The Involvement of SRSF1 in pre-mRNA splicing.

By Andrew Michael Jobbins.

Department of Molecular and Cell Biology, University of Leicester.

Supervisor: Professor Ian Eperon.

Student number: 119003682

Abstract.

The key splicing signals in pre-mRNA, the branch-point, 5' splice site and 3' splice site, are exceptionally poorly conserved in mammals. To compensate for this, key regulatory sequences are required to direct the reciprocal factors, U2AF, U1 snRNP and U2 snRNP respectively, to the correct sites. These regulatory sequences function by recruiting activators, of which the main family are the SR proteins, or repressors, of which the main family are the hnRNP proteins, which in turn stimulate or repress the binding of key spliceosomal factors. The archetypal SR protein is SRSF1. SRSF1 was the first non-snRNP factor identified and the first found to control alternative splicing. Its best-understood activity is to stimulate the inclusion of exons by binding to purine-rich exonic splicing enhancer (ESE) sequences. There is also some evidence suggesting its involvement in constitutive splicing, which began with demonstrations that it could compensate for depletion of the U1 snRNP. However, further investigations into both its recruitment via ESEs and its possible role in constitutive splicing have foundered due to apparent non-stoichiometric binding. Single molecule experiments allow us to look at the exact number of SRSF1 proteins that bind. The experiments outlined here indicate that the U1 snRNP can actually recruit SRSF1 in a stoichiometric manner. This implicates a possible recruitment mechanism for SRSF1 which would allow it to play a role in core splicing reactions and exon definition. Furthermore we demonstrate that with increasing numbers of enhancers, which sequentially increase splicing efficiency, the number of SRSF1 proteins bound does not change but the chance of a protein binding event increases. This fits a model in which the initial binding of SRSF1 is weak and transient. The same construct is also used to show that introducing a non RNA link in between an ESE and its target site does not silence the ESEs effect, indicating that ESEs exert their effect via RNA loops.

Acknowledgements.

I would like to start by thanking my supervisor, Professor Ian Eperon, for the opportunity to undertake this research as well as his unending daily support. I would also like to thank Dr Andrew Hudson, Dr Glenn Burley and Dr Linus Reichenbach for their vital collaborations which made all of this work possible. I am grateful to Christian Lucas, Dr Olga Makarova, Dr Lucy Eperon, Dr Andrey Revyakin, Dr Cyril Dominguez and Dr Dimitry Cherny for their help throughout my PhD and in some cases prior to it. I would also like to thank Christian, Li, Oksanna, Robert, Alia, Carika and Tilemachos for their help and friendship.

Unfortunately as I write this, it is with great sadness that I mourn the loss of my Grandpa. This is particularly difficult given the contribution he made to my passion for science at such a young age. This also brings into light the other person who is sadly no longer with us but without whom this work would not have been possible, my Mum. Despite not being able to witness the completion or even start of my PhD, her encouragement, kindness and positivity all still contribute to my life in many ways, including my work. It is for these reasons that I would like to dedicate this work to both my Grandpa and Mum.

Furthermore I would like to particularly thank my Dad and Grandma for their unending support and encouragement. Their unselfish nature continually serves to help me achieve all my goals whilst also serving as an example for the person I strive to be. I would also like to thank my brothers, Nan and Grandad whom have all contributed to me being where I am today in their own special way. Last but not least, I would like to extend my thanks to my girlfriend Rebecca and my best friend Jack who I couldn't do this without as well as all the other friends that have provided so many welcome distractions during my PhD.

Contents.

Abbreviations.

A – Adenine

AMPS – Ammonium persulfate

ATP - Adenosine Triphosphate

BP – Base pairs

BSA – Bovine serum albumin

C – Cytosine

CrPi – Creatine phosphate

Cy5 – Cyanine 5

DNA - Deoxyribonucleic Acid

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

EMCCD – Electron-multiplying charge-coupled device

ESE – Exonic splicing enhancer

ESS exonic splicing silencer

EtBr – Ethidium bromide

G – Guanine

GFP – Green fluorescent protein

HEPES – N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid

hnRNP – Heterogeneous nuclear ribonucleoproteins

IP – Immunoprecipitation

ISE – Intronic splicing enhancer

ISS - Intronic splicing silencer

KGlu – Potassium Glutamate

LED – Light emitting diode

mRNA - messenger Ribonucleic Acid

NP40 – Nonidet p-40

NT – Nucleotide

Oligo – Oligonucleotide

PAGE – Poly Acrylamide Gel Electrophoresis

PBS – Phosphate Buffered Saline

PCR – polymerase chain reaction

PEG – Polyethylene Glycol

Pre-mRNA – precursor messenger RNA

PTB – Polypyrimidine binding protein

R – Arginine

RNA – Ribonucleic Acid

RRM – RNA recognition motif

RS domain – Arginine serine domain

S – Serine

SDS – Sodium dodecyl sulphate

SF1 - Splicing Factor 1

SF3A/B - Splicing Factor 3A/B

SMN – Survival of the Motor Neurone protein

snRNA – small nuclear RNA

snRNP – small nuclear RiboNuclear Particle

SRPK1 – Serine/threonine-protein kinase 1

SRSF1 – Serine arginine splicing factor 1

T – Thymidine

TBS – Tris Buffered Saline

TBST – Tris Buffered Saline with TWEEN

TEMED – N,N,N’,N’-Tetramethylethylenediamine

TIRF – Total internal reflection fluorescence

TIRFM - Total internal reflection fluorescence microscopy

Tra2β – Transformer protein 2 beta

Tris – Tris(hydroxyethyl)aminomethane

tRNA – Transfer RNA

U – Uracil

U2AF – U2 auxiliary factor

W/V – weight by volume

Chapter 1. Introduction.

1.1. Splicing.

1.1.1. Pre-mRNA splicing.

In humans more than 99% of protein-coding gene transcripts are transcribed containing non-coding sequences. These non-coding sequences are known as introns whilst the coding sequences are called exons. In order for the transcript to be successfully exported from the nucleus and subsequently undergo translation, the introns need to be removed and the exons stitched back together (figure 1A). This process is known as RNA splicing (Berget et al. 1977; Chow et al. 1977).

There are on average 8.8 exons per gene and 6.3 different transcripts produced per gene (Dunham et al. 2012). It is this ability to produce a number of different transcripts from the a single gene that allows the relatively few protein coding genes found in the human genome, 20 687 vs 37 544 for the rice genome, to produce the level of complexity that is witnessed in humans (Dunham et al. 2012). In fact, the level and complexity of splicing patterns is one of the major differences found between species, with more complex splicing being found in more complex organisms (Barbosa-Morais et al. 2012).

For the splicing of pre-mRNA in the nucleus of nearly all higher organisms, macromolecular machines called spliceosomes are required. These spliceosomes assemble in an ordered manner on the RNA and consist of small nuclear ribo nucleoprotein particles (snRNPs) as well as over 200 extra proteins (Wahl et al. 2009). A single spliceosome consists of five snRNPs, U1, U2, U4, U5, and U6 which are formed of distinct small nuclear RNA (snRNA), Sm proteins and snRNP specific proteins. The U4, U5

and U6 snRNPs assemble and bind together as a single unit known as the tri-snRNP (Nguyen et al. 2015).

Introns spliced by the spliceosome can be placed into two specific categories depending on their splice sites and the snRNPs that assemble upon them. The first are the U2-type introns which constitute around 99.8% of all introns and are spliced by the major spliceosome which consists of the aforementioned U1, U2, U4, U5 and U6 snRNPs (Sheth et al. 2006). The second much rarer class are the U12-type introns, these are spliced in the same manner but utilize the minor spliceosome which consists of the U11, U12, U4atac, U5 and U6atac (Turunen et al. 2013). Both spliceosomes share the use of the U5 snRNP which is thought to be the snRNP that co-ordinates the catalytic activities of the spliceosomes (Patel & Steitz 2003). The catalytic component of the spliceosome also requires two magnesium ions, termed M1 and M2, which are coordinated by specific nucleotides from the U6 snRNA and play reciprocal roles during the two consecutive transesterification reactions (Fica et al. 2013).

In order for spliceosomes to assemble on pre-mRNA in the correct manner four key signals within the RNA need to be recognised. These are shown in figure 1B and consist of the 5’ splice site, 3’ splice site, Branch point and the polypyrimidine tract. The 5’ splice site constitutes the sequence at the 5’ end of an intron and crosses the intron/exon junction; the consensus for this site is G-GURAGU (where – is the intron/exon boundary and R is either A or G) (Sheth et al. 2006). The 3’ splice site similarly denotes the intron/exon boundary but at the 3’ end of the intron and the consensus sequence here is simply YAG- (where Y is either C or T and – is the intron/exon boundary). The branch point lies between 20 to 40 nucleotides upstream of the 3’ splice site and consists of a

weakly conserved A that is nonetheless crucial for the first step in splicing (Zhou & Reed 1998). In between the branch point and 3' splice site lies the polypyrimidine tract which, as the name suggests, consists of around 12 pyrimidines which are important for the correct recognition of the 3' splice site (Wagner & Garcia-Blanco 2001).

Once the key signals in the RNA have been recognised/bound by their corresponding components, the spliceosome can form. The spliceosome forms a series of complexes that ultimately manipulate the pre-mRNA into a conformation whereby two successive S_N2-type transesterification reactions can occur. These conformational changes are shown in a greatly simplified manner in figure 1C.

The first trans-esterification reaction occurs when the 2'OH group of the conserved A from the branch point performs a nucleophilic attack on the 5' end phosphate at the 5' splice site. This results in a free 5' exon, which is held by the spliceosome, and an intron lariat – 3' exon. The 3'OH that is now present on the 5' exon is then positioned so that it can perform a nucleophilic attach on the 5' end phosphate at the 3' splice site. This results in the two exons being ligated together and the intron being removed.

Once the pre-mRNA has been spliced into its correct mRNA it is subsequently targeted for export to the cytoplasm via the exon junction complexes (EJC) that form at the boundaries between exons after splicing has occurred (Tange et al. 2004). If splicing does not occur correctly however, then the RNA cannot be exported correctly and this results in it being degraded. The other product produced from splicing, the intron lariat, is also commonly targeted for degradation; however, recent data has suggested that a sub set of introns may be de-branched and exported to the cytoplasm to undergo other cellular functions (Garrey et al. 2014)(Talhouarne & Gall 2014).

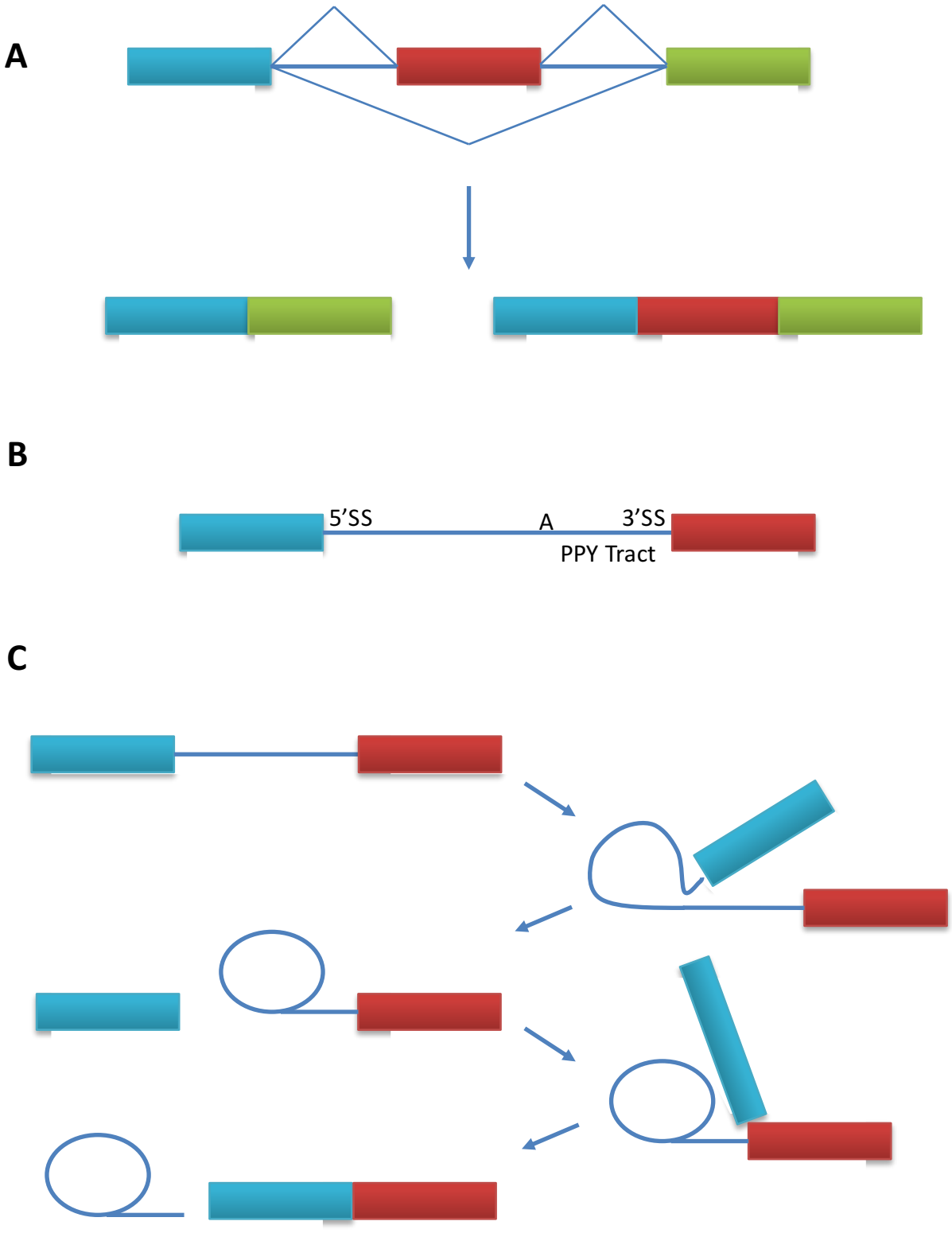


Figure 1. Basics of alternative splicing, the key signals involved and the steps in the reaction path. A) Diagram to show the possible spliced variants that can be produced via alternative splicing; exon skipping results in the exclusion of the red exon whilst exon inclusion results in the longer isoform that contains the central red exon. B) Diagram showing the location of the key signals that need to be identified for splicing to occur (from left to right; the 5' splice site, the branch point A, the polypyrimidine tract and the 3' splice site). C) Flow diagram showing the two sequential transesterification reaction, the intermediates that are produced and the final products.

1.1.2. Spliceosome assembly.

The spliceosome assembles in a series of complexes shown in figure 2A; A (pre-spliceosome), B (pre-catalytic complex), B^{act} (the activated spliceosome), B* (the catalytically activated spliceosome), C (post-step 1 spliceosome), C* (post-step 1 catalytically activated spliceosome), P (post splicing) and the intron lariat spliceosome (ILS) (Wahl et al. 2009). There are also additional complexes, H, I and E, that form before complex A that have been identified in vitro but are either not believed to form in vivo or form so rapidly and transiently that they cannot be observed (Larson & Hoskins 2017). Key differences between these complexes can be found in both their composition and conformation. The composition of the constituent snRNPs are shown in figure 2B.

As mentioned previously the key signals that initiate splicing are the 5'SS, branch-point, poly-pyrimidine tract and the 3' SS, which are recognized by select components of the spliceosome: the U1 snRNP (Mount et al. 1983), SF1 (Berglund et al. 1997), U2AF65 and U2AF35 (Zorio & Blumenthal 1999) respectively. Upon the binding of these factors the early spliceosome or E complex is said to have formed; at this point splicing is said to be committed but the splice sites are not fixed (Wu & Manley 1989; DONMEZ et al. 2004). However it is possible that E complex is an artefact of in vitro systems and a different complex, I, might be a better approximation for the step before complex A; this is discussed at the end of the section. The U2 snRNP then base pairs to the branch-point and the ATPases Prp5 (O'Day et al. 1996) and Sub2/UAP56 (Shen et al. 2007) drive A complex formation. The base pairing of the U2 snRNA and the branch-point is stabilized by the ATP-dependent binding of additional U2-associated proteins from the splicing factor SF3A/B complex to the anchoring site just

upstream of the branch point (Das et al. 2000; Gozani et al. 1996) and the RS domain of U2AF65 (Valcárcel et al. 1996). This stable association of U2 leads to the displacement of SF1 from the branch-point and the interaction at the branch-point A is replaced by SF3b14a (Will et al. 2001) whilst SF3b155 interacts with the C-terminal RNA recognition motif of U2AF65 (Gozani et al. 1998).

This conformation in turn serves as a binding platform for the recruitment of the U4, U5 and U6-containing tri-snRNP. Upon the binding of the tri-snRNP, B complex is said to have formed; this is the only time all the snRNPs are bound together. At this point and in A complex, the branch-point is recognized via base pairing by the U2 snRNA (Wassarman & Steitz 1993). The transition from B to B^{act} is predominantly driven by the ATPases Brr2 (Raghunathan & Guthrie 1998), which unwinds the base paired U4 and U6 snRNA, and Prp28 (Staley & Guthrie 1999), which displaces the U1 snRNP at the5' splice site. This in turn leads to U4 and U1 disassociating; U4 due to no longer being bound by the U6 snRNA and U1 due to being displaced as a result of Prp28 by U6 snRNA at the 5' SS. 30 additional proteins are also recruited at this point (Wahl et al. 2009). The new proteins recruited predominantly come in the form of the Prp19–CDC5L complex or the Nineteen complex (NTC) in yeast (Chan & Cheng 2005). Unlike the snRNPs, components of this complex are loosely associated with each other.

The complex is then reformed again, mainly by Prp2, and a catalytically active version of the spliceosome, B*, forms (Bessonov et al. 2010). The first transesterification reaction can then occur, giving C complex which contains a free exon 1 and an intron lariat - exon 2. Prior to the second catalytic step, additional rearrangements occur, mainly driven by Prp16 (Villa & Guthrie 2005) and an activated version of C complex forms, C*. After the

second catalytic step has been completed the post splicing complex (P) forms which contains the fully spliced mRNA and the intron lariat. The mRNA is then released with a number of proteins attached to it including those found at EJCs and the intron lariat remains attached to the ILS. The snRNPs are then recycled for further splicing reactions.

As mentioned before, these main splicing complexes are preceded by a sub-complex called complex H. This complex involves the binding of many regulatory proteins such as SR proteins or hnRNPs. These proteins combine to create an environment where either the early splicing components can bind e.g. U1, SF1 etc. or they can't. Upon U1 binding the complex becomes E or I complex.

As mentioned previously there is an argument that E complex is an artefact of in vitro systems as the concentration of ATP never reaches 0 in cells. More recently a new complex, complex I, has been proposed. This complex is generated by removing ATP but preventing the de-phosphorylation of proteins with phosphatases (Chen et al. 2016). In this manner the two properties of ATP in the early complexes are separated, the phosphorylation of proteins and driving ATPases. This is perhaps a better model for what happens before complex A formation as proteins such as SR proteins, which are open to extensive phosphorylation and are key to early splicing complex formation, remain in their phosphorylated form but the ATPases which drive further complex progression are inhibited. The argument for complex I over complex E is based on single molecule data that shows that many of the early splicing factors bind in a non-specific manner in E, even when they have been shown to via biochemical data, but not in I (Chen et al. 2016). Both complex E and I are looked at in this thesis as useful information can be gained

from both. It is within these early complexes, H, E or I, that splice site recognition, not selection, is believed to occur (Michaud & Reed 1991).

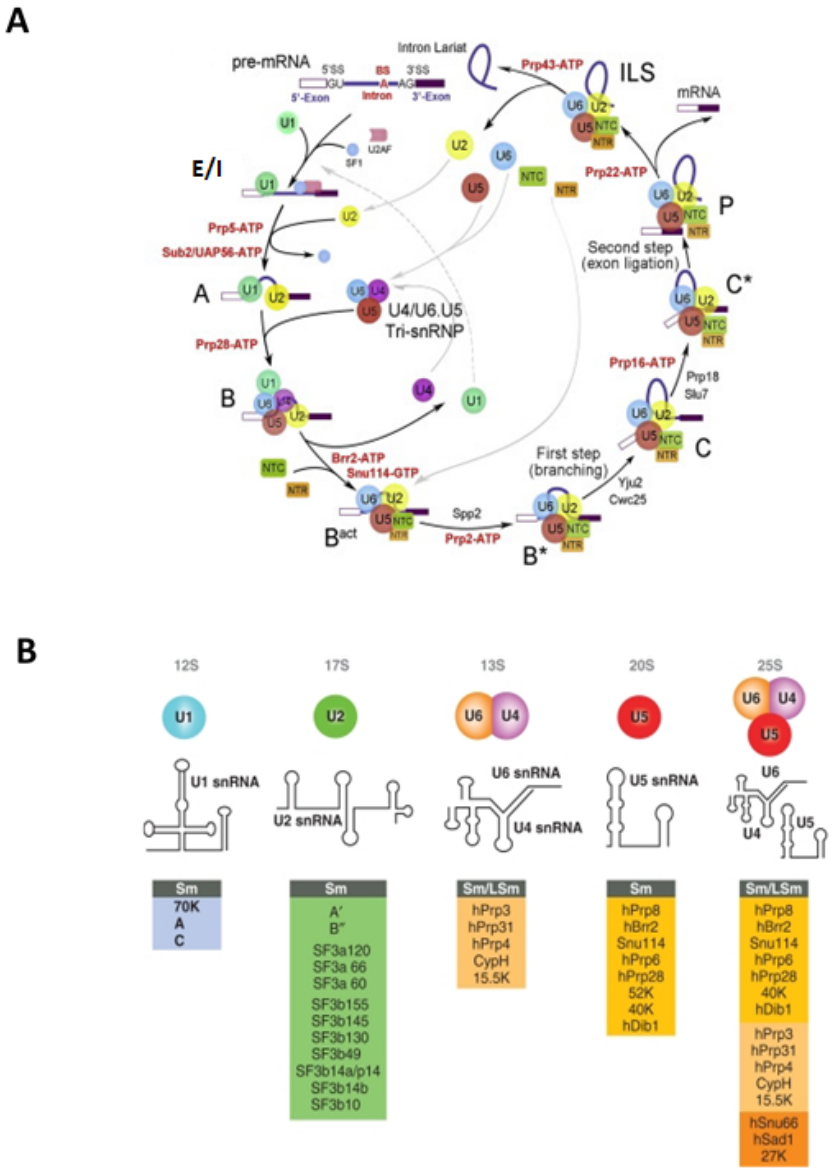


Figure 2. Progression of the spliceosome through the complexes as well as the key proteins involved (Images taken from Shi, 2017 and Will, 2011). (A) The cycle of the complexes in the spliceosome. (B) The main protein components of each of the snRNPs along with how the snRNA of the tri-snRNP assembles.

1.1.3. Structure of the Spliceosome

Whilst the assembly of the spliceosome has been established gradually over thirty plus years via ensemble techniques, the structures of the complexes of the spliceosome have been established over a relatively short period of time. The emergence of cryo-electron microscopy (EM) as an extremely powerful technique for visualising biological structures (Kühlbrandt 2014) has allowed the structure of the spliceosome to be revealed at near atomic resolution. Prior to this the use of standard EM has only been able to solve the structures of constituent parts; the U1 snRNP (Kondo et al. 2015) for example. Whilst these relatively small structures appear minor compared to the structure of the entire B complex for example, there contribution is to not be underestimated.

The earliest spliceosome complex in the assembly pathway to be solved is that of B complex, the first complex in figure 3. In this structure, the domain of the spliceosome that contains the U2 snRNP in figure 3, is connected to the main body of the spliceosome via three main protein bridges; SF3B1, SF3B3 and SF3A1. The U4/U6.U5 tri-snRNP proteins, which are located in the “body”, undergo dramatic rearrangements upon their binding. Most important of these is a closed pocket formed by Prp8 and Dim1; this appears to facilitate U6 snRNA interactions with the 5' splice site. Furthermore, as seen in figure 3, Brr2 is reformed in such a manner so that it is poised to contact the U4-U6 duplex and unwind it; a step that is essential for later spliceosomal complex progression. However Brr2 at this point appears to be held back from this unwinding step by a number of B specific proteins. Upon progression from B complex, when these proteins are released, Brr2 is allowed to unwind the helix (Bertram, Agafonov, Dybkov, et al. 2017).

The next complex in the series is B^{act}. The structure of the human B^{act} has not been solved at this point, but that of the yeast variant has been and given the similarities seen between other complexes the yeast B^{act} is thought to closely resemble that of the human. This structure is prior to the RNA helicase Prp2's remodelling of the spliceosome into B*. The structure here reveals that the catalytic core of the spliceosome, which consists of U2 snRNA, U6 snRNA and Prp8, has formed. The 5' splice site is positioned ready for the first transesterification reaction but is blocked by proteins. To further prevent early catalysis the branched A is sequestered by Hsh155 and held approximately 50 angstroms from the 5' splice site. Together these observations suggest that Prp2s remodelling involves the removal of Hsh155 (SF3B1) and the proteins blocking the 5' splice site in order to allow the Branch-point 5'SS interaction to occur (Yan et al. 2016).

There is no structure for B* and it is indeed possible that this will not be possible as once the aforementioned rearrangements have occurred then the first transesterification reaction is inevitable. However it is expected to be very similar to complex C except for the arrangements of bonds in the RNA.

Following the first transesterification reaction the RNA is formed into a lariat-exon 2 and a free exon 1; this is complex C, the final complex in figure 3. Once again there is as of yet no structure for the human C complex but there is one from yeast. In this complex/structure the 5' splice site is cleaved but remains within the catalytic pocket and the lariat has formed with the 5' phosphate of the 5' end of the intron is linked to the branch-point A 2'OH. Exon 1 is held by Prp8 and in fact base-pairs with the U5 snRNA loop 1. Non-Watson-Crick interactions between the branch-point and the 5' splice site hold the branched A within the catalytic site while base pairing with the U6 snRNA holds

the 5’ end of the intron in place. The intron immediately after the branch-point extends out of the catalytic pocket towards Prp16; the helicase involved in remodelling the spliceosome for step 2 splicing (Galej et al. 2016).

Immediately following Prp16 remodelling the spliceosome is activated for step 2; this is the C* complex. The structure of the human C* complex revealed that the core of the spliceosome is virtually unchanged from the yeast C complex described above. However, in this complex the branch-point A is approximately 20 angstroms from the catalytic core. Furthermore the RNA helicase PRP22, involved in causing the spliceosome to disassociate from spliced RNA, is located about 100 angstroms from the catalytic core; indicating that it may exert its affect at a distance. The predominant alteration driven by PRP16 following the first step of splicing appears to be a large-scale movement of the U2 snRNP (Bertram, Agafonov, Liu, et al. 2017).

Despite the wealth of knowledge that the structures of the spliceosomal complexes have both confirmed and revealed, there are still a number limitations. Due to the dynamic nature of the spliceosome large parts are not fixed and therefore appear unstructured in the structures. This is highlighted further by the number of proteins that have been shown to be present in these complexes but are not in the structures; SRSF1 for example. Furthermore there is a lack of structures for the early complexes of the spliceosome; this is important considering the high level of regulation that occurs in these complexes. It will be interesting to see if the structures of these complexes are solvable, again given their dynamic nature, and what information regarding regulation they will tell us.

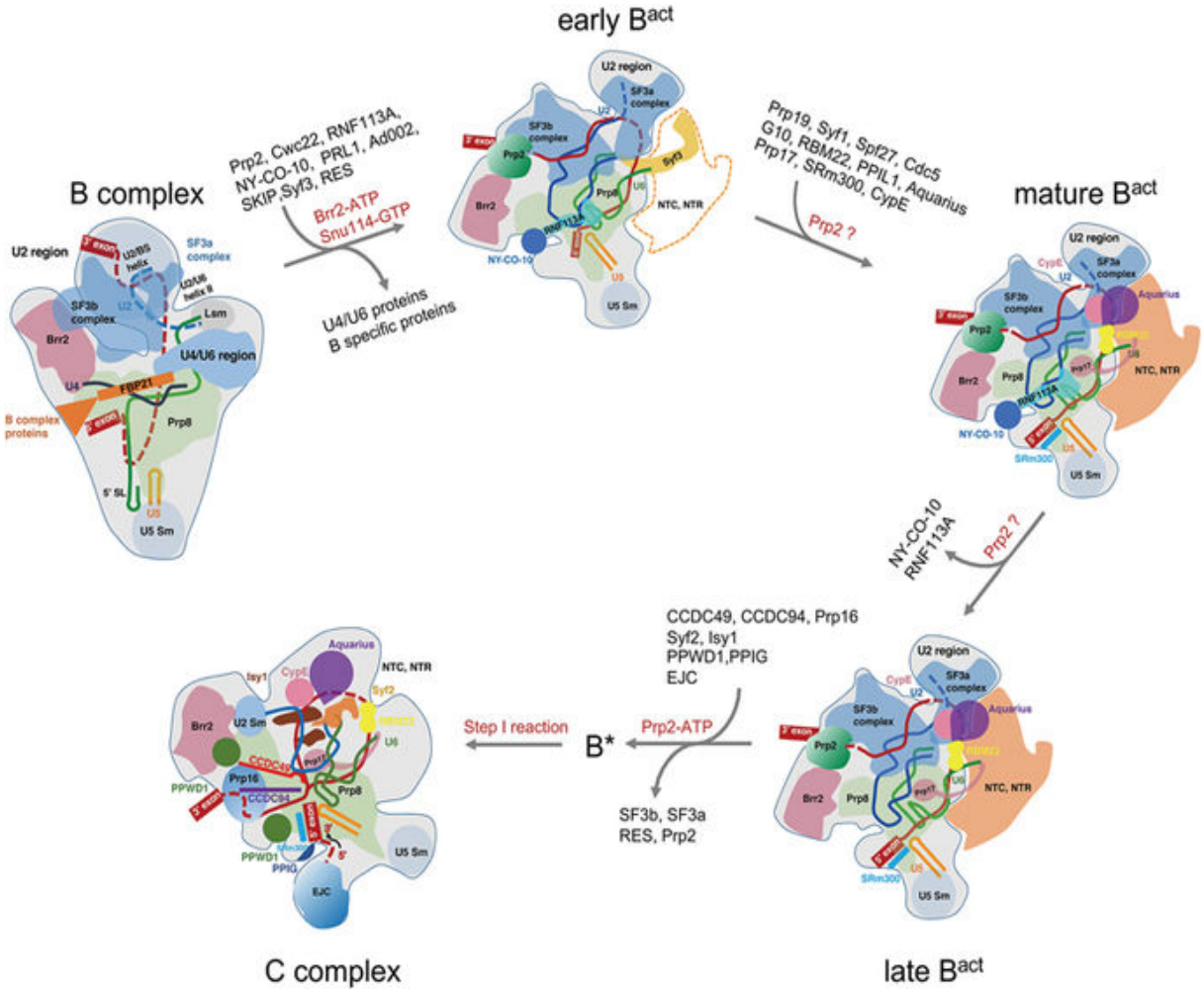


Figure 3. The remodelling from the B complex through to C complex based on the solved structures (Zhang et al. 2018).

1.1.4. The Structure of the U1 snRNP.

The predominant complex of the spliceosome that this thesis deals with is complex A. The U1 snRNPs major role in splicing involves the recognition of the 5' end of the intron which, along with U2 at the 3' end, allows the assembly of the rest of the spliceosome (Mount et al. 1983; Maroney et al. 2000). Crucial information as to how these two functions are fulfilled, RNA recognition and factor recruitment, as well as insights into additional functions, are revealed in the structure of this crucial factor.

U1 snRNPs are comprised of U1 snRNA, seven Sm proteins (SmB/SmB', SmD1, SmD2, SmD3, SmE, SmF and SmG), common to all snRNPs except U6, and three specific proteins (U1-C, U1-70K and U1-A) (Mount et al. 1983; Bringmann & Lührmann 1986). Early EM, using negative staining, revealed a globular core domain with two protrusions accounting for U1-70k and U1-A (Kastner & Lührmann 1989; Kastner et al. 1992). Further structural analysis revealed the doughnut shaped sm ring in between stem loops 3 and 4 and allowed the mapping of the structures of the three specific proteins onto the snRNA. This showed U1-C binding the core of the snRNP, U1-70K binding stem loop 1 and U1-A binding stem loop 2 (Stark et al. 2001).

The structure of the core of the snRNP was later mapped at 5.5 Å (Pomeranz Krummel et al. 2009). This revealed the N-terminal region of U1-70k extends from its RRM, bound to stem loop 1, in a long α-helix and wraps around the sm protein ring to make contact with U1-C. This accounted for the requirement of U1-70k for U1-C binding (Nelissen et al. 1994). However, the resolution was not high enough to map how the protein components of the snRNP influenced the 5' splice site interaction.

Due to difficulties stemming from the inherent mobility of the long RNA helices arranged in the four way junction, two sub-structures of the U1 snRNP were solved in higher resolution (Kondo et al. 2015). These structures are shown mapped onto the 5.5 Å electron density map in figure 4. U1-C was revealed to make no base specific contacts with the 5'SS sequence but it did hydrogen bond with the sugar phosphate backbone of the 5' splice site. This contact allowed U1-C to stabilize binding to some mismatched 5'SS oligonucleotides. Further to the previous structure, the N-terminus of the U1-70K protein was shown to bind to the subunit interfaces between SmD2 and SmF and between SmD3 and SmB.

The structure of the U1 snRNP has revealed a number of crucial bits of information about the binding of its core proteins and how it interacts with 5' splice sites. The organisation of its RNA and core proteins has also assisted in revealing the structures of later complexes. The structures along with biochemical data have also provided known roles for each of the sections of the snRNA except stem loop 3; the 5' end binds the 5' splice site (Krämer et al. 1984), stem loop 1 binds U1-70k, stem loop 2 binds U1-A (Wu & Maniatis 1993) and stem loop 4 can bind PTB (Sharma et al. 2011) and SF3A1 (Sharma et al. 2014a). This lack of interaction means stem loop 3 of the U1 snRNA is free and exposed as seen in figure 4.

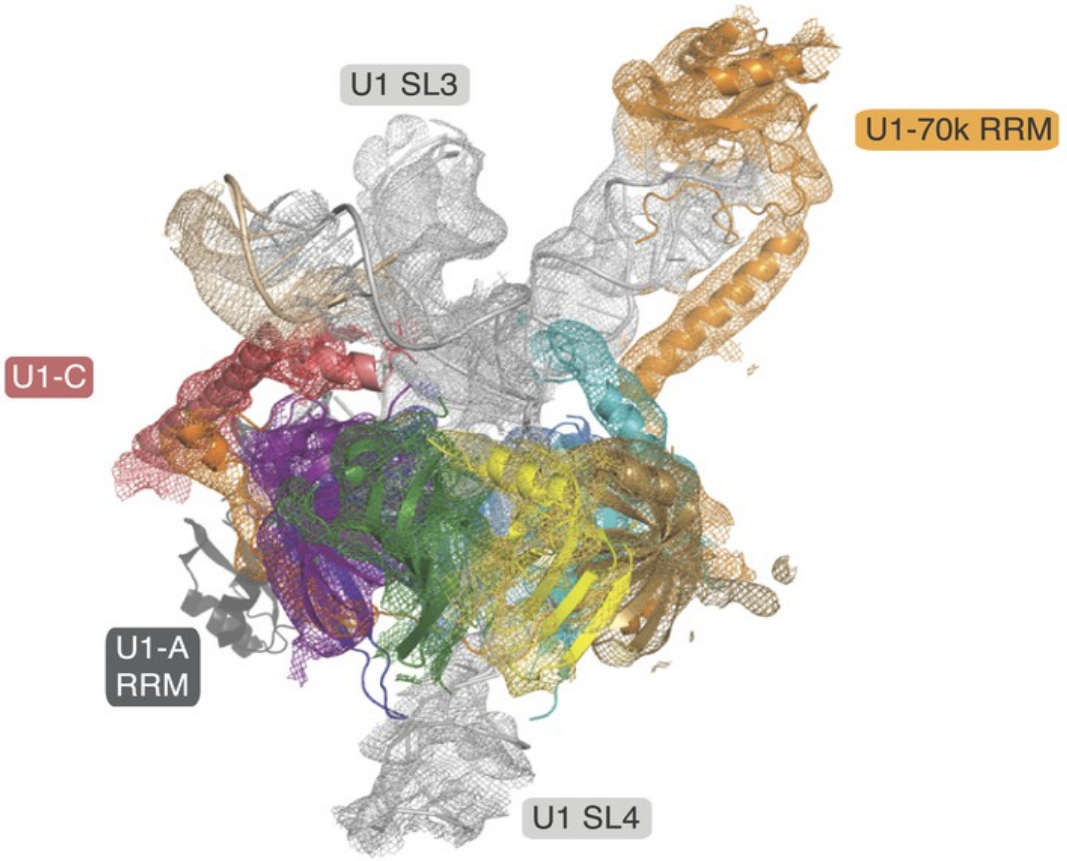


Figure 4. Crystal structures of the two structures (Kondo et al. 2015) **placed into the electron density map generated from the 5.5 Å structure** (Pomeranz Krummel et al. 2009) **displaying the three specific proteins and the exposed stem loops.**

1.1.5. Splice site selection.

One of the primary steps to occur in pre-mRNA splicing is the selection of the splice sites. At the 5' splice the U1 snRNP binds via the base pairing of its snRNA to the pre-mRNA whilst at the 3' splice site U2AF 65 recognises the poly pyrimidine tract, U2AF 35 recognises the 3' splice site and the U2 snRNP binds to the branch point via base pairing. However, selection is not as simple as recognition.

The 5' splice site in mammals is exceptionally poorly conserved with a vast range of sequences being accepted as functional splice sites; more than 9000 unique sequences for a 9nt motif (Roca et al. 2012). In fact there are even a sub-set of 5' splice sites that have been discovered which don't even contain the highly conserved GU; they contain GC instead (Kitamura-Abe et al. 2004). This lack of conservation means that the number of potential 5' splice sites far exceeds the number of used sites. Despite this the primary factor thought to affect which site is chosen is the sequence of the 5' splice site itself. Sites that contain sequences that perfectly match the sequence of the 5' tail of the U1 snRNA are deemed consensus sites. These sites are, as one would expect, able to bind U1 more stably and are therefore favoured for selection (Eperon et al. 1986; Lear et al. 1990). This however does not tell the whole story as intermediate sites with an intermediate affinity can be found to be used extensively (ROCA et al. 2005) and vice versa with strong sites being under-used (Eperon et al. 1993; Nelson & Green 1990).

There are a number of further determinants which determine which sites are actually selected and which are not; with the majority being found in the surrounding sequences; figure 5A. These usually correspond to sites for silencers or enhancers but may even code for certain RNA structural features (Buratti & Baralle 2004). It is often the dynamic

balance between the positive and negative features found in close proximity which determines the eventual outcome.

Similar factors can affect the use of 3’ splice sites. However due to the need for multiple factors to bind within a certain distance of each other, there are less non-used candidate sites than at 5’ splice sites. Although the relative lack of specific sequence requirements at 3’ splice sites also means that mutations, in the RNA or in the RNA recognition features of the proteins, can easily lead to the use of cryptic splice sites.

1.1.6. Exon vs Intron definition.

When the splice sites are selected they then communicate with each other to initiate the formation of the splicing complexes that lead to the excision of a designated intron. However there has been much discussion over whether the splice sites communicate across a designated exon for inclusion or across an intron designated for excision (De Conti et al. 2013).

Intron definition, figure 5B, is widely accepted to occur across short introns where the chance of a collision occurring between a selected 5' splice site and a selected 3' splice site is high enough (Roy et al. 2008). This is furthered by the fact that there a number of known factors which can bridge the splice sites and allow them to be held in close proximity (Abovich & Rosbash 1997; Shao et al. 2012; Kao & Siliciano 1996; Becerra et al. 2015; Wu & Maniatis 1993); although none have been shown to do so individually.

Exon definition, figure 5C, conversely is widely accepted to occur across short exons with long introns. Again this initially fits with models where the chance of an interaction between a downstream 5' splice site and an upstream 3' splice site, as opposed to a downstream 3' splice site, is now far more likely (Hollander et al. 2016). This model is also supported by data which indicates that a strong downstream 5' splice site can promote the use of a weak upstream 3' splice site (Hwang & Cohen 1996; Bateman et al. 1994). However one of the limitations of this model is the lack of a known direct interaction between 5' and 3' splice site factors across an exon. The known intronic bridging factors could not account for this as these factors, Prp40, SF3A1 etc., have not been shown to function across an exon. The generally accepted solution for this problem is that both sets of factors can interact with enhancer proteins and these could thus

bridge the sites and serve to enhance. This is why in cartoons depicting exon definition, exons are often depicted with exonic splicing enhancers that can recruit enhancer proteins in the centre of them. However a number of short exons found in vivo that are thought to undergo exon definition only contain weak potential enhancer sites (Anczuków et al. 2015; Kechris et al. 2008).

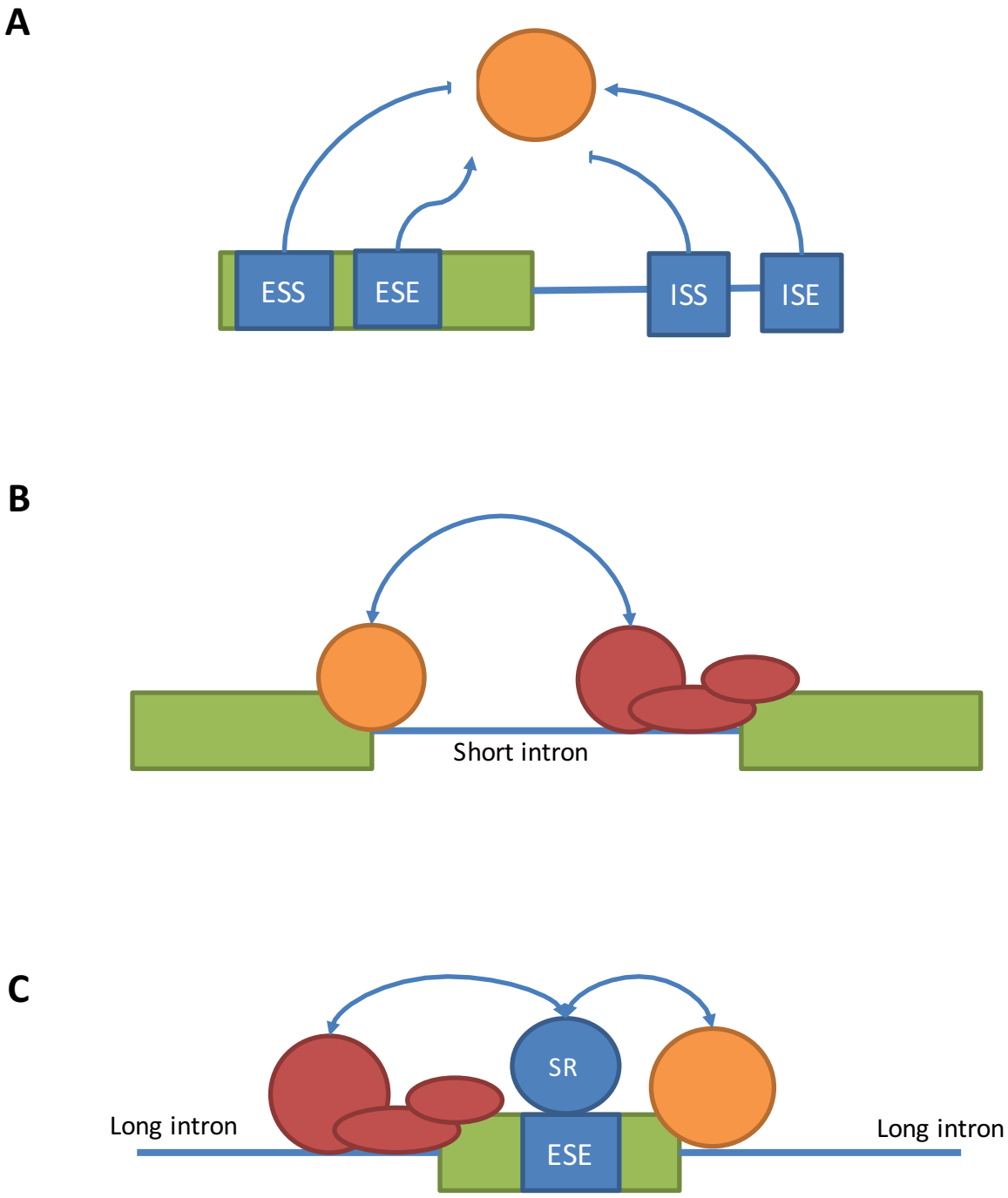


Figure 5. Influence of RNA sequence elements on splice site selection and intron/exon definition models. (A) Cartoon showing the effect of Exonic splicing silencers (ESS), exonic splicing enhancers (ESE), intronic splicing silencers (ISS) and intronic splicing silencers on the recruitment of the U1 snRNP to a 5' splice site. (B) Cartoon depicting two splice sites making contact over a small intron. (C) Cartoon depicting the two splice sites making contact over a short exon before making contact over an intron.

1.2. Splicing regulation.

1.2.1. Regulatory sequences.

The initiation of splicing is regulated by a variety of proteins binding to specific sequences in the RNA. These regulatory sequences are often found flanking weak splice sites or in exons that are alternatively spliced, but in many cases they are found in constitutive exons and near strong splice sites, perhaps as an evolutionary safeguard. The proteins that bind these sequences are then able to stimulate or support the binding of key spliceosomal factors such as U2AF/U1 or alternatively block their binding.

A number of mechanisms have been suggested by which these proteins may function (Xavier Roca et al. 2013; Erkelenz et al. 2013).Many activator/repressor proteins such as SR proteins or hnRNP proteins are able to exert their affects from sites distant from their targets (Lavigueur et al. 1993a; Graveley et al. 1998a). This suggests that long range regulation may well be a common occurrence, as it is in transcription (van Heyningen & Bickmore 2013). However the mechanisms of both long range and short range regulation are still poorly characterised.

The range of sequences that can serve to act in a regulatory manner has expanded greatly over the past decade, in part due to the onset of wide-spread transcriptomic level assays such as CLIP (Sanford et al. 2009; Fu & Ares 2014; Chang et al. n.d.; Ule et al. 2005). These assays allow the mapping of the binding sites of a number of key regulators. Mutational analyses have identified a number of ESEs found in vivo, disruption of which lead to changes in splicing patterns (Tejedor et al. 2015; Scotti & Swanson 2015). SNPs found in vivo have also contributed to our understanding of regulatory sequences. One of the most studied examples of this is in the SMN gene; here

a single C>T mutation results in exon 7 being skipped (Lorson et al. 1999). The reason for this is the disruption of an enhancer or/and the creation of a silencer. These SNPs serve to highlight the fine balance in which these regulatory sequences exist. These methods have identified regulatory sequences associated with 50 or so proteins, and thus potential regulatory sequences can be identified by computational analyses (Lim et al. 2011; Akerman et al. 2009). However, these can result in multiple proteins being mapped to a very small area. Even if the effects of the proteins are validated, for example by mini-gene assays, it is difficult to imagine any single exon being regulated under normal circumstances by so many proteins (Wee et al. 2014).

Whilst each result looking into the effects and binding to regulatory sequences reveals more information regarding protein preferences and the dynamic balance of protein binding to exons, they also serve to highlight just how little we know. Furthermore due to the lack of conservation of splice sites in higher organisms, with more complexity and increased splicing variation, the presence and importance of regulatory sequences is not to be underestimated.

1.2.2. Exonic splicing enhancers.

Splicing regulatory sequences can be split into four classes depending on their effect, (enhancers or silencers) and their location (exonic or intronic); figure 5A. Exonic splicing enhancers (ESEs) are often found near constitutive or alternative sites and in constitutive or alternative exons (Humphrey et al. 1995; Ramchatesingh et al. 1995). Due to their regularity ESEs are one of the most common and in turn most well-studied genetic elements.

They seem to function by recruiting activator proteins to the RNA that in turn promote the binding of key splicing factors, U1, U2AF, U2 etc (Lavigueur et al. 1993b; Staknis & Reed 1994). This is often in competition with silencers, which allows the use of the exons or splice sites which they control to be fine-tuned depending on the level of expression of the regulatory proteins (Anczuków et al. 2015). There are two major determinants of the strength of an enhancer's strength; these are the sequence of the enhancer itself and the number of repeats of the enhancer (Hertel & Maniatis 1998).

The sequences of ESEs have been widely studied using a range of techniques. Initially ESEs were discovered as sequences that were essential for certain exons to be included, which were subsequently established as being rich in purine nucleotides essential for binding enhancer proteins (Dirksen et al. 1994; Sun et al. 1993). Techniques have since advanced and have been used to establish the consensus binding site for multiple proteins. Developments include SELEX (Tuerk & Gold 1990), functional SELEX (Liu et al. 1998), mutational studies (Wang et al. 2005), CLIP (Ule et al. 2005) and RNA-seq (Anczuków et al. 2015). Whilst these methods have served to identify sequences that bind certain proteins, these sequences often don't map well to known enhancers

identified in vivo (Pandit et al. 2013a); which may be due to a number of techniques only identifying sequences if they influence alternative splicing.

In contrast to the effect of the RNAs sequence on the binding of proteins, the effect of repeats of an enhancer has not. This is despite a number of pieces of evidence that indicate that ESEs are repeated in exons that are constitutively included (Lim & Sharp 1998; Hedjran et al. 1997); in fact, there is a negative correlation between the number of repeats of an enhancer and the combined strength of the splice sites of an exon (Anczuków et al. 2015). There are two main reasons that this factor has been less studied, the first is the paradigm established by a key paper in 1998 (Hertel & Maniatis 1998) and the second is the lack of suitable techniques.

In 1998 Hertel and Maniatis indirectly tested the binding of SR proteins to the RNA by following the rates of splicing of a minigenes based on the *Drosophila* double sex (*dsx*) gene. This gene had previously been shown to be dependent on repeated elements (*dsx*RE) found three hundred nucleotides downstream of the target 3' splice site (Hertel et al. 1996). The proportion of spliced mRNA in vitro after a certain period of time was shown to increase with increasing amounts of the *Drosophila* proteins Tra and Tra2 protein until it reached a maximum, but the maximum level was dependent on the number of repeat elements. It was inferred that as protein concentration increased, the occupancy of the *dsx*REs increased until they were fully occupied. At this point the addition of more protein would make no difference. The concentration of Tra and Tra2 needed to reach half-maximal splicing did not depend on the number of *dsx*REs, so it was concluded that each enhancer element acted individually and there was no cooperation. Due to these combined observations it was concluded that the limiting

step when all the enhancers were occupied was the subsequent interaction with the target factor. The linear increase in splicing witnessed with each additional repeat element, led to the conclusion that all of the elements were targeting the same factor. This led to a model whereby binding of the enhancer proteins was not limiting but the chance of an interaction with the target was. This model is shown in figure 6A.

The *Drosophila* model, however, does not fit with a number of observations of enhancer proteins in mammals that have been made subsequently. Primarily it has regularly been found that one of the archetypal mammalian enhancer proteins, SRSF1, does not bind RNA strongly (Suhyung Cho et al. 2011; S. Cho et al. 2011; Anczuków et al. 2015). In fact a number of affinity purification experiments with isolated SRSF1 have found SR proteins be a minor component of the protein mixture despite these enhancers requiring SR proteins for their activity (Lindsay D. Smith et al. 2014). Furthermore Tra and Tra2 are not human proteins and whilst in *Drosophila* they have been shown to bind in a stable complex, this has not been shown for human enhancer proteins. Finally, whilst the experiments in the 1998 paper were an important achievement, the massive expansion of enhancer proteins and sequences they can bind that have been discovered since highlights the possibility that different enhancers may function in different ways (Hertel & Maniatis 1998). Other possibilities include SR proteins binding stably but contacting multiple targets (each of which has an approximately equal effect on the rate), SR protein binding being cooperative or the probability of binding by an SR protein being very low and each site contributes independently; represented in figure 6B-D respectively.

A lack of suitable techniques to look at the occupancy of ESEs has also meant that a progression beyond the 1998 paradigm has been difficult. Without an ability to look at ESE occupancy, the other models that could explain the effect of enhancer repeats cannot be conclusively tested and therefore refuted.

1.2.3. The mechanism of ESEs

Once an enhancer protein has bound to an ESE, it still needs to make contact with its target factor to exert its effect. This step following ESE binding, whereby an enhancer interacts with its target, is somewhat less studied, again due to a lack of suitable techniques. In the standard model, after binding to an ESE the enhancer interacts directly either with U2AF 35, which binds the 3'SS, or U2 snRNPs, which base pair to branch points (Lavigueur et al. 1993b; Zhu et al. 2001). This is made more complicated by the fact that ESEs can exert their affects at distances up to and including 300 bases away (Hedley & Maniatis 1991). The first model that could explain the mechanism for this is one that involves the intervening RNA, between the ESE and its target, looping out and a collision occurring in this manner as shown in figure 6E. Figure 6F shows the second possible model which involves the use of the intervening RNA to send the signal to the target, either through the propagation of proteins along the RNA or an alternative scanning system. Whilst this debate may seem like an essential one, it has remained relatively understudied; with only three papers looking at it and all three of those having their drawbacks.

Two lines of evidence support the first model, figure 6E, whereby the intervening RNA, between the binding site and target, is flexible and allows collisions to occur. One is that the efficacy of MS2-tethered RS domains depended on the distance from their binding site to the target 3' SS (Graveley et al. 1998a). However, the data is also consistent with a model based on the propagation of proteins along the RNA as in figure 6F (Lewis, Andrew J Perrett, et al. 2012). Similar results were found using MS2-Tra2 but the use of its individual RS domains did not stimulate splicing as expected, suggesting that intact

proteins behave differently to isolated domains (Sciabica & Hertel 2006). The other evidence in favour of 3D diffusion is that MS2-tethered RS domains could be cross-linked to the 3'SS whilst being bound elsewhere (Shen et al. 2004) but here too it is hard to exclude a role for protein interactions if the intact SR protein were used. Our lab has attempted to differentiate between these two models at a 5' ESE and showed that the effect of the ESE is lost if the ESE were connected by non RNA linkers, which supports figure 6F; an RNA dependent protein complex model (Lewis, Andrew J Perrett, et al. 2012). However the linkers were introduced using Click coupling, which introduces a triazole group into the RNA (Kolb et al. 2001), meaning it is difficult to establish whether the effect seen was down to the coupling method used or the blocking of the ESE. However, when a triazole group was introduced into an oligonucleotide used for shifting splicing, that contained an ESE, there was little to no effect so one would expect there to be little effect in this case (Perrett et al. 2013).

Therefore the mechanism by which an ESE exerts its effect is still not clearly known. This key mechanism lies at the heart of ESE function and the lack of a clear experimental proof in this respect hinders our advancement in understanding the basics of RNA splicing.

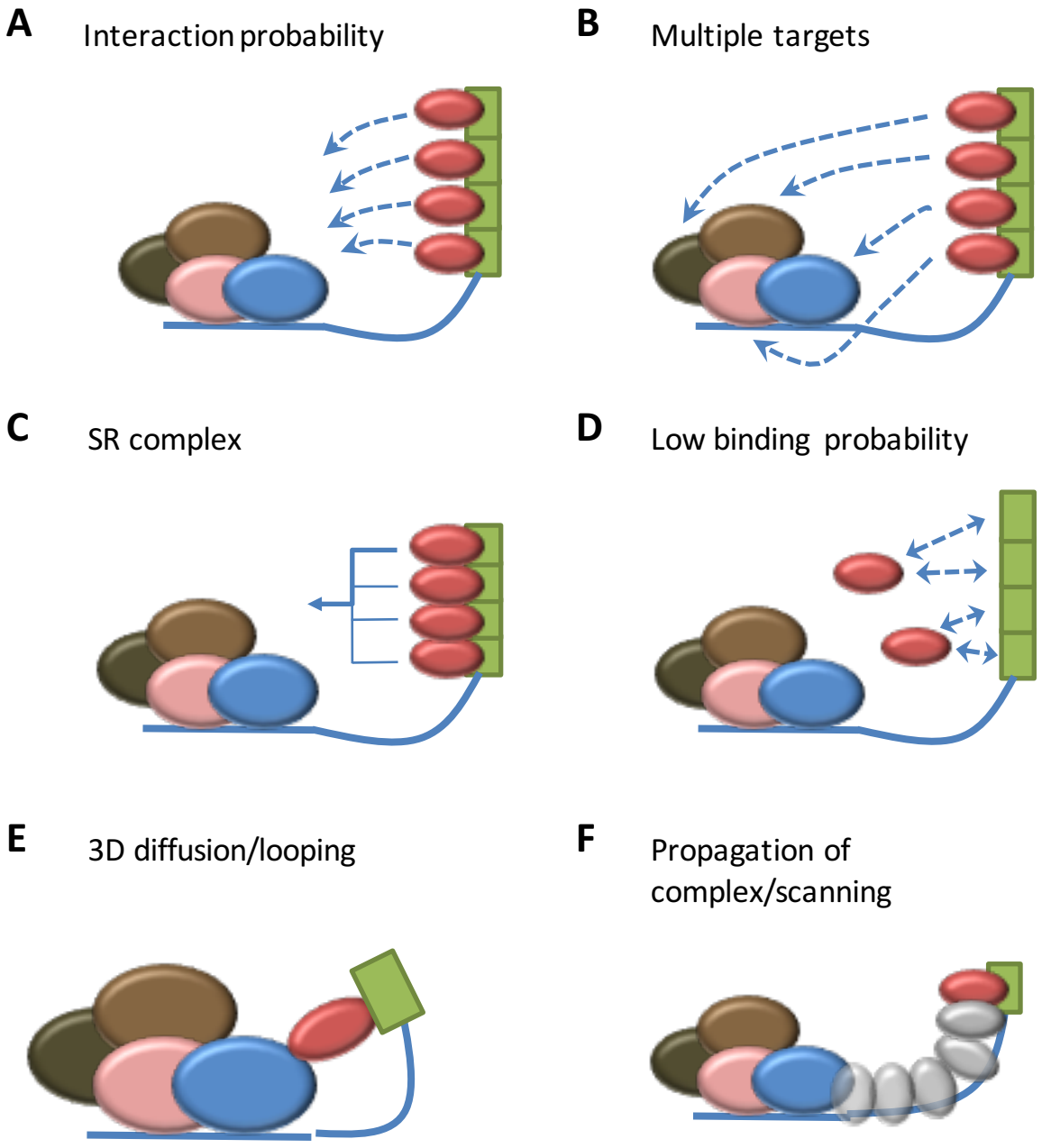


Figure 6. Possible SR protein binding patterns (A-D) and activation mechanisms (E & F) by which ESEs stimulate 3' splice site usage. (A) SR proteins bind stably, with multiple occupancy of repeated ESEs, and the probability of subsequent interactions by mechanisms E or F is the limiting factor. Solid blue line, pre-mRNA; red oval, SRSF1; other ovals, components at the 3'SS (U2AF, U2 snRNP, etc.); dashed lines, low probability interactions. (B) SR proteins bind stably, with multiple occupancy of tandemly-repeated ESEs, and are able to interact with multiple targets, each of which has an approximately equal effect on the rate. (C) SR protein binding is cooperative. (D) The probability of binding by an SR protein is very low and each site contributes independently. (E) Activation involves direct contact by three-dimensional diffusion between an ESE-bound SR protein and a 3'SS component; the intervening RNA is looped out. (F) Activation involves processes that maintain contact with the RNA between the ESE and the 3'SS, such as propagation of SR protein complexes or scanning along the RNA (for example, in conjunction with a helicase).

1.2.4. SR proteins.

ESEs function by recruiting enhancer proteins to the RNA and it is these enhancer proteins that enhance the binding of crucial splicing factors such as U1, U2AF, U2 etc. The main family of enhancer proteins in RNA splicing are the SR proteins. These are distinguished by the presence of a C-terminal RS domain rich in arginine-serine dipeptides in addition to one or two RNA recognition motifs (RRM) (Long & Caceres 2009). The standard model for 3'SS-activation by SR proteins is that the RRMs bind the ESE (van der Houven van Oordt et al. 2000) before the RS domains stimulate increased recruitment of U2AF or U2 snRNPs to the 3'SS. The SR proteins have roles in a wide variety of processes, including transcription (Mo et al. n.d.), nuclear export, translation and nonsense-mediated decay (Long & Caceres 2009; Maslon et al. 2014), but it is for their role in splicing that they are best known.

The first SR protein discovered was SRSF1 and this has since been studied in depth (Longman et al. 2000; Sun et al. 1993). SRSF1 in particular activates or represses the inclusion of hundreds of exons (Pandit et al. 2013b; Anczuków et al. 2015), and this activity is thought to be the primary reason why it is both essential (Wang et al. 1996; Longman et al. 2000; S. Lin et al. 2005) and an oncoprotein (Das & Krainer 2014). Accordingly, SRSF1 has provided the paradigm for understanding the effects of exonic splicing enhancers (ESEs) (X. Roca et al. 2013; Das & Krainer 2014; Bates et al. 2017). However, it has a series of additional properties that the other SR proteins do not have that suggest that it may have an important role in the core reactions of splicing. The major difficulty in understanding these properties is that they do not all fit with a model whereby SRSF1 is recruited to pre-mRNAs solely by ESEs.

The generally accepted model for activation by SRSF1 is that it binds ESEs (ESEs) (Graveley & Maniatis 1998; Sanford et al. 2009; Cléry et al. 2013; Pandit et al. 2013b; Anczuków et al. 2015), then interacts by 3D-diffusion via its RS domain or an RRM domain with the 70K protein of the U1 snRNP and with U2AF35 (Wu & Maniatis 1993; Anon 1994; Kohtz et al. 1994; Cao & Garcia-Blanco 1998; Cho et al. 2011), i.e., components found at the 5' and 3' splice sites whose binding is limiting for splicing activity. As a result it stabilizes U1 snRNPs at potential 5'SSs (Eperon et al. 1993; Kohtz et al. 1994; Staknis & Reed 1994; S. Cho et al. 2011; Jamison et al. 1995) and U2AF35 or U2 snRNP at 3'SSs (Lavigueur et al. 1993b; Staknis & Reed 1994; Smith et al. 2014; Tarn & Steitz 1995; Graveley et al. 2001; Martins de Araújo et al. 2009). The two properties that led to the isolation of SRSF1 originally fit with this model; its ability to modulate 5'SS selection and its ability to restore splicing activity to S100 cytoplasmic extracts (Ge & Manley 1990; Krainer et al. 1990; Krainer et al. 1991), since the latter could be accounted for by the ability to enhance the binding of scarce components and depended on exon sequences (Chandler et al. 1997; Mayeda et al. 1999).

Despite the considerable weight of evidence behind the above described ESE-based recruitment model, SRSF1 has a number of properties which do not fit with this model. First, SRSF1 was shown to stimulate splicing in S100 extracts of pre-mRNA substrates that lacked a 3' exon and contained only 1-3 nt of the 5' exon (Hertel & Maniatis 1999); this suggests that SRSF1 plays a crucial role in splicing but in an exonic sequence independent manner. Second, whereas the activation of splicing by SR proteins requires phosphorylation of their RS domains (Cáceres & Krainer 1993; Zuo & Manley 1993; Mermoud et al. 1994; Roscigno & Garcia-Blanco 1995; Cao & Garcia-Blanco 1998; Xiao

& Manley 1997; Xiao & Manley 1998; Graveley & Maniatis 1998; Zhu & Krainer 2000; Cartegni & Krainer 2003), the catalytic reactions of splicing require de-phosphorylation of the RS domains of SR proteins, and specifically SRSF1 (Mermoud et al. 1994; Mermoud et al. 1992; Xiao & Manley 1998); this suggests that the RS domain must undertake a secondary role later in splicing after activation. This fits with evidence that the RS domain of SRSF1 contacts the pre-mRNA in mature spliceosomal complexes B and C (Shen et al. 2004). Attempts to develop models whereby SRSF1 plays different roles in constitutive splicing from those in alternative splicing by identifying the domains responsible for each of the properties discussed above have foundered, possibly because the same domains are required for both activities (Cáceres & Krainer 1993; Zuo & Manley 1993; Zhu & Krainer 2000; S. Lin et al. 2005; Suhyung Cho et al. 2011; Wang & Manley 1995; Eperon et al. 2000).

	ESE Dependent	ESE independent
RS-dependent	• 3'SS activation. • Exon inclusion.	• Exon independent splicing. • Tri-snRNP recruitment.
RS-independent	• Counteract repressors. • Constitutive splicing (in the presence of a strong 3' SS).	• 5' Splice site selection.

Table 1. Table summarising the findings of Zhu and Krainer in 2000 regarding the functions of SRSF1 and the domains responsible for them.

Establishing whether SRSF1 really is part of the constitutive reactions of splicing is made particularly difficult because the SR proteins are found at variable and non-stoichiometric levels in spliceosome preparations (Schmidt et al. 2014). This is rather as might be expected, given that SRSF1 can bind to a wide range of pre-mRNA sequences and also might interact purely on the basis of electrostatic interactions via its RS domain.

Binding by multiple molecules that are involved in different interactions precludes the use of conventional methods for resolving different events.

Another well studied member of the SR protein family is Tra2β. Tra2β is considered an SR-like protein as opposed to a core member of the SR protein family for two key reasons. First it contains two RS domains, one at its N terminus and one at is C terminus, whereas the core SR proteins such as SRSF1 only contain a single RS domains at the C. Second, it cannot restore splicing to S-100 extracts on its own, unlike the other core SR proteins (Tacke et al. 1998). Tra2β was originally identified as one of two human homologues of the *Drosophila* Tra2 protein, the other being Tra2α (Dauwalder et al. 1996). It is known to have important roles in normal mammalian development and is essential for mouse embryonic and brain development (Grellscheid et al. 2011; Mende et al. 2010). The paralog Tra2α was created by a gene duplication of the Tra2 gene early in the vertebrate lineage and is found in nearly all vertebrates. It is an interesting enhancer protein because in *Drosophila* it forms stable complexes on purine rich ESEs along with other SR proteins. However Tra, unlike Tra2, is absent in humans despite being essential in *Drosophila*.

As hinted at earlier the *Drosophila* variant, Tra2, has been found to play a number of key roles in *Drosophila* splicing. One of these key roles is in sex determination; where its expression allows the activation of enhancers that stimulate the alternative splicing of key genes (McKeown et al. 1987; McKeown et al. 1988; Hoshijima et al. 1991; Boggs et al. 1987; Tian & Maniatis 1993; Baker 1989). Due to the effect of its orthologue on development, Tra2β has been extensively studied in human developmental states such as spermatogenesis (Venables et al. 2000; Venables et al. 2005). However its expression

is not limited to early developing tissues and a number of key genes have been found to have strong binding sites for the protein.

The major RNA binding site for Tra2β is an AGAA-rich sequence (Tacke et al. 1998; Cléry et al. 2011; Tsuda et al. 2011), although NGAA is also sufficient for binding but with lower efficiency (Cléry et al. 2011). A second minor form of RNA binding can also occur with Tra2β in which the RRM interacts with single-stranded CAA-rich motifs that are found within a stem loop (Tsuda et al. 2011).

This has led to a number of investigations Tra2β's role in developed cells and its effect on constitutive and alternative splicing. The well-studied SMN exon 7 is known to have a strong Tra2β (and Tra2α) binding site at its centre (Hofmann & Wirth 2002; Hofmann et al. 2000; Young et al. 2002). A number of studies have since shown that this site is crucial for the recruitment of the U2 snRNP to the 3' splice site of the upstream intron (Martins de Araujo et al. 2009). However its importance has been contested. This is in part given its lack of effect on splicing following the SMN 1>2 C>T mutation, which does not occur in the Tra2β site. Furthermore, studies looking at the effect of splice shifting oligos (SSO) and Tailed oligonucleotide enhancers of splicing (TOES) appear to show that the Tra2β site in SMN can be blocked as long as an alternative strong enhancer is present (Smith et al. 2014). Another gene that is affected by Tra2β is the CHEK1 gene. Here a number of enhancers within exon 3 influence its inclusion/exclusion. This thus allows the exon to be controlled depending on the expression level of Tra2β. However, similar to the SMN exon, it is not known whether Tra2β binds to all the enhancers, none of the enhancers or in a co-operative manner (Best et al. 2014). Whilst Tra2β appears to play

a number of crucial roles in controlling select splicing events, question marks over its exon specific binding patterns hinder further research.

1.2.5. Structure of SRSF1.

SRSF1 consists of two RRMs and an unstructured RS tail. RRM1 is a canonical RRM whilst RRM2 is referred to as a pseudo RRM. Psuedo RRMs are defined as such as they do not interact with RNA in the expected manner, via there beta sheets. The RS tail of SRSF1 contains many serine and arginine repeats and is unstructured and flexible.

The canonical RRM is an RNA-binding domain (RBD) first identified in the 1980s. Biochemical characterization of hnRNP C identified a consensus RBD of approximately 90 amino acids (Swanson et al. 1987; Dreyfuss et al. 1988). Two regions of high conservation were found and termed Ribonucleoprotein (RNP) domain 1 and RNP 2. RNP 1 had the sequence Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe/Tyr-Val/Ile/Leu-X-Phe/Tyr, where X can be any amino acid, whilst RNP 2 contained Ile/Val/Leu-Phe/Tyr-Ile/Val/Leu-X-Asn-Leu (Kenan et al. 1991). The RRM has a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology. The arrangement contains one four-stranded antiparallel β-sheet with the two α-helices packed against the β-sheet. The RNP 1 and 2 motifs are located in the central strands of the β-sheet. Many of the conserved residues, mentioned above, are in the hydrophobic core except four residues that contribute to RNA binding. These were RNP 1 Lys1, Tyr-Gly3 and Tyr-Val5 and RNP 2 Val2 (Nagai et al. 1990). However the central RNPs of an RRM are not the only determinant of a RRMs specificity or binding strength; N and C terminal extensions as well as additional RRMs or other domains can influence a RRMs properties (Maris et al. 2005).

A pseudo-RRM is able to bind to both RNA and protein via it's α helix 1. Unlike RRM 1, a canonical RRM, pseudo-RRMs have a conserved sequence consisting of Ser/Trp/Gln/Asp/Leu/Lys/Asp in α-helix 1 (Birney et al. 1993). Conversely they lack the

conserved aromatic residues in their β-sheet surface (Cléry et al. 2008). The structure revealed that the domain specifically binds the GGA motif, found in many ESEs, using the conserved residues located in the α-helix. More specifically the contact is with the GG dinucleotide. Interestingly, in contacting proteins, such as SRPK1, the exact same sites on the α-helix is occupied. This double involvement of α helix 1 implies that SRSF1 phosphorylation and RNA binding must be separate events (Cléry et al. 2013).

The pair of RRMs, one canonical and one pseudo, along with the RS domain likely contributes to both the RNA specificity of SRSF1 and its protein-protein contacts. Furthermore the glycine rich linker in between the RRMs has also been shown to contribute to selectivity (Cho et al. 2011). The presence of multiple domains opens up the protein to multiple possible contacts simultaneously; one with RNA and one with a factor that SRSF1 is recruiting for example (Cho et al. 2011; Smith et al. 2014; Martins de Araújo et al. 2009; Eperon et al. 2000). However with three domains it is possible SRSF1 could contact three separate factors, possibly two with RNA considering its two RRMs; something which has not been shown before.

1.2.6. DEXD/H proteins

There are around 40 different DEXD/H-box proteins found in the human proteome, all of which are known to play a role in RNA metabolism (Tanner & Linder 2001; Linder & Jankowsky 2011). There are three sub families of DEXD/H proteins which are the DEAD-box family of ATPases, which are typically not considered pure helicases as they destabilise both RNA-RNA and RNA-protein interactions (Jankowsky et al. 2001), the SKI family and the DEAH family. Seven of the DEXD/H helicases are known to be involved in splicing. The name is derived from the highly evolutionarily conserved (V/I)LDEADX(M/L)LXXGF amino acid sequence in the N-terminal ATP binding domain (NABD). Brr2 is often included in that list despite being a Ski2-like helicase.

DEAD-box protein 5 (DDX5 (p68)) is one of the most abundant members of the DEAD-box subfamily of RNA helicases (F. V Fuller-Pace & Ali 2008). DDX5 consists of two functional domains: an N-terminal ATP binding domains (NABD) and a C-terminal catalytic domain (CTCD) which provides the helicase activity (Bates et al. 2005; Meurs et al. 1992). Though generally considered a nuclear protein, DDX5 is known to be shuttled between the nucleus and the cytoplasm. DDX5 is highly multifunctional, influencing processes including cell cycle regulation (Choi & Lee 2012), ribosome biogenesis (Saporita et al. 2011), and transcription. However despite its numerous functions, its consensus target remains poorly defined. Furthermore whilst other functions have been heavily investigated, its role within splicing reactions remains unclear (Dardenne et al. 2014; Wongpalee et al. 2016).

Despite being poorly characterised in the context of splicing reactions, it has in fact been shown to be an essential factor. Depletion of endogenous p68 (DDX5) did not affect the

loading of U1 snRNPs on to the 5'ss, however it did arrest spliceosome assembly at the pre-spliceosome stage; suggesting that p68 may play a role in the progression of the early spliceosome (Liu 2002). Furthermore, studies looking at mutant versions of DDX5 suggested that the protein might play a role in the U4.U6/U5 addition but that this function does not require the ATPase and RNA unwinding activities of p68 (Lin et al. 2005).

There are a specific subset of 5' splice sites, such as the one found in Tau exon 10, which can form extensive secondary mRNA structures (Jiang et al. 2000). In cases where this secondary structure remains, the U1 snRNA cannot form a duplex with its complementary sequence which in turn often leads to skipping of the associated exon (Jankowsky et al. 2001). In the case of Tau Exon 10, there is a pyrimidine-rich sequence immediately downstream of the 5' splice site which is an ideal site for the RNA binding protein RBM4. This in turn can recruit DDX5 to the RNA which can destabilise the secondary structure and allow the U1 snRNP to associate (Kar et al. 2011). Similarly, DDX5 is also observed to increase inclusion of the intron D exon (IDX) in the cell-cycle control factor H-Ras. IDX forms a stem-loop structure with a downstream intronic silencer sequencer, which is stabilised by dimerised hnRNP H. Disruption of this stem loop by DDX5 displaces hnRNP H which dimerises to hold the IDX to the ISS (Guil, de La Iglesia, et al. 2003; Guil, Gattoni, et al. 2003).

The S. cerevisiae orthologue of DDX5, Dbp2, which is similarly poorly characterised but shows such high similarity to the human DDX5 that it is able to rescue Dbp2 knockout yeast cells (Barta & Iggo 1995). Dpb2 has been found to co-transcriptionally modulate the secondary structure of RNA to facilitate RNP assembly, prevent usage of cryptic

transcriptional start sites and interact with nuclear decay factor Rrp6 to repress aberrant transcription (Cloutier et al. 2012). Human orthologues may have developed alternative roles to yeast spliceosomal proteins in response to the more numerous, longer introns.

Hence, whilst a number of the roles of DDX5 outside of splicing have been elucidated, its role within splicing reactions and the spliceosome still remain unclear. A number of examples have demonstrated that it may play a role in de-stabilising secondary structures in order to allow the U1 snRNP to bind. However in many cases there is not a clear route as to how the DDx5 is recruited to the RNA. Alternatively, data also suggests that it plays a role in spliceosome progression. However in this case, there is no obvious way for it to be recruited to the RNA. Hence information as to how the DDX5 is recruited may reveal vital information about its function and what exons it may function on.

1.2.7. Splicing regulation and disease.

Control of the pattern of splicing for a particular gene involves the recognition and use of a vast number of RNA motifs, from splice sites to regulatory sequences, which all need to be recognised correctly in order for splicing to occur in the correct manner. Therefore single nucleotide polymorphisms (SNPs) in any of these key sequences can have drastic effects on the resulting pattern and subsequent protein that is translated.

When investigating mutations that cause diseases the effect of changes to the splicing pattern are often ignored completely despite estimates that 62% of these mutations affect splicing (López-Bigas et al. 2006). These mutations can have a range of effects depending on their location. Mutations in 5' or 3' splice sites can lead to exon skipping, use of cryptic splice sites or even intron retention (Ohno et al. 2017). Mutations in enhancers can lead to skipping, use of an alternative site or intron retention. Mutations in silencers can lead to exon inclusion or alternative splice site usage (Scotti & Swanson 2015). A good example of a disease-causing SNP that affects splicing is the SMN C>T mutation in exon 7; this results in the abolition of a weak SRSF1 binding site and the creation of a hnRNP A1 binding site which causes the exon to be skipped (Lorson et al. 1999; Kashima et al. 2007; Martins de Araujo et al. 2009). This in turn leads to the translation of a non-functional protein and a subsequent diseased state.

Equally, as mentioned, mutations in proteins involved in splicing can have similar effects on splicing patterns. These mutation can lead to enhanced/weakened protein-protein interactions or enhanced/weakened protein-RNA interactions. Due to the complicated assembly of the spliceosome, these relatively minor changes can have a major effect. Examples include the overexpression or post-translational modifications of SRSF1 in a

number of cancers (Gonçalves & Jordan 2015; Anczuków et al. 2015; Das et al. 2012; Anczuków et al. 2012), the mutation of Prp8 in retinitis pigmentosa (Boon et al. 2007; Pena et al. 2007) and the mutation of SF3B1 in a number of cancers (Maguire et al. 2015; Darman et al. 2015; Wan & Wu 2013).

To combat these diseases a number of different methods have been proposed, with the ability to shift splicing from an altered disease state back to a healthy state having great promise (Havens et al. 2013). Two current methods involve the use of complementary oligonucleotides or small molecules to shift splicing. Antisense oligonucleotides that can bind to certain sites in the RNA that affect splicing, such as silencers or enhancers, function by blocking or stimulating the binding of proteins to drive splicing in a new direction. In SMN this has been achieved by blocking a silencer that is found immediately downstream of the 5' splice site of exon 7; this received FDA approval in 2016 under the name SPINRAZA (Staropoli et al. 2015; Madipalli 2017; Rigo et al. 2014). The splicing of SMN has also been altered via the use of small molecules that stimulate the binding of key factors such as U1 snRNPs (Zhao et al. 2016; Palacino et al. 2015; Woll et al. 2016). These methodologies are still in their relative infancy and there is much promise for the future.

1.3. Single Molecule studies.

1.3.1. Single molecules studies of RNA splicing

The splicing of RNA has predominantly been studied using so called ensemble methods which involve looking at an entire population of molecules and witnessing the average events and making assumptions from these. Whilst these methods have been largely successful and have revealed a number of crucial bits of basic information, their inability to look at events on an individual basis limits them. This is particularly the case when looking at the binding of splicing factors, proteins that bind more or less indiscriminately and have functions in specific complexes or at specific sites. Single molecule studies however allow us to look at individual binding events and interpret the exact number of proteins that binds to a single strand of RNA. This thus allows us to glimpse the changes in protein binding that lead to the behaviour observed ensemble experiments.

Single molecule studies have been successfully used previously in a variety of studies associated with RNA splicing. In yeast systems of splicing the Moore lab has used single molecule Fluorescence Resonance Energy Transfer (FRET) methods to analyse conformational changes that occur during tri-snRNP/Prp19 complex recruitment (Hoskins et al. 2016) and before the 5’ splice site and branch point come into close proximity (Crawford et al. 2013). Furthermore they have shown, using co-localisation, that the spliceosome can assemble in a U1 or U2 first manner (Shcherbakova et al. 2013). Similar methods have been used by other labs to show that U4/U6 duplex unwinds in a multi-step process (Rodgers et al. 2016), how the U4/U6 snRNA assemble (Hardin et al. 2015), how helicases interact with their substrates (König et al. 2013) and how Group 2 introns fold (Steiner et al. 2008).

In mammalian systems the Eperon lab has used co-localisation and stoichiometric analysis, as is used in this study, to analyse the binding of key components of the early spliceosome to both the 3’ splice site (Chen et al. 2016) and 5’ splice site (Hodson et al. 2012). In collaboration with the Smith lab they have also analysed the number of molecules of the splicing repressor PTB that bind (Cherny et al. 2010; Gooding et al. 2013).

1.3.2. Excitation and emission.

There are two types of photoluminescence, fluorescence; which involves molecules being excited for relatively short periods of times, and phosphorescence, which involves molecules being excited for longer periods of times. Fluorescent molecules, or fluorophores, tend to emit their energy a lot faster; this makes them much more useful for analysing biological molecules rapidly.

Fluorophores absorb a photon at an energy required to raise an electron from a ground state (S_0) to a higher energy level or shell (S_1). Internal relaxation then occurs causing the electron to switch to an excited state of a slightly lower energy. The molecule then releases this energy and settles back to its ground state; this is shown in Jablonskis diagram (Figure 7A). However due to this internal relaxation the photon that is released is of a lower energy and thus a lower wavelength than from what was absorbed. The resulting difference in wavelengths is known as a stokes shift (Figure 7B) (Jablonski 1933).

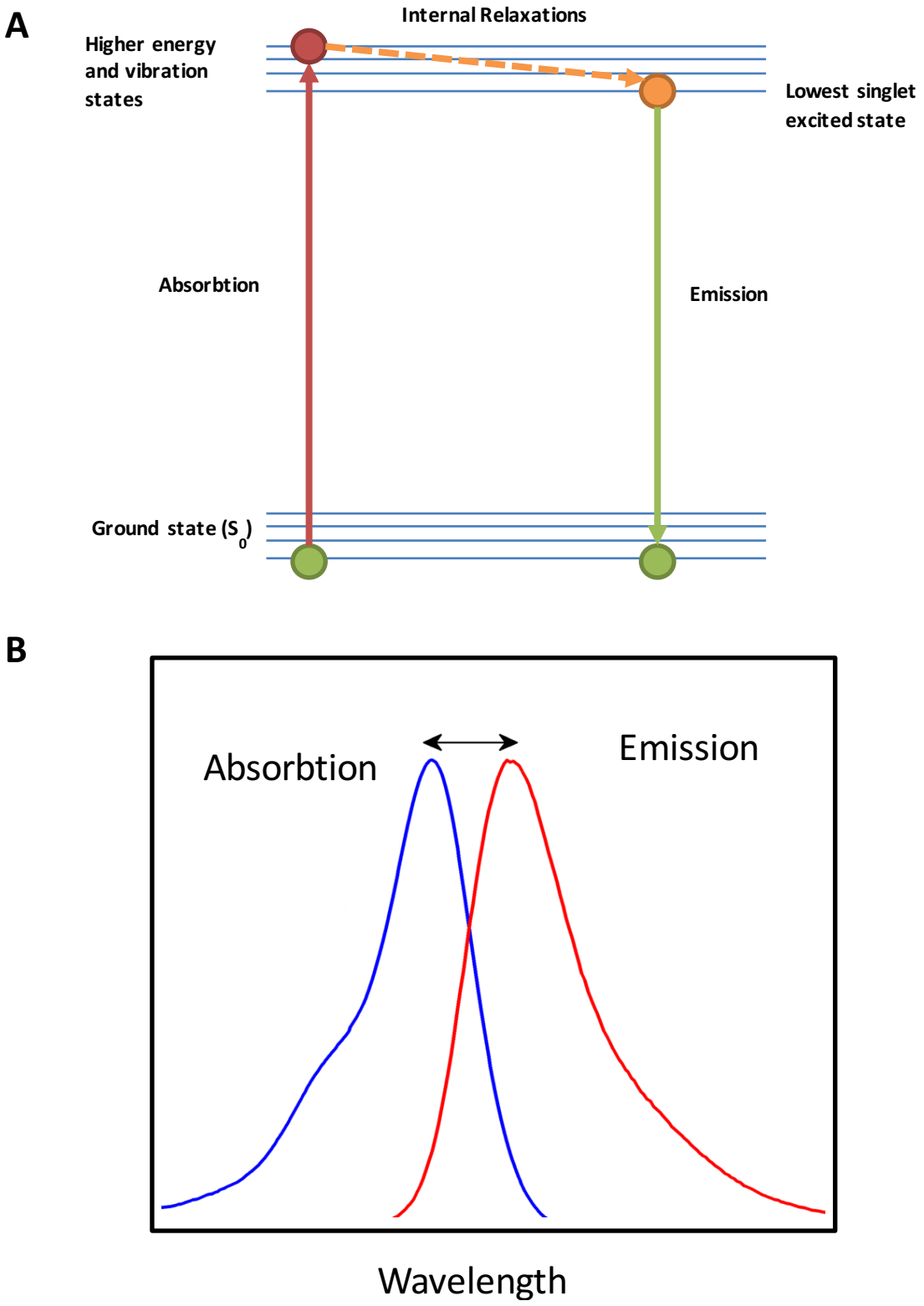


Figure 7. Mechanism of fluorescence. A) Simplified Jablonskis diagram showing an electron being excited to a higher energy state by absorbing energy. The electron then relaxes to a lower high energy state before releasing the energy and settling back to its ground state. B) Simplified stokes shift diagram showing the difference in energy from the absorbed to the emitted.

1.3.3. Co-localisation and Photo-bleaching.

The work described here is concerned with two key pieces of information that can be garnered from looking at single molecules; the co-localisation of two different molecules and the number of steps that it takes for a certain fluorophore to photo-bleach.

Co-localisation refers to the visualisation of a spatial overlap between signals from two fluorophores that emit at different wavelengths. One major issue with co-localisation based studies is Chromatic aberration. This results from a failure of a lens to focus different wavelengths to the same point. It occurs due to lenses having different refractive indices for different wavelengths of light. The result of these aberrations is that two molecules that may be bound to one another, and therefore would be expected to occupy the same pixel as one another in the resulting image, do not. This can in turn lead to confusion as to which molecules are actually co-localised and which are not. In order to overcome this calibration experiments must be done. These involve using two molecules that are known to be bound to each other in order to calculate the displacement (explained in detail in Materials and Methods). This can then be applied to spots that are not known to bind in order to determine whether they bind.

The second piece of information gained in this work from single molecule experiments is the number of steps in which co-localised fluorophores photo-bleach. Photo-bleaching is a photochemical alteration of a fluorophore (or dye) molecule that results in the molecule losing its fluorescence. This alteration is often irreversible due to the cleaving of covalent bonds.

In the single molecule experiments described here, the number of photo-bleaching steps in which a single complex bleaches correspond to the number of fluorophores that are present. For example, a single bleaching step is evidence for at least one fluorophore present whilst two steps indicates two and so forth (explained in detail in Materials and Methods). A flow-through of identification of spots through to the assignment of steps and how that correlates to the number of proteins is shown in figure 8.

The loss of fluorescence caused by photo-bleaching can be limited by reducing the intensity of the excitation laser or light source, increasing the number of fluorophore-containing molecules or employing better fluorophores that have longer life times.

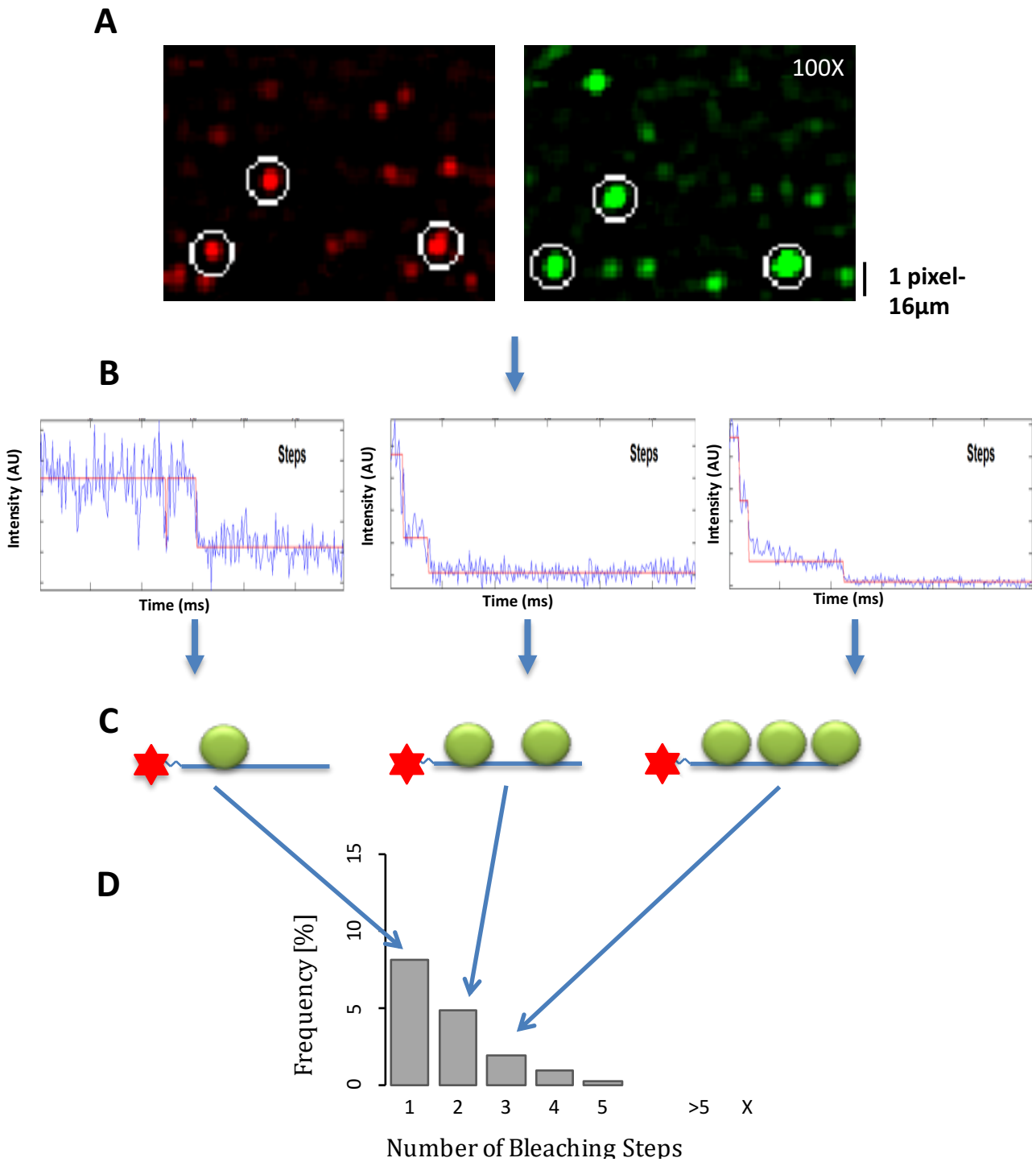


Figure 8. Schematic showing the assignment of co-localised spots followed by the assignment of steps according to their bleaching pattern and how this relates to the number of molecules that are bound to the RNA and how we display this. (A) Screen grabs showing the false coloured images from the illumination with 640nm light (Cy5) and 488nm light (GFP). (B) Screen grabs showing three example step profiles; the y axis is an arbitrary number for intensity and the x axis is time. (C) Cartoons showing how the relevant step profiles relate to the number of molecules that are bound to the labelled RNA. (D) Example histogram showing the graphical manner in which the analysed data is displayed; the y axis indicates the frequency of cases that have that number of bleaching steps as a percentage of the total number of RNA spots and the x axis shows the number of bleaching steps that correlates to that class.

1.3.4. TIRFM.

Total internal reflection (TIR) fluorescence (TIRF) microscopy (TIRFM) uses the generation of an evanescent wave of light to illuminate a small region of a sample located immediately above a surface. When a beam of light, or laser, passes between two separate mediums it is refracted according to Snell's law. If the beam is passing from a medium with a high refractive index, such as glass, into a medium with a lower index, such as water, the angle which the light is refracted is larger than the incidence angle; if this angle is large enough then the entire beam is reflected back. The smallest angle at which TIR can occur is known as the critical angle.

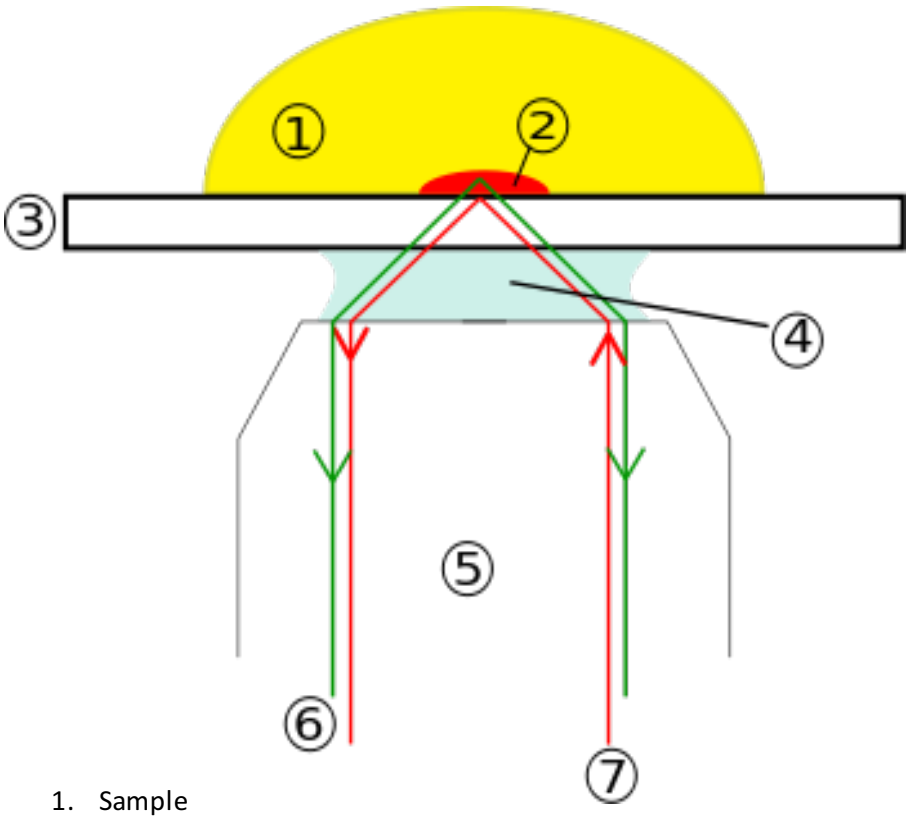
At angles higher than the critical angle the incident light is entirely reflected. An important feature for single molecule studies of this total reflection is the generation of an evanescent field on the side opposite from where the beam is reflected. This is the side with a lower refractive medium e.g. water. The generated field decreases in intensity with distance from the surface at an exponential rate. This rate depends on a number of things such as the wavelength of light, the refractive indices of both mediums etc. but typically the field ranges between 50 to 100 nm (Axelrod et al. 1984). The advantage of this method of illumination is that it allows one to visualise samples attached to the surface with minimum background from the rest of the sample. There are two common methods for utilising TIR for fluorescence microscopy; they are the use of a prism (Axelrod 2001; Thompson & Steele 2007) or the use of a microscope objective directly. Alternative methods have been suggested, including the use of both methods (Burghardt & Ajtai 2010) and the use of a parabolic mirror (Ruckstuhl & Seeger 2003), but these are currently not common practise.

Whilst the prism technique is easier to set up, it carries the disadvantage that the sample is located between the objective and the prism which means more elaborate methods of sample placement and subsequent movement often have to be used. An advantage of this method however is that the field generated with prism based-techniques is usually wider, typically up to 100µm at the beam waist, which produces a flatter illumination field. Prism based-techniques are also helped by the fact that the emission and excitation beam pathways are completely separate meaning that there is less chance of stray reflections.

The objective approach however carries the key advantage that the sample is placed on top of the objective and is therefore far easier to place and manipulate. This approach requires that light beam be introduced through filters or the use of micro mirrors within the microscope and the proximity of the excitation and emission pathways means there is potential for stray reflections. The closer proximity of the two beams also means that they are very close to the critical angle. This results in the width of the effervescent field being narrower and more intense. This can result in images appearing more intense but with more background. Overall the signal to background noise ratio is potentially better with prism techniques but the ease and speed of use as well as the intensity of images produced by the objective based techniques is far superior. For these reasons, an objective-based approach was chosen whilst designing the microscope used in this study. Figure 9 shows the setup for objective TIRFM.

TIRFM is especially well suited to investigate membranes (Thompson et al. 1993), cell-signalling (Sako et al. 2000), the movement of single molecules within the membrane (Schmidt et al. 1996) or the single steps of myosin moving along actin filaments (Yildiz

et al. 2003). However, here we will use this microscopy to observe individual molecules of RNA and protein, figure 8A, to observe how they interact in different complexes and with what stoichiometry, figure 8C.



1. Sample
2. Effervescent field
3. Slide
4. Immersion oil
5. Objective
6. Emission beam
7. Excitation beam

Figure 9. Setup of an objective based total internal reflection microscope and the direction of the excitation and emission laser beams. 1. The sample on top of the slide, 2. The effervescent field generated by the reflected laser beam, 3. The slide, 4. Immersion oil, 5. The objective leading to the camera, 6. Emission beam, 7. Excitation beam pathway.

1.4. Summary.

In all human cells there are a vast range of cellular machines that perform essential functions for the cell. These range from simplistic systems to some of the most complicated; of which the spliceosome, with all its hundreds of proteins that interact with each other, is definitely one. Due to this complicated nature and despite nearly 30 years of research, many of its mechanisms are wrapped in mystery. The selection of the splice sites and the functionality of enhancer proteins is one of these. SRSF1 is heavily involved in this part of splicing and consequently has been researched extensively. However despite this, many of the various findings conflict with each other. Its roles, the roles of its domains, the nature of its recruitment, the number of molecules involved and mechanism of action are all still unclear and further insight is required. Experiments that are done in functional conditions necessarily use nuclear extracts, with all the countless possible interactions of the dauntingly numerous proteins. Especially problematic is that different states of each machinery can be present at any given moment, as they all might progress at slightly different speeds and/or include different paths. All these deviations are easily hidden in ensemble measurements. TIRFM and co-localisation provides a tool that can be used to overcome a number of these issues. This in turn can provide us with a number crucial insights into the stoichiometry of protein binding and thus its functions. Combining these single molecule techniques with standard and novel ensemble techniques will allow us to build a clearer picture of the recruitment and mechanism of action of SRSF1.

1.5. Aims.

The aim of this investigation is to explore the avenues of recruitment of SRSF1 to pre-mRNA and look at the subsequent mechanisms of action.

Single molecule TIRF microscopy will be used to analyse the numbers of SRSF1 molecules that bind to splicing constructs. Modifications of these splicing constructs by addition or removal of select sequences would then reveal the sequence motifs that are important for the recruitment of these SRSF1 molecules. Co-localisation with U1 snRNPs will reveal how these two vital splicing factors interact and influence each other. Furthermore analysis of binding with different ESE sequences and variable numbers of ESEs will reveal whether SRSF1 binds stably and multiple molecules can bind concurrently.

The use of chemical bridging along with hybrid RNAs that contain linkers of non-RNA will be used to elucidate the mechanism whereby an ESE that recruits SRSF1 can exert an effect at a distance. These non-RNA stretches should interrupt propagation or scanning mechanisms but allow looping mechanisms to occur. Splicing assays would reveal whether the ESE could still exert an effect and which mechanism is correct.

1.6. Objectives

In order to achieve the aims of this study we will use a succession of single molecule experiments followed by chemical bridging experiments. This combinatorial approach will allow us to analyse both recruitment and mechanisms in depth. These are approaches that have not been applied extensively to this field and will require some development but will also provide a lot of insights.

One of the crucial aims of this study is to develop the methods. The chemical bridging involves the ligation of hybrid RNA to long pre-transcribed RNA, which posits a number of issues that will need to be overcome in order to obtain efficient ligation which will provide enough material for subsequent analysis. The single molecule analysis of splicing factors has been fairly well established by prior members of the lab, but it still requires optimisation for each factor studied.

Once the methods have been established and validated then they can be used to look at key events within the early regulation of splicing. One of these key events is the binding of enhancer proteins to ESEs; this will be studied using the single molecule techniques as well as established ensemble methods for the archetypal activator protein SRSF1. This will serve as a model protein for the ESE-driven activation of mammalian splicing. The mechanism of activation following binding will be studied using the chemical bridging method. Together these approaches should produce new insights into the binding and activation mechanisms of ESEs.

There are also a number of additional properties of SRSF1 that suggest it may have an alternative method of recruitment in the absence of ESEs and a subsequent alternative

role. This is something that the use of single molecule methods will allow us to analyse. By using a variety of RNA constructs we can analyse what exactly drives SRSF1's recruitment to RNA.

Chapter 2. Materials and Methods.

2.1. Construct synthesis techniques.

2.1.1. Polymerase chain reaction

DNA constructs were amplified from plasmids or prior PCR products. The reactions were done in a a G-Storm thermal cycler (G-Storm), using Red Taq DNA polymerase (Sigma-Aldrich) or Q5 high fidelity polymerase (New England Biolabs) according to the manufacturer's instructions. Reactions were checked on a 1 %(w/v) agarose gel containing 0:01 %(w/v) EtBr for impurities and to ensure the product was the correct size. The reactions were then extracted with phenol-chloroform extracted and precipitated with ethanol.

2.1.2. DNA/RNA purification

Samples were made up to 100 µl using T.E.1 (10 mM Tris pH 7:5 and 100 µM EDTA) or H_2O. Equal amounts of phenol:chloroform:isoamyl alcohol (Sigma-Aldrich) were added, then the sample was vortexed and centrifuged before the supernatant was removed and kept. Sodium acetate was added to a concentration of 300 mM and three volumes of ethanol were added before the sample was centrifuged for 15 min at full speed. The supernatant was removed and 200 µl 70% ethanol were added before a final centrifugation of 15 min. The supernatant was once again removed and the pellet was dried under vacuum for 30mins. The pellet was then disolved in an appropriate volume of TE.1.

2.1.3. In vitro Transcription of RNA

A transcription reaction containing 40 mM Tris HCl (pH 7.5), 20 mM $MgCl_2$, 10 mM NaCl, 2 mM spermidine HCL, 10 mM DTT, 4 mM NTPs, 0:05% NP-40, 5% T7 polymerase (1:20 diluted (homemade)), 2 units/µl RNaseOUT (Invitrogen) and 10 ng/µl PCR template was set up at room temperature and incubated for 4 hours at 37 °C. 1 unit of DNase (Promega) was added and the reaction incubated for a further 30 minutes at 37 °C to degrade the precursor DNA. The transcript was then purified with a MicroSpin G-50 or G-100 column (GE Healthcare) depending on the size of the transcript. The sample was finally purified via phenol-chloroform extraction and ethanol precipitation, as described in 2.1.3.

2.1.4. In vitro Transcription of Radiolabelled RNA

A reaction mixture containing 5 mM DTT, 1 mM diguanosine triphosphate sodium [G(5')ppp(5')G] (GE Healthcare), 0.5 mM ATP, 0.5 mM CTP, 0.5 mM UTP, 0.05 mM GTP, 40 mM Tris HCl pH 7.5, 6 mM MgCl, 10 mM NaCl, 2 mM spermidine HCL, 0.33 µM [α-^{32}P] GTP (10 mCi/ml, 3000 Ci/mmol (Perkin Elmer), 2 units/µl RNase OUT (Invitrogen), 5% T7 polymerase (1:20, homemade) was set up at room temperature. It was then incubated for 2 hours at 37 °C. The transcripts were polyacrylamide gel purified, ethanol precipitated and dissolved in TE.1.

2.1.5. In vitro transcription incorporating a 5' GMPS cap

A 50 µl transcription reaction containing 40 mM Tris-HCL (pH 7.5), 2 mM spermidine HCL, 20 mM MgCl, 10 mM NaCl, 10 mM DTT, 4 mM ATP, 4 mM CTP, 4 mM UTP, 1 mM GTP, 10 ng/ul PCR product, 10 mM Guanosine-5'-O-mono phosphorothioate (5-GMPS) (Biolog), 2 units/ul RNase OUT (Invitrogen) and 6% T7 (1:20, homemade) was set up on

ice and then incubated at 37 °C for 4 hours. 1 unit of DNase (Promega) was added and incubated for a further 15 minutes. The reaction was subsequently phenol-chloroform extracted, ran through a G-50 column and then ethanol precipitated.

2.1.6. Maleimide labelling of GMPS capped RNA

The pellet from 2.1.5 was disolved in 17 µl H_2O, 2 µl 1M $NaPO_4$ (pH 7.2) and 1 µl 10 mM Cyanine-5 maleimide in DMSO (Luminiprobe) and incubated at room temperature for 4 hours. The Cy5-labelled RNA was run on a polyacrylamide gel and bands of the correct height were visualised via UV shadowing and then excised and eluted. The splicing efficiency of Cap labelled RNA is comparable to uncapped RNA (Chen et al. 2016).

2.1.7. Denaturing Polyacrylamide Gel electrophoresis

A gel mix consisting of 7 M Urea, 6-12% polyacrylamide (19:1 bisacrylamide) solution (National diagnostics) and 1xTBE (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA) was mixed and the urea dissolved. The gel was polymerised with addition of 0.1% Ammonium persulfate and 0.002% TEMED and allowed to set in a gel cast with an appropriate comb. Samples to be loaded were mixed with an equal volume of 9 parts formamide and 1 part 0.5M EDTA containing xylene xyanol and bromophenol blue dyes. The samples were loaded and the gel run at 4-32 watts for 0.5-3 hours depending on the size and percentage of the gel and the size of the sample being run. Gels were subsequently transferred to 3mm card, dried and visualised with a phosphorimager screen and typhoon scanner.

2.1.8. UV shadowing

RNA to be purified was mixed with an equal volume of formamide dyes and loaded on to a polyacrylamide gel. Once run the gel was removed from the plates and wrapped in

saran wrap. The wrapped gel was then placed on a fluorescent TLC plate and irradiated with UV light. Shadows that appeared where the RNA. These bands were excised and the RNA eluted.

2.2. Bacterial cloning techniques.

2.2.1. Cloning

For the subsequent expression of proteins in mammalian cells, cloning was done using Exonuclease III (ExoIII). Plasmids for expression in mammalian cells, pLeic 23 and 25, were cut with restriction enzymes. When doing PCR, primers used were designed with a 20 to 25 nucleotide sequence complementary to the plasmid. 250 ng of digested plasmid and the same molar amount of PCR fragment were mixed along with 1 µl of buffer (appropriate NEB generic buffer) and 1 U of ExoIII (NEB). Reactions were done at 14°C for 1 min to digest the 3' ends of both the plasmid and the PCR fragment. The fragments were then ethanol precipitation and the pellet dissolved in TE.1 buffer (1 mM Tris, 0.1 mM EDTA). The fragments were annealed by heating to 55 °C for 3 min and then cooling to room temperature.

2.2.2. Preparation of chemically competent E.coli cells

1 ml of commercially bought TOP 10 *E.coli* cells (Sigma) were grown overnight on LB Agar with no antibiotics. A single colony was picked and grown in 5 ml of standard LB broth (1% (w/v) tryptone, 0.5% yeast extract, and 1% NaCl) medium without any antibiotics at 37°C shaking for 3-6 hours. This was then used to inoculate 100ml of LB broth which was grown at 37°C with shaking for 12 hours or until the OD_{600} of the culture reached 0.6. Cells were harvested by centrifugation at 4000 rpm for 10 min followed by re-suspension of the cells in 50 ml of ice-chilled 0.1 M $CaCl_2$ and left on ice for 30 mins. The cells were centrifuged as before and the pellet was re-suspended in 4 ml of ice-cold 0.1 M $CaCl_2$ and 15% glycerol. Cells were aliquoted and in snap frozen in liquid nitrogen before storage at -80°C.

2.2.3. Transformation of E.coli cells.

50 ng plasmid DNA or annealed cloning fragments were added to 50 µl of thawed chemically-competent cells and incubated on ice for 5-10 minutes. Cells were then heat-shocked at 42 °C for 30 seconds. 200 µl of LB broth was added. The mixture was incubated at 37 °C, shaking, for 1 hour. The cells were then spread on a LB agar plate containing an appropriate antibiotic and incubated at 37 °C for 12-24 hours.

2.2.4. Extraction of plasmid DNA

Plasmid DNA was extracted using a Qiagen miniprep or Endofree maxiprep kit according to the kits' instructions.

2.3. Mammalian Cell culture methods.

2.3.1. Transfection of HeLa Cells

An appropriate number, ~5x10^6, of HeLa cells were seeded in a 15cm dish to give ~70% confluence after a further 24 hours of growth. Following this they were transfected with Fugene 6 (Roche), Optimem and an endo-free purified DNA plasmid. Cells were then cultured for a further 14 to 48 hours before being harvested, depending on transfection efficiency and cell death.

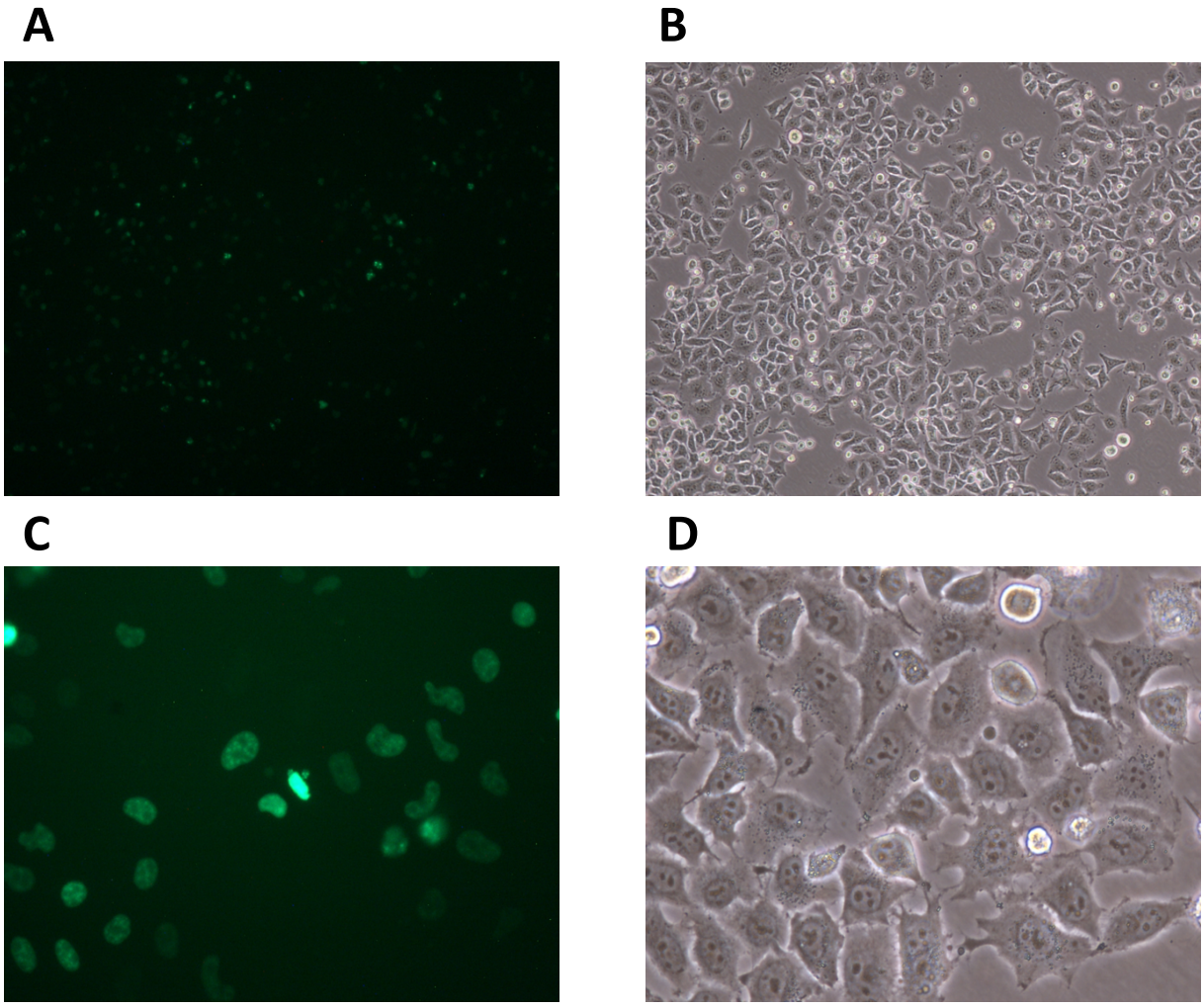


Figure 10. Images of HeLa cells following transfection with GFP tagged proteins. (A) 10X magnification with 460-500nm filter applied. (B) 10X magnification of HeLa cells. (C) 40X magnification of HeLa cells with 460-500nm filter applied. (D) 40X magnification of HeLa cells.

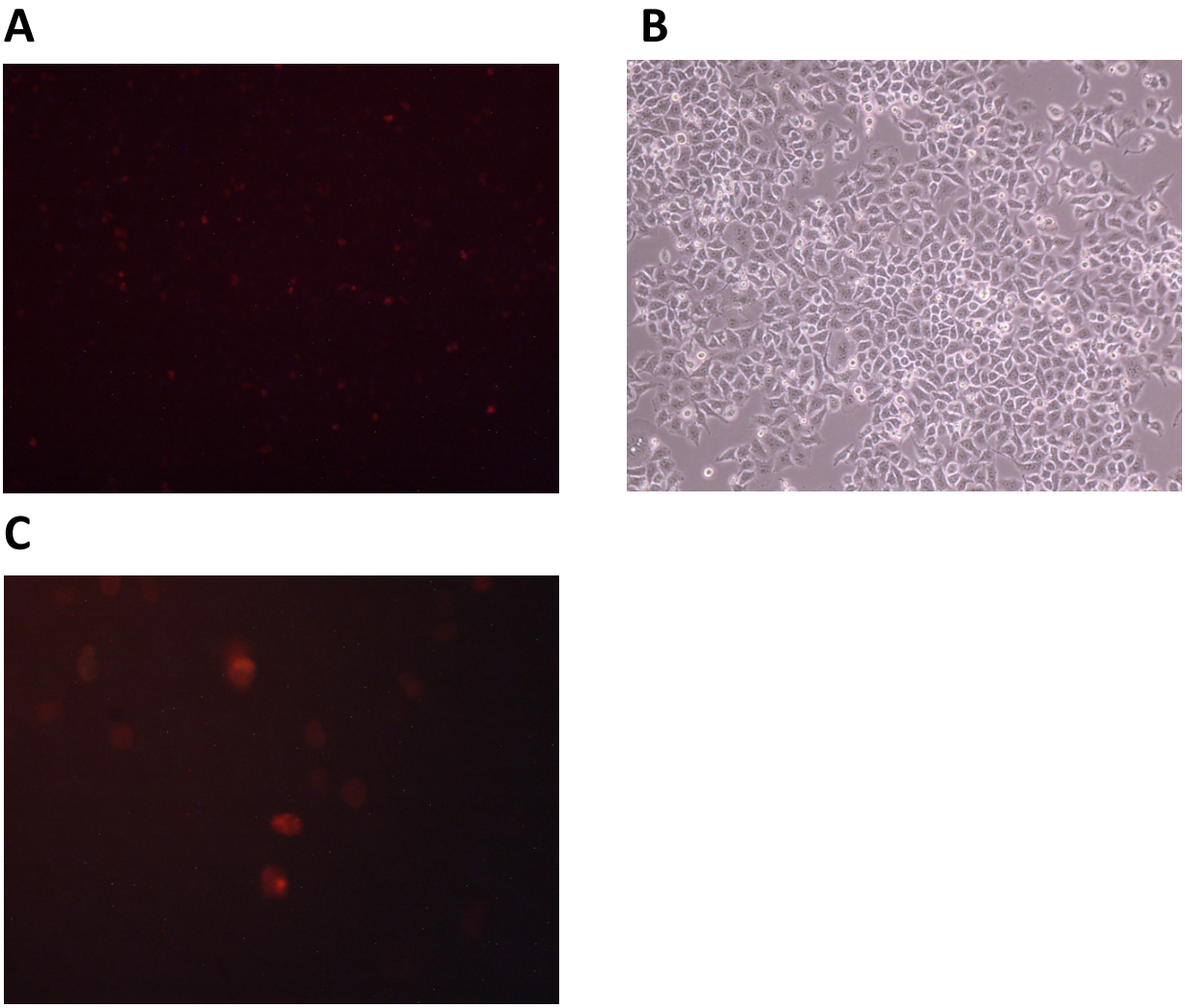


Figure 11. Images of HeLa cells following transfection with mCherry tagged proteins. (A) 10X magnification with 540-580nm filter applied. (B) 10X magnification of HeLa cells. (C) 40X magnification of HeLa cells with 540-580nm filter applied.

2.3.2. Nuclear Extract Preparation

Following transfection as described in 2.3.1., cells were washed once with chilled PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4 and 1.8 mM KH_2PO_4 (pH 7.4)), treated with trypsin and harvested in 10 ml PBS. The cells were pelleted by centrifugation at 1000 rpm for 5 minutes. The cell pellet was suspended in 1 ml PBS and centrifuged at 4 °C at 13,000 rpm for 5 minutes. The cell pellet was re-suspended in an equal volume (packed cell volume, PCV) of buffer A (10 mM Hepes (pH 8), 1.5 mM MgCl2, 10 mM KCl and 1

mM DTT) and left to swell on ice for 15 minutes. Cells were then lysed by forcing them through a 25 g needle ten times. The lysed cells were centrifuged at 13000 rpm at 4 °C for 1 min to pellet the nuclei. The nuclear pellet was re-suspended in 0.7 PCV of extraction buffer C (20 mM Hepes (pH 8), 25% glycerol, 420 mM NaCl, 0.2 mM EDTA and 1 mM DTT). To improve penetration of the buffer, the pellet plus buffer C was incubated at 4 °C with a magnetic stirrer for 1 hour. The suspension was centrifuged at 13000 rpm at 4 °C for 5 minutes to remove the nucleii. The supernatant is dialysed with buffer D (20 mM Hepes (pH 8), 10% glycerol, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT) for 2 hours at 4 °C. The dialysed nuclear extract was snap-frozen in liquid nitrogen and stored at -80 °C.

2.4. In vitro analysis techniques

2.4.1. SDS PAGE

SDS-PAGE gels were made of a resolving gel (370 mM tris base (pH 8.6), 12% acrylamide (ProtoGel, National Diagnostics) plus 0.1 %(w/v) SDS which was polymerized with 0.1 % AMPS and 0.14 % TEMED. Saturated butanol was then added to ensure a level top. A stacking gel (125 mM tris base (pH 6.8), 4% acrylamide, 0.1% SDS polymerized with 0.1% AMPS and 0.2% TEMED) was then added on top and a comb of an appropriate size used. The gels were run in SDS-PAGE running buffer (125 mM tris base, 0.96 M glycine, 0.1% SDS (pH 8:3)), plus Precision Plus Protein Kaleidoscope protein markers (Bio-Rad), at 80 V until the samples were in the resolving gel (~30 min) and at 150 V until the Bromophenol blue was at the bottom (~60 min).

2.4.2. Western Blot

Protein samples were incubated with an equal amount of loading buffer (400 mM Tris (pH 8.6), 3.6 % SDS, 19.2% glycerol, bromophenol blue, xylene cyanol, 1 % freshly added 2-mercaptoethanol) for 5 min at 80 °C. The sample was then run on an SDS-PAGE gel as per section 2.4.1. The gel was soaked in transfer buffer (48 mM tris base, 39 mM glycine, 0.0375% SDS, 10% methanol).

Eighteen pieces of 3 mm chromatography paper (Whatman) and a Nitrocellulose membrane (Amersham Hybond - ECL, GE Healthcare) were also soaked in transfer buffer. Nine pieces of paper, the membrane then nine more pieces of paper were stacked a Biometra Fast- Blot B33 (Bio-Rad). The semi dry apparatus was then ran at 10 W for 30 min. The membrane was incubated overnight at 4 °C in blocking buffer (20 mM

tris- HCl (pH 7.5), 150 mM NaCl, 0.1% Tween-20, 5% milk (w/v), spun down for 5 min at 4637 g). The membrane was then incubated with the relevant primary antibody in blocking buffer for 1 hour at 4 °C followed by 3 washes with blocking buffer for 5 min. This was repeated with a fluorescent secondary antibody and 5 further washes. The membrane was then visualised using an Odyssey (LiCor). Quantification of bands was done using OptiQuant (PerkinElmer).

2.4.3. In vitro Splicing

Reactions containing 5-100 counts/second of radiolabelled RNA, 20 mM creatine phosphate (CrPi), 3.2 mM MgCl2, 20 mM HEPES (pH 7.5), 1.5 mM ATP, 50 mM potassium glutamate (KGlu), 3% RNase- Out (Invitrogen), 0.05% NP-40 and 50% nuclear extract were set up in a volume of 10 µl and incubated at 30 °C. At designated time points, aliquots of 2 µl were removed and placed in a microtitre plate on dry ice. Alternatively triplet reactions were set up in a microtiter plate and incubated at 30 °C for 2 hours. All aliquots were Proteinase K-treated as described in 2.4.5. followed by ethanol precipitation. 10 µl of formamide dyes (containing 5 mM EDTA (pH 8) and 0.2% Xylene Cyanol and 0.2% Bromothenol Blue) were added. The samples were vortexed and heated to 80 °C for 1 minute before being run on a denaturing polyacrylamide gel. The gel was dried and exposed to a phosphorimager screen. Quantification of mRNA bands was done using OptiQuant (PerkinElmer).

2.4.4. In vitro complex formation

Splicing reactions were set up as per section 2.4.3. Time-points of appropriate length were once again taken and frozen on dry ice. For complex E, ATP and CrPi were omitted.

2.4.5. Proteinase K Treatment

50 µl of 0.4 mg/ml Proteinase K (Roche) in 100 mM tris HCL (pH 7.5), 12.5 mM EDTA (pH 8.2), 150 mM NaCL, 1 % SDS was added to each splicing reaction and incubated at 37 °C for 15 min.

2.4.6. Low melting point Agarose Gel Electrophoresis

Complexes other than E were treated with 1 µg/µl of Heparin and incubated for 30 min at room temperature. An equal volume of 50 mM tris, 50 mM glycine and 40 % glycerol were added and the samples were run on a native agarose gel (1 % to 2 % w/v UltraPure Low Melting Point Agarose (Invitrogen), 50 mM Tris and 50 mM Glycine) for 4 to 5 hours at 4 °C. The gel was crushed between four 3 mm paper sheets overnight and afterwards dried and exposed to a phosphor-imager screen. Quantification of bands was done using OptiQuant (PerkinElmer).

2.4.7. Biotin pulldown

40 µl of Neutravidin agarose beads in slurry were washed in 10 volumes of washing buffer (100 mM Tris-HCl (pH 7.5), 1% SDS, 1 mM DTT) five times and in 10 volumes of buffer D (20 mM HEPES (pH 8), 10% Glycerol, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT) once. Pre-washed beads were then combined with a ten-fold (compared to the binding capacity of the beads) excess of synthesised biotinylated RNA/linker and incubated on ice for 30 minutes with gentle agitation. A 200 µl standard splicing mixture (Eperon et al, 1994) minus the RNA was made up, containing 40% pre-cleared commercial nuclear extract (centrifugation at full speed for 1 min), and incubated at 30 °C for 15 min to allow any proteins to be phosphorylated. The splicing mixture was then combined with the RNA/linker pre-bound beads and incubated with gentle agitation for 30 minutes at 30

°C. The beads were subsequently washed 10 times with 5 volumes of FSP buffer (20 mM tris HCl (pH7.5), 60 mM KCl, 2.5 mM EDTA, 0.1% Triton, 100 mM NaCl). The bound proteins were analysed by mass spectrometry after trypsinisation (University of Leicester Protein and Nucleic Acid Chemistry Lab).

2.4.8. Crosslinking

The RNA samples for crosslinking were transcribed containing [α-^{32}P] GTP as described in 2.1.4. The radioactivity of the RNA was then measured using a scintillation counter (Beckman Coulter) and the amount of RNA used calculated depending on the number of Gs present in each sample. The RNA was incubated under splicing conditions, with 50% nuclear extract, at 30 °C for 15 minutes. The reaction mixtures were irradiated with UV for 2 minutes before being treated with a cocktail of RNaseT1/A. The crosslinked samples were ran on an SDS gel, 2.4.1. The membrane was then used for western blotting, to identify the height of a protein in question, and exposed to a phosphor imager screen to identify the amount of crosslinked RNA to the protein of interest.

2.5. TIRF microscope design.

The microscope used was designed and built by Dr Andrew Hudson and Dr. Robert Weinmeister and is described in much greater detail elsewhere (Weinmeister, 2015). The microscope design is that of an inverted microscope utilising total internal reflection. Four diode lasers of wavelengths 488, 532, 561 and 633 nm are available for use; in this study the 532 laser is not used. The different laser beams are combined using a beam combiner that puts them all in the same pathway. An optical fibre runs from the aforementioned beam combiner to the microscope. The laser beam then comes to the beam extender which widens the incoming beam from a 1mm diameter to a 15 mm diameter. The laser beam at this point has a Gaussian intensity profile so a blocking device that only permits the central part of the beam, a 9 mm diameter, through; widening the beam and then narrowing it again by passing it through this aperture produces a beam with a much more homogenous intensity profile. A number of adjustable mirrors guide the beam into the objective at an angle perpendicular to that of the objective. A TIR mirror which sits below the objective reflects the excitation beam into the objective. The laser is then focussed onto the back focal plane of the objective by the TIR lens. The beam then reaches the interface with the sample, where the laser illuminates an area with a diameter of around 25µm, at an angle in excess of the required critical angle for total internal reflection. An evanescent field is created which fluorophores near the surface.

The emission bean is then collected again by the objective and passed through a number of notch filters which are suited for the wavelengths of the lasers used to excite the sample; this reduces interference by scattering the excitation beam. An EMCCD captures

the images from the point at which TIR occurred. The reflected beam from the sample is separated from the emission beam by a small mirror. This beam is then directed onto a quadrant photodiode which is used to keep the microscope in TIR and focus.

2.6. TIRF data acquisition.

2.6.1. Sample Chamber Preparation

22 by 50 mm cover slides, #1 (Menzel-Gläser), were washed in 1 M KOH for four hours. This was followed by ten washes with distilled water then sonication in a water bath for 12 min; this was repeated at least five times. The slides are then dried using a vacuum and were sometimes further dried in a desiccator. The slides were further cleaned using an argon plasma cleaner (MiniFlecto-PC-MFC, Gala Instrumente) five times, at five minutes a time, with pure Argon at 0.15 mbar and an applied power of 80 W.

One to five channels were then created using double-sided tape. Each channel was 5 mm to 10 mm wide. 22 by 22 mm cover slips, #1.5, were then used to close the chambers.

2.6.2. Preparation of Samples for the Splicing Complexes E, A and I

Reaction mixes containing 50% nuclear extract, 3.2 mM MgCl2, 50 mM KGlu and 3 units RNase OUT (Invitrogen) were prepared. For complex E this mix was used. For complex A, 20 mM CrPi, 1.5 mM rATP and 20 mM Hepes (pH 7.5) were added to allow complex progression. In order to prevent progression past complex A 1 µM of an oligonucleotide complementary to the U6 snRNA was added. In cases where the binding of the U1 snRNP was blocked, 2 µM of an oligonucleotide complementary to the U1 snRNA was added. For experiments were the de-phosphorylation of proteins was required, 1xPhosStop (Roche) was added; it is provided as a 5x stock.

Reaction mixtures were incubated for 15 mins at 30 °C before the pre-mRNA was added at a final concentration of 31.25 nM or 62.5 nM and the reaction was incubated for a further 15 minutes at 30 °C.

2.6.3. Sample dilution for the microscope

The samples were serially diluted 10,000 or 20,000 times to 5 pM or 20 pM RNA, using buffer A2 (40 mM Hepes (pH 7.5), 3.2 mM $MgCl_2$, 50 mM K-Glu, 50 mM KCl, 0.1 mM EDTA and 0.5 mM DTT. The samples were then loaded into the sample chamber and incubated for 5 min.

2.6.4. Two/Three-Colour acquisition

One or two different splicing factors and the RNA were labelled with different fluorophores. The fluorophores used in this study were mEGFP (also referred to as GFP for the rest of the study), mCherry and Cy5; the excitation and emission spectra of these are sufficiently separate. These correspond to the available diode lasers of wavelength of 488 nm, 561nm and 633 nm respectively. The 488nm weakly excites mCherry as well as GFP and the 561nm laser also excites the Cy5 weakly; sequential illumination of the order 633>561>488 was used to prevent cross-excitation. Images for all the different fluorophores were obtained on the same camera using the same beam path. This reduced the variation of the acquired data and simplified the later data analysis.

For example, excitation with the 633 nm laser was used first to excite Cy5; this is unable to excite mCherry or GFP so the collected signal originated from the Cy5 labelled RNA only. This was done until complete bleaching of the fluorophore has occurred. The excitation laser was switched to the 561 nm laser, which excited mCherry. After

bleaching, the 488 nm laser was switched on to excite GFP. This laser remains on until bleaching was complete.

2.6.5. Automated Data Acquisition with LabVIEW

An individual experiment involving the microscope actually consisted of a large number of sequential acquisitions. This was done to obtain enough images and in turn enough spots to provide an accurate representation of the binding pattern of a single factor; fifty different images usually provide 100 to 200 co-localised spots. To collect each image, the sample cover slip had to be moved to explore a new, un-bleached, area. During each image acquisition the excitation lasers were turned on/off in a sequential order but the sample had to be kept in focus, which may be different for each laser. A LabView program written by Dr. Robert Weinmeister was used to control the microscope. This allowed for a degree of automatic data acquisition, requiring only an initial manual setup and a small number of prompts during acquisition.

The CCD chip was cooled to -80 °C. The exposure time, 0.1 or 0.05 seconds usually, and EM gain of the camera, usually 300, were input. The power of each laser was chosen and the focus point for that wavelength found and saved. The CCD chip of the camera is 512 by 512 pixels but the size of the part used was chosen was the central 250 by 250 pixels; this is so as to remove any variation in beam intensity across the chip. The excitation beam which was reflected back and out onto the quadrant photodiode was used to provide a feedback loop which kept the microscope in focus. Nine sequential acquisitions were automatically acquired, with the microscope cycling through the excitation lasers, at specified frames, and adjusting the focus in accordance with that laser. Images were saved, with a name that was input, as stacked TIFF files with a

corresponding text file attached which contained any input annotations describing the experiment.

2.7. TIRF data analysis

The analysis programs used for this work were written in Matlab by Dr Robert Weinmeister.

2.7.1. Spot detection

A moving average was calculated to subtract the underlying background followed by an adjustment of the brightness/contrast that gave an enhanced image that has less variability from image to image and also made spots more distinct and therefore easier to select. Spots from images were then identified by an algorithm as well as subsequently by eye. The spots were recognized and refused or accepted depending on a number of factors; intensity, shape and proximity of other intensity peaks. The algorithm used in this study utilises a box of 19 by 19 pixels which was moved across the image. If a pixel had an intensity of 1.119 times or more of the average of the rest of the pixels then it was considered a potential spot; in cases where there were two pixels that fulfilled this criterion, the box was reduced in size until only one remained. The intensity of the potential spot was then fitted by a Gaussian distribution. Checks on spot identification of Cy5 spots showed that no false positives are identified, with the only error coming in the form of two spots that were close together being identified as one as opposed to two; this error was easily rectified by eye.

2.7.2. Co-localisation and dealing with Chromatic aberrations

Following spot assignment, the co-localisation of a spot with a spot of another wavelength needed to be analysed. Spots derived from images at a given wavelength were identified as “marker spots”; this was nearly always Cy5 from the labelled RNA. The assessment of co-localisation can be done solely by eye; images from the two

different fluorophores were overlaid and due to false colouring, paired spots could be identified. However, determination of co-localisation by eye was open to a large amount of user error and bias; this is exacerbated by the chromatic aberration or colour shift of the spots from different wavelengths.

Different wavelengths emitted from a single point did not actually fit perfectly on top of each other. This phenomenon is known as chromatic aberrations and is an effect that results from the failure of a lens to focus different wavelengths onto a single same convergence point. This is the case because lenses have different refractive indices for different wavelengths of light. A transformation was applied to spots from non-reference wavelength to test whether they were co-localised with the marker spots. The transformation was calculated once per month by using an RNA transcript with two fluorescent oligonucleotides attached (Cy5 and Atto488 dyes were used in this case). These spots should be perfectly co-localised and enabled the shift needed to be calculated as Xc = A, SFx = B, Yc = C and SFy = D where A-D are integers. Xc and Yc are the co-ordinates of where no shift needs to be applied whilst SFx and SFy indicate the scale factor of the shift. The scale factor takes into account a weighting parameter which allows the shift to be used for different fluorophores, this is set to 0 for Cy5, as shifts are calculated relative to Cy5, 0.52 for mCherry and 1 for GFP.

Once the positions of the spots were transformed, the distance between the marker spots and those produced by a second laser were calculated. A threshold was set at two pixels for all of the work displayed here as even following the calculated transformation, perfectly co-localised spots were as much as two pixels apart. If more than one spot can be considered co-localised the closest is used.

The co-localisation of two fluorophores is displayed as the percentage of the marker spots that are co-localised.

2.7.3. Step detection

If two spots were considered co-localised the intensity traces over time were used to analyse the number of fluorophores present within a single complex. The irreversible bleaching of a number of fluorophores was detected as step-wise changes in intensity. Counting these steps allows us to determine the number of fluorescent proteins present.

The background intensity was subtracted by deducting the average intensity of a 13 by 13 pixel square from a 3 by 3 pixel square around the spot. The relatively small square used for the spot means that the chance of additional spots influencing the spot intensity was low and the relatively large area used for the background intensity means any additional spots nearby would not noticeably affect it. Distinct steps were initially assigned using an algorithm based on a Bayesian step detector (Weinmeister, 2015). By following a running average of intensity over time the algorithm detects points at which the average intensities before and after were significantly different. Each region was then re-examined by the same process. This was repeated until no additional changes were detected or the trace was too short. Re-activated fluorophores still resulted in the correct number of steps being assigned as the assessment is based on the resulting plateaus and not the location of those plateaus e.g. parts of the trace that belong to the same fluorophore that had been re-activated were assigned to a single step. The algorithm allowed the identification of up to 8 to 9 steps per intensity time trace. A subsequent manual review of the intensity time traces was then performed to ensure

all steps were counted or to remove any that have been miscounted. An annotation of "un decided" was assigned if the intensity time trace showed that co-localisation was correct but there were an unknown number of steps. The resulting frequencies of spots that had a set number of steps were collated and displayed in histograms.

2.7.4. Data correction

The data presented in this thesis is un-corrected. Potential correcting for the level of labelling, binding efficacy of labelled proteins vs un-labelled proteins, protein concentration and potential protein dimerization is possible but would be an uncertain procedure. The level of labelling analysed would not be a perfect measure, as GFP/mCherry that has bleached in the extract would lead to less labelled protein whilst the measure of labelled protein would remain the same (less actively fluorescently labelled proteins). The binding efficiency of proteins vs labelled proteins would depend on what it was binding to (i.e. labelling may effect RNA binding but not its protein interactions). Protein concentration would depend on the Kd of the protein and as before this is dependent on the binding partner. Dimerization may depend on the context i.e. in complex the protein may be a monomer but un-bound it may be a dimer. As such, the data is not presented as showing definitive numbers of proteins binding, instead changes to the patterns observed are discussed and quantitative observations such as changes to co-localisation, changes to ratios of numbers in bins and basic statistical fitting to simple expectations is used. Qualitative observations of changes to patterns, when different RNAs or complexes are used, are also discussed. This combination of basic quantitative observations along with qualitative observations allows us to analyse the effect of differing RNA sequences and complexes on protein binding.

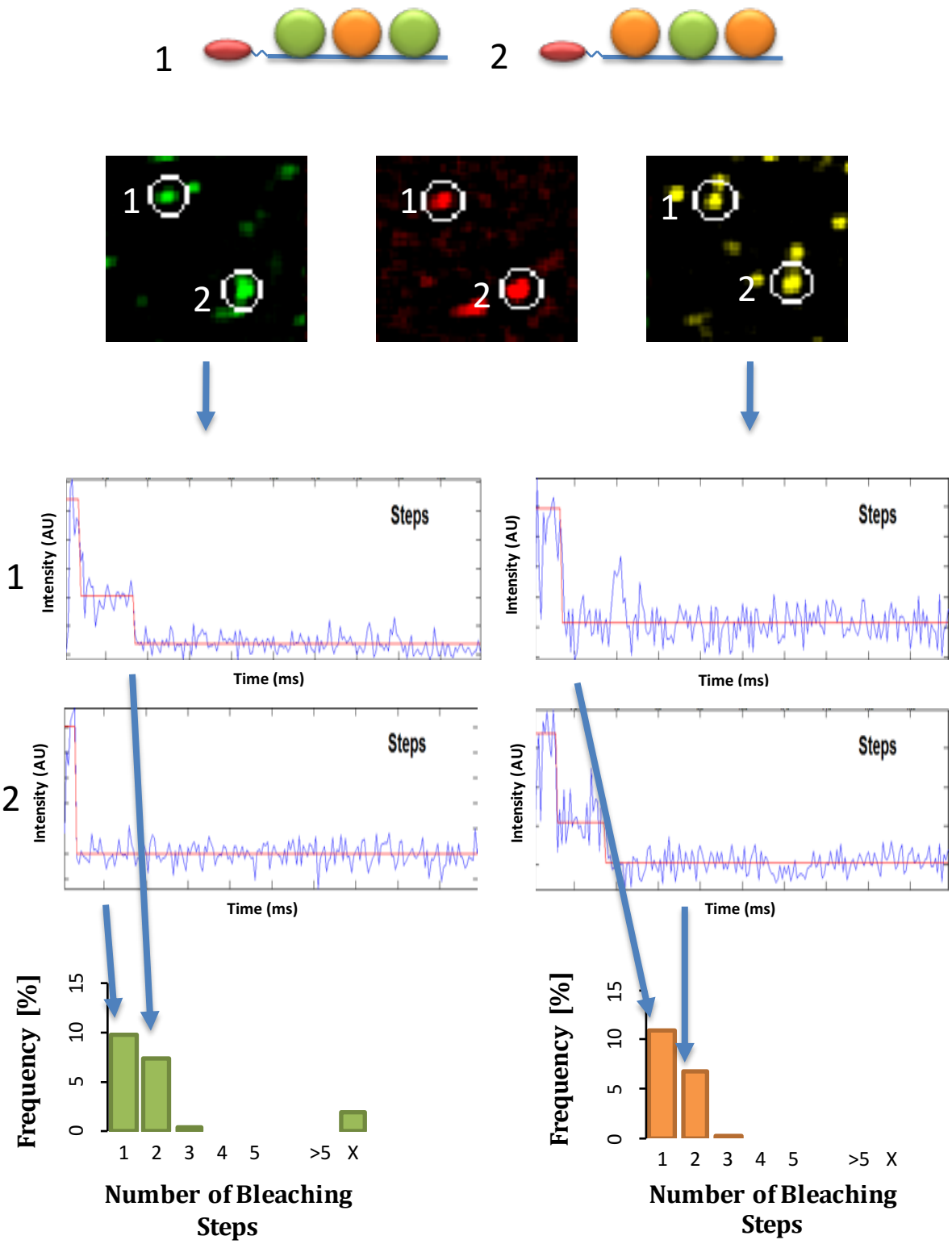


Figure 12. How a single molecule of RNA with different numbers of two different factors bound relates to the spots observed, bleaching steps analysed and the histograms generated.

Chapter 3. Relationship between SRSF1 and U1 at 5' splice sites.

3.1. Introduction.

The generally accepted model for 5'SS activation by SRSF1 is that it binds exonic splicing enhancer sequences (ESEs) (Tacke & Manley 1995), and then recruits U1 snRNPs by direct protein interactions (S. Cho et al. 2011). This model is commonly depicted in cartoon representations of splicing. This model is compatible with the first two, and most widely accepted, properties that led to the isolation of SRSF1 originally. The first is its ability to modulate 5'SS selection (Ge & Manley 1990) i.e. by binding to exonic sequences to drive splicing to specific sequences nearby. The second is its ability to restore splicing activity to S100 cytoplasmic extracts (Krainer et al. 1990; Krainer et al. 1991) i.e. by binding to exonic sequences and stimulating U1 binding constitutive 5' splice sites. Despite the extensive evidence supporting the ESE driven recruitment model, a number of observations (outlined in the introduction) seem to point towards SRSF1 playing a role in constitutive splicing reactions. In order for this to be plausible, and preliminary data from our lab would seem to suggest this, an alternative recruitment mechanism may be necessary.

Prior work by (Weinmeister, 2015) performed using the same TIRF microscope setup as used in this study showed that a single molecule of SRSF1 was recruited to pre-mRNAs in A complex. This work was done on RNA derived from Beta-Globin RNA (Globin) which contained a single strong 5' splice site and a strong 3' splice site. This RNA also contains a number of proposed exonic enhancer sequences in the 5' exon. The observation that a single molecule of SRSF1 was recruited to each RNA could fit with a number of models;

a single molecule of SRSF1 may be recruited per 3’ splice site, a single molecule of SRSF1 may be recruited per 5’ splice site, a single molecule of SRSF1 may be recruited per intron, a single molecule of SRSF1 may be recruited per A complex or a single SRSF1 may be recruited to the exonic sequences.

In order to look at this further, an RNA that contained two introns, created by duplicating the Globin RNA sequence, was used. In this case there was a strong increase in the binding of a second molecule of SRSF1. This observation however still fits with all the above models. To try to narrow down on the exact mechanism of recruitment, the Globin RNA was used but the binding of U1 was blocked using an oligonucleotide complimentary to the 5’ end of the U1 snRNA. Here the stoichiometric recruitment of a single molecule of SRSF1 was lost and replaced with a non-specific pattern that is characteristic of background binding for RNA that cannot form complex A. Further experiments using a mutated 3’ splice site still showed a single SRSF1 being recruited but with some background as would be expected with poor A complex formation but some stoichiometric SRSF1 recruitment. These results showed that the stoichiometric recruitment of SRSF1 relied on U1 binding to a 5’ splice site.

These experiments combined remove a number of the possibilities for SRSF1 recruitment. However we are still left with the possibility that the U1 snRNP is recruiting SRSF1 in a 1:1 stoichiometry (figure 13A) or that exonic sequences are recruiting SRSF1 and they need U1 to bind stably (figure 13B). In this chapter we show experiments that shed further light on how SRSF1 is recruited to pre-mRNAs.

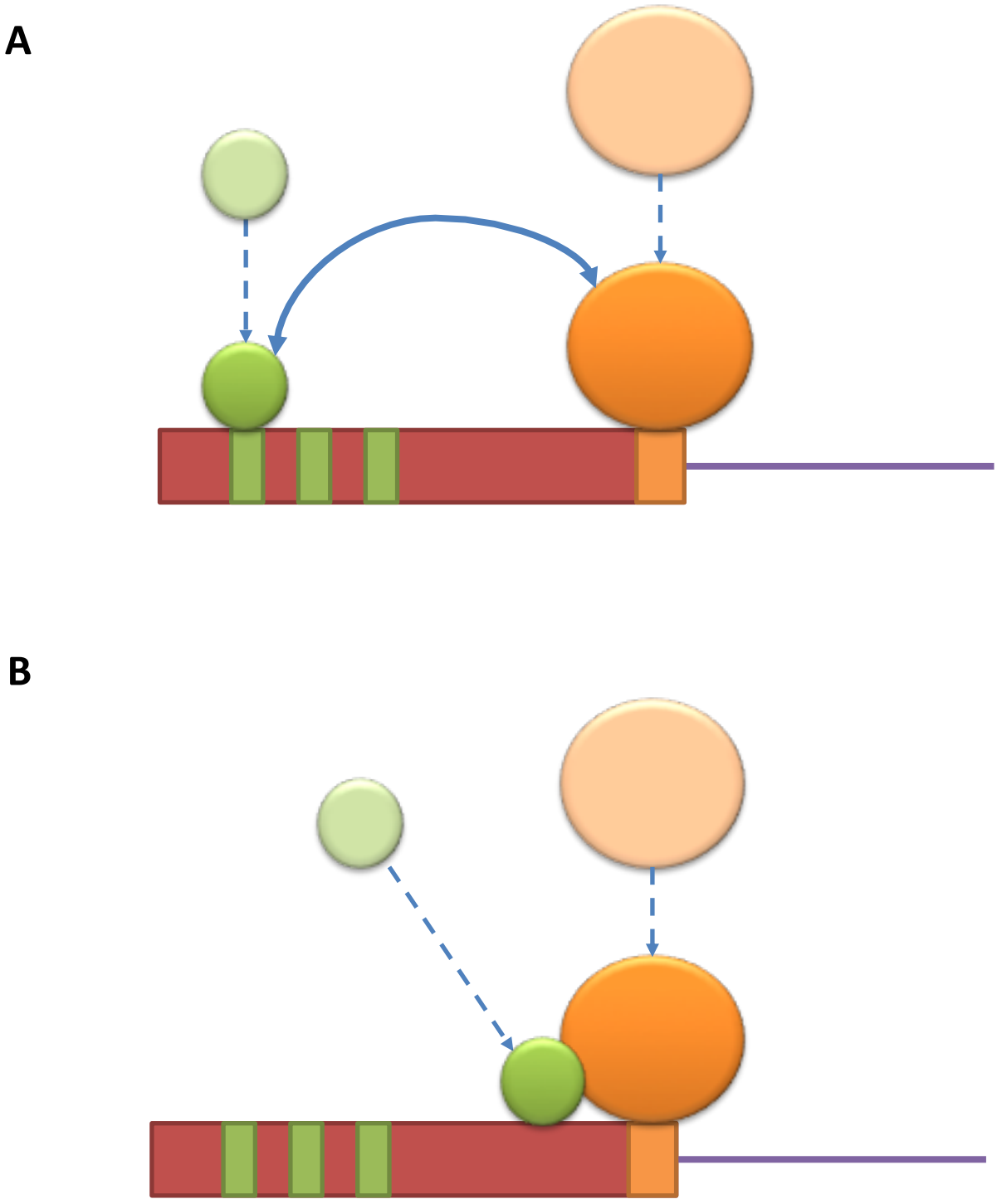


Figure 13. Possibilities for the recruitment of SRSF1 to RNA following Dr. Weinmeisters preliminary work. (A) A model whereby SRSF1 is recruited to ESEs but requires a bound U1 snRNP for stable binding. (B) A model whereby the U1 snRNP recruits SRSF1 itself.

3.2. In vitro analysis and validation of SRSF1, U1A/SRSF1, DDX5 and DDX5-NEAD nuclear extracts.

The U2B'' and U2AF35/65 nuclear extracts used in this chapter are described and validated in Chen et al (Chen et al. 2016).

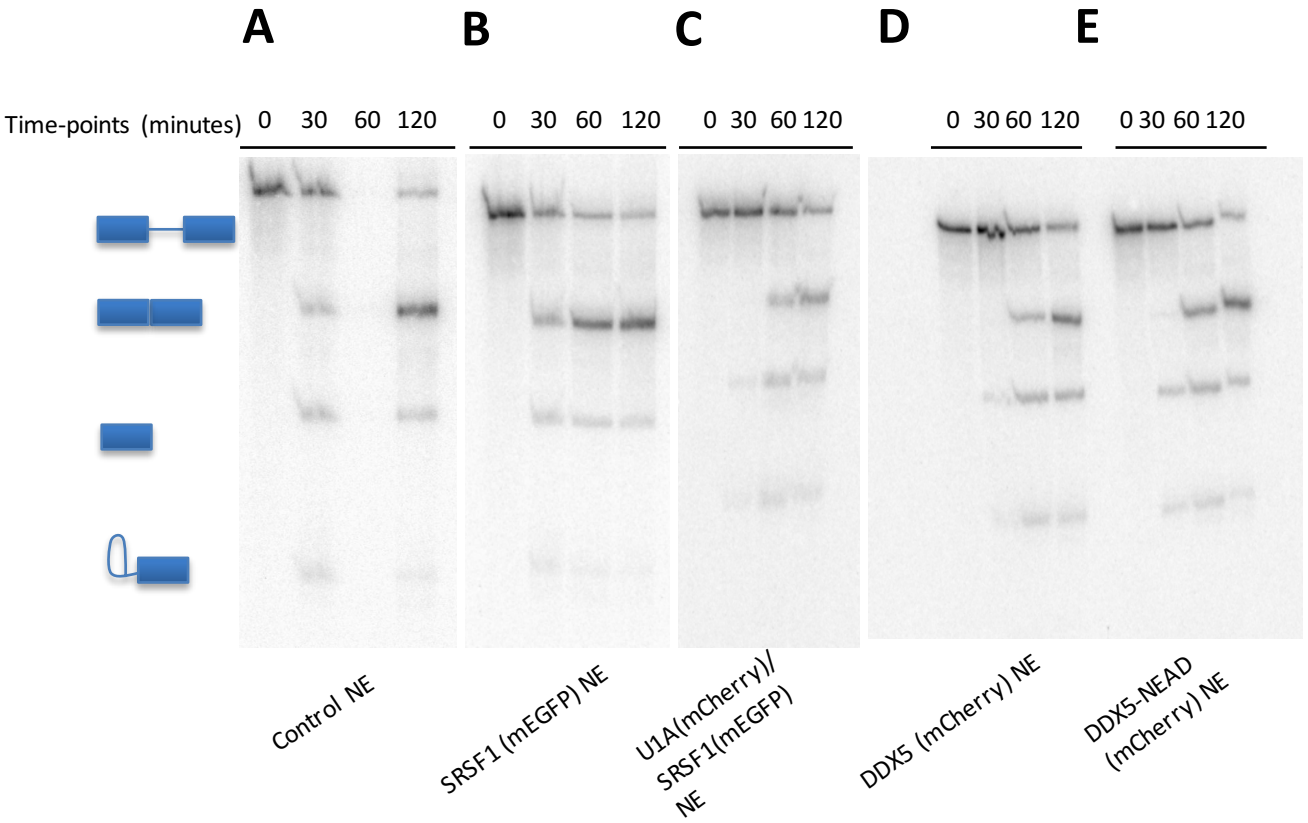
For the SRSF1, U1A/SRSF1, DDX5 and DDX5-NEAD nuclear extracts, splicing reactions were done with a Beta-Globin derived construct (Globin). Figure 14B to E show the levels of splicing compared with those produced by a commercial nuclear extract, (Figure 14A). All give good levels of splicing comparable to the commercial nuclear extract (Figure 14F).

To be able to analyse experiments correctly and model the number of proteins that are binding to our RNA, the ratio of labelled protein vs endogenous protein needs to be checked. For this, western blots along with the appropriate antibody for the protein being analysed are used. Figures 15A to D show the western blots for each extract with two for the dual labelled extract. The levels of labelling are shown in the table in Figure 15E. Interestingly whilst the DDX5 extract showed good levels of labelled protein, the DDX5-NEAD extract seemed to show low levels. One possibility was that the mutation would disrupt the antibodies binding to the protein, but this seems unlikely given the sites of antibody binding and the mutation, although changes in how the protein may fold could equally have an effect. Results described later seem to suggest however that the level of labelling of the two extracts is roughly the same. Use of a different antibody could also help to answer this question.

The concentration of labelled protein in the first extract used, mEGFP-SRSF1, was also tested to ensure that results obtained from transfected nuclear extracts were comparable to cellular conditions i.e. comparable protein levels. Figure 16 shows a western blot using an anti-GFP antibody and increasing concentrations of GFP protein which were used to create a standard concentration curve. The anti-GFP antibody was then used with the SRSF1 nuclear extract and the intensity used to obtain the concentration. Given that the concentration was comparable to estimated cellular conditions, 1.3 µM (Mayeda et al. 1993), it was inferred that binding results were comparable.

Figure 17 shows the splicing of all the RNA constructs that are used in this chapter. The transcript consisting of exon 2 of Beta-Globin, a hybrid intron and exon 7 of SMN2 (BG-SMN2), the transcript with two repeats of the Ron ESE added to the 3' end (BG-SMN2+ESEAx2) and the transcript with a U1 binding site at the 3' end (BG-SMN2+U1) are shown here and described in further detail in Chapter 4; both the U1 site and two repeats of the ESE increase splicing significantly. The Globin transcript shows strong splicing which increases significantly with the addition of a 3' U1 binding site (Globin+U1). Additionally, a transcript that contained 2nts of Beta-Globin exon 2 and 2nts of Beta-Globin exon 3 (Globin 2nt), Globin but with a deleted 5' splice site (Globin M) and an Adenovirus derived construct that contained two alternative 5'SS (Ad) were also used. Globin 2nt can form complexes and undertake the first step of splicing whilst Globin M does not splice or form complexes and Ad splices to both splice sites (data not shown). Figure 18 shows schematics of all the transcripts and the sequences are shown in the appendices.

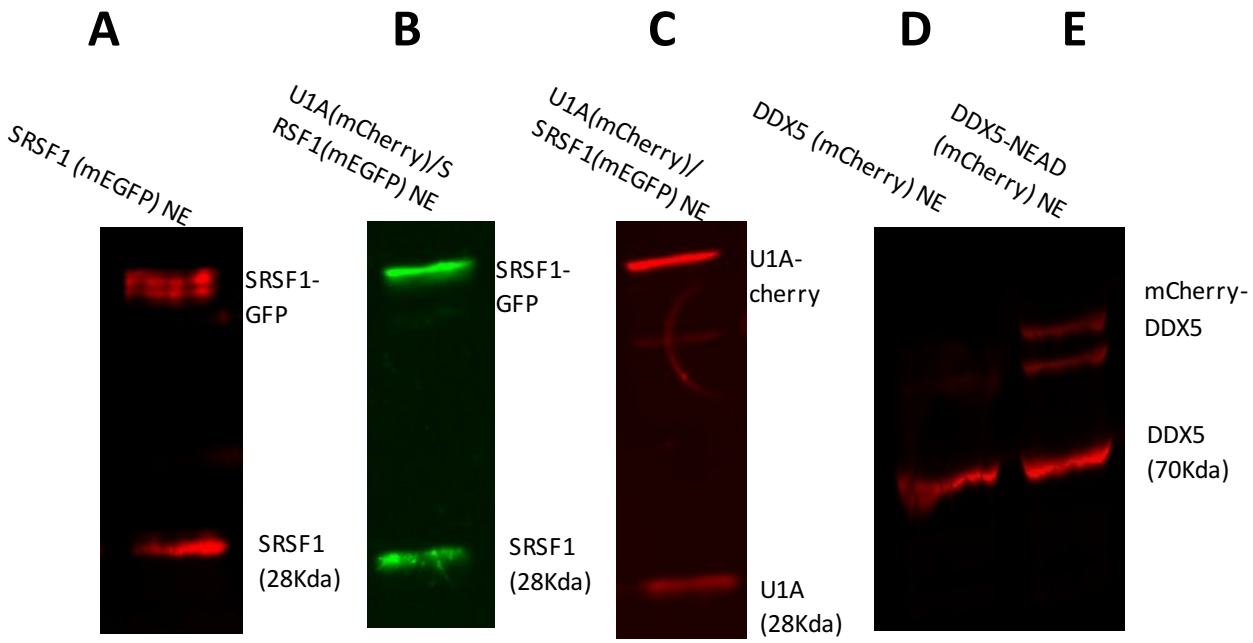
In order to analyse changes to protein binding, figure 18 shows idealised histograms for single or double occupancy when 60% of the proteins in the extract are labelled; which is the case for the level of labelling of SRSF1 in both the single and double labelled extract and for U1 in the double labelled extract. A histogram showing a geometric pattern, which is typical of non-specific binding of proteins, is also shown. In the following results sections, references to a geometric pattern refers to a pattern as shown in figure 18; which results from each successive bin being half the size of the previous bin.



F

NE	Control	SRSF1(mEGFP)	U1A(mCherry)/ SRSF1(mEGFP)	DDX5(mCherry)	DDX5-NEAD(mCherry)
Splicing efficiency after 2 hours (mRNA/(mRNA +pre-mRNA)	83	85	54	59	80

Figure 14. Polyacrylamide gels showing the activity of splicing in transfected nuclear extracts used for single molecule studies in this chapter. (A) Splicing time course of Globin with 50% commercial nuclear extract. (B) Splicing time course of Globin with 50% mEGFP-SRSF1 nuclear extract. (C) Splicing time course of Globin with 50% mCherry-U1A/mEGFP-SRSF1 nuclear extract. (D) Splicing time course of Globin with 50% mCherry-DDX5 nuclear extract. (E) Splicing time course of Globin with 50% mCherry-DDX5 (NEAD) nuclear extract. (F) Quantification of splicing efficiency taken from two hour time point.

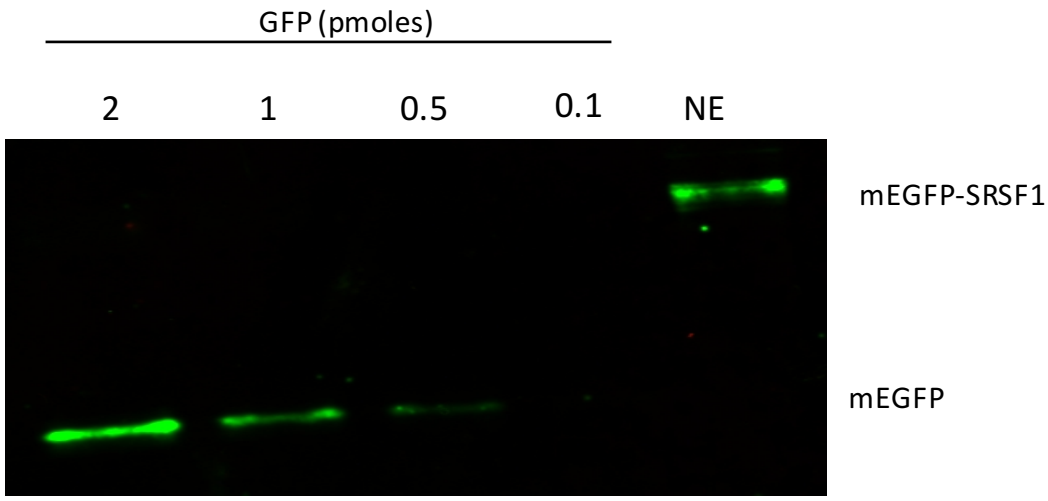


F

NE	SRSF1(mEGFP)	U1A(mCherry)/ SRSF1(mEGFP) (Anti SRSF1)	U1A(mCherry)/ SRSF1(mEGFP) (Anti U1A)	DDX5-NEAD(mCherry)	DDX5(mCherry)
% labelled	59	58	57	18	40

Figure 15. Western blots demonstrating the level of labelled protein present in each nuclear extract. (A) Western blot using anti-SRSF1 antibodies (CSH) on the nuclear extract containing mEGFP-SRSF1 (Christian Lucas). (B) Western blot using anti-SRSF1 antibodies on the nuclear extract containing mEGFP-SRSF1 and mCherry-U1A (Christian Lucas). (C) Western blot using anti-U1A antibodies (Abcam) on the nuclear extract containing mEGFP-SRSF1 and mCherry-U1A (Christian Lucas). (D) Western blot using anti-DDX5 antibodies (Abcam) on the nuclear extract containing mCherry-DDX5. (E) Western blot using anti-DDX5 antibodies (Abcam) on the nuclear extract containing mCherry-DDX5 (NEAD). (F) Quantification of the percentage of labelled protein present in each extract.

A



B

NE	mEGFP-SRSF1
Concentration (µM)	0.72

Figure 16. Western blot to determine the concentration of labelled protein in the mEGFP-SRSF1 nuclear extract. (A) Western blot using anti-SRSF1 antibodies (CSH) and increasing concentrations of GFP protein in comparison to the mEGFP-SRSF1 nuclear extract. (B) Quantification of the concentration of the labelled protein in the extract.

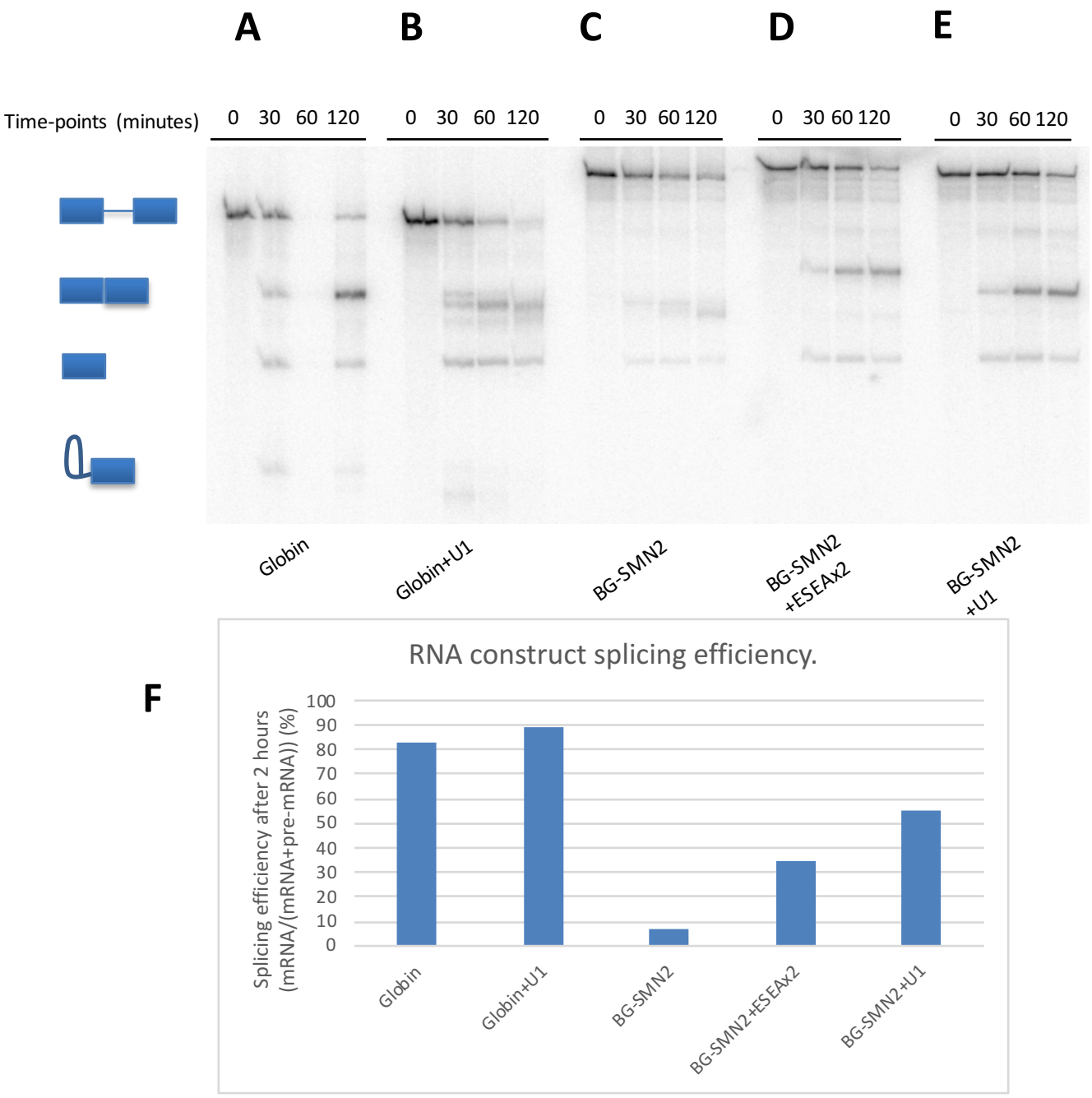


Figure 17. Polyacrylamide gel showing the splicing of the different RNA constructs used in this chapter. (A) Splicing time course of Globin. (B) Splicing time course of Globin+U1. (C) Splicing time course of BG-SMN2. (D) Splicing time course of BG-SMN2+ESEAx2. (E) Splicing time course of BG-SMN2+U1. (F) Quantification of splicing efficiency from the two hour time point.

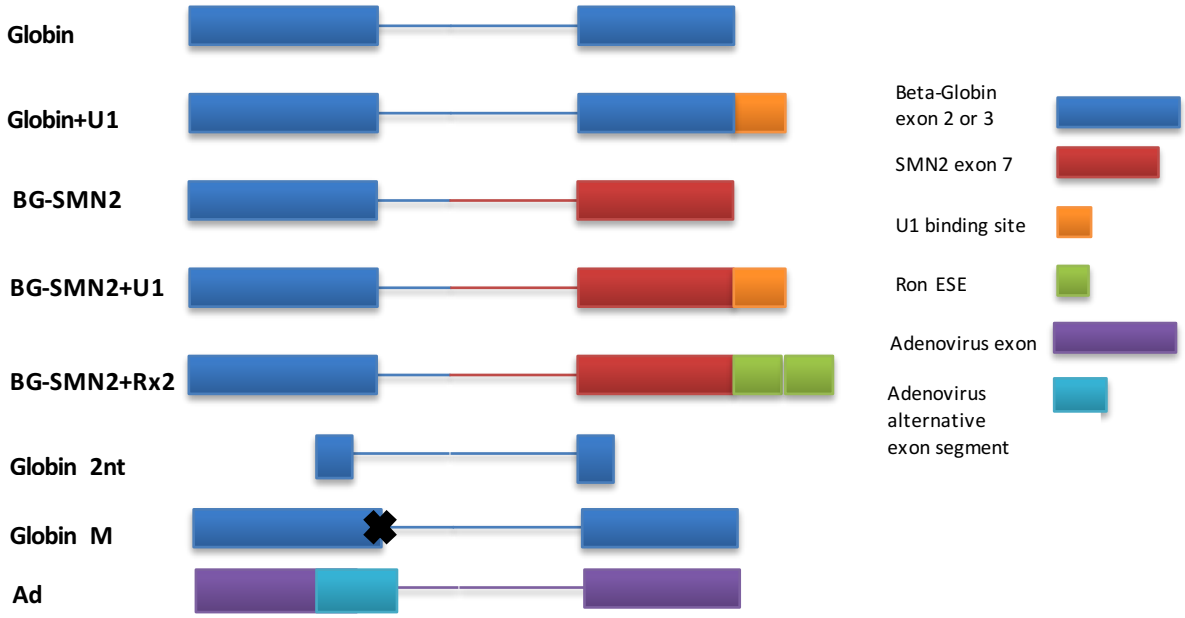


Figure 18. Schematics of RNA constructs used.

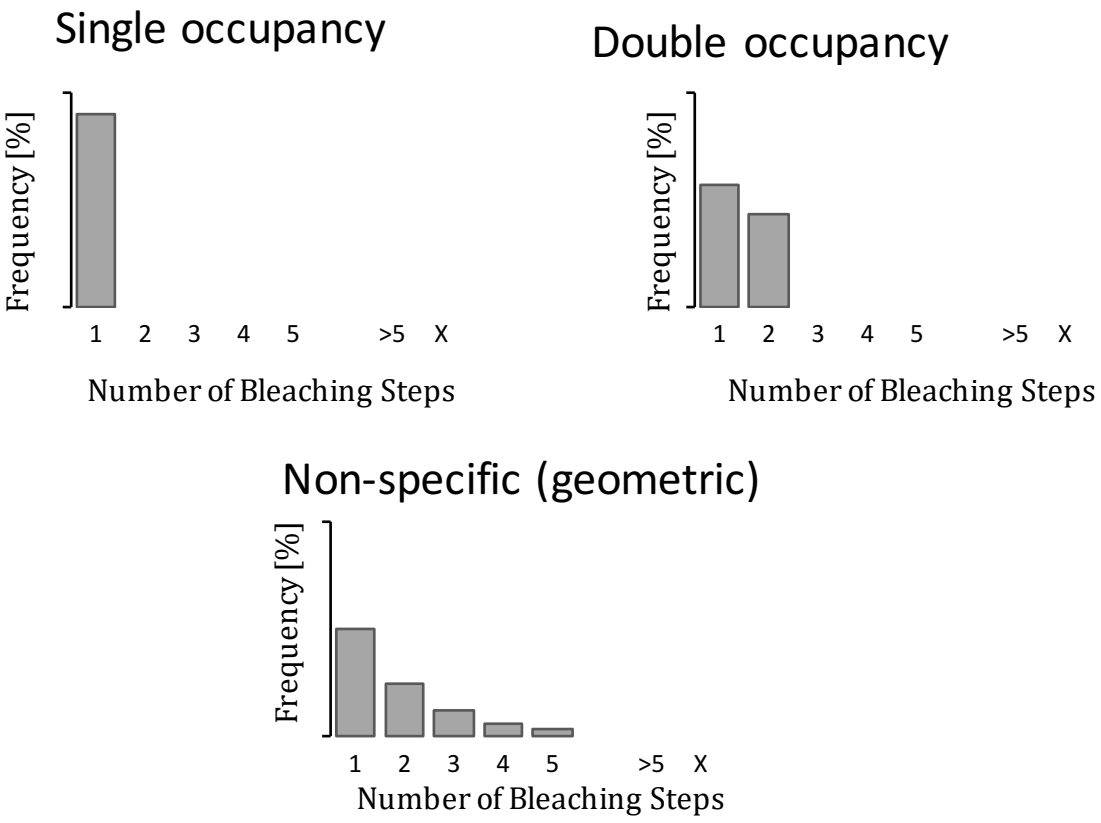


Figure 19. Idealised histograms for a single protein binding or two proteins binding when 60% of the proteins in the extract are labelled. A histogram showing non-specific binding that shows a geometric distribution.

3.3. Does U1 recruit SRSF1 to the 5'SS?

In order to look at the relationship between SRSF1 and U1 snRNPs at 5' splice sites during complex A, it was first necessary to look at the binding of both to our Globin splicing construct that contains a single strong 5' splice. The construct splices efficiently and forms complex A rapidly. Figure 20A shows that mCherry-labelled U1A, which is known to bind only as part of the full snRNP (Will et al. 1993; Hodson et al. 2012), binds in a 1:1 ratio with the cy5 labelled RNA in complex A conditions. This is to be expected from an RNA that contains a single strong binding site. Similarly Figure 20B shows the GFP-SRSF1 seems to bind in 1:1 ratio with the RNA also as shown by Dr Weinmeister. In both cases the small amount of RNA that binds two or three molecules of SRSF1 or U1 are expected to come from RNAs that have not progressed to complex A and therefore still have some non-specific background binding as mentioned before or protein molecules that have dimerised.

We then further analysed whether the GFP-SRSF1 and mCherry-U1 could co-exist on the RNA at the same time. This is to show that the two are not mutually exclusive or that the two patterns aren't coming from two separate populations of cy5-RNA. Figure 20C shows the binding pattern of mCherry-U1A to the RNA when GFP-SRSF1 is also present (three way co-localisation). The pattern is the same ($P_{20C=20A}$=0.93), showing that the binding of SRSF1 does not affect the binding of U1. Furthermore the co-localization percentage is half that of the co-localisation in figure 20A which is what one would expect if all the RNA molecule that had a SRSF1 bound also had a U1 bound given the level of labelled mCherry-U1A to unlabelled endogenous U1A. Once again, similarly the pattern for SRSF1 does not change ($P_{20D=20B}$=0.76) when we analyse cases where both

RNA and U1 are present. Also similarly the co-localization percentage is once again halved. As before, given the percentage of SRSF1 that is labelled vs the percentage that is unlabelled, this indicates that there is always an SRSF1 present when there is a U1 present.

These results confirm that a single U1 binds to Globin RNA, a single SRSF1 binds to Globin RNA and that both are present at the same time on the RNA. However these findings could still fit to several of the models outlined previously; SRSF1 may bind to exonic sequences and is stabilized by binding to U1, a single SRSF1 may be recruited per complex A that is formed or a single SRSF1 may be recruited per 5'SS/U1snRNP that binds.

In order to rule out that a single SRSF1 binds per complex A, an RNA that contains two strong 5' splice sites was used, Ad. Once again we looked at the binding of both U1 and SRSF1 to the RNA in complex A. Figure 21B shows the U1 binding and as expected with two 5' splice sites we see an increase in the number of RNAs with two molecules of U1 bound (figure 18, double occupancy). Likewise with SRSF1 we see an increase in the number of molecules with two bound, figure 21A ($P_{\text{double occupancy}}$=0.75), compared to the RNA with a single 5' splice site, figure 20B. This shows a single molecule of SRSF1 is recruited per 5' splice site/U1. This is supported by figure 21C and D which shows Globin M and the Globin construct in the presence of an oligonucleotide that is complementary to the 5' end of the U1 snRNA, respectively. In both cases the level of SRSF1 binding to the RNA is severely reduced, 19% to 11% and 9% respectively, and the background binding is increased; a pattern that is consistent with poor complex formation (figure 21D, P_{geo}=0.72). Therefore it appears that SRSF1 binds in a one to one stoichiometric

manner with 5’ splice sites/U1 snRNPs. However we still can’t rule out that SRSF1 is recruited by exonic sequences and is only stabilized by an interaction with U1.

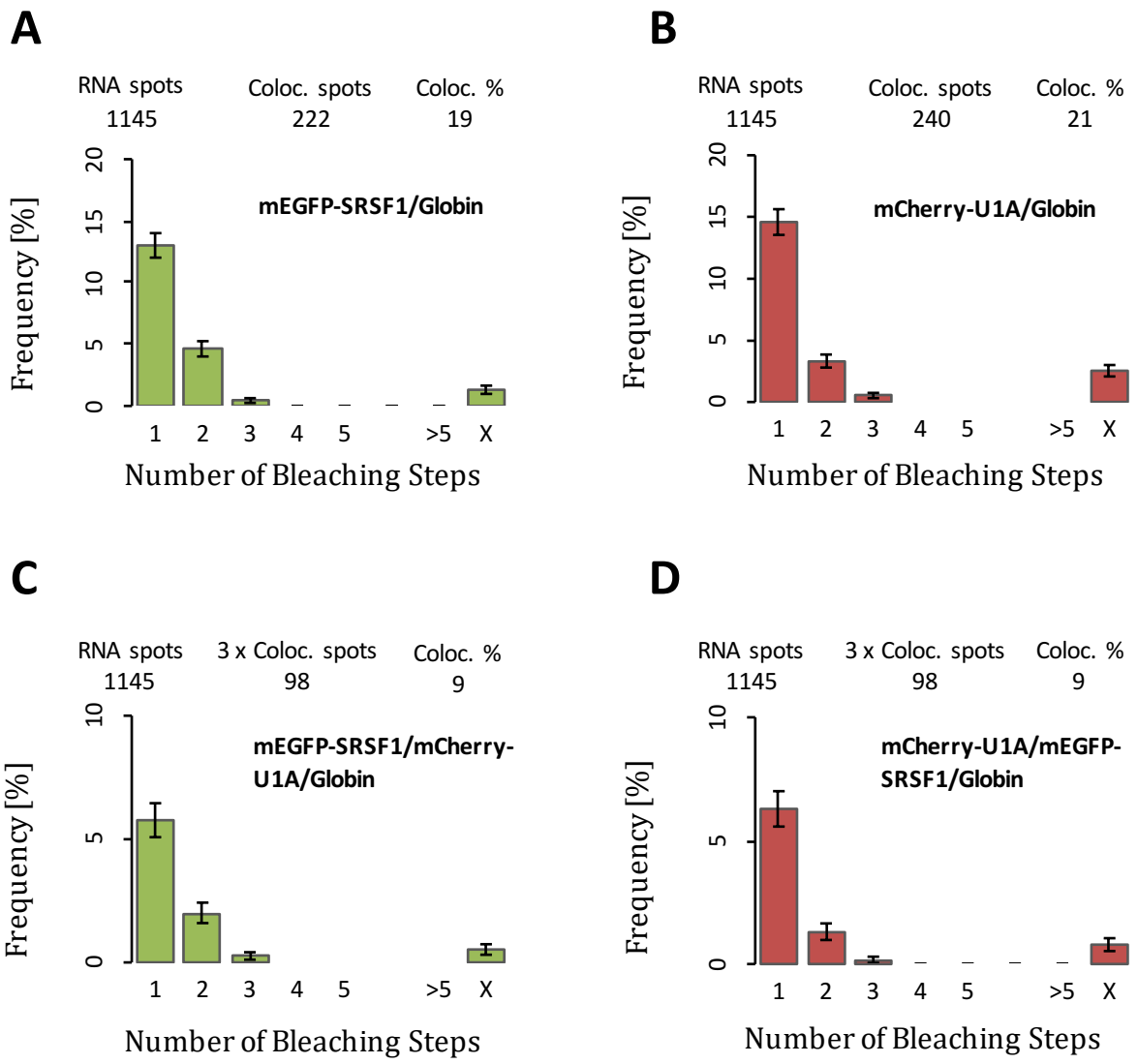


Figure 20. Single molecule multicolour co-localization studies on the levels and stoichiometries of mEGFP-SRSF1 and mCherry U1A binding to molecules of labelled pre-mRNA in nuclear extracts. Histograms show the frequencies (%) of pre-mRNA molecules showing bleaching of co-localized labelled protein in n steps. > refers to complexes where more than 5 bleaching steps were measured, x represents complexes where the number could not be determined. The percentage value above each histogram is the percentage of labelled pre-mRNA molecules (RNA spots) that were associated with mEGFP-SRSF1/mCherryU1A (Coloc. spots). Nuclear extracts contained ATP and an oligonucleotide complementary to U6 snRNA that blocks progression beyond complex A. (A) Binding of mEGFP-SRSF1 to Globin RNA. (B) Binding of mCherry-U1A to Globin RNA. (C) Binding of mEGFP-SRSF1 to Globin RNA in the presence of mCherry-U1A. (D) Binding of mCherry-U1A to Globin RNA in the presence of mEGFP-SRSF1.

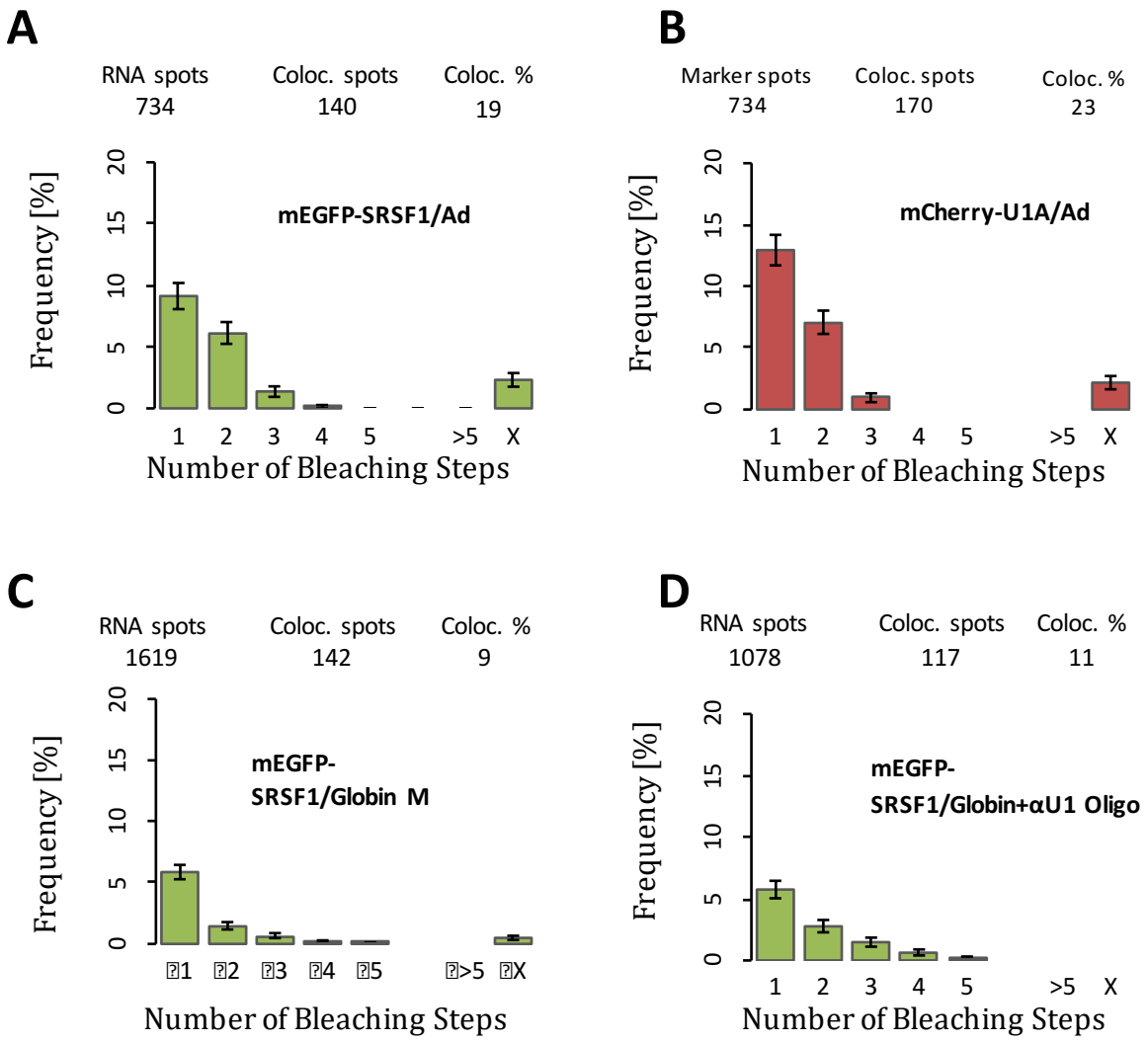


Figure 21. Single molecule studies on the binding pattern of mEGFP-SRSF1 and mCherry-U1A to Globin and Ad RNA. (A) Binding of mEGFP-SRSF1 to Ad (2x5'SS) RNA. (B) Binding of mCherry-U1A to Ad (2x5'SS) RNA. (C) Binding of mEGFP-SRSF1 to Globin M. (D) Binding of mEGFP-SRSF1 to Globin RNA in the presence of an anti U1 Oligonucleotide.

3.4. Does the recruitment of SRSF1 by U1 depend on ESEs or exonic sequences?

In order to determine whether the recruitment of SRSF1 in the cases above is independent of the known ESE driven recruitment of SRSF1 discussed in chapter 4, the BG-SMN2 transcript was used. This transcript on its own gives very low levels of splicing due to the weak SMN 2 exon 7 3' splice site (7%), however the addition of two copies of the Ron enhancer, also used in chapter 4, enhances splicing 5-fold (35%). These RNAs, plus and minus the ESEs, were taken and the binding of SRSF1 analysed, plus and minus the anti-U1 oligo used previously, with the nuclear extract containing just the mEGFP-SRSF1. Figure 22A and B show the binding patterns with and without the ESE; here one can see that the addition of the ESE stimulates the binding of a second SRSF1 molecule. Figure 22C and D show the same RNAs except in the presence of the anti-U1 oligo. Here the pattern in the absence of the ESE shows that there is almost no SRSF1 binding whilst the case with the ESE shows a pattern similar to the one without an ESE. These results show that the ESE-driven recruitment of this additional SRSF1 is not affected by the blocking of U1 binding and the apparent U1-driven recruitment is a separate mechanism.

However it is still possible that SRSF1 may bind to exonic sequences upstream of the 5' splice site and that it then contacts the U1 snRNP for stabilization. To answer this, a Globin 2nt was used. This RNA surprisingly still forms complex A effectively and undergoes at least the first step in splicing; the second step is hard to examine considering that the product would only be four nucleotides long. When this RNA was analysed with the mCherry-U1A and mEGFP-SRSF1 nuclear extract the binding pattern

of U1, Figure 23A, remains unchanged showing that a single molecule of U1 can still bind to the 5' splice site. Similarly figure 23B shows the pattern of SRSF1 remains unchanged. Once again when the three-way co-localization is looked at, figure 23C and D, the patterns remain unchanged and the co-localization is roughly halved, 30%/23%>12%, indicating that the two molecules are co-existing on the same RNAs.

These experiments collectively confirm that the U1 snRNP can recruit SRSF1 to the RNA on its own. This is a separate mechanism from the known ESE-driven method of recruitment and is independent of exonic sequences. This mechanism of recruitment also fits with a raft of the properties that were highlighted in the introduction that suggest SRSF1 has a separate role in splicing reactions from its role in alternative splicing. Here we are providing a mechanism whereby these properties can be undertaken. This also suggests another role for U1 in splicing reactions; positioning SRSF1 at the centre of spliceosome.

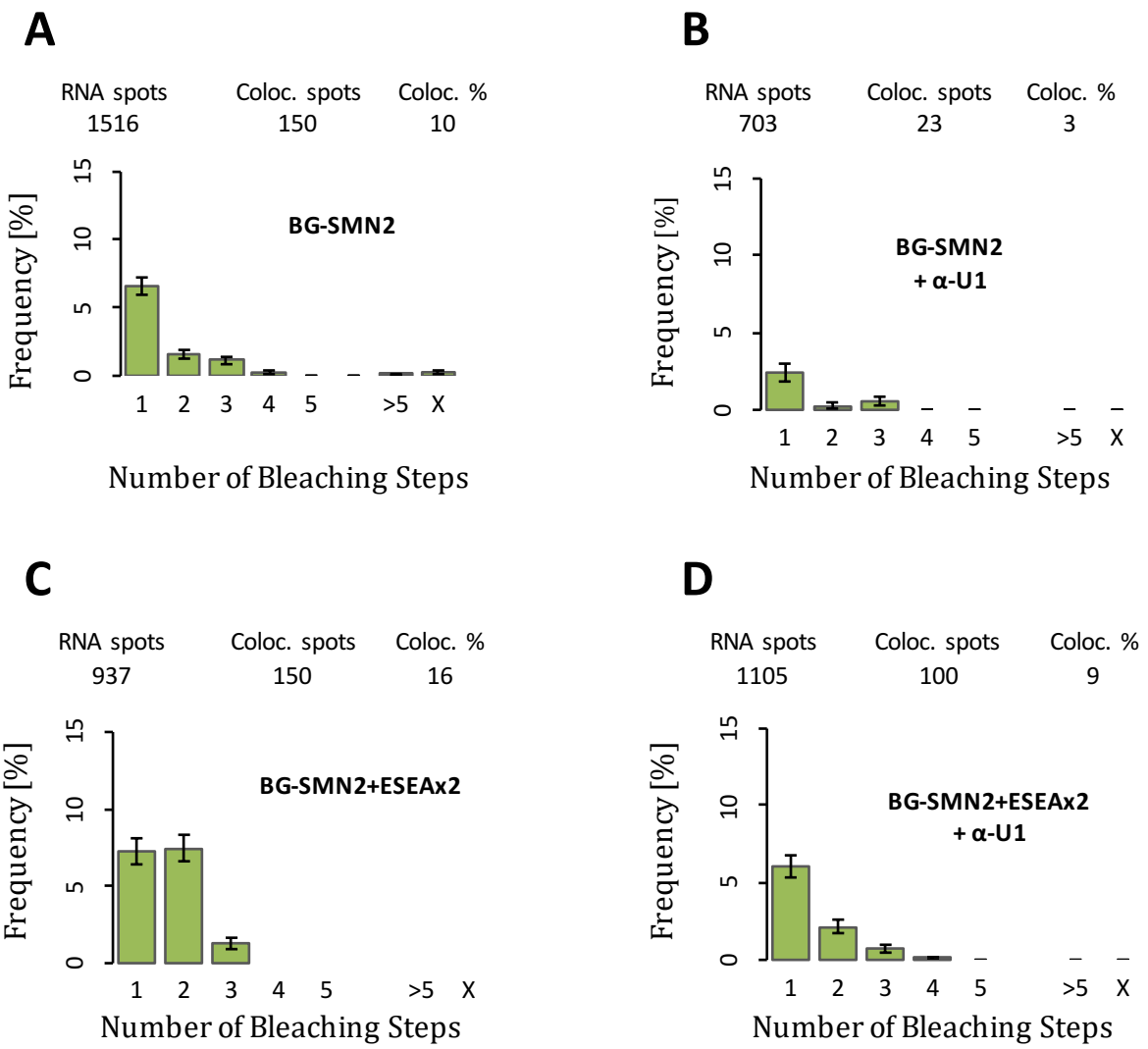


Figure 22. Single molecule studies looking at the binding pattern of mEGFP-SRSF1 to BG-SMN2 hybrid RNA. (A) Binding of mEGFP-SRSF1 to BG-SMN2. (B) Binding of mEGFP-SRSF1 to BG-SMN2 in the presence of the anti U1 oligo. (C) Binding of mEGFP-SRSF1 to BG-SMN2+ESEAx2. (D) Binding of mEGFP-SRSF1 to BG-SMN2+ESEAx2 in the presence of the anti U1 Oligo.

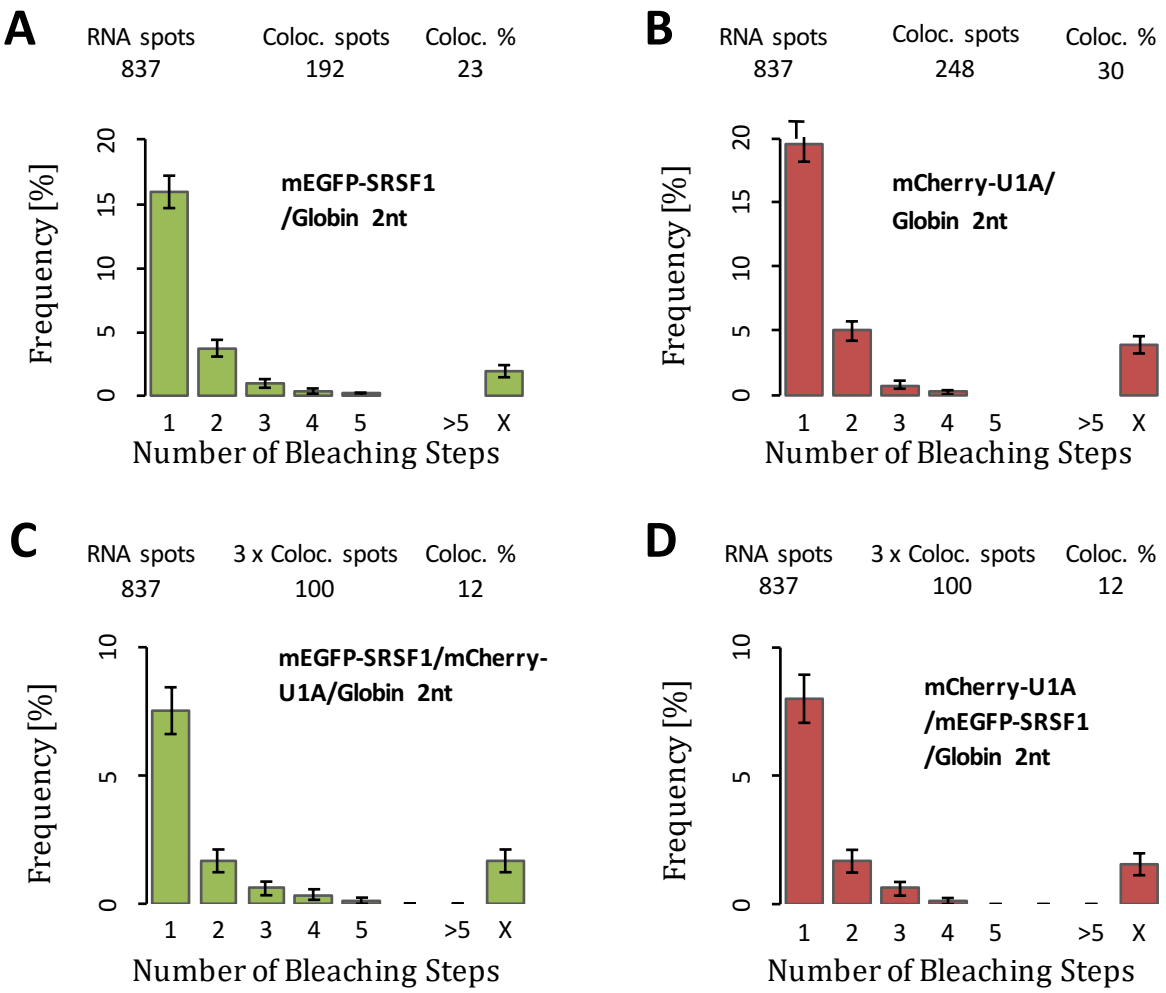


Figure 23. Single molecule studies looking at the binding pattern of mEGFP-SRSF1 and mCherry-U1A to Globin 2nt. (A) Binding of mEGFP-SRSF1 to Globin 2nt. (B) Binding of mCherry-U1A to Globin 2nt. (C) Binding of mEGFP-SRSF1 to Globin 2nt in the presence of mCherry-U1A. (D) Binding of mCherry-U1A to Globin 2nt in the presence of mEGFP-SRSF1.

3.5. Do U1 and SRSF1 bind in an RNA independent manner?

Considering the above observation, that the U1 snRNP recruits SRSF1 to the RNA, it is next worth considering if the two interact prior to their binding to the RNA. This is an interesting proposition because it would suggest that the two may bind together in a complex prior to RNA binding, it would also suggest that the SRSF1 is recruited in a completely RNA-independent manner.

In order to investigate this the nuclear extract containing both labelled factors, mCherry-U1A and mEGFP-SRSF1, was incubated under complex A conditions but without any labelled RNA present. Figure 24A and B show the binding patterns of U1 and SRSF1 when the other component is present. As expected the binding patterns show that they bind in a one to one ratio ($P_{24A=24B}$=0.5). Surprisingly, the co-localization for both is near 25% which given the labelling percentages of each factor suggests that roughly half of the factors are bound to one another at each point in time.

One drawback to this experiment is the possible presence of endogenous RNAs in the nuclear extract which the factors could bind to together. The presence of these RNAs should be minimal as most are estimated to be lost during the nuclear extract preparation. However in order to attempt to remove these the co-localisation experiments were repeated in the presence of ribonuclease A and absence of ribonuclease inhibitor. Interestingly the patterns and co-localization percentage does not change significantly, figure 24C and D. Given that the ribonuclease should digest the U1 snRNA as well and thus cause the snRNP to fall apart, these results would suggest a

strictly protein-protein interaction between the U1A protein and SRSF1. However given that numerous experiments have found that U1A does not interact with SRSF1, it is more likely that the lack of change is due to incomplete digestion of the snRNA and the interaction is still between the snRNP and SRSF1.

The results without ribonuclease do however further the previous interesting findings. These suggest that the SRSF1 and U1 snRNP can interact in an RNA-independent manner. This also suggests that rather than the U1 snRNP binding and then bringing an SRSF1 to the RNA, it may bind the SRSF1 and then bind the RNA with it.

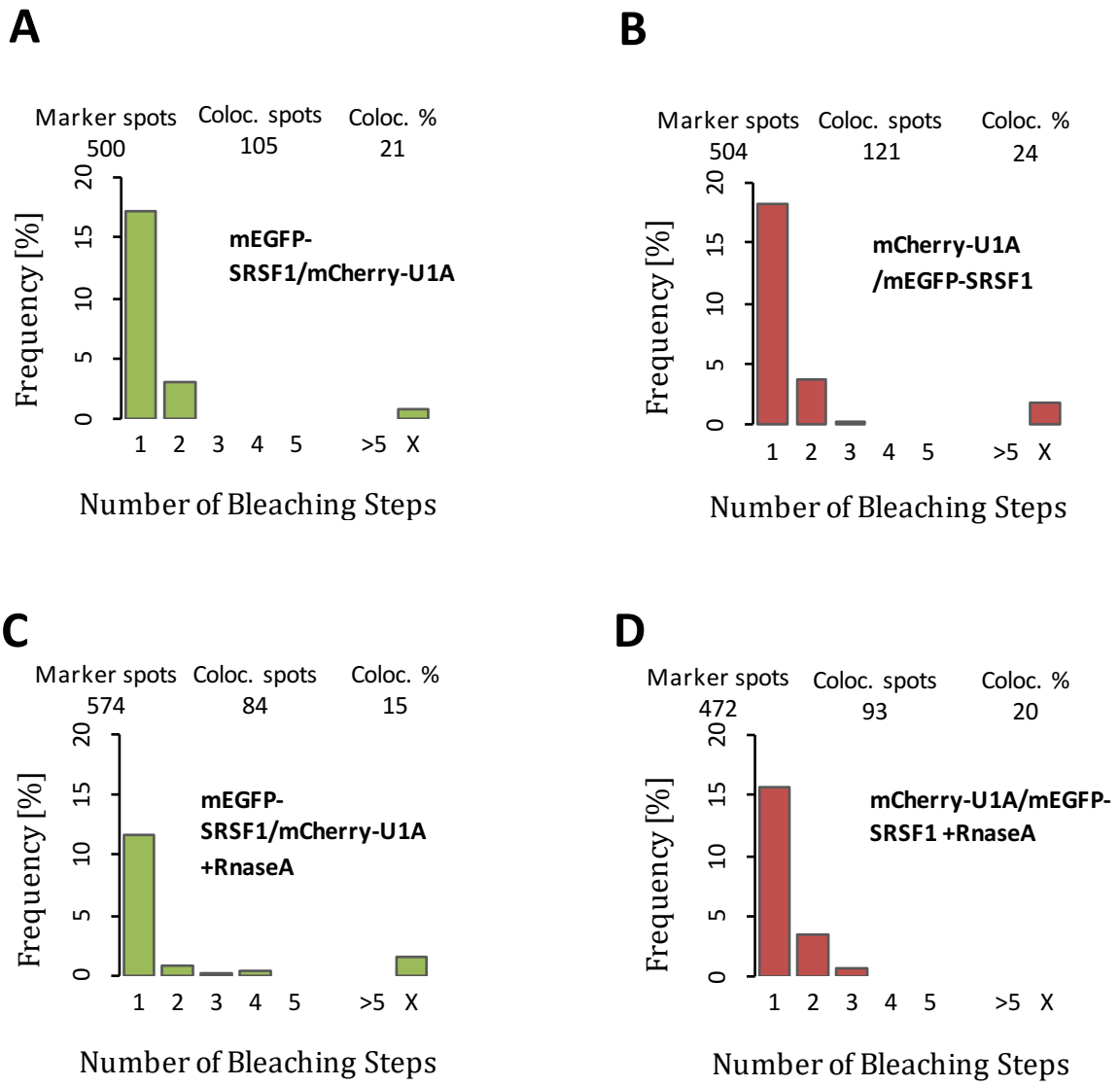


Figure 24. Single molecule studies showing the binding pattern of mEGFP-SRSF1 to mCherry-U1A in the absence of any RNA. (A) Binding pattern of mEGFP-SRSF1 to mCherry-U1A. (B) Binding pattern of mCherry-U1A to mEGFP-SRSF1. (C) Binding pattern of mEGFP-SRSF1 to mCherry-U1A in the presence of Ribonuclease A. (D) Binding pattern of mCherry-U1A to mEGFP-SRSF1 in the presence of Ribonuclease A.

3.6. A 3' U1 binding site stimulates splicing and recruits SRSF1.

Aside from the possible role that SRSF1 may play in core splicing reactions, the recruitment by U1 also raises another interesting prospect. Whilst it is widely known and accepted that a strong downstream 5' splice site stimulates the use of the 3' splice site immediately upstream of it (Hwang & Cohen 1996), the mechanism for how this happens is not clear. Despite a number of cross-intron interactions between the U1 snRNP and 3' splice site being known (Shao et al. 2012), SF3A, PRPF40A etc., none of these have been shown to function across exons. One possibility is that there is an ESE in the centre of every exon and that this recruits an SR protein which in turn can interact with both splice sites and bridge the gap. However, this is clearly either not the case or the ESE is not sufficient when splicing has been shown to depend on a downstream 5' splice site (Hwang & Cohen 1996).

Given that we have shown the U1 snRNP can recruit SRSF1 on its own to 5' splice sites, one possible purpose of this interaction could be to answer this question. If the U1 snRNP can recruit SRSF1 to a downstream 5' splice site, this SRSF1 could then serve to stabilise the binding of crucial factors to the 3' splice site.

In order to test this, Globin+U1 was used. The presence of this additional sequence increases the splicing (83%>89%) and complex A formation of the construct, although both were considerably high regardless. The binding of U1 and SRSF1 were then analysed for this transcript. Figure 25A shows that, as expected with two strong 5' splice sites present, one or two molecules of mCherry-U1A are bound. Once again Figure 25B

shows that with an extra 5’ splice site an additional mEGFP-SRSF1 is recruited. This shows that the U1 snRNP can recruit SRSF1 to the RNA regardless of its location and supports the idea that U1 could recruit SRSF1 downstream of 3’ splice sites and thus act similarly to an ESE.

The Globin transcript, however, splices particularly well regardless of the presence of the downstream 5’ splice site, Globin+U1, as it has a strong 3’ splice site. Thus, this doesn’t necessarily make it a particularly good model as the 3’ splice site doesn’t need exon definition to function. In order to test that this system functions even with a weak 3’ splice site, the previously used BG-SMN2 hybrid transcript was used. Once again a consensus 5’ splice site was added to the 3’ end of the second exon. Again this strongly stimulated splicing, but this time the effect was far more significant, splicing going from 7% to 55%; this is an increase which is comparable to the effect seen with four additional ESEs (Chapter 4). When the binding was analysed, Figure 25A and B, there was an increase in the number of molecules bleaching in two steps for both mCherry-U1A and mEGFP-SRSF1. This additional recruitment of SRSF1 in conjunction with the substantial increase in splicing strongly supports that the downstream 5’ splice site acts as an ESE for a weak upstream 3’ splice site by recruiting SRSF1. The substantial increase in splicing despite the relatively short sequence added (10nt), compared to the four extra ESEs (48nt), indicates a very stable mechanism for recruitment. This is the opposite of what is seen with ESEs in chapter 4, once again reinforcing that this is a distinct mechanism of recruitment.

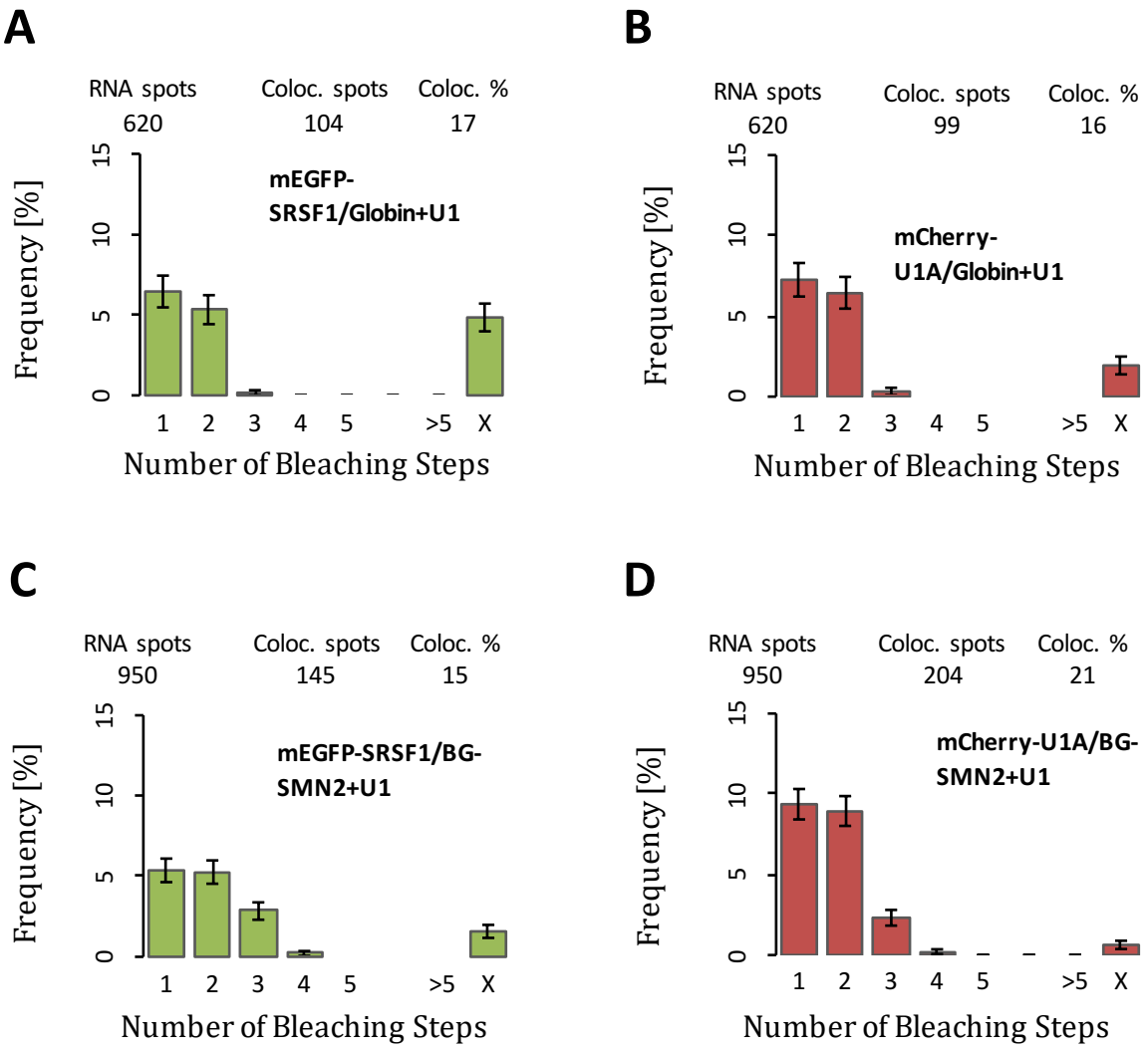


Figure 25. Single molecule studies showing the binding of mEGFP-SRSF1 and mCherry-U1A to Globin or BG-SMN2 hybrid RNA. (A) Binding pattern of mEGFP-SRSF1 to Globin+U1. (B) Binding pattern of mCherry-U1A to Globin+U1. (C) Binding pattern of mEGFP-SRSF1 to BG-SMN2+U1. (D) Binding pattern of mCherry-U1A to BG-SMN2+U1.

3.7. A 3' U1 binding site that recruits SRSF1 stabilises the binding of key factors to the 3' Splice site.

The previous evidence strongly suggests that a U1 snRNP recruited to a 5' splice site can recruit SRSF1. Furthermore this is shown to occur at sites both at the 5' end of an intron and at the 3' end of an exon which stimulates the splicing of the upstream intron. However to fully verify that the recruitment of this SRSF1 by the U1 snRNP is playing a role in an exon definition like system, we need to analyse the binding of crucial factors to the 3' splice site.

The classical model for exon definition as described in the introduction is that an SRSF1 is recruited to the centre of an exon and this then contacts both the U1 snRNP and the 3' splice site complex (De Conti et al. 2013). Our data has shown that the SRSF1 is not in fact recruited to the exon itself but is recruited by the U1 snRNP. However the final piece of the puzzle is to show that this subsequently exerts an effect at the 3' splice site. The potent effect seen on splicing is good indirect evidence for this but with our single molecule techniques we have a way of assessing it directly.

The nuclear extracts used here have been validated previously by Chen et al (Chen et al. 2016). The constructs are BG-SMN2 RNA and BG-SMN2+U1. Figure 26B and E shows us the recruitment of U2AF 35 to these constructs, here it can be seen that the co-localisation percentage nearly doubles with the addition of the U1 binding site, 12%-20%, indicating a strong up-surge in recruitment. The recruitment of U2AF 65, as seen in figure 26C and F, also shows a strong increase, 10%>14%, but not to the level seen

with U2AF 35. Finally figure 26A and D shows the recruitment of U2B''. Here, we again see an increase in recruitment, 10%>27%.

These results show that the SRSF1 that is recruited, by the 3' bound U1 snRNP, may then go on to stabilise the 3' splice site complex either via a U2 snRNP component or through direct binding to U2AF35, which in turn stabilises the other factors. Furthermore this evidence in conjunction with the previous section illustrates that if a downstream 5' splice site is strong then it can serve to stimulate exon definition without any intervening exonic sequences.

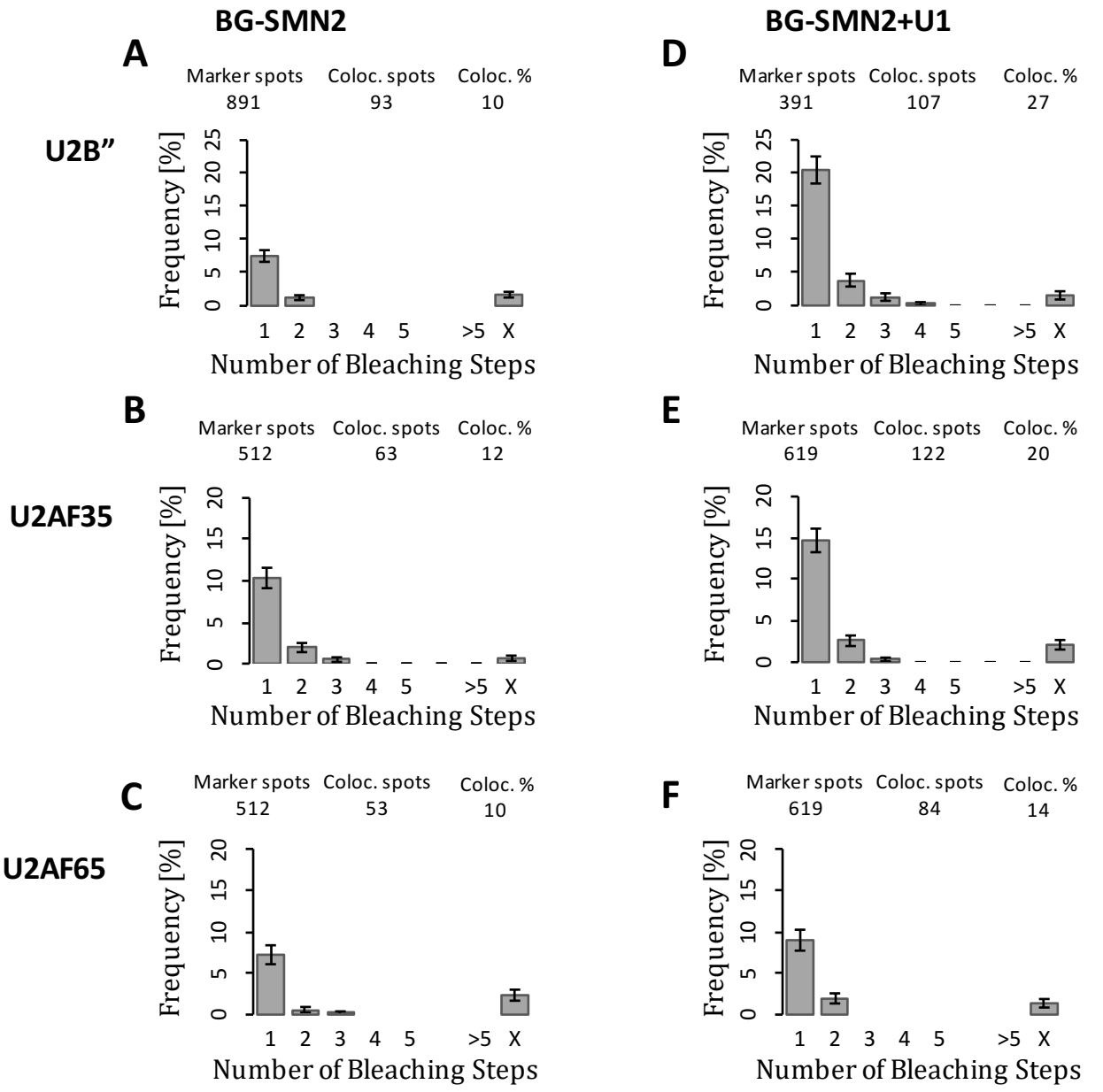


Figure 26. Single molecule experiments showing the recruitment of the key 3' splice site factors in the presence and absence of a 3’ 5’ splice site. (A-C) BG-SMN2 RNA with mEGFP-U2B'' (A), mCherry-U2AF35 (B), mEGFP-U2AF65(C). (D-F) BG-SMN2+U1 with mEGFP-U2B'' (D), mCherry-U2AF35 (E), mEGFP-U2AF65 (F).

3.8. Potential role of DDX5 at 5' splice sites.

Having analysed the recruitment of SRSF1 to the pre-mRNA via single molecule studies and identified a novel recruitment mechanism, the binding of another poorly characterised splicing factor was looked at. DDX5 is a DEAD-box helicase that is thought to play a role in unwinding/destabilising secondary structures around 5' splices sites. However whilst in a number of cases the recruitment of DDX5 is thought to occur via other RNA binding proteins such as RBM5, in many cases there are no known sites for these co-factors.

In order to look at DDX5s recruitment to Globin, a nuclear extract containing mCherry-labelled DDx5 was used. Figure 27A shows that the majority of complexes contain a single molecule of DDX5. However, there are also a notable number of complexes with two molecules, which can be explained by the dimerization of DDX5 (Figure 27E). The number of DDX5 molecules that associate with one another in the absence of any RNA is much the same as in the case where the protein was co-localised with the RNA. This is perhaps not surprising given prior work identifying DDX5s activity in a homodimer or a heterodimer with DDX17 (F. V. Fuller-Pace & Ali 2008).

The recruitment of a single molecule or a pair of dimerised molecules of DDX5 to the RNA is perhaps what one would expect. In order to look in more detail at the recruitment mechanism for this molecule the binding was analysed in the presence of the anti-U1 oligonucleotide which inhibits U1 snRNP binding and in turn, as shown previously, prevents stoichiometric SRSF1 recruitment. When one looks at the DDX5 binding pattern, Figure 27B, we can see that the pattern shifts from a defined one or two molecules to a more non-specific geometric pattern with up to 5 molecules being

recruited. This shift is strikingly similar to the change seen for SRSF1 in the presence of the anti-U1 oligo.

In order to look at whether DDX5 was being recruited to the RNA via its helicase domain, a mutant version of DDX5 with no helicase activity was tested; it is termed a NEAD mutant as it has had its catalytic aspartic acid switched to an asparagine (from here on this mutant will be referred to as DDX5-NEAD). The nuclear extract containing the DDX5-NEAD was used with Globin and the binding pattern is seen in figure 27D. Here one can see that the binding pattern is nearly identical to the non-mutant version, seemingly indicating that its recruitment has nothing to do with its helicase domain. However in the case of the DDX5-NEAD, when the anti-U1 oligo was used we see no shift to a non-specific geometric pattern.

The data looking at DDX5 binding with and without the anti-U1 oligo provides us with some interesting insights. The shift from a single/dimerised molecule to multiple binding as seen with SRSF1, seems to indicate that the DDX5 binds in a 1:1 stoichiometry with SRSF1. This could possibly be with another SR protein that is also recruited stoichiometrically by U1 snRNPs that we have not analysed yet and which would also show geometric binding in the absence of U1 binding. However, the work using the DDX5-NEAD mutant does not fit with this model. The mutation should disrupt RNA binding as opposed to any protein-protein interactions that might occurring with an SR protein so one would expect to see the same shift towards a geometric pattern as with the native protein. Further work is needed to really elucidate the mechanism for DDX5 recruitment to 5' splice sites and what the effect of the D>N mutation has on recruitment. Despite this, the shift seen with the native pattern in the presence of the

anti-U1 oligo does seem to suggest that U1 may be at the heart of co-ordinating a number of factors during the early stages of splicing, either directly or indirectly. Furthermore it raises the possibility that SRSF1 may be able to recruit DDX5 to 5' splice sites as RBM5 does.

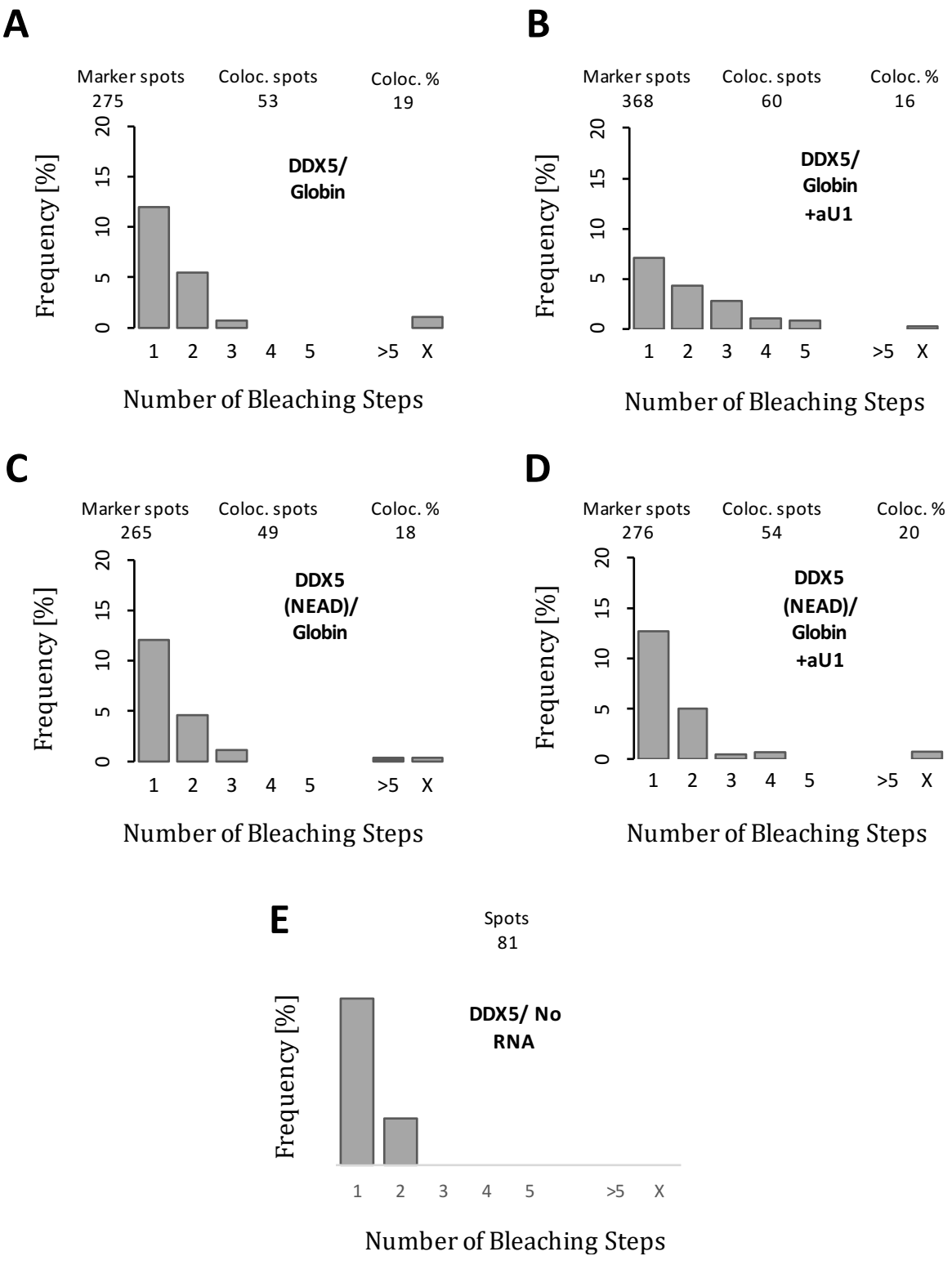


Figure 27. Single molecule studies showing the binding pattern of DDX5 and mutant DDX5 to Globin RNA and in the absence of RNA. (A) Binding of mCherry-DDX5 to Globin. (B) Binding of mCherry-DDX5 to Globin in the presence of an anti U1 oligo. (C) Binding of mCherry-DDX5 (NEAD) to Globin. (D) Binding of mCherry-DDX5 (NEAD) to Globin in the presence of an anti U1 Oligo. (E) Dimerisation of mCherry-DDX5 in the absence of any RNA.

3.9. Summary.

The conventional model for the recruitment and action of SRSF1 in pre-mRNA splicing involves the binding of SRSF1 to ESEs and the subsequent formation of interactions with U1 snRNPs at 5′ splice sites and U2-asociated factors at 3′ splice sites. This is consistent with extensive data from transcriptome-wide analyses of binding sites by CLIP, which have revealed that SRSF1 binding sites are particularly enriched upstream of 5′ splice sites and, in some cases much more strongly, at the 5′ end of exons just downstream of the 3′ splice sites (Pandit et al. 2013b). Here we show that there is another completely different mode of recruitment for SRSF1 which may indicate new potential roles for SRSF1 and U1 snRNPs.

By using single molecule methods that distinguish between complexes containing different numbers of molecules of SRSF1, we have been able to show that a single molecule of SRSF1 is recruited to splicing-competent pre-mRNA in a process that is independent of exons but dependent on U1 snRNPs bound to 5′ splice sites. This occurs whether the 5′ splice site is at the 5′ end of a spliced intron, one of two alternative sites or at the 3′ end of an exon which is involved in stimulating the splicing of the upstream intron. Strikingly, even in the absence of any RNA significant proportions of SRSF1 molecules and U1 snRNPs are associated with each other. From this one can infer that the binding of a U1 snRNP to a 5′ splice site concomitantly recruits a molecule of SRSF1. This may have been missed by conventional cross-linking assays or transcriptome-wide CLIP analyses of SRSF1 if the protein does not directly contact the pre-mRNA or does so via surfaces that do not contain amino acids that cross-link efficiently.

An additional and less distinct form of binding produces the significant background of molecules bound by more than one or two molecules of SRSF1, as seen in Chapter 4 as well. When U1 snRNA was sequestered, the distribution was geometric and thus might be accounted for by stochastic interactions. However, if the U1-dependent association involved binding of an additional molecule of SRSF1 to RNA molecules that already had SRSF1 bound via this background binding, then one would expect the patterns seen to reflect a shift to complexes with more molecules bound. In contrast to this the binding of the U1 snRNP in fact restricts the number of complexes with higher numbers of bleaching steps and stimulates those with one bleaching step per site. Hence, it is likely that these distributions arise from two mutually exclusive processes; specific recruitment of SRSF1 molecules via U1 snRNPs or ESEs vs the association of background molecules to pre-mRNAs that are not destined to form functional splicing complexes.

The positioning of SRSF1 in a stoichiometric manner by the U1 snRNP at 5' splice sites has interesting ramifications for exon definition systems and splicing of downstream introns. The lack of a known bridge over exons has been a major stumbling block for the exon definition models and has resulted in the assumption that there are ESEs in the centre of nearly all exons. The recruitment of a molecule of SRSF1 by a U1 snRNP provides a possible bridge for how the two ends of an exon could interact. This, in conjunction with the data displaying the improved recruitment of crucial 3' splice site factors when a downstream 5' splice site is present, provides a possible route for how a 5' splice site can act in a similar manner to an ESE.

The potential role for SRSF1 in the splicing of a downstream intron is perhaps more unforeseen but not without supporting evidence. As described in detail in the

introduction and interrogated in the discussion, there are a number of properties of both U1 snRNPs and SRSF1 which suggest both have additional roles in splicing above and beyond what is commonly accepted. Is one of the roles of U1 snRNPs to position SRSF1 at the heart of the spliceosome so that it can play a role later in splicing? Or is the recruitment of SRSF1 by U1 snRNPs strictly for the enhancement of exon definition and the recruitment in the absence of an upstream 3' splice site simply an artefact?

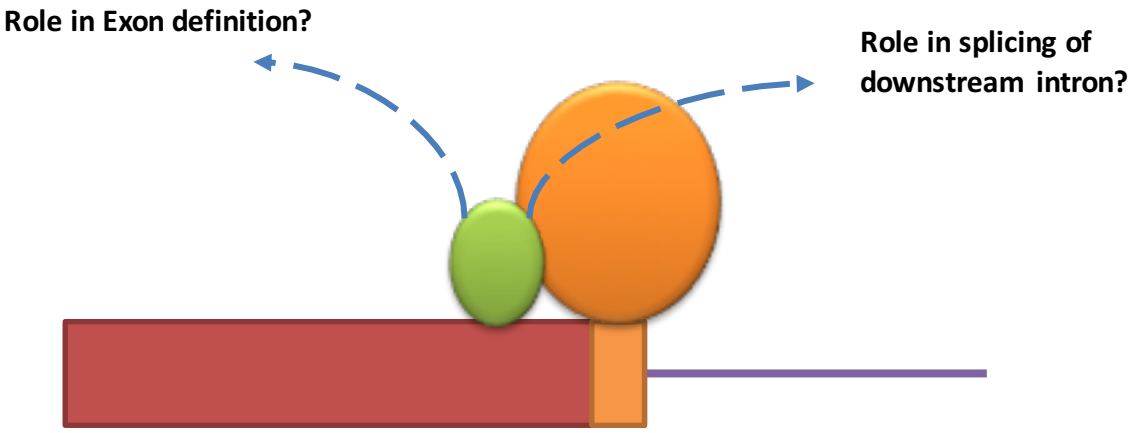


Figure 28. Model showing the recruitment of SRSF1 to the 5' splice site by the U1 snRNP and its possible subsequent functions.

Chapter 4. The connection between ESE activity and SRSF1 binding.

4.1. Introduction.

Whilst chapter 3 interrogates an alternative recruitment mechanism for SRSF1 via U1 snRNPs, this chapter focuses on the established ESE-driven mechanism. In chapter 3 data looking at RNAs with ESEs showed that the two recruitment mechanisms were distinct. Models of the mechanisms of action of ESEs are generally resolved into two steps, either of which might be limiting or regulated. The first is binding to the ESE (this chapter) whilst the second is the step in which the ESE-bound SR protein activates its target splice site (chapter 5).

The first part of this model, the binding of the SR protein to the RNA, was tested indirectly in an important paper in 1998 (Hertel & Maniatis 1998). Here Hertel and Maniatis followed the rates of splicing dependent on the *dsx*REs and found the proportion of *dsx* pre-mRNA that spliced in vitro increased with the addition of the *Drosophila* SR proteins Tra and Tra2, but the level of splicing approached a maximum that was dependent on the number of ESEs and not protein concentration. It was assumed that the *dsx*REs were concurrently occupied and the limiting step was strictly the interaction between the bound proteins and their target. This paradigm has since been accepted as the mechanism for how ESEs function in all metazoans, as shown by the lack of further study in this area since then. However this data whilst appearing to fit with a number of observations for the function of *Drosophila* proteins and enhancers, does not fit with an array of data concerning human SR proteins and enhancers.

Principally, whilst in 1998 the number of known enhancer sequences was relatively small, with ESEs only being discovered a few years earlier (Sun et al. 1993), the number of ESEs now known to function in splicing is vast. This expansion of sequences, simply based on its massive nature, opens the possibility that not all sequences may function in the same manner. How a sequence will act is likely to depend on the proteins that it recruits and the nature of this proteins interaction with RNA. Whilst the proteins, Tra and Tra2 studied previously had been shown in *Drosophila* to bind stably to RNA in a enhancer complex (Tian & Maniatis 1993), the Tra protein is not found in humans and a number of human enhancer proteins have been shown to bind weakly to RNA (Cho et al. 2011; Anczuków et al. 2015). Chief amongst these is SRSF1. SRSF1 was the first SR protein to be discovered and has all the properties of a standard SR protein as well as a range of additional ones, as discussed in chapter 3.

SRSF1, due to its significance in the splicing of a vast range of exons, has provided the paradigm for the study of human SR proteins and ESEs in turn. However in contrast to the *Drosophila* Tra/Ta2 proteins, SRSF1 is regularly found to bind to RNA weakly and in a somewhat transient manner with Kds ranging from 0.2 µM to above 1 µM (Cho et al. 2011; Anczuków et al. 2015). Furthermore it is often found to be a minor component of protein mixes that bind to ESE sequences despite these sequences being shown to depend on SRSF1 for their functionality (as highlighted in chapter 5) (Smith et al. 2014; Ghigna et al. 2005).

Due to the vast range of ESE sequences now discovered in humans along with the clear difference in binding mechanics between SRSF1 and Tra/Tra2, we believe this field is worth further study. Single molecule multi-colour co-localization methods have been

used extensively to determine directly the numbers of protein molecules bound to a molecule of pre-mRNA in mammalian and yeast nuclear extracts and to follow the dynamics of association and dissociation of splicing factors; providing mechanistic evidence that would otherwise be unattainable (Cherny et al. 2010; Larson & Hoskins 2017; Hoskins & Moore 2012). The use of these techniques will allow fresh light to be shed upon the problems outlined above due to its innate ability to observe the occupancy of ESEs. In this section we analyse the binding of SRSF1 to multiple different ESEs and multiples of the most potent one in order to look at the relationship between SRSF1 binding and an ESE's apparent effect, demonstrated by experiments looking at the complex formation and splicing efficiency. This is supported by experiments looking at the binding of the key 3' splice site factors, U2AF 35 and 65 and the U2 snRNP (represented by U2B''), in order to validate the action of said ESEs. Furthermore the binding of Tra2β is also analysed to show that it is SRSF1 that is crucial for the effects seen, not just any SR protein.

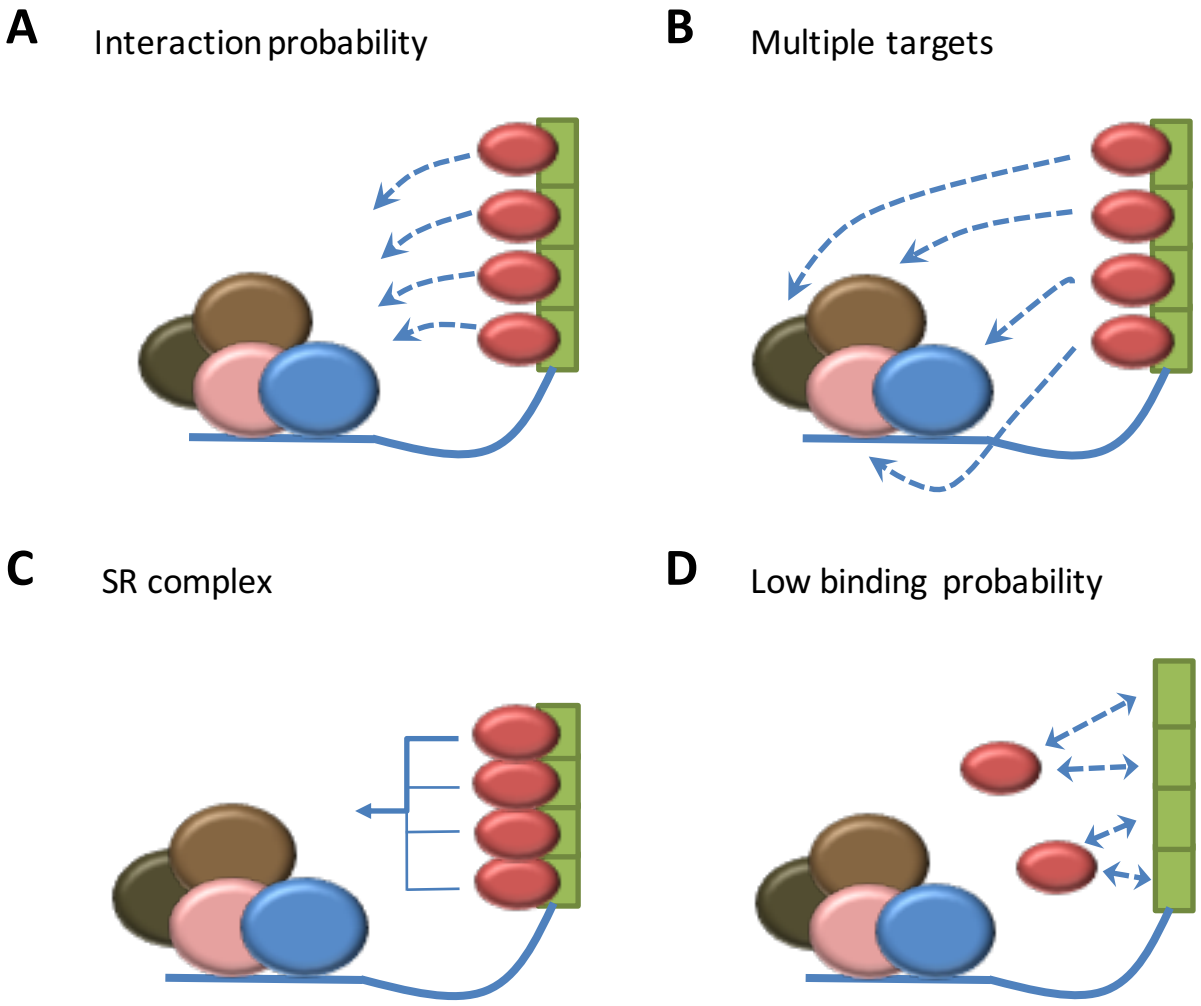


Figure 29. Models representing possible SR protein binding steps by which ESEs stimulate 3' splice site usage. (A) SR proteins bind stably, with multiple occupancy of repeated ESEs, and the probability of subsequent interactions. (B) SR proteins bind stably, with multiple occupancy of tandemly-repeated ESEs, and are able to interact with multiple targets, each of which has an approximately equal effect on the rate. (C) SR protein binding is cooperative. (D) Protein binding is transient and increased ESEs increases the chance of binding.

4.2. Validation and analysis of Tra2β and SRSF1/Tra2β nuclear extracts.

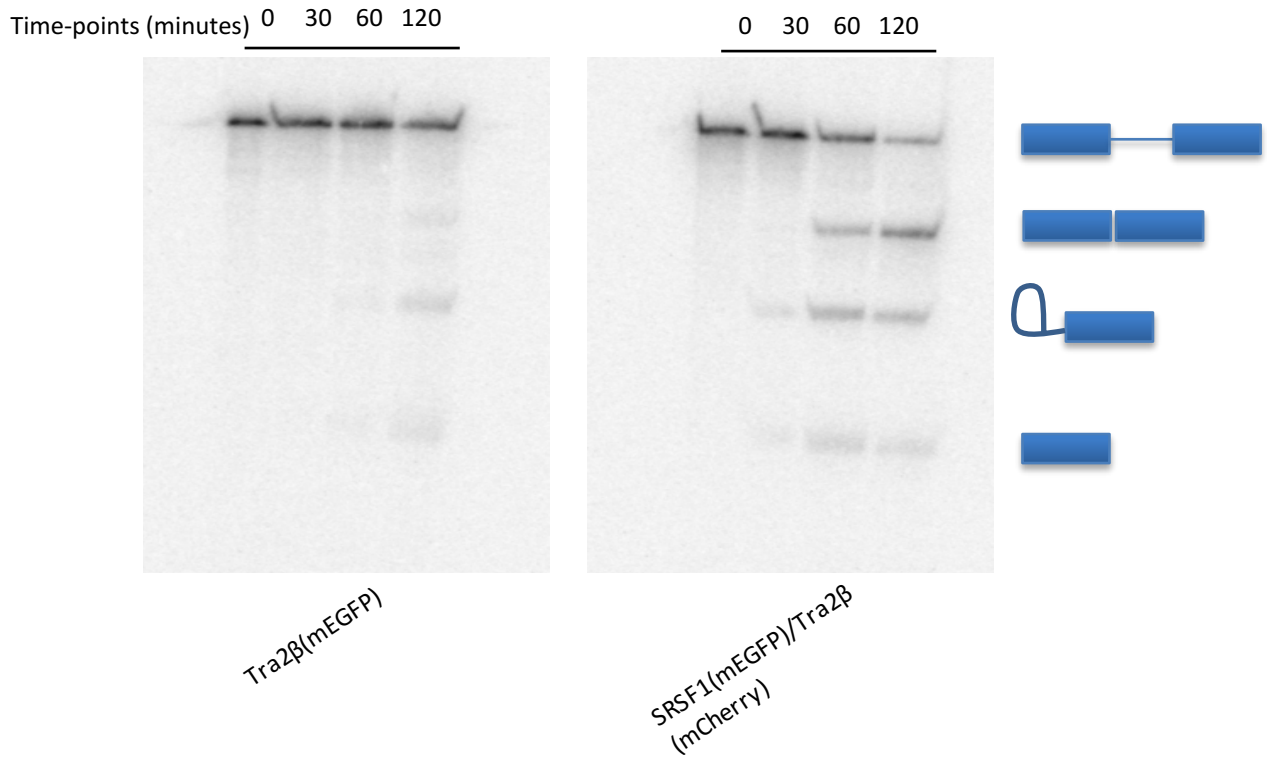
In order for the system that we use to be entirely valid a number of preliminary experiments need to be undertaken; the nuclear extract preparations that we make and use need to be shown to be competent in splicing assays and the amount of labelled protein vs the endogenous protein needs to be checked.

In order to analyse the nuclear extracts' abilities to perform splicing reactions, the Globin construct is used. Figure 30A and B show the levels of splicing for two of the four extracts used in this section compared with a commercial nuclear extract; the U2B'' and U2AF35/65 nuclear extracts have been used and validated previously (Chen et al. 2016) and the U1A/SRSF1 nuclear extract was used in chapter 3. Whilst both show splicing, the Tra2β-GFP containing nuclear extract prepared from 293T cells shows a significantly lower level (figure 30C). This may possibly be due to the exceptionally high level of labelled protein, as detailed below.

To be able to analyse the experiments correctly and model the number of proteins that are binding to our RNA then the level of labelled protein vs the level of endogenous protein needs to be analysed. Western blots with the appropriate antibody for the protein being analysed are used. In the cases where there are two labelled proteins transfected, two separate western blots were done. As stated earlier it can be seen that in the case of the Tra2β nuclear extract there is an exceptionally high level of labelling with almost no endogenous protein being identified. The levels of labelling are shown in the table in figure 31D.

A

B

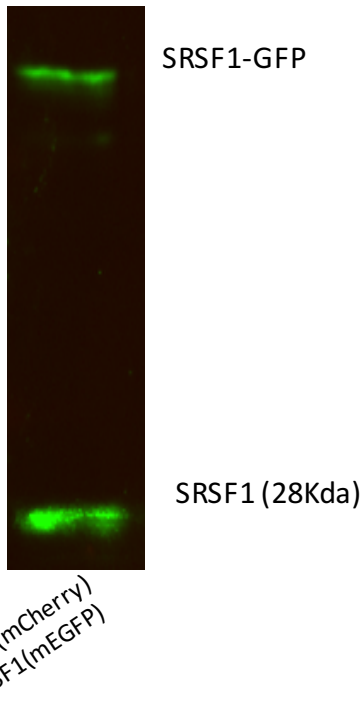


C

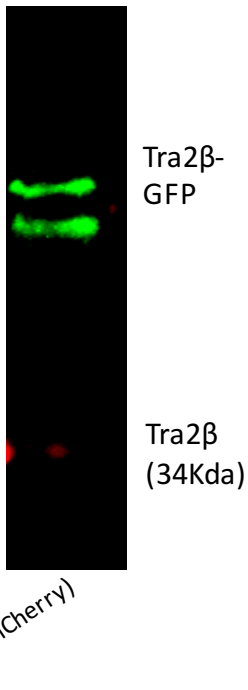
NE	Tra2β(mEGFP)	SRSF1(mEGFP)/Tra2β (mCherry)
Splicing efficiency after 2 hours (mRNA/(mRNA+pre-mRNA)	6	59

Figure 30. Polyacrylamide gel showing the level of splicing observed with Globin RNA in nuclear extracts containing labelled proteins. (A) Splicing time course of Globin RNA with 50% mEGFP-Tra2β transfected nuclear extract. (B) Splicing time course of Globin RNA with 50% mEGFP-SRSF1/mCherry-Tra2β transfected nuclear extract. (C) Quantification of the level of splicing after two hours for each extract.

A



C



D

NE	Tra2β(mCherry)/SRSF1(mEGFP) (Anti-SRSF1)	Tra2β(mCherry)/SRSF1(mEGFP) (Anti-Tra2β)	Tra2β (Anti Tra2β)
%labelled	37		100

Figure 31. Western blots showing the level of labelled protein present in the transfected nuclear extracts vs the level of endogenous un-labelled protein. (A) Anti SRSF1 antibodies used against the nuclear extract containing mEGFP-SRSF1 and mCherry-Tra2β. (C) Anti Tra2β antibodies used against the nuclear extract containing mEGFP-Tra2 β. (D) Quantification of the level of labelled protein present in each extract compared to the endogenous.

4.3. The dependence of splicing on ESE sequences is linked to the level of ESE-dependent binding by SRSF1.

Whilst it is known that purine-rich sequences within exons stimulate the use of the exons and this occurs through the recruitment of SR proteins such as SRSF1, it has proved difficult to show a direct link between increased recruitment and an increased enhancing effect. Being able to quantitatively link increased recruitment to ESEs with an increased effect has most likely proven difficult due to the effect discussed in chapter 5; the recruitment of SRSF1 by U1. This additional mechanism of recruitment means that all splicing competent constructs would recruit SRSF1 regardless of the presence of an ESE.

Single molecule methods enable the number of molecules of SRSF1 in complexes to be determined, as opposed to the proportion of RNA molecules bound. In order to do this, the chimeric RNA construct, BG-SMN2, also used in chapter 3, was used. This RNA on its own gives very limited levels of splicing (~2%). However the addition of an extra enhancer sequence increases the level of splicing significantly but to varying levels depending on the sequence.

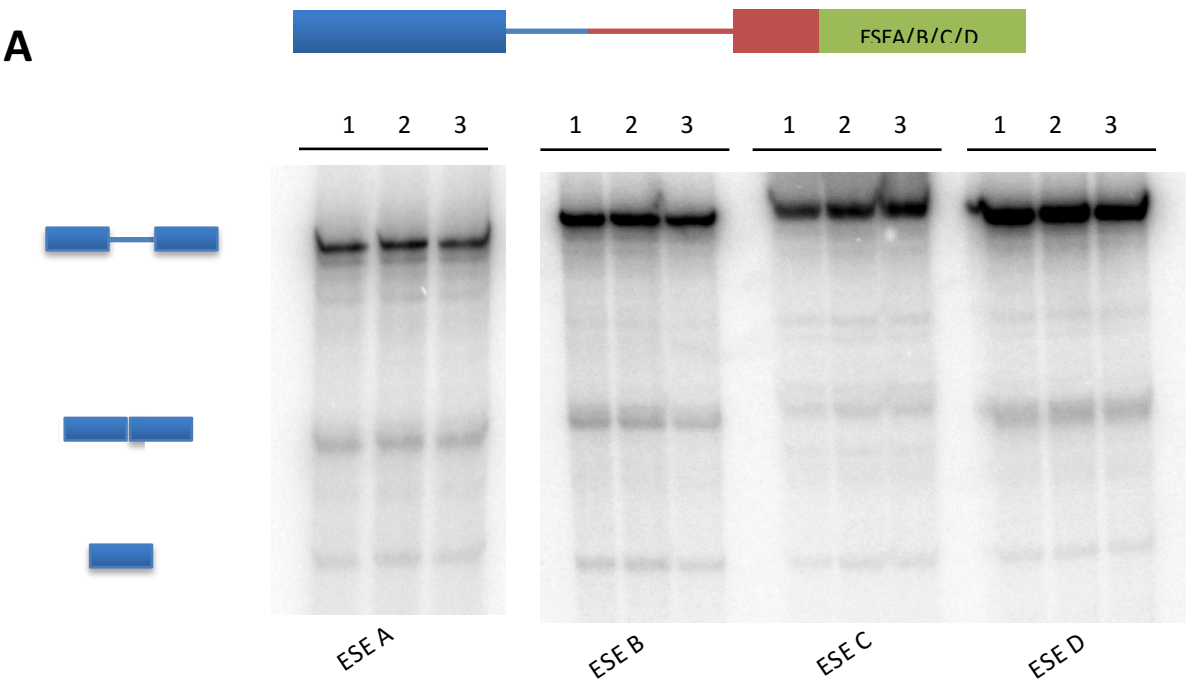
To test whether there was a correlation between the efficacy of an ESE and binding of SRSF1, four different sequences were tested: ESE A (BG-SMN2+ESEA)- the known SRSF1 binding site from Ron exon 12 (Ghigna et al. 2005; S. Cho et al. 2011), ESE B (BG-SMN2+ESEB)– an SRSF1 binding motif from functional SELEX (Sheth et al. 2006), ESE C (BG-SMN2+ESEC)- the known Tra2B binding site from SMN exon 7, which has been shown to promote splicing but should not bind SRSF1 (Hofmann et al. 2000; Cartegni & Krainer 2003) and ESE D (BG-SMN2+ESED)- an SRSF1 motif established via RNA-seq

(Anczuków et al. 2015) (Sequences in the appendices). ESEs B-D were incorporated as double copies so as to more closely match ESE As length. Each ESE was attached to the end of BG-SMN2 (figure 32A). All four sequences were tested in in vitro splicing reactions and all four enhanced splicing significantly ($P_{ESEA}=3x10^{9}$, $P_{ESEB}=0.001$, $P_{ESEC}=0.01$, $P_{ESED}=0.0002$). However the ESE derived from Ron exon 12, ESE A, was by far the most effective (Figure 32B).

Each transcript was then labelled and analysed in the nuclear extract containing mEGFP-SRSF1 and mCherry-U1A, under conditions permitting complex A formation. Each experiment was repeated on three separate days to ensure the co-localisation percentage observed was reproducible. The distributions displayed are an accumulation of the spots from each of the three separate experiments.

In the absence of an ESE, there was a strong peak of complexes containing a single molecule of SRSF1, followed by a non-specific geometric distribution from two onwards ($P_{Geo\ (n2-5)}=0.15$) (Figure 33A). We showed in chapter 3 that the U1 snRNP can recruit a single SRSF1 to a 5'SS in complex A. Previous data also showed that failure to form complex A leads to a non-specific geometric distribution with up to 5 molecules of SRSF1 bound, resembling that in figure 33A. The distribution of complexes in figure 33A therefore probably represents a mixture of complex A with a single U1-bound SRSF1 and complexes that, due to the weak 3'SS, have not formed complex A. The ratio of two steps to one step is 37% for the construct with no ESE and this increases to 58% for ESE-A and 40-43% for the other ESEs; showing their effect on recruiting an additional SRSF1. The results of adding an ESE, figure 33B-E, show that there is a link between the efficacy of the ESEs, the recruitment of a second molecule of SRSF1, and the loss of the

background geometric distribution representing non-specific complexes. This is the first time that a quantitative link has been demonstrated between an ESE's effect on splicing and its recruitment of SRSF1. The presence of additional molecules of SRSF1, from the 5'SS and the non-specific background, on a valid splicing substrate would contribute to the difficulties encountered in showing ESE-dependent SRSF1 recruitment using ensemble methods. Furthermore, the results suggest that SRSF1 is, at least in this case, a limiting factor in an ESE-dependent system.



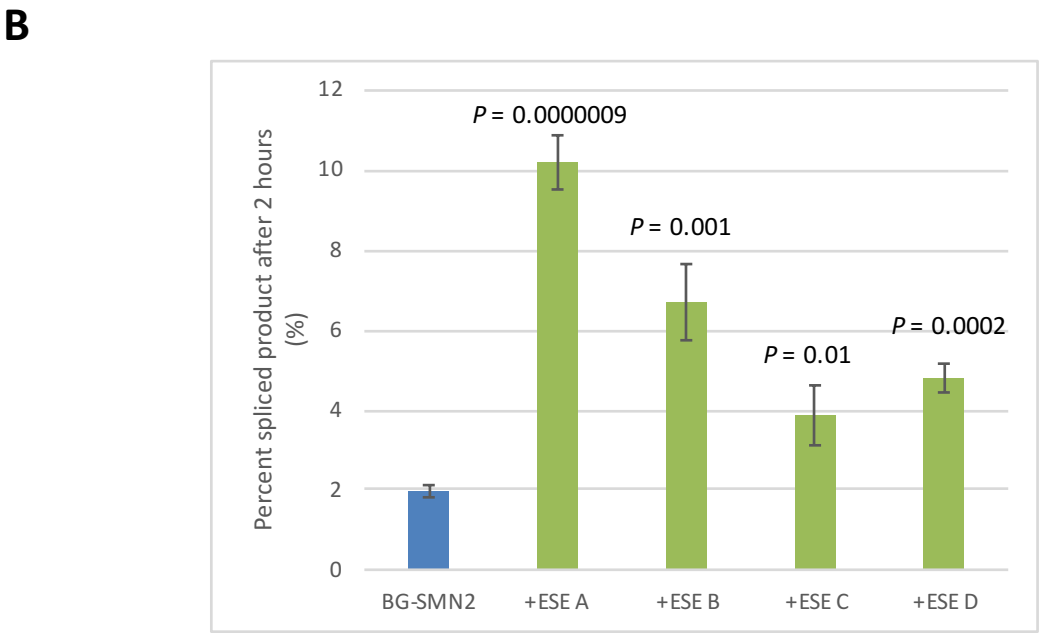


Figure 32. Effects of different ESE sequences on splicing and the association of SRSF1. (A) Splicing activity *in vitro* of pre-mRNA with different 3'ESEs. The pre-mRNA is represented by a diagram showing the sequences from Beta-Globin exon 2 (blue), SMN2 exon 7 (red) and the ESE under test (green). (B) The level of spliced mRNA (%, [mRNA]/[mRNA + pre-mRNA]) after incubation for two hours was calculated for each of three reactions done in triplicate (5 reactions for ESE A as one of the triplicates gave an anomalous results), and the mean and standard deviation are shown. The probabilities (*P*) that results with the ESE-containing reactions are from the same population as the results from the pre-mRNA lacking an extra ESE (BG-SMN2) were calculated by a Student's t test.

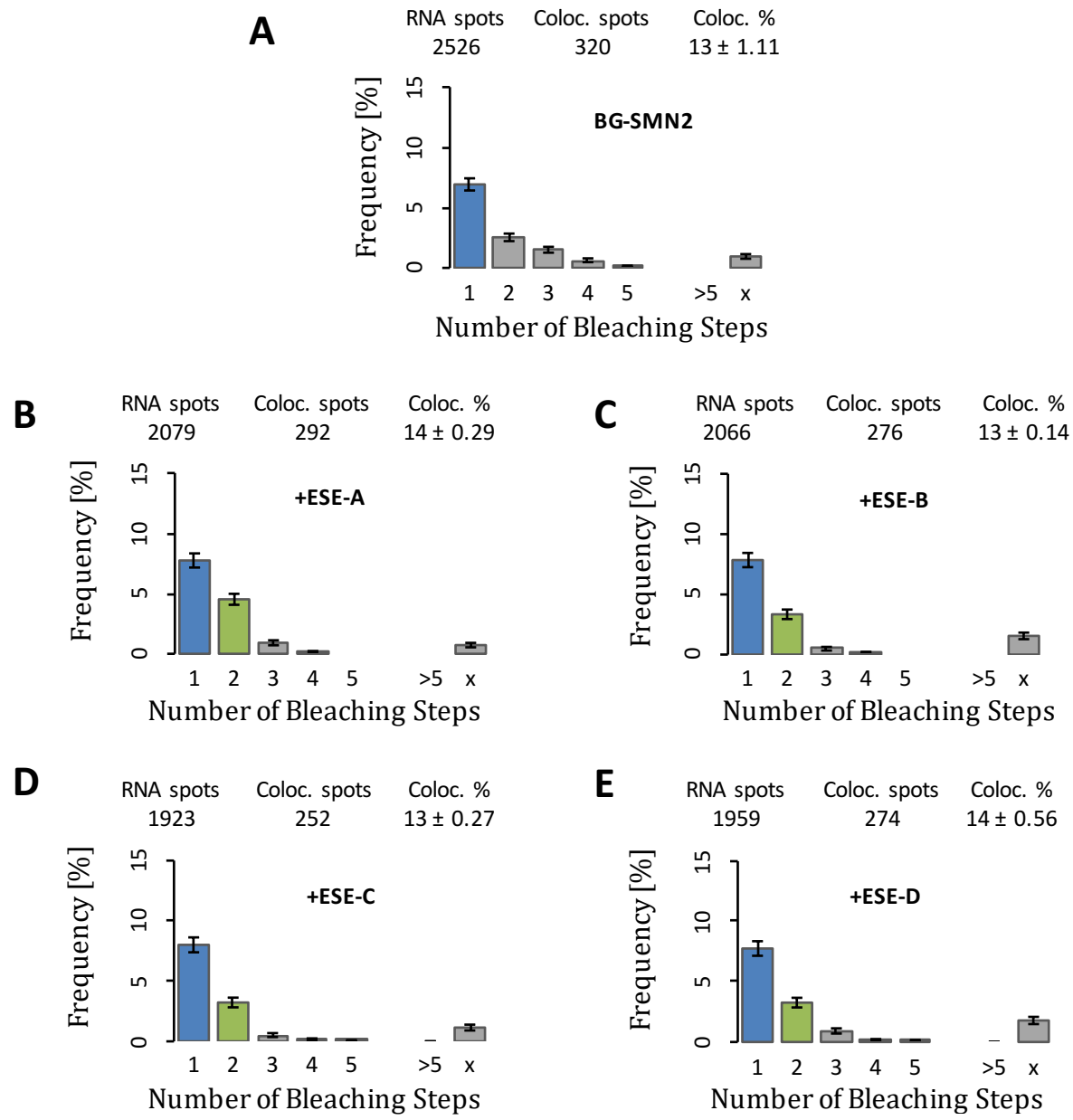


Figure 33. Single molecule studies showing frequency distributions (A) mEGFP-SRSF1 binding to BG-SMN2 with no 3'ESE. (B) mEGFP-SRSF1 binding to BG-SMN2+ESEA. (B) mEGFP-SRSF1 binding to BG-SMN2+ESEB. (C) mEGFP-SRSF1 binding to BG-SMN2+ESEC. (D) mEGFP-SRSF1 binding to BG-SMN2+ESED.

4.4. Multiple ESEs produce additive effects on splicing but do not increase the number of SRSF1 molecules bound.

The results in section 4.3 looking at the link between the potency of ESEs and the binding of SRSF1 confirms prior ideas surrounding the recruitment to and function of ESEs. Another widely accepted model is that an ESE provides a stable binding site for SRSF1. In this case, it would be predicted that additional ESEs of the same type would have no additional effects. The Maniatis laboratory in 1998 argued on the basis of measurements of splicing efficiency that binding of a *Drosophila* Tra/Tra2 ESE was saturated but that additional copies of the ESE enhanced splicing because they increased the chance of a bound SR protein making the contacts required for activation. Since then it has been assumed that this is how all enhancers, including mammalian ones, function. However for mammalian enhancers there have been no direct tests of their mechanisms and in fact a number of models could still explain the phenomenon.

To test whether multiple copies of an optimal ESE improved splicing in mammals, the number of Ron ESEs (ESEA) was increased. The Ron ESE was chosen as this produced the largest effect on splicing and SRSF1 recruitment. The increased number of ESEs produced an approximately linear, R^2=0.98, increase in the level of splicing, as expected (figure 34).

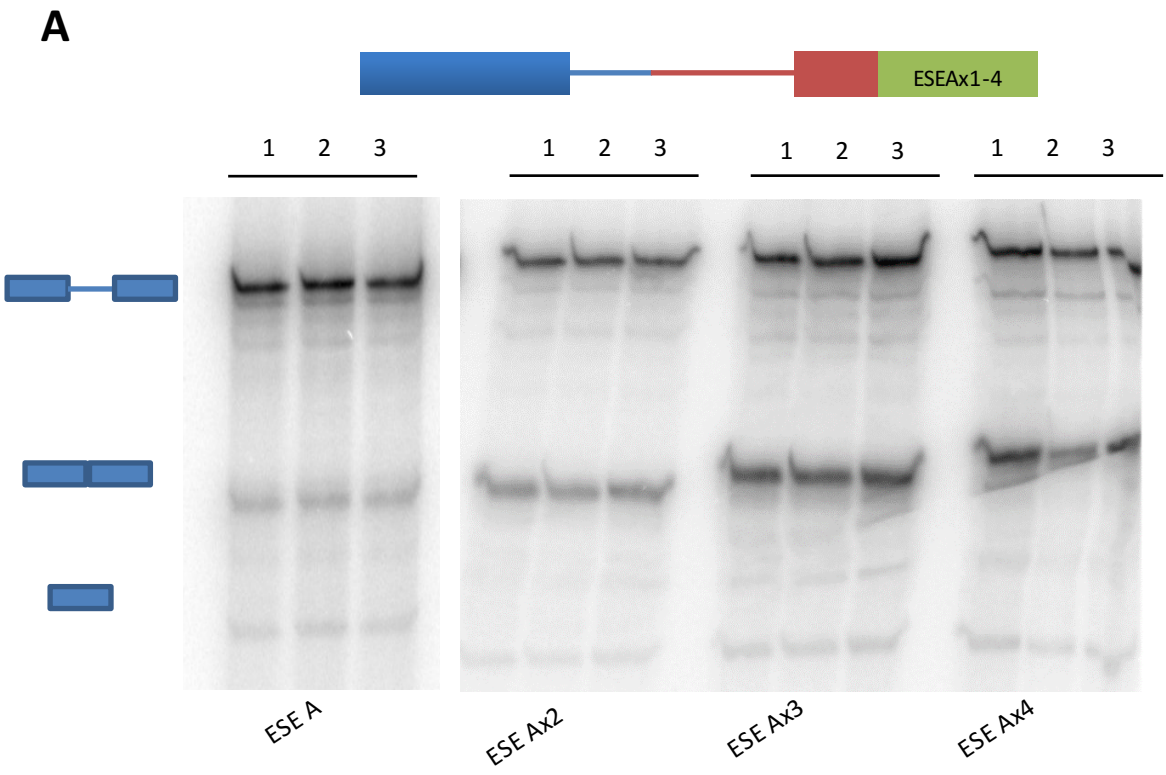
This increase in splicing was also reflected in an increased rate of A complex formation and decrease in H complex formation (figure 35A-D), again as one would expect. However, the ESEs also stimulated an increase in complex B and post splicing complex formation, showing the enhancers are having an effect after the activation of splicing. This is perhaps linked to chapter 3 and another possible role for SRSF1 later in splicing

reactions. SRSF1 has been shown to enhance tri-snRNP binding which would enhance complex B formation.

However, in this case the activation of splicing by SRSF1 was analysed. Complexes were halted in complex A by the use of oligonucleotide complementary to the U6 snRNA. The five constructs with zero to four ESEs were analysed using single molecule techniques with the mCherry-U1A and GFP-SRSF1 nuclear extract. These constructs showed that the increase in the number of ESEs was accompanied by increased recruitment of the second SRSF1; as seen by the increase in the ratio of spots bleaching in two steps to one step. This was also mirrored in an increased co-localisation percentage (0xESEs>4xESEs, 13%>18%) and a further decrease in the background geometric binding (figure 36A-E). However, strikingly, the increased numbers of ESEs led to no notable increase in the number of molecules of SRSF1 that were recruited. This rules out the *Drosophila* model as multiple proteins binding to multiple ESEs are required. The changes in co-localisation percentage and the relative percentage of complexes that contain two molecules of SRSF1 both correlate with the increased splicing efficiency (R^2=0.92) and complex formation efficacy (R^2=0.89) (Figure 37A-B). This data fits with the prior data showing that splicing efficiency is linked to the recruitment of a second molecule of SRSF1.

These results show that the activity of repeated ESEs is connected with the binding of a single extra SRSF1. The increased effect on splicing that is linked to this increased recruitment suggest that each ESE has a low probability of recruiting SRSF1. Therefore by adding more ESEs one is increasing the chance of this single low probability binding event occurring. Furthermore the recruitment of a single molecule would also indicate that there is a single target factor at the 3’ splice site as observed with the *Drosophila*

enhancers (Hertel & Maniatis 1998). We conclude that the affinity of SRSF1 for an ESE is low and that the probability of binding SRSF1 is rate-limiting. The concurrent loss of background binding supports the model that binding of the second SRSF1 strengthens the 3'SS stimulating complex A formation.



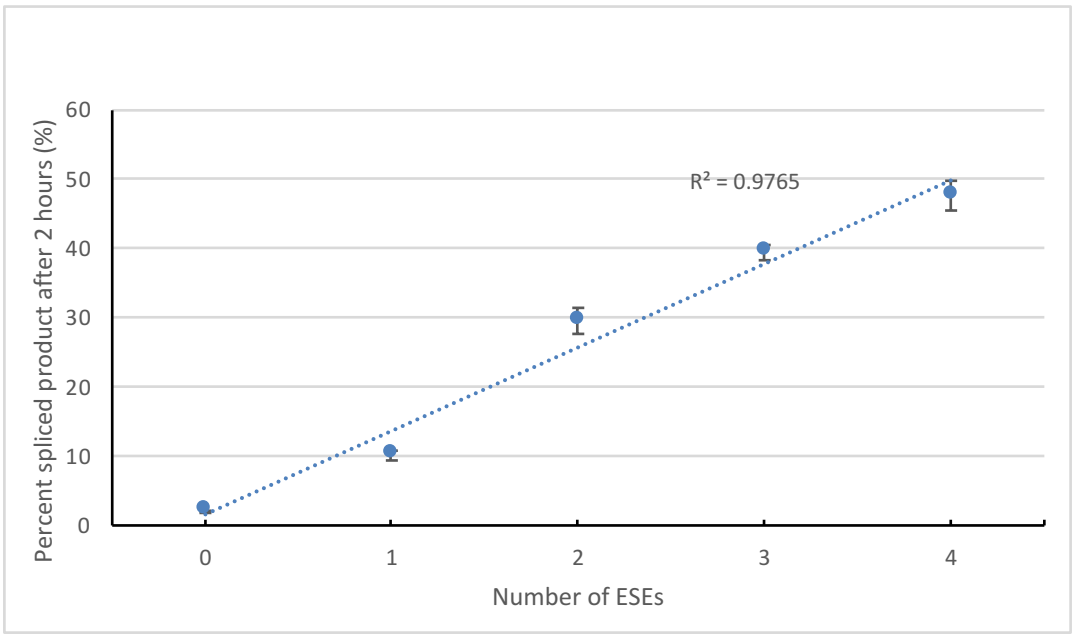


Figure 34. Effects of multiple copies of the Ron ESE-A sequence on splicing. (A) Triplicate Splicing for 2 hours of BG-SMN2 pre-mRNA terminating in 1, 2, 3 or 4 copies of ESE-A. (B) Mean of splicing reaction shown above plus standard deviation and R squared value.

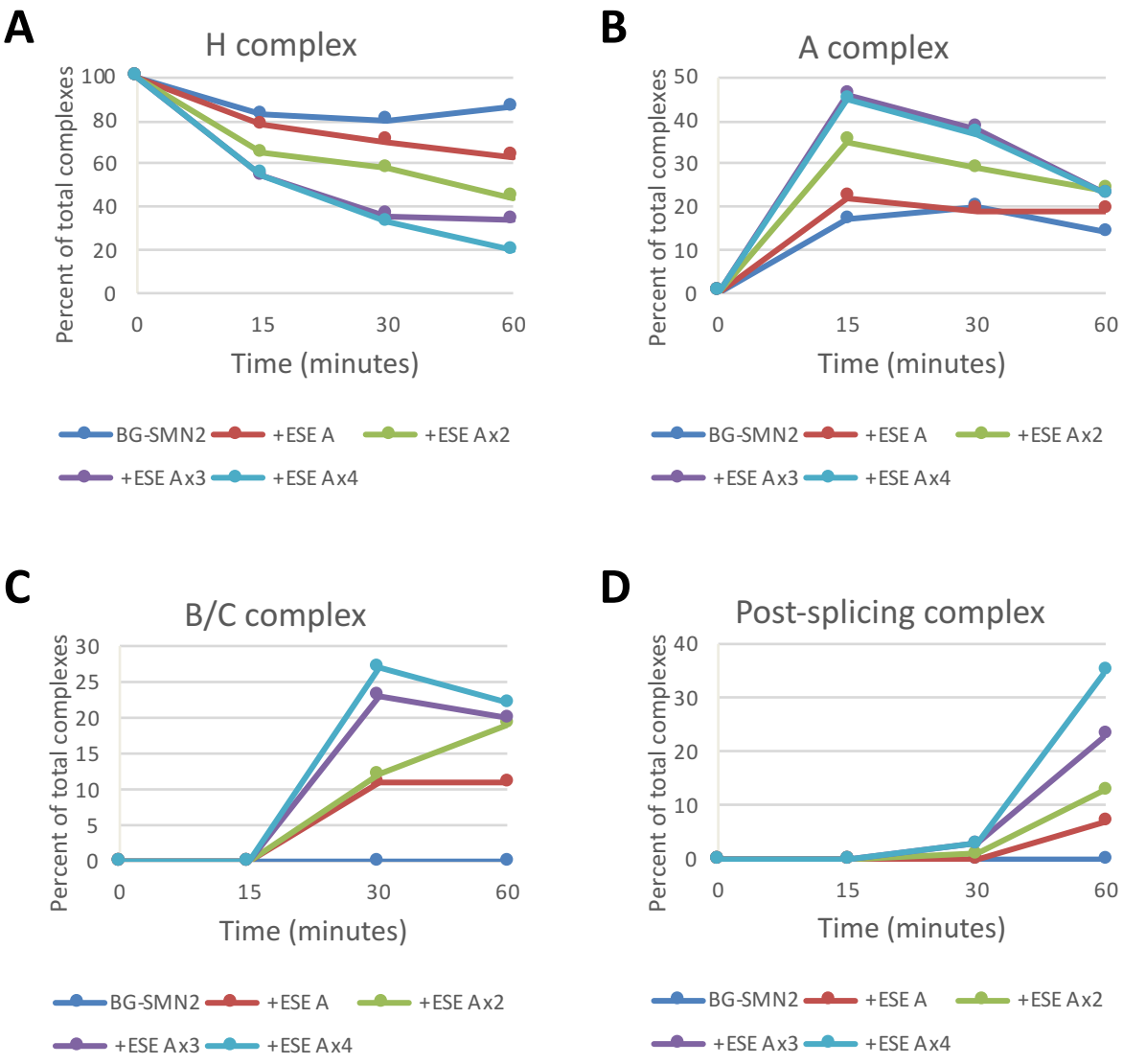


Figure 35. Relative intensities of complexes containing labelled RNA after analysis by native gel electrophoresis of splicing reactions containing the indicated pre-mRNA that had been incubated for the times shown. (A) Formation of H complex, (B) formation of A complex, (C) formation of higher complexes, B or C etc. (D) formation of post splicing complex.

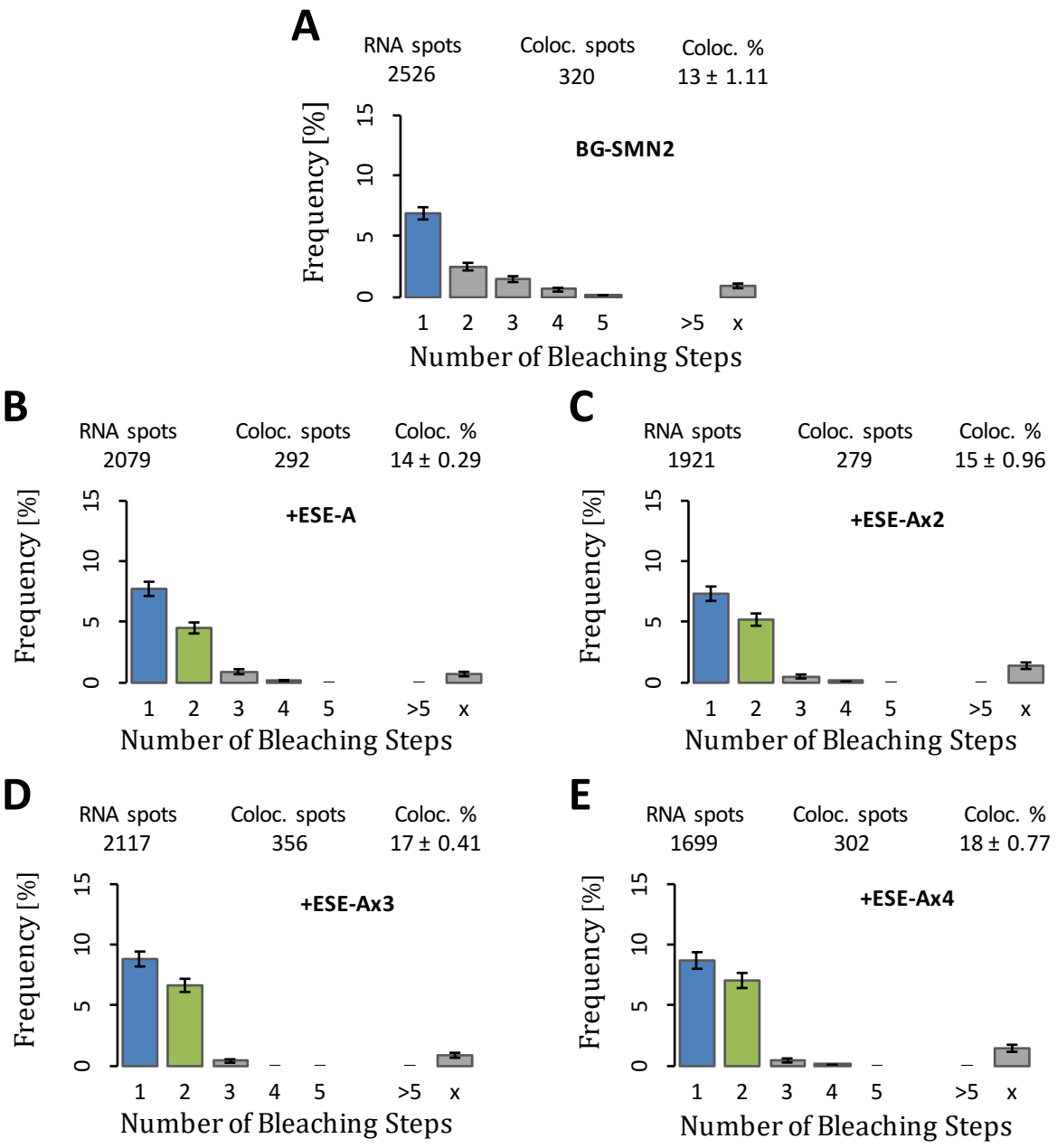
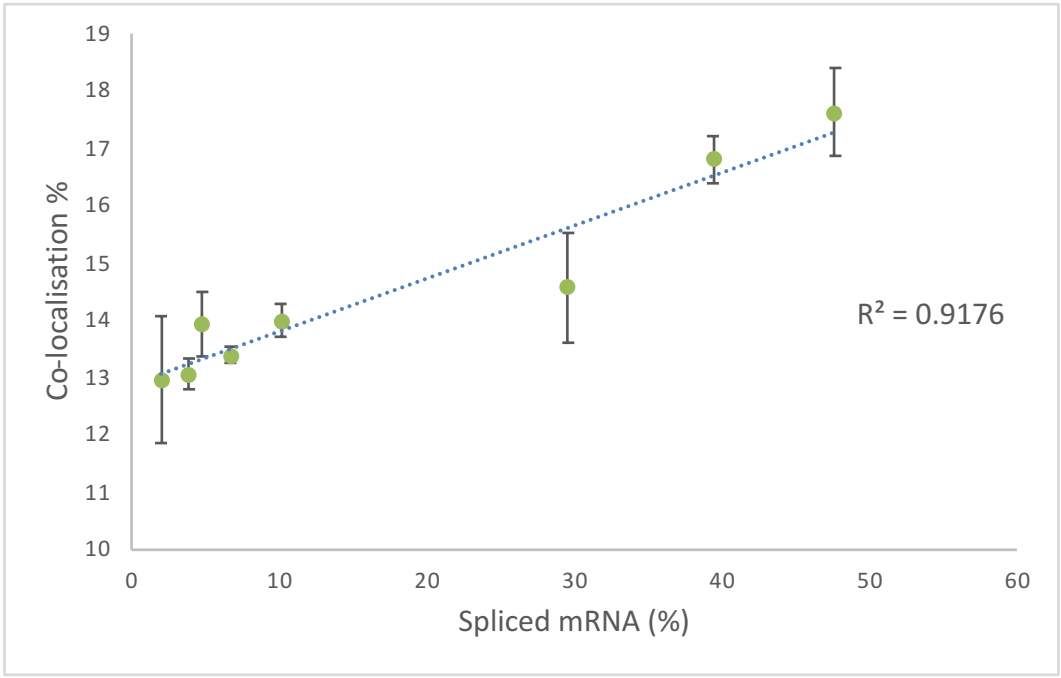


Figure 36. Single molecule multicolour studies on the levels and stoichiometries of mEGFP-SRSF1 binding in nuclear extracts to molecules of BG-SMN2 pre-mRNA terminating in 0, 1, 2, 3 or 4 copies of ESE-A. (A-E) pre-mRNA terminating in 0, 1, 2, 3 or 4 copies of ESE-A respectively.

A



B

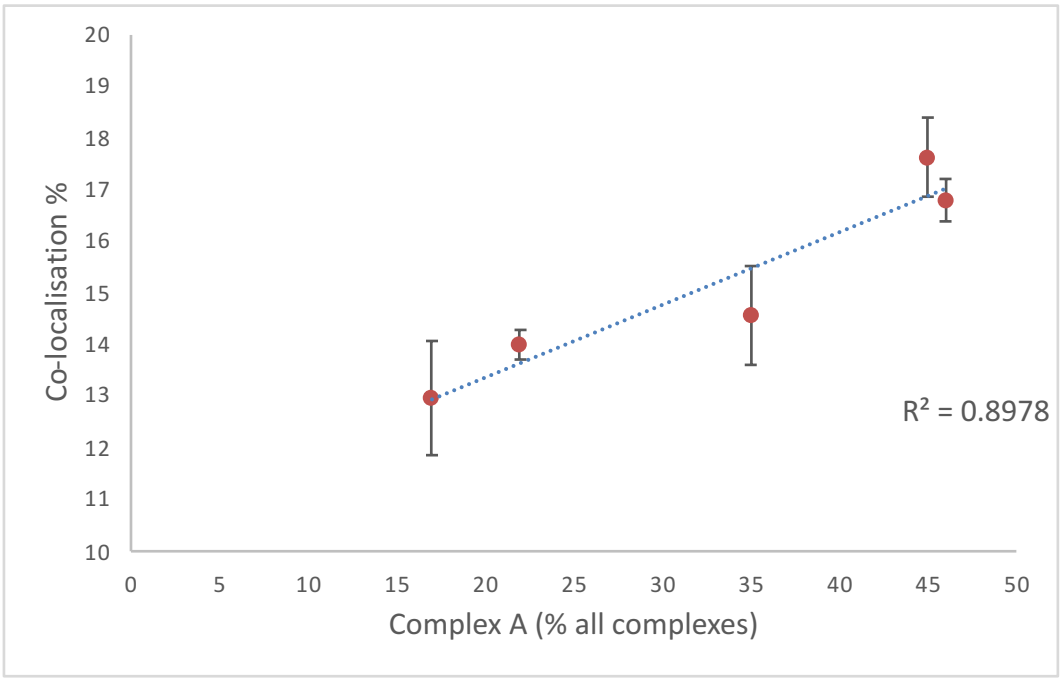


Figure 37. The total proportion of pre-mRNA molecules co-localised with molecules of mEGFP-SRSF1 for the pre-mRNA with no ESE, ESEs A-D and ESE Ax2, 3 and 4 compared with the efficiencies of splicing (A) or complex A formation (B) (only ESE Ax0, 1, 2, 3 and 4).

4.5. Increased repeats of the Ron ESE increases SRSF1 recruitment in cross linking assays.

The single molecule assays in section 4.4 suggest that increased numbers of repeats of an ESE increases the recruitment of SRSF1 but not the number of molecules recruited. In order to validate that multiples of the Ron ESE do indeed stimulate increased recruitment of SRSF1, cross linking assays using 1-4 copies of the ESE were performed.

Figure 38A shows the location of SRSF1 on the membrane, identified by western blotting using the previously used anti-SRSF1 antibody (CSH). Figure 38B shows the western blot overlaid onto the phosphor image in order to show which band corresponds to SRSF1. Figure 38C shows that the SRSF1 band increases with increasing numbers of ESEs. This is further supported by the quantification in figure 38D.

Whilst this data does not rule out any of the originally proposed models it does support that with increased ESEs there is increased recruitment of SRSF1. When taken with the single molecule data it supports the model that the increased recruitment simply comes from an increased chance of recruiting one extra molecule.

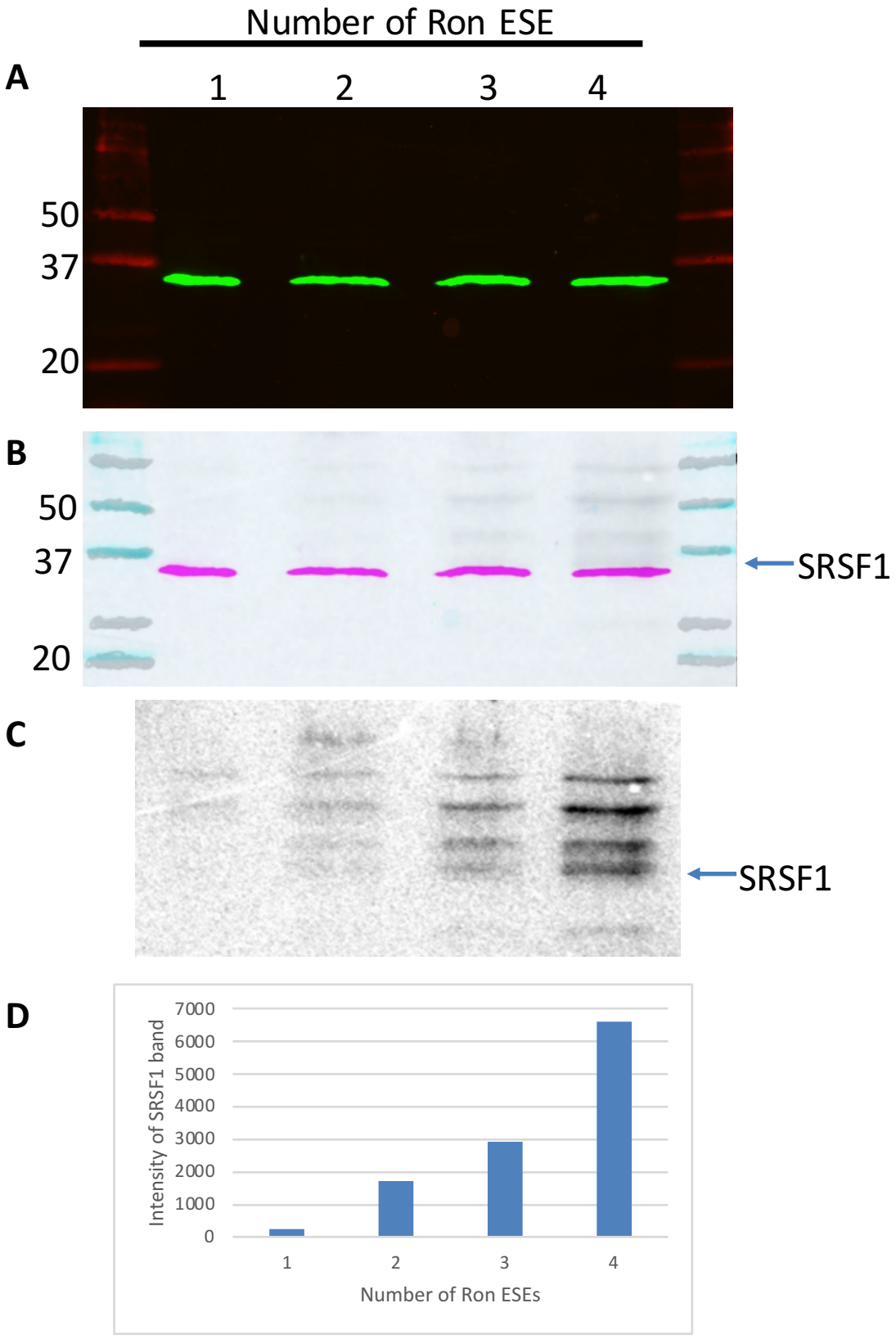


Figure 38. Cross linking assay showing the increased recruitment of SRSF1 by increased repeats of the Ron ESE. (A) Western blot using anti-SRSF1 anti-bodies. (B) Western blot overlaid onto the phosphor image of the membrane. (C) The phosphor image of the membrane with SRSF1 indicated. (D) Quantification of the intensity of the SRSF1 band.

4.6. The potent Ron ESE increases the association of U2AF and U2.

The well-known stimulatory effect of ESEs has been widely shown to be due to the enhancement of the binding of key 3' splice site factors such as U2 and U2AF. To confirm that our ESEs and in turn the bound SRSF1 are acting as expected, the binding of the three key 3' splice site factors was analysed via single molecule methods. These nuclear extracts have previously been validated and used to identify novel complexes that can form prior to A complex (Chen et al. 2016).

The BG-SMN2 hybrid construct, which splices poorly (~7% after 2 hours) and forms complexes very slowly and inefficiently, and the version with four additional ESEs (BG-SMN2+ESEAx4) attached to the 3' end, which splices efficiently (48% after two hours) and forms complexes rapidly, were used and analysed with two separate extracts. The first extract used contained mEGFP-labelled U2B'', a major component of the U2 snRNP complex which has been shown to only bind as part of the complex. Figure 39A shows the binding of the labelled U2B'' to the RNA construct with no ESEs where the binding is very weak, shown by a very low level of co-localisation of 10%. In contrast Figure 39B shows the binding of U2B'' to the construct with four additional ESEs where the binding is much improved with a much higher co-localisation percentage, 25%. A similar pattern can be seen for U2AF35 in figures 39C and D, again, the construct with no ESEs gives weak binding whilst the construct with the ESEs shows much improved binding, 12%>25% co-localisation. Figures 39E and F show the binding of U2AF 65; in this case, although the binding is improved, it is not to the level as seen with U2 and U2AF35. In

all three cases the pattern of binding does not change, with nearly all complexes showing a single binding event.

These results show that the ESEs and the subsequently recruited SRSF1 are functioning by improving the recruitment of key factors to the 3' splice site. However whilst these experiments are revealing in that they show the improved binding in a clear quantitative way, we cannot pinpoint the exact protein contact that the SRSF1 is making as the stabilisation of one would stabilise the others. The contact could be either of the U2AF proteins, the SF3B complex or the SF3A complex. This result is to be expected and is in line with the literature and common understanding of how ESEs work. This work does however, validate the previous results that are shown and also demonstrates that the additional ESEs are simply increasing the chance of recruitment of SRSF1 which stabilises the 3' splice site and is not acting by another unforeseen route.

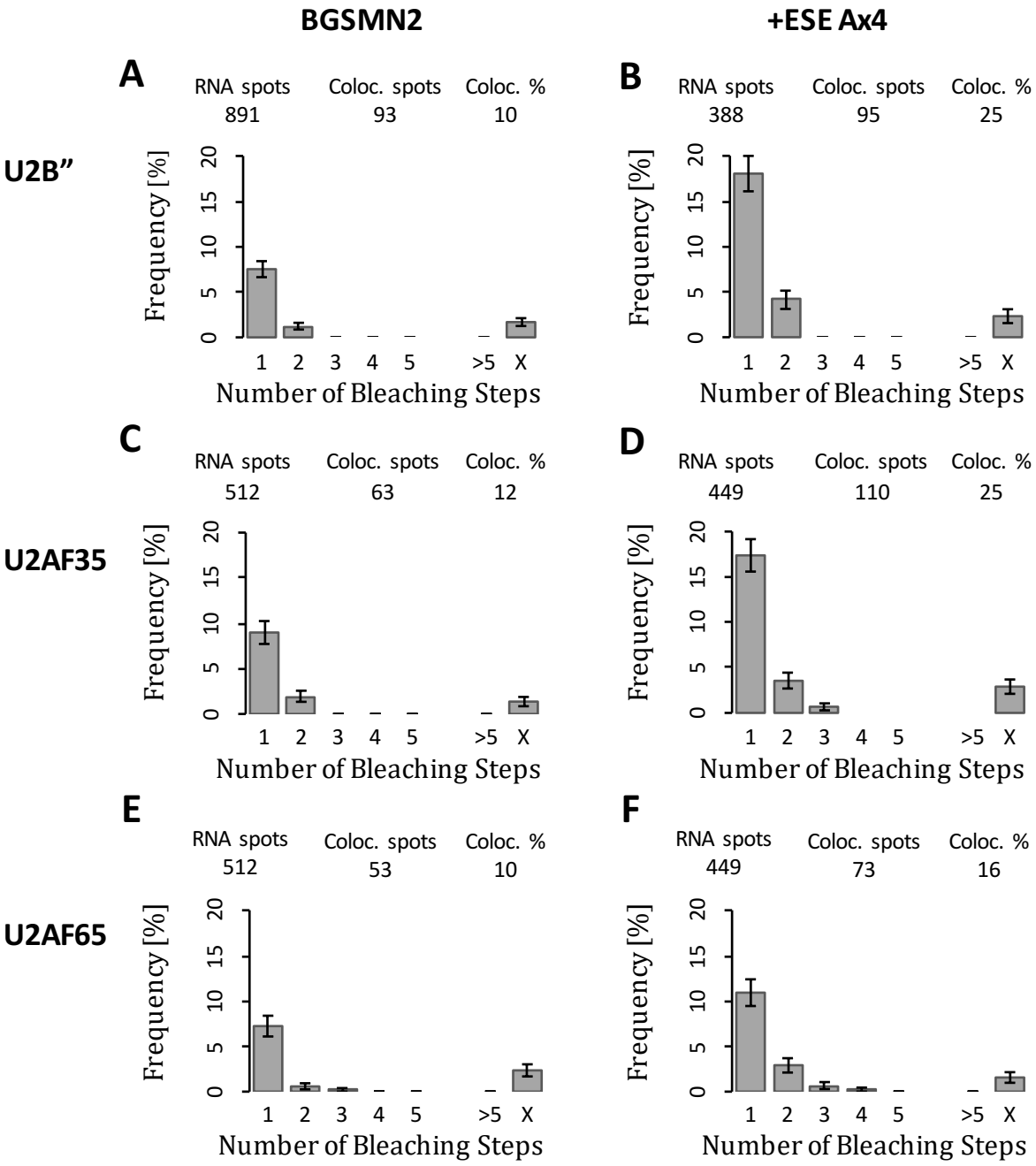


Figure 39. Single molecule colocalisation studies on the levels and stoichiometries of mEGFP-U2B'', mEGFP-U2AF65 and mCherry-U2AF35 binding in nuclear extracts to molecules of BG-SMN2 pre-mRNA terminating in 0 or 4 copies of ESE-A. (A-B) U2B'' (C-D) U2AF35 (E-F) U2AF65.

4.7. Does Tra2B binding correlate with splicing?

Another important SR protein involved in the regulation of splicing is Tra2β, one of two human homologues of the aforementioned *Drosophila* Tra2 protein. As in *Drosophila*, Tra2β has been extensively implicated in splicing pathways which are activated in early developmental processes such as spermatogenesis (Venables et al. 2000; Venables et al. 2005; Grellscheid et al. 2011). However, Tra2β is still expressed in a number of adult tissues leading to speculation that it may play a role in controlling alternative splicing in mature mammalian tissues. A number of studies seem to support this idea. CLIP studies further support this idea, showing a number of binding sites for the protein in alternatively spliced exons (Grellscheid et al. 2011). The apparent binding preference is for RNA tracts that contain GAA repeats. One such exon that contains a binding site for Tra2β is SMN exon 7 (Martins de Araújo et al. 2009; Tsuda et al. 2011).

Due to our use of the SMN 2 exon 7 in our model system and the presence of GAA motifs in some of the enhancers we tested, we decided to analyse the binding of Tra2β to our RNA constructs. This would serve a dual purpose: to see if the ESEs used could recruit more than one SR protein and subsequently if the binding of the protein correlated with the splicing changes seen.

The BG-SMN2 hybrid construct without an ESE and the four constructs with the four different ESEs were tested with a nuclear extract that contained mEGFP-labelled Tra2β prepared from 293T cells (extract prepared by Dr L.P. Eperon). Figure 40A shows, that in the absence of an ESE, the binding pattern is similar to that given with SRSF1, indicating that there might be one or two molecules bound with a higher probability as well as an apparent background. This fits with what one might expect, a single Tra2β

being recruited to the site in the centre of the SMN exon and non-specific background binding driven by the RS domain as with SRSF1. However the results did not fit the prediction when the ESEs are added as seen in figure 40B-E. Each of the ESEs stimulated the binding of Tra2β but to variable levels. It is clear that these variable levels of binding did not correlate at all well with the different splicing patterns witnessed in each case. However, whilst the binding of the protein itself did not correlate with splicing patterns seen, it is possible that both Tra2β and SRSF1 are needed for a potent effect to be seen and that Tra2β may act as a co-factor to enhance the effect of SRSF1.

In order to analyse this the RNA constructs with tandem repeats of the potent Ron ESE were also tested. This would serve to show if the proteins recruitment correlated with the increased recruitment of SRSF1 seen with the increasing number of ESEs. Figure 41A-D shows that the binding patterns of Tra2β with the constructs that had an increasing number of ESEs did have an effect on the recruitment and binding pattern. What is striking though, is that the construct with one ESE gave the same pattern as the one with two ESEs and similarly three ESEs matched four ESEs. This rules out any correlation between Tra2β binding and both splicing and SRSF1 recruitment as both increase linearly with the increasing number of ESEs. One possibility to explain the step-wise changes in binding is that there is some steric hindrance between ESEs i.e. if the Tra2β binds to the first ESE it cannot bind to the second but can to the third and so on.

In order to look at whether the binding of Tra2β and SRSF1 were competing for similar sites, a co-transfected nuclear extract with mEGFP-SRSF1 and mCherry-Tra2β was used. In this case only the RNA with four ESEs was used. Figure 42A and B show that the binding of SRSF1 and Tra2β are much the same as we have seen in previous extracts.

There does seem to have been a reduction in the number of complexes with 3 or more Tra2β proteins present but there was a significant proportion of complexes with two molecules bound. This shift may be due to the relative amount of protein that is labelled in the extract. The complexes containing two molecules of mEGFP-SRSF1 were analysed. The binding patterns seem to remain relatively unchanged, figure 42C and D. This shows that both proteins can bind concurrently and do not exclude or restrict each other.

This data collectively indicates that, whilst nearly all the ESEs can bind Tra2β, this binding does not correlate to the different splicing patterns seen and does not assist in the recruitment or activity of SRSF1. However it does not act as a competitor either, with both proteins being bound to the same extent regardless of the number of the other respective protein. To conclude, the ESEs can recruit Tra2β but surprisingly this seems to have almost no effects, either positive or negative. Due to the somewhat surprising nature of these results, further experiments in order to explore this relationship are required.

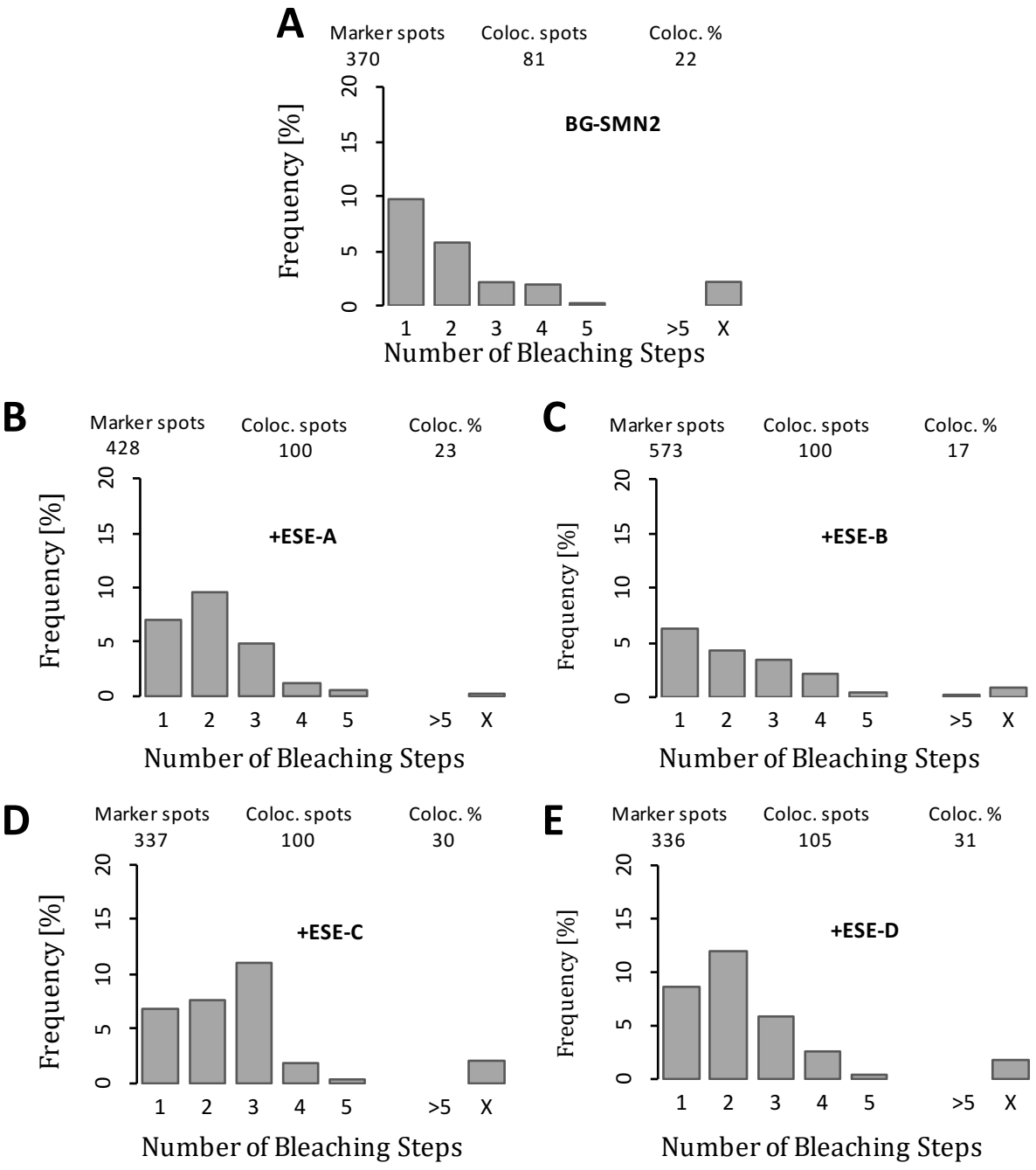


Figure 40. Single molecule co-localisation studies on the levels and stoichiometries of mEGFP Tra2β binding in nuclear extracts to molecules of BG-SMN2 pre-mRNA terminating in no ESE or ESE-A-D. (A) No ESE, (B) ESE-A, (C) ESE-B, (D) ESE-C, (E) ESE-D.

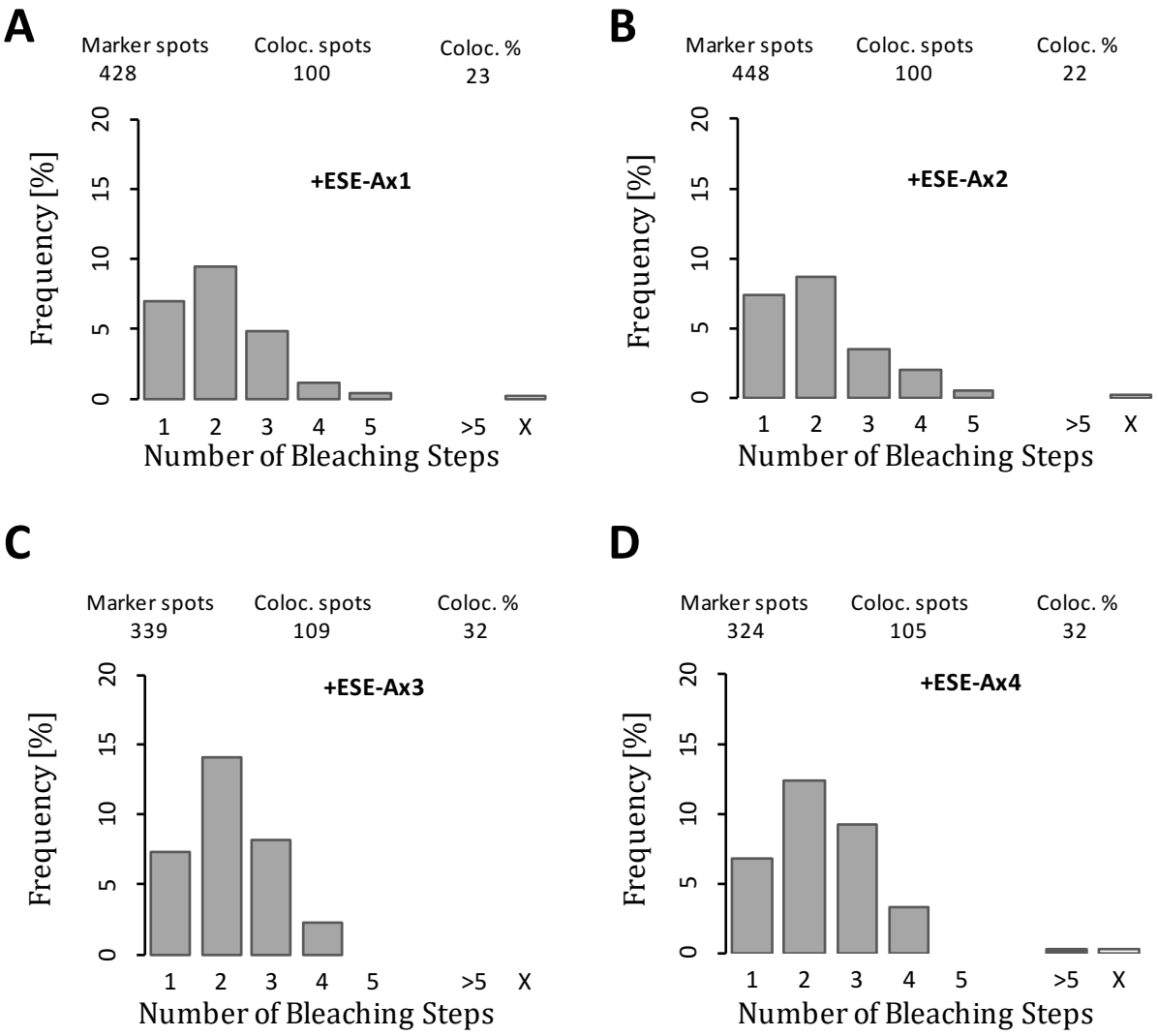


Figure 41. Single molecule co-localisation studies on the levels and stoichiometries of mEGFP Tra2β binding in nuclear extracts to molecules of BG-SMN2 pre-mRNA terminating in no ESE or ESE-Ax1, 2, 3, 4. (A) ESE-Ax0, (B) ESE-Ax1, (C) ESE-Ax2, (D) ESE-Ax3.

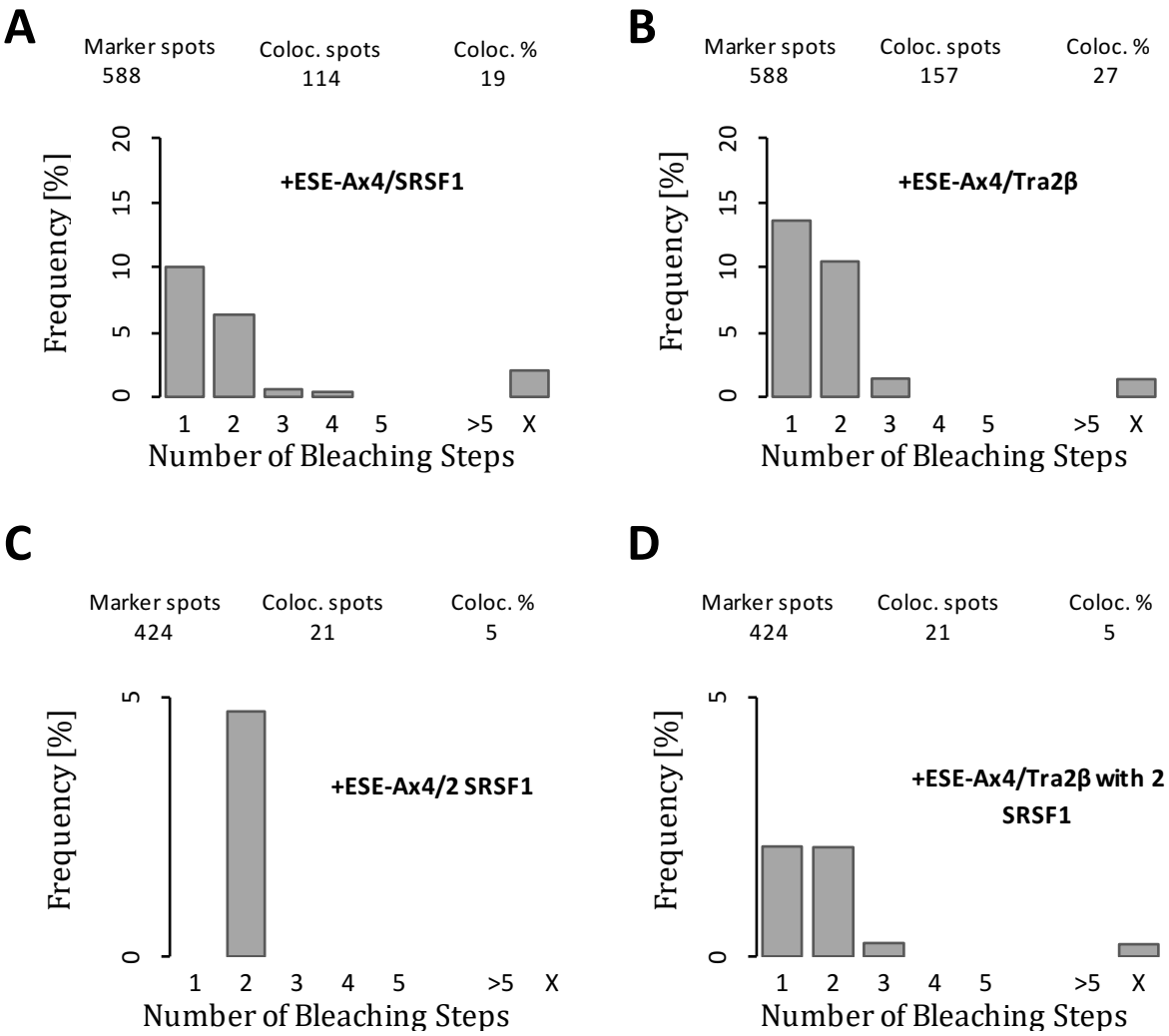


Figure 42. Single molecule co-localisation studies on the levels and stoichiometries of mEGFP-SRSF1 and mCherry-Tra2β binding in nuclear extracts to molecules of BG-SMN2 pre-mRNA terminating in ESE-Ax4. (A) mEGFP SRSF1, (B) mCherry Tra2β, (C) Two molecules of SRSF1, (D) Tra2β when there are two molecules of SRSF1.

4.8. Summary.

In contrast to a spate of results describing the transcriptome-wide binding sites and protein interactions of proteins that activate or repress splicing, the molecular mechanisms by which they effect splicing have received little attention in recent years. Such investigations are not amenable to ensemble methods, which provide little information about the heterogeneity of complexes or the stoichiometry's of components. This is particularly true of activator proteins, which exert effects at a range of distances from their binding sites.

The comparison of four different ESEs showed that they all stimulated splicing, but the Ron exon 12 sequence was by far the best. This might have been explained had it recruited more than one molecule of SRSF1, but in fact the single molecule results show that all the ESEs stimulated the recruitment of a second SRSF1, additional to the single SRSF1 recruited to the 5' splice site, and also suppressed the geometric distribution of 3 4, 5 etc. molecules. The fact that only one SRSF1 was seen is interesting as it indicates that there were no additional SRSF1 proteins propagating out from the initial site, at least in A complex. This however does not rule out the recruitment of other SR proteins or a heterogeneous complex. The percentage of co-localisation did not increase significantly for the stronger enhancers. This is most likely due to the first molecule of SRSF1 being recruited at the 5'SS regardless of the presence of an ESE (Chapter 3). Together with the geometric distribution associated with complexes that fail to form complex A, this explains why most experiments attempting to quantify SRSF1 binding have only been successful on short segments rather than full splicing units. The effect of the ESEs is clearer when we look at the proportion of RNA molecules associated with

two molecules of mEGFP-SRSF1. This correlates well with the splicing efficiency of each of the constructs. What this correlation allows us to establish is that splicing efficiency is directly linked to the ESE's ability to recruit the second SRSF1.

Whilst the data looking at the four different enhancers supported previous ideas surrounding enhancers, the data looking at multiple ESEs provided strong evidence for a mechanism. As expected we saw that additional repeats of the Ron ESE increased splicing sequentially in a near linear manner. An increase in splicing with additional ESE repeats could be accommodated for by at least four models, three of which involve the binding of additional molecules of SRSF1. However, we saw no evidence for more than one additional SRSF1 molecule binding. There was a continuation of the trend witnessed with the four different ESEs, whereby the strongest splicing construct, with four Ron ESEs, gave the highest proportion of molecules bleaching in two steps. However in this this case there were substantial changes in the percentage of co-localisation between each construct, possibly because the level of binding initiated by the ESE was at a sufficiently high level to be detectable above the binding of the 5'SS-bound SRSF1 and the background. We conclude that each additional ESE increases the chance of a single binding event occurring. One explanation is that SRSF1 binds in a transient manner initially (Cho et al. 2011) and it is only maintained on the RNA if an interaction occurs with an effector such as U2AF, U2 snRNP (Martins de Araujo et al. 2009; Smith et al. 2014; Anczuków et al. 2015) or an intermediary.

Furthermore we show that this is not the case with another important SR protein, Tra2β. In this case binding can occur but this has no clear effect on splicing and does not seem to effect the binding of SRSF1. This further supports that the effect we are seeing is

solely down to the increased chance of SRSF1 recruitment. The lack of effect of Tra2β and the striking effect of SRSF1 may be explained by their target factor. It is possible that Tra2β may serve to stabilise a 3'SS factor that can already stably bind and therefore is not limiting whilst SRSF1 may act on one that is.

The transience of SRSF1 binding inferred here fits with prior observations of SRSF1 binding to RNA. However the confirmation of this also concurrently rules out the *Drosophila* model of ESE function in which the presence of multiple enhancers leads to multiple copies of the activator protein. An alternative model that fits our results is one wherein binding is limiting and not the subsequent interaction. As stated previously this may not be the case with every enhancer protein, or even every SR protein, or every ESE. What this does show is that with the archetypal enhancer protein found in humans, then with increased ESEs there is simply an increased chance of a single molecule binding. Other proteins may use one of the other models outlined earlier.

The evidence here shows that the effects of a 3' ESE are intrinsically linked to their ability to recruit a single additional molecule of SRSF1 and that a stronger site or additional repeats act by increasing the chance of this extra SRSF1 binding. The binding of this single molecule of SRSF1 is shown to enhance splicing by enhancing the binding of key 3' splice site factors. SM methods also reveal that there is no homogenous complex, containing multiple molecules of SRSF1, present and that another key SR proteins, Tra2β, binding does not correlate with the increases in splicing that we see.

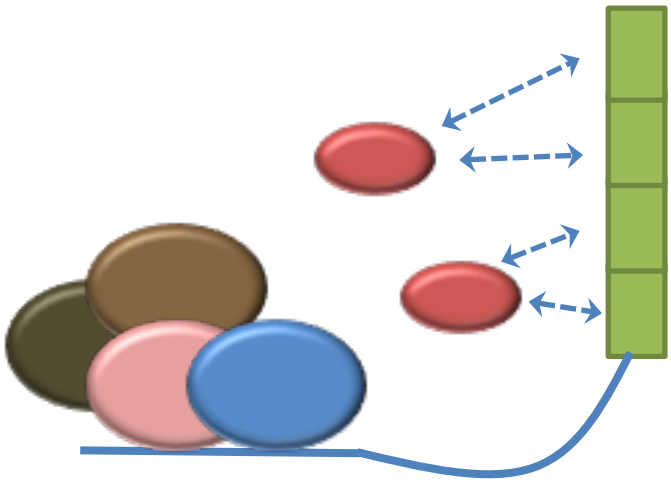


Figure 43. Model showing how increased numbers of ESEs increases the effect on splicing by increasing the chance of a single binding event occurring.

Chapter 5. An ESE does not require an RNA connection to its target.

5.1. Introduction.

The second step of ESE activity whereby the bound protein makes contact with its target immediately follows the binding of the enhancer protein. However whilst numerous studies have tried to characterise the RNA binding preferences for enhancer proteins such as SRSF1 (SELEX (Tacke & Manley 1995), functional SELEX (Liu et al. 1998; Sheth et al. 2006), CLIP studies (Sanford et al. 2008; Pandit et al. 2013b) or RNA-seq (Anczuków et al. 2015)), the interaction step is far less studied. Similarly to the lack of research into the effect of multiples of enhancers, this is in part due to a lack of suitable techniques.

Previously only three assays had looked to answer how an ESE bound enhancer protein can contact its target at a distance and all have had major drawbacks that mean a definitive model cannot be drawn. Despite this, cartoons depicting splicing regularly depict ESEs looping to their target. The incorporation of these cartoons in text books has likely led to an assumption that this question has been addressed and does not need further research.

The first piece of work to look at this mechanism came in 1998 where an RS domain based on SRSF1's RS domain was tethered at variable distances to the 3' splice site upon which it acted (Graveley et al. 1998b). This study, whilst being a clever and informative for its time, did have some drawbacks. First of all the data generated from the experiments was fitted to a mathematical model for the interaction of two points on a freely moving chain, but could also fit to a mathematical model for the propagation of

proteins along the RNA (Lewis, Andrew J. Perrett, et al. 2012). Second, the use of tethered proteins or domains is not an ideal system for the study of ESEs as it eliminates the vital binding step which as shown in the previous chapter is limiting in some cases. One might expect that if an RS domain was tethered to the RNA then at some point it would inevitably contact its target site. This however does not necessarily mean that this is how an enhancer protein contacts its target in cells. As discussed previously the transient nature of the RNA protein interaction might mean the protein is not bound long enough for this system to be viable.

The next piece of work that looked to address this issue was in 2004 where Shen et al found that an RS domain tethered at a distance could also be cross-linked to near a 3' splice (Shen et al. 2004). This was an exceptionally clever experiment and seemingly provides strong evidence for the looping based model. However this paper also suffers from the assumptions made in 1998 about the stable binding of enhancer proteins to ESEs (Hertel & Maniatis 1998). Given that proteins are likely to bind transiently, it is possible that their lifetime on the RNA is not long enough for a looped interaction to occur and that by tethering the RS domain, one is generating an artificial system whereby the bound lifetime of the protein-RNA complex is effectively infinite.

The final study to look at this mechanism came in 2012. Here Lewis et al, from this lab, used click chemistry to introduce a HEG linker in between an ESE and two alternative 5' splice sites (Lewis et al. 2012). In this case if the ESE was functioning the splicing was shifted to the upstream 5' splice site whilst without it, the downstream site was used. This raised the question of whether ESEs function differently at 5' splice sites, possibly due to the recruitment mechanism outlined in chapter 3. The major drawback of this

assay, however, was the use of the click chemistry. This particular reaction resulted in the introduction of triazole linkage into the RNA which might have unknown effects on the RNA and may inhibit the ESE by some unseen route.

Whilst three different studies have attempted to answer the question of how ESEs function, we still do not have a conclusive mechanism. It is, similarly to the binding of ESEs, possible that different ESEs may function differently; depending on distance etc.

Therefore whilst this question is central to our understanding the basic mechanism of how ESEs function, it remains relatively unanswered. In order to try to avoid the major drawbacks of the studies outlined above a new approach is needed. In this chapter we outline the development of a cohesive method for the synthesis (in collaboration with Dr. Burley and Dr. Reichenbach) and attachment of non-RNA strands into larger RNA constructs. This will allow us to discern how an ESE or stimulatory 5' splice site makes contact with its target. Using this strategy will avoid the use of click chemistry and avoid the need to tether proteins/domains and thus will represent better what happens in cells.

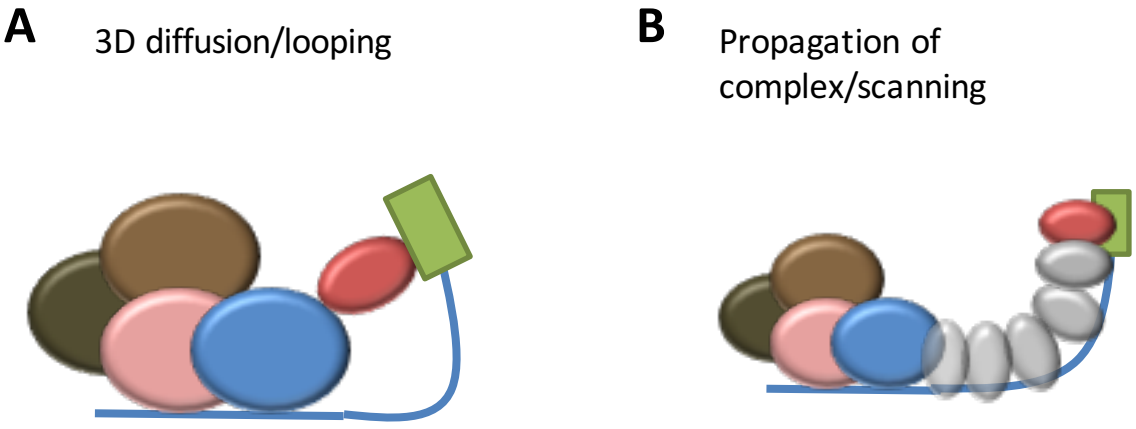


Figure 44. Models showing the two proposed systems whereby an ESE can exert its effect from a distance. (A) The ESE binds the SRSF1 and the intervening RNA loops out and a contact is made in that manner. (B) The ESE binds a molecule of SRSF1 and this stimulates the binding of other proteins which propagate the signal along the RNA to the target.

5.3. Synthesis, purification, attachment of hybrid RNA strands.

To test the influence of non-RNA building blocks on RNA splicing, we needed to prepare pre-mRNA constructs of a sufficient length for efficient splicing that also contained non-natural bases. Previously, this was achieved by using click chemistry to attach synthetic chains to a longer RNA prepared by transcription. However, since we wanted to avoid any disruption to the RNA backbone apart from the intended insertion of a polymer chain, we adopted a different approach from that used previously to avoid any residual triazoles from click linkers, amide linkages, etc., and instead used phosphoramidites to synthesise small oligonucleotide strands (up to 50 nt) containing linkers by solid phase synthesis. We envisaged that these strands could later be ligated to the larger RNAs obtained by transcription to form the full-length splicing constructs.

The system we tested was one in which a downstream 5' splice site or an ESE stimulate a weak upstream 3' splice site (the BG-SMN2 RNA construct used in the previous chapters). The modifications tested were a hexaethylene glycol (HEG) spacer as well as two abasic modifiers, one with and one without the 2' OH group characteristic of RNA (Figure 45). The phosphoramidites used were synthesized via solid-phase oligonucleotide synthesis protocols on an ABI 394 synthesizer by Dr Linus Reichenbach. After cleavage from the solid support and deprotection of the protecting groups, the crude oligonucleotides were purified using gel electrophoresis and UV shadowing.

The ligation procedure, splinted ligation using RNA ligase 2, was modified from Crawford et al. This procedure involves the use of a DNA splint to bring two RNA fragments into

close proximity in a duplex, RNA ligase 2 can then attach the two fragments. In order for the fragments to be ligated, the 3' fragment must have a 5' phosphate and the 5' fragment must have a 3' OH. These conditions are generated using polynucleotide kinase (PNK) (NEB) and Antarctic phosphatase (AP) (NEB). The ligated fragment is purified from the un-ligated reactants by gel electrophoresis.

Preliminary experiments were performed using short DNA oligonucleotides in order to identify the shortest possible splint necessary for efficient duplex formation. Efficient duplex formation is important for splinted ligation; shown in figure 47. A short splinted region is desirable so as to minimise the length of synthesised oligonucleotide required; synthesis efficiency decreases with length. Figure 46 shows that efficient duplex formation is possible with a twenty base splint, ten base pairs with each fragment.

Initial attempts to perform splinted ligation on transcribed RNA and commercially purchased RNA fragments resulted in smearing in the gel. This was identified to stem from RNA degradation due to heating in the presence of magnesium, found in the buffer, following PNK treatment. This was circumvented by removing the kinase using phenol-chloroform extraction rather than heat inactivation.

Antarctic phosphatase was omitted to begin with as ligation was being performed using a transcribed 5' fragment, which should terminate in a 3'OH, and a synthesised/commercial 3' RNA fragment, which has a 5'phosphate following PNK treatment. However ligation was unsuccessful unless Antarctic phosphatase was used. One possibility is that the transcribed RNA may have phosphates added to its 3' end, which the Antarctic phosphatase would also be able to remove.

The final issue encountered was the temperature at which the ligation was performed. This was initially 37 °C, as dictated by the manufacturer's instructions, however this gave poor ligation which was likely because this temperature was above the melting temperature of the duplex. To overcome this, the fragments were heated to 70 °C and cooled to 10 °C below the melting temperature of the lower of the melting temperatures of the two 10 base double-stranded regions. Ligation was then performed at 2 °C below the melting temperature of the combined 20 base duplex region.

The final ligation procedure used was as follows. The 5' transcript for ligation was transcribed (as described in 2.1.3) and gel purified using polyacrylamide gel electrophoresis. The RNA was then mixed with the chemically synthesized 3' synthetic fragment that contained one of the three non-RNA linkers, or none, between two stretches of RNA. The RNA molecules were mixed in a 1:5 ratio of transcript to synthetic fragment. Any 5' or 3' phosphates were then removed by incubation with 10 units of Antarctic Phosphatase (NEB) at 37°C for 45 min. The enzyme was inactivated by heating to 70°C for 5 min and the protein was removed using 3 consecutive extractions with phenol-chloroform. A phosphate was added to the 5' end of the 3' fragment using 20 units of polynucleotide kinase (NEB) with a 10 fold excess (compared to RNA) of ATP. Protein was removed by three extractions with phenol-chloroform. The 20 base DNA splint was added at a ratio of 1:1 with the 5' fragment. The reaction was then heated to 70 °C for 5 min and cooled slowly to 10°C colder than the lower of the melting temperatures of the two 10 base double-stranded regions. 20 units of RNA ligase 2 were added and the reaction mixture was incubated at 2 °C below the melting temperature of the 20 base duplex or 37°C, whichever was lower (reactions were successful at

temperatures as low as 20 °C) for 2 h. The full-length RNA was purified by running on a polyacrylamide gel and visualized using X-ray film; two bands were visible, corresponding to the ligated and un-ligated RNA.

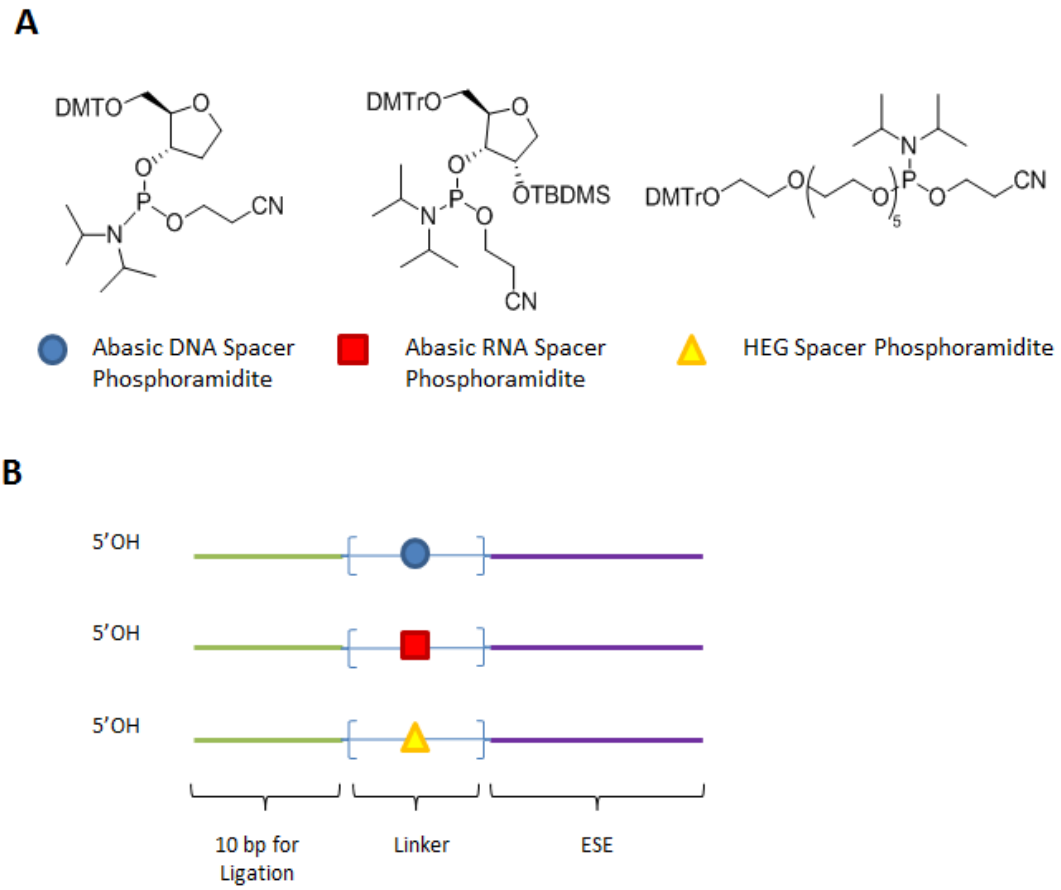


Figure 45. Schematics of the three different phosphoramidites used in this study and how they fit in to the synthesised strands.

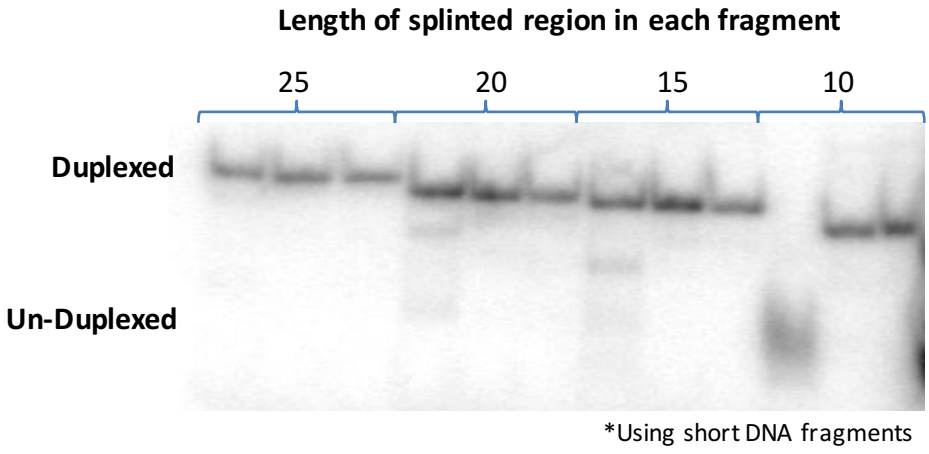


Figure 46. Native (no Urea) polyacrylamide gel showing DNA duplexes containing two DNA fragments and a DNA splint.

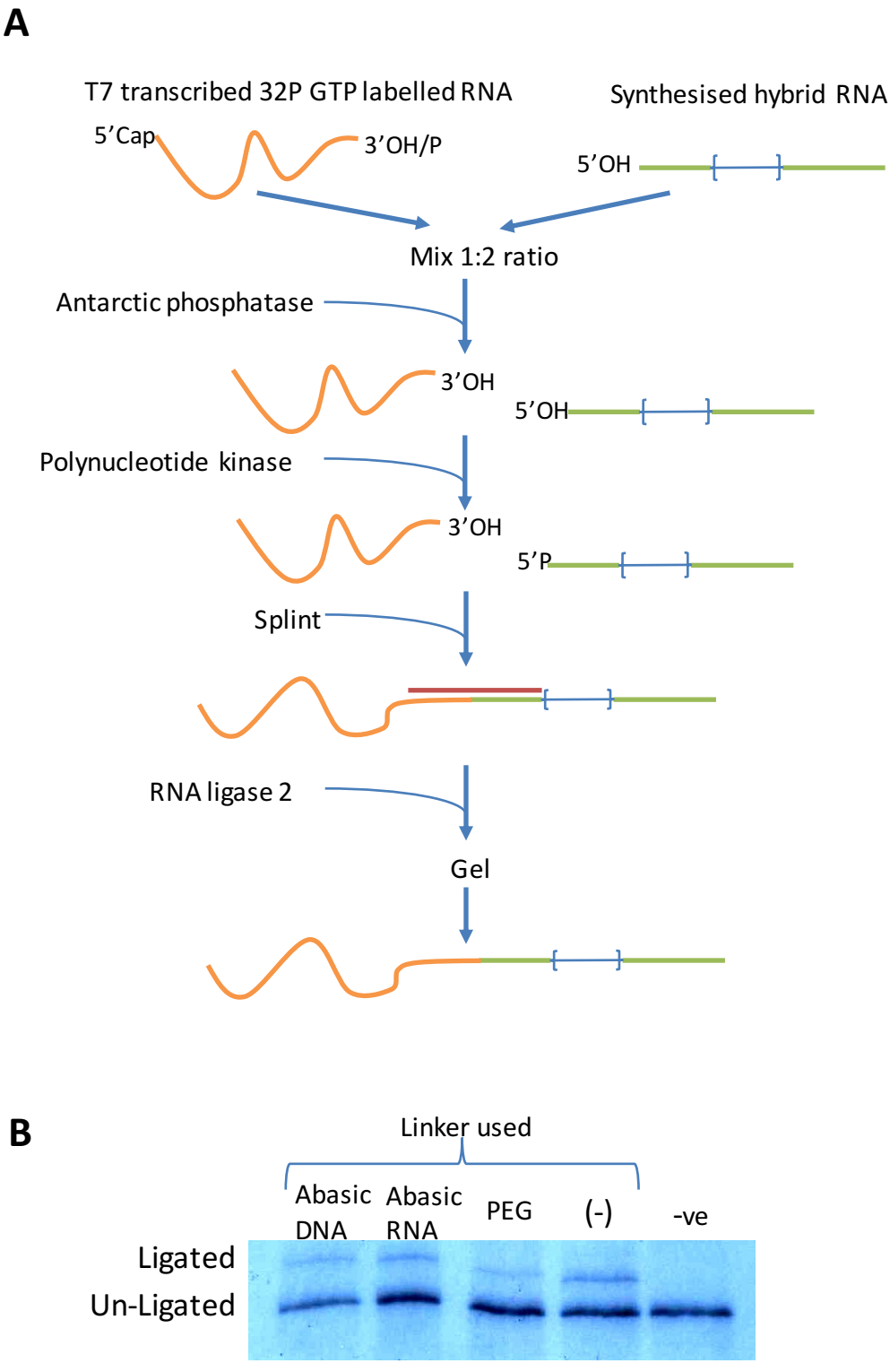
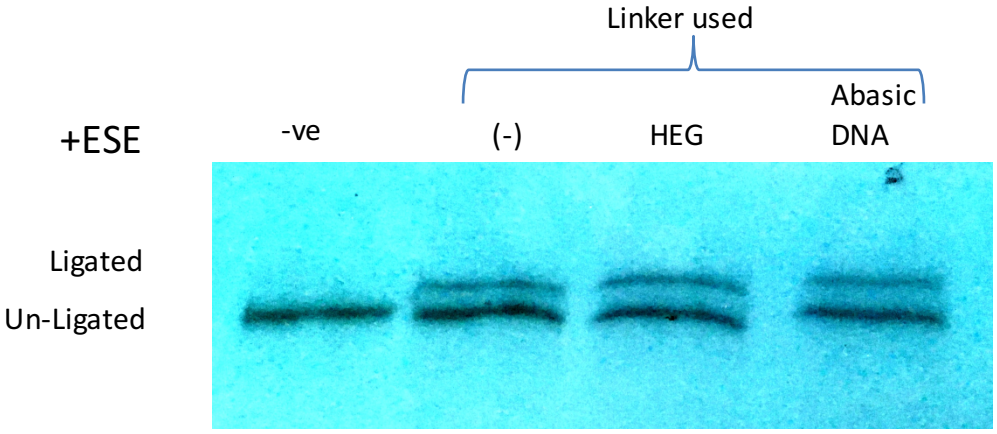


Figure 47. Schematic showing the splinted RNA ligation procedure. (A) Cartoon representation of the RNA ligation procedure used. (B) Example radiograph showing the ligated and un-ligated bands.



+ESE

Linker	Ligation efficiency (%)
(-)	32.44
HEG	33.56
Abasic DNA	37.4

Figure 48. Example of the level of ligation efficiency used following the final procedure used.

5.4. A Strong ESE functions by looping to its target site.

Following the development of the system outlined previously, hybrid RNA strands where constructed and tested in in vitro splicing assays. The first element which was tested was a potent 3' ESE.

For this we used the HEG, abasic RNA and abasic DNA linkers. The HEG linker should be at least as flexible as RNA, since its length of ~2.2nm is about four times the persistence length (Knowles et al. 2011) whilst an unfolded RNA chain of 12 nucleotides would have a contour length of ~7nm which is only 3.4 persistence lengths (Caliskan et al. 2005). Therefore it would easily be able to transfer signals via a looping mechanism but it lacks the charged backbone and the bases that characterise RNA, so would not be able to transfer any signals via mechanisms that relied on the RNA i.e. sliding or propagation models. The abasic DNA may be close to the flexibility of RNA and so again allow looping, but it would prevent some forms of propagation due to its deoxy ribose backbone and lack of bases. The abasic RNA contains the same charged backbone but no bases, so allowing looping and propagation via the backbone but no propagation via the bases. In combination, the three linkers should indicate whether the nature of the link between an ESE and its target is important and if so which property.

Each linker was synthesized with a portion of the 3' exon of SMN2 exon 7 at the 5' end and two copies of the potent Ron ESE (ESE-A) at the 3' end. Two copies of the ESE were chosen so as to maximise the efficiency of synthesis, which decreases with increasing length, and so that the change in splicing due to the effect of the ESE was significant. The linkers were attached, by the ligation system outlined above, to the body of the BG-

SMN2 exon 7 transcript. The distance from the 3' splice site to the ESE was maintained by deleting twelve nucleotides from the exon. The sequence removed contained the Tra2β enhancer region, which has no effect on splicing in the presence of an additional potent SRSF1-dependent enhancer (Smith et al. 2014)(p = 0.12, figure 49).

Splicing assays with each assembled construct showed that the ESE could function effectively over the abasic RNA and HEG linkers but not over the abasic DNA (Figure 50). It is possible that the abasic DNA may be binding superfluous proteins that inhibit the enhancers function; this hypothesis is explored in detail later in the chapter. Whilst the effect of the ESE was significantly decreased compared to the control in the other two cases, the splicing after 2 hours was ~5 times higher than cases with no ESE. One explanation for the decreased level of splicing is the proximity of the linkers to the ESE. Previous studies have shown that enhancer sequences function better and bind SRSF1 better when they are found in an extended non-specific RNA chains (Cho et al. 2011). Given the proximity of the ESE to the linker, it is possible that the linkers are decreasing the ESEs potency in this manner.

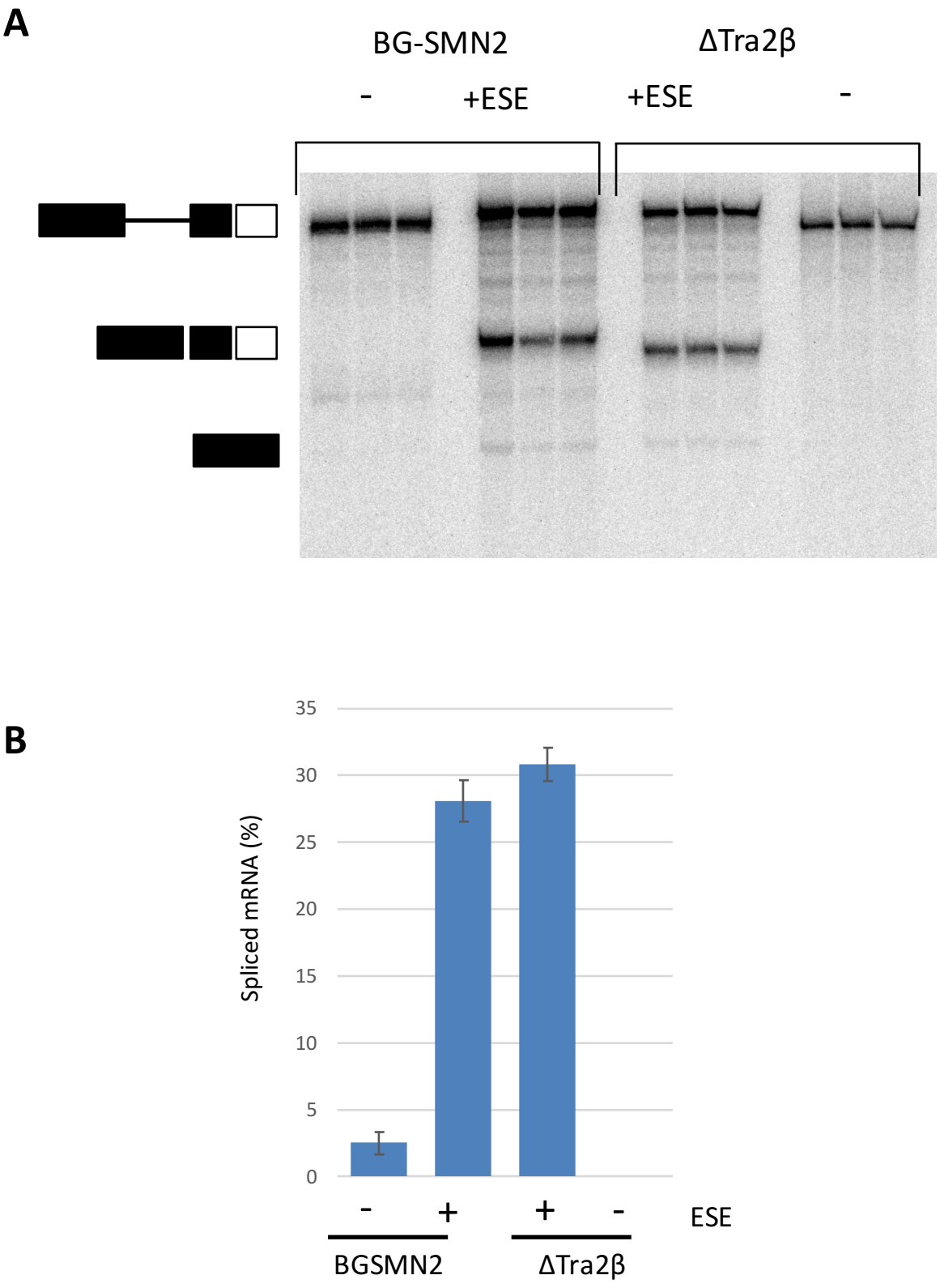
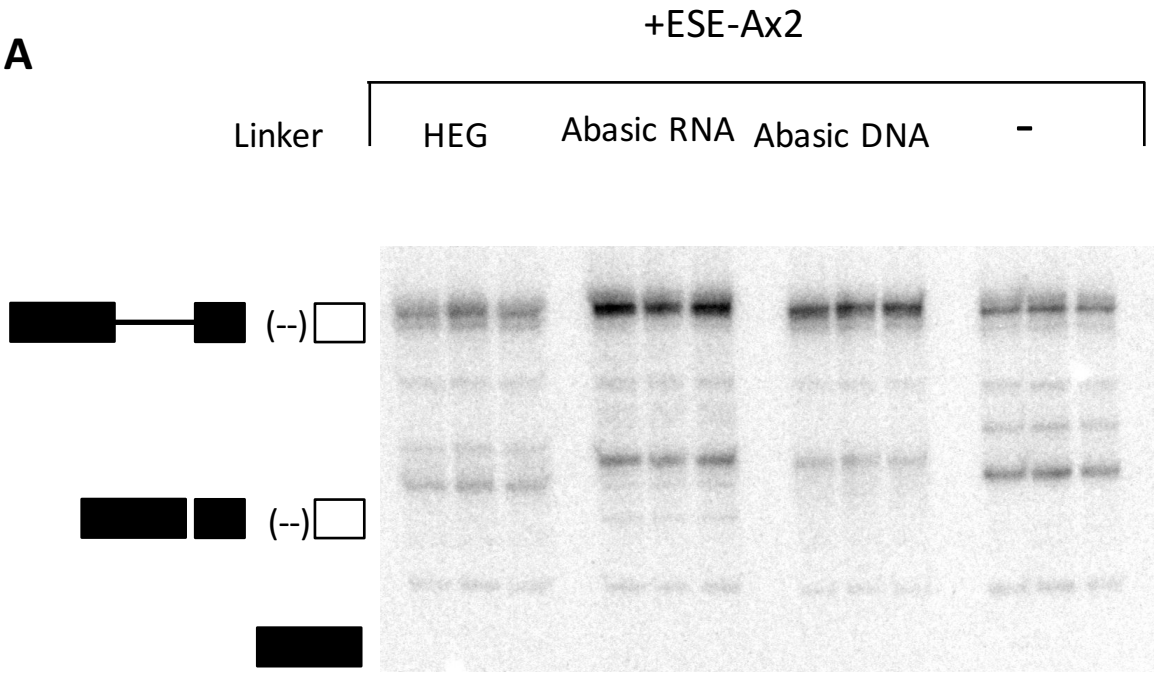


Figure 49. Figure showing the splicing of the BG-SMN2 hybrid transcript +/- the known Tra2β binding site and plus and +/- two copies of the Ron ESE. (A) Triplicate splicing assay after incubation for two hours. (B) Quantification of splicing after two hours.



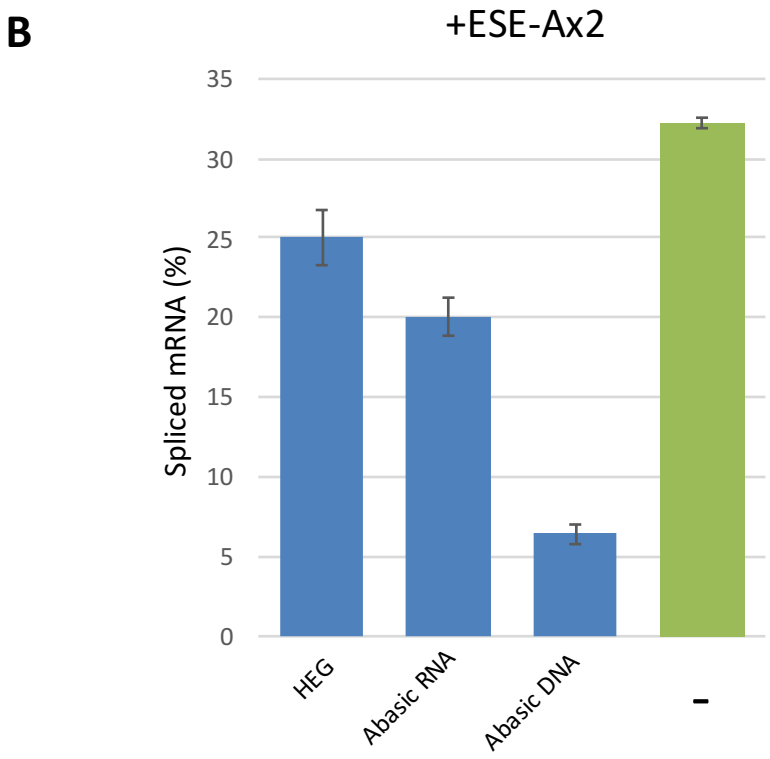
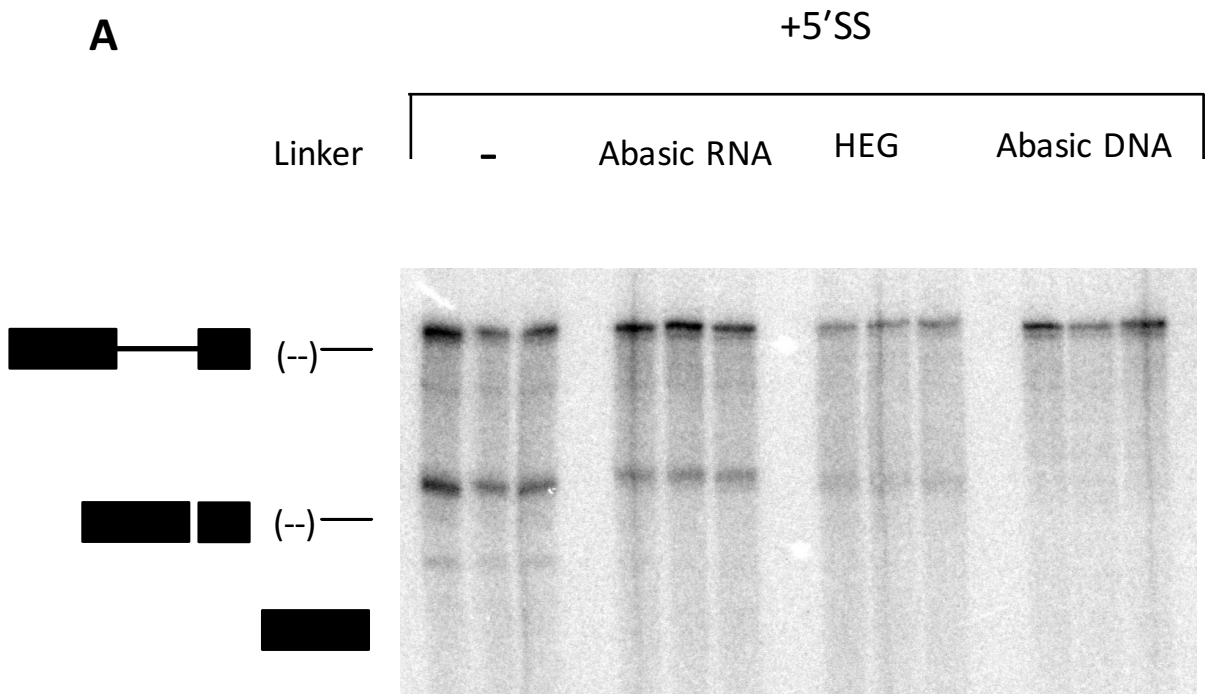


Figure 50. Figure showing the splicing of the ligated constructs containing one of the three linkers or no linker before 2 copies of the Ron ESE. (A) Triplicate splicing assay showing splicing after two hour incubations. (B) Quantification of splicing after two hours.

5.5. The activity of a downstream 5'SS across non-RNA linkers.

Another RNA element that enhances the use of an upstream 3' splice site is a strong U1 binding site. As described in chapter 3, the presence of this site recruits a U1 snRNP which in turn recruits a molecule of SRSF1. To test whether this mode of enhancement also functions regardless of the composition of the RNA the above system was again used.

The BG-SMN2 hybrid construct with the twelve nucleotide deletion was used and all three non-RNA linkers were also used. However this time each of the linkers had a ten base consensus 5' splice site on their 3' side in place of the ESE. Ligation proceeded as described previously and once again the assembled RNA constructs were tested in splicing/g assays. Similarly to the ESE, the abasic RNA and HEG linkers whilst limiting the effect of the U1 binding site, still allowed a strong stimulatory effect to be exerted. The abasic DNA again blocked any stimulatory effect. These results indicate that the 5' splice site bound U1 loops to interact with the 3' splice site complex via the recruited SRSF1. The reduced level of effect may indicate that the RNA immediately preceding or following the 5' splice site is important for SRSF1 mediated enhancement by U1.



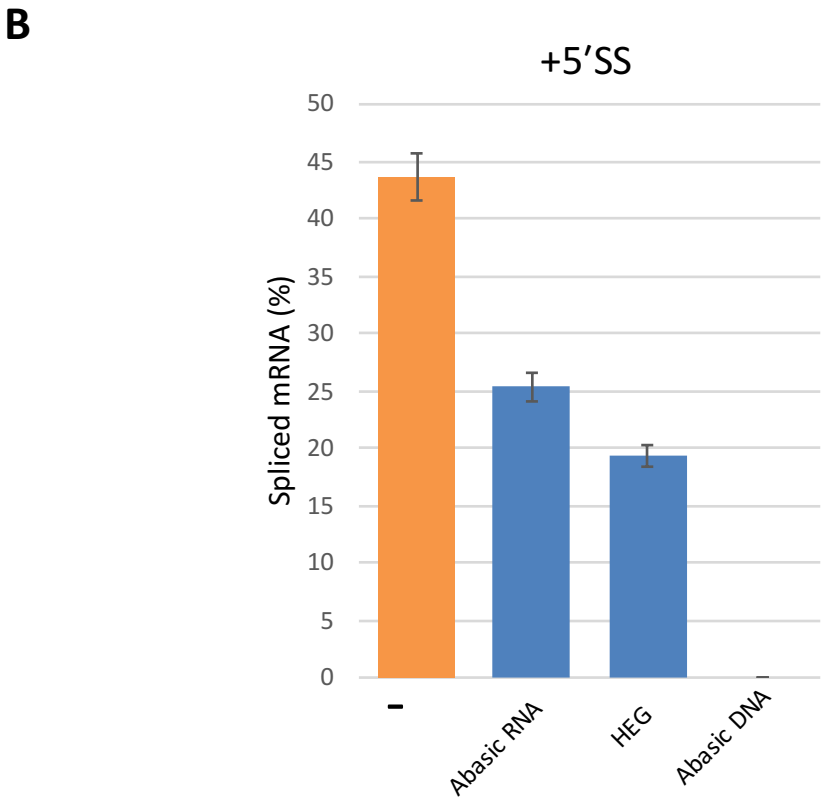


Figure 51. Figure showing the splicing of the ligated constructs containing one of the three linkers or no linker before a 3' consensus 5' splice site. (A) Triplicate splicing assay showing splicing after two hour incubations. (B) Quantification of splicing after two hours.

5.6. Analysis of superfluous binding to non-RNA linkers.

One possible explanation of the behaviour of the abasic DNA linker is that it was bound by proteins that prevented free diffusion. To get insight into the behaviour of the linkers used for these strands, four short sequences consisting of only one base, a block of one of the three non-RNA linkers or the ESE and biotin were synthesized. The compounds were purified by RP-HPLC or by binding affinity and used for affinity purification from nuclear extracts.

The protein mix pulled down was first analysed via PAGE and silver staining. Figure 52 shows that the protein mixtures are clearly different with a significantly darker band appearing between 25 and 30 KDa in the ESE sample and a darker band appearing around 37KDa in the abasic DNA sample. The band in the ESE sample could relate to a number of SR proteins, SRSF1 is 28 KDa, whilst the band in the abasic DNA sample could fit to the known abasic DNA binding protein APEX1, 35 KDa. In order to analyse the proteins in more detail the samples were sent for mass spectrometry.

The list of proteins recovered from the beads was refined by removing any examples in which a protein was more abundant in the control (no ligand) compared to all of the samples (abundance was assessed using scaffold Top3TIC). Subsequently, non-RNA/DNA binding proteins were removed and the abundances of the remaining proteins in the ESE or linker affinity purifications were compared (Figure 53). In most cases, the protein was bound most abundantly to the RNA oligonucleotide compared with the linkers. Unsurprisingly, these cases include a range of RNA binding proteins, including a number of SR proteins and hnRNP proteins as well as other splicing factors. The binding of the archetypal SR protein SRSF1, which mediates the activity of this ESE

(Ghigna et al. 2005), is displayed in figure 53A and shows that it binds preferentially to the ESE compared with the linkers. However, the binding of SRSF1 was still low in comparison to other proteins found, although this is a known feature of SRSF1 binding, which has been shown previously to be hard to detect by affinity purification (Smith et al. 2014; Staknis & Reed 1994). Although it was still the highest SR protein present. The HEG linker bound the least number of proteins, 20, of which none were specific to the HEG sample and only one was slightly enhanced in the HEG sample, DDX46. The abasic RNA sample bound 22 proteins after the refinements of which only one was not found in the other samples, SREK1, and only one other was enhanced over the other samples, NEDD8. In the case of the NEDD8, the level of protein was roughly similar to the RNA sample suggesting it may have an affinity for the charged backbone, which may relate to its role in DNA repair. The level of SREK1, on the other hand, was very low. The abasic DNA sample also pulled out 22 proteins of interest, of which only one was not found in the other samples, but nine of them showed preferential binding to abasic DNA. The most notable of these are APEX-1 and FEN1. These are two proteins that are known to function during DNA damage repair processes and, as can be seen in Figure 54B and C, are far more prevalent in the abasic DNA sample. This suggests that the abasic DNA may be able to recruit these proteins when in nuclear extract and these may account for the inhibition of splicing witnessed.

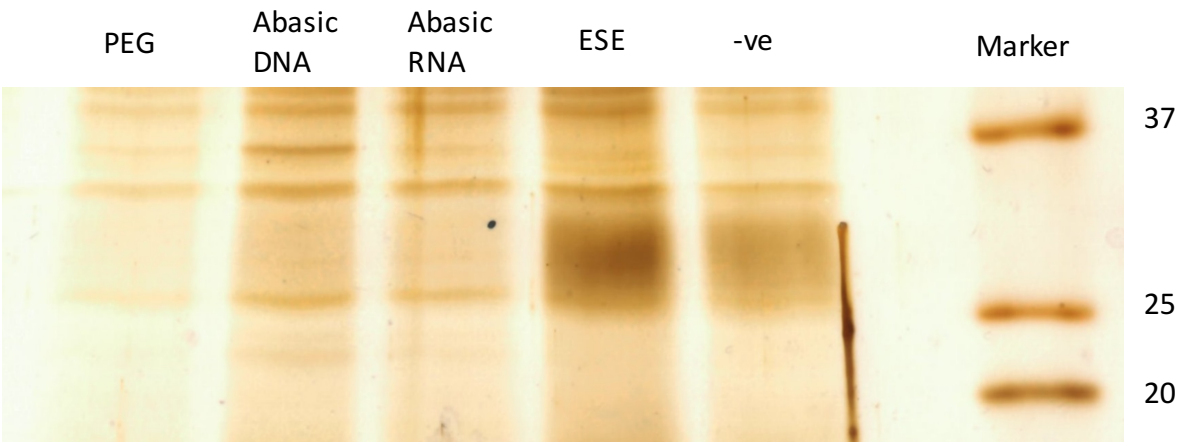


Figure 52. Silver stained polyacrylamide gel showing the proteins pulled down by each of the biotinylated non-RNA strands and the biotinylated ESE as well as size markers to allow identification of proteins.

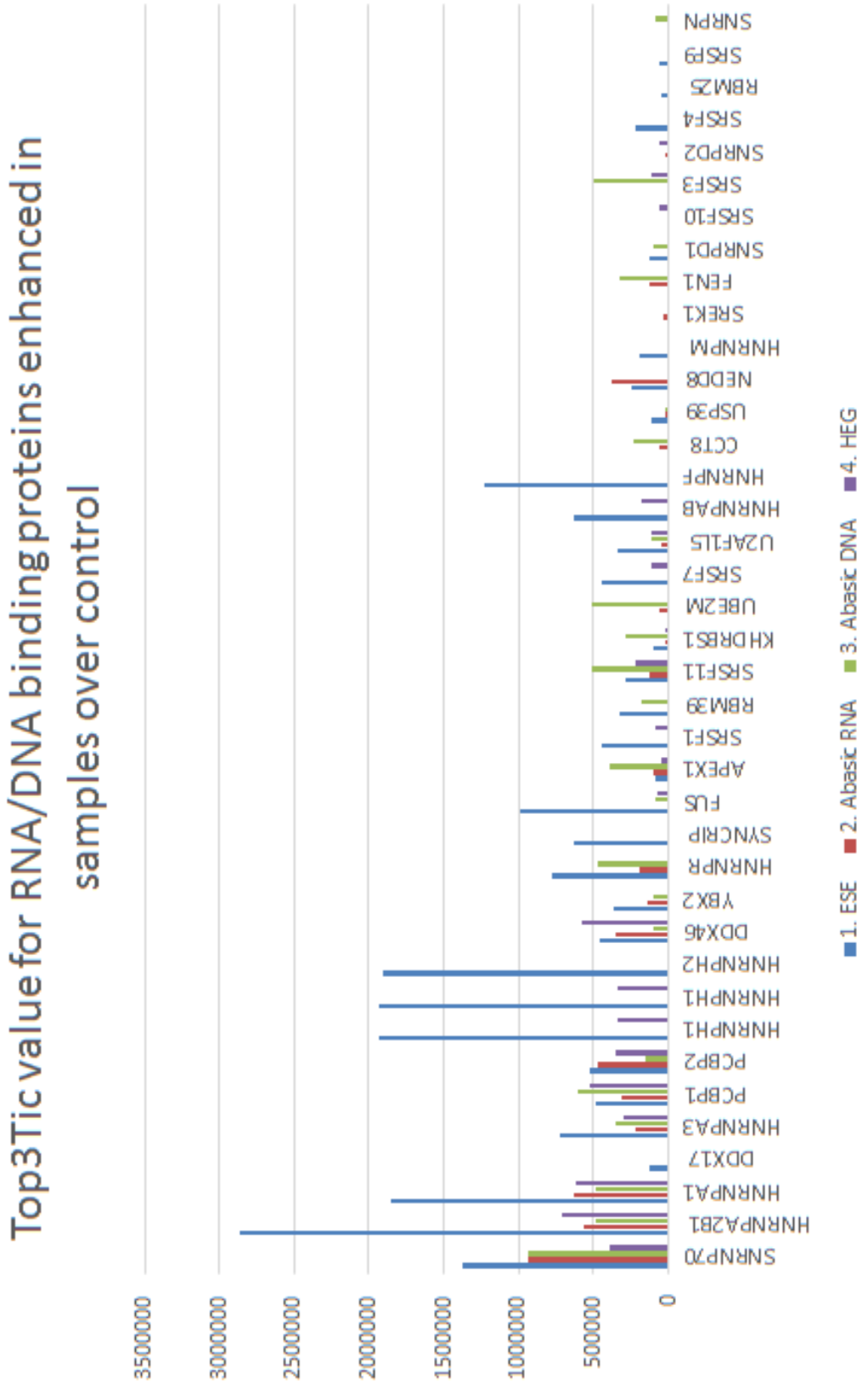


Figure 53. Graph showing the proteins pulled down and their relative quantities compared to one another.

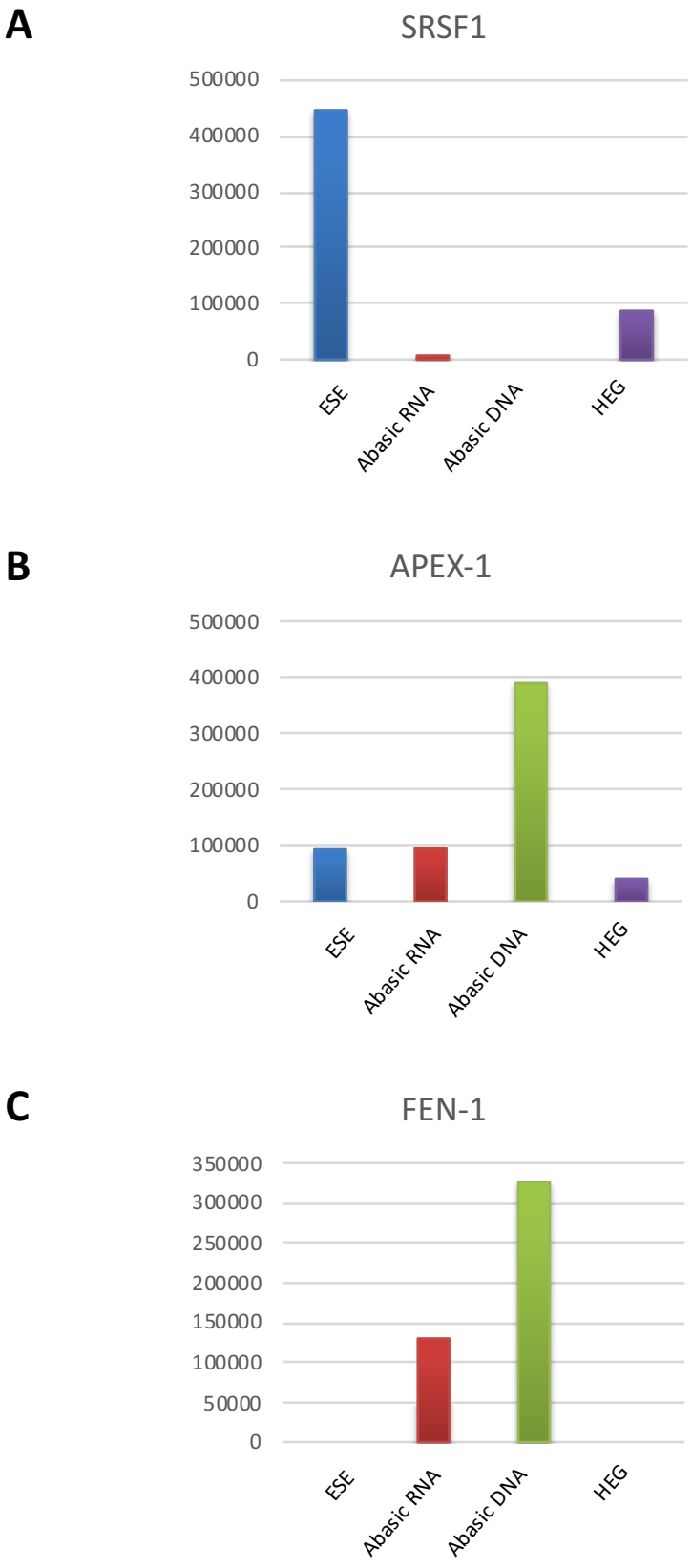


Figure 54. Graphs showing the relative binding of select proteins to each ligand. (A) The binding of SRSF1 to each ligand. (B) The binding of APEX1 to each ligand. (C) The binding of FEN1 to each ligand.

5.7. Summary.

The work described here describes a methodology for the synthesis and attachment of RNA molecules to include stretches of non-RNA strands in a larger RNA molecule. Furthermore we outline preliminary results which show two of the three linkers tested still allow a 3' ESE or U1 binding site to exert an enhancing effect on the upstream 3' splice site. Additional experiments looking at the binding patterns of each of the linkers reveals that the abasic DNA linker has the potential to bind non-splicing factors effectively, which may inhibit splicing.

Whilst the synthesis of hybrid RNAs and the ligation of RNA molecules have both been described previously, their effective use in conjunction preciously has been limited. This combination provides an effective system to analyse how two points connected by a strand of RNA can interact with one another. Furthermore this approach removes a number of the flaws from previous studies i.e. the need to tether proteins or introduce chemical groups into the RNA (Lewis et al. 2012; Shen et al. 2004; Graveley et al. 1998b).

In the example of an ESE, two possible models have been proposed previously, the looping of the RNA via 3D diffusion or the propagation of proteins along the RNA to send the signal. These two models, using this system are thus easily differentiated. It can be seen that the ESE is still able to exert an effect over the HEG or the abasic RNA linkers. However in both cases the level of splicing after two hours is significantly less than in the absence of a linker. One explanation for this is that the proximity of the linkers to the ESE may serve to limit enhancer protein binding to at least the initial parts of the ESE. Experiments performed by Cho et al have showed previously that the enhancer protein SRSF1 binds preferentially to its consensus site when it is found in extended RNA

sequences (Cho et al. 2011). Therefore potential binding to the first ESE in the pair may be limited and thus the effect seen reduced.

Another problem in the RNA splicing field is that of exon vs intron definition. In chapter 3 there are significant results that seem to help elucidate how an exon definition system may function and more specifically how a downstream U1 binding site can enhance the use of an upstream 3' splice site. The technique used here along with the preliminary results once again allow us to look at how these two separate sites might interact with each other. It can be seen that the U1 site can exert its enhancing effect across the abasic RNA and HEG linkers. Once again, though, the effect is significantly less than with no linker. One possibility is that this SRSF1 recruited by the U1 snRNP may need to make contact with the RNA around the 5' splice site, which is limited in this case, in order to exert an enhancing effect.

The one linker that in both cases does not allow either the ESE or the 5' splice site to exert their stimulatory effects is the abasic DNA linker. In order to look at why this linker behaves differently from the other two, the binding pattern of each linker was analysed via pulldowns and mass spectrometry. An ESE sequence was also used and showed a binding pattern that one would expect. In the cases of the HEG and abasic RNA then the binding patterns showed no binding preferences that were surprising or significant thus explaining why they were able to allow the ESE and U1 site to exert there effects. However in the case of the abasic DNA sequence two notable proteins were found. These were FEN-1 and APEX-1, both of which are known DNA damage repair proteins. This binding pattern suggests that the abasic DNA may bind these proteins in such a manner that may inhibit the enhancing signal to be passed and thus splicing inhibited.

These results together show that this strategy can provide a useful tool for looking at how two separate sites in RNA can communicate. Further experiments using different stretches of RNA between the ESE/U1 binding site and the linker may allow for more conclusive arguments to be drawn, whilst the application of this system to other situations has great promise. One such situation may be the silencing of exons that have intronic silencers flanking them. Additional situations outside of splicing may also be applicable. However we also show here that the use of abasic DNA stretches, at least for experiments conducted with RNA and in nuclear extract, are not viable.

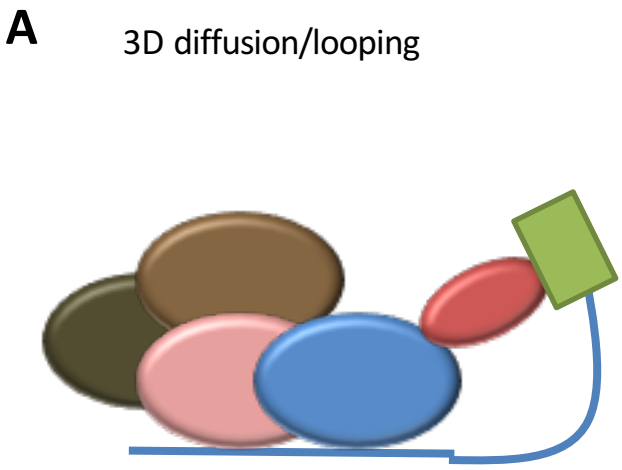


Figure 55. Model showing the correct method for how an ESE or 5' splice site, which enhances the use of an upstream 3' splice site, contacts its target.

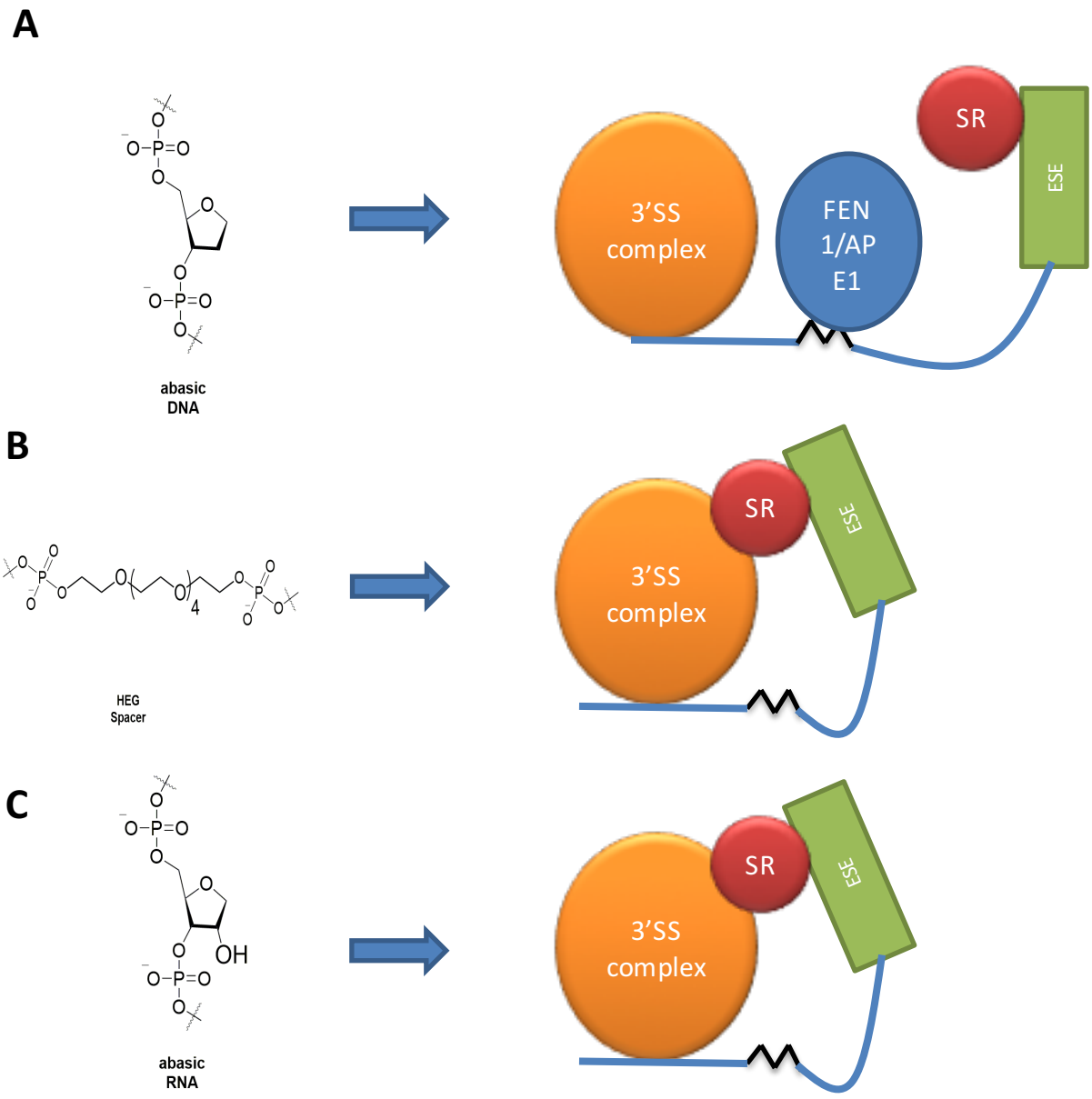


Figure 56. Models showing the effects of each of the linkers on the activity of the ESE. (A) The abasic DNA binds non splicing factors which interrupt the signal from the ESE being passed along. (B) and (C) the HEG and abasic RNA linkers allow the ESE bound SRSF1 to travel freely through space and contact the target.

Chapter 6. A single molecule look at SMN exon 7.

6.1. Introduction.

In conducting the experiments in chapter 3-5 a number of experiments where undertaken that involved the use of RNAs derived from SMN exon 7. This exon has been extensively studied and targeted by therapeutics but an exact understanding of the control of the exon is still incomplete.

SMN1 exon 7 is nearly entirely included in cells but SMN2 exon 7 is nearly completely excluded resulting in a non-functional protein being translated (Kolb & Kissel 2015). This dramatic change however is simply driven by a single silent mutation in exon 7 that mutates a C to T (Lorson et al. 1999). A number of theories have attempted to explain why this single nucleotide change drives such a dramatic shift in splicing. The first is that the mutation abolishes a binding site for SRSF1 (Cartegni et al. 2006). The second is that the mutation causes the creation of an hnRNPA1 binding site (Kashima et al. 2007). Whilst both are supported by a number of experiments, neither has been excluded or concretely proven. What most studies do agree on is that the exon has a particularly weak 3' splice site that needs stabilising for splicing to occur (Martins de Araújo et al. 2009). This is either by the SRSF1 enhancer, which is lost in SMN2, or the downstream 5' splice site, which may be repressed by the hnRNPA1 binding site in SMN2. Another motif within the exon that has been extensively studied is a proposed binding site for Tra2β. The importance of this site is contentious, however, with some studies showing that it has a significant effect on the recruitment of U2AF65 whilst other showing that it

can be blocked and no notable effect on splicing observed (Martins de Araújo et al. 2009; Owen et al. 2011).

Therapeutic strategies have mostly focussed on targeting the mutated ESE, the downstream ISS or the 5' splice site itself. One strategy involves using tailed oligonucleotides (TOES) that had a complementary region to the exon and an ESE in its tail. These TOES strategies aimed to increase splicing by increasing recruitment of SRSF1 to the exon and thus stimulating the 3' splice site (Owen et al. 2011; Smith et al. 2014). The next strategy involves using oligonucleotides that block a hnRNPA1 binding site in the downstream intron. This strategy aimed to prevent the hnRNPA1 site, possibly in conjunction with the site created in the exon, from suppressing the use of the 5' splice site (Rigo et al. 2014). The final strategy involved using small molecules which stabilise the binding of the U1 snRNP to the 5' splice site in a particularly context specific manner (Palacino et al. 2015). This strategy stabilises U1 binding which in turn stabilises the upstream 3' splice site.

Despite a number of therapeutics that shift splicing being developed, it is still not known exactly how the SNP drives the changes to splicing that we see. Here, a single molecule approach has been used to look at the binding of the key factors SRSF1 and U1.

6.2. Does the C>T mutation in SMN exon 7 directly alter the recruitment of SRSF1 in early splicing complexes?

In order elucidate whether the SMN1>2 mutation results in the loss of an SRSF1 ESE, the mEGFP-SRSF1 nuclear extract from chapter 3 was used. The BG-SMN2 transcript was again used except this time we used SMN 1 exon 7 (BG-SMN1) as well as the SMN 2 version used previously. These two transcripts were tested in a number of complexes in order to look at whether SRSF1 was recruited during any of the early complexes. The complexes tested were E complex, A complex and I complex (Chen et al. 2016).

The first complex that forms during in vitro splicing assays is complex E. This complex forms in the absence of ATP and precedes A complex. To stall complex progression at this point, ATP is omitted during the creation of the splicing mix. In figure 57A and B it can be seen that in both cases, BG-SMN 1 and 2, that both transcripts having a non-specific geometric pattern of binding. This occurs due to de-phosphorylated SRSF1 or due to transcripts being unable to form complex A effectively (Weinmeister 2015)(Chapter 3 and 4). In this case due to the omission of ATP, it is likely that the proteins are de-phosphorylated so this explains the pattern seen in both cases. BG-SMN1 transcript does show enhanced binding of one or two molecules of SRSF1 but this is hard to interpret due to the background binding.

In order to try to look at a pre-A like complex, where the binding pattern isn't masked by non-specific binding by de-phosphorylated proteins, complex I was looked at. This complex forms in the absence of ATP but when proteins are phosphorylated. This is

achieved by using phosphatase inhibitors. Previous work has shown that in this complex there is stoichiometric recruitment of 3' splice site factors and that the U2 snRNA is base paired to the branch point (Chen et al. 2016). Figure 57C and D show that under these conditions BG-SMN2 binds a single molecule of SRSF1 whilst BG-SMN1 has a strong upsurge in the number of complexes with two molecules bound. By using the anti-U1 oligonucleotide described previously we can show that the single molecule recruited by BG-SMN2 transcript and one of the molecules in the BG-SMN1 transcript stem from the U1 snRNP bound at the 5' splice site of the Globin exon, figure 60A and B, as discussed in chapter 3.

The last of the early splicing complexes that have been analysed is complex A. This forms in the presence of ATP and we can ensure complex progression does not continue by using an anti-U6 oligonucleotide which blocks tri-snRNP association. Under these complex conditions the patterns of BG-SMN 1 and 2 (Figure 60C and D) are highly similar with a peak of complexes that contain a single molecule followed by a background binding of RNAs with multiple proteins bound. As the proteins in these conditions are likely to be phosphorylated, the background binding here is likely coming from the transcripts inability to form complex A effectively. This is to be expected as both RNAs splice very weakly.

These results seem to indicate that the C>T mutation from SMN1>2 results in the removal of an ESE that can recruit SRSF1 in early complexes but not efficiently or strongly enough for the bound protein to survive in to A complex or even promote A complex formation itself, as is seen with the Ron enhancer in chapters 3 and 4. This fits

with the in vitro splicing results which show that BG-SMN 1 splices no more efficiently than BG-SMN2.

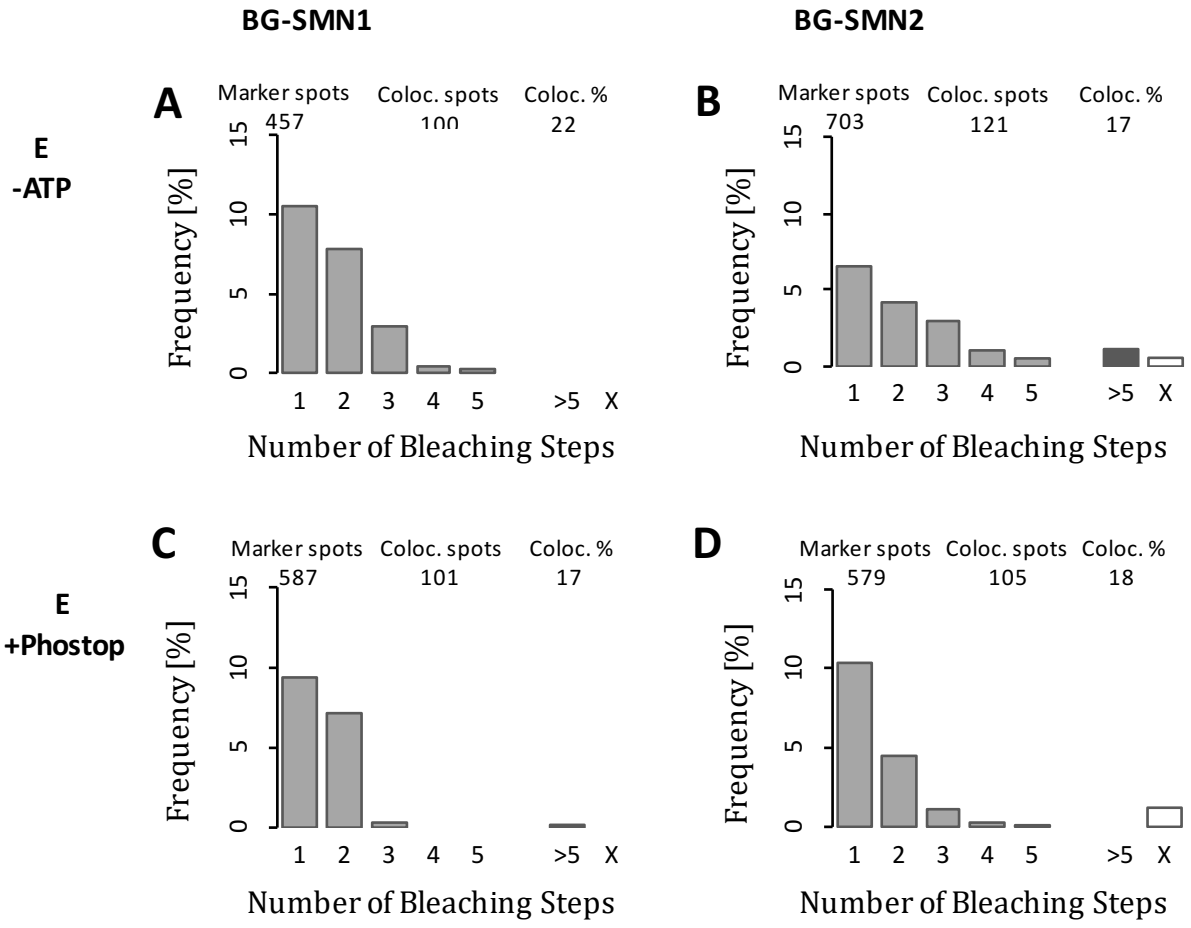


Figure 57. Single molecule experiments analysing the binding of SRSF1 to BG-SMN1/SMN2 hybrid RNAs. (A) mEGFP-SRSF1 binding to BG-SMN1 under E complex conditions (minus ATP). (B) mEGFP-SRSF1 binding to BG--SMN2 under E complex conditions. (C) mEGFP-SRSF1 binding to BG-SMN1 under I complex conditions (minus ATP, plus Phos-stop). (D) mEGFP-SRSF1 binding to BG-SMN2 under I complex conditions.

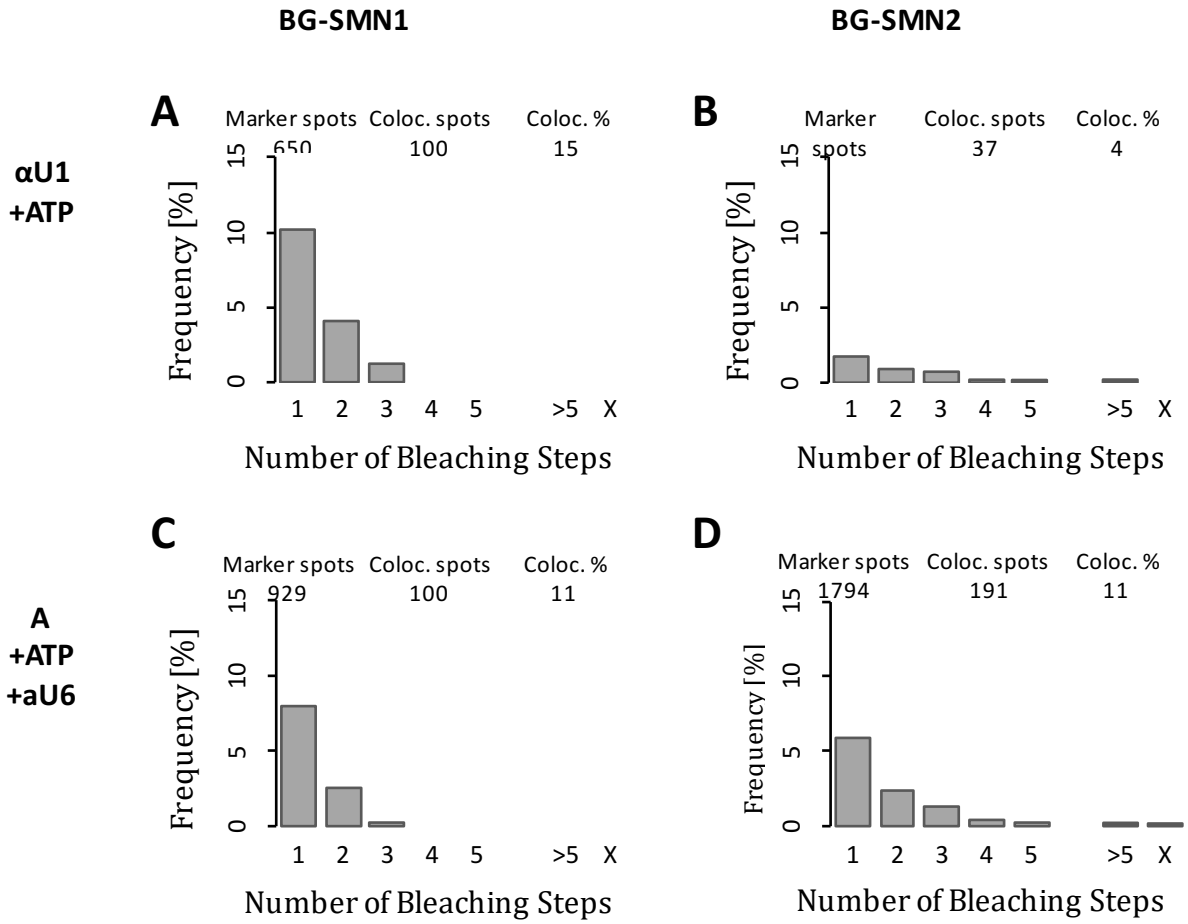


Figure 58. Single molecule experiments analysing the binding of SRSF1 to BG-SMN1/SMN2 hybrid RNAs. (A) mEGFP-SRSF1 binding to BG-SMN1 under A complex conditions (plus ATP) plus the anti-U1 Oligo. (B) mEGFP-SRSF1 binding to BG-SMN2 under A complex conditions plus the anti-U1 oligo. (C) mEGFP-SRSF1 binding to BG-SMN1 under A complex conditions (plus ATP). (D) mEGFP-SRSF1 binding to BG-SMN2 under A complex conditions.

6.3. Does the C>T mutation in SMN exon 7 directly alter the recruitment of U1 in A complex.

To look at the recruitment of U1 snRNPs to SMN exon 7, a three exon transcript consisting of the SMN 1 or 2 exon 7 cloned in between Beta-Globin exon 2 and 3 was used (BG-SMN1-BG and BG-SMN2-BG). This transcript has three 5' splice sites; a consensus site at the ends of Globin exon 2 and 3 and the SMN 5' splice site under study. A three exon transcript was preferred over a two exon transcript so as to include the intronic hnRNPA1 binding site which is reported to supress U1 binding. The nuclear extract described in chapters 3 and 4 containing GFP-SRSF1 and mCherry-U1A was used.

Figure 59A and B show the binding of U1A to the BG-SMN1/2-BG. BG-SMN2-BG shows a strong spike in the number of complexes with two molecules of U1 bound, presumably from the two Globin sites, whilst BG-SMN1-BG shows less of a spike with two bound but with more complexes with a third molecule bound presumably from the SMN 5' splice site. The binding of SRSF1 was also looked at in this experiment (figure 59C and D). Once again, as expected from the results in chapter 3, the binding pattern of SRSF1 mimics that of U1. These results support the idea that the mutation in SMN 2 results in decreased U1 recruitment to the exon.

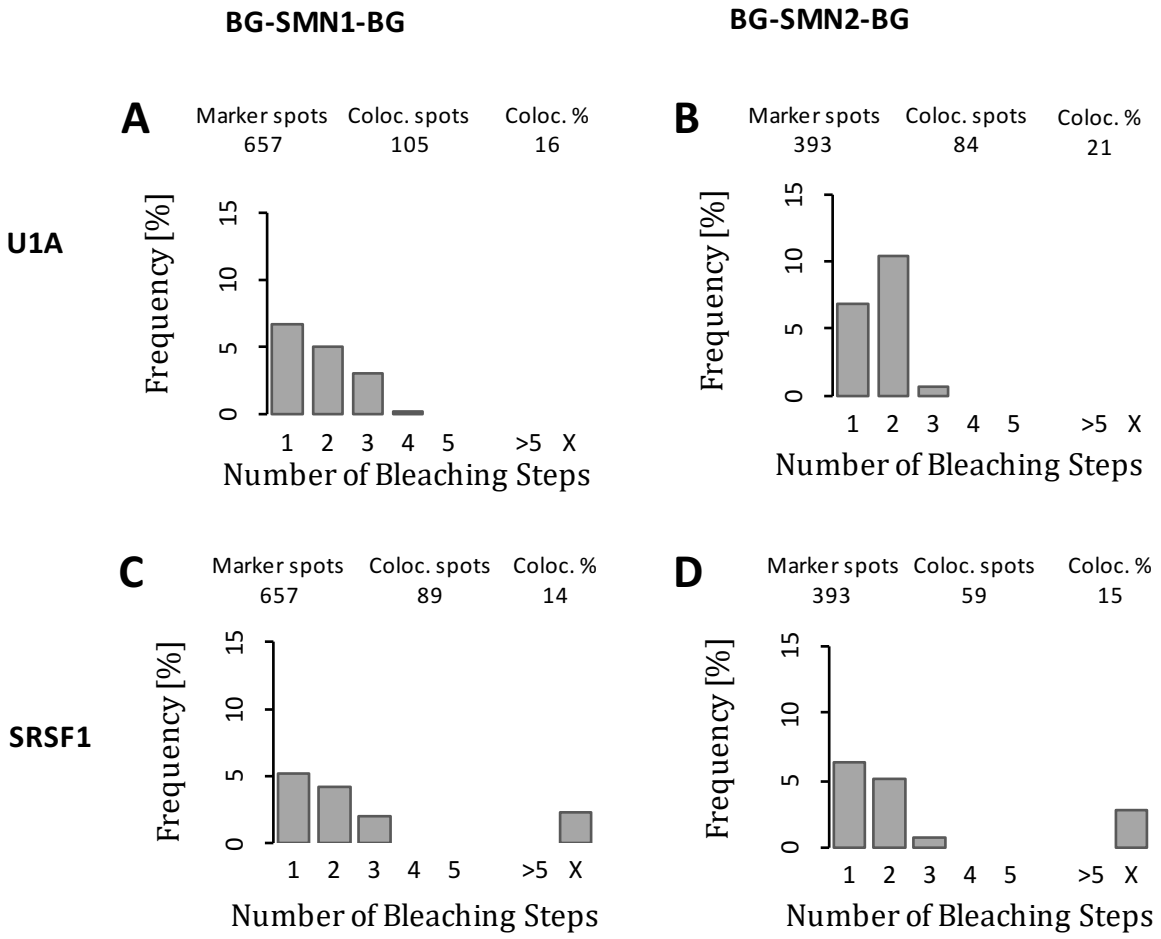


Figure 59. Histograms showing the binding pattern of mCherry U1A and mEGFP SRSF1 to 3 exon RNA constructs consisting of SMN1/2 exon 7 in between exon 2 and 3 of Globin. (A) Binding of U1A to BG-SMN1-BG under A complex conditions. (B) Binding of U1A to BG-SMN2-BG under A complex conditions. (C) Binding of SRSF1 to BG-SMN1-BG under A complex conditions. (D) Binding of SRSF1 to BG-SMN2-BG under A complex conditions.

6.4. Summary.

These results, whilst not offering any new striking discoveries, do serve to support prior ideas and explain some of the confusion that has surrounded SMN exon 7. We observe here that the SMN 1 exon 7 can recruit a molecule of SRSF1 transiently but this does not survive into A complex and does not seem to be sufficient to promote splicing on its own. Furthermore we show that SMN 1 can recruit U1 better than SMN 2 and that this results in an additional SRSF1 being recruited.

A possibility that is supported by this work is that in SMN 1 there is some transient binding of SRSF1 to the exon but it is not significant for splicing activation and what actually activates SMN splicing is the binding of the U1 snRNP to the exon which recruits SRSF1 strongly. Whilst in SMN 2 the mutation creates an hnRNPA1 binding site which can function along with the downstream site in the intron to repress the recruitment of U1 to the exon and thus inhibit splicing. This explanation fits with why therapies that either block one of the hnRNPA1 binding sites or strengthen U1 binding are effective (Palacino et al. 2015; Zhao et al. 2016). The transient SRSF1 binding and the previously uncharacterized recruitment of SRSF1 by U1 snRNPs also likely explains why a number of studies find SRSF1 binding to the exon. However we postulate here that the SRSF1 that is significant for splicing activation is the one recruited by the U1 snRNP in A complex conditions.

If this hypothesis were correct then it would suggest that the reason SMN 2 gives such low levels of splicing is due to the ISS and ESS, that is created, working together to inhibit U1 recruitment. However, how hnRNPA1 suppresses splicing is still poorly characterised.

It is not known if the sites loop together to prevent binding by hiding the 5’ splice site or whether proteins propagate from one site to the next and thus blanket the intervening area. Single molecule studies using labelled hnRNPA1 could help serve to prove that the protein binds at the specified sites whilst further work using the system outlined in chapter 5 for creating hybrid RNAs could help to elucidate how the two sites interact with one another.

Chapter 7. Discussion.

7.1. Implications for the mechanism of enhancer function at the 3' splice site.

In order to assess the implications of the results observed and there implications for enhancer function, it is first necessary to look at the prior models and the experiments that led to their development. Following this, new models that build upon the old ones can be developed or new ones that oppose them postulated. Most models for the activation of splicing by ESEs resolves there action into two defined steps. The first is the binding of the enhancer protein itself and the second is the mechanism of the interaction between the enhancer protein and its target factor. Both parts of the model were tested indirectly by the Maniatis lab in 1998 in two key papers (Graveley et al. 1998b; Hertel & Maniatis 1998).

The first part of the model was tested using the *Drosophila* double sex (*dsx*) gene which contains a series of established 13 nucleotide enhancer sequences (*dsx*RE) that are found 300 nucleotides away from the 3' splice site which they activate (Hedley & Maniatis 1991). These enhancers were previously shown to bind the *Drosophila* enhancer proteins Tra, Tra2 and one other SR protein in a stable enhancement complex (Tian & Maniatis 1993). It was subsequently found that with these enhancers, based on inferences from rates of splicing, that there was concurrent occupancy of all sites. This led to the conclusion that the rate of splicing was determined by the probability of an interaction occurring between one of the enhancer bound protein complexes and the target factor (Hertel & Maniatis 1998).

This conclusion was supported by the second paper in 1998 from the Maniatis lab. This paper showed that a tethered RS domains potential for activation depended on the distance from its target site. This data also allowed the formation of a model for the second step. The data was found to fit to a model for the probability of interactions between two sites on a freely moving chain. It was concluded that after a stable protein had bound, that it then interacted with its target via an RNA loop (Graveley et al. 1998b).

However the most informative experiments for the second step of enhancer activity came in 2004 from Shen and Green. RS domains that were stably tethered to a 3' ESE were found to be able to be cross-linked near a 3' splice site. While these results support the idea of direct interactions or looping, the use of tethering meant that the contribution of binding to the rate-determining step could not be assessed. Furthermore, the numbers of SR proteins bound were not known so propagation could not be excluded. Thus, the models inferred from the rates of *Drosophila* Tra/Tra2-dependent splicing are still generally used to represent or interpret mammalian ESE activity.

In contrast to a spate of recent results describing the transcriptome-wide binding sites and protein interactions of proteins that activate or repress splicing, the molecular mechanisms by which they effect splicing have received little attention since the *dsx* experiments. Such investigations are not amenable to ensemble methods, which provide little information about the heterogeneity of complexes or the stoichiometry of components. This is particularly true of activator proteins, which exert effects at a distance from their binding sites. We sought in this work to use a combination of single

molecule imaging and chemical biology to test models for the binding and signalling activities of SRSF1 dependent enhancers.

The comparison of four different ESEs showed that they all stimulated splicing, but the Ron exon 12 sequence was by far the best. This might have been explained had it recruited more than one molecule of SRSF1, but in fact the single molecule results show that all the ESEs stimulate the recruitment of a second SRSF1 (the first binding via the U1 snRNP (Chapter 3)). The ESEs also suppress the geometric distribution of 3 4, 5 etc. molecules that is characteristic of RNAs which cannot form complex A. The fact we see only one additional SRSF1 is interesting as it indicates that there are no additional SRSF1 proteins propagating out from the initial site, at least in A complex. This however does not rule out the recruitment of other SR proteins or a heterogeneous complex. The percentage of co-localisation does not increase substantially for the stronger enhancers. The lack of correlation is most likely due to the first molecule of SRSF1 being recruited at the 5'SS regardless of the presence of an ESE. Together with the geometric distribution associated with complexes that fail to form complex A, this explains why most experiments attempting to quantify SRSF1 binding have been only been successful on short segments rather than full splicing units. The effect of the ESEs is clearer when we look at the proportion of RNA molecules associated with two molecules of mEGFP SRSF1. This correlates well with the splicing efficiency of each of the constructs. What this correlation allows us to establish is that splicing efficiency is directly linked to the ESEs ability to recruit the second SRSF1.

Whilst the data looking at the four different enhancers supported previous ideas surrounding enhancers, the data looking at multiple ESEs provided strong evidence for

a new mechanism. As expected we saw that additional repeats of the Ron ESE increased splicing sequentially in a near linear manner. An increase in splicing with additional ESE repeats could be accommodated for by at least four models, three of which involve the binding of additional molecules of SRSF1. However, we clearly saw no evidence for more than one additional SRSF1 molecule binding even with four copies of the ESE. There was a continuation of the trend witnessed with the four different ESEs, whereby the strongest splicing construct, with four Ron ESEs, gave the highest proportion of molecules bleaching in two steps. However in this this case there were substantial changes in the percentage of co-localisation between each construct, possibly because the level of binding initiated by the ESE was at a sufficiently high level to be detectable above the binding of the 5'SS-bound SRSF1 and the background. We conclude that each additional ESE increases the chance of a single binding event occurring. One explanation is that SRSF1 binds in a transient manner initially (Suhyung Cho et al. 2011; Anczuków et al. 2015) and it is only maintained on the RNA if an interaction occurs with an effector such as U2AF, U2 snRNP (Fu et al. 1992; Chen et al. 2016; Lindsay D. Smith et al. 2014; Lavigueur et al. 1993b) or an intermediary.

Two models have been proposed to explain how the binding of a single SRSF1 produces an effect on a target some 50 nucleotides upstream; (i) that the SRSF1 binds and interacts with its target via an RNA loop (Graveley et al. 1998a; Shen et al. 2004), and (ii) that the SRSF1 binds and stimulates the binding of additional proteins which propagate towards the target and exert the effect seen (Lewis, Andrew J. Perrett, et al. 2012). An additional model, in which SRSF1 prevents propagation of hnRNP complexes towards the splice site, does not apply here because of the location of the ESE. The single

molecule data excludes propagation of stable complexes of SRSF1, although it does not rule out heterogeneous protein complexes or propagation in other complexes other than A. To resolve this more definitively we created chimeric RNA constructs that contain stretches of non-RNA in between the potent Ron enhancer and the 3' SS. The non-RNA in question is either (i) PEG, which should stop propagation but allow RNA looping, (ii) abasic RNA, which should allow looping or propagation via electrostatic interactions but not via the recognition of bases, or (iii) abasic DNA, which should behave the same as abasic RNA but prevent backbone binding by RNA binding proteins. It is important that flexibility of the linkers is the same or more than that of RNA so as to rule out stiffness being a factor; abasic RNA and abasic DNA are assumed to be at least as flexible as RNA with a Kuhn statistical segment length of ~4nm whilst PEG is more flexible at 1.2nm (Knowles et al. 2011).

The abasic DNA linker initially appeared to suggest that the intervening strand is important whilst the abasic RNA and PEG suggested the opposite. Mass spec data looking at protein binding to each of the strands shows that the abasic DNA can in fact bind APEX-1 and FEN 1, known DNA damage repair proteins, which may actually either stiffen the strand or inhibit splicing factor binding by binding associated DNA damage repair proteins such as YB-1 or YB-2. The abasic RNA and PEG therefore indicate that the intervening strand is not important for an ESE to exert an effect as long as it is both flexible and does not recruit non splicing factors. However, the splicing of the constructs containing the PEG and abasic RNA is significantly less efficient than the construct with no linker (P=0.0045 and 0.00016 respectively). This may indicate that the presence of the non-RNA in close proximity to the ESEs could weaken SRSF1 binding. Future work

testing different lengths of PEG/abasic RNA and different lengths of RNA before the ESE but after the PEG/abasic RNA may help to answer this.

These results strongly indicate that intramolecular loop formation is crucial and thus they are consistent with previous inferences from site specific cross linking (Shen & Green 2004) or the effects of varying the distance between the 3'SS and the ESE (Graveley et al. 1998b). Work from this lab has shown previously the activity of an ESE upstream of an alternative 5'SS was blocked when it was attached via a short non-RNA linker using click chemistry. Further previous work on ESE-coupled nano-particles had shown that a triazide linkage at the 3' side of an ESE did not block activity, although it did at the 5' side. It was inferred therefore that the outcome was not the result of the coupling method but rather reflected interference with propagation by a non-RNA linker (Lewis, Andrew J. Perrett, et al. 2012). The current findings with a 3' ESE suggest that ESEs function differently at 5'SS than at 3'SS or that the introduction of the triazide link into the RNA did have an inhibitory effect. However, unlike abasic DNA, there is no obvious explanation by which a triazide group could interfere with looping, and we conclude that it is possible that 5' and 3' ESEs do work by different mechanisms. A significant difference is that the ability of SRSF1 to stimulate U1 snRNP binding does not require an RS domain (Eperon et al. 2000; S. Cho et al. 2011; X. Roca et al. 2013), whereas the function of a 3' ESE can be satisfied by tethering of the RS domains alone (Graveley et al. 1998b; Shen & Green 2004).

These results and conclusions serve to, in some aspects, support prior models for the activation of splicing but also reject previous models. The second step of ESE activity, where an enhancer bound protein interacts with its target, was previously postulated to

occur through the looping of RNA. This however was inconclusive due to the drawbacks and limitations of prior methods. The use of tethered proteins, considering other results outlined in this thesis, is clearly not appropriate, and in fact other data seemed to contradict this. Here we use a system which removes a number of these limitations and the limitation of the contradictory data. Our conclusions are that 3' ESEs, after binding proteins, do interact with their target via the looping of RNA through space. This is in agreement with the previous models.

However our data strongly opposes the previous models for the binding of proteins to ESEs. The original models, which were based on *Drosophila* proteins, suggested multiple ESEs recruit multiple proteins in stable complexes. Using single molecule methods, which allow us to monitor occupancy, we can clearly see that for a mammalian SRSF1 dependent ESE that this is not the case. This raises two possibilities, (i) that mammalian ESEs work differently to *Drosophila* ones or (ii) that different ESEs function differently depending on their corresponding factor. The first possibility would fit with data showing there is more alternative splicing humans (Barbosa-Morais et al. 2012); where a system whereby the expression of a particular protein can be fine-tuned to achieve a particular level of splicing may be beneficial. Whereas in an organism such as *Drosophila*, where alternative splicing is less prevalent, splicing may be controlled by whether the protein is simply present or not. The second possibility is perhaps more likely due to the wide ranging binding affinities and sequence specificities enhancer proteins have. It is possible that some proteins, such as SRSF1, may bind weakly initially so the number of copies of an ESE and the expression of the protein mean that alternative splicing can be controlled to variable levels. Whilst other proteins may bind stably, such as Tra/Tra2,

and simply switch splicing on or off. In order to explore these possibilities further and elucidate how the vast range of enhancer proteins in humans function, further experiments looking at additional SR proteins and ESE sequences are needed.

7.2. SRSF1 recruitment by U1 snRNPs and its implication on exon definition.

Whilst the recruitment of SRSF1 by ESEs is well established, its recruitment by U1 snRNPs is not. The positioning of a single molecule of SRSF1 at a 5' splice site by a U1 snRNP has interesting implications for both the splicing of the upstream exon or in the splicing of the downstream intron.

The effects on exon definition of a downstream 5' splice site over a short exon are widely reported (Hwang & Cohen 1996; Kreivi et al. 1991) and are similar to those seen for a potent ESE. Indeed, there is a negative correlation between the combined strength of splice sites in cassette exons and the density of ESEs within them (Anczuków et al. 2015). This would seemingly indicate that the presence of a strong 5' splice site may preclude the need for ESEs in an exon. In light of our results, ESEs and U1 snRNP binding may be alternative ways to recruit SRSF1 for exon definition. This possibility is strengthened by the current absence of any clear evidence for a direct involvement of U1 snRNPs in exon definition. While the presence of a downstream 5' splice site that recruits a U1 snRNP has been shown to produce an increase in U2AF65 binding to the upstream 3' splice site (Hoffman & Grabowski 1992), the functional interactions have not been identified. The previous solution for this problem is to assume that exon definition requires an ESE to recruit an SR protein to the centre of the exon, and that this molecule interacts with both the U1 snRNP and U2AF. This idea does not however fit with the apparent equivalence of either a strong 5' splice site or a dense number of ESEs shown by the negative correlation described previously. Instead, following our findings it seems likely

that in cases with a strong 5' splice site, and a low number of ESEs, that the U1 snRNP will recruit an SRSF1 which will interact directly with a 3' splice site factor, exactly as would be expected for an ESE-bound molecule.

This is supported by experiments here where the binding of the key 3' splice site factors was analysed. These showed that the downstream 5' splice site acted in a nearly identical manner to the addition of four portent Ron ESEs in terms of their effect of U2AF and U2 binding. What was perhaps the most interesting thing here was that the four repeats of the Ron ESE amounted to 48 nucleotides whilst the 5' splice site added here was only 10, indicating that whilst they may function in a near identical manner the stability of recruitment is distinct.

Direct interactions are also consistent with our finding that only one SRSF1 is recruited in complex A per 5' splice site and there is no evidence of cooperative association of additional molecules of SRSF1 across the exon. Furthermore we show, by introducing non-RNA linkers in between a 3' 5' splice site and its upstream target 3' splice site, that the enhancing effect of the site is still witnessed.

An alternative model for how a single SRSF1 molecule, recruited by a U1 snRNP, can bridge the 5' and 3' ends of an exon is revealed when the structure of SRSF1, and the interactions of its domains, is dissected. Previously, the possibility of SRSF1 being required for exon definition was limited by the apparent need for the presence of strong ESEs to be recruited. However, in light of the evidence presented here, SRSF1 may be recruited to the exon initially by U1 snRNPs and then make contact with the RNA and 3' splice site components subsequently. This prior stabilisation may mean the RNA

sequence does not need to be a consensus sequence which in turn may explain SRSF1s apparent spurious binding in CLIP experiments (Pandit et al. 2013a).

The contact with the 3' splice site factors is known to occur via the RS domain (Zhu & Krainer 2000; Lindsay D. Smith et al. 2014; Graveley et al. 2001; Martins de Araujo et al. 2009) whilst the RNA interaction has been established to predominantly require RRM2 (Cléry et al. 2013). However, the interaction with U1 is less clear. An RRM-RRM interaction between SRSF1 RRM1 and U1-70K (S. Cho et al. 2011) and an RRM-RNA interaction between RRM1 of SRSF1 and stem loop 3 of the U1 snRNA (unpublished data, Allain lab) have been proposed. The RNA driven interaction would explain the apparent lack of an interaction for the stem loop 3 of the U1 snRNA, with it always being free for SRSF1 binding and subsequent exon definition. Meta-analysis of published CLIP data for SRSF1 also supports SRSF1 binding the U1 snRNA in stem loop 3; it is likely this interaction was previously missed as standard CLIP analysis pipelines remove reads that can be mapped to multiple locations in the genome so the duplicates of the U1 gene would lead to U1 snRNA reads being removed.

When these models are reconciled with the data from the ESE sections then a clearer picture of the different recruitment pathways of SRSF1 is presented. In cases where there is a strong ESE or multiple ESEs then the probability of an interaction is high enough for recruitment and a subsequent interaction, with either U1 (via RRM1) or 3' splice site factors (via the RS domain), to occur. However, in cases where there are no or weak ESEs but a strong 5' splice site, then SRSF1 binds to the U1 snRNP first and either, contacts nearby RNA (via RRM2) before stabilising 3' splice site factors or, directly stabilises 3' splice site factors.

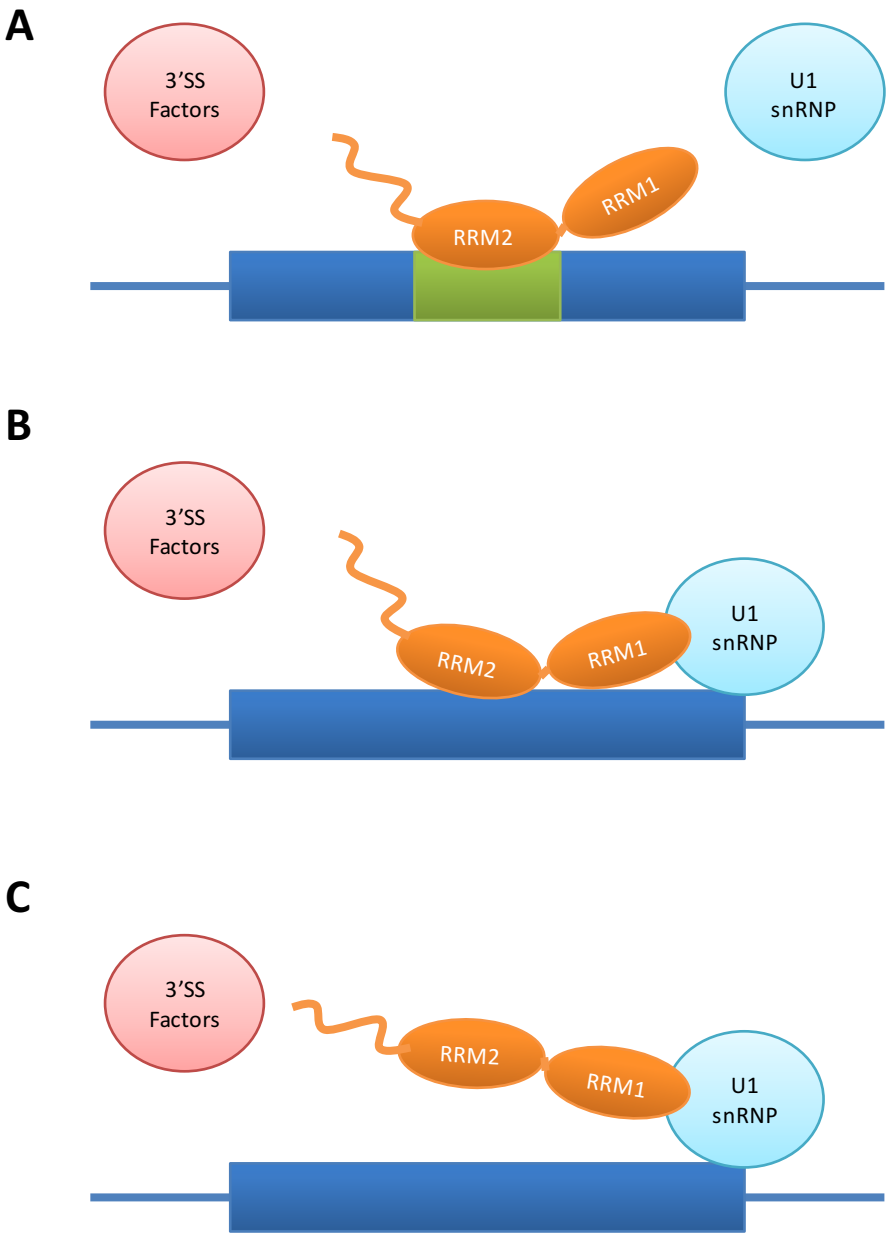


Figure 60. Pathways for SRSF1 recruitment to exons. A) ESE driven recruitment model, B) U1 driven recruitment model with intermediate RNA binding prior to interactions with 3' splice site factors, C) U1 driven recruitment with direct interactions with the 3' splice site factors.

7.3. SRSF1 recruitment by U1 snRNPs and its implication on the core splicing reaction.

Whilst the recruitment of SRSF1 by U1 snRNPs seems highly likely to play a role in exon definition, whether it is important for the splicing of the downstream intron is less clear. In order to look at this possibility it is first necessary to re-look at the previous data of the role of U1 snRNPs and SRSF1 in splicing.

The U1 snRNP was the first splicing factor discovered, and it has well characterized roles in the selection of 5' splice sites (X. Roca et al. 2013). Despite this its status as a splicing reaction component was originally in doubt due to its weak association with splicing complexes (Konarska & Sharp 1986). It has been suggested that it forms cross-intron interactions via association with the proteins SF3a, Prp5 or Prp40, although none of these have been shown to bind as a single molecule and interact directly with both ends of the intron simultaneously (Shao et al. 2012; Abovich & Rosbash 1997; Sharma et al. 2014b). The possibility that the U1 snRNP itself plays a direct role in the splicing reaction is weakened by three observations. (i) It has been shown to enable splicing even when binding some nucleotides away from the 5' splice site; the exact position of the 5' splice site is determined by the U6 snRNA base-pairing (Hwang & Cohen 1996; Brackenridge et al. 2003; Hang et al. 2015). (ii) The tri-snRNP can bind directly to a 5' splice site regardless of the presence of a U1 snRNP; interestingly the recruitment of the tri-snRNP is also enhanced by SR proteins (Roscigno & Garcia-Blanco 1995; Konforti & Konarska 1994; Maroney et al. 2000; Wan et al. 2016). (iii) The splicing of some introns has been found to be independent of the U1 snRNP (Crispino et al. 1996; Raponi et al. 2009). Furthermore, some introns that are U1-dependent have remarkably been shown in vitro

to splice without a U1 snRNP if the concentration of SR proteins is high enough (Tarn & Steitz 1994). One possibility that could be inferred is that the U1 snRNP might not play a direct role in splicing reactions but acts to co-ordinate the binding of other key factors, such as SRSF1 or the tri-snRNP, in a stoichiometric and site specific manner.

Direct evidence for a role of SRSF1 in core splicing reactions is also limited, but not necessarily non-existent. As noted above, SR proteins can substitute for U1 snRNP or recruit the tri-snRNP, and the catalytic reactions of splicing require de-phosphorylation of SR proteins whilst the activation by SR proteins requires them to be phosphorylated (Mermoud et al. 1994). Indeed when SRSF1 was added to an S100 cytoplasmic extract to restore splicing, it was found that its RS domain contacted the 5'SS in both complexes B and C, subsequent to when ESE driven activation is believed to occur, and that it was required for U6 snRNA base-pairing to the 5'SS (Shen et al. 2004). Furthermore splicing was found to not need the addition of SRSF1 if an RS domain was tethered near to the 5'SS or when the base-pairing between U6 snRNA and the 5' splice site had been improved (Shen & Green 2007). These results suggest that SRSF1 may actually play a role in stimulating splicing reactions at a much later stage than first thought. The recruitment by a U1 snRNP to the 5'SS, which we demonstrate here, may well be how an SRSF1 is positioned to undertake this role in splicing reactions.

Additional experiments looking at the relationship between U1 snRNP binding and DDX5 binding also provided interesting insights. These results seem to show that DDX5 was recruited in a stochastic manner unless a U1 snRNP could bind, in which case it bound in a 1:1 stoichiometry. This is not too distant from what we see with SRSF1. This suggests one of two things, the SRSF1 is responsible for recruiting DDX5 and when it binds

stochastically (background binding) then so does DDX5, or similarly to SRSF1, stochastic background binding happens on RNAs which are not destined for complex formation, as designated by the lack of a U1 snRNP, and stoichiometric binding happens on those that are. If the former is the case then this may be another role for SRSF1 in core splicing reactions, to position DDX5. Whilst if the latter is true then this suggests that the U1 snRNP may act as a hub to bring in many factors that play key roles in splicing reactions and that SRSF1 is just one of many factors that come in via U1. However the results using the NEAD mutant of DDX5 do not fit to this model as we expect to see the same shift in the presence of the anti-U1 oligo regardless of the loss of the proteins helicase activity. In order to assess this further and to differentiate between the possibilities outlined above i.e. is it SRSF1, U1 or another protein that is recruiting DDX5, more conclusive experiments are required. Furthermore, analysis of additional proteins such as RBM5 or other SR proteins may help to answer key questions regarding DDX5 recruitment and also might assist in answering questions regarding whether U1 co-ordinates any other splicing regulators during early spliceosome assembly.

7.4. Limitations of single molecule studies.

This study examines in depth the mechanisms for the recruitment of the archetypal SR protein SRSF1 to pre-mRNA in the early splicing complexes. Here we use single molecule methods as well as ensemble techniques to examine in detail how SRSF1 is recruited to ESEs as well as to reveal a separate mode of recruitment mediated by U1 snRNPs.

However whilst the methods used here have allowed a number of major questions, which would otherwise been unanswerable, to be answered, they are not without their limitations. One of the limitations in our results is that we cannot perfectly determine the exact proportion of complexes that associate with two or more molecules of SRSF1. This is highlighted in experiments where there is a spike in the number of complexes with two molecules bound but there is an under-representation of complexes that bleach in three steps. We would expect to see this increase as a result of the intrinsic dimerization of mEGFP-SRSF1. One possibility is that the interactions required for SRSF1 binding in complex A, either via U1 or via ESEs, prevent the surfaces that would be required for dimerization from being exposed. Another possibility is that there is a level of bleaching that occurs prior to the imaging of the complexes under the microscope. Exposure to light sources prior to imaging, is minimised so this is less likely and if this were the case then decreased levels of complexes that contain two molecules would also be observed. Furthermore, bleaching prior to imaging but after the laser has been switched on can also be excluded as fluorescence was recorded continuously during the switch between lasers.

Single molecule methods might also underestimate the level of complexes containing multiple molecules of SRSF1 due to a proportion of mEGFP molecules not being fluorescently active. This could result from two key factors: (i) a significant proportion of mEGFP molecules being in a dark state (blinking) or (ii) a significant proportion of mEGFP molecules being mis-folded. The first factor, blinking, can be ruled out since the dark state lifetime of mEGFP is only 1-2 s (Dickson et al. 1997; Garcia-Parajo et al. 2000; Vámosi et al. 2016) and measurements for experiments here were made for between 15-30 s, so the probability of being in a dark state throughout is minimal.

The second factor, mis-folding, is harder to estimate accurately and therefore might be significant. One assay looking at GFP labelled protein mis-folding estimated around 20% of the GFP was mis-folded (Ulbrich & Isacoff 2007). However, in this case the EGFP sequence was attached to the C terminal side of the protein which would lead to increased mis-folding since the folding of the EGFP would be affected by any mis-folding of the upstream sequence (Waldo et al. 1999; Wang & Chong 2003). The fluorescent labels used in this study, both mCherry and mEGFP, are attached to the N termini of the target protein so mis-folding is less likely. Therefore we considered the contribution of mis-folding to be near negligible.

These limitations, even if they weren't negligible in there effect, do not affect the observations that the proportions of complexes containing a specified number of molecules strongly depends on the identity of the RNA i.e. the presence of an ESE, number of ESEs or number of consensus 5' splice sites.

Bibliography.

Abovich, N. & Rosbash, M., 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, 89(3), pp.403–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9150140 [Accessed May 1, 2014].

Akerman, M. et al., 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biology*, 10(3), p.R30. Available at: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r30 [Accessed November 3, 2017].

Anczuków, O. et al., 2015. SRSF1-Regulated Alternative Splicing in Breast Cancer. *Molecular Cell*, 60(1), pp.105–117. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26431027 [Accessed November 3, 2017].

Anczuków, O. et al., 2012. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature Structural & Molecular Biology*, 19(2), pp.220–228. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22245967 [Accessed November 8, 2017].

Anon, 1994. The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell*, 76(4), pp.735–746. Available at: http://www.sciencedirect.com/science/article/pii/0092867494905126 [Accessed November 6, 2017].

Axelrod, D., 2001. Total internal reflection fluorescence microscopy in cell biology. *Traffic (Copenhagen, Denmark)*, 2(11), pp.764–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11733042 [Accessed November 8, 2017].

Axelrod, D., Burghardt, T.P. & Thompson, N.L., 1984. Total Internal Reflection Fluorescence. *Annual Review of Biophysics and Bioengineering*, 13(1), pp.247–268. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6378070 [Accessed November 8, 2017].

Baker, B.S., 1989. Sex in flies: the splice of life. *Nature*, 340(6234), pp.521–524. Available at: http://www.nature.com/doifinder/10.1038/340521a0 [Accessed November 8, 2017].

Barbosa-Morais, N.L. et al., 2012. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science*, 338(6114), pp.1587–1593. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23258890 [Accessed November 3, 2017].

Barta, I. & Iggo, R., 1995. Autoregulation of expression of the yeast Dbp2p "DEAD-box" protein is mediated by sequences in the conserved DBP2 intron. *The EMBO journal*, 14(15), pp.3800–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7641698 [Accessed November 8, 2017].

Bateman, J.F. et al., 1994. A 5' splice site mutation affecting the pre-mRNA splicing of two upstream exons in the collagen COL1A1 gene. Exon 8 skipping and altered definition of exon 7 generates truncated pro alpha 1(I) chains with a non-collagenous insertion destabilizing the triple helix. *The Biochemical journal*, 302 (Pt 3), pp.729–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7945197 [Accessed November 3, 2017].

Bates, D.O. et al., 2017. Pharmacology of Modulators of Alternative Splicing. *Pharmacological reviews*, 69(1), pp.63–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28034912 [Accessed November 4, 2017].

Bates, G.J. et al., 2005. The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *The EMBO journal*, 24(3), pp.543–53. Available at: http://emboj.embopress.org/cgi/doi/10.1038/sj.emboj.7600550 [Accessed November 8, 2017].

Becerra, S. et al., 2015. Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing in vivo. *RNA (New York, N.Y.)*, 21(3), pp.438–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25605964 [Accessed November 3, 2017].

Berget, S.M., Moore, C. & Sharp, P.A., 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8), pp.3171–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/269380 [Accessed November 3, 2017].

Berglund, J.A. et al., 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, 89(5), pp.781–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9182766 [Accessed November 3, 2017].

Bertram, K., Agafonov, D.E., Liu, W.-T., et al., 2017. Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature 2017 542:7641*, 542(7641), p.nature21079. Available at: http://www.nature.com/doifinder/10.1038/nature21079 [Accessed November 3, 2017].

Bertram, K., Agafonov, D.E., Dybkov, O., et al., 2017. Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell*, 170(4), pp.701–713.e11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28781166 [Accessed November 3, 2017].

Bessonov, S. et al., 2010. Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA (New York, N.Y.)*, 16(12), pp.2384–403. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2995400&tool=pmcentrez&rendertype=abstract [Accessed April 6, 2014].

Best, A. et al., 2014. Human Tra2 proteins jointly control a CHEK1 splicing switch among alternative and constitutive target exons. *Nature communications*, 5, p.4760. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25208576 [Accessed November 8, 2017].

Birney, E., Kumar, S. & Krainer, A.R., 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic acids research*, 21(25), pp.5803–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8290338 [Accessed May 15, 2018].

Boggs, R.T. et al., 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell*, 50(5), pp.739–47. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2441872 [Accessed November 8, 2017].

Boon, K.-L. et al., 2007. prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nature Structural & Molecular Biology*, 14(11), pp.1077–1083. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17934474 [Accessed November 8, 2017].

Brackenridge, S., Wilkie, A.O.M. & Screaton, G.R., 2003. Efficient use of a "dead-end" GA 5' splice site in the human fibroblast growth factor receptor genes. *The EMBO journal*, 22(7), pp.1620–31. Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdg163 [Accessed December 13, 2017].

Bringmann, P. & Lührmann, R., 1986. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *The EMBO journal*, 5(13), pp.3509–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2951249 [Accessed May 13, 2018].

Buratti, E. & Baralle, F.E., 2004. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Molecular and Cellular Biology*, 24(24), pp.10505–10514. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15572659 [Accessed November 3, 2017].

Burghardt, T.P. & Ajtai, K., 2010. Single-molecule fluorescence characterization in native environment. *Biophysical reviews*, 2(4), pp.159–167. Available at: http://link.springer.com/10.1007/s12551-010-0038-z [Accessed November 8, 2017].

Cáceres, J.F. & Krainer, A.R., 1993. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *The EMBO journal*, 12(12), pp.4715–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8223480 [Accessed November 6, 2017].

Caliskan, G. et al., 2005. Persistence Length Changes Dramatically as RNA Folds. *Physical Review Letters*, 95(26), p.268303. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16486414 [Accessed November 30, 2017].

Cao, W. & Garcia-Blanco, M.A., 1998. A Serine/Arginine-rich Domain in the Human U1 70k Protein Is Necessary and Sufficient for ASF/SF2 Binding. *Journal of Biological Chemistry*, 273(32), pp.20629–20635. Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.273.32.20629 [Accessed November 6, 2017].

Cartegni, L. et al., 2006. Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, SMN1 and SMN2. *The American Journal of Human Genetics*, 78(1), pp.63–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16385450 [Accessed December 5, 2017].

Cartegni, L. & Krainer, A.R., 2003. Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nature Structural Biology*, 10(2), pp.120–125. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12524529 [Accessed November 6, 2017].

Chan, S.-P. & Cheng, S.-C., 2005. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *The Journal of biological chemistry*, 280(35), pp.31190–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15994330 [Accessed November 3, 2017].

Chandler, S.D. et al., 1997. RNA splicing specificity determined by the coordinated action of RNA recognition motifs in SR proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp.3596–601. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9108022 [Accessed November 6, 2017].

Chang, X., Li, B. & Rao, A., RNA-binding protein hnRNPLL regulates mRNA splicing and stability during B-cell to plasma-cell differentiation Divisions of a Signaling and Gene Expression and. Available at: http://www.pnas.org/content/112/15/E1888.full.pdf [Accessed November 3, 2017].

Chen, L. et al., 2016. Stoichiometries of U2AF35, U2AF65 and U2 snRNP reveal new early spliceosome assembly pathways. *Nucleic Acids Research*, 45(4), p.gkw860. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27683217 [Accessed November 8, 2017].

Cherny, D. et al., 2010. Stoichiometry of a regulatory splicing complex revealed by single-molecule analyses. *The EMBO journal*, 29(13), pp.2161–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2905242&tool=pmcentrez&rendertype=abstract [Accessed January 7, 2015].

Cho, S. et al., 2011. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences*, 108(20), pp.8233–8238. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21536904 [Accessed November 4, 2017].

Cho, S. et al., 2011. The SRSF1 linker induces semi-conservative ESE binding by cooperating with the RRMs. *Nucleic acids research*, 39(21), pp.9413–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21852328 [Accessed November 4, 2017].

Choi, Y.-J. & Lee, S.-G., 2012. The DEAD-box RNA helicase DDX3 interacts with DDX5, co-localizes with it in the cytoplasm during the G2/M phase of the cycle, and affects its shuttling during mRNP export. *Journal of cellular biochemistry*, 113(3), pp.985–96. Available at: http://doi.wiley.com/10.1002/jcb.23428 [Accessed November 8, 2017].

Chow, L.T. et al., 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1), pp.1–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/902310 [Accessed November 3, 2017].

Cléry, A. et al., 2013. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 110(30), pp.E2802–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23836656 [Accessed November 4, 2017].

Cléry, A. et al., 2011. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nature Structural & Molecular Biology*, 18(4), pp.443–450. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21399644 [Accessed November 8, 2017].

Cléry, A., Blatter, M. & Allain, F.H.-T., 2008. RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, 18(3), pp.290–298. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18515081 [Accessed May 15, 2018].

Cloutier, S.C. et al., 2012. The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. *The Journal of biological*

chemistry, 287(31), pp.26155–66. Available at: http://www.jbc.org/lookup/doi/10.1074/jbc.M112.383075 [Accessed November 8, 2017].

De Conti, L., Baralle, M. & Buratti, E., 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp.49–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23044818 [Accessed November 3, 2017].

Crawford, D.J. et al., 2013. Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. *Proceedings of the National Academy of Sciences*, 110(17), pp.6783–6788. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23569281 [Accessed May 22, 2018].

Crispino, J.D. et al., 1996. Cis-acting elements distinct from the 5' splice site promote U1-independent pre-mRNA splicing. *RNA (New York, N.Y.)*, 2(7), pp.664–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8756409 [Accessed December 13, 2017].

Dardenne, E. et al., 2014. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Reports*, 7(6), pp.1900–1913. Available at: http://linkinghub.elsevier.com/retrieve/pii/S2211124714003854 [Accessed November 8, 2017].

Darman, R.B. et al., 2015. Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point. *Cell Reports*, 13(5), pp.1033–1045. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26565915 [Accessed November 8, 2017].

Das, R., Zhou, Z. & Reed, R., 2000. Functional Association of U2 snRNP with the ATP-Independent Spliceosomal Complex E. *Molecular Cell*, 5(5), pp.779–787. Available at: http://www.sciencedirect.com/science/article/pii/S1097276500803184 [Accessed February 13, 2014].

Das, S. et al., 2012. Oncogenic Splicing Factor SRSF1 Is a Critical Transcriptional Target of MYC. *Cell Reports*, 1(2), pp.110–117. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22545246 [Accessed November 8, 2017].

Das, S. & Krainer, A.R., 2014. Emerging Functions of SRSF1, Splicing Factor and Oncoprotein, in RNA Metabolism and Cancer. *Molecular Cancer Research*, 12(9), pp.1195–1204. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24807918 [Accessed November 4, 2017].

Dauwalder, B., Amaya-Manzanares, F. & Mattox, W., 1996. A human homologue of the Drosophila sex determination factor transformer-2 has conserved splicing regulatory functions. *Proceedings of the National Academy of Sciences of the United States of America*, 93(17), pp.9004–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8799144 [Accessed November 8, 2017].

Dickson, R.M. et al., 1997. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640), pp.355–8. Available at: http://www.nature.com/doifinder/10.1038/41048 [Accessed December 13, 2017].

Dirksen, W.P. et al., 1994. A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *The Journal of biological chemistry*, 269(9), pp.6431–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8119993 [Accessed November 4, 2017].

DONMEZ, G., Hartmuth, K. & Lührmann, R., 2004. Modified nucleotides at the 5' end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA*, 10(12), pp.1925–1933. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15525712 [Accessed November 3, 2017].

Dreyfuss, G., Swanson, M.S. & Piñol-Roma, S., 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends in biochemical sciences*, 13(3), pp.86–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3072706 [Accessed May 15, 2018].

Dunham, I. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74. Available at: http://www.nature.com/doifinder/10.1038/nature11247 [Accessed November 3, 2017].

Eperon, I.C. et al., 1993. Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *The EMBO journal*, 12(9), pp.3607–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8253084 [Accessed November 3, 2017].

Eperon, I.C. et al., 2000. Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Molecular and cellular biology*, 20(22), pp.8303–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11046128 [Accessed November 6, 2017].

Eperon, L.P., Estibeiro, J.P. & Eperon, I.C., 1986. The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature*, 324(6094), pp.280–282. Available at: http://www.nature.com/articles/324280a0 [Accessed November 3, 2017].

Erkelenz, S. et al., 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA (New York, N.Y.)*, 19(1), pp.96–102. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3527730&tool=pmcentrez&rendertype=abstract [Accessed January 7, 2015].

Fica, S.M. et al., 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature*, 503(7475), pp.229–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24196718 [Accessed November 3, 2017].

Fu, X.D. et al., 1992. General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. *Proceedings of the National Academy of Sciences of the United States of America*, 89(23), pp.11224–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1454802 [Accessed November 17, 2017].

Fu, X.-D. & Ares, M., 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature Reviews Genetics*, 15(10), pp.689–701. Available at: http://www.nature.com/doifinder/10.1038/nrg3778 [Accessed November 3, 2017].

Fuller-Pace, F. V & Ali, S., 2008. The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. *Biochemical Society transactions*, 36(Pt 4), pp.609–12. Available at: http://biochemsoctrans.org/lookup/doi/10.1042/BST0360609 [Accessed November 8, 2017].

Fuller-Pace, F.V. & Ali, S., 2008. The DEAD box RNA helicases p68 (Ddx5) and p72 (Ddx17): novel transcriptional co-regulators. *Biochemical Society Transactions*, 36(4), pp.609–612. Available at: http://biochemsoctrans.org/lookup/doi/10.1042/BST0360609 [Accessed November 8, 2017].

Galej, W.P. et al., 2016. Cryo-EM structure of the spliceosome immediately after branching. *Nature*, 537(7619), pp.197–201. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27459055 [Accessed November 3, 2017].

Garcia-Parajo, M.F. et al., 2000. Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), pp.7237–42. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10860989 [Accessed December 13, 2017].

Garrey, S.M. et al., 2014. A homolog of lariat-debranching enzyme modulates turnover of branched RNA. *RNA*, 20(8), pp.1337–1348. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/24919400 [Accessed November 3, 2017].

Ge, H. & Manley, J.L., 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell*, 62(1), pp.25–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2163768 [Accessed November 6, 2017].

Ghigna, C. et al., 2005. Cell Motility Is Controlled by SF2/ASF through Alternative Splicing of the Ron Protooncogene. *Molecular Cell*, 20(6), pp.881–890. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16364913 [Accessed November 4, 2017].

Gonçalves, V. & Jordan, P., 2015. Posttranscriptional Regulation of Splicing Factor SRSF1 and Its Role in Cancer Cell Biology. *BioMed Research International*, 2015, pp.1–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26273603 [Accessed November 8, 2017].

Gooding, C. et al., 2013. MBNL1 and PTB cooperate to repress splicing of Tpm1 exon 3. *Nucleic Acids Research*, 41(9), pp.4765–4782. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23511971 [Accessed May 22, 2018].

Gozani, O., Feld, R. & Reed, R., 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes & development*, 10(2), pp.233–43.

Gozani, O., Potashkin, J. & Reed, R., 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Molecular and cellular biology*, 18(8), pp.4752–60. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9671485 [Accessed November 3, 2017].

Graveley, B.R., Hertel, K.J. & Maniatis, T., 1998a. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *The EMBO journal*, 17(22), pp.6747–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9822617 [Accessed November 3, 2017].

Graveley, B.R., Hertel, K.J. & Maniatis, T., 1998b. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *The EMBO journal*, 17(22), pp.6747–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9822617 [Accessed November 4, 2017].

Graveley, B.R., Hertel, K.J. & Maniatis, T., 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA (New York, N.Y.)*, 7(6), pp.806–18.

Available at: http://www.ncbi.nlm.nih.gov/pubmed/11421359 [Accessed November 6, 2017].

Graveley, B.R. & Maniatis, T., 1998. Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Molecular cell*, 1(5), pp.765–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9660960 [Accessed January 7, 2015].

Grellscheid, S. et al., 2011. Identification of Evolutionarily Conserved Exons as Regulated Targets for the Splicing Activator Tra2β in Development W. A. Bickmore, ed. *PLoS Genetics*, 7(12), p.e1002390. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22194695 [Accessed November 8, 2017].

Guil, S., de La Iglesia, N., et al., 2003. Alternative splicing of the human proto-oncogene c-H-ras renders a new Ras family protein that trafficks to cytoplasm and nucleus. *Cancer research*, 63(17), pp.5178–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14500341 [Accessed November 8, 2017].

Guil, S., Gattoni, R., et al., 2003. Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Molecular and cellular biology*, 23(8), pp.2927–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12665590 [Accessed November 8, 2017].

Hang, J. et al., 2015. Structural basis of pre-mRNA splicing. *Science*, 349(6253), pp.1191–1198. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26292705 [Accessed December 13, 2017].

Hardin, J.W. et al., 2015. Assembly and dynamics of the U4/U6 di-snRNP by single-molecule FRET. *Nucleic Acids Research*, 43(22), pp.10963–10974. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26503251 [Accessed May 22, 2018].

Havens, M.A., Duelli, D.M. & Hastings, M.L., 2013. Targeting RNA splicing for disease therapy. *Wiley Interdisciplinary Reviews: RNA*, 4(3), pp.247–266. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23512601 [Accessed November 8, 2017].

Hedjran, F. et al., 1997. Control of alternative pre-mRNA splicing by distributed pentameric repeats. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23), pp.12343–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9356451 [Accessed November 4, 2017].

Hedley, M.L. & Maniatis, T., 1991. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell*, 65(4), pp.579–86. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/1674449 [Accessed November 4, 2017].

Hertel, K.J. et al., 1996. Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. *RNA (New York, N.Y.)*, 2(10), pp.969–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8849774 [Accessed November 4, 2017].

Hertel, K.J. & Maniatis, T., 1999. Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 96(6), pp.2651–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10077565 [Accessed November 6, 2017].

Hertel, K.J. & Maniatis, T., 1998. The function of multisite splicing enhancers. *Molecular cell*, 1(3), pp.449–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9660929 [Accessed November 3, 2017].

Van Heyningen, V. & Bickmore, W., 2013. Regulation from a distance: long-range control of gene expression in development and disease. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1620), p.20120372. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23650642 [Accessed November 3, 2017].

Hodson, M.J. et al., 2012. The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites. *Nucleic acids research*, 40(14), pp.6850–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22505580 [Accessed November 28, 2017].

Hoffman, B.E. & Grabowski, P.J., 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes & development*, 6(12B), pp.2554–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1285125 [Accessed December 13, 2017].

Hofmann, Y. et al., 2000. Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). *Proceedings of the National Academy of Sciences*, 97(17), pp.9618–9623. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10931943 [Accessed November 8, 2017].

Hofmann, Y. & Wirth, B., 2002. hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. *Human molecular genetics*, 11(17), pp.2037–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12165565 [Accessed November 8, 2017].

Hollander, D. et al., 2016. How Are Short Exons Flanked by Long Introns Defined and Committed to Splicing? *Trends in Genetics*, 32(10), pp.596–606. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27507607 [Accessed November 3, 2017].

Hoshijima, K. et al., 1991. Control of doublesex alternative splicing by transformer and transformer-2 in Drosophila. *Science (New York, N.Y.)*, 252(5007), pp.833–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1902987 [Accessed November 8, 2017].

Hoskins, A.A. et al., 2016. Single molecule analysis reveals reversible and irreversible steps during spliceosome activation. *eLife*, 5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27244240 [Accessed May 22, 2018].

Hoskins, A.A. & Moore, M.J., 2012. The spliceosome: a flexible, reversible macromolecular machine. *Trends in Biochemical Sciences*, 37(5), pp.179–188. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22480731 [Accessed November 3, 2017].

Van der Houven van Oordt, W. et al., 2000. Role of SR protein modular domains in alternative splicing specificity in vivo. *Nucleic Acids Research*, 28(24), pp.4822–4831. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/28.24.4822 [Accessed November 4, 2017].

Humphrey, M.B. et al., 1995. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5' splice sites in a differential internal exon. *Molecular and cellular biology*, 15(8), pp.3979–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7623794 [Accessed November 4, 2017].

Hwang, D.Y. & Cohen, J.B., 1996. Base pairing at the 5' splice site with U1 small nuclear RNA promotes splicing of the upstream intron but may be dispensable for slicing of the downstream intron. *Molecular and cellular biology*, 16(6), pp.3012–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8649413 [Accessed November 3, 2017].

Jablonski, A., 1933. Efficiency of Anti-Stokes Fluorescence in Dyes. *Nature*, 131(3319), pp.839–840. Available at: http://www.nature.com/articles/131839b0 [Accessed November 8, 2017].

Jamison, S.F. et al., 1995. U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic acids research*, 23(16), pp.3260–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7667103 [Accessed November 6, 2017].

Jankowsky, E. et al., 2001. Active Disruption of an RNA-Protein Interaction by a DExH/D RNA Helicase. *Science*, 291(5501), pp.121–125. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/11141562 [Accessed November 8, 2017].

Jiang, Z. et al., 2000. Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17. *Molecular and cellular biology*, 20(11), pp.4036–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10805746 [Accessed November 8, 2017].

Kao, H.Y. & Siliciano, P.G., 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Molecular and cellular biology*, 16(3), pp.960–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=231078&tool=pmcentrez&rendertype=abstract [Accessed January 21, 2014].

Kar, A. et al., 2011. RNA Helicase p68 (DDX5) Regulates tau Exon 10 Splicing by Modulating a Stem-Loop Structure at the 5' Splice Site. *Molecular and Cellular Biology*, 31(9), pp.1812–1821. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21343338 [Accessed November 8, 2017].

Kashima, T. et al., 2007. hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Human Molecular Genetics*, 16(24), pp.3149–3159. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17884807 [Accessed December 5, 2017].

Kastner, B. et al., 1992. Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *The Journal of cell biology*, 116(4), pp.839–49. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1531145 [Accessed May 13, 2018].

Kastner, B. & Lührmann, R., 1989. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *The EMBO journal*, 8(1), pp.277–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2469573 [Accessed May 13, 2018].

Kechris, K., Yang, Y.H. & Yeh, R.-F., 2008. Prediction of alternatively skipped exons and splicing enhancers from exon junction arrays. *BMC genomics*, 9, p.551. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19021909 [Accessed November 3, 2017].

Kenan, D.J., Query, C.C. & Keene, J.D., 1991. RNA recognition: towards identifying determinants of specificity. *Trends in biochemical sciences*, 16(6), pp.214–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1716386 [Accessed May 15, 2018].

Kitamura-Abe, S. et al., 2004. Characterization of the splice sites in GT-AG and GC-AG introns in higher eukaryotes using full-length cDNAs. *Journal of*

bioinformatics and computational biology, 2(2), pp.309–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15297984 [Accessed November 3, 2017].

Knowles, D.B. et al., 2011. Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31), pp.12699–704. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1103382108 [Accessed November 30, 2017].

Kohtz, J.D. et al., 1994. Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, 368(6467), pp.119–124. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8139654 [Accessed November 6, 2017].

Kolb, H.C., Finn, M.G. & Sharpless, K.B., 2001. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie (International ed. in English)*, 40(11), pp.2004–2021. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11433435 [Accessed November 4, 2017].

Kolb, S.J. & Kissel, J.T., 2015. Spinal Muscular Atrophy. *Neurologic Clinics*, 33(4), pp.831–846. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26515624 [Accessed December 5, 2017].

Konarska, M.M. & Sharp, P.A., 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, 46(6), pp.845–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2944598 [Accessed December 13, 2017].

Kondo, Y. et al., 2015. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife*, 4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25555158 [Accessed November 3, 2017].

Konforti, B.B. & Konarska, M.M., 1994. U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP. *Genes & development*, 8(16), pp.1962–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7958870 [Accessed December 13, 2017].

König, S.L.B. et al., 2013. Helicase-mediated changes in RNA structure at the single-molecule level. *RNA Biology*, 10(1), pp.133–148. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23353571 [Accessed May 22, 2018].

Krainer, A.R. et al., 1991. Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell*, 66(2), pp.383–94. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/1830244 [Accessed November 6, 2017].

Krainer, A.R., Conway, G.C. & Kozak, D., 1990. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes & development*, 4(7), pp.1158–71. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2145194 [Accessed November 4, 2017].

Krämer, A. et al., 1984. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell*, 38(1), pp.299–307. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6235919 [Accessed November 3, 2017].

Kreivi, J.P. et al., 1991. A U1 snRNA binding site improves the efficiency of in vitro pre-mRNA splicing. *Nucleic acids research*, 19(24), p.6956. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=329340&tool=pmcentrez&rendertype=abstract [Accessed June 16, 2015].

Kühlbrandt, W., 2014. Cryo-EM enters a new era. *eLife*, 3, p.e03678. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25122623 [Accessed November 3, 2017].

Larson, J.D. & Hoskins, A.A., 2017. Dynamics and consequences of spliceosome E complex formation. *eLife*, 6, p.e27592. Available at: https://elifesciences.org/articles/27592 [Accessed November 3, 2017].

Lavigueur, A. et al., 1993a. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes & development*, 7(12A), pp.2405–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8253386 [Accessed January 7, 2015].

Lavigueur, A. et al., 1993b. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes & development*, 7(12A), pp.2405–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8253386 [Accessed November 3, 2017].

Lear, A.L. et al., 1990. Hierarchy for 5′ splice site preference determined in vivo. *Journal of Molecular Biology*, 211(1), pp.103–115. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2299664 [Accessed November 3, 2017].

Lewis, H., Perrett, A.J., et al., 2012. An RNA Splicing Enhancer that Does Not Act by Looping. *Angewandte Chemie International Edition*, 51(39), pp.9800–9803. Available at: http://doi.wiley.com/10.1002/anie.201202932 [Accessed November 4, 2017].

Lewis, H., Perrett, A.J., et al., 2012. An RNA splicing enhancer that does not act by looping. *Angewandte Chemie (International ed. in English)*, 51(39), pp.9800–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22936639 [Accessed February 10, 2015].

Lim, K.H. et al., 2011. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(27), pp.11093–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21685335 [Accessed November 3, 2017].

Lim, L.P. & Sharp, P.A., 1998. Alternative splicing of the fibronectin EIIIB exon depends on specific TGCATG repeats. *Molecular and cellular biology*, 18(7), pp.3900–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9632774 [Accessed November 4, 2017].

Lin, C. et al., 2005. ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Molecular and cellular biology*, 25(17), pp.7484–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16107697 [Accessed December 19, 2017].

Lin, S. et al., 2005. Dephosphorylation-Dependent Sorting of SR Splicing Factors during mRNP Maturation. *Molecular Cell*, 20(3), pp.413–425. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16285923 [Accessed November 4, 2017].

Linder, P. & Jankowsky, E., 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nature reviews. Molecular cell biology*, 12(8), pp.505–16. Available at: http://www.nature.com/doifinder/10.1038/nrm3154 [Accessed November 8, 2017].

Liu, H.X., Zhang, M. & Krainer, A.R., 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes & development*, 12(13), pp.1998–2012. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9649504 [Accessed November 4, 2017].

Liu, Z.-R., 2002. p68 RNA helicase is an essential human splicing factor that acts at the U1 snRNA-5' splice site duplex. *Molecular and cellular biology*, 22(15), pp.5443–50. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12101238 [Accessed December 19, 2017].

Long, J.C. & Caceres, J.F., 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, 417(1), pp.15–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19061484 [Accessed November 4, 2017].

Longman, D., Johnstone, I.L. & Cáceres, J.F., 2000. Functional characterization of SR and SR-related genes in Caenorhabditis elegans. *The EMBO Journal*, 19(7), pp.1625–1637. Available at: http://emboj.embopress.org/cgi/doi/10.1093/emboj/19.7.1625 [Accessed November 4, 2017].

López-Bigas, N., Blencowe, B.J. & Ouzounis, C.A., 2006. Highly consistent patterns for inherited human diseases at the molecular level. *Bioinformatics*, 22(3), pp.269–277. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16287936 [Accessed November 8, 2017].

Lorson, C.L. et al., 1999. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proceedings of the National Academy of Sciences of the United States of America*, 96(11), pp.6307–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10339583 [Accessed November 3, 2017].

Madipalli, S., 2017. Spinraza—the patient perspective. *Gene Therapy*, 24(9), pp.501–502. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28880018 [Accessed November 8, 2017].

Maguire, S.L. et al., 2015. *SF3B1* mutations constitute a novel therapeutic target in breast cancer. *The Journal of Pathology*, 235(4), pp.571–580. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25424858 [Accessed November 8, 2017].

Maris, C., Dominguez, C. & Allain, F.H.-T., 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal*, 272(9), pp.2118–2131. Available at: http://doi.wiley.com/10.1111/j.1742-4658.2005.04653.x [Accessed May 15, 2018].

Maroney, P.A., Romfo, C.M. & Nilsen, T.W., 2000. Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. *Molecular cell*, 6(2), pp.317–28. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10983979 [Accessed December 13, 2017].

Martins de Araujo, M. et al., 2009. Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA*, 15(4), pp.515–523. Available at: http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.1273209 [Accessed November 6, 2017].

Martins de Araújo, M. et al., 2009. Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA (New York, N.Y.)*, 15(4), pp.515–23. Available at: http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.1273209 [Accessed November 6, 2017].

Maslon, M.M. et al., 2014. The translational landscape of the splicing factor SRSF1 and its role in mitosis. *eLife*, 3, p.e02028. Available at: https://elifesciences.org/articles/02028 [Accessed November 4, 2017].

Mayeda, A. et al., 1999. Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Molecular and cellular biology*, 19(3), pp.1853–63. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10022872 [Accessed November 6, 2017].

Mayeda, A., Helfman, D.M. & Krainer, A.R., 1993. Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Molecular and cellular biology*, 13(5), pp.2993–3001. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8474457 [Accessed December 20, 2017].

McKeown, M., Belote, J.M. & Baker, B.S., 1987. A molecular analysis of transformer, a gene in Drosophila melanogaster that controls female sexual differentiation. *Cell*, 48(3), pp.489–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3100051 [Accessed November 8, 2017].

McKeown, M., Belote, J.M. & Boggs, R.T., 1988. Ectopic expression of the female transformer gene product leads to female differentiation of chromosomally male Drosophila. *Cell*, 53(6), pp.887–95. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2454747 [Accessed November 8, 2017].

Mende, Y. et al., 2010. Deficiency of the splicing factor Sfrs10 results in early embryonic lethality in mice and has no impact on full-length SMN/Smn splicing. *Human Molecular Genetics*, 19(11), pp.2154–2167. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20190275 [Accessed November 8, 2017].

Mermoud, J.E., Cohen, P. & Lamond, A.I., 1992. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic acids research*, 20(20), pp.5263–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1331983 [Accessed November 6, 2017].

Mermoud, J.E., Cohen, P.T. & Lamond, A.I., 1994. Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *The EMBO journal*, 13(23), pp.5679–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7988565 [Accessed November 6, 2017].

Meurs, E.F. et al., 1992. Constitutive expression of human double-stranded RNA-activated p68 kinase in murine cells mediates phosphorylation of eukaryotic initiation factor 2 and partial resistance to encephalomyocarditis virus

growth. *Journal of virology*, 66(10), pp.5805–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1382142 [Accessed November 8, 2017].

Michaud, S. & Reed, R., 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & development*, 5(12B), pp.2534–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1836445 [Accessed November 3, 2017].

Mo, S., Ji, X. & Fu, X.-D., Unique role of SRSF2 in transcription activation and diverse functions of the SR and hnRNP proteins in gene expression regulation. *Transcription*, 4(5), pp.251–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24406341 [Accessed November 4, 2017].

Mount, S.M. et al., 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell*, 33(2), pp.509–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6190573 [Accessed November 3, 2017].

Nagai, K. et al., 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, 348(6301), pp.515–520. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2147232 [Accessed May 15, 2018].

Nelissen, R.L. et al., 1994. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *The EMBO journal*, 13(17), pp.4113–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8076607 [Accessed May 13, 2018].

Nelson, K.K. & Green, M.R., 1990. Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America*, 87(16), pp.6253–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2143583 [Accessed November 3, 2017].

Nguyen, T.H.D. et al., 2015. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature*, 523(7558), pp.47–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26106855 [Accessed November 3, 2017].

O'Day, C.L., Dalbadie-McFarland, G. & Abelson, J., 1996. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *The Journal of biological chemistry*, 271(52), pp.33261–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8969184 [Accessed November 3, 2017].

Ohno, K., Takeda, J. & Masuda, A., 2017. Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdisciplinary Reviews: RNA*,

p.e1451. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28949076 [Accessed November 8, 2017].

Owen, N. et al., 2011. Design principles for bifunctional targeted oligonucleotide enhancers of splicing. *Nucleic acids research*, 39(16), pp.7194–208. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkr152 [Accessed November 8, 2017].

Palacino, J. et al., 2015. SMN2 splice modulators enhance U1–pre-mRNA association and rescue SMA mice. *Nature Chemical Biology*, 11(7), pp.511–517. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26030728 [Accessed November 8, 2017].

Pandit, S. et al., 2013a. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell*, 50(2), pp.223–35. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23562324 [Accessed November 4, 2017].

Pandit, S. et al., 2013b. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Molecular cell*, 50(2), pp.223–35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3640356&tool=pmcentrez&rendertype=abstract [Accessed January 12, 2015].

Patel, A.A. & Steitz, J.A., 2003. Splicing double: insights from the second spliceosome. *Nature Reviews Molecular Cell Biology*, 4(12), pp.960–970. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14685174 [Accessed November 3, 2017].

Pena, V. et al., 2007. Structure of a Multipartite Protein-Protein Interaction Domain in Splicing Factor Prp8 and Its Link to Retinitis Pigmentosa. *Molecular Cell*, 25(4), pp.615–624. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17317632 [Accessed November 8, 2017].

Perrett, A.J. et al., 2013. Conjugation of PEG and gold nanoparticles to increase the accessibility and valency of tethered RNA splicing enhancers. *Chem. Sci.*, 4(1), pp.257–265. Available at: http://xlink.rsc.org/?DOI=C2SC20937C [Accessed December 19, 2017].

Pomeranz Krummel, D.A. et al., 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, 458(7237), pp.475–480. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19325628 [Accessed May 13, 2018].

Raghunathan, P.L. & Guthrie, C., 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current biology : CB*, 8(15), pp.847–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9705931 [Accessed November 3, 2017].

Ramchatesingh, J. et al., 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Molecular and cellular biology*, 15(9), pp.4898–907. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7651409 [Accessed November 4, 2017].

Raponi, M. et al., 2009. Low U1 snRNP dependence at the NF1 exon 29 donor splice site. *FEBS Journal*, 276(7), pp.2060–2073. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19292874 [Accessed December 13, 2017].

Rigo, F. et al., 2014. Pharmacology of a Central Nervous System Delivered 2'-O-Methoxyethyl-Modified Survival of Motor Neuron Splicing Oligonucleotide in Mice and Nonhuman Primates. *Journal of Pharmacology and Experimental Therapeutics*, 350(1), pp.46–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24784568 [Accessed November 8, 2017].

Roca, X. et al., 2012. Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes & Development*, 26(10), pp.1098–1109. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22588721 [Accessed November 3, 2017].

Roca, X., Krainer, A.R. & Eperon, I.C., 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes & Development*, 27(2), pp.129–144. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23348838 [Accessed November 3, 2017].

Roca, X., Krainer, A.R. & Eperon, I.C., 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes & development*, 27(2), pp.129–44. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3566305&tool=pmcentrez&rendertype=abstract [Accessed December 18, 2014].

ROCA, X., Sachidanandam, R. & Krainer, A.R., 2005. Determinants of the inherent strength of human 5' splice sites. *RNA*, 11(5), pp.683–698. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15840817 [Accessed November 3, 2017].

Rodgers, M.L. et al., 2016. A multi-step model for facilitated unwinding of the yeast U4/U6 RNA duplex. *Nucleic Acids Research*, 44(22), pp.10912–10928. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27484481 [Accessed May 22, 2018].

Roscigno, R.F. & Garcia-Blanco, M.A., 1995. SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA (New York, N.Y.)*, 1(7), pp.692–706. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7585254 [Accessed November 6, 2017].

Roy, M. et al., 2008. The effect of intron length on exon creation ratios during the evolution of mammalian genomes. *RNA (New York, N.Y.)*, 14(11), pp.2261–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18796579 [Accessed November 3, 2017].

Ruckstuhl, T. & Seeger, S., 2003. Confocal total-internal-reflection fluorescence microscopy with a high-aperture parabolic mirror lens. *Applied optics*, 42(16), pp.3277–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12790480 [Accessed November 8, 2017].

Sako, Y., Minoghchi, S. & Yanagida, T., 2000. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nature cell biology*, 2(3), pp.168–72. Available at: http://www.nature.com/doifinder/10.1038/35004044 [Accessed November 8, 2017].

Sanford, J.R. et al., 2008. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Research*, 19(3), pp.381–394. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19116412 [Accessed November 4, 2017].

Sanford, J.R. et al., 2009. Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome research*, 19(3), pp.381–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19116412 [Accessed November 3, 2017].

Saporita, A.J. et al., 2011. RNA Helicase DDX5 Is a p53-Independent Target of ARF That Participates in Ribosome Biogenesis. *Cancer Research*, 71(21), pp.6708–6717. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21937682 [Accessed November 8, 2017].

Schmidt, C. et al., 2014. Analyzing the Protein Assembly and Dynamics of the Human Spliceosome with SILAC. In *Methods in molecular biology (Clifton, N.J.)*. pp. 227–244. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25059615 [Accessed November 6, 2017].

Schmidt, T. et al., 1996. Imaging of single molecule diffusion. *Proceedings of the National Academy of Sciences of the United States of America*, 93(7), pp.2926–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8610144 [Accessed November 8, 2017].

Sciabica, K.S. & Hertel, K.J., 2006. The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing. *Nucleic acids research*, 34(22), pp.6612–20. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1747189&tool=pmcentrez&rendertype=abstract [Accessed February 10, 2015].

Scotti, M.M. & Swanson, M.S., 2015. RNA mis-splicing in disease. *Nature Reviews Genetics*, 17(1), pp.19–32. Available at:

http://www.nature.com/doifinder/10.1038/nrg.2015.3 [Accessed November 3, 2017].

Shao, W. et al., 2012. A U1-U2 snRNP interaction network during intron definition. *Molecular and cellular biology*, 32(2), pp.470–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22064476 [Accessed November 3, 2017].

Sharma, S. et al., 2014a. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes & development*, 28(22), pp.2518–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25403181 [Accessed May 13, 2018].

Sharma, S. et al., 2014b. Stem-loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes & development*, 28(22), pp.2518–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25403181 [Accessed December 13, 2017].

Sharma, S. et al., 2011. U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Molecular cell*, 41(5), pp.579–88. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21362553 [Accessed May 13, 2018].

Shcherbakova, I. et al., 2013. Alternative Spliceosome Assembly Pathways Revealed by Single-Molecule Fluorescence Microscopy. *Cell Reports*, 5(1), pp.151–165. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24075986 [Accessed May 22, 2018].

Shen, H. & Green, M.R., 2004. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Molecular cell*, 16(3), pp.363–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15525510 [Accessed January 7, 2015].

Shen, H. & Green, M.R., 2007. RS domain–splicing signal interactions in splicing of U12-type and U2-type introns. *Nature Structural & Molecular Biology*, 14(7), pp.597–603. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17603499 [Accessed December 13, 2017].

Shen, H., Kan, J.L.C. & Green, M.R., 2004. Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Molecular cell*, 13(3), pp.367–76. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14967144 [Accessed November 4, 2017].

Shen, J., Zhang, L. & Zhao, R., 2007. Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. *The Journal of biological chemistry*, 282(31), pp.22544–50. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/17562711 [Accessed November 3, 2017].

Sheth, N. et al., 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Research*, 34(14), pp.3955–3967. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16914448 [Accessed November 3, 2017].

Smith, L.D. et al., 2014. A Targeted Oligonucleotide Enhancer of SMN2 Exon 7 Splicing Forms Competing Quadruplex and Protein Complexes in Functional Conditions. *Cell Reports*, 9(1), pp.193–205. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25263560 [Accessed November 4, 2017].

Smith, L.D. et al., 2014. A targeted oligonucleotide enhancer of SMN2 exon 7 splicing forms competing quadruplex and protein complexes in functional conditions. *Cell reports*, 9(1), pp.193–205. Available at: http://www.cell.com/article/S221112471400727X/fulltext [Accessed February 10, 2015].

Staknis, D. & Reed, R., 1994. SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Molecular and cellular biology*, 14(11), pp.7670–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7935481 [Accessed November 4, 2017].

Staley, J.P. & Guthrie, C., 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Molecular cell*, 3(1), pp.55–64. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10024879 [Accessed November 3, 2017].

Stark, H. et al., 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, 409(6819), pp.539–542. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11206553 [Accessed May 13, 2018].

Staropoli, J.F. et al., 2015. Rescue of gene-expression changes in an induced mouse model of spinal muscular atrophy by an antisense oligonucleotide that promotes inclusion of SMN2 exon 7. *Genomics*, 105(4), pp.220–228. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25645699 [Accessed November 8, 2017].

Steiner, M. et al., 2008. Single-molecule studies of group II intron ribozymes. *Proceedings of the National Academy of Sciences*, 105(37), pp.13853–13858. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18772388 [Accessed May 22, 2018].

Sun, Q. et al., 1993. General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes & development*, 7(12B), pp.2598–608. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8276242 [Accessed November 4, 2017].

Swanson, M.S. et al., 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Molecular and cellular biology*, 7(5), pp.1731–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3110598 [Accessed May 15, 2018].

Tacke, R. et al., 1998. Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell*, 93(1), pp.139–48. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9546399 [Accessed November 8, 2017].

Tacke, R. & Manley, J.L., 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO journal*, 14(14), pp.3540–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7543047 [Accessed November 17, 2017].

Talhouarne, G.J.S. & Gall, J.G., 2014. Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes. *RNA (New York, N.Y.)*, 20(9), pp.1476–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25051970 [Accessed November 3, 2017].

Tange, T.Ø., Nott, A. & Moore, M.J., 2004. The ever-increasing complexities of the exon junction complex. *Current Opinion in Cell Biology*, 16(3), pp.279–284. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15145352 [Accessed November 3, 2017].

Tanner, N.K. & Linder, P., 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Molecular cell*, 8(2), pp.251–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11545728 [Accessed November 8, 2017].

Tarn, W.Y. & Steitz, J.A., 1995. Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. *Proceedings of the National Academy of Sciences of the United States of America*, 92(7), pp.2504–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7708674 [Accessed November 6, 2017].

Tarn, W.Y. & Steitz, J.A., 1994. SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes & development*, 8(22), pp.2704–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7958927 [Accessed December 13, 2017].

Tejedor, J.R. et al., 2015. Role of six single nucleotide polymorphisms, risk factors in coronary disease, in *OLR1* alternative splicing. *RNA*, 21(6), pp.1187–1202. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25904137 [Accessed November 3, 2017].

Thompson, N.L., Pearce, K.H. & Hsieh, H. V, 1993. Total internal reflection fluorescence microscopy: application to substrate-supported planar membranes. *European biophysics journal : EBJ*, 22(5), pp.367–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8112222 [Accessed November 8, 2017].

Thompson, N.L. & Steele, B.L., 2007. Total internal reflection with fluorescence correlation spectroscopy. *Nature Protocols*, 2(4), pp.878–890. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17446873 [Accessed November 8, 2017].

Tian, M. & Maniatis, T., 1993. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell*, 74(1), pp.105–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8334698 [Accessed November 8, 2017].

Tsuda, K. et al., 2011. Structural basis for the dual RNA-recognition modes of human Tra2-β RRM. *Nucleic Acids Research*, 39(4), pp.1538–1553. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20926394 [Accessed November 8, 2017].

Tuerk, C. & Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968), pp.505–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2200121 [Accessed November 4, 2017].

Turunen, J.J. et al., 2013. The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*, 4(1), pp.61–76. Available at: http://doi.wiley.com/10.1002/wrna.1141 [Accessed November 3, 2017].

Ulbrich, M.H. & Isacoff, E.Y., 2007. Subunit counting in membrane-bound proteins. *Nature methods*, 4(4), pp.319–21. Available at: http://www.nature.com/doifinder/10.1038/nmeth1024 [Accessed December 13, 2017].

Ule, J. et al., 2005. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*, 37(4), pp.376–386. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16314267 [Accessed November 4, 2017].

Valcárcel, J. et al., 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science (New York, N.Y.)*, 273(5282), pp.1706–9. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/8781232 [Accessed November 3, 2017].

Vámosi, G. et al., 2016. EGFP oligomers as natural fluorescence and hydrodynamic standards. *Scientific reports*, 6(1), p.33022. Available at: http://www.nature.com/articles/srep33022 [Accessed December 13, 2017].

Venables, J.P. et al., 2000. RBMY, a probable human spermatogenesis factor, and other hnRNP G proteins interact with Tra2beta and affect splicing. *Human molecular genetics*, 9(5), pp.685–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10749975 [Accessed November 8, 2017].

Venables, J.P. et al., 2005. Up-regulation of the ubiquitous alternative splicing factor Tra2β causes inclusion of a germ cell-specific exon. *Human Molecular Genetics*, 14(16), pp.2289–2303. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16000324 [Accessed November 8, 2017].

Villa, T. & Guthrie, C., 2005. The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes & development*, 19(16), pp.1894–904. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16103217 [Accessed November 3, 2017].

Wagner, E.J. & Garcia-Blanco, M.A., 2001. Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Molecular and Cellular Biology*, 21(10), pp.3281–3288. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11313454 [Accessed November 3, 2017].

Wahl, M.C., Will, C.L. & Lührmann, R., 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell*, 136(4), pp.701–18.

Waldo, G.S. et al., 1999. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*, 17(7), pp.691–695. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10404163 [Accessed December 13, 2017].

Wan, R. et al., 2016. The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science*, 351(6272), pp.466–475. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.aad6466 [Accessed December 13, 2017].

Wan, Y. & Wu, C.J., 2013. SF3B1 mutations in chronic lymphocytic leukemia. *Blood*, 121(23), pp.4627–4634. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23568491 [Accessed November 8, 2017].

Wang, H. & Chong, S., 2003. Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. *Proceedings of the National Academy of Sciences*, 100(2), pp.478–483. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12515863 [Accessed December 13, 2017].

Wang, J. et al., 2005. Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Research*, 33(16), pp.5053–5062. Available at: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki810 [Accessed November 4, 2017].

Wang, J. & Manley, J.L., 1995. Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. *RNA (New York, N.Y.)*, 1(3), pp.335–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7489505 [Accessed November 6, 2017].

Wang, Z. et al., 1996. Characterization of the Mouse Cyclin D3 Gene: Exon/Intron Organization and Promoter Activity. *Genomics*, 35(1), pp.156–163. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8661116 [Accessed November 4, 2017].

Wassarman, D.A. & Steitz, J.A., 1993. A base-pairing interaction between U2 and U6 small nuclear RNAs occurs in > 150S complexes in HeLa cell extracts: implications for the spliceosome assembly pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 90(15), pp.7139–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8346227 [Accessed November 3, 2017].

Wee, C.D. et al., 2014. Targeting SR Proteins Improves SMN Expression in Spinal Muscular Atrophy Cells M. Caputi, ed. *PLoS ONE*, 9(12), p.e115205. Available at: http://dx.plos.org/10.1371/journal.pone.0115205 [Accessed November 3, 2017].

Will, C.L. et al., 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *The EMBO journal*, 20(16), pp.4536–46. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11500380 [Accessed November 3, 2017].

Will, C.L., Behrens, S.E. & Lührmann, R., 1993. Protein composition of mammalian spliceosomal snRNPs. *Molecular biology reports*, 18(2), pp.121–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8232294 [Accessed November 17, 2017].

Woll, M.G. et al., 2016. Discovery and Optimization of Small Molecule Splicing Modifiers of Survival Motor Neuron 2 as a Treatment for Spinal Muscular Atrophy. *Journal of Medicinal Chemistry*, 59(13), pp.6070–6085. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27299569 [Accessed November 8, 2017].

Wongpalee, S.P. et al., 2016. Large-scale remodeling of a repressed exon ribonucleoprotein to an exon definition complex active for splicing. *eLife*, 5. Available at: https://elifesciences.org/articles/19743 [Accessed November 8, 2017].

Wu, J. & Manley, J.L., 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes & development*, 3(10), pp.1553–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2558966 [Accessed November 3, 2017].

Wu, J.Y. & Maniatis, T., 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, 75(6), pp.1061–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8261509 [Accessed November 6, 2017].

Xiao, S.H. & Manley, J.L., 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes & development*, 11(3), pp.334–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9030686 [Accessed November 6, 2017].

Xiao, S.-H. & Manley, J.L., 1998. Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *The EMBO Journal*, 17(21), pp.6359–6367. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9799243 [Accessed November 6, 2017].

Yan, C. et al., 2016. Structure of a yeast activated spliceosome at 3.5 A resolution. *Science*, 353(6302), pp.904–911. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27445306 [Accessed November 3, 2017].

Yildiz, A. et al., 2003. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science (New York, N.Y.)*, 300(5628), pp.2061–5. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1084398 [Accessed November 8, 2017].

Young, P.J. et al., 2002. SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. *Human molecular genetics*, 11(5), pp.577–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11875052 [Accessed November 8, 2017].

Zhang, X. et al., 2018. Structure of the human activated spliceosome in three conformational states. *Cell Research*, 28(3), pp.307–322. Available at: http://www.nature.com/doifinder/10.1038/cr.2018.14 [Accessed May 15, 2018].

Zhao, X. et al., 2016. Pharmacokinetics, pharmacodynamics, and efficacy of a small-molecule *SMN2* splicing modifier in mouse models of spinal muscular atrophy.

Human Molecular Genetics, 25(10), pp.1885–1899. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26931466 [Accessed November 8, 2017].

Zhou, Z. & Reed, R., 1998. Human homologs of yeast Prp16 and Prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. *The EMBO Journal*, 17(7), pp.2095–2106. Available at: https://reed.med.harvard.edu/sites/reed.med.harvard.edu/files/publications/1998 Zhou EMBO.pdf [Accessed November 3, 2017].

Zhu, J. & Krainer, A.R., 2000. Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes & development*, 14(24), pp.3166–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11124808 [Accessed November 6, 2017].

Zhu, J., Mayeda, A. & Krainer, A.R., 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Molecular cell*, 8(6), pp.1351–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11779509 [Accessed January 7, 2015].

Zorio, D.A.R. & Blumenthal, T., 1999. Both subunits of U2AF recognize the 3′ splice site in Caenorhabditis elegans. *Nature*, 402(6763), pp.835–838. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10617207 [Accessed November 3, 2017].

Zuo, P. & Manley, J.L., 1993. Functional domains of the human splicing factor ASF/SF2. *The EMBO journal*, 12(12), pp.4727–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8223481 [Accessed November 6, 2017].

Appendix.

8.1. Pre-mRNA Sequences.

Exon 1

Exon 2

Alternative exon 1

ESE sequence

Additional U1 sequence

Ad.

GGGAGGAGGACGGAGGACGGAGGACAUCGCUGUCUGCGAGGGCCAGCUGUUGGGGU

GAGUACUCCCUCUCAAAAGCGGGCAUGACUUCUGCGCUAAGAUUGUCAGUUUCCAA

AAACGAGGAGGAUUUGAUAUUCACCAGCUGUUGGGGUGAGUCCUUUGAGGGUGGCC

GCGUCCAUCUGGUCAGAAAAGACAAUCUUUUUGUUGUCAAGCUUGCUGCACGUCUAGG

GCGCAGUAGUCCAGGGUUUCCUUGAUGAUGUCAUACUUAUCCUGUCCCUUUUUUUUCC

ACAGCUCGCGGUUGAGGACAAACUCUUCGCGGUCUUUCCAGUACUCUUGGAUC

Globin.

GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGGGGACCTGTCCTCT

GCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTT

CAGTGAGGGTCTGAGTCACCTGGACAACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCA

CTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGTTC

TTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACTCTGCTAACCATGTTCATGCCTTCT

TCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGG

Globin M

GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGGGGACCTGTCCTCT
GCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTT
CAGTGAGGGTCTGAGTCACCTGGACAACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCA
CTGTGACAAGCTGCACGTGGATCCTGAGAACTT~~CAGGGTGAG~~TTTGGGGACCCTTGATTGTTC
TTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACTCTGCTAACCATGTTCATGCCTTCT
TCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGG

BG-SMN2

GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGGGGACCTGTCCTCT
GCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTT
CAGTGAGGGTCTGAGTCACCTGGACAACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCA
CTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGTTC
TTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACAGACTATCAACTTAATTTCTGATC
ATATTTTGTTGAATAAAATAAGTAAAATGTCTTGTGAAACAAAATGCTTTTTAACATCCATATA
AAGCTATCTATATATAGCTATCTATATCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTAC
AGGGTTTTAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATCAGGA

Additional ESE sequences

ESE A (1xRON) – **CAAGGCGGAGGAAG**

ESE B (2xcombined SELEX/GGA) – **CACACAGGACCACACAGGAC**

ESE C (2xSMN Tra2B) – **AAAAAGAAAGAAAAAAAGAAAGAA**

ESE D (2xKrainer) – **UCAGAGGAUCAGAGGA**

BG-SMN2+U1

GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGGGGACCTGTCCTCT
GCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTT
CAGTGAGGGTCTGAGTCACCTGGACAACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCA
CTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGTTC
TTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACAGACTATCAACTTAATTTCTGATC
ATATTTTGTTGAATAAAATAAGTAAAATGTCTTGTGAAACAAAATGCTTTTTAACATCCATATA
AAGCTATCTATATATAGCTATCTATATCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTAC
AGGGTTTTAGACAAAATCAAAAAGAAGGAAGGTGCTCACATTCCTTAAATCAGGTAAGT

Globin+U1

GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGGGGACCTGTCCTCT
GCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTCATGGCAAGAAGGTGCTGGCTGCCTT
CAGTGAGGGTCTGAGTCACCTGGACAACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCA
CTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGTTC
TTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACTCTGCTAACCATGTTCATGCCTTCT
TCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAGGTA
AGT

8.2. Sequences for ligation.

Pre-mRNA body for ligation

1 GGGCTGCTGGTTGTCTACCCATGGACCCAGAGGTTCTTCGAGTCCTTTGG

51 GGACCTGTCCTCTGCAAATGCTGTTATGAACAATCCTAAGGTGAAGGCTC

101 ATGGCAAGAAGGTGCTGGCTGCCTTCAGTGAGGGTCTGAGTCACCTGGAC

151 AACCTCAAAGGCACCTTTGCTAAGCTGAGTGAACTGCACTGTGACAAGCT

201 GCACGTGGATCCTGAGAACTTCAGGGTGAGTTTGGGGACCCTTGATTGTT

251 CTTTCTTTTTCGCTATTGTAAAATTCATGTTATATGGTCGACAGACTATC

301 AACTTAATTTCTGATCATATTTTGTTGAATAAAATAAGTAAAATGTCTTG

351 TGAAACAAAATGCTTTTTAACATCCATATAAAGCTATCTATATATAGCTA

401 TCTATATCTATATAGCTATTTTTTTTAACTTCCTTTATTTTCCTTACAGG

451 GTTTTAGACAAAATCGGTGCTCACA

Synthesised RNA strands

+ESE - UUCCUUAAAU – CAAGGCGGAGGAAGCAAGGCGGAGGAAG

+P+ESE - UUCCUUAAAU – 2HEG – CAAGGCGGAGGAAGCAAGGCGGAGGAAG

+abasic DNA +ESE - UUCCUUAAAU – 12 abasic DNA – CAAGGCGGAGGAAGCAAGGCGGAGGAAG

+abasic Ribo +ESE - UUCCUUAAAU – 12 abasic ribo – CAAGGCGGAGGAAGCAAGGCGGAGGAAG

+U1 - UUCCUUAAAU – CAGGUAAGUU

+PEG+U1 - UUCCUUAAAU – 2 HEG – CAGGUAAGUU

+Abasic DNA+U1 - UUCCUUAAAU – 12 abasic DNA – CAGGUAAGUU

+Abasic Ribo+U1 - UUCCUUAAAU – 12 abasic ribo – CAGGUAAGUU

(G/U on the end is because this was favorable for synthesis)

8.3. Sequence of SRSF1.

MSGGGVIRGP AGNNDCRIYV GNLPPDIRTK DIEDVFYKYG AIRDIDLKNR RGGPPFAFVE FEDPRDAEDA VYGRDGYDYD GYRLRVEFPR SGRGTGRGGG GGGGGGAPRG RYGPPSRRSE NRVVVSGLPP SGSWQDLKDH MREAGDVCYA DVYRDGTGVV EFVRKEDMTY AVRKLDNTKF RSHEGETAYI RVKVDGPRSP SYGRSRSRSR SRSRSRSRSN SRSRSYSPRR SRGSPRYSPR HSRSRSRT

RRM 1 is highlighted in red and RRM 2 is highlighted in cyan. The RS domain is highlighted in Green (Ge et al. 1991; Zuo and Manley 1993).

8.4. Sequences of oligonucleotides used in single molecule work.

2’O methyl modified nucleotides are shown in bold. Cy5 denotes a Cyanine-5-Dye. Biotin denotes a biotin group.

- Globin 5’ Cy5 Oligo = Cy5 - UAG ACA ACC AGC AGC CC - Biotin
- Ad 5’ Cy5 Oligo = Cy5 - ACC UGC AGG CAU GCA - Biotin
- Anti U1 Oligo = GCC AGG UAA GUA U - Biotin
- Anti U6 Oligo = CUG UGU AUC GUU CCA AUU UU

8.5. Single molecule Raw data.

Caption for the generated table	No. of Marker spots	No. of colocalized spots	Coloc. %	1	2	3	4	5	6	7	8	9	10	+	n.d.	Sum
BGSMN2 E (SRSF1)	614	100	0.163	50.0	30.0	12.0	6.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
BGSMN2 A (SRSF1)	1516	150	0.099	100.0	24.0	17.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	4.0	150.0
BGSMN2 A αU1 (SRSF1)	703	23	0.033	17.0	2.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	23.0
BGSMN2 A αU2 (SRSF1)	695	54	0.078	33.0	13.0	6.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	54.0
BGSMN2+E E (SRSF1)	656	100	0.152	41.0	34.0	15.0	8.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	100.0
BGSMN2+E A (SRSF1)	937	150	0.160	68.0	70.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	150.0
BGSMN2+E A αU1 (SRSF1)	1105	100	0.090	67.0	24.0	8.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
BGSMN2+E A αU2 (SRSF1)	1200	91	0.076	74.0	12.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	91.0
GloC (O) A (SRSF1)	820	172	0.210	136.0	29.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	172.0
GloC (M) A (SRSF1)	549	113	0.206	77.0	27.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	1.0	113.0
GloC (O) A αU1 (SRSF1)	1078	117	0.109	62.0	30.0	16.0	7.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	117.0
GloC (O) A αU1 αU2 (SRSF1)	607	36	0.059	22.0	9.0	3.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	36.0
GloM (O) A (SRSF1)	1619	142	0.088	95.0	24.0	11.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	142.0
SRSF1 Dimers +Rnase A	-	203	-	157.0	33.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	203.0
SRSF1 Dimers -Rnase A	-	129	-	118.0	9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	129.0
GloC(O) (SR/U1) SRSF1 A	1145	222	0.194	149.0	53.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	222.0
GloC(O) (SR/U1) U1A A	1145	240	0.210	167.0	38.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	29.0	240.0
GloC(O) (SR/U1) SRSF1 A (3 way)	1145	98	0.086	66.0	23.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	98.0
GloC(O) (SR/U1) U1A A (3 way)	1145	98	0.086	72.0	15.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	98.0
Glo2NT(C) (SR/U1) SRSF1 A	837	192	0.229	133.0	31.0	8.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	16.0	192.0
Glo2NT(C) (SR/U1) U1A A	837	248	0.296	164.0	42.0	7.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.0	248.0

Glo2NT(C) (SR/U1) SRSF1 A (3 way)	837	100	0.119	63.0	14.0	5.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	14.0	100.0
Glo2NT(C) (SR/U1) U1A A (3 way)	837	100	0.119	67.0	14.0	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	100.0
GloC17(M) (SR/U1) SRSF1 A	620	104	0.168	40.0	33.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.0	104.0
GloC17(M) (SR/U1) U1A A	620	99	0.160	45.0	40.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	99.0
GloC17(M) (SR/U1) SRSF1 A (3 way)	680	40	0.059	17.0	12.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	40.0
GloC17(M) (SR/U1) U1A A (3 way)	680	40	0.059	15.0	13.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	40.0
BGSMN2(M) (SR/U1) SRSF1 A	961	117	0.122	64.0	26.0	15.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	117.0
BGSMN2(M) (SR/U1) U1A A	961	199	0.207	151.0	36.0	6.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	199.0
BGSMN2(M) (SR/U1) SRSF1 A (3 way)	918	76	0.083	39.0	14.0	13.0	2.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	76.0
BGSMN2(M) (SR/U1) U1A A (3 way)	918	76	0.083	48.0	20.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	76.0
BGSMN2+U1(M) (SR/U1) SRSF1 A	950	145	0.153	51.0	50.0	27.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	145.0
BGSMN2+U1(M) (SR/U1) U1A A	950	204	0.215	89.0	85.0	22.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	204.0
BGSMN2+U1(M) (SR/U1) SRSF1 A (3 way)	982	99	0.101	39.0	26.0	21.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	99.0
BGSMN2+U1(M) (SR/U1) U1A A (3 way)	982	99	0.101	38.0	37.0	12.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	99.0
U1A/SRSF1	500	105	0.210	86.0	15.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	105.0
SRSF1/U1A	504	121	0.240	92.0	19.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	121.0
U1A/SRSF1 +Rnase A	574	84	0.146	67.0	5.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	84.0
SRSF1/U1A +Rnase A	472	93	0.197	74.0	16.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	93.0
SRSF1 Dimers -Rnase A	-	159	-	138.0	18.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	159.0
U1A Dimers -Rnase A	-	143	-	124.0	15.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	143.0
AdCC(O) (SR/U1) SRSF1 A	734	140	0.191	67.0	45.0	10.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.0	140.0
AdCC(O) (SR/U1) U1A A	734	170	0.232	95.0	52.0	7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.0	170.0
AdCC(O) (SR/U1) SRSF1 A (3 way)	816	86	0.105	41.0	20.0	7.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.0	86.0
AdCC(O) (SR/U1) U1A A (3 way)	816	86	0.105	49.0	26.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	86.0
GloC (O) A (SRSF1)	446	97	0.217	70.0	20.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	97.0
GloC (O) A aU1 (SRSF1)	551	66	0.120	40.0	13.0	7.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	66.0
GloM (O) A (SRSF1)	681	78	0.115	55.0	15.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	78.0

GloC (O) A (SRSF1)	506	102	0.202	68.0	19.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	102.0
SRSF1 Dimers	-	166	-	129.0	32.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	166.0
BGSMN2+U1 U2B''	391	107	0.274	80.0	15.0	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	107.0
BGSMN2+U1 U2AF65	619	84	0.136	65.0	12.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	84.0
BGSMN2+U1 U2AF35	619	122	0.197	91.0	16.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	122.0
BGSMN2 (SRSF1)	961	117	0.122	64.0	26.0	15.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	117.0
BGSMN2 (1) (SRSF1)	654	102	0.156	54.0	19.0	13.0	9.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	102.0
BGSMN2 (2) (SRSF1)	911	101	0.111	57.0	19.0	10.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	101.0
BGSMN2+R1 (SRSF1)	739	101	0.137	56.0	33.0	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	101.0
BGSMN2+R1 (1) (SRSF1)	632	95	0.150	50.0	32.0	6.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	95.0
BGSMN2+R1 (2) (SRSF1)	708	96	0.136	55.0	29.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	96.0
BGSMN2+R2 (SRSF1)	733	100	0.136	51.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	100.0
BGSMN2+R2 (1) (SRSF1)	590	100	0.169	49.0	34.0	7.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	100.0
BGSMN2+R2 (2) (SRSF1)	598	79	0.132	41.0	26.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	79.0
BGSMN2+R3 (SRSF1)	849	146	0.172	75.0	63.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	146.0
BGSMN2+R3 (1) (SRSF1)	580	101	0.174	52.0	38.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	101.0
BGSMN2+R3 (2) (SRSF1)	688	109	0.158	60.0	40.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	109.0
BGSMN2+R4 (SRSF1)	629	119	0.189	56.0	52.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	119.0
BGSMN2+R4 (1) (SRSF1)	583	106	0.182	51.0	40.0	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	106.0
BGSMN2+R4 (2) (SRSF1)	487	77	0.158	41.0	28.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	77.0
BGSMN2+E2 (SRSF1)	762	100	0.131	58.0	24.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	100.0
BGSMN2+E2 (1) (SRSF1)	741	99	0.134	58.0	26.0	3.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	99.0
BGSMN2+E2 (2) (SRSF1)	563	77	0.137	46.0	19.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	77.0
BGSMN2+E3 (SRSF1)	507	64	0.126	44.0	14.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	64.0
BGSMN2+E3 (1) (SRSF1)	732	100	0.137	58.0	25.0	4.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	100.0
BGSMN2+E3 (2) (SRSF1)	684	88	0.129	52.0	23.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.0	88.0
BGSMN2+E4 (SRSF1)	635	85	0.134	45.0	17.0	8.0	2.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	85.0

BGSMN2+E4 (1) (SRSF1)	618	81	0.131	48.0	17.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	81.0
BGSMN2+E4 (2) (SRSF1)	706	108	0.153	59.0	30.0	5.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	108.0
BGSMN2 (U2)	891	93	0.104	67.0	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0	93.0
BGSMN2+R4 (U2)	388	95	0.245	70.0	16.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	95.0
BGSMN2 (35)	512	63	0.123	46.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	63.0
BGSMN2(65)	512	53	0.104	37.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.0	53.0
BGSMN2+R4 (35)	449	110	0.245	78.0	16.0	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	110.0
BGSMN2+R4 (65)	449	73	0.163	49.0	13.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	73.0