in the western blot in Figure 5.3 has previously been shown to be due to phosphorylation of serine 144, and substitution of this residue to either alanine (phospho-null) or glutamate (phosphomimetic) had no effect on splicing activity (Barnhart *et al*, 2022). The S144E mutant did, however, show reduced mobility on SDS-PAGE as the phosphorylated S144 wild-type RBPMS-A does (Barnhart *et al*, 2022). The double bands for the RBPMS-B protein and its mutants are likely also from phosphorylation, although this had not been verified. The upper bands for RBPMS-A, RBPMS-B and the cysteine mutant are a similar intensity, but the ΔC20, ΔVFVTY and ΔQQ mutants have different intensities (Figure 5.5A, anti-FLAG panel, lanes 7 & 8). The less aggregation prone ΔVFVTY deletion mutant has a higher intensity upper band, whereas the double glutamine deletion (ΔQQ)

mutant has a lower intensity band. Although S144 phosphorylation has been shown to not affect activity of RBPMS, meaning it is unlikely that there is a direct correlation between activity and the intensity of this band, it may be linked to aggregation propensity. The predicted aggregation prone deletion of the two glutamine residues in RBPMS-B may make S144 less accessible to kinases, or their binding sites, due to its aggregation behaviour, whereas the Δ VFVTY phosphorylation sites may become more accessible. To confirm whether the upper bands seen for RBPMS-B and its mutants were due to phosphorylation, a recombinant phosphatase was added to the lysates of these protein samples and a western blot was performed (Fig.5.5B). In the +phosphatase lanes, the upper band for RBPMS (anti-FLAG panel) disappears. This result demonstrates that the upper band seen for RBPMS-B and its mutants, as well as for Δ C20 and RBPMS-A, is due to phosphorylation. Furthermore, this result links higher potential levels of oligomerisation/aggregation and accessibility of S144 to kinases, meaning the phosphor band of RBPMS-B may be a potential readout for oligomerisation state.

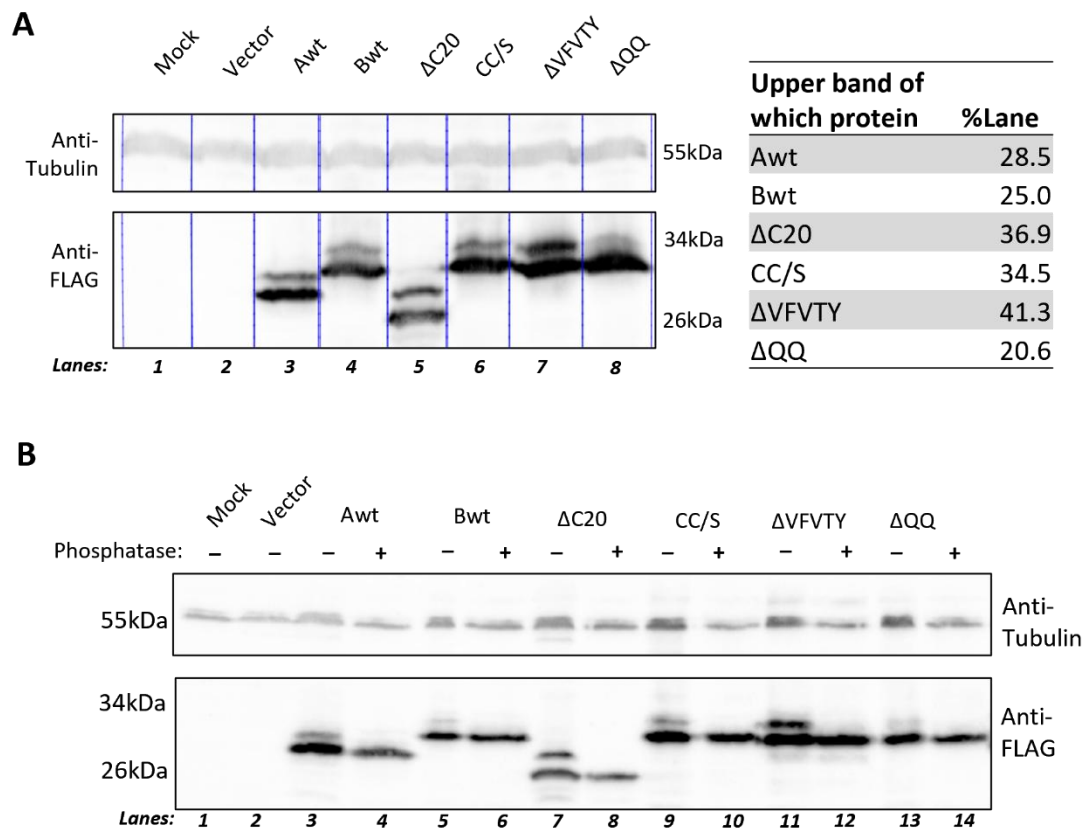


Figure 5.5 Phosphatase study confirms upper band in RBPMS western blot is due to phosphorylation:
A: Western blot image with Tubulin displayed as a loading control and bands for RBPMS highlighted with pink lines. Suspected phosphorylated RBPMS-B bands have been quantitated relative to the upper band on RBPMS-B wild-type (Bwt) using the Chemidoc software. **B:** Western blot image showing the effect of phosphatase treatment on the bands seen for RBPMS (anti-FLAG).

Aggregation breaking mutations increase activity of RBPMS-B to the level of RBPMS-A

The C-termini of both RBPMS-A and RBPMS-B after the RRM domain are predicted to be unstructured. Recent advances with AlphaFold have allowed for more accurate predictions of protein structure (Varadi et al, 2022, 2024; Abramson et al, 2024). RBPMS-B was modelled with AlphaFold 3, and the C-terminal tail was mainly unstructured except for the HSA region. In both the full-length predicted structure and with the HSA peptide alone, this region was predicted to form antiparallel β strands (Figure 5.6). The probability score of this secondary structure prediction was low for the full-length protein, suggesting low confidence, but nonetheless showed that this region may once again be behaving differently to the rest of the C-terminus. Based on the finding that the HSA region may form anti-parallel β sheet structures, two residues within this region were mutated to proline (I196/V205P, P2 & I196/V203/V205P, P3). Prolines are reported to be “breakers” of aggregation, in particular for amyloid fibrils formed from anti-parallel β sheets, due to their restricted rotational movement about the N – C α bond (Conchillo-Solé *et al*, 2007). The AlphaFold models (Fig.5.6, lower panel) of these mutants show that the β sheets have been partially (P2) or fully (P3) disrupted. AGGRESCAN predicted that these mutants all had much lower aggregation propensity scores than wild-type protein (Figure 5.7A). In addition to this mutant, AGGRESCAN was used to predict *in silico* which residues would affect the aggregation propensity if substituted to alanine, and nine residues were identified as being “pro” aggregation and were subsequently mutated to alanine in the mutant HSA/A (L192/L193/I196/F198/V199/V203/F204/V205/Y207A).

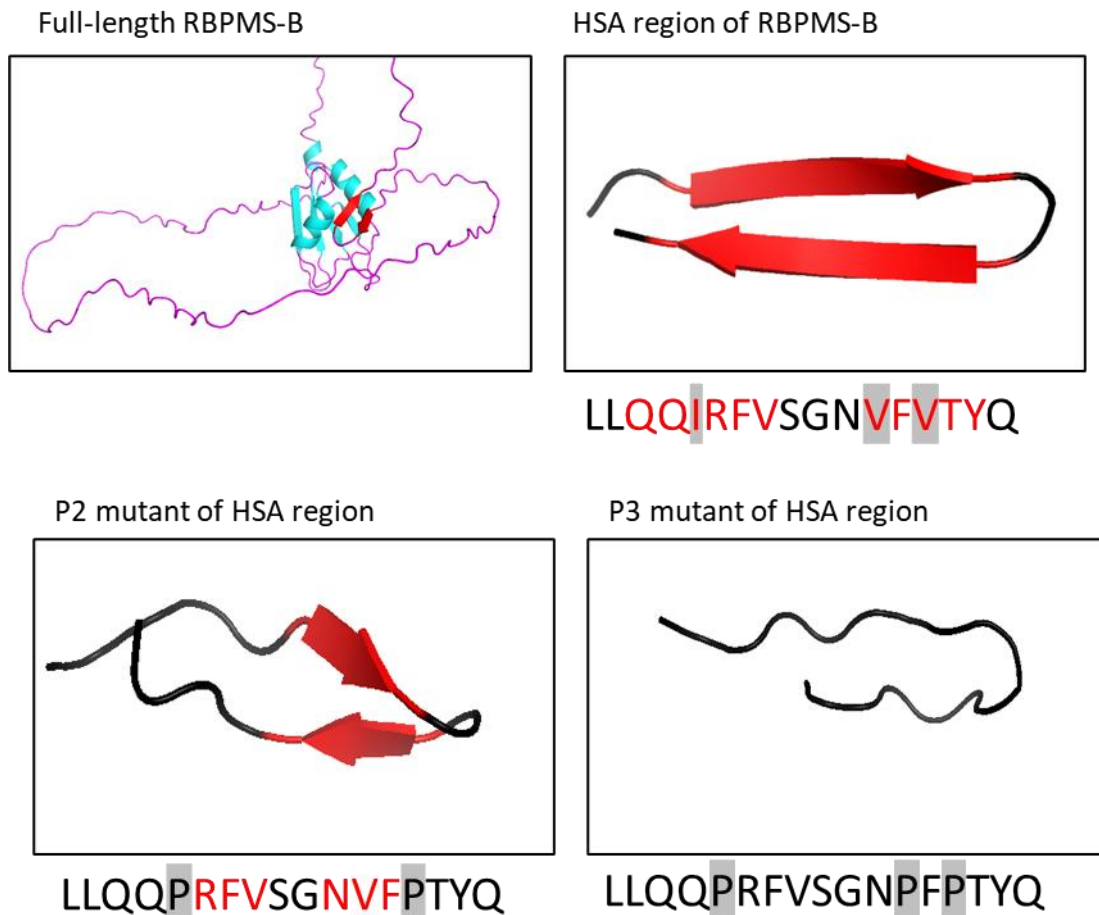


Figure 5.6 AlphaFold predicts that the HSA region of RBPMS-B forms antiparallel beta sheets: AlphaFold models of full-length RBPMS-B (upper left) and the HSA region by itself (upper right), and the two proline mutants P2 (bottom left) and P3 (lower left). The RRM is shown in cyan, the β sheet forming section of the HSA region is shown in red in all images, the rest of the C-terminus is shown in pink. HSA sequence is shown beneath with residues in red that are forming the antiparallel β sheets. Grey highlight shows residues substituted to proline in the P2 mutant. Images taken with Pymol.

To see whether this is correlated with a difference in activity, these mutants were co-transfected into HEK293T cells with *FLNB* and *Actn1* minigene reporters (Figure 5.7B). With *FLNB*, both RBPMS isoforms have high levels of inclusion of exon H1 relative to the mock control and showed similar activity to each other. Overexpression of RBPMS-A and RBPMS-B led to almost 100% inclusion of exon H1, so it was unlikely that an increase in activity caused by the HSA substitutions would be observable with this minigene. Surprisingly, both the HSA/A mutant with a decreased predicted aggregation score and the Δ QQ mutant with an increased score, showed reduced activity relative to wild-type RBPMS-B (p-values = <0.001 for HSA/A mutant and 0.023 for the Δ QQ). With the repression target *Actn1*, the effect of the mutations was more pronounced. Here, the Δ C20 mutant is inactive, whereas RBPMS-B has some residual activity relative to basal levels of repression of the NM exon (p-values relative to vector control = >0.999 for Δ C20 and <0.001 for Bwt). As before, the Δ QQ deletion mutant did not have any activity,

as expected from its increased aggregation score (p-value = 0.976 relative to vector control). The HSA/A mutant protein again showed no repression activity, despite the fact it was designed to decrease aggregation (p-value = >0.999). This suggests that some of the residues mutated in this region may be important for the splicing activity of RBPMS-B. The two proline mutants were both significantly more active than wild-type RBPMS-B (p-values relative to Bwt = <0.001 for both proline mutants) and the P2 mutant protein was as active as RBPMS-A. The P2 mutant was also more active than the Δ VFVITY deletion mutant protein (p-value = <0.001 for P2 vs Δ VFVITY). These results suggest that the predicted aggregation behaviour may be linked to the lower splicing activity of RBPMS-B, but that some residues of the HSA region may still be important for activity. The RBPMS-B P2 mutant is likely to be the ideal mutant to explore the effect of the HSA on activity and aggregation as it increased activity relative to wild-type but had fewer substitutions made. Additionally, its predicted aggregation propensity was reduced relative to wild-type RBPMS-B but was close to RBPMS-A (Fig.5.7A).

A

Full length protein scores	Awt	Bwt	Δ HSA	Δ VFVTY	Δ QQ	P2	P3	HSA/A
Normalized aa aggregation propensity Sequence Sum for 100 residues:	-6.9	-4.4	-8.2	-7.3	-3.3	-6.3	-7.2	-11.0

B

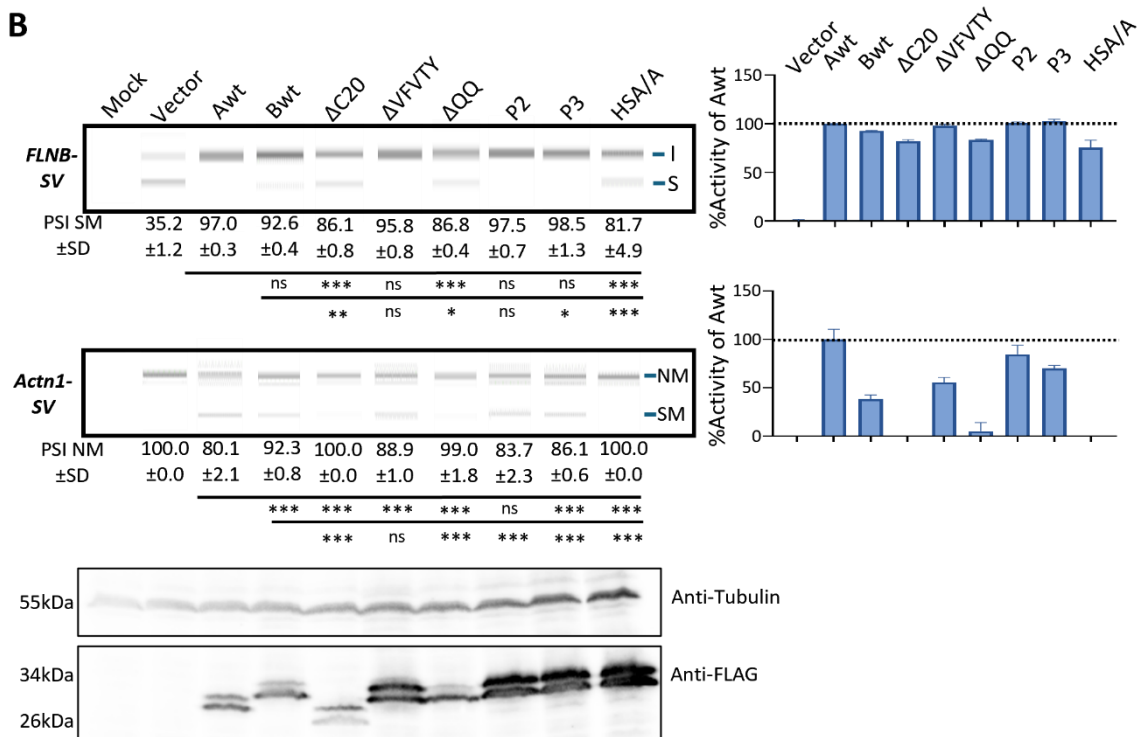


Figure 5.7 Proline substitutions in the HSA region enhance the splicing activity of RBPMS-B: **A:** AGGRESCAN scores showing aggregation propensity. Values for aggregation propensities closer to 0 are predicted to be more aggregation prone. **B:** Qiaxcel results from the co-transfection of RBPMS-B HSA mutants and the effect they have on splicing activation and repression with minigene constructs (SV signifies the backbone vector of the minigenes). Statistical test was a one-way ANOVA followed by Tukey's multiple comparison test, comparing activity of each isoform to the mutants (top row = RBPMS-A, bottom row = RBPMS-B), * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test used to check variance, QQ plot for normal distribution. PSI = percentage spliced in, I = inclusion, S = skipped, Awt = RBPMS-A wild-type, Bwt = RBPMS-B wild-type.

5.4 AROMATIC RESIDUES OF RBPMS-B C-TERMINUS ARE IMPORTANT FOR SPLICING

FUNCTION

Transfection studies of aromatic substitution mutants show RBPMS-B requires aromatic residues for its splicing activity

In chapter 4, it was discovered that there were two clusters of aromatic residues important for the activity of RBPMS-A (Figures 4.3 – 4.5, & 4.8). These clusters were the YYYW residues from 141 – 154 before the unique C-terminal tail, and the WWF residues (from 178 – 195) within the extreme RBPMS-A C-tail. The RBPMS-B unique C-terminal tail has five aromatic residues, four phenylalanines and one tyrosine. Due to the necessity of the aromatic residues for splicing activation and repression by RBPMS-A, a Part II undergraduate student under my supervision, Georgina Withers, created a series of aromatic mutants to test the importance of these residues for splicing regulation by RBPMS-B (Withers, Part II Dissertation, 2023). The four phenylalanines were mutated to produce the FFFF/A mutant (F178/189/198/204A, FFFF/A) and the Y207 residue was also included to make the mutant FFFFY/A (Fig.5.8). In addition, the YYY and YYYW residues were also mutated (Y141/151/154A and Y141/151/154/W148A). Since the proline mutant (P2) had more activity than RBPMS-B wild-type, in particular with repression targets, the previous mutations were also made in the context of the P2 mutant. This was to assess how these residues may be aiding splicing repression as RBPMS-B wild-type itself has very little activity, meaning decreases in activity beyond this would not be observable.

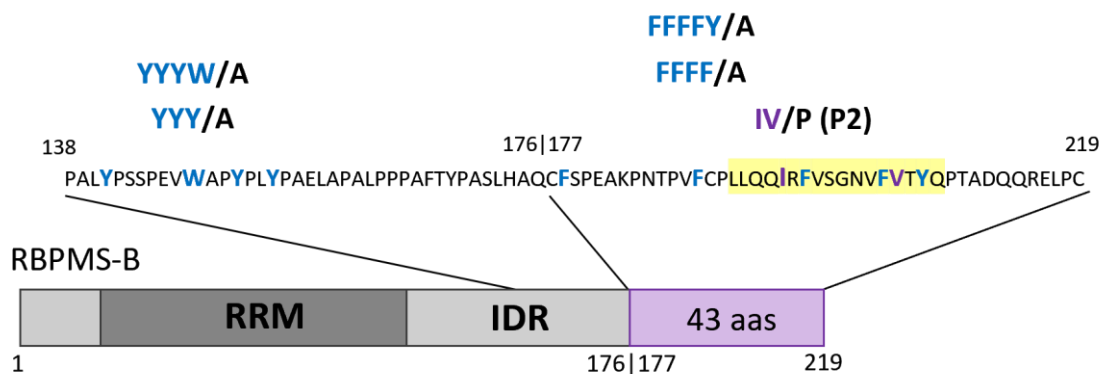


Figure 5.8 Aromatic residues in RBPMS-B C-terminus were mutated to alanine: Schematic of domain structure of RBPMS-B with residues that have been mutated shown in different colours (aromatics in blue, sites of proline substitution in purple, and the HSA region has been highlighted in yellow).

Transfection studies with these mutant proteins were performed by Georgina Withers in HEK293T cells, and the splicing outcomes of endogenous RBPMS targets were observed via RT-PCR (Fig.5.9B). Each protein had similar expression, allowing fair comparison of the different activities (Fig.5.9A). With the activation target *PPFIBP1* (Fig.5.9B, upper panel), wild-type RBPMS-B activates exon inclusion to a similar extent as RBPMS-A, and for P2, inclusion was significantly higher than RBPMS-A but not RBPMS-B (lanes 3 – 5, p-value of P2 vs RBPMS-A = <0.001 and vs wild-type RBPMS-B = 0.260). The YYY mutants in the context of wild-type and P2 proteins, both had reduced activity, as did the YYYW mutants (lanes 6 – 9, p-values = <0.001 for all compared to wild-type RBPMS-B). This suggests that these residues are

required for splicing activation for both RBPMS isoforms. The Bwt-FFFF mutant retained most of its ability to activate exon inclusion, with similar activity to RBPMS-A but slightly reduced activity relative to wild-type RBPMS-B and the P2 mutant (lane 10, p-values = 0.002 relative to wild-type RBPMS-B and <0.001 relative to P2). This suggests that these phenylalanine residues are not important for splicing activation by RBPMS-B. The P2-FFFF mutant, however, lost activity completely, which was surprising given that its wild-type context counterpart had increased activity (p-value = <0.001 relative to wild-type RBPMS-B and P2). Two of the phenylalanine residues are within the HSA region. The HSA/A mutant was previously shown to lose activity also, in which the two phenylalanines were also mutated (Fig.5.7). This result again suggests that some of the residues within the HSA region are required for activity. The FFFFY mutants lost activity, suggesting that this residue Y207, is important for splicing activation (Fig.5.9B, lanes 12 & 13, p-values = <0.001 for both FFFFY mutants relative to wild-type RBPMS-B). With the repression target *ACTN1* (Fig.5.9B, middle panel), wild-type RBPMS-B had no activity and neither did the aromatic mutants with the wild-type HSA region (p-value = >0.999 for all proteins except the RBPMS-A and P2 proteins where the p-values were <0.001). The P2 mutant had increased activity relative to wild-type RBPMS-B, but none of the mutants in the P2 context had activity. This would suggest that the YYY/W and FFFF/Y residues are important for splicing repression, as these mutants can no longer repress the NM exon. Wild-type RBPMS-B had some residual function with the repression target *TPM1* (Fig.5.9B, lower panel) but it was still reduced relative to RBPMS-A and the P2 mutant (p-values of wild-type RBPMS-B relative wild-type RBPMS-A or the P2 mutant = <0.001). Similarly to *ACTN1*, most of the aromatic mutants have lost activity relative to wild-type and P2, suggesting that these aromatic residues are important for splicing repression (p-values of all the aromatic mutants relative to mock = >0.999 except the Bwt-FFFF mutant which was <0.001). The Bwt-FFFF mutant has some residual activity, similar to wild-type RBPMS-B, once again demonstrating that these residues are not important for activity, but if the Y207 residue is also mutated this activity is lost.

Overall, it would appear the aromatic residues (Y141/151/153 and W148A) before the unique C-terminal tail of RBPMS-A and RBPMS-B are important for the splicing activity of both isoforms. The phenylalanine residues in the unique C-terminal tail of RBPMS-B do not appear to be necessary for splicing activation or repression in the wild-type context but may still be important when the proline substitutions have been introduced. The subsequent mutation of the Y207 residue in RBPMS-B, however, does abrogate activity with the targets observed here in both the wild-type and P2 context. Whether this reduction in activity is observed if this residue is mutated alone or requires the phenylalanine residues to be mutated too remains to be investigated

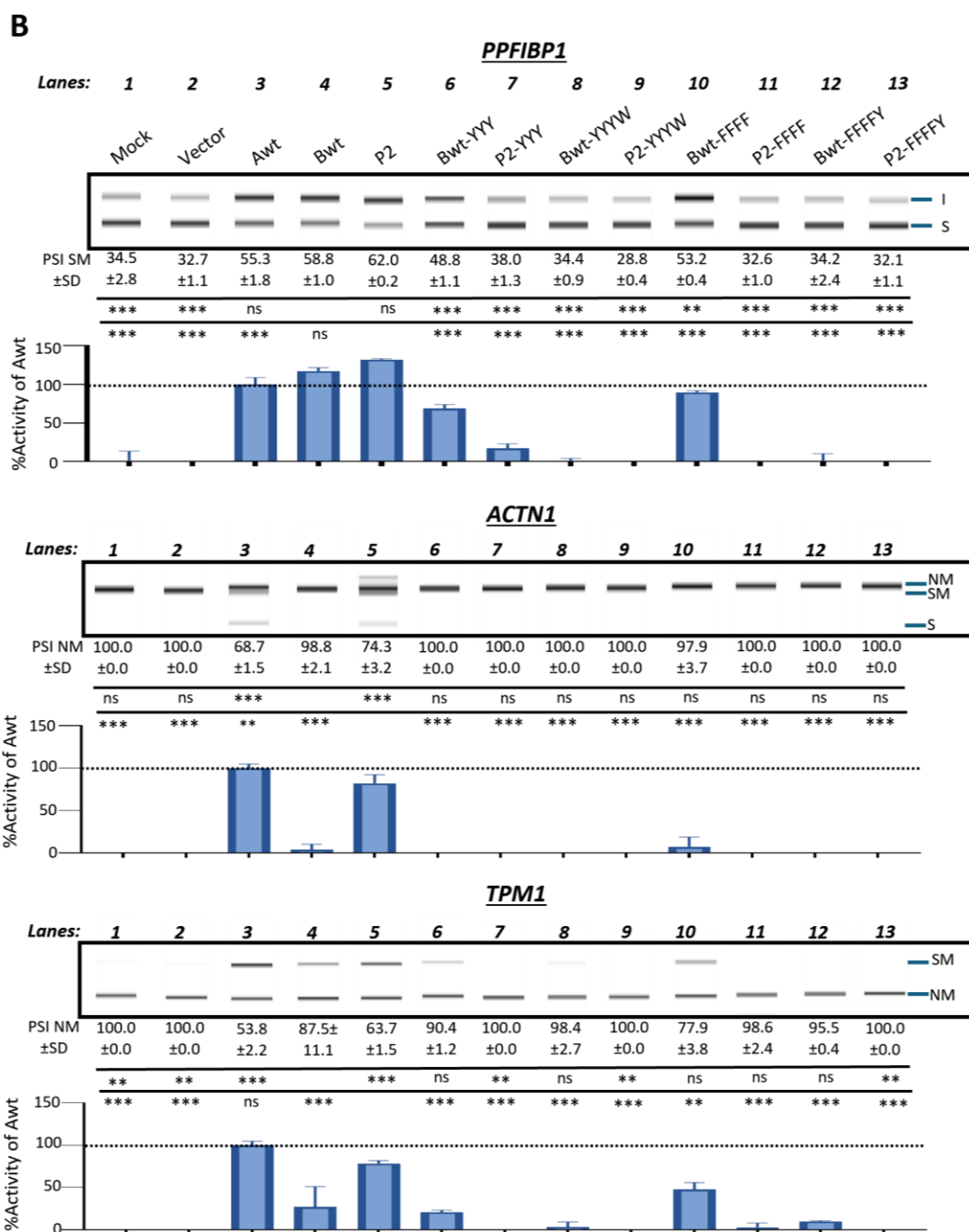
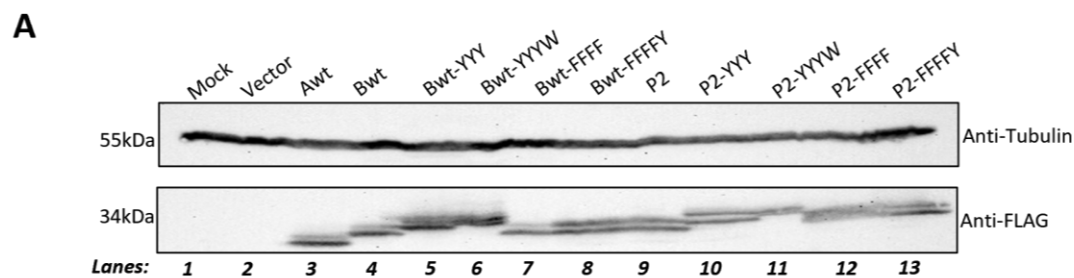


Figure legend on next page.

Figure 5.9 Aromatic residues aid splicing activity of RBPMS-B: **A:** Western blot results from the substitution mutant transfection into HEK293T cells. **B:** Qiaxcel RT-PCR data results from the substitution mutant transfection into HEK293T cells, with endogenous targets PPFIBP1, ACTN1 and TPM1 used to measure differences in splicing activity. %Activity refers to the percentage activity of RBPMS-A. Statistical test was a one-way ANOVA followed by a Dunnett's multiple comparison test, comparing activity of wild-type to the mutants, * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$. Brown-Forsythe test for equal variance and QQ plot for normality. PSI = percentage spliced in, SM = smooth muscle, NM = non-muscle, S = skipping, I = inclusion, Awt = RBPMS-A wild-type, Bwt = RBPMS-B wild-type, P2 = proline mutant.

Protein	Splicing activity	Aggregation propensity prediction
RBPMS-A	++++	+
RBPMS-B	++	+++
RBPMS-B CC/S mutant	++++	++
RBPMS-B Δ VFVTV	++++	+
RBPMS-B Δ QQ	+	++++
RBPMS-B P2	++++	+
RBPMS-B P3	++++	+
RBPMS-B YYY/A	-	++
RBPMS-B P2 YYY/A	-	+
RBPMS-B YYYW/A	-	++
RBPMS-B P2 YYYW/A	-	+
RBPMS-B FFFF/A	++	+
RBPMS-B P2 FFFF/A	-	+
RBPMS-B FFFFY/A	-	+
RBPMS-B P2 FFFFY/A	-	+

Table 5.1 Splicing activity and predicted aggregation propensity of RBPMS-B and its mutants. Predicted aggregation propensity from the AGGRESCAN tool.

5.5 THE HSA PEPTIDE HAS FIBRIL-LIKE PROPERTIES

Biophysical studies of HSA peptide reveal it forms large stable structures

Further investigation into the structure of RBPMS-B and its HSA region was required to investigate its relationship to activity. Purification attempts of the full-length protein proved difficult as the protein

was mainly insoluble, therefore studies were initially performed with a synthetic peptide of the HSA region (residues 192 – 208, Fig.5.3A) to investigate the structure and biophysical properties of RBPMS-B. To assess whether the AlphaFold prediction about the secondary structure of this protein was correct, Circular Dichroism (CD) experiments were performed (Fig.5.10). For this experiment, the peptide was dialysed into a buffer containing 32.5% Acetonitrile, 10 mM phosphate buffer and 50 mM NaF. The acetonitrile was used to keep the peptide soluble as when it was not added the peptide started to precipitate. Sodium fluoride was used as an alternative to NaCl or KCl, as chloride ions absorb light at the wavelengths used for CD ($\lambda = 185 - 250$ nm) (Greenfield, 2007). CD spectra were taken at 25°C, then the peptide was heated to 90°C (“melted”) and CD measured again, and spectra were observed a third time post-melt, at 25°C again. The CD spectra were analysed with the CDNN software and are displayed with molar ellipticity, which corrects for concentration and refers to the absorption of light by the protein relative to its molecular weight and concentration (Fig.5.10). The spectra at each temperature look very similar, with the CDNN software predicting mainly a mix of antiparallel beta sheet structure and random coil (largest percentages being: 35% antiparallel β sheet, 25% random coil, 17% β turn). This suggests that the peptide is not completely antiparallel β sheet in this context, but nevertheless it is still partially forming this secondary structure type. The most surprising aspect of the CD spectra shown here is the consistency at the different temperatures the peptide was subjected to. The HSA peptide retains the secondary structure even after heating to 90°C, which is normally used to denature or “melt” a protein/peptide meaning it would become 100% random coil (Tiffany & Krimm, 1972; Greenfield, 2007). Being able to retain its secondary structure at these high temperatures demonstrates that the HSA peptide structures are extremely stable.

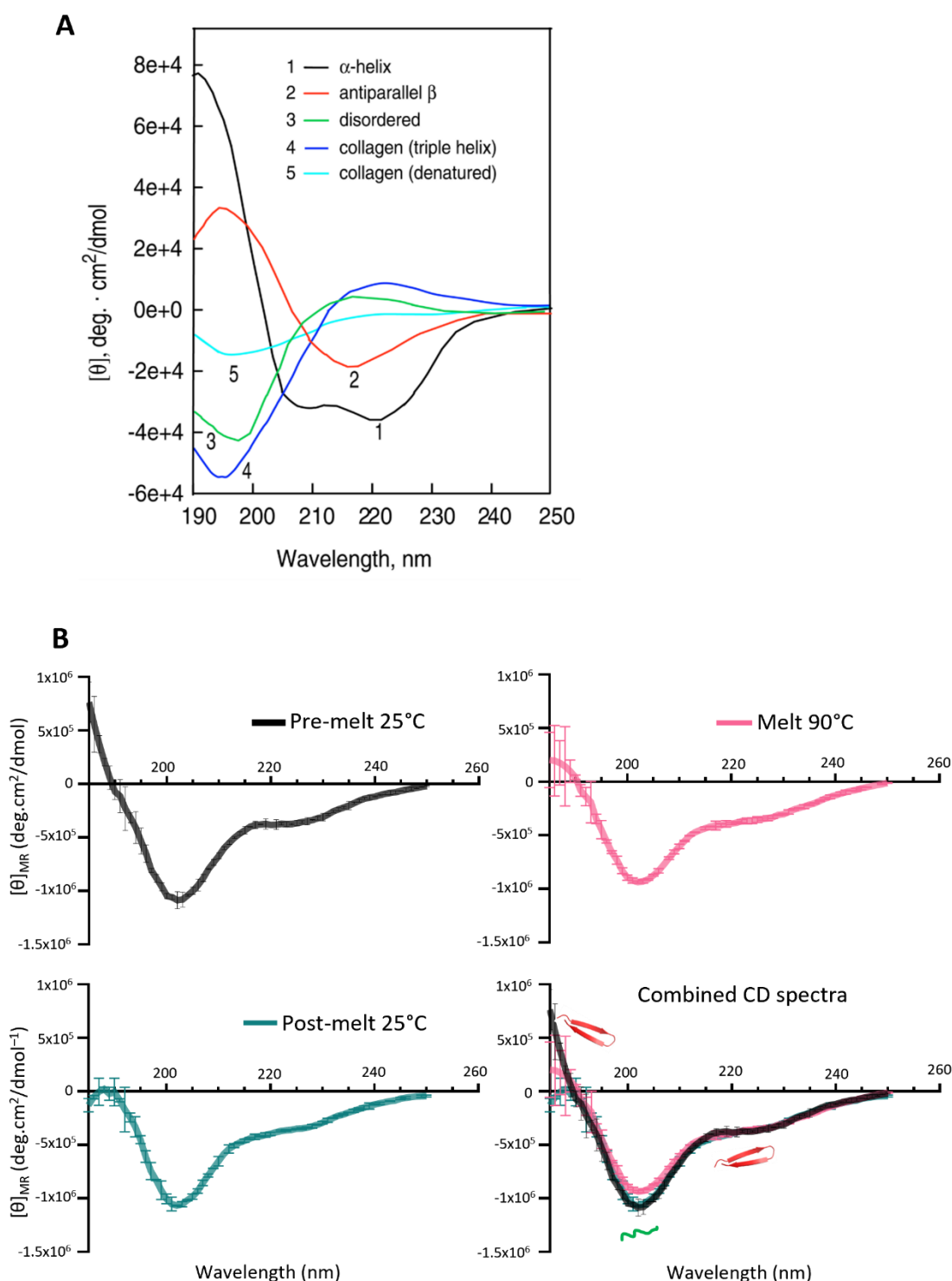


Figure 5.10 Circular dichroism confirms that the HSA region of RBPMS-B can form highly stable anti-parallel beta sheets: **A:** Standard circular dichroism spectra for various secondary structures, taken from Greenfield, 2007. **B:** Smoothed Circular Dichroism spectra of the HSA peptide at 0.1 mg/ml in a 0.1 cm cell. Spectra were taken at different temperatures, 25°C, 90°C and 25°C again after heating (melting). Spectra were analysed with the AVIV CD software. Small diagrams of either an antiparallel beta sheet (red) or random coil (green) have been added to the graph on the left of the lower panel to demonstrate which regions of the spectra resemble which secondary structure types. CDNN software used to report secondary structure features. $[\theta]_{\text{MR}}$ = mean residue molar ellipticity, refers to the molar circular dichroism of individual protein residues.

Negative stain images of peptide show large structures forming

Dynamic light scattering was used to measure the size of the particles that the peptide is forming, but the particles were shown as $>1\ \mu\text{m}$ in diameter, which is close to the size limit of DLS and so the size was not able to be accurately estimated using this method (data not shown). To observe the large structures the HSA peptide may be forming, a pilot negative stain electron microscopy (EM) experiment was performed, with the peptide in the same buffer that had been used for CD (Fig. 5.11). The peptide alone has a molecular weight of just over 2 kDa and thus would be too small to observe by itself. The DLS attempts, however, suggested that the peptide may be forming large structures and would therefore be visible. The EM images indeed show a variety of structures forming, many of them well over 100 nm in diameter. The most interesting of these structures are the long fibre structures seen in images 3 and 4. These appear to be similar to the amyloid fibril structures that have previously been observed with negative stain EM (Cooke & Brown, 2011; Kollmer *et al*, 2016). The buffer used in this EM experiment was the same as the CD buffer and the sample was prepared at 4°C, which is not ideal for fibril formation, however, even in these conditions the peptide was seen to form fibrils. These images, in conjunction with the stability of the anti-parallel β sheet secondary structure of the peptide seen after high temperatures with CD, point towards this peptide being able to undergo fibril formation.

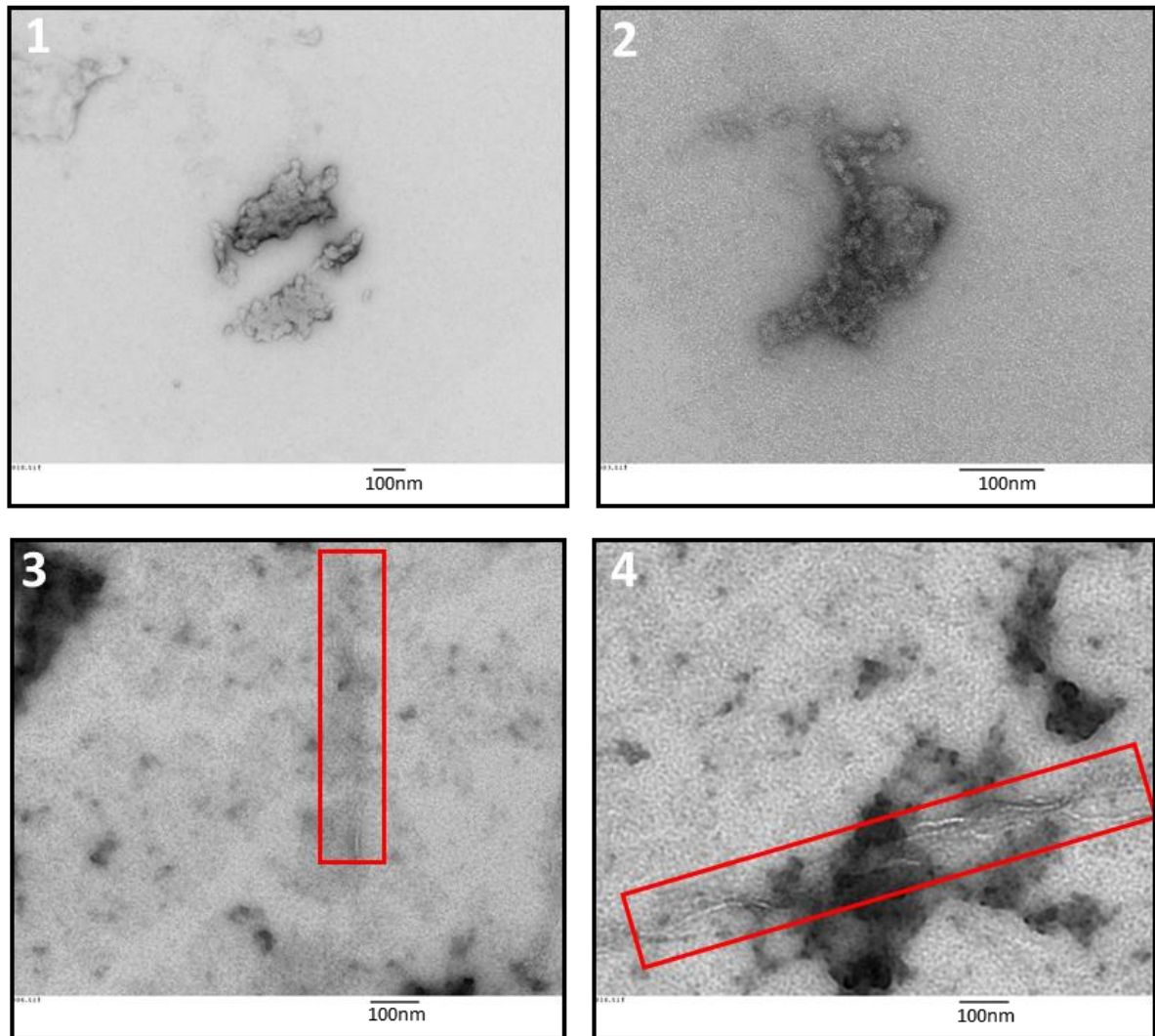


Figure 5.11 Negative stain electron microscopy images show HSA peptide of RBPMS-B forms large structures: Negative stain transmission electron microscopy images of the HSA peptide at 0.1 mg/ml. Images were taken with the Tecnai G2 80-200keV transmission electron microscope. Copper grids were used with a carbon film and Uranyl Acetate was used to stain samples. Red boxes indicate possible fibril structures.

HSA peptide binds ThT and behaves as a fibril

To assess whether the HSA peptide has a propensity to form fibrils, fluorescent plate reader assays were performed using a molecule called Thioflavin T (ThT). ThT binds to fibril structures and emits fluorescence at 480 nm when excited at 440 nm in a highly specific manner and has long been used to study amyloid fibres (Naiki *et al*, 1990; Levine, 1999; Biancalana *et al*, 2009; Gade Malmos *et al*, 2017; Xue *et al*, 2017). Samples were prepared in PBS, and experiments were performed at 37°C as this has been shown to be ideal conditions for fibril formation. To confirm whether ThT could accurately detect fibrils, an initial time-course assay was performed where proteins were incubated with and without

ThT for 18 hr (Fig.5.12). Carbonic anhydrase was used as a negative control as it does not form fibril structures, and the amyloid peptide A β 40 was used as a positive control as it is known to form fibrils and bind ThT (Lindskog, 1997; Xue *et al*, 2017). Fluorescence at 480 nm increases with time for the sample containing ThT and A β 40, but not for ThT or protein alone, or ThT with carbonic anhydrase (Fig.5.12), suggesting that ThT can indeed be used to specifically detect fibril formation.

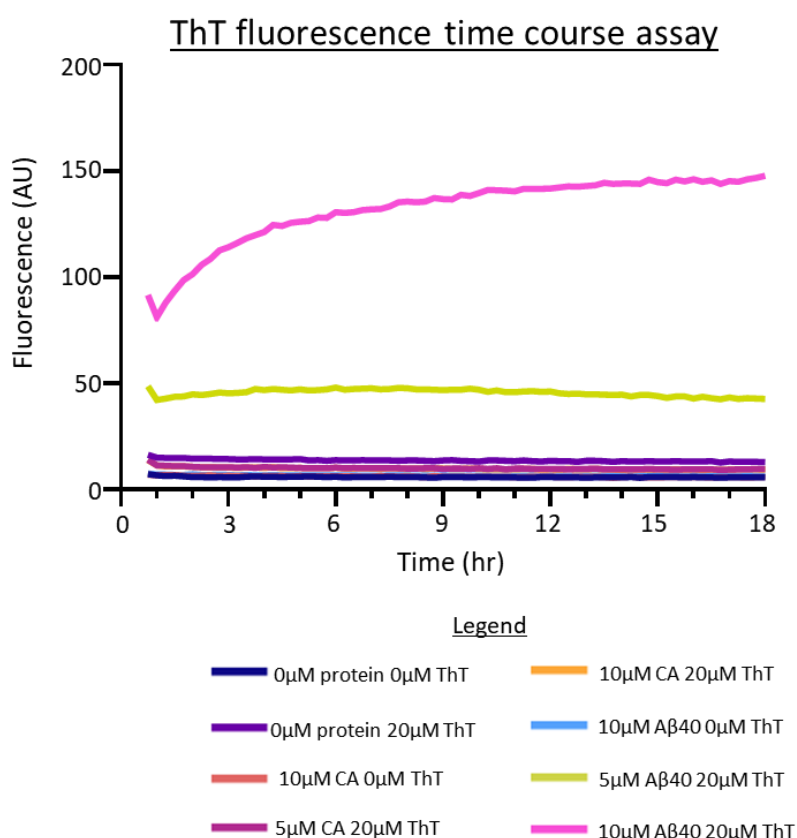


Figure 5.12 Fluorescence plate reader studies shows ThT binds specifically to fibrils: Pilot time-course data measuring fluorescence emission at 480 nm of Thioflavin T (ThT) in the presence or absence of increasing concentrations of A β 40 and carbonic anhydrase (CA) at 37°C. Time-course was between 0 – 18 hr and measurements were taken every 15 min at $\lambda = 480$ nm.

The HSA peptide was tested alongside the amyloid A β 40 peptide, the negative control carbonic anhydrase, along with another negative control, rat islet amyloid polypeptide (riAPP) which also does not form fibrils unlike its human homologue (Nanga *et al*, 2009; Press *et al*, 2019). As before, this experiment was carried out at 37°C for 18 hr (Fig.5.13). Once again, the amyloid peptide A β 40 showed increased fluorescence relative to the negative controls and ThT by itself. The fluorescence for the A β 40 and ThT sample was several orders of magnitude lower than for that of the HSA peptide with ThT. This suggests that the HSA peptide is indeed forming fibril structures, and it appears to have a much higher propensity than even the A β 40 peptide. There is no increase in fluorescence observed as

time increases, unlike for the amyloid peptide. This indicates that the fibrils have already formed before the experiment began. Experimental set up was performed at 4°C to slow fibril formation as generally amyloid fibres require temperatures closer to 37°C to form, as shown by both time courses for this peptide (Fig.5.12 & 5.13). The HSA peptide, however, appears to form fibrils quicker and at lower temperatures than the A β 40 peptide. Future tests would be needed to examine the kinetics of this fibrillisation.

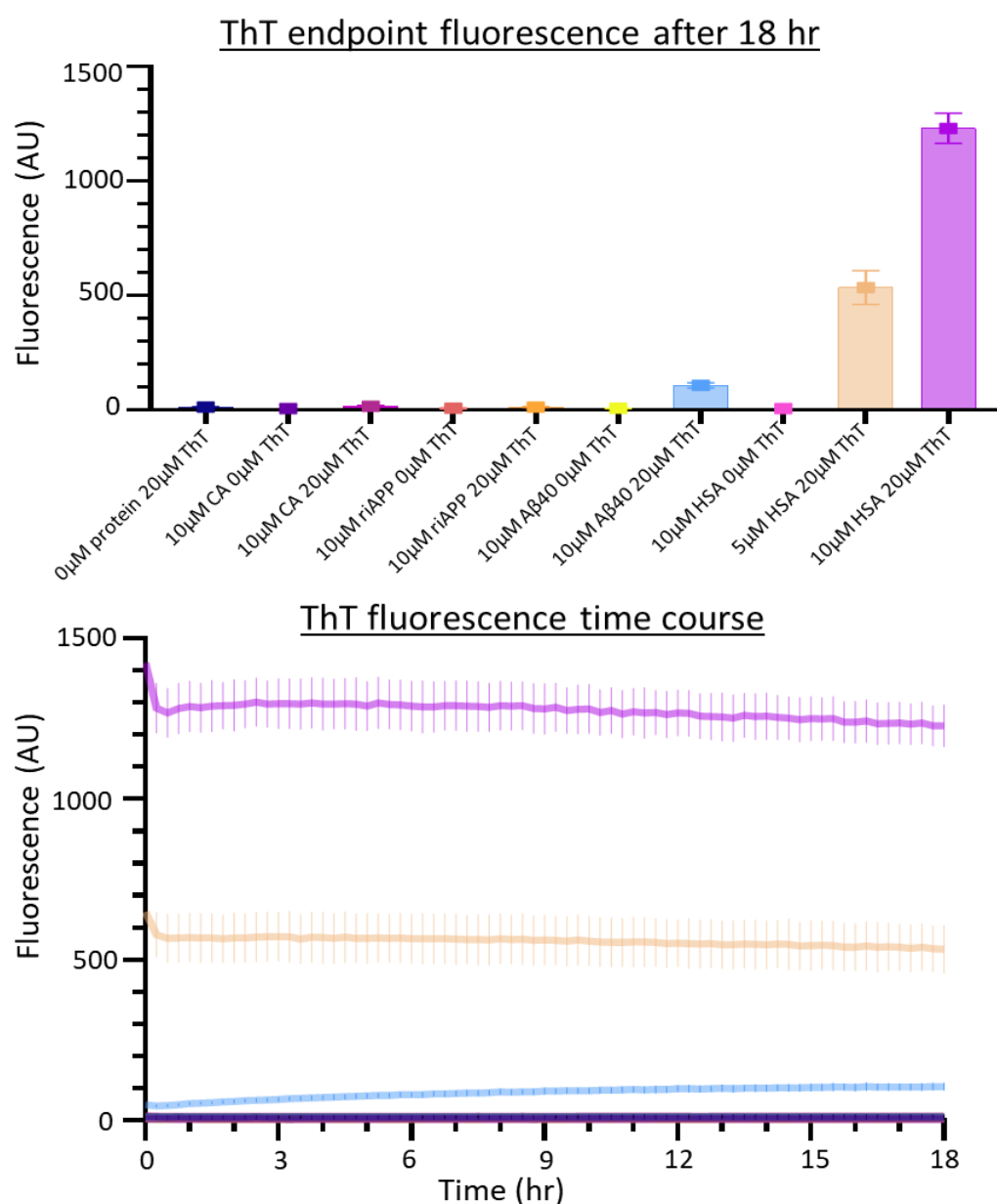


Figure 5.13 Fluorescence plate reader studies show ThT binds to the HSA peptide: Time-course and endpoint data of ThT plate reader assay: 10 μ M of Carbonic Anhydrase (CA), rat islet amyloid polypeptide (riAPP), A β 40, 5 – 10 μ M HSA peptide or no protein $\pm$ 20 μ M of ThT, with fluorescence emission of ThT at λ = 480 nm. Time-course was between 0 – 18 hr and measurements were taken every 15 min at λ = 480 nm. Three technical repeats were performed.

To investigate how stable these apparent fibril structures were and whether they could be denatured, a pilot experiment with the ThT fluorescent plate reader assay was performed with Guanidine HCl (Gdn HCl) for seven hours at 37°C. The HSA peptide was dialysed into a buffer containing 0 – 6 M Gdn HCl and then ThT was added at 20 µM (or 0 µM, data not shown) and the plate reader assay was performed (Fig.5.14). Guanidine HCl is used to denature proteins, and amyloid proteins are susceptible at high concentrations (Naiki *et al*, 1990; Narimoto *et al*, 2004; Xue *et al*, 2017). ThT fluorescence decreases as [Gdn HCl] increases (Fig.5.14), but there is no change overtime suggesting that the Gdn HCl initially partially denatures the fibrils, but that this process occurs quickly, before the start of the experiment. This is likely due to delays in sample preparation and the start of the measurements. At 6 M Gdn HCl, the HSA peptide fibrils have apparently been fully denatured, with ThT fluorescence at a similar level to the no protein sample. At 3 M Gdn HCl, the endpoint fluorescence of the HSA peptide is at a similar level to the Aβ40 peptide in 0 M Gdn HCl. This experiment shows that the fibrils formed by HSA are susceptible to denaturation and denature at similar Gdn concentrations to amyloid proteins, which require 6 – 7 M Gdn HCl for full denaturation also (Narimoto *et al*, 2004). In future experiments, the Aβ40 peptide would be included at each Gdn HCl concentration to see if this peptide required the same concentration of Gdn HCl to be denatured as the peptide in this context.

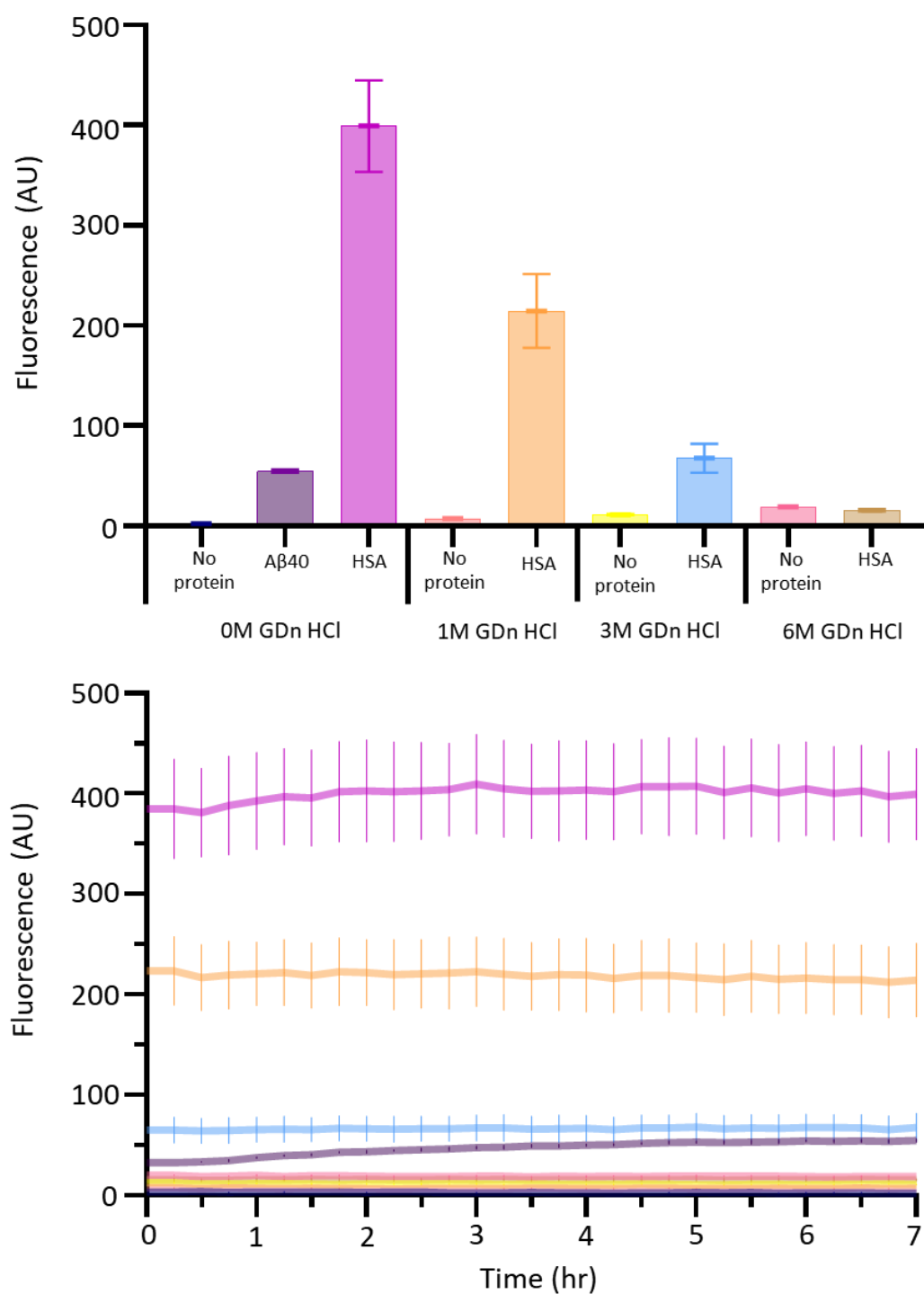
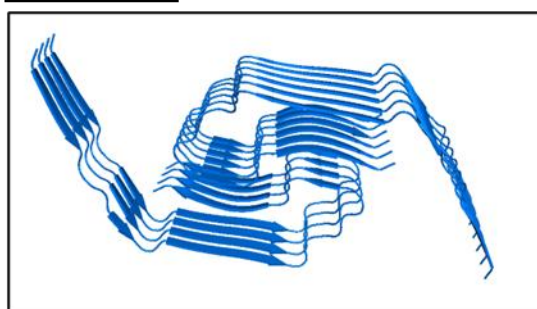


Figure 5.14 Guanidine HCl can denature HSA fibrils at high concentration: Time-course and endpoint data of ThT plate reader assay with increasing concentration of guanidine HCl $\pm$ 10 μ M HSA peptide and 20 μ M of ThT, with fluorescence emission of ThT at λ = 480 nm. Aβ40 was included as a positive control with 0 M guanidine HCl. Time-course was between 0 – 7 hr and measurements were taken every 15 min at λ = 480 nm. Three technical repeats were performed.

AlphaFold predicts that the HSA peptide behaves similarly to the A β 40 during oligomerisation

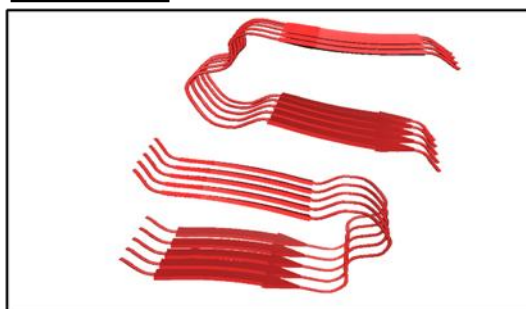
Given that the ThT assays demonstrated the HSA peptide is likely forming fibrils in a similar manner to the amyloid peptide, AlphaFold 3 was used to model how these peptides might oligomerise with ten copies of each peptide (Fig.5.15) (Abramson *et al*, 2024). The P2 HSA mutant was modelled along with a segment of the RRM of RBPMS known to form antiparallel β sheets (Y50 – V63) to observe if these form different structures to the wild-type HSA and amyloid peptides. The A β 40 peptide and HSA peptide form stacked antiparallel β sheet structure whereas the P2 mutant and RRM peptide formed two separate sheets of beta-sheet structures. The probability scores for all these structures were low suggesting low confidence in the structure predictions, but HSA and A β 40 peptides still showed a clear difference from the structures formed by the P2 and RRM peptides. This suggests that the P2 mutant cannot form similar structures to the wild-type HSA peptide and therefore is unlikely to undergo fibrillisation, but this would need to be confirmed by performing the ThT assays, negative stain and CD experiments with a P2 mutant peptide.

A β 40 peptide



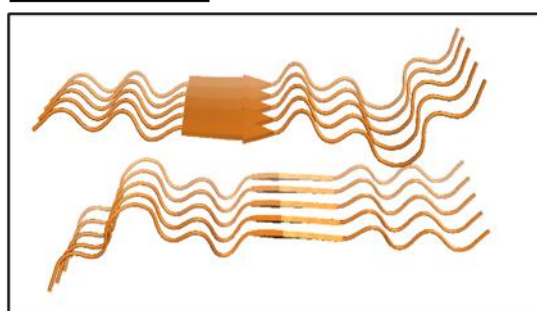
DAEFRHDSGYEVHHQKLVFFAEDVGSNK
GAIIGLMVGGVVIA

HSA peptide



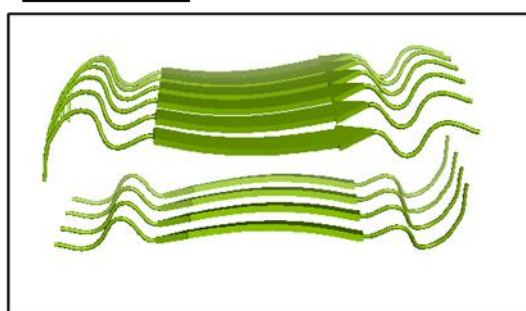
LLQQRVSGNVFVTYQ

HSA P2 peptide



LLQQ**P**RVSGNV**F**PTYQ

RRM peptide



YEGSLIKLTSKQPV

Figure 5.15 Amyloid peptide and HSA form similar structures with AlphaFold: AlphaFold 3 models of 10 copies of each peptide: A β 40 (top left), wild-type HSA (top right), P2 HSA peptide (bottom left) and a section of the RRM seen to form antiparallel β sheet (bottom right). The amino acid sequences of each peptide are given below their models, and the location of the proline substitutions in the P2 are in bold. Images taken with Pymol.

5.6 FULL LENGTH RBPMS-B PROTEIN FORMS LARGE SPHERICAL STRUCTURES

Recombinant RBPMS-B forms large complexes during purification

Recombinant RBPMS-B had previously been purified in a similar manner to wild-type RBPMS-A as described in chapter 3.2. Only very small amounts of protein were able to be purified with this method as most of the protein formed large complexes which were highly stable and were above the void volume of the columns used in purification (>7MDa) (Yang, unpublished). As an alternative approach, purification of recombinant N-terminally STREP-II 6xHis tagged RBPMS-B was attempted by glycerol gradient sedimentation followed by STREP- and His-tagged bead purification. The *E. coli* lysate following benzonase treatment and centrifugation was subjected to a 10 – 60% glycerol gradient with 15.5 hr centrifugation at 32 000 rpm, and fractions were run on an SDS-PAGE gel which was Coomassie Blue stained (Fig.5.16A). The band above the 29kD marker in fractions 20 – 25 (indicated by the red box) was the correct size for tagged recombinant RBPMS-B. As expected from the previous purification attempts, RBPMS-B was found in the heaviest fractions for this gradient. Sedimentation profiles of the *E. coli* 30S and 50S ribosomal subunits to confirm the gradient sedimentation was successful; 2.5 mM EDTA was added to keep the subunits from forming the 70S complex (Fig.5.16B). These results indicate that RBPMS-B forms very large complexes when expressed in *E. coli*, as not only is it found in the heaviest fractions of the sedimentation gradient, it also sediments into heavier fractions than the 50S ribosome subunit.

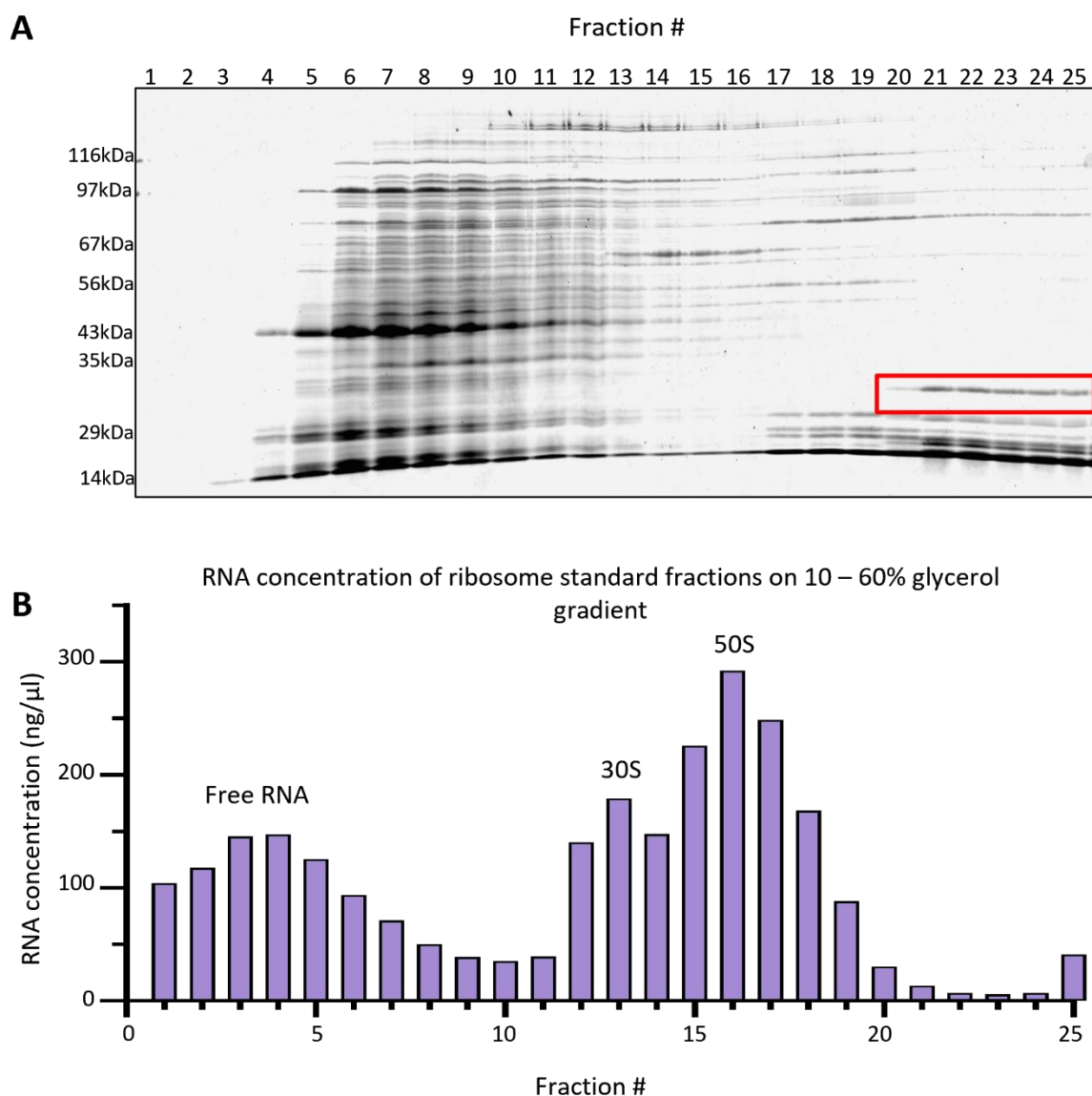


Figure 5.16 Recombinant RBPMS-B sediments into heaviest fractions of 10-60% glycerol gradient: A: Coomassie stained SDS-PAGE gel displaying the fractions from a 10-60% glycerol gradient containing Benzonase treated RBPMS-B wild-type lysate from *E. coli*. Fractionation was performed from top to bottom of the gradient, with the first fractions containing the lightest material. Red-box highlights the likely mobility of RBPMS-B. **B:** Sedimentation profile of *E. coli* ribosomal subunits used as standards to test sedimentation speed of the 10-60% glycerol gradient. RNA concentration was measured on via UV absorption at $\lambda = 260$ nm.

The sample was then run through a capto-700 column which has a size-exclusion limit of 700 kDa, which corresponds to ~30 subunits of RBPMS-B. This step was to remove the smaller proteins, whilst RBPMS-B oligomers should be large enough to go into the flow-through, as is seen in Figure 5.17A (Fig.5.17A, lanes 1 – 6). STREP agarose beads were used to try and capture RBPMS-B via its STREP tag, but as can be seen from Figure 5.17B, there was no RBPMS-B seen in the elution fractions and instead it appeared to still be in the flow-through (lane 1). This is likely due to RBPMS-B being too large to bind to these beads, which have an exclusion limit of 6 MDa, which would correspond to ~200 subunits of RBPMS-B. His Dynabeads were also tested for their ability to purify RBPMS-B via its His-tag (Fig.5.17C). Sample was eluted from the beads with imidazole, and anything left on the beads was eluted via boiling the beads and run on an SDS-PAGE gel to see if the imidazole elution was efficient. The lysate sample (lane 1) was used to compare how clean the whole purification process had been, and how efficient it was. RBPMS-B can be seen to be pulled down specifically with these beads and eluted with imidazole (lanes 5 & 6). The elution was not 100% efficient as there is still RBPMS-B that appears to be left on the beads (lanes 2 & 3), but the eluates are clean showing purification of RBPMS-B was successful.

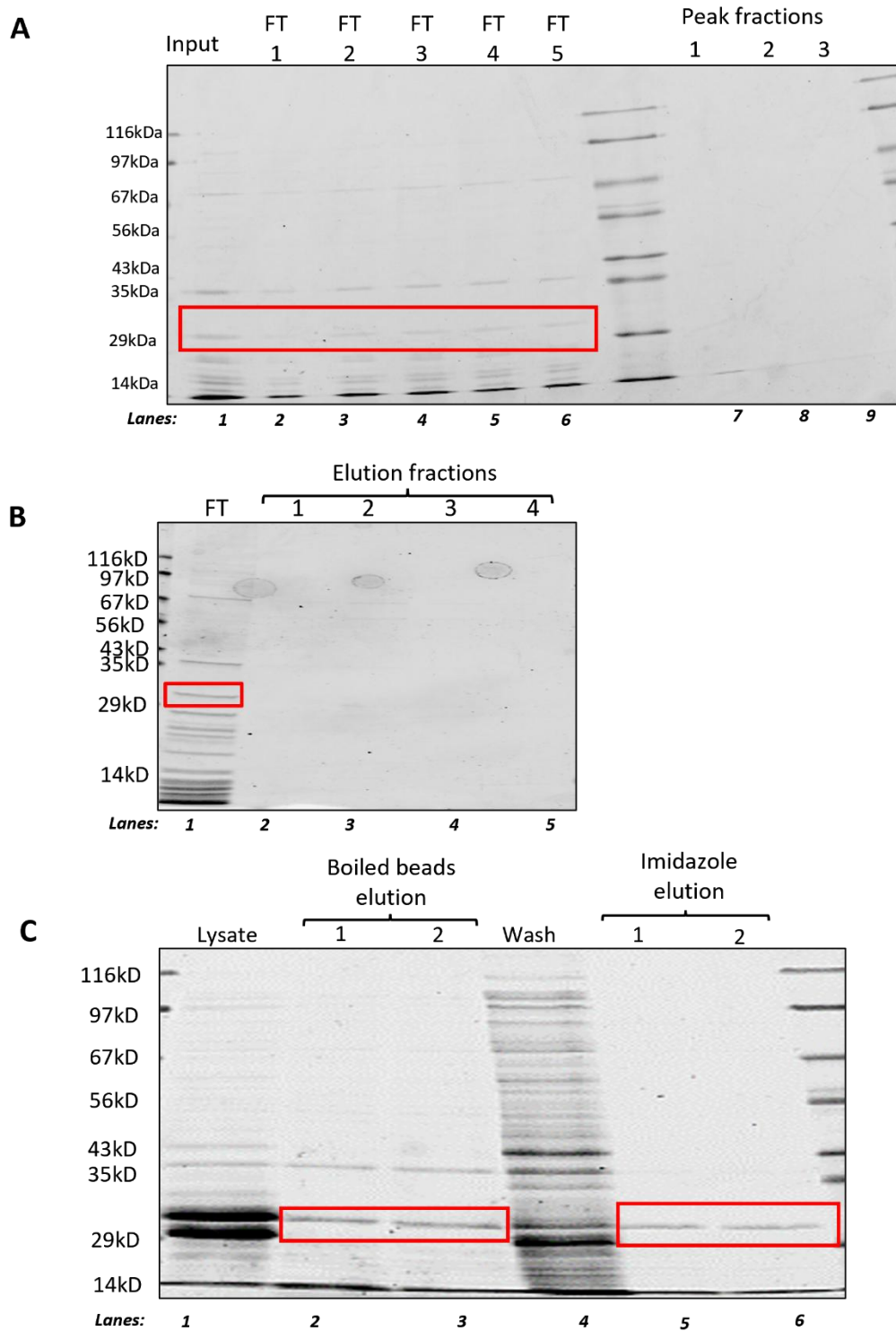


Figure 5.17 Recombinant RBPMS-B can be successfully purified: **A:** Flow-through (FT) and peak fractions on a Coomassie stained SDS-PAGE gel after RBPMS-B sedimentation fractions were run through a capto-700 column. **B:** Flow-through and peak fractions on a Coomassie stained SDS-PAGE gel after recombinant RBPMS-B capto-700 flow-through fractions were run through a column containing STREP-Tactin agarose beads. **C:** Benzonase treated lysate from *E. coli* expressing RBPMS-B was run on an SDS-PAGE with sample after His Dynabead elution. Beads refers to sample that was left on the beads after imidazole elution, the sample was boiled to remove it from the beads. Elute refers to the imidazole eluted sample and wash is the sample that did not bind to the beads and was collected after the first wash step. Red-boxes highlight the likely mobility of RBPMS-B.

Despite the successful purification of RBPMS-B, the yield at the end of the process was quite low (200 μ l of 0.5 mg/ml protein, versus RBPMS-A wild-type which was purified with 1 ml of $\sim$ 2-3 mg/ml protein). Two different preparations of RBPMS-B protein lysates were run on an SDS-PAGE gel and Coomassie Blue stained (Bwt1 and Bwt2), before and after the centrifugation step, which occurs after Benzonase treatment, alongside an unrelated lysate from the RBPMS-A YYYW/A mutant, which runs higher than RBPMS-B (Figure 5.18). Here, most of the RBPMS-B is in the pellet after centrifugation rather than the supernatant, meaning most of it is insoluble (lanes 4 & 5 versus 6 & 7).

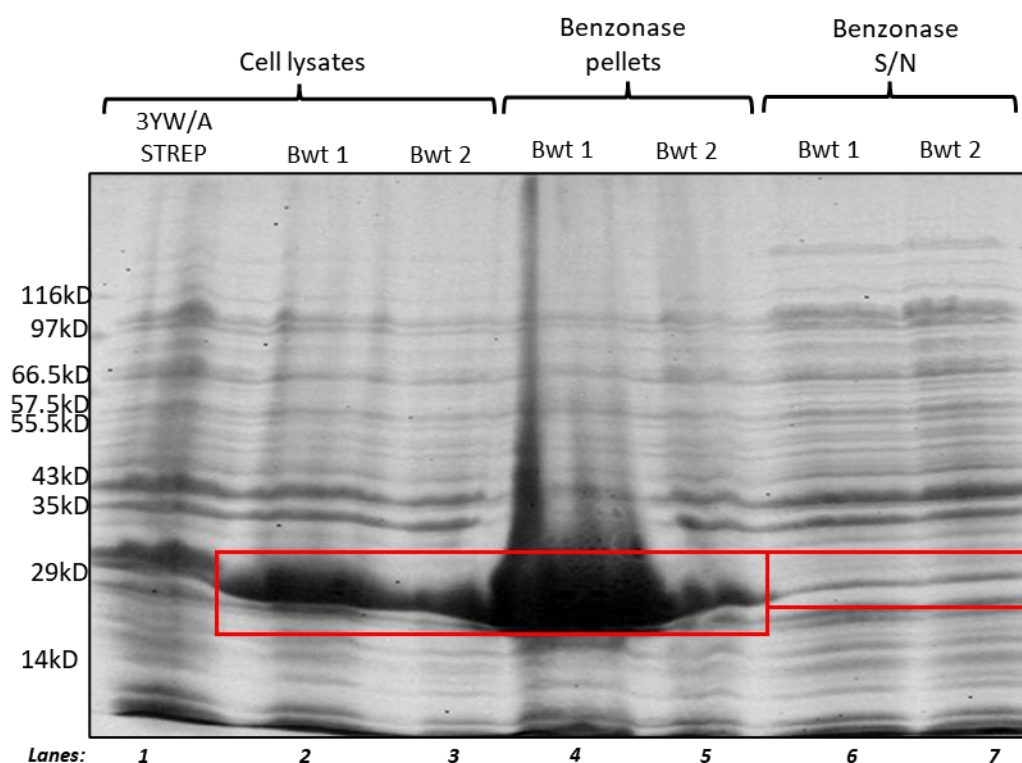


Figure 5.18 Recombinant RBPMS-B is mainly insoluble after protein induction: Coomassie stained SDS-PAGE gel demonstrating the expression of recombinant RBPMS-B in cell lysates, and how much is insoluble after centrifugation and benzonase treatment (pellet) and how much is soluble (in the supernatant, S/N). RBPMS-A YYYW/A STREP lysate was run next to the RBPMS-B lysates for comparison as the two proteins have different mobilities.

A pilot experiment was performed testing three different induction temperatures, 18°C, 37°C and 42°C (Fig.5.19). Despite different levels of total induction (lanes 2 – 3), an induction temperature of 42°C gave the highest proportion of soluble recombinant RBPMS-B.

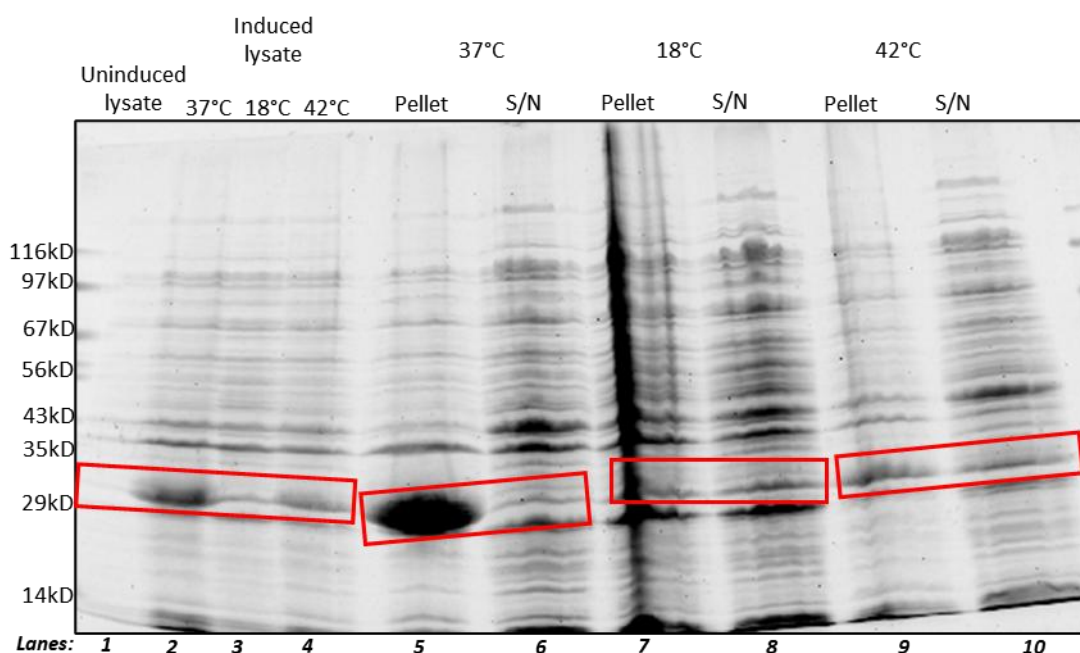


Figure 5.19 42°C protein induction incubation temperature leads to more soluble RBPMS-B being produced: Coomassie stained SDS-PAGE gel comparing protein expression levels of recombinant RBPMS-B and amounts of insoluble (pellet) versus soluble (supernatant, S/N) protein with protein induction at different temperatures.

Negative stain images show RBPMS-B forms large spherical structures

Purified recombinant RBPMS-B from Figure 5.17C (lanes 5 and 6) was imaged by negative stain EM to examine whether the full-length protein formed similar structures to the HSA peptide (Fig 5.20A). The structures formed by RBPMS-B full-length appear to be quite different to the HSA peptide (Figure 5.11). These structures are mainly circular, whereas the structures formed by the HSA peptide alone were more heterogeneous. Furthermore, there are no fibrils visible with the full-length protein. The circular structures formed by full-length RBPMS-B have a diameter of ~50 nm on average (Fig.5.20B), which is approximately triple the average diameter of the RBPMS-A structures seen by EM (~14 – 15 nm) (Yang *et al*, 2023). These images suggest that the full-length protein may not form fibrils, or at least not under these conditions. The full-length protein was in a different buffer (20 mM HEPES pH 7.9, 100 mM KCl, 0.5 mM DTT) compared to the peptide which was in the CD buffer. The buffer of the full-length RBPMS-B protein was the same as was used to assess RBPMS-A by EM but does not allow for fair comparison with the HSA peptide. The HSA peptide should be tested in a similar buffer and conditions to full-length RBPMS-B protein to allow for proper comparison of the structures formed. These results, however, do indicate that RBPMS-B forms large complexes, larger than that observed for RBPMS-A with a similar

method. Given that negative stain EM gives 2D projections, it can be assumed that these structures are spherical in shape, and therefore a rough estimate of their volume and the number of subunits they contain can be made. The rough estimated radius of an RBPMS dimer is ~3 nm (approximated in pymol), and the approximate radii of the RBPMS-B structures seen here are ~25 nm. Using the volume of a sphere equation $V = \frac{4}{3}\pi r^3$, the volume of a dimer of RBPMS, if assumed to be spherical, is ~113 nm³ (molecular weight = 44 kDa) and the volume of the RBPMS-B structures is ~65 450 nm³. This corresponds to the ~579 dimers of RBPMS-B and is ~25 MDa in molecular weight, far beyond the exclusion limit of the STREP agarose beads tested in Figure 5.17B. For comparison, the approximate radii of RBPMS-A structures seen by EM were ~7.5 nm, corresponding to a volume of 1767 nm³ or ~16 dimers with a molecular weight of ~660 – 690 kDa. The nature and stability of these structures should be investigated to uncover how they may be linked to the different splicing activity of RBPMS-B relative to RBPMS-A.

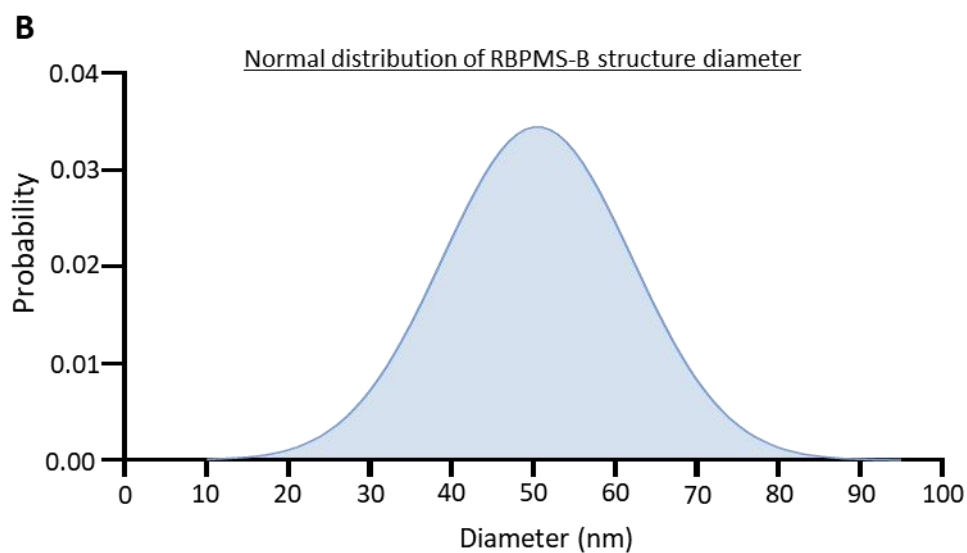
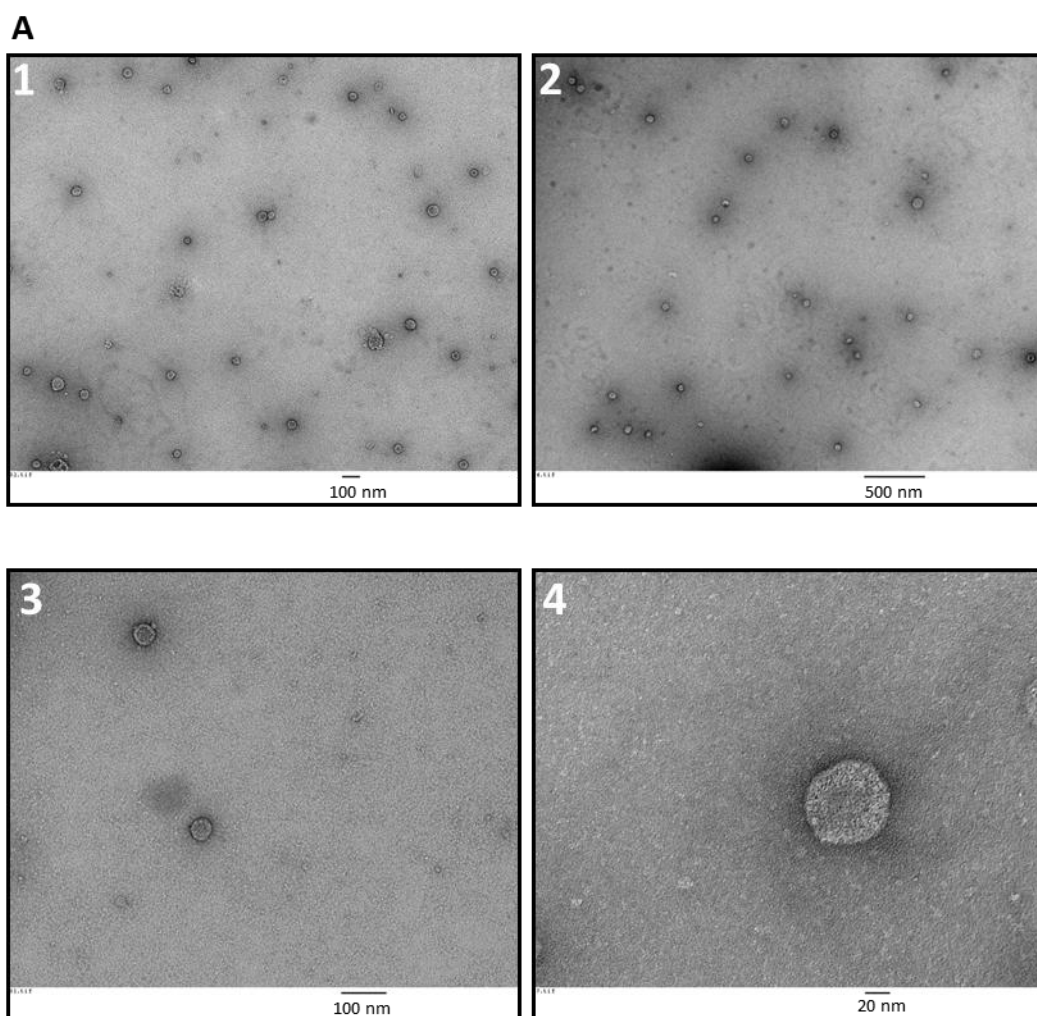


Figure 5.20 Negative stain EM images show RBPMS-B forms large circular structures in vitro: A: Negative stain transmission electron microscopy images of recombinant RBPMS-B at 0.5 mg/ml. Images were taken with the Tecnai G2 80-200keV transmission electron microscope. Copper grids were used with a carbon film and Uranyl Acetate was used to stain samples. **B:** Normal distribution of the diameters of the structures that RBPMS-B is observed to form from the negative stain EM images.

The proline RBPMS-B mutant behaves like RBPMS-A during purification

The RBPMS-B proline mutant protein or peptide was not able to be included in the above experiments during the timeframe of this study. The recombinant protein was purified, however, (Fig.5.21) and as it had similar predicted aggregation propensity to wild-type RBPMS-A and similar activity, it was purified as described for the STREP 6xHis tagged RBPMS-A (Section 2.31 & Fig.4.7). These gel images and AKTA traces show that RBPMS-B P2 mutant was able to be purified in a similar manner to RBPMS-A, and its peak fractions were comparable to wild-type RBPMS-A. This suggests that RBPMS-B P2 and wild-type RBPMS-A may form similar size complexes, compared to wild-type RBPMS-B which was not able to be purified efficiently by this method due to its large size. More biophysical and structural analyses comparing RBPMS-A, RBPMS-B and its P2 mutant protein would need to be carried out to confirm this observation.

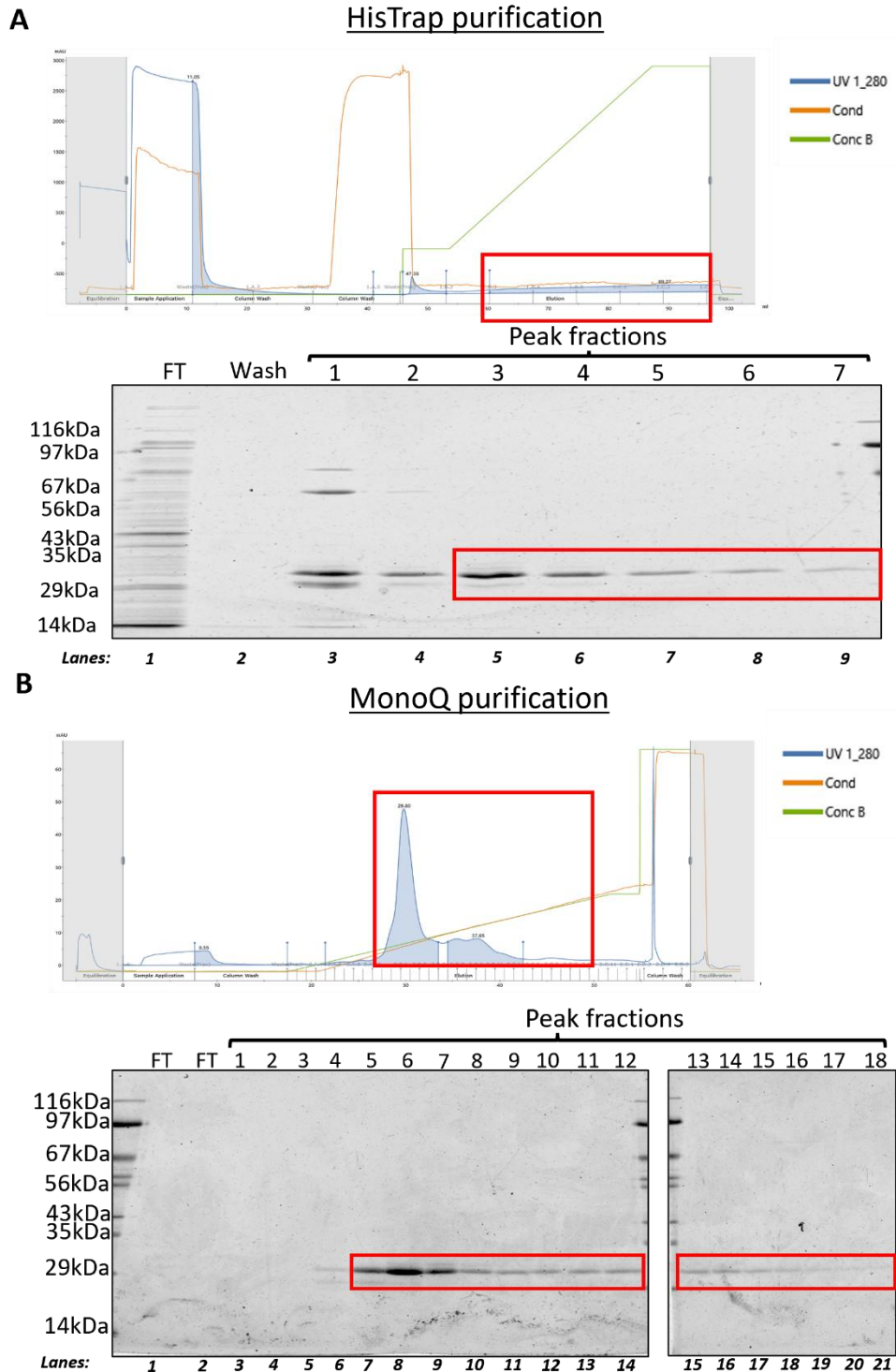


Figure 5.21 Recombinant RBPMS-B P2 mutant protein can be purified in the same manner as RBPMS-A: **A:** HisTrap column purification of the recombinant RBPMS-B P2 mutant. Top = UV/Vis spectra traces produced by the UNICORN AKTA software. Bottom = Coomassie Blue stained SDS-PAGE gels showing the RBPMS-B P2 mutant purification after HisTrap. **B:** Mono Q column purification of the recombinant RBPMS-B P2 mutant. Top = UV/Vis spectra traces produced by the UNICORN AKTA software. Bottom = Coomassie stained SDS-PAGE gels showing the RBPMS-B P2 mutant purification after Mono Q purification. The red box indicates the clean fractions. FT = flow-through during column purification, Cond = conductivity (salt concentration), Conc B refers to concentration of either HisB buffer or QB (section 2.3), UV 1_280 refers to UV absorbance at $\lambda = 280$ nm.

5.7 CONCLUDING REMARKS

From these results, it appears that RBPMS-B has reduced splicing activity relative to RBPMS-A with repression targets, but not the activation targets except for *PTPRF* (Figures 3.3 & 5.2C). The pilot knockdown identified two targets (*Myocd* and *Ppfibp1*) in which RBPMS-B knockdown had a greater reduction in inclusion of the activated exons relative to RBPMS-A (Fig.5.2C). The results from this chapter indicate the difference in the ability of RBPMS-B to regulate splicing events compared to RBPMS-A is likely to be due to the seventeen amino acid HSA region in its unique C-terminus. Deletion mutations showed that certain residues antagonise activity (VFVTY) whilst others may be conferring activity (QQ) (Figures 5.3 & 5.4). The online tool AGGRESCAN predicted that the presence of the VFVTY residues increases aggregation propensity, whilst the glutamine residues appear to be reducing this predicted aggregation behaviour. This suggests that aggregation behaviour and activity may be linked. This was also demonstrated by the cysteine to serine mutant (CC/S), which had also been predicted to have reduced aggregation propensity compared to wild-type, and this mutant had increased splicing activity, particularly with repression target *ACTN1* (Figure 5.3). The isoleucine/valine to proline mutant (P2 and P3) proteins were produced as proline is predicted to “break” β strands in amyloid-like fibrils, and this was shown in aggregation propensity prediction scores (Figure 5.7A). Although the P2 mutant still showed some β strand character in the AlphaFold model (Fig.5.6), when AlphaFold was used to test how the oligomerisation of the peptides might occur (Fig.5.15), the P2 mutant showed a different structure to the wild-type HSA peptide and amyloid peptide. Both proline mutant proteins had increased activity relative to wild-type RBPMS-B and the P2 protein had no significant difference to the activity of RBPMS-A. The HSA/alanine mutant (HSA/A) had low predicted aggregation propensity but had reduced activity relative to wild-type RBPMS-B protein. This suggests that some of the residues included in this substitution mutant are required for the splicing activity, two of which (F198 & 204) were also in the FFFF mutants, which lost activity in the P2 context. This means that the predicted aggregation propensity cannot always be a predictor of splicing activity for RBPMS-B. Another example of how aggregation propensity and activity are not always linked for RBPMS-B, was how the HSA mutants behaved with the *PPFIBP1* minigene. Here, RBPMS-B and the CC/S mutant protein had the highest activity, but the Δ VFVTY had reduced activity compared to the wild-type protein, although it was still similar to RBPMS-A. This target may represent a splicing event which requires the C-terminus of RBPMS-B and perhaps its propensity to form larger structures than RBPMS-A for activation, hence the wild-type protein can increase inclusion more than the deletion mutant protein which behaves more similarly to RBPMS-A. Furthermore, it indicates that these VFVTY residues may be required for the splicing activity of RBPMS-B for certain targets. Indeed, when Y207 was included in the aromatic

substitution mutants tested (Fig.5.8 & 5.9), activity was lost in both the wild-type and P2 contexts. Whether the substitution of this residue is solely responsible for this additional loss in the wild-type context or whether it requires the mutation of the phenylalanine residues, thus demonstrating redundancy between the aromatic residues, remains to be seen. The phenylalanine residues were shown to be important for activity for the P2 mutant, but not the wild-type protein. The YYYW residues were found to be important for activity in both contexts, suggesting these residues may be playing a similar role in this isoform as they are with RBPMS-A. The loss of activity upon mutation of tyrosine and tryptophan residues in the wild-type context versus the retention of activity with phenylalanine substitutions hints that there may be a preference with RBPMS-B towards tyrosine and tryptophan residues over phenylalanine residues. Phenylalanine residues are known to form weaker cation – π and π – π interactions than tryptophan and tyrosine residues, suggesting that these residues of RBPMS-B may be forming similar interactions as they are in their RBPMS-A counterpart (Wang *et al*, 2018). Further investigation into the importance of these aromatic residues as well as the HSA region in activity, higher order assembly and protein – protein interactions of RBPMS-B is required to confirm these observations.

The recombinant full-length protein studies show that RBPMS-B forms very large complexes *in vitro* (Figures 5.16A & 5.20), and the HSA peptide studies demonstrated that there is a propensity of this region to undergo fibrillisation (Figures 5.11 – 5.15). Such large structures of RBPMS have not been observed with immunofluorescence studies of RBPMS in PAC1 cells, suggesting that in a cellular environment the very extreme higher order assembly behaviour of RBPMS-B observed by EM *in vitro* does not occur and may be prevented by other molecules present in the cell (Nakagaki-Silva *et al*, 2019; Huang, unpublished). The *in vitro* data of both the full-length and HSA peptide does, however, indicate that RBPMS-B does have a propensity to form large structures, which may be linked to its lower activity compared to RBPMS-A. The full-length protein was not observed to form fibrils in the negative stain EM experiment (Figure 5.20A), which may suggest that RBPMS-B is protected against this phenomenon by the rest of the protein preventing the HSA region from driving fibril formation. These results, however, do not mean that the full-length protein is unable to form fibrils. The fibril forming conditions used in the ThT plate reader study were not used for the negative stain EM experiment, where the samples were prepared on ice. It will be interesting to observe whether the full-length protein is able to form fibrils under these conditions, and whether this can be shown by ThT fluorescence and negative stain EM. Additionally, including the P2 full-length protein and P2 version of the HSA peptide in these future studies will be useful controls to confirm that this mutant cannot form the large structures seen nor the fibril structures that the wild-type HSA peptide can form. From the purification of this proline RBPMS-B mutant, it behaved much more similarly to the RBPMS-A

isoform than to wild-type RBPMS-B which formed complexes too large for this purification method. This indicates it may have similar assembly properties to RBPMS-A which may explain its similar activity to this isoform, and greater activity to RBPMS-A. The AlphaFold structure of full-length RBPMS-B (Figure 5.6, left model), shows the anti-parallel β sheet structure of the HSA region packing against the RRM, although it packs against the alpha helices of the RRM which are not directly involved in RNA binding. The likelihood of the location of the HSA being so close to the RRM in this model is low, as it most probably due to the software being inaccurate with where to “organise” the unstructured tail and so places any similar structures together.

Overall, it appears from the results of this chapter that the extreme higher order assembly behaviour seen by wild-type RBPMS-B protein (Figure 5.16A & 5.20) is related to its reduced of activity in comparison to RBPMS-A. The results from the HSA mutants, in particular the P2 mutant, demonstrate that by disrupting this behaviour, RBPMS-B activity can be rescued. *In vitro* studies such as splicing assays, and biophysical analyses comparing RBPMS-A, RBPMS-B and the proline mutant would confirm this observation about activity and higher order structures.

5.8 FUTURE PLANS

Further analysis into the HSA region, and how it alters the assembly and activity of RBPMS-B is required to elucidate the structure – function relationship of RBPMS-B and how this relates to RBPMS-A:

- Further repeats of the isoform specific knockdown followed by RNA sequencing experiments to identify global effects each isoform has on splicing in differentiated PAC1 cells
- Immunofluorescence (IF) experiments comparing differentiated PAC1 cells with cells which have been treated with isoform specific siRNA to identify the localisation of each isoform
- Further aromatic residue mutation in RBPMS-B C-terminus to investigate the importance of the tyrosine and tryptophan residues (individual mutation of Y207 and in combination with YYY and YYYW mutants)
- Seeding experiments: Adding the HSA peptide to full-length RBPMS-B protein and observing whether the full-length protein has any propensity to form fibrils under different conditions
- *In vitro* splicing assays with recombinant full-length RBPMS-B and P2 mutant compared to wild-type RBPMS-A with *PTPRF*, *Tpm1*, and *FLNB*
- Negative stain EM experiments with HSA, RBPMS-B, RBPMS-B P2 and RBPMS-A in similar conditions (fibril forming conditions) to see how their assembly/structures compare
- CD experiments with full-length protein and proline P2 mutant

- ThT assays with wild-type RBPMS-B and P2 mutant proteins and P2 peptide to see if this assay can detect their fibrilisation
- Biophysical assays (such as mass photometry and AUC) with RBPMS-A, RBPMS-B, RBPMS-B P2 and the aromatic mutants to elucidate the higher order assembly differences between the isoforms
- Pull-down and proteomic analysis of RBPMS-A, RBPMS-B and the P2 mutant proteins to investigate the interaction partners of the two isoforms and whether the proline mutant interacts with similar proteins to RBPMS-A versus RBPMS-B.

GENERAL DISCUSSION

6.1 STRUCTURE VERSUS FUNCTION OF RBPMS

The research described in this thesis has focussed on the oligomerisation of RBPMS, and how this relates to its function. These aspects of RBPMS behaviour were investigated as they have previously shown correlation. RBPMS-A was seen *in vitro* to form dynamic higher order assemblies of >20mer, but unexpectedly, these larger oligomers were identified to not be required for function. RBPMS-B was shown previously to be less active than RBPMS-A and potentially form larger oligomers, but this behaviour had not been experimentally validated. Here, RBPMS-B was indeed found to form much larger oligomers of potentially ~25MDa in size *in vitro*, likely due to a seventeen amino acid region (HSA) in its unique C-terminus and its propensity to form anti-parallel β sheets. A proline substitution mutant which disrupted this region, was found to have greater activity and likely had reduced higher order assembly propensity. The findings about RBPMS-B do suggest a link between assembly and activity, but here, the higher order assembly appears to be detrimental to activity.

RBPMS-A was previously thought to require self-assembly for activity

Previously, it was demonstrated that RBPMS-A can undergo dynamic self-assembly into higher order oligomers, which was found to be dependent on protein concentration and electrostatic interactions (Peacey, Part III Dissertation, 2020; Yang *et al*, 2023). The results in chapter 4 also confirmed that wild-type RBPMS-A undergoes higher order assembly (Figures 4.16 – 4.21) and can even undergo LLPS in the presence of crowding reagents (Figures 4.14 & 4.15). Deletion of the final twenty amino acids of RBPMS-A (Δ C20) was shown to result in a loss of this higher order assembly, demonstrating that this C-terminal tail can drive this assembly (Yang *et al*, 2023). The Δ C20 mutant also lost splicing repression function with *Tpm1*, which was also observed in the present study, where this mutant could no longer repress exon 3 of *Tpm1*, or the NM exon of *ACTN1*, or indeed fully activate the inclusion of exon 23 of *PTPRF* (Figures 4.1, & 4.3 – 4.5). These results identified that the final twenty amino acids of RBPMS-A are necessary for both its function and higher order assembly, and thus this was interpreted to indicate that the higher order assembly of RBPMS-A is required for its function. This was further suggested by the phosphomimetic (TT/E) and phosphonull (TT/A) studies of RBPMS-A carried out in the lab, where TT/E lost activity and higher order assembly, whereas TT/A retained both activity and oligomerisation to similar levels to the wild-type protein (Barnhart *et al*, 2022).

Many RBPs have been shown to undergo oligomerisation to perform their splicing activity. SR proteins can localise to the phase separated nuclear speckles, which were once thought to be purely for storage

purposes, but SR proteins have recently been suggested to actively regulate splicing at the surface of these speckles (Sabari *et al*, 2020; Liao & Regev, 2021; Paul *et al*, 2024). RBPs that regulate alternative splicing have been shown to undergo localised assembly to perform their functions, such as TDP-43 and its formation of so-called binding-region condensates on multivalent RNA targets, and RBFOX1 and its requirements for self-assembly for splicing activity (Ying *et al*, 2017; Hallegger *et al*, 2021). Furthermore, it was proposed that for proteins containing one or two RNA binding domains (such as FUS, TDP-43, and RBFOX1/2), multimerization of the proteins on a multivalent RNA requires self-assembly via their IDRs (Banani *et al*, 2017; Ule & Blencowe, 2019). Therefore, given the observations that RBPMS-A can undergo higher order assembly, and a mutant disrupted both assembly and splicing activity, it was proposed that RBPMS-A requires assembly to perform its function.

The other major isoform found in differentiated VSMCs, RBPMS-B, was predicted to be aggregation prone, specifically due to the seventeen amino acid HSA region in its unique C-terminal tail (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019). This protein was found to be less active than RBPMS-A with a number of targets and was suggested *in vitro* to form large complexes of >7MDa (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019; Yang, unpublished). Consequently, it was hypothesised that extreme higher-order or aggregation-like assembly of RBPMS may be detrimental to its splicing activity.

6.11 RELATING OLIGOMERISATION TO FUNCTION

Higher order assembly and splicing activity can be uncoupled for RBPMS-A

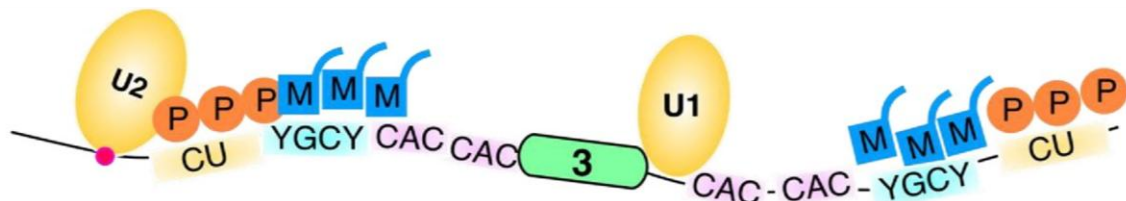
It was expected that the mutants of RBPMS-A which affected higher order assembly would have reduced activity, as had been previously demonstrated with the Δ C20 and TT/E mutant proteins (Barnhart *et al*, 2022; Yang *et al*, 2023). Indeed, the aromatic mutants exhibited this behaviour, with each having reduced higher order assembly to a similar extent and reduced splicing activity for both activation and repression events in transfection assays and *in vitro* (Figures 4.3 – 4.5, 4.8, 4.16 & 4.20). Unexpectedly, the basic mutant RKR/A had reduced higher order assembly relative to the wild-type protein despite showing full activity for splicing activation and repression (Figures 4.3 – 4.5, 4.9, 4.17 & 4.20). Furthermore, this basic mutant also had reduced UV crosslinking to RNA ability, suggesting it had reduced RNA binding (Figures 4.12 & 4.13). This shows that the self-assembly of RBPMS-A into higher order oligomers is not required for its splicing activity but may be linked to increased RNA binding. It also suggests that $\pi - \pi$ interactions but not cation – π homotypic interactions are required for activity. The issue that remains to be resolved is why the RKR/A mutant is still able to regulate splicing while the aromatic mutants and the RKR/E mutant cannot. RBPMS-A was also observed to

form hetero-oligomers with various other proteins in HeLa nuclear extract, as well as forming PPIs with various other RBPs, which the $\Delta C20$ mutant was unable to do (Yang *et al*, 2023). Thus, it is possible that the disrupted activities of the aromatic mutants and the RKR/E mutant may not solely be due to lack of self-assembly, but also due to a loss of PPIs/hetero-oligomer formation. This was suggested by the pilot study with the WWW/A and YYY/A mutants carried out by a former undergraduate Part III student which found that these mutants, and particularly the WWW/A mutant, lost interactions with a number of co-regulators (discussed in Section 4.7) (Lee, Part III Dissertation, 2022). It is possible that the retention of activity by the RKR/A mutant is due to it maintaining these heterotypic interactions and forming hetero-oligomers. An extensive proteomic study of the RBPMS-A mutants was not achieved during the timeframe of this study. The western blot results from the pull-down in Section 4.6, however, did hint at a difference between the basic mutants and the aromatic mutants, as well as between the RKR/A/E mutants. The RKR/A mutant was shown to maintain an interaction with RBFOX2 whereas the aromatic and RKR/E mutants did not (Fig.4.24). The RKR/E mutant was, however, shown to still interact with RBM14. RBFOX2 has been observed to co-regulate a number of events with RBPMS, whereas RBM14 was not shown to have any effect on splicing on RBPMS regulated events when knocked down (Jacob *et al*, 2022; Yang *et al*, 2023). Thus, one reason for RKR/A being active whilst the RKR/E mutant is inactive is the maintenance of heterotypic interactions with co-regulator proteins. The phosphomimetic TT/E mutant here was shown to still interact with RBFOX2, despite this protein being inactive. Previously, this mutant was shown to cause an occlusion of the RRM of RBPMS-A via an NMR experiment whereby the TT/E region looped back and interacted strongly with the RNA-binding interface of the RRM and thus was predicted to exclude RNA binding, whereas the same region in the TT/A mutant did not form this structure (Barnhart *et al*, 2022). This occlusion is likely the reason why this mutant is inactive despite being able to maintain some heterotypic interactions, whilst the TT/A protein can still regulate splicing.

Yang *et al* hypothesised that the heterotypic interactions formed by RBPMS-A at its repression target RNA, *Tpm1*, were required for its function (Yang *et al*, 2023). It was proposed that RBPMS-A can form “binding-region condensates” in an analogous manner to those described for TDP-43, for its repression of exon 3 (Figure 6.1). These condensates were proposed to consist of co-regulator proteins such as MBNL1 and RBFOX2, and was anticipated to require 5-10 subunits of RBPMS-A. The requirement of this scale of self-assembly suggests that the large complex which the wild-type protein can form *in vitro* may not actually be required for activity, and instead smaller homo-oligomers that can interact with co-regulators may be a more important requirement. Therefore, the distinguishing characteristics between the active versus inactive mutants may well be their ability to interact with co-regulators and form hetero-oligomers with them. The formation of hetero and homo-typic interactions are often

linked, but not always, as was demonstrated by the tyrosine mutants of RBFOX1 (Ying *et al*, 2017). The authors identified tyrosine mutants of RBFOX1 which had impaired self-assembly but could still interact with the LASR complex. Although for these mutants of RBFOX1 their splicing activity was mainly impaired, the authors did find that for a subset of exons, these self-assembly impaired mutants did retain splicing repression and some activation activity. RBPMS-A may behave similarly to how RBFOX1 does at such exons, whereby it is not required to form large homo-oligomers for its splicing function but does require the formation of PPIs.

-RBPMS



+RBPMS

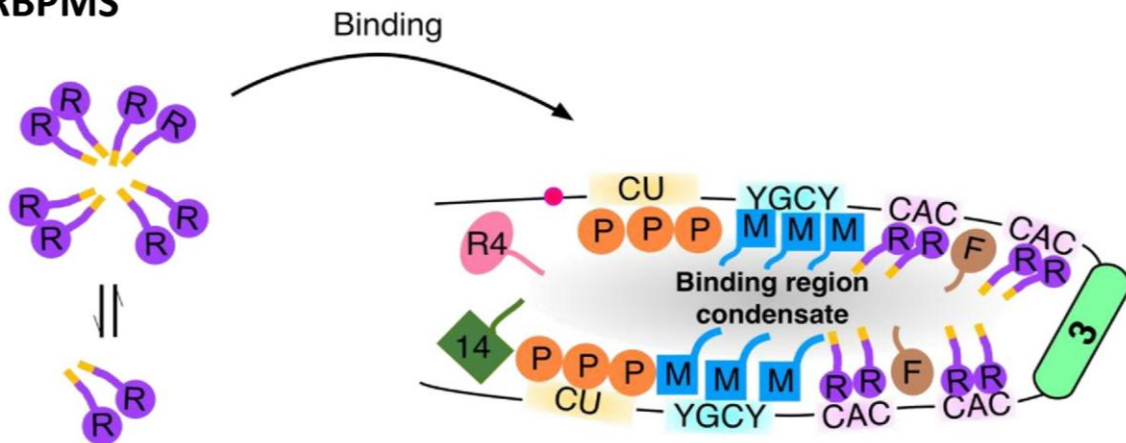


Figure 6.1 Splicing repression by RBPMS: The proposed model for splicing repression by RBPMS-A at Tpm1 exon 3 adapted from Yang *et al*, 2023. Without RBPMS-A, the co-repressors PTBP1 and MBNL1 are shown to still bind around exon 3 but cannot inhibit the binding of U1 and U2 snRNPs. RBPMS-A is proposed to form a “binding region condensate” with MBNL1 and RBFOX2 and other RBPs which blocks the binding of U1 and U2 snRNPs. The shapes represent the different proteins, P = PTBP1, M = MBNL1, R = RBPMS, F = RBFOX2, R4 = RBM4, 14 = RBM14, and the yellow ovals are U1 and U2 snRNPs. The binding sites of the RBPs are given, and the red dot represents the branchpoint.

In addition to the effects that homotypic and heterotypic interactions may have on the activity of these mutant proteins, likewise, their ability to bind RNA is also important. UV crosslinking studies were employed to assess whether these mutants were still able to interact with target RNA. These experiments were performed either in buffer only, or in HeLa nuclear extract to recreate an

environment where RBPMS must compete for binding to RNA with other proteins. Surprisingly, UV crosslinking efficiency was reduced across all mutants in buffer and nuclear extract (Figures 4.12 & 4.13). Even the two active mutants, TT/A and RKR/A showed reduced crosslinking to RNA in HeLa nuclear extract relative to wild-type. The RKR/A mutant also had reduced crosslinking in the buffer only samples when there was no competition from other proteins to bind to the RNA. Some RBPs have been shown to have poor crosslinking efficiency, thus UV crosslinking can be an inefficient technique, meaning it may not always capture all RBPs bound to an RNA, and the amount of each protein bound to this construct may be an underestimate (Esteban-Serna *et al*, 2023). This inefficiency often gives rise to high signal-noise ratios, as can be seen from the large error bars from the UV crosslinking studies in Figures 4.12 and 4.13. Despite these limitations, it appears that wild-type RBPMS-A has greater RNA binding capability than the RKR/A and TT/A proteins. Given that these experiments were conducted in similar conditions to the *in vitro* splicing assays, in which these two mutants were active, it is likely that a large fraction of the binding detected by UV crosslinking is not necessary for activity. The RKR/E mutant showed no RNA crosslinking in buffer or nuclear extract, which indicates why this mutant is not active, despite having similar assembly behaviour to the RKR/A protein and potentially retaining some PPIs. The reduced RNA binding ability of the TT/A and RKR/A mutants suggests that RBPMS-A itself may not require multimerization on multivalent RNAs and instead can form smaller complexes and interact with other proteins that are bound in similar locations along the RNA. Further tests measuring direct RNA binding, such as electrophoretic mobility shift assays (EMSAs), are required to fully understand the RNA binding abilities of RBPMS-A wild-type and its mutants. The previous work on RBPMS-A identified that the loss of the final twenty amino acids led to a reduction in RNA binding (via UV crosslinking studies) and this was interpreted to be linked to the loss of activity and the loss of heterotypic interactions seen. This interpretation was due to the observation that for this Δ C20 mutant, RNA binding was possible in a non-competitive environment (buffer only) but was lost in the nuclear extract (Yang *et al*, 2023). This was hypothesised to mean that this mutant could no longer interact with other proteins bound on the RNA, due to the loss of PPIs seen in the proteomics study and thus lost RNA binding ability and therefore activity. The data described here partially brings this hypothesis into question, as we see that the level of UV crosslinked related RNA binding observed for wild-type is in fact not required for activity.

Phosphorylation impacts oligomerisation and activity

The TT/E mutant had reduced activity with activation and repression splicing reporters compared to the wild-type protein and was also shown to have reduced oligomerisation in the mass photometry (MP) experiments (Figures 4.5, 4.9, 4.17 & 4.20). Previously, the T113E and T118E residues were shown by NMR to interact with the RRM of RBPMS, causing a looping back of the C-terminal tail over the RRM

(Barnhart *et al*, 2022). This looping back of the C-terminus may limit the interaction it has with other subunits. Thus, the assembly of the TT/E mutant would be affected, as well as a reduction in RNA binding as mentioned in the above section. Additionally, the loss of oligomerisation may be due to increased negative charge, leading to repulsion between other negatively charged residues of RBPMS subunits. In the MP experiments carried out here, the TT/A mutant had reduced higher order assembly of >20mer formation compared to the wild-type protein, but previous experiments carried out with AUC showed that this mutant behaves more similarly to wild-type in terms of higher order assembly (Barnhart *et al*, 2022). This inconsistency may be purely due to MP having a higher sensitivity to minor changes in assembly which AUC may not have detected. AUC and MP experiments can be repeated with the same set of purified proteins to confirm whether the difference between assembly of wild-type and TT/A is consistent. Given that self-assembly was found here to not be required for activity, it is likely that it is not this aspect of the behaviour of the TT/E mutant that results in a loss of activity. This mutant was also shown to still maintain interactions with RBFOX2. Thus, the most likely explanation for the lack of activity of this mutant, and indeed the RKR/E mutant, is the loss of RNA binding, shown here (Fig.4.12) and previously in the NMR experiment and EMSA experiments (Barnhart *et al*, 2022).

Extreme higher order assembly of RBPMS-B negatively impacts splicing activity

With RBPMS-B, mutation of the region predicted to be responsible for its extreme higher order assembly (the HSA region) generally led to an increase in activity compared to wild-type and became closer to RBPMS-A, particularly with splicing repression events (Figures 5.3, 5.4 & 5.7). This was true for the P2/P3 mutants and the Δ VFVTY mutant but not for the Δ QQ mutant or the HSA/A mutant, as these mutants had little activity. This shows that some of the residues of this region may be involved in the splicing activity of RBPMS-B, and this was again implied by the effect of mutating the aromatic residues of RBPMS-B (Fig.5.9). The increased activity of the P2 mutant and reduced potential higher order assembly (Figures 5.7 & 5.21) suggests that the β sheet forming capabilities of the unique C-terminus of RBPMS-B antagonise splicing activity by causing the formation of extreme higher order oligomers, thereby reducing the effective available concentration of RBPMS-B. Whether these oligomers are as extreme in cells as the large spherical structures of the full-length protein *in vitro* or indeed the fibril-like structures that the peptide alone can form requires further investigation. Immunofluorescence studies observing the localisation of *pan* RBPMS has been performed in PAC1 cells, and RBPMS was mainly nuclear and evenly dispersed rather than large foci being observed, which suggests that perhaps RBPMS-B may not be forming large aggregates in cells (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019; Huang, unpublished). This extreme higher order assembly potential, however, does point towards RBPMS-A and RBPMS-B having different biophysical properties

linked to their splicing activities. The purification of the RBPMS-B P2 mutant showed it behaved very similarly to RBPMS-A (Fig.5.21), whilst RBPMS-B wild-type formed complexes too large to be purified in the same way. This suggests it may only be this HSA region causing a difference in biophysical properties of the two isoforms, and that the rest of the C-terminus of RBPMS-B may act in a similar manner to RBPMS-A, i.e. may also be forming cation – π and/or π – π interactions. The C-termini of RBPMS-A and RBPMS-B show little similarity of either amino acid sequence or compositions of amino acids (Fig.5.1). They both contain aromatic residues, however, with RBPMS-A having WWF residues and RBPMS-B having FFFFY residues. In both isoforms, the aromatic residues in their unique C-termini, and the Y141 – Y151 cluster before the unique tails were found to be important for activity. This hints that π – π interactions may also be important for activity of RBPMS-B, but the importance of these interactions for its higher order assembly remains to be investigated.

AlphaFold models of different oligomerisation states of the RBPMS isoforms and the RBPMS-B P2 mutant suggest that the HSA region of wild-type RBPMS-B may be forming important interactions between different subunits (Fig.6.2) and may be stabilising these assemblies. The HSA region is observed to form anti-parallel β sheet structures in the AlphaFold models, and although they are not seen in the tetrameric form, the region still maintains similar interactions between subunits. These interactions are mainly between the polar amino acids, glutamine and asparagine, although they also include non-polar interactions from F204 and L192. For the P2 mutant, at each oligomerisation stage there is no observable anti-parallel β sheet structure forming with the HSA region, although this region is still mediating interactions between subunits in the 10mer model. The interactions formed here appear to be between phenylalanine and arginine residues of different subunits, which resembles the types of interactions observed between the subunits in the higher order assembly of RBPMS-A (Fig.6.2), lower panel). These models further indicate similarities between the P2 mutant of RBPMS-B and wild-type RBPMS-A. The importance of these residues for assembly and activity can be investigated for assembly with mutagenesis studies in a similar manner as was done with the RBPMS-A mutants. It should be noted that the AlphaFold modelling of the C-terminus of the proteins showed very low confidence at each oligomerisation stage, hence further experimental investigation into the nature of the assemblies of the proteins and the necessary residues is needed.

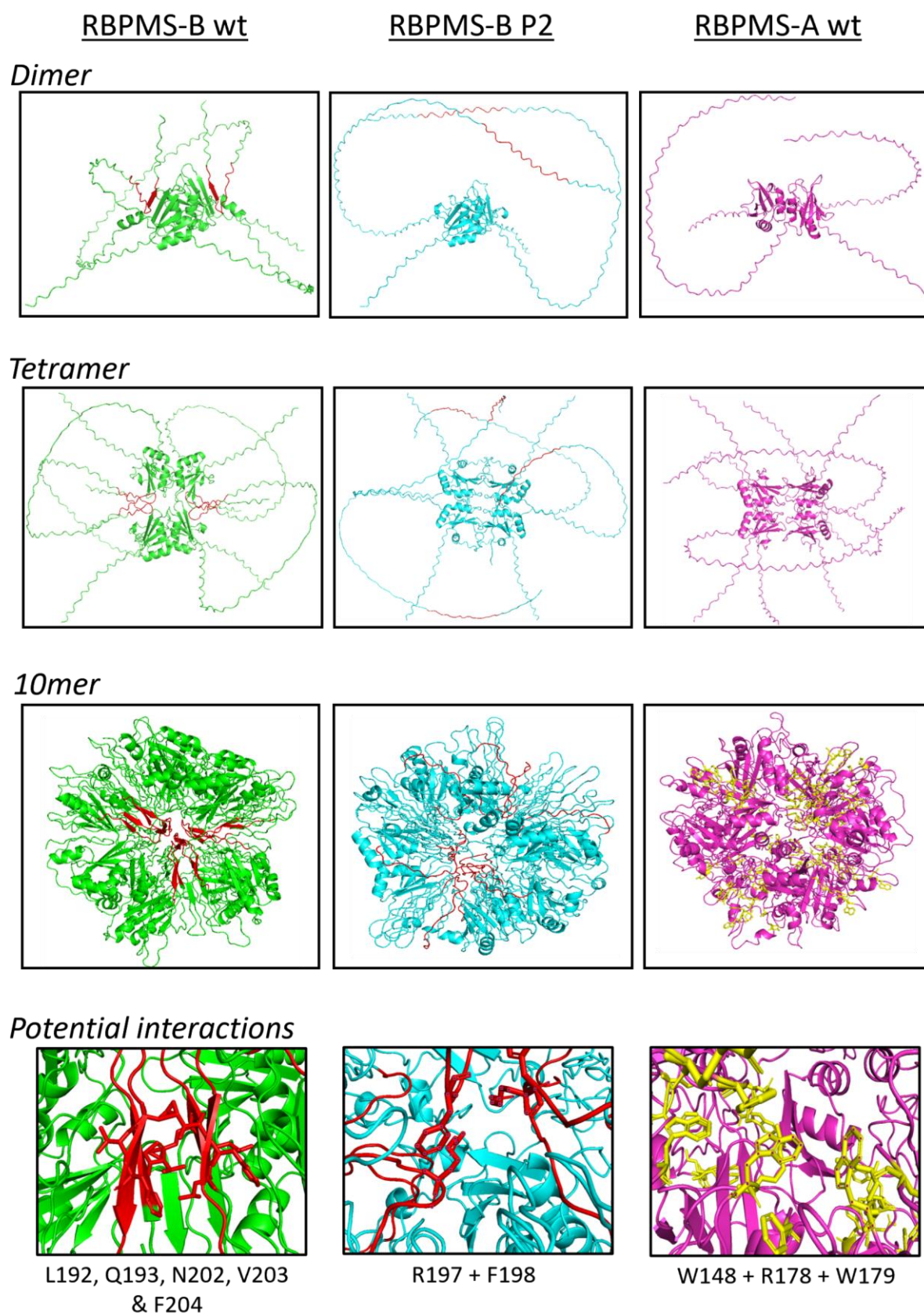


Figure legend on the next page.

Figure 6.2 Assemblies of RBPMS-A and RBPMS-B: AlphaFold 3 models of wild-type RBPMS-B (wt) (green), RBPMS-B P2 mutant (cyan) and wild-type RBPMS-A (magenta) at different oligomerization states. For wild-type RBPMS-B and the P2 mutant, the HSA regions have been highlighted in red. For the 10mer assemblies of RBPMS-A, the aromatic and basic residues found to be required for assembly have been highlighted in yellow. Interactions between residues of different subunits for the HSA regions (RBPMS-B and P2) or the aromatic/basic residues for the 10mer species are shown in the bottom panel, with the exact residues shown to be interacting across subunits displayed underneath the images. Images taken with Pymol.

6.12 TYPES OF INTERACTIONS REQUIRED FOR RBPMS OLIGOMERISATION

Cation – π and π – π interactions may be required for oligomerisation of RBPMS-A

Assembly of RBPs is often mediated by multivalent interactions such as electrostatic, cation – π and π – π interactions from the so-called “sticker” regions of IDRs (Fig.1.10) (Mittag & Pappu, 2022). These regions are thought to drive phase separation, and other residues of IDRs are thought of as “spacer” regions that connect the different stickers together. For example, in the low-complexity region of hnRNP A1, evenly spaced aromatic residues were identified to be “stickers” and driving assembly, whilst the other non-aromatic residues were “spacers” (Ginell & Holehouse, 2023). Mutagenesis of these regions can identify which residues are important for the interactions (i.e. are forming the “stickers”), and testing the functions of such mutants provides insight into how important this assembly behaviour is for physiological activity (Alberti *et al*, 2019). This strategy was employed to identify which residues of the C-terminus of RBPMS-A were required for this higher order oligomerisation behaviour and if the same residues were also essential for activity (Fig.4.2). The differences between wild-type and the basic and aromatic mutants were already suggested during their purification. The purification of the wild-type RBPMS-A protein was optimised due to the observation that it formed higher order oligomers that had increased avidity for the agarose matrix of the HisTrap columns used in purification. The aromatic and basic mutants were observed to be fully eluted in the first fractionation step, likely due to decreased oligomerisation resulting in lower avidity for the chromatography matrix (data not shown). Wild-type, TT/A and TT/E proteins were on the other hand, eluted over five 7 ml fractions (Fig.3.8). The basic/aromatic mutants were therefore purified using a protocol that had originally been adapted for the Δ C20 mutant which had smaller fractionation steps to ensure clean fractions of these mutants could be selected (see section 4.3). This suggested that the basic and aromatic residues were needed for assembly, but T113 and T118 were not. The aromatic and basic residues of the C-terminus of RBPMS-A were indeed found to be required for higher order assembly in the mass photometry (MP) experiments, at both low and high protein concentration (Figures 4.16 – 4.20). Aromatic residues can take part in cation – π and π – π interactions and basic

residues take part in cation – π interactions, thus it is possible that these residues are involved in similar interactions in RBPMS-A for its self-assembly (Gallivan & Dougherty, 1999; Gromiha *et al*, 2004; Nott *et al*, 2015; Wang *et al*, 2018; Sabari *et al*, 2020; Mittag & Pappu, 2022). To further investigate if these residues participate in cation – π and/or π – π interactions, the residues could be substituted for other residues that form weaker cation – π / π – π interactions, such as mutating W/Y to F, and R to K residues. These mutants would be expected to form weaker interactions and thus oligomerisation would still be impaired, but not to the same extent as the alanine mutants. It is also possible that the YYYYW region may be interacting with the residues of the final twenty amino acid tail (RKR and WWF) intramolecularly. If this was the case, however, then assembly of RBPMS-A into higher order oligomers would likely not be affected by mutations. Additionally, the AlphaFold predictions with the 10mer model of RBPMS-A show that these residues are interacting intermolecularly rather than intramolecularly (Fig.6.2). The basic residues mutated in the RKR mutants are the only positively charged residues in the C-terminus of RBPMS-A, which contains five glutamine residues (Fig.4.2) meaning without these basic residues the C-terminus becomes much more negatively charged, particularly with the RKR/E mutant. This alteration of the overall charge of the C-terminus may disrupt the intramolecular interactions (and intermolecular interactions) that the RKR/A and RKR/E mutant form leading to a loss of higher order assembly, rather than a disruption specific of cation – π bonds. The potential mutation of the R residues to K discussed above may address whether loss of intra- and intermolecular interactions is due to loss of cation – π interactions or an increase in the overall negative charge of the C-terminus of RBPMS-A.

Non-specific interactions with RNA can affect higher order assembly of RBPMS

RNA is known to affect higher order assembly of RBPs and drive LLPS and it is present in many biocondensates in eukaryotic cells (Ule & Blencowe, 2019; Sabari *et al*, 2020; Mittag & Pappu, 2022). The effect of RNA on phase separation/assembly of RBPs is often biphasic, with low concentrations aiding or inhibiting phase separation, and increasing concentrations having the opposite effect (Gasior *et al*, 2019; Babinchak *et al*, 2020; Dutagaci *et al*, 2021). Often, RNA initially leads to increased assembly, before a decrease at a certain critical concentration due to increased repulsion from negative charges. The valency of the RNA also affects assembly in a biphasic manner, where multivalent RNA promotes higher order RNP networks, but at very high RNA concentrations, each protein will be more likely to be binding to a different RNA. This results in higher order assemblies being disrupted. In the case of RBPMS-A, increasing RNA appeared to diminish the more extreme levels of higher order assembly observed (>20mer) (Fig.4.21), but with the multivalent longer RNA, at 1800 nM, these >20mer species became more prevalent again, demonstrating a biphasic response with the longer substrate. At 1800 nM of RNA, for the 3xCAC₂ RNA, binding of 10 μ M RBPMS-A to RNA should be

stoichiometric as the 1800 nM 3xCAC₂ RNA has three binding sites for RBPMS dimers, allowing most RBPMS-A molecules to bind to the RNA at this concentration. Sequence specificity did not seem to be a factor in the response of RBPMS-A to RNA, however, as the 3xCCC₂, which RBPMS has been shown not to bind, shows a similar effect as 3xCAC₂ (Fig.4.21) (Yang *et al*, 2023). This points towards RBPMS-A forming non-specific interactions with RNA, and its these interactions that affect its higher order assembly, which may become more numerous with longer RNA, aiding higher order assembly when introduced at a higher concentration. The initial decrease in assembly upon lower concentration of RNA, or when using the shorter RNA substrate suggests lower concentration and shorter RNA substrates may disrupt the protein-protein interactions, due to the formation of new RNA-protein interactions. It is possible that as concentration of the longer RNA increases, the RNA-protein interactions formed along the RNA molecule may bring smaller assemblies of RBPMS-A together and thus allows for more protein-protein interactions to form. The WWW/A mutant did not show any change upon addition of RNA, likely due to the mutations introduced diminishing any higher order assembly beyond the effect that RNA may have (Fig.4.22). This and the wild-type results suggest that the non-specific interactions that RBPMS-A has with RNA may only effect more extreme higher order assembly of >20mer. In the UV crosslinking studies, wild-type and the TT/A mutant showed non-specific crosslinking with the mCAC mutant of TM3 RNA in the crosslinking experiment when performed in buffer, but the RKR/A mutant did not (Fig.4.13, lanes 26 – 28). This points towards the RKR residues forming non-specific interactions with RNA, hence these interactions are lost when these residues are mutated. Since the amount of RNA binding observed is likely not required for function, it suggests that these non-specific interactions are also not required. Further experiments investigating RBPMS-A wild-type and RKR/A protein binding to non-specific RNA, such as EMSAs and further UV crosslinking experiments, may help to answer if these residues are indeed mediating non-specific RNA-protein interactions.

RBPMS-B extreme higher order assembly is dependent on its β -sheet structured tail

For RBPMS-B, its unique C-terminal tail was predicted to be aggregation prone due to its seventeen amino acid HSA sequence (residues 192 – 208, Figure 5.3A). AlphaFold predicted that the HSA region formed antiparallel β sheet structures, whereas the rest of the C-terminus for RBPMS-A and RBPMS-B was predicted to be unstructured (Fig.5.6). Proline substitutions were introduced to this region as these have been reported to be breakers of amyloid fibrils formed from antiparallel β sheets, and these mutations (P2 and P3) were predicted to disrupt this β sheet structure (Fig.5.6) (Conchillo-Solé *et al*, 2007). A peptide of this region was indeed found to be forming anti-parallel β sheet secondary structure via circular dichroism experiments and was observed to form fibril-like structures *in vitro* (Figures 5.10 – 5.14). AlphaFold predicted that this HSA peptide may stack in a similar way as the A β 40

peptide which is been well-documented in forming fibrils, whereas the P2 mutant and β sheet forming peptide from the RRM of RBPMS did not stack in the same way (Fig.5.15) (Xue *et al*, 2017). This further confirms that this region of the HSA peptide has a propensity to form fibrils *in vitro*. The full-length protein was not observed to form these fibrils *in vitro*, and instead formed large, spherical structures that may contain as many as ~579 dimers of RBPMS-B (Fig.5.20), whereas similar methods for RBPMS-A gave structures that likely contained ~16 dimers. This demonstrates an extreme difference between the potential higher order oligomerisation capabilities of these two isoforms. To confirm this, experiments for each isoform observing their assembly need to be performed in parallel under the same conditions.

6.13 ALTERNATIVE SPLICING OF IDR REGIONS OFTEN MEDIATES CHANGES IN ASSEMBLY BEHAVIOUR OF RBPS

Various splicing factors/RBPs have been identified where the isoforms have different activities, such as RBFOX2, whose isoforms were identified in adult rat brain tissue and determined to have different splicing targets and therefore different functions (Partridge & Carter, 2017). The difference in activity of the isoforms of RBPMS is due to their unique C-terminal tails, encoded for by alternative exons (Fig.5.1), suggesting that the activity of RBPMS can be modulated via AS of its exons 7 and 8. The isoforms produced differ in splicing activity, likely due to a difference in oligomerisation and biophysical properties. AS of IDRs is common and often conserved among regulators of splicing and can result in alterations to splicing activity, and thus represents another layer of AS regulation (Gueroussov *et al*, 2017a). These changes to IDRs can frequently result in alterations in propensity for phase separation and aggregation, as was seen for the splicing regulator hnRNP DL (Batlle *et al*, 2020; Garcia-pardo *et al*, 2023; Giudice & Jiang, 2024). For this protein, AS results in three isoforms with either two IDRs (N-terminal and C-terminal), one IDR (C-terminal) or none, which all have different phase separation behaviours. The double-IDR isoform (hnRNP DL1) can undergo LLPS whilst the single IDR isoform (hnRNP DL2) forms amyloid-like fibrils, and the isoform with no IDRs (hnRNP DL3) does not undergo LLPS or aggregation. The LLPS behaviour of hnRNP DL1 was found to be mediated by potential cation – π interactions between arginine and tyrosine residues from its N-terminal IDR (Batlle *et al*, 2020). The fibril formation of hnRNP DL2 occurs in physiological conditions and is mainly driven by polar contacts from its Y/G rich C-terminal IDR which is well conserved across vertebrate species, as is the fibril forming HSA region of RBPMS-B (Fig.1.9) (Garcia-pardo *et al*, 2023). These two isoforms have been shown to have different functions, with hnRNP DL2 being the predominant form and regulating AS networks. This isoform was shown to be able to bind RNA in its fibril form, suggesting that the amyloid-like structures are functional. Disease-associated mutations of the fibril-forming domain were

shown to reduce the formation of these fibril structures, further pointing towards these structures being the functional state of the protein (Garcia-pardo *et al*, 2023). The existence of these two isoforms with different assembly behaviour was hypothesized to be a way in which different structural assemblies can evolve for proteins. These two hnRNP DL isoforms therefore show similarity with the isoforms of RBPMS found in differentiated VSMCs, whereby one can form dynamic assemblies (hnRNP DL1 & RBPMS-A) and the other may be forming more stable, larger structures (hnRNP DL2 & RBPMS-B). The AS of RBPMS at exon 7/8 may therefore serve to regulate the activity of RBPMS through the introduction of the antagonistic C-terminal tail of RBPMS-B. Understanding the conditions and regulators behind exon 7 inclusion over exon 8 may reveal the purpose this differential isoform production serves.

6.2 MECHANISMS OF SPLICING ACTIVATION VERSUS REPRESSION

6.21 SUCCESSFUL PRODUCTION OF AN *IN VITRO* SPLICING ACTIVATION ASSAY

The previous mechanistic splicing studies of RBPMS have focused on splicing repression using the *Tpm1* construct (Barnhart *et al*, 2022; Yang *et al*, 2023). The first results chapter (chapter 3) focused on the production of a complementary splicing activation assay with *PTPRF*. This assay was successful in demonstrating the ability of RBPMS-A to activate splicing *in vitro*. Hence, using this assay in conjunction with the repression assay with *Tpm1* previously established provides a unique chance to elucidate and compare the mechanisms of both splicing activation and repression for a single protein. *In vitro* splicing assays allow for the detailed analysis of how an RBP may be regulating a particular splicing event (Coehlo & Smith, 2014). These assays permit the control of both the *trans* and *cis* elements introduced, without issues such as differing expression between proteins/RNAs. This allows for any observational changes in splicing to be discerned as directly related to changes made to the proteins/RNA, which allows for great insight into the mechanisms of splicing. The *in vitro* splicing assays for RBPMS also allow the direct comparison of splicing activation/repression with biophysical properties of the proteins and their PPIs. One limitation of *in vitro* splicing assays is that they use pre-transcribed RNA, which means any alternative splicing that occurs co-transcriptionally will not be observed, and co-transcriptional splicing is a well-documented phenomenon (as discussed in sections 1.2 & 1.3). RBPMS-A and its mutants, however, behaved very similarly in both the transfection assays and the *in vitro* splicing assays. Wild-type RBPMS-A was able to activate and repress splicing events *in vitro*, showing that these assays worked well for this purpose and the lack of co-transcriptional splicing did not appear to be an issue. This is despite the fact that *Tpm1* was one of the first documented cases where RNA Polymerase pausing was shown to affect alternative splicing (Roberts *et al*, 1998). Another potential limitation of the use of these *in vitro* assays to study the mechanism of action of the tissue specific splicing regulator RBPMS is that they are performed in HeLa nuclear extract, which is not derived from SMC tissue. It is true that there may be some interactions with VSMC specific proteins that are missed, but the proteomics study with RBPMS in HeLa nuclear extract previously identified interactions with many core splicing factors that are widely expressed in many cell types, including VSMCs (Yang *et al*, 2023). Furthermore, many of the interactors were also observed in the BioID study of RBPMS performed in human induced pluripotent stem cell (hiPSC) derived cardiomyocytes, including many hnRNPs, core spliceosomal proteins, and MBNL1 (Gan *et al*, 2024). These interactions imply that RBPMS acts by interacting with various ubiquitously expressed splicing proteins to perform its function. Thus, using HeLa nuclear extract is a sufficient model to understand how RBPMS regulates splicing in differentiated VSMCs. This was confirmed by the cell-type specific splicing pattern switch of

Tpm1 exon 3 repression upon addition of RBPMS-A to HeLa nuclear extract (Yang *et al*, 2023). Whether addition of RBPMS-A would be sufficient for a cell-type specific switch activation targets was unknown. It was possible that for specific activation, RBPMS-A would require cell-type specific co-activators, or indeed co-transcriptional regulation. In the activation assays produced here, however, we were able to show the RBPMS-A addition to HeLa nuclear extract was able to reproduce the VSMC specific splicing pattern of *PTPRF* exon 23 inclusion. Furthermore, this was shown to be a specific response, as the mutation of possible RBPMS-A RNA binding sites in the construct abrogated this RBPMS-dependent behaviour.

6.22 POTENTIAL MECHANISMS OF SPLICING ACTIVATION

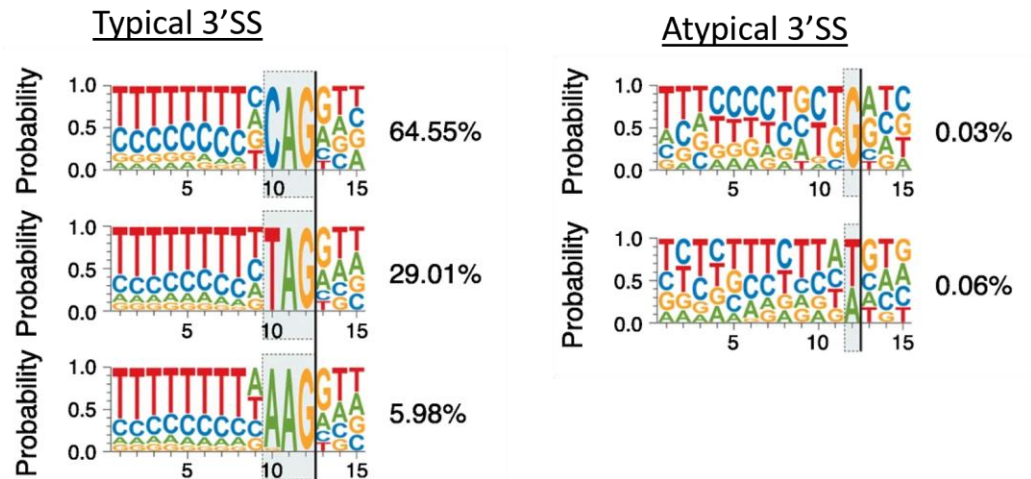
What the PTPRF assay reveals about the mechanism of splicing activation by RBPMS-A

The splicing activity of RBPMS has been shown to be RNA position dependent (Nakagaki-Silva *et al*, 2019; Jacob *et al*, 2022). RBPMS adheres to similar position-dependent rules of other RBPs, whereby the binding of RBPMS upstream and within regulated exons is associated with exon skipping, whilst the binding in the downstream intronic region is associated with exon inclusion (Ule & Blencowe, 2019). These position dependent rules point towards RBPMS having different mechanisms for splicing activation and repression. None of the mutants of RBPMS-A produced in Chapter 4, however, showed a preference for activation over repression or vice versa. Hence, these mutants are not useful for the investigation into the mechanisms of activation versus repression. This result demonstrates that the effect that these mutants had, i.e. on homotypic interactions, RNA binding, and potentially on the formation of heterotypic interactions, was universal for both activation and repression. Splicing repression was hypothesised to be aided by the formation of localised condensates formed of multiple copies of RBPMS, MBNL1 and RBFOX2 (Yang *et al*, 2023). MBNL had binding sites co-localising with RBPMS binding sites on the *Tpm1* construct, and knockdown experiments showed that both proteins often regulated the same events. These localised condensates were proposed to prevent proper U1 and U2 binding to the 5'SS and 3'SS of *Tpm1* exon 3, and thus the exon would not be recognised by the spliceosome and is skipped (Fig.6.1). This idea came from observations that the addition of RBPMS-A to nuclear extract containing a *Tpm1* construct consisting of exons 2 and 3 and the downstream intronic region prevented the formation of any ATP-dependent spliceosome complexes, showing assembly was stalled and did not progress. Furthermore, U1 and U2 binding to this transcript was shown to be impaired upon addition of RBPMS-A. The location of the potential binding sites of RBPMS on *Tpm1* agree with this model as they are located within and around exon 3. The CACs found to be important for splicing activation of exon 23 (e23) of *PTPRF* are located further downstream, towards the constitutive exon 24 (Figures 3.1 & 6.3). This means that bound RBPMS-A is located far from the

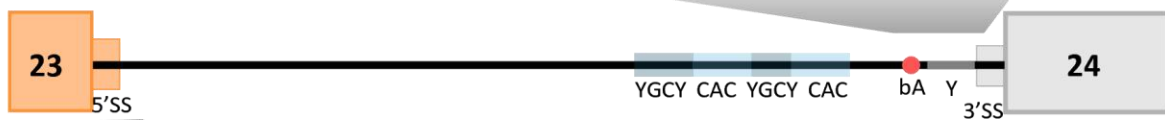
5'SS of exon 23 (~600bp) and thus RBPMS-A is unlikely to be directly interacting with U1 snRNP binding at this site unless there is looping out of the downstream intron. The 5'SS sequence of e23 is non-consensus (Fig.6.3), thus U1 snRNA may not interact as efficiently as it would at a consensus 5'SS. Looping out of intronic regions to promote exon inclusion has been suggested for exon activation by the RBFOX proteins, where they bind to intronic ISEs over 500bp downstream from the exon, and RNA:RNA contacts allow the formation of stem-loop structures which brings the RBFOX/LASR complex into close proximity with the splicing machinery around the alternative exon (Lovci *et al*, 2013; Ying *et al*, 2017). This mechanism, however, is unlikely in the case of *PTPRF* e23, as the deletion of 400bp of this downstream intronic region did not affect basal or RBPMS-dependent activation of exon 23 (Figures 3.5 & 3.7).

Given the proximity of the CAC clusters which RBPMS-A requires for its activity to the branch point, polypyrimidine tract and 3'SS of the constitutive exon 24 (Fig.6.3), RBPMS-A may instead act to promote splicing of e23 by recruiting the U2AF and the U2 snRNP components to this site to promote the splicing of this downstream exon (Fig.6.4). Equally, promoting of U2 snRNP/U2AF binding could result in increased inclusion of exon 22, although this is not what was observed. The RBP hnRNP A1 was observed to bind downstream of exon 11 of the proto-oncogene *RON*, in the constitutive exon 12 (Bonomi *et al*, 2013). This was found to promote the inclusion of exon 11 through the antagonisation of the activity of SRSF1, which also bound to exon 12 but promoted exon 11 skipping. This example shows that there is precedence for such a model where binding of an RBP downstream of a regulated exon, towards or within the constitutive exon can result in exon inclusion. With the repression target *Tpm1*, RBPMS-A was shown to displace U2AF and U2 snRNP from their strong branchpoint/polypyrimidine tract binding sites around exon 3 (Mullen *et al*, 1991; Yang *et al*, 2023). The polypyrimidine tract downstream from e23 of *PTPRF* is less U-rich and is interrupted with purines (Fig.6.3) and thus may be a weaker polypyrimidine tract, resulting in U2AF65 having reduced binding affinity (Coolidge *et al*, 1997; Sickmier *et al*, 2006; Shao *et al*, 2014). Thus, in this case, RBPMS-A may act to recruit U2AF/U2 snRNP to the branchpoint and polypyrimidine tract to promote exon inclusion, and this would explain why this exon is mainly skipped if RBPMS-A is not present (Figures 3.3, 3.6 & 3.12 – 3.14). Promotion of U2AF binding to weak polypyrimidine tracts to increase exon inclusion has been observed for the RBP SAM68 which recruits U2AF to a weak polypyrimidine tract upstream of the v5 exon of *CD44* leading to exon inclusion, and another RBP YB-1 acts in a similar manner to also promote exon v5 inclusion (Tisserant *et al*, 2008; Wei *et al*, 2012). This model of activation by RBPMS-A is supported by the proteomic screen of RBPMS-A that identified interactions with U2 snRNP components, including SF3B1 and the branch point binding protein SF1. The CAC sites found to be important for RBPMS-dependent activation of *PTPRF* exon 23 are ~40bp upstream of the branchpoint

sequence. U2 snRNP has been found to have a footprint on the RNA that extends ~40bp upstream of the branchpoint sequence, and thus its feasible that RBPMS-A bound at the second CAC cluster could be interacting with U2 snRNP (Nelson & Green, 1989). Moreover, the radio-labelled splicing time-course assays with *PTPRF* indicate that that the downstream intron of e23 may be spliced first (Fig.3.9, lanes 5 – 8). The band for the intermediate product with the downstream intron spliced out (22–23|24) appears to be stronger than the band for the intermediate product with the other intron spliced out (22|23–24) at the earliest time-point (lane 6) and then appears to be fainter at the later time-points than the other intermediate product (lanes 7 & 8). This band (22–23|24) appears to be stronger in the +RBPMS-A lane than the -RBPMS-A, suggesting it may be RBPMS dependent. More time-course assays would need to be performed to confirm if this observation is consistent. RBPMS-A may also be promoting the splicing of this intron and therefore exon 23, by antagonising other proteins that may be binding to these sites, e.g. hnRNP proteins. RBPMS-A was observed to interact with various hnRNPs in the proteomic screenings and thus may sterically inhibit their binding to the downstream intron of *PTPRF* (Fig.6.4) (Yang *et al*, 2023; Gan *et al*, 2024). RBPMS may therefore act to activate splicing by recruiting positive splicing factors, and/or preventing the binding of negative regulators.



CTAATGCTGCCTGTGTATGCCCTACCTCCCC TAG | GTA



CAA | GTTAG

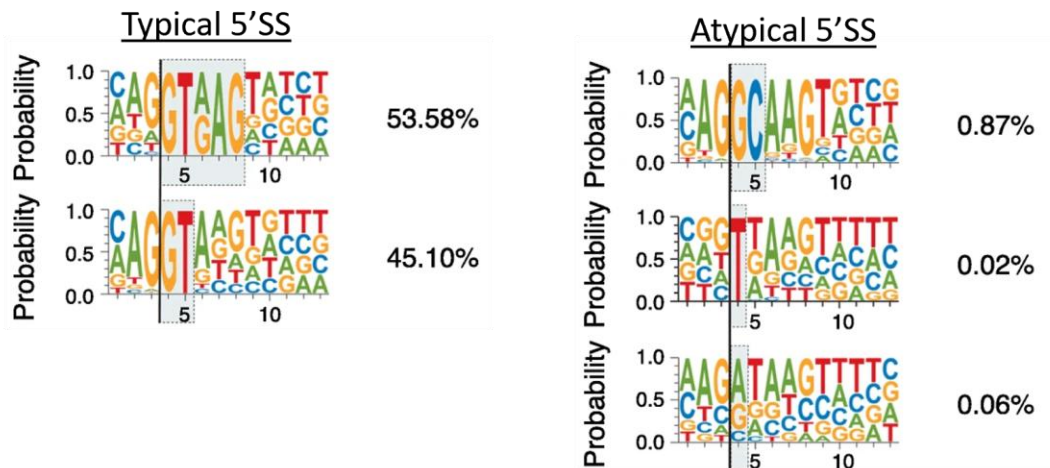


Figure 6.3 PTPRF has a weak polypyrimidine tract downstream from exon 23: Typical and atypical 5' splice sites (5'SS) and 3' splice site (3'SS) sequences and polypyrimidine tract taken from (Wan & Larson, 2018) and are compared with the 5'SS, branchpoint sequence, polypyrimidine tract, and 3'SS of PTPRF, downstream from its alternative exon 23. The suspected branchpoint sequence of this exon is underlined. Branchpoint location on PTPRF is signified with a red box (bA) and the polypyrimidine tract location is shown with a grey line and a "Y".

MBNL binding motifs are co-associated with the CAC motifs that were found to be important for RBPMS-A activation of PTPRF e23, and their mutation resulted in reduced RBPMS-dependent activation of the exon. This suggests that MBNL1/2 may be co-regulating the inclusion of this exon with RBPMS-A. This was confirmed with knockdowns of MBNL1/2 and RBPMS in differentiated PAC1 cells which showed they regulated this exon in the same direction (Yang *et al*, 2023). In fact, MBNL1/2 knockdown was found to be more potent than RBPMS knockdown for disrupting the activation of this exon. RBFOX2 was also identified by knockdown to be regulating e23 in the same direction as RBPMS-A, and a canonical RBFOX2 binding motif (GCAUG) was identified between the first and second CAC clusters. The importance of this binding site was not investigated here, but its mutation may well also lead to a reduction in RBPMS-dependent activation of e23. These observations suggest that RBPMS-A may be forming a common complex composed of RBPMS subunits and MBNL (and possibly RBFOX2), as proposed for splicing repression, but the activity of this complex on exon skipping/inclusion may purely depend on the location of its binding. The location of the RBPMS binding motifs for different activated exons is not always so far downstream and may be closer to the 5'SS (*FLNB*), or for some targets, the downstream intron is much longer (*PPFIBP1*) (Fig.3.1). This implies that RBPMS may act by a variety of different mechanisms for activation, depending on where its *cis* binding motifs are located. A recent study measuring RNA:RNA contacts mediated by RBPs observed positional dependent effects of PTBP1 splicing through facilitated by RNA looping (Ye *et al*, 2023). The authors identified that PTBP1 promoted repression of exons by forming complexes across an exon to prevent splice site binding. Exon activation was promoted via the formation of cross-intron complexes. They found that PTBP1 promotes exon inclusion when its bound exclusively in the upstream intron, or exclusively in the downstream one. PTBP1 promotes asymmetrical intron removal where the intron that the PTBP1 complex has formed across is removed first by the spliceosome. RBPMS may be regulating splicing in a similar manner, with repressed exons, the exon is looped out and splice site recognition is lost (Fig.6.1), whilst with exon inclusion, binding to the downstream intron may promote the formation of a cross-intron complex to bring splice sites closer together. This cross-intron complex may promote U1 and U2 snRNP interactions, as well as promoting the binding of U2 and U2AF binding through recruitment of these proteins. Further analysis of where RBPMS binds on more activation targets and how this binding affects splicing outcomes is required to fully elucidate these mechanisms.

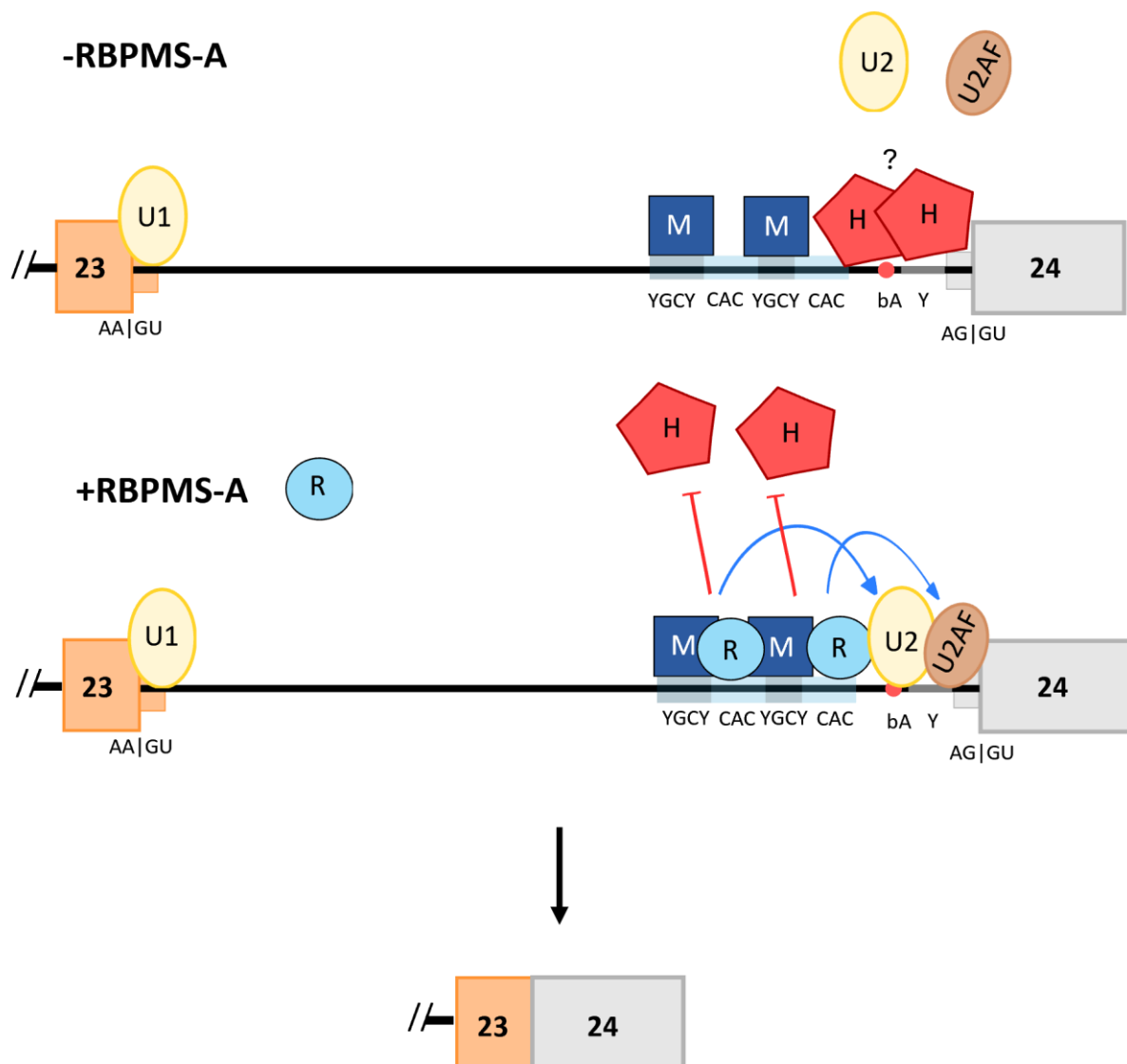


Figure 6.4 Potential activation model with PTPRF and RBPMS: A hypothetical model for how RBPMS-A may be activating inclusion of PTPRF exon 23 via promoting of splicing of the downstream intron. Given that the length of the downstream intron did not affect RBPMS-dependent activation, intron looping is unlikely, and instead RBPMS-A may be interacting with core splicing machinery at the 3' end of the intron. RBPMS-A has been shown to interact with members of the core splicing machinery including the U2 snRNP components and thus may act to recruit these and/or the U2AF proteins (U2AF65 and U2AF35) to the branchpoint, polypyrimidine tract and 3' splice site. It may also be acting to prevent the activity of inhibitory proteins binding to these regions. U1 and U1 snRNPs are depicted with yellow ovals, MBNL1/2 is depicted with a blue square (M) and RBPMS with a blue circle (R), the possible inhibitory protein(s) are indicated with a red pentagon and an H to represent hnRNP. The binding sites for RBPMS and MBNL are shown with pale blue and dark boxes with the motifs underneath. The branchpoint (bA) is shown with a red circle, the polypyrimidine tract is shown with (Y) and a grey line. The 5' and 3' splice sites of PTPRF are shown with sequences.

What the activity of RBPMS-B can reveal about splicing activation by RBPMS

Although none of the RBPMS-A mutants showed differential activity for activation versus repression, RBPMS-B does show differential activity to RBPMS-A, as it was observed to not be able to repress splicing events. RBPMS-B was not able to be produced recombinantly to be tested in the *in vitro* assays. Thus, although expression levels between RBPMS-A and RBPMS-B appeared to be similar in the transfection assays, it cannot be ruled out that differences in expression/turnover are what is causing the difference in activity. It may simply be that splicing repression requires a higher concentration of RBPMS than splicing activation does. From the titration of RBPMS-A in the activation and repression assays in Figure 3.14, however, splicing activation with *PTPRF* exon 23 actually appears to require a higher dose of RBPMS-A than *Tpm1* exon 3 repression, although this observation is not reproduced in later titration experiments (Fig.4.10). The picture gets more complex, however, as RBPMS-B also has different activities with different activated exons. For example, RBPMS-B was not able to activate *PTPRF*, but overexpression led to greater inclusion of exon 10 of *PPFIBP1* than overexpression of RBPMS-A (Fig.3.3). Overexpression RBPMS-A and B, however, showed similar levels of exon inclusion for *FLNB*. The knockdown data for each isoform showed that for two of the activation targets (*Ptpnf* and *Ppfia1*), knockdown of RBPMS-B had less of an effect than RBPMS-A knockdown, whilst with the other two targets (*Myocd* and *Ppfibp1*) knockdown of RBPMS-B led to greater skipping of the regulated exons than RBPMS-A knockdown (Fig.5.2B). This demonstrates that at least for RBPMS-B, there may be different mechanisms of activation for different exons. Different mechanisms of splicing regulation at different exons have been observed for a number of splicing regulators. For example, PTBP1 is thought to inhibit the N1 exon of *c-src* by causing a hyper-stable U1 snRNA:5'SS interaction and preventing productive cross-intron interactions, whereas the SM exon of α -actinin is repressed by PTBP1 preventing U2AF65 binding (Matlin *et al*, 2007; Wongpalee *et al*, 2016). With RBFOX1, activation of different exons apparently requires different oligomerisation states, with a 10Y/S mutant that could not self-assemble only able to activate a subset of exons, whilst the rest of the regulated exons required self-assembly behaviour (Ying *et al*, 2017). Given that RBPMS-A and RBPMS-B have been suggested to have different degrees of higher order assembly in the *in vitro* studies carried out here, this oligomerisation behaviour may be affecting which exons can be regulated by RBPMS-B. The stability of the HSA peptide structures indicates that this region of RBPMS-B may mediate the formation of stable higher order assemblies of RBPMS-B which cannot disassemble as readily as the dynamic assemblies of RBPMS-A. Since the RBPMS-A mutagenesis studies indicated that higher order self-assembly may not be a requirement for splicing activity, it might be more necessary for the subunits of RBPMS-A to be available to interact with other co-regulators and bind to RNA. If the subunits of RBPMS-B are caught up in large stable assemblies, then they will be less readily available

to interact with other proteins/RNA, thus splicing behaviour is reduced. This particular requirement may be more important at certain exons (repressed exons and some activated exons), thus leading to differential activity of RBPMS-B (Fig.6.5). Nevertheless, this does not explain why RBPMS-B activity leads to greater levels of inclusion of *PPFIBP1* exon 10 compared to RBPMS-A. In fact, the experiment with RBPMS-B and its HSA mutants with the *PPFIBP1* minigene showed that the $\Delta VFVTY$ mutant which was predicted to have reduced aggregation propensity and was normally more active than wild-type, was instead shown in this instance to have reduced activity, more like RBPMS-A (Fig.5.4). Whereas the ΔQQ mutant, which was predicted to be more aggregation prone and was normally less active, showed similar activity to the wild-type protein with this target. This would suggest that perhaps the more extreme and potentially less dynamic higher order assembly is more suitable for exon activation in this instance. This difference in the ability of RBPMS-B to regulate different exons may be due to the different locations of *cis* elements that RBPMS binds to. The binding sites of RBPMS for *PPFIBP1* have not been elucidated, but there are potential binding sites in the downstream exon that do not seem to be clustered to one particular area, unlike to the *PTPRF* CAC clusters (Fig.3.1). The downstream intron of *PPFIBP1* is also much longer than the *PTPRF* exon (~2kb versus 0.8kb), thus perhaps in this instance, RBPMS-B may be binding at various sites across the intron and forming a large complex which may bring the 5'SS and 3'SS of this downstream intron closer together to promote cross-intron recognition, leading to more efficient exon inclusion. RBPMS-A may be playing a similar role but may not assemble as efficiently as RBPMS-B, or it may be acting in a similar manner to the mechanism proposed for exon 23 inclusion of *PTPRF*. Mutants of TDP-43 that affected condensation properties were found to have differential RNA binding preferences (Hallegger *et al*, 2021). Mutants that had increased condensation propensity were found to be associated with binding motifs that were more dispersed across RNA, with separations between clusters of >100 nt. Conversely, mutants with perturbed condensation behaviour preferentially bound to clusters that were closer together. Thus, RBPMS-A/B may display similar oligomerisation dependent behaviour at certain target exons where CAC clusters are more disperse (*PPFIBP1*) thus requiring more extreme assembly, whilst at other targets, where CAC clusters are closer, this assembly may not be necessary (*PTPRF*). Testing the RBPMS-A mutants that affected higher order assembly but not activity (RKR/A) with *PPFIBP1* may help to answer whether this target exon requires oligomerisation for its behaviour. If the intron is being looped out due to multimerization of RBPMS along the RNA, the RKR/A mutant protein may not be able to do this properly as it lacks oligomerisation behaviour. If it is dependent on heterotypic interactions as proposed for *PTPRF*, then the RKR/A mutant would likely still be active. Global analysis of RBPMS-A and RBPMS-B binding targets by CLIP and their effects on splicing as well as their binding sites may shed more light on this differential activity.

6.3 POTENTIAL ROLES OF RBPMS-A VERSUS RBPMS-B

BPMS-A and its homologue BPMS2 both showed similar activities with several targets in transfection studies but although BPMS2 was also expressed in differentiated VSMCs, its expression levels (at mRNA level) were 10-fold lower than BPMS levels (Nakagaki-Silva *et al*, 2019). This suggests that BPMS2 may not be playing as an important a role in the regulation of the AS network of VSMCs as BPMS. Furthermore, knockdown of BPMS2 in PAC1 cells had only a minimal effect on splicing compared to BPMS knockdown. On the other hand, knockdown of BPMS-A versus BPMS-B were both observed to influence the splicing of the targets analysed (Fig.5.2), and their expression at the protein level appeared to be similar in the control (C2) cells, which agrees with previous data showing they had similar expression at the mRNA level in differentiated PAC1 cells (Nakagaki-Silva *et al*, 2019). The question that remains is why both isoforms are expressed at a similar level in these cells if BPMS-B has generally been observed to be less active. The different expression profiles of BPMS-A and BPMS-B in human tissues at the protein level has not been investigated, although GTEx multi-tissue RNA-Seq data indicates that both isoforms are expressed to a similar extent in adult tissues, but on average BPMS-B was slightly more abundant (Kim-hellmuth *et al*, 2020). Additionally, they were both shown to be expressed in various breast cancer cell lines, with BPMS-B (named as BPMS1-C in these studies) being the predominant form (Fu *et al*, 2015). Overexpression and knockdown studies of each isoform identified that both BPMS-A and BPMS-B were acting as tumour suppressors in these cells via binding to cFos and Smad3 to regulate transcription and reduce cell proliferation. BPMS-B was the isoform previously identified as regulating transcription through binding to the Smad proteins (Sun *et al*, 2006). BPMS-A and BPMS-B were also discovered to promote tumour suppression of ovarian cancer in overexpression studies, where they were again playing a role in transcription regulation (Rabelo-Fernández *et al*, 2022a). Crucially, each isoform was found to regulate specific groups of transcripts, with targets of BPMS-A being associated with chemoresistance, and targets of BPMS-B being associated with tumour suppression. Additionally, there were transcripts that were identified as being regulated by both isoforms, and these transcripts were associated with collagen remodelling, ion transport and the immune response. These studies indicate the similar and distinct roles these isoforms can play and hence it is possible they may have distinct targets in AS regulation of differentiated VSMCs also. *Pan* BPMS knockdown was found to not alter gene expression levels which indicates that it is not regulating transcriptional events in differentiated VSMCs (Nakagaki-Silva *et al*, 2019). But given that BPMS-B appears to be less active as a splicing regulator it is possible that this isoform may have other roles. The RBP Rim4 in yeast was shown to form amyloid-like structures and in this state it was able to bind RNA more efficiently than in its monomeric state (Ottoz *et al*, 2023). By binding RNA, it was able to repress translation by sequestering mRNA away from the translation

machinery. RBPMS-B may be playing a similar role in repressing RNA processing/translation in cells, as it is potentially also forming large structures. More work investigating the RNA binding capability of RBPMS-B as well as its RNA targets and their outcome in cells will help to elucidate its potential role.

Given that the RRM of the isoforms are identical, these proteins may be forming heterodimers with one another and even hetero-oligomers. This phenomenon has not been investigated but if they are able to form these complexes, the outstanding question is how will this affect their function? As previously shown with the chimeric mutants where the C-terminus of RBPMS-A was fused onto RBPMS-B caused RBPMS-B to gain repressive function; whilst the fusion of the RBPMS-B C-terminus onto RBPMS-A caused RBPMS-A to lose repressive function, may provide some clues (Nakagaki-Silva, PhD Thesis, University of Cambridge, 2019). Whether addition of RBPMS-B to oligomers/dimers of RBPMS-A causes more extreme higher order assembly of RBPMS-A and therefore a reduction in activity remains to be explored.

6.4 CONCLUSION

The work described here provides the framework to understand the mechanisms by which RBPMS regulates splicing. The successful production of an *in vitro* splicing activation assay with *PTPRF* along with the repression assay previously developed with *Tpm1*, provides a unique opportunity for an in-depth dissection of the mechanisms of both activation and repression for a single splicing regulator. *Cis* element analysis of the *PTPRF* construct demonstrated that the activation of *PTPRF* by RBPMS-A was specific. Furthermore, it was determined that RBPMS-dependent activation also required MBNL binding sites proximal to the RBPMS motifs.

The development of the working *in vitro* splicing activation and repression assays allowed for the assessment of the regions of RBPMS that were required to regulate these splicing events, and for this activity to be directly compared to biophysical behaviour. Initially, aromatic residues were found to be important for splicing activity in transfection assays, whilst the basic residues in the same area were not necessary. Wild-type RBPMS-A was observed to undergo phase separation in the presence of crowding reagents as well as smaller higher order assemblies of >20 subunits on its own. Surprisingly, both the aromatic residues and the basic residues were found to be necessary for self-assembly of RBPMS-A into larger oligomers (>10mer). This finding asserted that higher order assembly of RBPMS-A may not be necessary for its activity. Additionally, it was found that RBPMS-A may not require as extensive RNA binding as observed for the wild-type protein in order for it to perform its splicing function. Preliminary data looking at the heterotypic interactions of the aromatic and basic mutants of RBPMS-A showed that the aromatic mutants had impaired interactions compared to the basic mutants and to the wild-type protein. Therefore, it appears that the homotypic interactions that RBPMS-A can form may not be necessary for its activity and the higher order self-assembly behaviour of RBPMS-A may be a side-effect from its ability to form heterotypic interactions with co-regulators, which are required for activity. Further studies investigating true RNA binding abilities of the mutants as well as proteomic screens is required to fully elucidate the mechanism of action of RBPMS.

The other isoform, RBPMS-B, was identified here to be forming extreme higher order assemblies, much larger than those observed for RBPMS-A. The region found to be responsible for this extreme oligomerisation behaviour was a section of the RBPMS-B unique C-terminus previously predicted to be aggregation prone and shown here to be forming anti-parallel β sheet structures. This region was found to have fibrillisation propensity, although the full-length protein was only observed to form large spherical oligomers. Certain residues within this region were found to be important for the residual activity of RBPMS-B. In particular, the aromatic residues of the C-terminus of RBPMS-B were found to be important, showing similarities with RBPMS-A, and indicating they may share a common mode of

activation at some exons. The different levels of inclusion observed for the *PTPRF* and *PPFIBP1* exons for each isoform suggest they may not always act by the same mechanism, and intron length and the location of *cis* elements may play a role. A model for the potential assemblies of each isoform and their mutants and the impact on activity is shown in a summary figure below (Figure 6.5).

Overall, the data shown here demonstrates the importance of the aromatic residues of the C-terminus of RBPMS for its regulation of splicing activation and repression. Higher order assembly of RBPMS-A was found to not be required for activity whilst extreme higher order assembly of RBPMS-B was found to be detrimental for its splicing function. Finally, the production of a function *in vitro* splicing activation and repression assay provides the basis for fundamental analysis of the different mechanisms by which RBPMS achieves exon skipping versus exon inclusion.

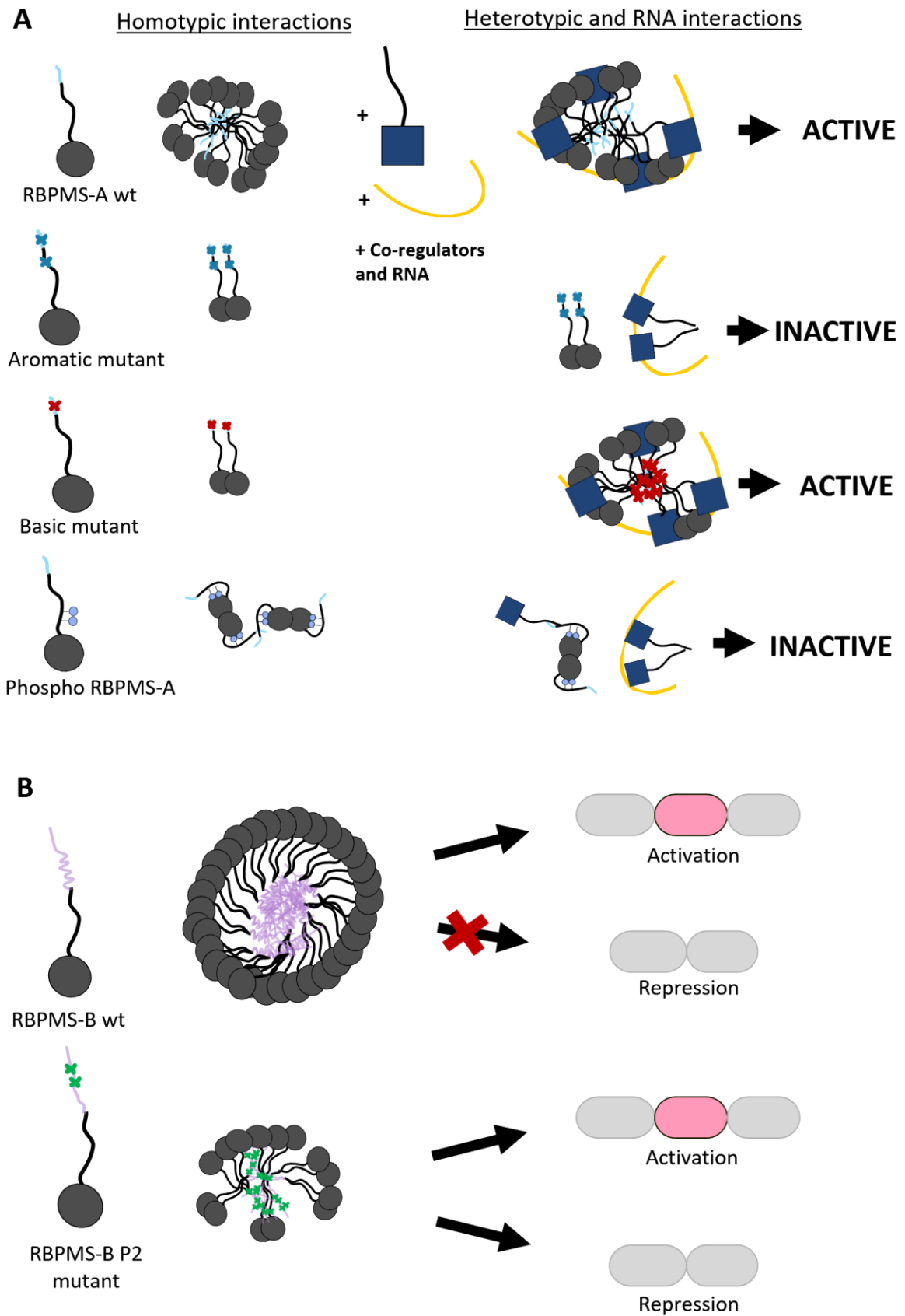


Figure legend on the next page.

Figure 6.5 Summary of the findings about the structure-function relationship of RBPMS: **A:** RBPMS-A self-assembly is affected by aromatic and basic residues on its extreme C-terminus, as well as by phosphorylation at T113 and T118. Wild-type (wt) and the basic mutant proteins can likely still interact with RNA, co-regulators such as RBFOX2 and possibly MBNL1, and are active, but the aromatic mutants cannot, therefore they lack splicing activity. The TT/E mutant was shown in this study to potentially maintain heterotypic interactions but was previously shown to not bind RNA due to interaction with negative charge at 113 and 118 interacting with the RRM RNA binding interface (Barnhart et al, 2022). **B:** RBPMS-B was found in this study to form very large complexes, predicted to be mediated by its anti-parallel β sheet structured HSA region. Two proline substitutions at this region produced a mutant that was now able to repress splicing events and likely has an assembly behaviour more closely resembling RBPMS-A wild-type. Heterotypic interactions with RBPMS-B have not been explored, nor has RNA binding, and thus these are not shown in the RBPMS-B model. Mutations are shown with crosses (X) in either blue (aromatic), red (basic) or green (proline substitution). Phosphorylation is indicated with a blue circle and a stick.

6.5 FUTURE DIRECTIONS

Importance of heterotypic interactions for the splicing activity of RBPMS

The previous proteomic screen of RBPMS-A revealed that the final twenty amino acids were required to mediate PPIs, activity and self-assembly (Yang *et al*, 2023). Given the importance of the IDR aromatic residues for activity and self-assembly; and the basic residues for self-assembly only, it will be interesting to observe which residues are also important for heterotypic interactions. Moreover, it will be interesting to observe how phosphorylation of T113 and T118 may affect PPIs by using the TT/E mutant protein. This was attempted in section 4.6, but greater efficiency of pull-down is required for a successful proteomics study. This would involve the addition of STREP-tagged RBPMS-A and its mutants to HeLa nuclear extract, followed by pull-down with STREP-tagged beads and LC-MS/MS analysis.

Additionally, wild-type RBPMS-A was observed to form hetero-oligomers in nuclear extract, whilst the final twenty amino acid deletion mutant did not (Yang *et al*, 2023). This experiment should be performed with the aromatic, basic and phospho-mutants to observe if they can form hetero-oligomers in a similar manner to the wild-type protein or if this is also impaired like their self-assembly behaviour. These experiments would involve the addition of fluorescently labelled protein to HeLa nuclear extract, or a bovine serum albumin (BSA) control, followed by sedimentation with a 10 – 60% glycerol gradient and fractionation with fluorescence detection. This would determine if formation of hetero-oligomers with co-regulators is required for activity of RBPMS-A, as suggested by the binding-condensate model for splicing repression by RBPMS-A (Fig.6.1).

Identification of which regions of RBPMS interact with other subunits/proteins

To observe where on RBPMS interactions with other subunits or co-regulator proteins are occurring, Click-chemistry can be employed (Kolb *et al*, 2001; Parker & Pratt, 2020). This technique involves the addition of a chemical moiety onto a cysteine, lysine, or methionine residue of the protein of interest which can then interact with proteins with a complementary chemical moiety when they come into close proximity. Photoactivatable reagents with maleimide or methanethiosulfonate groups can specifically react with the thiol groups of Cysteine residues, giving precise positioning of this group on a particular protein (Walko *et al*, 2019). Labelled proteins can then be introduced to partner proteins and followed by excitation at a particular wavelength, the photoactivatable group can form a cross-link with a partner protein. Reduction of the disulphide bond can then result in the transfer of the thiol group to the partner protein which can then be detected via mass spectrometry. Based on where the chemical moieties are for a particular protein and which proteins they have reacted with will reveal

not only PPI information but will specifically show which regions the protein interacts with. Given that the sole cysteine-196 of RBPMS-A has been shown to not be required for activity, this would be a good initial target to chemically modify, then further mutants could be made with the cysteine in different locations (providing this does not affect activity or assembly) to see where else may be important for forming heterotypic (and homotypic) interactions.

Investigation of the extreme higher order assemblies formed by RBPMS-B

Further investigation is required into the biophysical behaviour of RBPMS-B and its assemblies to reveal their stability and how they form. By observing the oligomerisation behaviour of RBPMS-B under different conditions, the nature of the interactions required, whether these assemblies are dynamic, and if the full-length protein has any propensity to form fibrils with and without the HSA peptide acting as a seed. Such experiments would involve negative stain EM, mass photometry and the ThT assays and should include wild-type RBPMS-B, its P2 mutant and RBPMS-A. Additionally, it will be interesting to observe if RBPMS-B assembly formation is affected by RNA in a similar manner to RBPMS-A, as well as assessing if RBPMS-B can bind to RNA substrates with similar ability as RBPMS-A. Lastly, the formation of possible heterodimers and hetero-oligomers should be assessed to observe: 1) if such structures can form and 2) what the biophysical properties of such assemblies may be.

Activation versus repression complex assembly

To observe how many subunits of RBPMS assemble on a transcript in an activation versus repression context, single molecule techniques can be employed. Single molecule total internal reflection fluorescence (TIRF) microscopy can identify the stoichiometry of a fluorescently labelled protein on a single transcript, and has previously been used to assess how many PTBP1 and MBNL1 proteins bind to *Tpm1* surrounding exon 3 (Cherny *et al*, 2010; Gooding *et al*, 2013). This could be employed with the *in vitro* constructs *Tpm1* and *PTPRF* to observe how many RBPMS subunits are assembled, as well as investigating the size of each complex. This approach can elucidate how RBPMS affects the binding of core snRNP/splicing factors to particular RNAs by addition of recombinant RBPMS to nuclear extract where the core splicing factors have been fluorescently labelled. If a more efficiently spliced activation construct can be produced, then full investigation of the activation assembly can be achieved to understand which other proteins are interacting specifically in an activation context, by using RNA-assisted pull-down in HeLa nuclear extract and performing mass spectrometry analysis. These studies would help to investigate how splicing activation and repression by RBPMS occur, shedding light on their mechanisms.

Global analyses of the targets of RBPMS-A and RBPMS-B

The two isoforms of RBPMS have been shown to have differential activity, with RBPMS-A appearing as the main splicing regulator given it can both activate and repress splicing. This assumption is based on the activity of each isoform with only a subset of the exons observed to be affected by *pan* RBPMS knockdown in differentiated PAC1 cells (Nakagaki-Silva *et al*, 2019). Isoform specific knockdown followed by RNA sequencing will reveal if these isoforms follow the suspected trend of RBPMS-B being less effective as a splicing regulator, or if each isoform can regulate different exons to different extents, as suggested by the studies on each isoform with transcriptional control in Ovarian cancer (Rabelo-Fernández *et al*, 2022a). UV crosslinking and immunoprecipitation (CLIP) techniques could be employed for each isoform (either through isoform specific knockdown or overexpression) to better identify where they are binding, and to observe if this is indeed related to their activity upon a particular exon. In addition, comparison of the PPI networks of each isoform via proteomic screening may further elucidate the different activities of these proteins. The different binding partners of RBPMS-A and RBPMS-B may reveal which interactions of RBPMS are required for activation versus those required for repression.

In summary, these future studies should aim to further elucidate the mechanisms of splicing by RBPMS, focusing on which residues are important for each isoform and why, the differences between activation and repression of splicing, as well as the differences between each isoform. These findings will enable further understanding of the mechanisms behind phenotypic plasticity of VSMCs and may prove useful for understanding the role of each isoform may be playing in other contexts, such as the in heart and in cancer.

REFERENCES

- Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630: 493–500
- Acharya P, Parkins S & Tranter M (2024) RNA binding proteins as mediators of pathological cardiac remodeling. *Front Cell Dev Biol* 12: 1–12
- Agirre E, Oldfield AJ, Bellora N, Segelle A & Luco RF (2021) Splicing-associated chromatin signatures: a combinatorial and position-dependent role for histone marks in splicing definition. *Nat Commun* 12: 1–16
- Aguero T, Zhou Y, Kloc M, Chang P, Houliston E & King M Lou (2016) Hermes (Rbpms) is a critical component of RNP complexes that sequester germline RNAs during oogenesis. *J Dev Biol* 4: 13–18
- Akagi T, Kamei D, Tsuchiya N, Nishina Y, Horiguchi H, Matsui M, Kamma H & Yamada M (2000) Molecular characterization of a mouse heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its tissue-specific expression. 245: 267–273
- Akerberg AA, Burns CE & Burns CG (2019) Exploring the Activities of RBPMS Proteins in Myocardial Biology. *Pediatr Cardiol* 40: 1410–1418
- Akerberg AA, Trembley M, Butty V, Schwertner A, Zhao L, Beerens M, Liu X, Mahamdeh M, Yuan S, Boyer L, *et al* (2022) RBPMS2 Is a Myocardial-Enriched Splicing Regulator Required for Cardiac Function. *Circ Res*: 980–1000
- Alberti S, Gladfelter A & Mittag T (2019) Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* 176: 419–434
- Allahverdian S, Chaabane C, Boukais K & Francis GA (2018) Smooth muscle cell fate and plasticity in atherosclerosis. *Cardiovasc Res* 114: 540–550
- Anvar SY, Allard G, Tseng E, Sheynkman GM, de Klerk E, Vermaat M, Yin RH, Johansson HE, Ariyurek Y, den Dunnen JT, *et al* (2018) Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol* 19: 1–18
- Babinchak WM, Dumm BK, Venus S, Boyko S, Putnam AA, Jankowsky E & Surewicz WK (2020) Small molecules as potent biphasic modulators of protein liquid-liquid phase separation. *Nat Commun*

- Bampton A, Gatt A, Humphrey J, Cappelli S, Bhattacharya D, Foti S, Brown AL, Asi Y, Low YH, Foiani M, *et al* (2021) HnRNP K mislocalisation is a novel protein pathology of frontotemporal lobar degeneration and ageing and leads to cryptic splicing. *Acta Neuropathol* 142: 609–627
- Banani SF, Lee HO, Hyman AA & Rosen MK (2017) Biomolecular condensates: Organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18: 285–298
- Baralle FE & Giudice J (2017) Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 18: 437–451
- Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ & Frey BJ (2010) Deciphering the splicing code. *Nature* 465: 53–59
- Barnhart MD, Yang Y, Nakagaki-Silva EE, Hammond TH, Pizzinga M, Gooding C, Stott K & Smith CWJ (2022) Phosphorylation of the smooth muscle master splicing regulator RBPMS regulates its splicing activity. *Nucleic Acids Res* 50: 11895–11915
- Bartsch D, Kalamkar K, Ahuja G, Lackmann JW, Hescheler J, Weber T, Bazzi H, Clamer M, Mendjan S, Papantonis A, *et al* (2023) mRNA translational specialization by RBPMS presets the competence for cardiac commitment in hESCs. *Sci Adv* 9: 1–17
- Batlle C, Yang P, Coughlin M, Kim HJ, Taylor JP & Ventura S (2020) hnRNPD Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Article hnRNPD Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. *CellReports* 30: 1117-1128.e5
- Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat Publ Gr*
- Berger SL (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev*: 142–148
- Berget SM (1995) Exon recognition in vertebrate splicing. *J biological Chem* 270: 2411–2414
- Berget SM, Moore C & Sharp PA (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA *. *Proc Natl Acad Sci* 74: 3171–3175
- Biancalana M, Makabe K, Koide A & Koide S (2009) Molecular Mechanism of Thioflavin-T Binding to the Surface of β - Rich Peptide Self-Assemblie. *J Mol Biol* 385: 1052–1063
- Blanc S (2016) Arginine Methylation : The Coming of Age. *Mol Cell* 65: 8–25
- Blanchette M & Chabot B (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites

- and by intron elements that repress splice site utilization. *18*: 1939–1952
- Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 0004: 106–110
- Boija A, Klein IA, Sabari BR, Dall’Agnese A, Coffey EL, Zamudio A V., Li CH, Shrinivas K, Manteiga JC, Hannett NM, *et al* (2018) Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175: 1842-1855.e16
- Bonnal S, Martinez C, Forch P, Bachi A, Wilm M & Valcarcel J (2008) RBM5/Luca-15/H37 Regulates Fas Alternative Splice Site Pairing after Exon Definition Sophie. *Cell* 32: 81–95
- Bonomi S, Matteo A, Buratti E, Cabianca DS, Baralle FE, Ghigna C & Biamonti G (2013) HnRNP A1 controls a splicing regulatory circuit promoting mesenchymal-to-epithelial transition. *Nucleic Acids Res* 41: 8665–8679
- Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M & Black DL (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* 21: 1636–1652
- Bratek-skicki A, Pancsa R, Meszaros B, Lindt J Van & Tompa P (2020) A guide to regulation of the formation of biomolecular condensates. *287*: 1924–1935
- Bregman DB, Pestell GR & Kidd J V (2000) Cell cycle regulation and RNA Polymerase II. *Front Biosci* 5: 244–257
- Brozovich F V., Nicholson CJ, Degen C V., Gao YZ, Aggarwal M & Morgan KG (2016) Mechanisms of vascular smooth muscle contraction and the basis for pharmacologic treatment of smooth muscle disorders. *Pharmacol Rev* 68: 476–532
- Buchan JR & Parker R (2009) Eukaryotic Stress Granules: The Ins and Outs of Translation. *Mol Cell* 36: 932–941
- Bugg D, Bailey LRJ, Bretherton RC, Beach KE, Reichardt IM, Robeson KZ, Reese AC, Gunaje J, Flint G, DeForest CA, *et al* (2022) MBNL1 drives dynamic transitions between fibroblasts and myofibroblasts in cardiac wound healing. *Cell Stem Cell* 29: 419-433.e10
- Buljan M, Chalancon G, Eustermann S, Wagner GP, Fuxreiter M, Bateman A & Babu MM (2012) Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Mol Cell* 46: 871–883
- Buratti E, Muro F, Giombi M, Gherbassi D, Iaconcig A & Baralle FE (2004) RNA Folding Affects the

- Recruitment of SR Proteins by Mouse and Human Polypurinic Enhancer Elements in the Fibronectin EDA Exon. 24: 1387–1400
- Cajal S (1903) Un sencillo metodo de coloracion seletiva del reticulo pro- toplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. *Trab Lab Invest Biol Univ Madrid* 2: 129–221
- Cartegni L, Chew SL & Krainer AR (2002) Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat Rev Genet* 3: 285–298
- Cartegni L & Krainer AR (2002) Disruption of an SF2 / ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. 30: 377–384
- Chen M & Manley JL (2009) Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* 10: 741–754
- Cherny D, Gooding C, Eperon GE, Coelho MB, Bagshaw CR, Smith CWJ & Eperon IC (2010) Stoichiometry of a regulatory splicing complex revealed by single-molecule analyses. *EMBO J* 29: 2161–2172
- Chkheidze AN & Liebhaber SA (2003) A Novel Set of Nuclear Localization Signals Determine Distributions of the α CP RNA-Binding Proteins. *Mol Cell Biol* 23: 8405–8415
- Chow LT, Gelinas RE & Roberts RJ (1977) An Amazing Sequence Arrangement of Adenovirus RNA at the 5' Ends. 12: 1–8
- Coelho MB & Smith CWJ (2014) Regulation of Alternative Pre-mRNA Splicing. *Methods Mol Biol* 1126: 257–269
- Collier NC, Heuser J, Levy MA & Schlesinger MJ (1988) Ultrastructural and biochemical analysis of the stress granule in chicken embryo fibroblasts. *J Cell Biol* 106: 1131–1139
- Collier NC & Schlesinger MJ (1986) The dynamic state of heat shock proteins in chicken embryo fibroblasts. *J Cell Biol* 103: 1495–1507
- Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X & Ventura S (2007) AGGRESCAN: A server for the prediction and evaluation of 'hot spots' of aggregation in polypeptides. *BMC Bioinformatics* 8
- De Conti L, Baralle M & Buratti E (2013) Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* 4: 49–60

- Cooke JA & Brown LJ (2011) Distance measurements by continuous wave EPR spectroscopy to monitor protein folding
- Coolidge CJ, Seely RJ & Patton JG (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *25*: 888–895
- Cooper TA, Wan L & Dreyfuss G (2009) RNA and Disease. *Cell* 136: 777–793
- Cosma MP (2002) Ordered Recruitment: Gene-Specific Mechanism of Transcription Activation. *Mol Cell* 10: 227–236
- Creemers EE, Sutherland LB, Oh J, Barbosa AC & Olson EN (2006) Coactivation of MEF2 by the SAP Domain Proteins Myocardin and MASTR. *Mol Cell* 23: 83–96
- Damianov A, Ying Y, Lin CH, Lee JA, Tran D, Vashisht AA, Bahrami-Samani E, Xing Y, Martin KC, Wohlschlegel JA, *et al* (2016) Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* 165: 606–619
- David CJ, Boyne AR, Millhouse SR & Manley JL (2011) The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev* 25: 972–982
- Ding F, Su CJ, Edmonds KHK, Liang G & Elowitz MB (2022) Dynamics and functional roles of splicing factor autoregulation. *Cell Rep* 39: 110985
- Ding J, Xu X, Yang D, Chu P, Dalton ND, Ye Z, Yeakley JM, Cheng H, Xiao R, Ross J, *et al* (2004) Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J* 23: 885–896
- Dutagaci B, Nawrocki G, Goodluck J & Ashkarran AA (2021) Charge-driven condensation of RNA and proteins suggests broad role of phase separation in cytoplasmic environments. 1–36
- Dynan WS (1989) Modularity in promoters and enhancers. *Cell* 58: 1–4
- Early P, Rogers J, Davis M, Calame K, Bond M, Wall R & Hood L (1980) Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. *Cell* 20: 313–319
- Eckhart L, Ban J, Ballaun C, Weninger W & Tschachler E (1999) Reverse transcription-polymerase chain reaction products of alternatively spliced mRNAs form DNA heteroduplexes and heteroduplex complexes. *J Biol Chem* 274: 2613–2615
- Ellis JD, Barrios-Rodiles M, Çolak R, Irimia M, Kim TH, Calarco JA, Wang X, Pan Q, O’Hanlon D, Kim PM, *et al* (2012) Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks.

- Erkelenz S, Mueller WF, Evans MS, Busch A & Scho K (2013) Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *Rna* 19: 96–102
- Esteban-Serna S, McCaughan H & Granneman S (2023) Advantages and limitations of UV cross-linking analysis of protein–RNA interactomes in microbes. *Mol Microbiol* 120: 477–489
- Expert-Bezancon A, Caer JP Le & Marie J (2002) Heterogeneous nuclear ribonucleoprotein (hnRNP) K is a component of an intronic splicing enhancer complex that activates the splicing of the alternative exon 6A from chicken β -tropomyosin pre-mRNA. *J Biol Chem* 277: 16614–16623
- Farazi TA, Leonhardt CS, Mukherjee N, Mihailovic A, Li S, Max KEA, Meyer C, Yamaji M, Cekan P, Jacobs NC, *et al* (2014) Identification of the RNA recognition element of the RBPMS family of RNA-binding proteins and their transcriptome-wide mRNA targets. *Rna* 20: 1090–1102
- Frismantiene A, Philippova M, Erne P & Resink TJ (2018) Smooth muscle cell-driven vascular diseases and molecular mechanisms of VSMC plasticity. *Cell Signal* 52: 48–64
- Fu J, Cheng L, Wang Y, Yuan P, Xu X, Ding L, Zhang H, Jiang K, Song H, Chen Z, *et al* (2015) The RNA-binding protein RBPMS1 represses AP-1 signaling and regulates breast cancer cell proliferation and migration. *Biochim Biophys Acta - Mol Cell Res* 1853: 1–13
- Gade Malmos K, Blancas-Mejia LM, Weber B, Buchner J, Ramirez-Alvarado M, Naiki H & Otzen D (2017) ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. *Amyloid* 24: 1–16
- Gallivan JP & Dougherty DA (1999) Cation- π interactions in structural biology. *Proc Natl Acad Sci* 96: 9459–9464
- Gan P, Wang Z, Bezprozvannaya S, McAnally JR, Tan W, Li H, Bassel-Duby R, Liu N & Olson EN (2024) RBPMS regulates cardiomyocyte contraction and cardiac function through RNA alternative splicing. *Cardiovasc Res* 120: 56–68
- Gan P, Wang Z, Morales MG, Zhang Y, Bassel-Duby R, Liu N & Olson EN (2022) RBPMS is an RNA-binding protein that mediates cardiomyocyte binucleation and cardiovascular development. *Dev Cell* 57: 959-973.e7
- García-Moreno JF & Romão L (2020) Perspective in alternative splicing coupled to nonsense-mediated mrna decay. *Int J Mol Sci* 21: 1–21
- Garcia-pardo J, Bartolomé-nafría A, Chaves-sanjuan A, Gil-garcia M, Visentin C, Bolognesi M, Ricagno S & Ventura S (2023) Cryo-EM structure of hnRNPD-2 fibrils, a functional amyloid associated

- with limb-girdle muscular dystrophy D3. *Nat Commun* 14: 1–12
- Gasior K, Zhao J, McLaughlin G, Forest, Gergory F, Gladfelter AS & Nerby J (2019) Partial demixing of RNA-protein complexes leads to intra-droplet patterning in phase-separated biological condensates. *Physiol Rep* 176: 139–148
- Gerber W V, Yatskievych TA, Antin PB, Correia KM, Conlon RA & Krieg PA (1999) The RNA-binding protein gene, hermes, is expressed at high levels in the developing heart. *Mech Dev* 80: 77–86
- Gilbert W (1978) Why genes in pieces? *Nature* 271: 501
- Ginell GM & Holehouse AS (2023) An Introduction to the Stickers-and-Spacers Framework as Applied to Biomolecular Condensates. In *Methods in molecular biology* pp 95–117.
- Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K & Luhrmann R (2012) Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun* 3
- Giudice J & Jiang H (2024) Splicing regulation through biomolecular condensates and membraneless organelles. *Nat Rev Mol Cell Biol*
- Goers ES, Purcell J, Voelker RB, Gates DP & Berglund JA (2010) MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res* 38: 2467–2484
- Gong TT, Liu FH, Xiao Q, Li YZ, Wei YF, Xu HL, Cao F, Sun ML, Jiang FL, Tao T, *et al* (2024) SH3RF2 contributes to cisplatin resistance in ovarian cancer cells by promoting RBPMS degradation. *Commun Biol* 7: 1–15
- Gooding C, Edge C, Lorenz M, Coelho MB, Winters M, Kaminski CF, Cherny D, Eperon IC & Smith CWJ (2013) MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3. *Nucleic Acids Res* 41: 4765–4782
- Gooding C & Smith CWJ (2008) Tropomyosin exons as models for alternative splicing. *Adv Exp Med Biol* 644: 27–42
- Graveley BR, Hertel KJ & Maniatis T (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J* 17: 6747–6756
- Greenfield NJ (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 1: 2876–2890
- Gromak N, Matlin AJ, Cooper TA & Smith CWJ (2003) Antagonistic regulation of α -actinin alternative

- splicing by CELF proteins and polypyrimidine tract binding protein. *Rna* 9: 443–456
- Gromiha MM, Santhosh C & Ahmad S (2004) Structural analysis of cation – pi interactions in DNA binding proteins. 34: 203–211
- Guan W-L, Jiang L-L, Yin X-F & Hu H-Y (2023) PABPN1 aggregation is driven by Ala expansion and poly(A)-RNA binding, leading to CFIm25 sequestration that impairs alternative polyadenylation. *J Biol Chem* 299: 105019
- Guérousov S, Weatheritt RJ, Hanlon DO, Lin Z, Narula A, Gingras A, Blencowe BJ, Guérousov S, Weatheritt RJ, Hanlon DO, *et al* (2017a) Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* 170: 324-339.e23
- Guérousov S, Weatheritt RJ, O'Hanlon D, Lin ZY, Narula A, Gingras AC & Blencowe BJ (2017b) Regulatory Expansion in Mammals of Multivalent hnRNP Assemblies that Globally Control Alternative Splicing. *Cell* 170: 324-339.e23
- Hallegger M, Chakrabarti AM, Lee FCY, Lee BL, Amalietti AG, Odeh HM, Copley KE, Rubien JD, Portz B, Kuret K, *et al* (2021) TDP-43 condensation properties specify its RNA-binding and regulatory repertoire. *Cell* 184: 4680-4696.e22
- Hamid FM & Makeyev E V (2017) A mechanism underlying position-specific regulation of alternative splicing. *Nucleic Acids Res* 45: 12455–12468
- Hartmuth K, van Santen MA & Lührmann R (2012) Ultracentrifugation in the Analysis and Purification of Spliceosomes Assembled In Vitro. In *Alternative pre-mRNA Splicing* pp 147–153. John Wiley & Sons, Ltd
- Hofweber M, Hutten S, Bourgeois B, Niessing D, Madl T, Dormann D, Hofweber M, Hutten S, Bourgeois B, Spreitzer E, *et al* (2018) Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation Article Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* 173: 706-713.e13
- Horton ER, Byron A, Askari JA, Ng DHJ, Millon-Frémillon A, Robertson J, Koper EJ, Paul NR, Warwood S, Knight D, *et al* (2015) Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nat Cell Biol* 17: 1577–1587
- Horvathova I, Voigt F, Kotrys A V., Zhan Y, Artus-Revel CG, Eglinger J, Stadler MB, Giorgetti L & Chao JA (2017) The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. *Mol Cell* 68: 615-625.e9

- House AE & Lynch KW (2006) An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat Struct Mol Biol* 13: 937–944
- Houseley J & Tollervey D (2009) The Many Pathways of RNA Degradation. *Cell* 136: 763–776
- Hughes MP, Sawaya MR, Boyer DR, Goldschmidt L, Rodriguez JA, Cascio D, Chong L, Gonen T & Eisenberg DS (2018) Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. *Science* (80-) 359: 698–701
- Hunter T (2007) The Age of Crosstalk: Phosphorylation, Ubiquitination, and Beyond. *Mol Cell* 28: 730–738
- Iakoucheva LM, Radivojac P, Brown CJ, Connor TRO, Sikes JG, Obradovic Z & Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32: 1037–1049
- Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-pournatzis T, Tapial J, Raj B, Hanlon DO, Barrios-rodiles M, Babor M, *et al* (2014) A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains. *Cell* 159: 1511–1523
- Jackson RJ, Hellen CUT & Pestova T V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 11: 113–127
- Jacob AG, Moutsopoulos I, Petchey A, Mohorianu I, Sinha S & Smith CWJ (2022) RBPMS promotes contractile smooth muscle splicing and alters phenotypic behaviour of human embryonic stem cell derived vascular smooth muscle cells. *bioRxiv*: 2022.11.27.516868
- Jacob F & Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3: 318–356
- Jangi M, Boutz PL, Paul P & Sharp PA (2014) Rbfox2 controls autoregulation in RNA-binding protein networks. *Genes Dev* 28: 637–651
- Jangi M & Sharp PA (2014) Building robust transcriptomes with master splicing factors. *Cell* 159: 487–498
- Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R & Burge CB (2013) SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. *Cell* 153: 855–868
- Jiang Y, Fu X, Zhang Y, Wang SF, Zhu H, Wang WK, Zhang L, Wu P, Wong CCL, Li J, *et al* (2021) Rett syndrome linked to defects in forming the MeCP2/Rbfox/LASR complex in mouse models. *Nat Commun* 12

- Kao C & Carsten M (1997) Morphology of smooth muscle. In *Cellular Aspects of Smooth Muscle Function* pp 1–47.
- Kashima T, Rao N, David C & Manley JL (2007) hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum Mol Genet* 16: 3149–3159
- Kfir N, Lev-Maor G, Glaich O, Alajem A, Datta A, Sze SK, Meshorer E & Ast G (2015) SF3B1 Association with Chromatin Determines Splicing Outcomes. *Cell Rep* 11: 618–629
- Kim-hellmuth S, Liang Y, Oliva M, Parsana P, Muñoz-aguirre M, Park Y, Saha A, Segrè A V & Strober BJ (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* (80-) 369: 1318–1330
- Kim KS, Lee Y, Lee JH, Lee SS, Chung JM & Jung HS (2024) Optimizing protein crosslinking control: Synergistic quenching effects of glycine, histidine, and lysine on glutaraldehyde reactions. *Biochem Biophys Res Commun* 702: 149567
- Kolb HC, Finn MG & Sharpless KB (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chemie - Int Ed* 40: 2004–2021
- Kollmer M, Meinhardt K, Haupt C, Liberta F, Wulff M, Linder J, Handl L, Heinrich L, Loos C, Schmidt M, *et al* (2016) Electron tomography reveals the fibril structure and lipid interactions in amyloid deposits. *Proc Natl Acad Sci U S A* 113: 5604–5609
- Krainer AF, Conway GC & Kozak D (1990) The Essential Pre-mRNA Splicing Factor SF2 Influences 5' Splice Site Selection by Activating Proximal Sites. *Cell* 62: 35–42
- Krainer AR, Maniatis T, Ruskin B & Green MR (1984) Normal and mutant human β -globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36: 993–1005
- Kristensen LS, Ebbesen KK & Hansen TB (2019) The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* 20
- Kwek KY, Murphy S, Furger A, Thomas B, O’Gorman W, Kimura H, Proudfoot NJ & Akoulitchiev A (2002) U1 snRNA associates with tfiih and regulates transcriptional initiation. *Nat Struct Biol* 9: 800–805
- Ladd AN & Cooper TA (2002) Finding signals that regulate alternative splicing in the post-genomic era. *Genome Biol* 3: 1–16
- Lafontaine DLJ, Riback JA, Bascetin R & Brangwynne CP (2021) The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Biol* 22: 165–182

- Lagier-Tourenne C & Cleveland DW (2009) Rethinking ALS: The FUS about TDP-43. *Cell* 136: 1001–1004
- Larsen SC, Sylvestersen KB, Mund A, Lyon D, Mullari M, Madsen M V, Daniel JA, Jensen LJ & Nielsen ML (2016) Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *9*: 1–14
- Lee KY, Seah C, Li C, Chen YF, Chen CY, Wu CI, Liao PC, Shyu YC, Olafson HR, McKee KK, *et al* (2022) Mice lacking MBNL1 and MBNL2 exhibit sudden cardiac death and molecular signatures recapitulating myotonic dystrophy. *Hum Mol Genet* 31: 3144–3160
- Lenz S, Lohse P, Seidel U & Arnold HH (1989) The Alkali Light Chains of Human Smooth and Nonmuscle Myosins Are Encoded by a Single Gene. *J Biol Chem* 264: 9009–9015
- LeVine H (1999) Quantification of β -sheet amyloid fibril structures with thioflavin T. *Methods Enzymol* 309: 274–284
- Li J, Choi PS, Chaffer CL, Labella K, Hwang JH, Giacomelli AO, Kim JW, Ilic N, Doench JG, Ly SH, *et al* (2018) An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer. *Elife* 7: 1–28
- Li P, Banjade S, Cheng H, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth J V, King DS, Banani SF, *et al* (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483: 386–341
- Liao SE & Regev O (2021) Splicing at the phase-separated nuclear speckle interface: A model. *Nucleic Acids Res* 49: 636–645
- Liao SM, Du QS, Meng JZ, Pang ZW & Huang RB (2013) The multiple roles of histidine in protein interactions. *Chem Cent J* 7: 1
- Libri D, Balvay L & Fiszman MY (1992) In Vivo Splicing of the , 3 Tropomyosin Pre-mRNA : A Role for Branch Point and Donor Site Competition. *12*: 3204–3215
- Lin S, Coutinho-mansfield G, Wang D, Pandit S & Fu X (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* 15: 819–826
- Lin Y, Protter DSW, Rosen MK & Parker R (2015) Formation and Maturation of Phase Separated Liquid Droplets by RNA Binding Proteins. *Mol Cell* 60: 208–219
- Lindskog S (1997) Structure and Mechanism of Carbonic Anhydrase. *Pharmacol Ther* 74: 1–20
- Liu Q, Fang L & Wu C (2022) Alternative Splicing and Isoforms: From Mechanisms to Diseases. *Genes (Basel)* 13

- Llorian M, Gooding C, Bellora N, Hallegger M, Buckroyd A, Wang X, Rajgor D, Kayikci M, Feltham J, Ule J, *et al* (2016) The alternative splicing program of differentiated smooth muscle cells involves concerted non-productive splicing of post-transcriptional regulators. *Nucleic Acids Res* 44: 8933–8950
- Llorian M, Schwartz S, Clark TA, Hollander D, Tan LY, Spellman R, Gordon A, Schweitzer AC, De La Grange P, Ast G, *et al* (2010) Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* 17: 1114–1123
- Logan J, Falck-Pedersen E, Darnell JE & Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* 84: 8306–8310
- Lovci MT, Ghanem D, Marr H, Arnold J, Gee S, Parra M, Liang TY, Stark TJ, Gehman LT, Hoon S, *et al* (2013) Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat Struct Mol Biol* 20: 1434–1442
- Lunde BM, Moore C & Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8: 479–490
- Lundquist EA, Herman RK, Rogalski TM, Mullen GP, Moerman DG & Shaw JE (1996) The mec-8 gene of *C. Elegans* encodes a protein with two RNA recognition motifs and regulates alternative splicing of unc-52 transcripts. *Development* 122: 1601–1610
- MacArthur MW & Thornton JM (1991) Influence of proline residues on protein conformation. *J Mol Biol* 218: 397–412
- Malki I, Liepina I, Kogelnik N, Watmuff H, Robinson S, Lightfoot A, Gonchar O, Bottrill A, Fry AM & Dominguez C (2022) Cdk1-mediated threonine phosphorylation of Sam68 modulates its RNA binding, alternative splicing activity and cellular functions. *Nucleic Acids Res* 50: 13045–13062
- Mandal SS, Chu C, Wada T, Handa H, Shatkin AJ & Reinberg D (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A* 101: 7572–7577
- Mao AH, Lyle N & Pappu R V (2013) Describing Sequence-Ensemble Relationships for Intrinsically Disordered Proteins. *Biochem J* 31: 1713–1723
- Marasco LE, Dujardin G, Sousa-luís R & Liu YH (2023) Counteracting chromatin effects of a splicing-correcting antisense oligonucleotide improves its therapeutic efficacy in spinal muscular atrophy. *Cell* 185: 2057–2070

- Mardon HJ, Sebastio G & Baralle FE (1987) A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res* 15: 7725–7733
- Martín G, Márquez Y, Mantica F, Duque P & Irimia M (2021) Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biol* 22: 1–26
- Mata M De, Alonso CR, Fededa JP, Pelisch F, Cramer P, Bentley D & Kornblihtt AR (2003) A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Mol Cell* 12: 525–532
- Matlin AJ, Southby J, Gooding C & Smith CWJ (2007) Repression of a -actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. *Rna* 13: 1214–1223
- Maurin M, Ranjouri M, Megino-Luque C, Newberg JY, Du D, Martin K, Miner RE, Prater MS, Wee DKB, Centeno B, *et al* (2023) RBFOX2 deregulation promotes pancreatic cancer progression and metastasis through alternative splicing. *Nat Commun* 14
- McDonald OG & Owens GK (2007) Programming smooth muscle plasticity with chromatin dynamics. *Circ Res* 100: 1428–1441
- Miano JM (2015) Myocardin in biology and disease. *J Biomed Res* 29: 3–19
- Misteli T, Cáceres JF, Clement JQ, Krainer AR, Wilkinson MF & Spector DL (1998) Serine Phosphorylation of SR Proteins Is Required for Their Recruitment to Sites of Transcription In Vivo. 143: 297–307
- Mittag T & Pappu R V. (2022) A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol Cell* 82: 2201–2214
- Mittnacht S & Weinberg RA (1991) G1/S Phosphorylation of the Retinoblastoma Protein Is Associated with an Altered Affinity for the Nuclear Compartment. *Cell* 65: 381–393
- Modrek B & Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* 34: 177–180
- Monahan Z, Ryan VH, Janke AM, Burke KA, Rhoads SN, Zerze GH, Meally RO, Dignon GL, Conicella AE, Zheng W, *et al* (2017) Phosphorylation of the FUS low-complexity domain disrupts phase separation , aggregation , and toxicity. *EMBO J* 36: 2951–2967
- Montgomery TS (1898) Comparative cytological studies, with especial regard to the morphology of the nucleolus. *J Morphol* 15: 265–582

- Mullen MP, Smith CWJ, Patton JG & Nadal-Ginard B (1991) α -Tropomyosin mutually exclusive exon selection: Competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev* 5: 642–655
- Nabeshima Y, Nabeshima Y, Nonomura Y & Fujii-Kuriyama Y (1987) Nonmuscle and smooth muscle myosin light chain mRNAs are generated from a single gene by the tissue-specific alternative RNA splicing. *J Biol Chem* 262: 10608–10612
- Naiki H, Higuchi K, Matsushima K, Shimada A, Chen WH, Hosokawa M & Takeda T (1990) Fluorometric Determination of Amyloid Fibrils in vitro Using the Fluorescent Dye, Thioflavine T. *Lab Invest* 62: 768–773
- Nakagaki-Silva EE (2019) Identification of RBPMS as a smooth muscle master splicing regulator via association of its gene with super-enhancers.
- Nakagaki-Silva EE, Gooding C, Llorian M, Jacob AG, Richards F, Buckroyd A, Sinha S & Smith CWJ (2019) Identification of RBPMS as a mammalian smooth muscle master splicing regulator via proximity of its gene with super-enhancers. *Elife* 8: 1–33
- Nanga RRP, R. Brender J, Xu J, Hartman K, Subramanian V & Ramamoorthy A (2009) Three-Dimensional Structure and Orientation of Rat Islet Amyloid Polypeptide Protein in a Membrane Environment by Solution NMR Spectroscopy. *J Am Chem Soc* 131: 1–7
- Narimoto T, Sakurai K, Okamoto A, Chatani E, Hoshino M, Hasegawa K, Naiki H & Goto Y (2004) Conformational stability of amyloid fibrils of β 2- microglobulin probed by guanidine-hydrochloride-induced unfolding. *FEBS Lett* 576: 313–319
- Narlikar GJ, Fan H & Kingston RE (2002) Cooperation between Complexes that Regulate Chromatin Structure and Transcription. *Cell* 108: 475–487
- Nelson KK & Green MR (1989) Mammalian U2 snRNP has a sequence- specific RNA-binding activity. *Genes Dev* 3: 1562–1571
- Nilsen TW & Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. 463
- Nirenberg MW & Matthaei JH (1961) The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci U S A* 47: 1588–1602
- Nojima T, Dienstbier M, Murphy S, Proudfoot NJ & Dye MJ (2013) Definition of RNA Polymerase II CoTC Terminator Elements in the Human Genome. *Cell Rep* 3: 1080–1092
- Notarnicola C, Rouleau C, Le Guen L, Virsolvy A, Richard S, Faure S & De Santa Barbara P (2012) The

- RNA-binding protein RBPMS2 regulates development of gastrointestinal smooth muscle. *Gastroenterology* 143: 687-697.e9
- Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, *et al* (2015) Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* 57: 936–947
- Olson S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley M, Rio DC & Graveley BR (2007) A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol* 14: 1134–1140
- Ottoz DS, Tang LC, Dyatel AE, Jovanovic M & Berchowitz LE (2023) Assembly and function of the amyloid-like translational repressor Rim4 is coupled with nutrient conditions. *EMBO J*: 1–18
- Owens GK (1995) Regulation of differentiation of vascular smooth muscle cells. *Physiol Rev* 75: 487–517
- Owens GK, Kumar MS & Wamhoff BR (2004) Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rev* 84: 767–801
- Padgett RA, Konarska MM, Grabowski PJ, Hardy SF & Sharp PA (1984) Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* (80-) 225: 898–903
- Pan Q, Shai O, Lee LJ, Frey BJ & Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413–1415
- Panigrahi A & O'Malley BW (2021) Mechanisms of enhancer action: the known and the unknown. *Genome Biol* 22: 1–30
- Park E, Pan Z, Zhang Z, Lin L & Xing Y (2018) The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am J Hum Genet* 102: 11–26
- Parker CG & Pratt MR (2020) Click Chemistry in Proteomic Investigations. *Cell* 180: 605–632
- Partridge LMM & Carter DA (2017) Novel Rbfox2 isoforms associated with alternative exon usage in rat cortex and suprachiasmatic nucleus. *Sci Rep* 7: 1–11
- Patel A, Lee HO, Jawerth L, Maharana S, Jahnel M, Hein MY, Stoyanov S, Mahamid J, Saha S, Franzmann TM, *et al* (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 162: 1066–1077
- Paul S, Arias MA, Wen L & Fei J (2024) RNA molecules display distinctive organization at nuclear speckles. *ISCIENCE* 27: 109603

- Philips A V., Timchenko LT & Cooper TA (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science (80-)* 280: 737–741
- Press M, Jung T, König J, Grune T & Höhn A (2019) Protein aggregates and proteostasis in aging: Amylin and β -cell function. *Mech Ageing Dev* 177: 46–54
- Proudfoot NJ, Furger A & Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108: 501–512
- Qamar S, Wang GZ, Randle SJ, Ruggeri FS, Varela JA, Lin JQ, Phillips EC, Miyashita A, Williams D, Ströhl F, *et al* (2018) FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell* 173: 720-734.e15
- Rabelo-Fernández RJ, Noriega Rivera RA, Rivera YS, Tous-Beveraggi J, Valiyeva F & Vivas-Mejia PE (2022a) Increased Expression of the RBPMS Splice Variants Inhibits Cell Proliferation in Ovarian Cancer Cells. *Int J Mol Sci* 23
- Rabelo-Fernández RJ, Santiago-Sánchez GS, Sharma RK, Roche-Lima A, Carrion KC, Rivera RAN, Quiñones-Díaz BI, Rajasekaran S, Siddiqui J, Miles W, *et al* (2022b) Reduced RBPMS Levels Promote Cell Proliferation and Decrease Cisplatin Sensitivity in Ovarian Cancer Cells. *Int J Mol Sci* 23: 535
- Raj B, Irimia M, Braunschweig U, Sterne-Weiler T, O’Hanlon D, Lin ZY, Chen GI, Easton LE, Ule J, Gingras AC, *et al* (2014) A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol Cell* 56: 90–103
- Rasmussen EB & Lis JT (1993) In vivo transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc Natl Acad Sci U S A* 90: 7923–7927
- Robberson BL, Cote GJ & Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10: 84–94
- Roberts GC, Gooding C, Mak HY, Proudfoot NJ & Smith CWJ (1998) Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res* 26: 5568–5572
- Roesch A, Becker B, Meyer S, Hafner C, Johannes P, Landthaler M & Vogt T (2005) Overexpression and hyperphosphorylation of retinoblastoma protein in the progression of malignant melanoma. 565–572
- Rogers DM, Jasim SB, Dyer NT, Auvray F, Réfrégiers M & Hirst JD (2019) Electronic Circular Dichroism Spectroscopy of Proteins. *Chem* 5: 2751–2774

- Sabari BR, Dall'Agnese A & Young RA (2020) Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* 45: 961–977
- Sagnol S, Yang Y, Bessin Y, Allemand F, Hapkova I, Notarnicola C, Guichou JF, Faure S, Labesse G & De Santa Barbara P (2014) Homodimerization of RBPMS2 through a new RRM-interaction motif is necessary to control smooth muscle plasticity. *Nucleic Acids Res* 42: 10173–10184
- Schwartz S, Meshorer E & Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16: 990–995
- Shamnas v M, Singh A, Kumar A, Mishra GP & Sinha SK (2024) Exitrons: offering new roles to retained introns—the novel regulators of protein diversity and utility. *AoB Plants* 16: plae014
- Shankman LS, Gomez D, Cherepanova OA, Salmon M, Alencar GF, Haskins RM, Swiatlowska P, Newman AAC, Greene ES, Straub AC, *et al* (2015) KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med* 21: 628–637
- Shanmugaapriya S, van Caam A, de Kroon L, Vitters EL, Walgreen B, van Beuningen H, Davidson EB & van der Kraan PM (2016) Expression of TGF- β Signaling Regulator RBPMS (RNA-Binding Protein With Multiple Splicing) Is Regulated by IL-1 β and TGF- β Superfamily Members, and Decreased in Aged and Osteoarthritic Cartilage. *Cartilage* 7: 333–345
- Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Qiu J, Jiang L, Li H, Chen G, *et al* (2014) Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol* 21: 997–1005
- Sharma S, Maris C, Allain FH. & Black DL (2011) U1 snRNA Directly Interacts with Polypyrimidine Tract Binding Protein During Splicing Repression. *Mol Cell* 41: 579–588
- Sharp PA (1987) Splicing of Messenger RNA Precursors. *Science (80-)* 235: 766–771
- Shi D-L (2024) RNA-Binding Proteins in Cardiomyopathies. *J Cardiovasc Dev Dis* 11: 88
- Shimamoto A, Kitao S, Ichikawa K, Suzuki N, Yamabe Y, Imamura O, Tokutake Y, Satoh M, Matsumoto T, Kuromitsu J, *et al* (1996) A unique human gene that spans over 230 kb in the human chromosome 8p11-12 and codes multiple family proteins sharing RNA-binding motifs. *Proc Natl Acad Sci U S A* 93: 10913–10917
- Shmueli MD, Hizkiahou N, Peled S, Gazit E & Segal D (2017) Total proteome turbidity assay for tracking global protein aggregation in the natural cellular environment. *J Biol Methods* 4: e69
- Shorrock HK, Gillingwater TH & Groen EJN (2018) Overview of Current Drugs and Molecules in

- Development for Spinal Muscular Atrophy Therapy. *Drugs* 78: 293–305
- Shukla S & Fisher SA (2008) Tra2 β as a novel mediator of vascular smooth muscle diversification. *Circ Res* 103: 485–492
- Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR & Kielkopf CL (2006) Structural Basis for Polypyrimidine Tract Recognition by the Essential Pre-mRNA Splicing Factor U2AF65. 49–59
- Song HW, Cauffman K, Chan AP, Zhou Y, King M Lou, Etkin LD & Kloc M (2007) Hermes RNA-binding protein targets RNAs-encoding proteins involved in meiotic maturation, early cleavage, and germline development. *Differentiation* 75: 519–528
- Soranno A, Koenig I, Borgia MB, Hofmann H, Zosel F, Nettels D & Schuler B (2014) Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc Natl Acad Sci U S A* 111: 4874–4879
- Soufari H & MacKereth CD (2017) Conserved binding of GCAC motifs by MEC-8, couch potato, and the RBPMS protein family. *Rna* 23: 308–316
- Southby J, Gooding C & Smith CWJ (1999) Polypyrimidine Tract Binding Protein Functions as a Repressor To Regulate Alternative Splicing of α -Actinin Mutually Exclusive Exons. *Mol Cell Biol* 19: 2699–2711
- Spellman R, Llorian M & Smith CWJ (2007) Crossregulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Mol Cell* 27: 420–434
- Spellman R & Smith CWJ (2006) Novel modes of splicing repression by PTB. 31: 2–5
- Spitz F & Furlong EEM (2012) Transcription factors: From enhancer binding to developmental control. *Nat Rev Genet* 13: 613–626
- Sun Y, Ding L, Zhang H, Han J, Yang X, Yan J, Zhu Y, Li J, Song H & Ye Q (2006) Potentiation of Smad-mediated transcriptional activation by the RNA-binding protein RBPMS. *Nucleic Acids Res* 34: 6314–6326
- Sun Z, Diaz Z, Fang X, Hart MP, Chesi A, Shorter J & Gitler AD (2011) Molecular determinants and genetic modifiers of aggregation and toxicity for the als disease protein fus/tls. *PLoS Biol* 9
- Tange T, Damgaard CK, Guth S & Valca J (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J* 20: 5748–5758
- Tao Y, Zhang Q, Wang H, Yang X & Mu H (2024) Alternative splicing and related RNA binding proteins

in human health and disease. *Signal Transduct Target Ther* 9

- Tauber D, Tauber G & Parker R (2020) Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem Sci* 45: 764–778
- Teplova M, Farazi TA, Tuschl T, Patel DJ, Program B & Rockefeller T (2016) Structural basis underlying CAC RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. *Q Rev Biophys*: 1–20
- Tholen J & Galej WP (2022) Structural studies of the spliceosome : Bridging the gaps. *Curr Opin Struct Biol* 77: 102461
- Thompson JR, Marcelino LA & Polz MF (2002) Heteroduplexes in mixed-template amplifications: Formation, consequence and elimination by ‘reconditioning PCR’. *Nucleic Acids Res* 30: 2083–2088
- Tiffany ML & Krimm S (1972) Effect of Temperature on the Circular Dichroism Spectra of Polypeptides in the Extended State. *Biopolymers* 11: 2309–2316
- Timmis A, Aboyans V, Vardas P, Townsend N, Torbica A, Kavousi M, Boriani G, Huculeci R, Kazakiewicz D, Scherr D, *et al* (2024) European Society of Cardiology : the 2023 Atlas of Cardiovascular Disease Statistics. *Eur Heart J* 00: 1–44
- Tisserant A, Ko H & Gmbh FK (2008) Signal-Regulated Pre-mRNA Occupancy by the General Splicing Factor U2AF. *PLoS One* 3: 1–7
- Tress ML, Abascal F & Valencia A (2017) Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem Sci* 42: 98–110
- Ule J & Blencowe BJ (2019) Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* 76: 329–345
- Uversky VN (2009) Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, Counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28: 305–325
- Veer EP Van Der, Bruin RG De, Kraaijeveld AO, Vries MR De, Bot I, Pera T, Segers FM, Trompet S, Gils JM Van, Roeten MK, *et al* (2013) Quaking, an RNA-Binding Protein, Is a Critical Regulator of Vascular Smooth Muscle Cell Phenotype. *Circ Res* 113: 1065–1075
- Wagner EJ & Garcia-Blanco MA (2001) Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Mol Cell Biol* 21: 3281–3288

- Walko M, Hewitt E, Radford SE & Wilson AJ (2019) Design and synthesis of cysteine-specific labels for photo-crosslinking studies. *RSC Adv* 9: 7610–7614
- Wan Y & Larson DR (2018) Splicing heterogeneity: Separating signal from noise. *Genome Biol* 19: 1–10
- Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP & Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470–476
- Wang J, Choi JM, Holehouse AS, Lee HO, Zhang X, Jahnel M, Maharana S, Lemaitre R, Pozniakovsky A, Drechsel D, *et al* (2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* 174: 688-699.e16
- Wang S, Sun Z, Lei Z & Zhang H (2022) RNA-binding proteins and cancer metastasis. *Semin Cancer Biol* 86: 748–768
- Wang Z & Burge CB (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *Rna* 14: 802–813
- Wei W, Mu S, Heiner M, Fu X, Cao L, Gong X, Bindereif A & Hui J (2012) YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. *40: 8622–8636*
- Westbrook ER, Ford HZ, Antolović V & Chubb JR (2023) Clearing the slate: RNA turnover to enable cell state switching? *Dev* 150: 1–5
- Whitelaw E & Proudfoot N (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* 5: 2915–2922
- Wieczorek DF, Smith CW & Nadal-Ginard B (1988) The rat alpha-tropomyosin gene generates a minimum of six different mRNAs coding for striated, smooth, and nonmuscle isoforms by alternative splicing. *Mol Cell Biol* 8: 679–694
- Will CL & Luhrmann R (2011) Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol*: 1–23
- Wongpalee SP, Vashisht A, Sharma S, Chui D, Wohlschlegel JA & Black DL (2016) Large-scale remodeling of a repressed exon ribonucleoprotein to an exon definition complex active for splicing. *Elife* 5: 1–26
- Woodruff RC, Tong X, Khan SS, Shah NS, Jackson SL, Loustalot F & Vaughan AS (2024) Trends in Cardiovascular Disease Mortality Rates and Excess Deaths, 2010–2022. *Am Jounral Prev Med* 4:

- Woodsmith J, Casado-Medrano V, Benlasfer N, Eccles RL, Hutten S, Heine CL, Thormann V, Abou-Ajram C, Rocks O, Dormann D, *et al* (2018) Interaction modulation through arrays of clustered methyl-arginine protein modifications. *Life Sci Alliance* 1: 1–14
- Wright CJ, Smith CWJ & Jiggins CD (2022) Alternative splicing as a source of phenotypic diversity.
- Xu S, Lai S-K, Sim DY, Ang WSL, Li HY & Roca X (2022) SRRM2 organizes splicing condensates to regulate alternative splicing. *Nucleic Acids Res* 50: 8599–8614
- Xu X, Yang D, Ding J-H, Wang W, Chu P-H, Dalton ND, Wang H-Y, Bermingham JR, Ye Z, Liu F, *et al* (2005) ASF/SF2-Regulated CaMKII δ Alternative Splicing Temporally Reprograms Excitation-Contraction Coupling in Cardiac Muscle. *Cell* 120: 59–72
- Xue C, Lin TY, Chang D & Guo Z (2017) Thioflavin T as an amyloid dye: Fibril quantification, optimal concentration and effect on aggregation. *R Soc Open Sci* 4
- Yang P, Mathieu C, Kolaitis R-M, Zhang P, Messing J, Yurtsever U, Yang Z, Wu J, Li Y, Pan Q, *et al* (2020) G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* 181: 325-345.e28
- Yang Y, Lee GC, Nakagaki-silva E, Huang Y, Peacey M, Partridge R, Gooding C & Smith CWJ (2023) Cell-type specific regulator RBPMS switches alternative splicing via higher -or der oligomerization and heterotypic interactions with other splicing regulators. *Nucleic Acids Res*: 1–22
- Yano M, Hayakawa-Yano Y, Mele A & Darnell RB (2010) Nova2 Regulates Neuronal Migration through an RNA Switch in Disabled-1 Signaling. *Neuron* 66: 848–858
- Yao Z, Xu N, Shang G, Wang H, Tao H, Wang Y, Qin Z, Tan S, Feng J, Zhu J, *et al* (2023) Proteogenomics of different urothelial bladder cancer stages reveals distinct molecular features for papillary cancer and carcinoma in situ. *Nat Commun* 14
- Ye R, Hu N, Ye R, Hu N, Cao C, Su R, Xu S, Yang C & Zhou X (2023) Capture RIC-seq reveals positional rules of PTBP1- associated RNA loops in splicing regulation Article Capture RIC-seq reveals positional rules of PTBP1-associated RNA loops in splicing regulation. *Mol Cell*: 1–17
- Ying Y, Wang XJ, Vuong CK, Lin CH, Damianov A & Black DL (2017) Splicing Activation by Rbfox Requires Self-Aggregation through Its Tyrosine-Rich Domain. *Cell* 170: 312-323.e10
- Young G, Hundt N, Cole D, Fineberg A, Andrecka J, Tyler A, Olerinyova A, Ansari A, Marklund EG, Collier MP, *et al* (2018) Quantitative mass imaging of single biological macromolecules. *Science* (80-)

360: 423–427

- Yu M, Yi J, Qiu Q, Yao D, Li J, Yang J, Mi C, Zhou L, Lu B, Lu W, *et al* (2023) Pan-cancer tRNA-derived fragment CAT1 coordinates RBPMS to stabilize NOTCH2 mRNA to promote tumorigenesis. *Cell Rep* 42: 113408
- Zaborowska J, Egloff S & Murphy S (2016) The pol II CTD: New twists in the tail. *Nat Struct Mol Biol* 23: 771–777
- Zhang H & Fisher SA (2007) Conditioning effect of blood flow on resistance artery smooth muscle myosin phosphatase. *Circ Res* 100: 730–737
- Zhang S, Aibara S, Vos SM, Agafonov DE, Lührmann R & Cramer P (2021) Structure of a transcribing RNA polymerase II-U1 snRNP complex. *Science (80-)* 371: 305–309
- Zhang XHF & Chasin LA (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18: 1241–1250
- Zhang Z, Kumar V, Dybkov O, Will CL, Zhong J, Ludwig SEJ, Urlaub H, Kastner B, Stark H & Lührmann R (2024) Structural insights into the cross-exon to cross-intron spliceosome switch. *Nature* 630: 216–221
- Zhong Q, Xiao X, Qiu Y, Xu Z, Chen C, Chong B, Zhao X, Hai S, Li S, An Z, *et al* (2023) Protein posttranslational modifications in health and diseases: Functions, regulatory mechanisms, and therapeutic implications. *MedComm* 4: 1–112
- Zhou D, Couture S & Scott MS (2021) RBFOX2 alters splicing outcome in distinct binding. *8000*: 1–14
- Zhu J, Mayeda A & Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8: 1351–1361