Regulation of Human TMnm alternative pre-mRNA splicing

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Neil Thornton M.A. *Cantab*
Department of Biochemistry
University of Leicester

January 1998

UMI Number: U106138

UMI®

Dissertation Publishing

ProQuest®

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Contents

Chapter 3   Role of cis-acting sequences in the regulation of human TMnm pre-mRNA splicing

Chapter 4   Repression of exon SK in non muscle cells

Chapter 5   Identification of SK ESE binding proteins (SEBPs)

Chapter 6   Discussion

Appendix 1

Appendix 2 PhosphorImager data

References

Regulation of human TMnm alternative pre-mRNA splicing
Neil Thornton M.A. *Cantab*

Tissue-specific alternative splicing of pre-messenger RNA is a common strategy for regulating mammalian gene expression. This type of regulation is exemplified by several of the tropomyosin genes. One of these genes, the human tropomyosin$_{nm}$ (hTM$_{nm}$) gene, contains two alternative central exons, exon NM and exon SK. These exons are expressed tissue specifically and mutually exclusively; exon NM is expressed in non muscle cells and exon SK is expressed in skeletal muscle cells. Exon SK is specifically repressed in non muscle cells. This repression is dependent on sequence at the 5'-end of the exon that closely resembles a 5' splice site sequence. It is demonstrated here that U1 snRNP binds to this inhibitory sequence and mediates repression of exon SK in non muscle cells. In muscle cells, expression of exon SK is shown to require sequence within exon SK adjacent to the inhibitory sequence. Experiments *in vivo* showed that the activity of this enhancer sequence is dependent on a constitutively expressed factor(s). The sequence responded to SR proteins *in vitro*, thus highlighting a potential role for these proteins in regulation of hTM$_{nm}$ expression. Expression of mutants in muscle cells demonstrated that the inactivity of exon NM in these cells is dependent on the activity of the SK exon and on a reduction in the ability of the NM exon to be recognised by the splicing apparatus. It is proposed that interaction of U1 snRNP with the 5'-end of exon SK blocks recognition of the adjacent enhancer sequence in non muscle cells. In muscle cells, repression of exon SK is lifted and the activity of the enhancer sequence, together with a reduction in the activity of exon NM, allows exon SK to outcompete exon NM for recognition by the splicing apparatus.

Chapter 1
Introduction

Mammalian constitutive and alternative pre-messenger RNA splicing

Primary transcripts (pre-messenger RNA) produced from most protein coding genes in eukaryotic nuclei contain sequences (intron sequences) located between sequences comprising the mature messenger RNA (exon sequences). These intron sequences must be removed efficiently, and the exons ligated without disrupting the reading frames, to generate mature messenger RNA (mRNA) products which can be transported to the cytoplasm. This process is known as pre-mRNA splicing. The accuracy required by the splicing process is emphasised in some cases which involve ligation of exon sequences located far apart (the first intron of c-abl is over 200Kb long, Bernards *et al* 1987) and/or removal of several introns (the dystrophin gene contains 65 exons extending over more than two megabases, Feener *et al*, 1989).

The splicing reaction

Splicing of pre-mRNAs occurs in a two-step pathway (figure 1.1, reviewed in Sharp 1994):
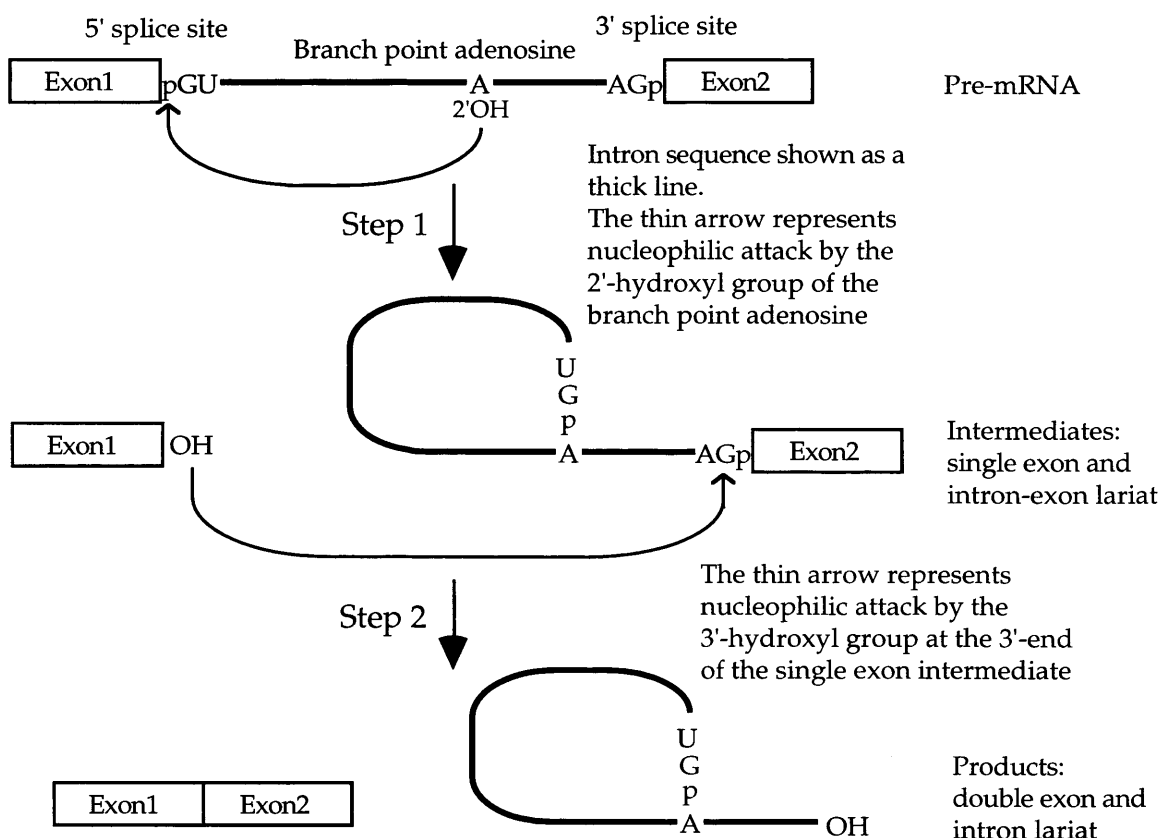
Figure 1.1 The catalytic steps of nuclear pre-mRNA splicing

In the first step the 2'-hydroxyl group of an adenosine near the 3'-end of the intron (the branch point adenosine) makes a nucleophilic attack on the 3', 5'-phosphodiester bond at the 5' splice site. Cleavage at the 5' splice site occurs at the same time as formation of a 2',5'-phophodiester bond between the 5' terminal guanosine of the intron and the branch point adenosine. A free 5' exon and a branched circular intron-3' exon known as a lariat is formed. In the second step the 3'-hydroxyl group of the free 5' exon attacks the phosphodiester bond at the 3' splice site. The lariat intron is displaced and the two exons are ligated together.

Cis-sequence requirements for mammalian pre-mRNA splicing

Comparison of several mammalian intron sequences has shown that the splice site junctions and the branch point adenosine are defined by consensus sequences. These consensus sequences and other *cis*-requirements are shown in figure 1.2.

Cis element	Sequence	Comment
5' splice site (5'ss)	(A/C)AG|GURAGU	"|" denotes the exon-intron boundary. R = purine. The GU at the 5'-end of the intron is invariant, the other positions vary.
3' splice site (3'ss)	YAG|G	Y = pyrimidine. The AG at the 3'-end of the intron is invariant.
Branch point sequence (BPS)	YNYURAY	N = any nucleotide. A denotes the branch point adenosine.
Polypyrimidine tract (PPT)	Pyrimidine-rich	Adjacent to and downstream of the BPS, before the 3'ss AG. In general, long PPTs are stronger than short ones.
Spatial requirements	Minimum distance between the 5'ss and branch point adenosine (BP) is ~48 nucleotides. Minimum distance between the BP and 3'ss is ~18 nucleotides. Thus, minimal mammalian intron size is ~66 nucleotides.	

Figure 1.2 Mammalian *cis*-sequence requirements for splicing[1]

(reviewed in Smith *et al*, 1989)

---

[1]A further class of pre-mRNA introns has recently been discovered. They have 5'-AT and 3'-AC boundaries and are referred to as AT-AC introns. The AT-AC introns studied so far share highly conserved sequences at the 5' splice site (\|ATATCCTT), branch site (TCCTTAAC), and 3' splice site (CCAC\|). These introns lack the polypyrimidine tract characteristic of most conventional 3' splice sites (Tarn and Steitz, 1996). This class of intron is beyond the scope of this text and they are not further discussed.

Spliceosome assembly

The splicing reaction does not occur free in solution but takes place within a multi-component complex known as the spliceosome. This comprises RNA-protein subunits known as small nuclear ribonucleoprotein particles (snRNPs) and several non-snRNP proteins. There are five snRNPs involved in splicing: U1, U2, U4, U5, and U6. Each snRNP comprises a molecule of RNA; a small nuclear RNA or snRNA (referred to as U1 snRNA, U2 snRNA, U4 snRNA, U5 snRNA, and U6 snRNA, respectively). The snRNAs are associated with several proteins, some of which are common to each snRNP (the core, or Sm proteins) and others which are specific. Each snRNP and several of the non-snRNP proteins are essential for splicing.

When pre-mRNA is added to nuclear extract it is immediately coated with proteins to form a complex. This complex is known as the H complex and its formation is independent of ATP and functional splice sites (Konarska and Sharp, 1987; Grabowski *et al*, 1985; Frendeway and Keller, 1985). Isolated H complexes cannot be chased into spliced products when incubated in nuclear extract in the presence of excess competitor RNA (Michaud and Reed, 1991) and are not, therefore, thought to form part of the spliceosome assembly pathway. A very basic outline of the mammalian spliceosome assembly pathway is given in figure 1.3 and described below (reviewed in Green, 1991; Wise, 1993; Lamond, 1993; Kramer, 1996).

5'ss   BP   PPT   3'ss

GU————————————A———————AG   Pre-mRNA

U1 snRNP   SR proteins
U2AF

U1   U2AF

SR   35
A————AG   Commitment complex E
65

U2 snRNP   ATP

U2
—————AG   Pre-splicing complex A

U4/U6.U5
tri-snRNP   ATP

U5
U4/U6

AG   Early splicing complex B

ATP

Late splicing complex C

ATP

Post-splicing complex
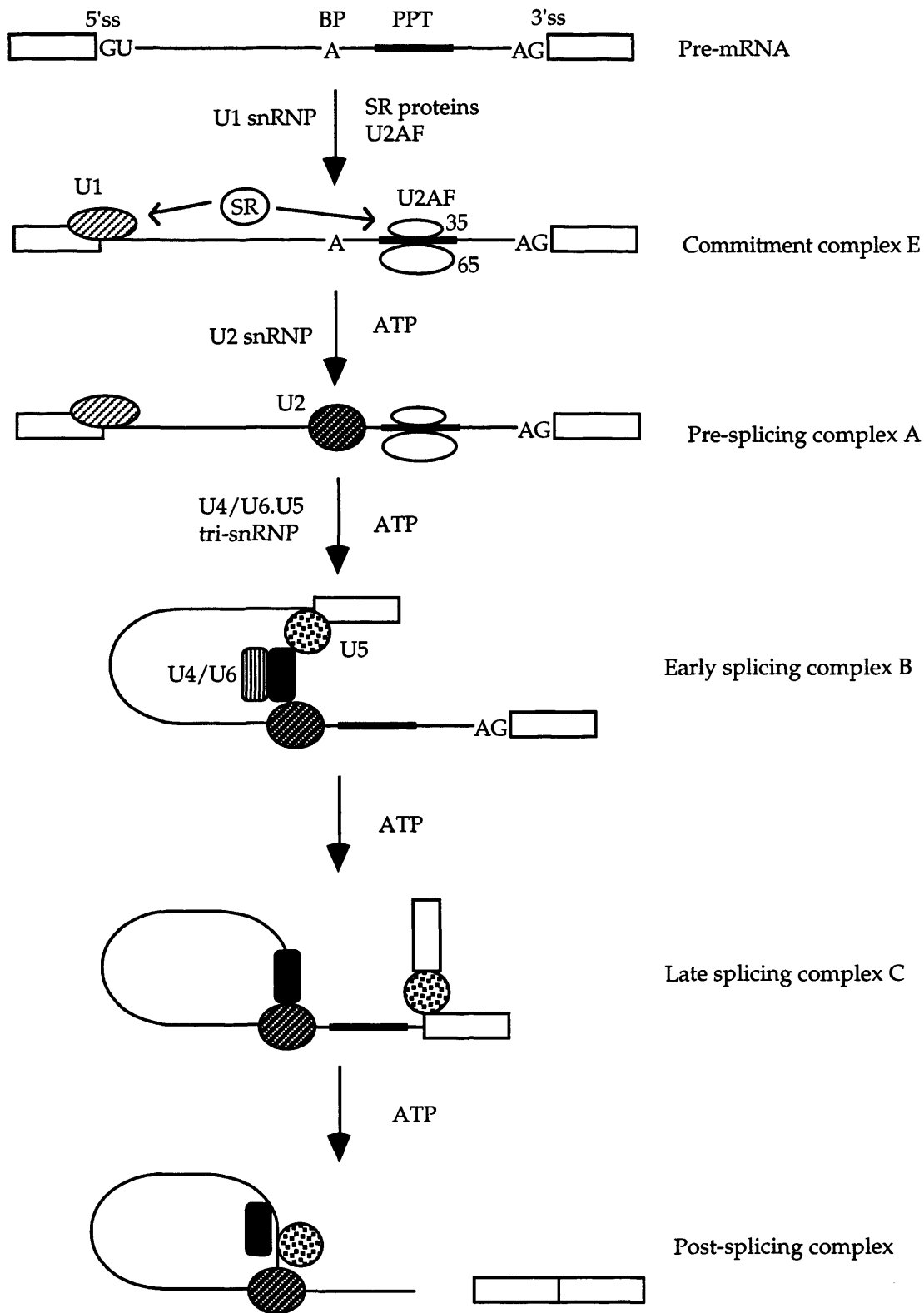
Figure 1.3 The spliceosome assembly pathway (adapted from Kramer, 1996)

See text for details

The first step in the pathway is the association of U1 snRNP with the 5'ss (mediated by base-pairing of the 5'-end of U1 snRNA with the 5'ss sequence) and binding of the 35Kd and 65Kd subunits of U2AF (a non-snRNP splicing factor) to the polypyrimidine tract (PPT). A stable complex called the commitment complex

1.4

(E complex) is formed. Formation of the commitment complex is dependent on the 5'ss and the branch point, but does not require ATP. U2AF is essential for the binding of U2 snRNP to the branchpoint sequence (BPS) which results in formation of the A complex. This association is mediated by base-pairing of an internal region of U2 snRNA with the BPS and, like all subsequent steps of the assembly process, requires ATP. The branch point adenosine does not base-pair with U2 snRNA and is bulged out from the helix formed by base-pairing of adjacent nucleotides with U2 snRNA. It is thought that this presents the adenosine as the nucleophile required for the first step of the splicing reaction. Assembly of the A complex is dependent on the PPT and a functional 3'ss. The 5'ss is not required.

The U4, U5, and U6 snRNPs join the spliceosome as a tri-snRNP complex to form complex B. An extensive base-pairing interaction between the U4 and U6 snRNAs is disrupted during a conformational rearrangement of the spliceosome that converts it into the catalytically active complex C. In this complex U6 snRNA forms contacts in the 5'-region of U2 snRNA and with the 5'ss sequence. U2 makes an additional contact with the first nucleotide of the second exon. Thus, it is thought that U2 and U6 snRNAs position the 5'ss and BP sequences adjacent to each other prior to catalysis. The interaction of U6 snRNA with the 5'ss is also important because it means that the 5'ss is recognised twice; once by U1 snRNA and once by U6 snRNA, enhancing the fidelity of 5'ss selection.

After the first cleavage reaction the 3' splice site sequence must be re-aligned relative to the 5'ss sequence to allow the second cleavage reaction to occur. A non-Watson Crick interaction between the invariant G nucleotides at the 5'- and 3'-ends of the intron is required for the second step of catalysis. It is proposed that U5 snRNP could promote this interaction and align the splice sites. U5 snRNA contacts the last nucleotide of the 5' exon before the first cleavage reaction. This interaction persists through both steps of splicing and an interaction with the first nucleotide of the 3' exon is formed after the first step. Alternatively, U5 snRNA may merely function to hold the 5' exon in the spliceosome after the first cleavage step. Following or during the second cleavage reaction U5 snRNA dissociates from the exon sequences and contacts nucleotides in the excised intron lariat. After catalysis the spliced mRNA is released from the spliceosome as a low molecular weight complex. The intron remains associated with U2, U5, and U6 snRNAs which are subsequently removed in an active process followed by degradation of the intron.

Thus, U2, U5 and U6 snRNAs are essential to the catalytic core of the spliceosome. The exact point at which U1 snRNP leaves the spliceosome is not known, but it may be displaced by the interaction of U6 snRNA with the 5'ss. The proteins which form part of the spliceosome are not thought to be directly involved in catalysis.

This discussion is primarily concerned with alternative pre-mRNA splicing which is principally regulated during the early stages of spliceosome assembly. These early stages are discussed in more detail below.

E complex

The stage in spliceosome assembly during which the 5' and 3' splice sites are first recognised is critical to splice site choice and is therefore a target step for the regulation of alternative splice site choice. The earliest complex in which recognition of the 5'ss and 3'ss sequences has been shown to occur is the E complex. This was first identified by Reed (1990). Complexes formed on pre-mRNA in HeLa cell nuclear extract in the absence of ATP were compared with complexes formed under similar conditions on RNA lacking splice sites. A complex unique to pre-mRNA (termed the early, or E complex) was shown to be enriched in U1 snRNA. Subsequent work (Michaud and Reed, 1991) demonstrated that the pre-mRNA in this complex could be chased into spliced products on the addition of splicing extract and excess competitor RNA. Thus, the pre-formed complex was stable to excess competitor RNA and contained pre-mRNA that was committed to the splicing pathway. Another group (Jamison et al 1992) also identified a complex with splicing commitment activity.

The sequence requirements for formation of the commitment activity were examined by both Michaud and Reed and Jamison et al. The commitment complex isolated by Jamison *et al* (1992) required both the 5' and 3' splice sites as well as the polypyrimidine tract (PPT). Michaud and Reed (1993) showed that a complex similar to E complex was present on RNA containing the 5'ss and surrounding sequence. The 5'ss sequence was necessary but not sufficient for formation of this E-like complex. A complex similar to E complex was also purified on RNA containing the 3'ss and surrounding sequence. The PPT was necessary for assembly of this complex. Given the similarity in sequence requirements and properties of the commitment complex identified by Michaud and Reed (1991) and Jamison *et al* (1992) it was proposed that these commitment complexes are identical.

Bennett *et al* (1992) investigated the protein composition of the E complex. They affinity purified complexes isolated by gel filtration then analysed the protein composition of these complexes on 2D gels. The E complex was shown to contain the U1 snRNP-specific proteins U1 70K and U1A as well as two snRNP core proteins B and B'. The complex also contained the non-snRNP protein U2AF and five other spliceosome associated proteins (SAPs). The E-like complexes formed on RNA containing a 5'ss or a 3'ss were demonstrated to be enriched in U1 snRNP and U2AF, respectively (Michaud and Reed, 1993). It was known that U1 snRNP binds to the 5'ss (Mount *et al*, 1983) and that U2AF binds

to the 3'ss (Zamore and Green, 1989), so it was proposed that E complex contains U1 snRNP bound to the 5'ss and U2AF bound to the 3'ss (Michaud and Reed, 1993).

U2AF was first identified (Ruskin *et al*, 1988) as a component of micrococcal nuclease-treated HeLa cell nuclear extract that was required for the binding of U2 snRNP to the branch point sequence (BPS). Other experiments demonstrated that U2AF bound to the pre-mRNA before U2 snRNP and that the PPT and 3'ss are required for this binding. U2AF was shown to have high affinity for poly(U) sequences (Zamore and Green, 1989), so this binding specificity was exploited in order to deplete HeLa cell nuclear extract of U2AF activity. These nuclear extracts were then inactive for splicing and pre-spliceosome formation. Addition of purified U2AF to these extracts restored their ability to splice and form prespliceosomes (Zamore and Green, 1991). The sole involvement of U2AF in this reconstitution activity was later confirmed with recombinant U2AF (Zamore *et al*, 1992).

Michaud and Reed (1993) showed that E complex assembly occurred more efficiently on RNA that had both splice sites intact. Thus, a functional interaction between the two sites appears to occur during formation of E complex. Where alternative 5' and/or 3' splice sites exist, regulation of alternative splice site choice is expected to occur at or before this stage of spliceosome assembly. It was proposed that the interaction between 5'ss and 3'ss sequences in E complex could be mediated by a direct or indirect interaction between U1 snRNP bound at the 5'ss and U2AF bound at the PPT. An indirect interaction could be mediated by one of the SAPs shown to be present in E complex (Bennett *et al*, 1992). Evidence that an indirect interaction occurs between U1 snRNP and U2AF came from an observation that a protein known as SC35 was required for interaction of U1 snRNP with the 3' splice site (Fu and Maniatis, 1992). SC35 is a member of a growing family of highly conserved proteins known as SR proteins.

SR proteins

There are currently nine members of the human SR protein family that have been cloned and sequenced (Table 1.1). The first SR protein to be identified was SF2/ASF. It was purified independently by two groups based on its ability to complement S100 extracts (a post nuclear cytoplasmic extract) for splicing activity (Krainer *et al*, 1990, 1991) and to change the relative use of two alternative 5' splice sites in SV40 pre-mRNA (Ge *et al* 1990, 1991). Subsequently, SC35 was identified by a monoclonal antibody raised against purified spliceosomes and was demonstrated to be required for splicing and spliceosome assembly *in vitro* (Fu and Maniatis 1990). The sequences of the two proteins revealed extensive regions of homology between them (Krainer et al, 1991; Ge *et al*, 1991). Further, studies *in*

Table 1.1 Human SR proteins

Name	Reference(s) in which protein cloned	Comment
SRp20	Zahler et al (1992)	Identical to mouse X16 (Ayane et al, 1991). Originally identified by mAb104 after two-step salt precipitation procedure (Zahler et al 1992). Cloned using probes based on peptide sequence.
SRp30a/SF2/ASF	Krainer et al (1991) Ge et al (1991)	Purified as a factor that reconstitutes S100 extracts for splicing ativity (Krainer et al, 1991) and as a factor that alters relative 5' splice use in SV40 pre-mRNA (Ge et al, 1991). Cloned using probes based on peptide sequence.
SRp30b/SC35	Fu and Maniatis (1992)	Identified by a monoclonal antibody raised against partially purified spliceosomes. Cloned using probes based on peptide sequence Identical to chicken PR264 clone.
SRp30c	Screaton et al (1995)	Obtained after screening of a cDNA library with a probe containing conserved sequences from previously characterised SR proteins.
SRp40	Screaton et al (1995)	Originally identified using same procedure as SRp20 (Zahler et al, 1992) then cloned in same screen as SRp30c.
SRp55	Screaton et al (1995)	Originally identified using same procedure as SRp20 (Zahler et al, 1992) then cloned in same screen as SRp30c. Homologous to drosophila B52 (dSRp55).
SRp75	Zahler et al (1993)	Originally identified using same procedure as SRp20 (Zahler et al, 1992) then cloned using probes based on peptide sequence.
9G8	Cavaloc et al (1994)	Recognised by a monoclonal antibody that inhibited splicing in vitro.
p54	Chaudhary et al (1991)	Originally isolated after screening an expression library with anti-lamin B antibodies.

vitro showed that SC35 was functionally interchangeable with the S100 complementing and alternative splice site switching activites of ASF/SF2 (Fu *et al* 1992). SF2/ASF and SC35 were shown to be members of a highly conserved family of proteins now known as SR proteins (Zahler *et al*, 1992). Six SR proteins were purified using a simple two-step salt precipitation procedure (with ammonium sulphate, then magnesium chloride, Zahler *et al*, 1992). The proteins were visualised during purification with a monoclonal antibody (mAb104) previously shown to interact with transcriptionally active sites in *Xenopus* lampbrush and *Drosophila* polytene chromosomes (Roth *et al* 1991). Partial amino acid sequence revealed that the proteins were all highly related and that two were identical to ASF/SF2 and SC35 (designated SRp30a and SRp30b, respectively, by Zahler *et al*, 1992). The remaining proteins were visible as bands running at 20, 40, 55 and 75Kd and were named SRp20, SRp40, SRp55 and SRp75 respectively. The conservation of SR protein activity throughout evolution was demonstrated when the *Drosophila* homologue of human SRp55 (dSRp55) was tested for its ability to function in the S100 complementation and alternative splice site switching assays of Krainer et al (1990a, 1990b) and Ge *et al* (1990). The dSRp55 protein had activity indistinguishable from ASF/SF2 (Mayeda *et al*, 1992).

The SR proteins share a common domain structure. The N-terminus contains sequences with strong homology to the 80 amino acid RNP-type recognition motif (RRM). This motif is present in many known RNA binding proteins. Typically, it has an 80 amino acid region containing two internal consensus elements RNP-1 and RNP-2 which are eight and six nucleotides long, respectively. Zahler *et al* (1992) also noted that human SRp30a (ASF/SF2), SRp40, SRp55 and SRp75 contained a repeat of the RRM motif. The sequence of this repeat is not as highly conserved as that of the N terminal domain (RRM homology, or RRMH). At the C-terminal end of each protein, several clusters of alternating arginine (R) and serine (S) residues comprise an RS domain. This region was noted to have homology to genetically identified splicing regulators in *Drosophila* and to a phylogenetically conserved region of the U1 70K protein (Krainer *et al* 1991; Ge *et al* 1991).

The activity of the individual proteins purified after the two-step purification of SR proteins was tested by Zahler *et al* (1992) in the S100 complementation assay (Krainer *et al*, 1990, 1991). They were all shown to restore splicing activity. It was concluded that there were several defining criteria for SR proteins:

    1) recognition by mAb104;

    2) copurification using a two-step salt precipitation procedure;

    3) the presence of at least one N terminal RRM domain and a C terminal RS domain;

    4) complementation of S100 extract for splicing activity.

Since the initial identification of a family of SR proteins, several more have been discovered. Details of these are given in table 1.1 and their domain structure is shown in figure 1.4.
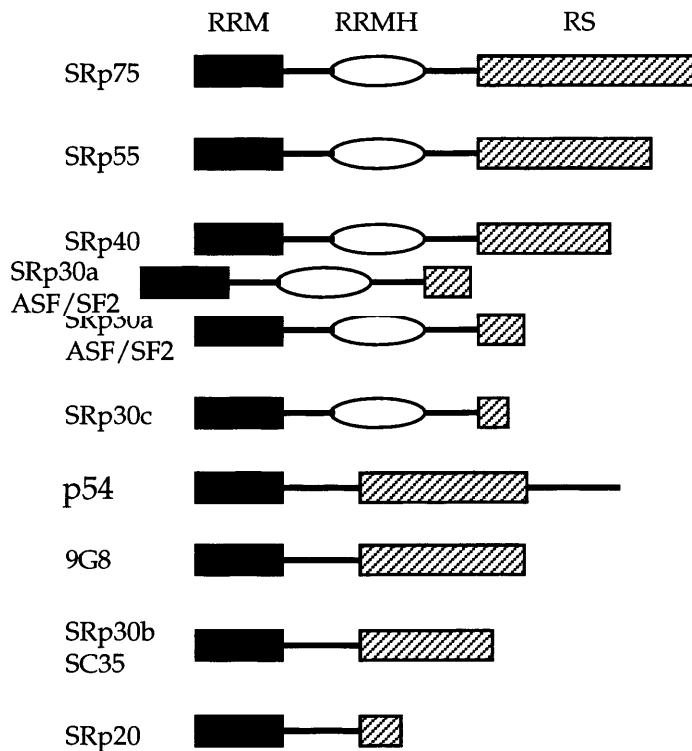
Figure 1.4 Domain structure of human SR proteins

RRM domain shown as black box, RRMH domain as open oval, RS domain as shaded box. Other sequence shown as thin line. Further details in text.

SR proteins and the early stages of spliceosome assembly

Wu and Maniatis (1993) noted that the RS domain present in SR proteins was also present in the U1 70K protein specific to U1 snRNP and in both subunits of the U2AF protein. They proposed that the functional interaction shown to occur between the 5' and 3' splice site sequences during E complex assembly (Michaud and Reed, 1993) could be mediated by interaction of the RS domain of SR proteins with the RS domain of U1 70K as part of U1 snRNP bound to the 5'ss, and with one, or both of the RS domains of U2AF bound to the PPT. Using the yeast two-hybrid assay they demonstrated that SC35 and SF2/ASF can interact simultaneously with U1 70K and U2AF^{35} proteins and that U2AF^{35} can interact simultaneously with U2AF^{65} and SC35 or SF2. The specificity of these interactions was also demonstrated because U2AF^{65} did not interact with U1 70K or with SC35 or SF2/ASF but did interact strongly with U2AF^{35} (Wu and Maniatis 1993; Zhang *et al*, 1992). A model for 5' and 3' splice site pairing was proposed:

1.9
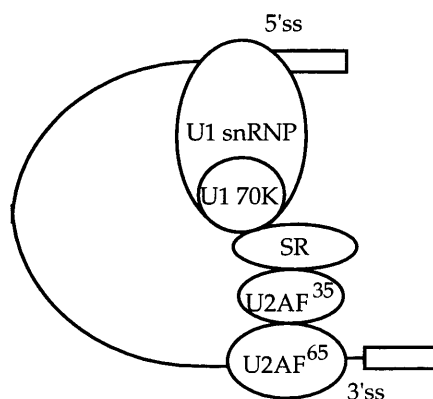
Figure 1.5 Model for 5' and 3' splice site pairing in E complex

The model proposed by Wu and Maniatis (1993) predicts that E complex assembly will be stimulated by SR proteins under conditions in which SR proteins are limiting. The effect of SR proteins on E complex formation was examined by Staknis and Reed (1994). Addition of recombinant SC35 to nuclear extracts significantly enhanced the formation of E complex. This stimulation was dependent on the presence of functional splice sites. Recombinant SC35 also enhanced the formation of complexes similar to E complex formed on RNA containing only a 5' or 3' splice site and surrounding sequence. This observation was interesting because it suggests that SR proteins may be able to recruit U1 snRNP and U2AF to the 5'ss and 3'ss respectively, and act as bridging components between the 5' and 3' splice sites.

The conclusion that SR proteins are part of the E complex prompted Staknis and Reed (1994) to investigate why these proteins had not been purified as components of this complex before (Michaud and Reed 1993). The original protocol involved a two-step purification procedure in which complexes assembled on biotinylated pre-mRNA were isolated by gel filtration followed by avidin affinity chromatography. E complex was known to be functional after gel filtration (Michaud and Reed, 1991) so it was proposed that SR proteins were still associated with the complex at this stage, but dissociated during the affinity purification step. Staknis and Reed (1994) modified the purification procedure by UV-crosslinking the complex after purification by gel filtration. The affinity purification step was missed out because it was found that this step led to heavy contamination of E complex by H complex; the latter complex is affinity purified much more efficiently. Use of this modified procedure showed that SR proteins can be crosslinked to the pre-mRNA in E complex and that this signal is dependent on functional 5' and 3' splice site sequences. A slight alteration to the model of Wu and Maniatis (1993) for the protein-RNA and protein-protein interactions involved in splice site pairing was proposed (figure 1.6).

1.10
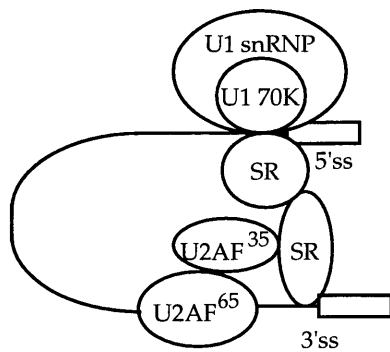
Figure 1.6 Modified model for 5' and 3' splice site pairing in E complex (Staknis and Reed, 1994)

Recent work has shown that U2 snRNP is also associated with E complex (Hong *et al*, 1997). Antibodies specific to protein components of U2 snRNP were able to efficiently immunoprecipitate E complex. The association of U2 snRNP however, is thought to be weak because E complex could not be immunoprecipitated by a U2 snRNP component-specific antibody in high salt. The functional relevance of U2 snRNP in E complex is yet to be understood.

Recruitment of U1 snRNP to the 5'ss by SR proteins

Mammalian 5'ss sequences are very diverse in sequence (Mount, 1982). It has been estimated that a candidate 5'ss would be found approximately every 100 nucleotides in a random RNA sequence (Eperon *et al*, 1993). Because mammalian introns can be as long as 200Kb (Bernards *et al*, 1987) and on average are ~1Kb in length (Hawkins, 1988; Traut, 1988) the selection of a 5'ss must be dependent on more than the degree to which the sequence of the site matches the consensus sequence. The demonstration that SR proteins promote specific recognition of the 5'ss sequence by U1 snRNP (Staknis and Reed, 1994) provides a basis for an explanation of 5'ss selection. The domain structure of SR proteins is highly suggestive of a role for these proteins in recruitment of U1 snRNP to the 5'ss. Their RRM domain could recognise specific splicing signals in RNA; either the 5'ss sequence itself or sequences nearby, and their RS domain could be involved in an interaction with the U1 70K protein specific to U1 snRNP. Consistent with this proposal, both the RRM and RS domains of ASF/SF2 have been shown to be required for the activity of this protein in constitutive splicing (Caceres and Krainer, 1993; Zuo and Manley, 1993).

Fu (1993) demonstrated that a single SR protein was sufficient to commit pre-mRNA to the splicing pathway. This work suggested that SR proteins could function in the earliest stages of spliceosome assembly by recruiting components of E complex to the pre-mRNA. Evidence that SR proteins are able to recruit U1 snRNP to the 5'ss has now come from several studies. SF2/ASF has been shown

to stabilise the interaction of U1 snRNP at 5' splice sites (Eperon *et al*, 1993). Crosslinking of U1 snRNP to two alternative 5' splice sites was shown to be dependent on the presence of SR proteins (Zahler and Roth 1995). Kohtz *et al* (1994) have reported that ASF/SF2 and purified U1 snRNP were involved in a cooperative interaction in binding RNA containing an intact 5'ss. This cooperation was dependent on the RS domain of both the U1 70K protein and SF2/ASF and on the addition of SF2/ASF to the pre-mRNA before the addition of U1 snRNP. Subsequent experiments examined the sequence and protein requirements for the formation of this ternary complex (Jamison *et al*, 1995). It was established that a 26 nucleotide RNA spanning a consensus 5'ss sequence was sufficient and that either one of the RRM domains within ASF/SF2 were required.

The ability of SR proteins to recruit U1 snRNP to the 5'ss suggests that they are able to bind specifically to the 5'ss or to sequences nearby. The identity of a selected splice site would then be dependent on the affinity of the SR protein for sequences at or near the 5' splice site and the extent to which the 5'ss sequence matched the consensus sequence. Consistent with this, ASF/SF2 was shown to have specific affinity for the 5'ss sequence of two different pre-mRNAs (Zuo and Manley, 1994). RNA sequences with high affinity to SC35 and ASF/SF2 were identified *in vitro* using a repetitive selection procedure (Tacke and Manley, 1995). Both proteins selected RNA sequences rich in purine residues. This was consistent with SR proteins being able to bind specifically to 5'ss sequences because these sequences are themselves purine-rich (the consensus 5'ss sequence consists of 7 purine residues out of 9). In fact, one of the two consensus sequences selected by SC35 contained a 7 out of 9 match to the consensus 5'ss sequence. Commitment of HIV tat pre-mRNA to splicing by ASF/SF2 (Fu 1993) could be explained by the presence of several RNA sequence motifs that closely resemble the consensus ASF/SF2 sequence selected by Tacke and Manley (1995).

However, not all the evidence is consistent with SR proteins binding to high affinity sites at or near the 5'ss and promoting the binding of U1 snRNP to the 5'ss sequence. β-globin pre-mRNA which was committed to splicing by SC35 (Fu 1993) does not contain any obvious matches to the sequence with high affinity for SC35 identified by Tacke and Manley (1995). Further, in an example of alternative 5'ss selection in which upstream 5'ss use is promoted by SRp40 and downstream 5'ss use is promoted by SRp30b, 50 nucleotides of sequence containing each 5'ss sequence has been shown not to be involved in the differential selection (Zahler and Roth, 1995).

No detectable complex is formed between SF2/ASF and the 26 nucleotide RNA spanning a 5'ss sequence prior to formation of the stable ternary complex after addition of U1 snRNP (Jamison *et al*, 1995). It was concluded that SF2/ASF can stabilise the interaction of U1 snRNP with the 5'ss and that U1 snRNP can

stabilise the interaction of ASF/SF2 with the RNA (Jamison *et al*, 1995). Thus, there may be no requirement for SR proteins to bind specifically to RNA at or near the 5'ss. Indeed, the cooperative interaction between U1 snRNP and SF2/ASF in RNA binding suggests that U1 snRNP acts to recruit SR proteins to the pre-mRNA. This is consistent with experiments showing that a high concentration of SR proteins can substitute for the function of U1 snRNP in splicing (Tarn and Steitz, 1994; Crispino *et al*, 1994). It has been proposed that for weak 5'ss sequences, high affinity SR protein binding sites are required in order to increase the local concentration of SR proteins for recruitment of U1 snRNP. For strong 5'ss sequences, U1 snRNP binding to the 5'ss is less dependent on SR proteins, but U1 snRNP then acts acts to recruit SR proteins which are required for the subsequent interaction with U2AF during E complex assembly (Jamison *et al*, 1995).

Recently, approximately 35 more proteins related to the SR family have been discovered (Blencowe *et al*, 1996). These proteins, in common with the known SR proteins, are selectively precipitated in magnesium chloride and remain soluble after magnesium precipitation in high salt. Many are recognised by three different monoclonal antibodies which react with distinct phosphoepitopes on the known SR proteins. The identity of many of these SR-related proteins is not yet known, although two have been shown to be nuclear matrix antigens. The existence of so many SR family members implies that their role may extend beyond one in the early stages of spliceosome assembly. Consistent with this, distinct subsets of SR-related proteins were demonstrated to remain bound to intron and to exon sequences following the second step of the splicing reaction.

Exon definition

Robberson *et al* (1990) compared the ability of different RNAs to assemble specific complexes in HeLa cell nuclear extract. They found that RNA containing 3' intron sequence adjacent to a downstream exon complete with 5'ss sequence assembled complexes much more efficiently than similar RNA lacking the 5'ss sequence. It was proposed that factors bound to the 5' and 3' splice sites could interact across the exon. The functional significance of the 5'ss-dependent stimulation of complex assembly on RNA containing an upstream 3'ss was demonstrated when it was shown that the splicing efficiency of pre-mRNA containing two exons separated by an intron (single intron pre-mRNA) was stimulated if the downstream exon also contained a 5'ss sequence. Subsequent experiments using pre-mRNA containing three exons separated by two introns (double intron pre-mRNA) also demonstrated that mutations in the 5'ss

sequence of the second intron could inhibit splicing of the first exon to the second exon (Talerico and Berget 1990).
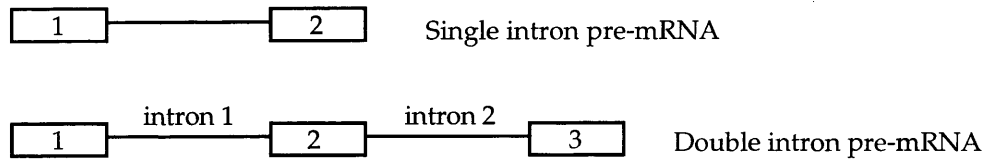
It was proposed that the splicing machinery searches downstream for a 5'ss after recognition of the 3'ss. Interactions between factors bound to the 5' and 3' splice sites then define the exon. Subsequent interactions between the 5' and 3' splice sites across introns must then occur before catalysis. Thus, exons rather than introns are the intial recognition units for assembly of the spliceosome.

Robberson *et al* (1990) noted that several features of splicing are consistent with the exon definition model for splice site selection. The model predicts that 5' and 3' splice site sequences will be found close together. Consistent with this, the length of nearly all exons is limited to 300 nucleotides. Further, the results of an analysis of phenotypes of known mutations of vertebrate 5'ss sequences terminating internal exons revealed that skippng of the upstream exon usually occurs. This is expected if exons must be defined before efficient spliceosome assembly can occur.

Further evidence for interactions across exons between components bound to 5' and 3' splice sites came from studies of preprotachykinin pre-mRNA. The 3' splice site of exon 4 is inherently unreactive. It was shown, however, that removal of the intron upstream of exon 4 occurred when the 5' splice site of exon 4 was mutated to improve its complementarity to the 5'-end of U1 snRNA or when the 5'ss of exon 5 was brought into proximity with the exon 4 3' ss by the removal of the intron separating exons 4 and 5 (Nasim *et al*, 1990). The exact nature of the preprotachykinin exon 4-spanning interactions were investigated by Hoffman and Grabowski (1992). They observed that crosslinking of the 65Kd subunit of U2AF to the polypyrimidine tract upstream of exon 4 was enhanced by mutations in the exon 4 5'ss sequence that improved its match to the consensus sequence. This crosslinking was also dependent on the 5' end of U1 snRNA and U1 snRNP. A complementation assay demonstrated that purified U1 snRNP and U2AF was not sufficient for U1 snRNP-facilitated binding of U2AF to the PPT. This activity was, however, restored by the addition of micrococcal nuclease-treated extracts deficient in U1 snRNP and U2AF to these components.

The work on preprotachykinin provides evidence for interactions across exons but argues against the involvement of a 5' to 3' scanning mechanism in 5' splice site definition. The requirement that the intron separating exons 4 and 5 must be spliced before removal of the intron separating exons 3 and 4 can occur clearly shows that the 5'ss of exon 4 is functional before the 3'ss of exon 4. It is

difficult to see how an undefined 3'ss could direct the identification of a downstream 5'ss. Instead interactions across the exon, at least in this pre-mRNA, seem to be required for definition of the upstream 3'ss.

The involvement of U1 snRNP and U2AF in exon definition suggests that the interactions proposed to occur across introns within E complex (Wu and Manitais, 1993) could also be responsible for the interaction of 5' and 3' splice sites across exons. Thus an involvement of SR proteins in these interactions is predicted (figure 1.7). Direct biochemical evidence for such interactions is currently lacking.
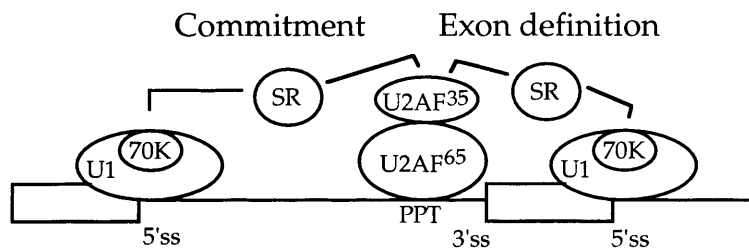
Figure 1.7 Early recognition of splice sites

Exon definition is likely to be most important in the selection of weak splice sites. These are often associated with regulated exons. Binding of U1 snRNP or U2AF to a weak 5'ss or PPT, respectively, will be enhanced if they are constrained to the vicinity of the weak site by interaction (via an SR protein(s), figure 1.7) with U2AF or U1 snRNP, respectively, bound to a splice site nearby. In a number of cases, splicing to weak 3' splice sites (these tend to have short polypyrimidine tracts and/or branchpoint sequences which do not match the consensus sequence very well) has been shown to be dependent on the presence of a splicing signal located in the adjacent exon known as an exon splicing enhancer (ESE). Here, it appears that the exon-bridging interactions are not required because SR proteins are able to bind directly to the ESE and stimulate interaction of U2AF with the PPT. ESEs are dicussed in detail later on.

The 5'-end of RNA polymerase II transcripts all contain a monomethyl guanosine cap. This is recognised by a nuclear cap-binding complex (CBC) comprising two cap binding proteins (CBP), CBP80 and CBP20 (Izaurralde *et al*, 1994). The exon definition model predicts that the 5'-terminal exon of mammalian pre-mRNAs is defined via exon-bridging interactions involving U1 snRNP bound at the 5'ss sequence and a component(s) bound upstream. Evidence that this upstream component could involve CBC first came when the splicing efficiency of an adenoviral pre-mRNA was shown to be reduced in extracts immunodepleted of CBC (Izaurralde *et al*, 1994). Subsequent experiments using a single-intron pre-mRNA demonstrated that CBC was required for efficient recognition of the 5' splice site by U1 snRNP during E complex

formation. In contrast, CBC was not required for splicing of the cap distal intron of a double-intron pre-mRNA (Lewis *et al*, 1996).

Details of the mechanism by which CBC affects recognition of the 5'ss by U1 snRNP are not known. Readdition of CBC, or addition of excess SR proteins to CBC-depleted extracts was demonstrated to stimulate the U1 snRNP-5'ss interaction. However, neither CBP80 or CBP20 contain an RS domain, and in a highly purified reconstituted system, addition of SR proteins had no effect on the dependence of the U1 snRNP-5'ss interaction on CBC. This argues against a role for SR proteins in mediating the effect of CBC on U1 snRNP. Involvement of hnRNP F has recently been proposed based on the observation that this protein can interact with CBC20 and CBC80 *in vitro*, and that hnRNP F binds preferentially to CBC-RNA complexes over naked RNA (Gamberi *et al*, 1997).

Little is known about the interactions that define 3'-terminal exons.

Alternative pre-mRNA splicing

Changes in splicing pattern can occur in tissue-specific or developmentally regulated fashion. Alternative splicing is, therefore, recognised as an important mechanism involved in the regulation of gene expression. Unlike regulation at the transcriptional level, this mechanism allows changes in expression which can add functional diversity to the product of a single gene. Gene rearrangement also generates protein diversity from single genes, but at the cost of permanent changes to the cell's genetic content. Consequently, this process is only observed in cells which are committed to terminal differentiation. Examples are the production of antibodies and T cell receptors. Combinatorial rearrangement of exons (V exons, D exons, and J exons) has the potential to generate vast numbers of antibodies or receptors, respectively. Alternative splicing represents an efficient way of using genomic information without the need for changes in genetic content. The alternative transcripts produced often differ in coding capacity and give rise to proteins with subtly or extremely altered properties. Some of the modes and consequences of alternative splicing are shown in figure 1.8 and table 1.2, respectively.

Effect	Gene	Comments
Protein localisation	Immunoglobulin μ heavy chain gene	Alternative 3' exons. M1 and M2 exons code for hydrophobic membrane-binding segment. These are used in early B lymphocytes (membrane bound immunoglobulin μ, $Ig\mu_m$). After antigen stimulation, S exon used and Igμ secreted ($Ig\mu_s$). The alternative terminal exon use is regulated by polyadenylation. The M2 exon poly (A) signal is stronger than the S exon poly (A) signal because it has a more extensive GU/U-rich region which is bound with higher affinity by the CstF polyadenylation factor. In early B cells the level of CstF is limiting and only the GU/U-rich region of the M2 exon poly (A) site is bound by CstF. Consequently only the M2 poly (A) site is used. On B cell maturation, the level of Cst F increases and the GU/U-rich region of the S exon poly (A) site also becomes bound by CstF. Use of the S exon poly (A) site is then favoured (Reviewed in Proudfoot, 1996).
Disruption of protein activity	*Drospophila* P element transposase	Alternative intron retention. See below.
Subtle changes in protein function	Pyruvate kinase	Mutually exclusive exons. M1 isoform found in muscle and has Michaelis-Menten kinetics with phosphoenolpyruvate (PEP) substrate. M2 isoform is allosterically regulated by PEP, ATP, alanine, fructose-1, 6-biphosphate.
Change in activity of protein	Calcitonin (CT)/calcitonin gene related peptide (CGRP) gene	Alternative splicing and poly(A) site use. In thyroid C cells exon 4 used. Contains CT coding sequence and poly(A) site. In certain neural cells exon 4 skipped. Exons 5 and 6 contain CGRP coding sequence and poly(A) site. CT functions as an inhibitor of bone resorption to decrease circulating levels of calcium and phosphate. CGRP functions in the peripheral and central nervous sytems as a neurotransmitter.
mRNA stability	Colony stimulating factor-1 (CSF-1) gene	Alternative 3' UTRs, one has AUUUA motif associated with rapidly turning over mRNAs. Suggests that RNA stability may be regulated by alternative splicing

Table 1.2 Functional consequences of alternative splicing (reviewed in Smith *et al*, 1989)
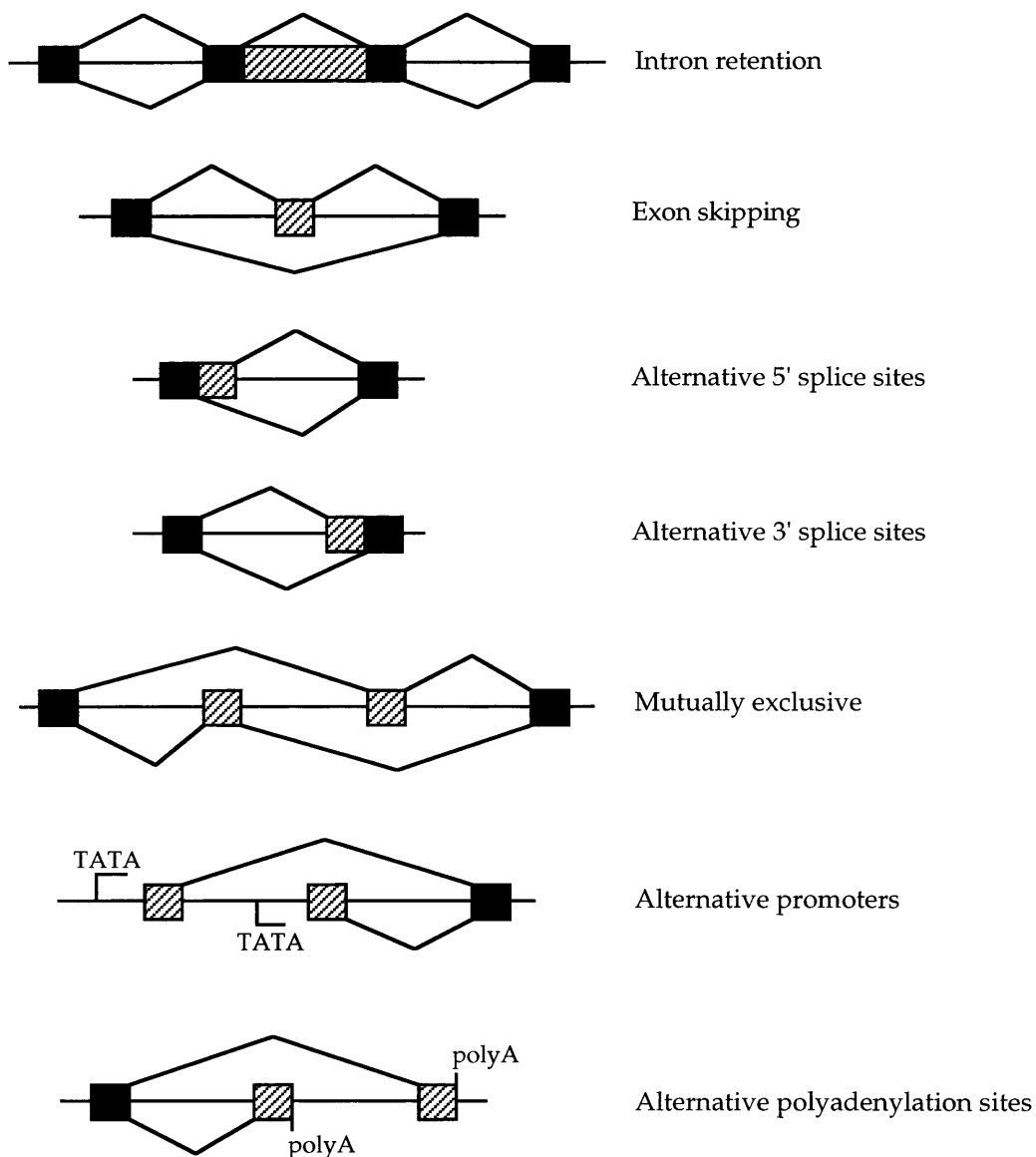
Figure 1.8 Modes of alternative splicing

Exons are represented as boxes, introns as horizontal lines. Splicing pathways are shown by the diagonal lines. Constitutive exons are shown in black, alternative exons are shaded.

Modes and mechanisms of alternative splicing

It can be predicted that a major step in the spliceosome assembly pathway which will be targetted in the regulation of alternative splicing is the formation of the E complex. It is at this stage that splice site pairing is first thought to occur. As discussed above, there are a number of protein-protein, protein-RNA, and RNA-RNA interactions that are required for the formation of this complex. There are potentially several ways in which these interactions could be prevented, reduced or enhanced in different cell types or at different developmental stages. These are likely to include changes in the concentration or activity of the constitutive factors comprising E complex (U1 snRNP, U2AF, SR proteins) or the use of specific factors in a particular cell type(s) or at a certain

1.18

developmental stage to either enhance or inhibit these interactions. The simplest mechanisms for regulation could involve either specific activation or specific inhibition of splice site pairing in a particular cell type or at a certain developmental stage. More complex mechanisms would involve specific activation of pairing in one cell type/developmental stage and specific inhibition in another, or vice versa.

A change in the actual concentration of U1 snRNP or U2AF would have effects on the expression of many genes unless this can be regulated so that there is a change in the local concentration of these factors. In the absence of local variations in concentration it is expected that the ability of these components to form protein-protein, protein-RNA, or RNA-RNA interactions will be disrupted in a specific way for specific pre-mRNAs. A change in the concentration of individual members of the SR protein family is likely to be involved in the regulated expression of at least some genes. SF2/ASF and SC35 have been shown to have similar, but distinct binding specificities (Tacke and Manley, 1995) and to have functional specificities for different pre-mRNAs in a splicing commitment assay (Fu, 1993). Therefore, pairing of splice sites in different pre-mRNAs is likely to involve different SR proteins. Thus, a change in the concentration of one SR protein is expected to have an effect on pairing of some, but not all splice sites (figure 1.9).
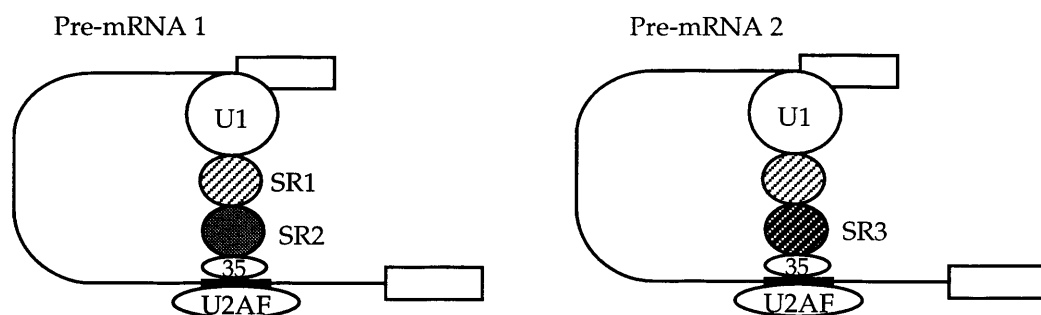
Figure 1.9 Specificity of SR protein function in bridging interactions

A change in the concentration of SR3 would disrupt splice site pairing of pre-mRNA 2, but would have no effect on pairing of pre-mRNA 1. The bridging interaction between U1 and U2AF is shown here to involve two SR proteins. The number of SR proteins involved in the bridging interaction has not been established.

Consistent with this idea, the level of individual SR proteins has been demonstrated to vary in different cell types (Zahler *et al*, 1993) and at different developmental stages (Screaton *et al*, 1995). Such a mechanism is predicted to be involved in the regulation of exon skipping or intron retention in a developmental stage- or tissue-specific manner. *In vivo*, changes in the level of individual SR proteins during T cell activation have been shown to correlate with changes in the expression patterns of CD44 and CD45 pre-mRNAs involving

skipping or inclusion of exons (Screaton *et al*, 1995). Further, expression of SF2/ASF in HeLa cells promoted inclusion of clathrin light chain B neuron-specific exon and prevented abnormal skipping ot two mutually exclusive β-tropomyosin exons (Caceres *et al*, 1994).

The activity of constitutive components of E complex could, in theory, be altered in a variety of ways. For example, alternatively spliced isoforms of some of the SR proteins have been reported (Zuo and Manley, 1993; Screaton *et al*, 1995). Two additional isoforms of SF2/ASF (ASF-2 and ASF-3) both lack the RS domain (Zuo and Manley, 1993). These isoforms could compete with SF2/ASF for binding to similar RNA sequences. The absence of an RS domain suggests that they would not be able make several of the protein-protein interactions made by the full length isoform and could, therefore, inhibit its function. Consistent with this, both isoforms were inactive in splicing (Zuo and Manley, 1993).

SR proteins are phosphorylated at residues in the RS domain. Splicing of human β-globin pre-mRNA was progressively inhibited by increasing amounts of a kinase specific for SR proteins *in vitro* (Gui *et al*, 1994). It was proposed that phosphorylation of these proteins had occurred and that their dephosphorylation is required for splicing. Phosphatases have been shown to be required for splicing but not for spliceosome assembly (Mermoud *et al*, 1992). Subsequent studies demonstrated that treatment of splicing extracts with a phosphatase prevented E complex assembly and that this inhibition could be relieved by the addition of SR proteins (Mermoud *et al*, 1994). Thus, phosphorylation of SR proteins may be required for spliceosome assembly. Dephosphorylation is then necessary before catalysis can occur.

Two recent studies have examined the effect of SR protein phosphorylation on their function *in vitro*. Binding of recombinant SF2/ASF to *in vitro* translated U1 70K protein was enhanced by *in vitro* phosphorylation of the RS domain of SF2/ASF. The relevance of this to splicing was suggested when the splicing efficiency of a pre-mRNA substrate was shown to be enhanced by phosphorylated, but not unphosphorylated SF2/ASF (Xiao and Manley, 1997). The sequence specificity of hSRp40 RNA binding is affected by its phosphorylation state. Sequences with high affinity to the unphosphorylated and *in vitro* phosphorylated protein were identified using an iterative selection procedure. No clear consensus sequence emerged for the unphosphorylated protein, however, a consensus motif was obtained for phosphorylated SRp40. It was proposed that phosphorylation of the RS domain of SRp40 was able to enhance the specificity of RNA binding by this protein (Tacke *et al*, 1997). Thus, protein-protein and protein-RNA interactions of SR proteins are influenced by their degree of phosphorylation. It is possible, therefore, that regulation of SR protein phosphorylation is used to control aternative splice site selection by influencing the RNA sequences or proteins that they are able to interact with.

A number of SR protein kinases have so far been identified (table 1.3). The exact role of these kinases *in vivo* and in alternative splicing is yet to be established.

Kinase	Comment	Reference
SRPK1	Phosphorylates several SR proteins and is specific for serines in the RS domain	Gui *et al*, 1994 a, b
Clk/Sty	Interacts with SF2/ASF in yeast two-hybrid assay and phosphorylates SF2/ASF almost exclusively in RS domian	Colwill *et al*, 1996
Not known	Associated with U1 snRNP	Woppmann *et al*, 1993
Topoisomerase I	Specific inhibitors change the phosphorylation state of SRs *in vivo*	Rossi *et al*, 1996

Table 1.3 SR protein kinases

A number of the more extensively studied examples of alternative splicing are now discussed to illustrate the variety of ways in which this process is regulated. As suggested above, these have all turned out to involve changes in the levels of constitutive splicing factors or the use of specific factors to enhance or inhibit the protein-protein, protein-RNA, or RNA-RNA interactions associated with the formation of the E complex.

Intron retention

Intron retention could potentially be regulated by a mechanism similar to that suggested in figure 1.9. However, there are other ways in which the interaction between U1 snRNP bound at the 5'ss and U2AF bound at the PPT could be disrupted. The most extensively studied example of intron retention is found in *Drosophila* and represents an example of inhibition of the interaction of U1 snRNP with the 5'ss sequence by a cell-type specific factor. *Drosophila* are particularly suited to analysis of splicing mechanisms because genetic screens and *in vivo* work can be combined with biochemical characterisation *in vitro*.

In the P transposable element of *Drosophila*, messenger RNA in which the intron has been spliced codes for a positively-acting protein. Intron retention results in the production of a protein that inhibits the function of the positively-acting protein. In this way alternative splicing acts as an on/off switch. P elements were discovered as the causative agents of a series of genetic traits in *Drosophila* known as hybrid dysgenesis. The symptoms of the disease are limited to the germline cells of the affected individuals and include sterility due to abnormal gonadic development, high rates of mutation, and chromosomal

rearrangements. These symptoms are brought about by the action of a mobile genetic element known as P transposable element. P element DNA contains four open reading frames which code for a transposase protein. This protein binds to inverted repeat sequences internal to larger terminal inverted repeats within the P transposable element and catalyses its transposition (reviewed in Rio, 1990, 1991). Transposition only occurs in germline tissue because functional transposase is not expressed in somatic cells.
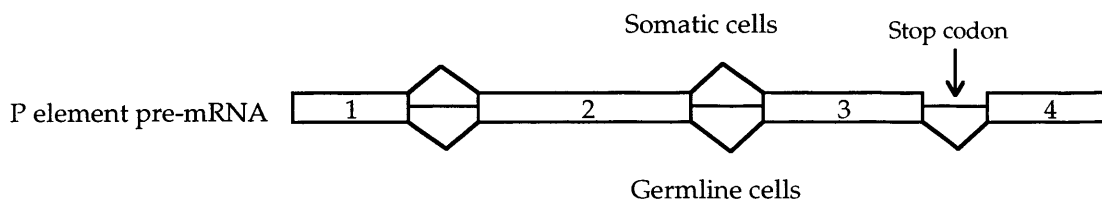
Figure 1.10 Structure of Drosophila P transposable element pre-mRNA

In germline cells, introns 1, 2, and 3 are removed and mature mRNA coding for the 87Kd functional transposase is produced. Somatic cell mRNA, however, retains the third intron (IVS3). A stop codon within this intron results in the production of a 66Kd protein that can function as a negative regulator of transposition in both germline and somatic cells (Robertson and Engels, 1989; Misra and Rio, 1990).

IVS3 splicing is inhibited in somatic cells (Siebel and Rio, 1990). Mutation of two sequences (F1 and F2) which closely resemble a consensus 5'ss sequence (pseudo 5'ss sequences) located within the exon upstream of IVS3 (figure 1.11) activated IVS3 splicing in splicing extracts made from *Drosophila* somatic cells (Siebel and Rio, 1990). Binding of U1 snRNP to the wild-type 5'ss sequence of exon 3 (the authentic 5'ss sequence) and to the F1 and F2 pseudo 5'ss sequences within exon 3 was examined in *Drosophila* somatic cell splicing extract (IVS3 is not spliced in these extracts) and in human somatic cell (HeLa cell) splicing extract (IVS3 is spliced in these extracts). In HeLa cell extracts, U1 snRNP bound to the authentic 5'ss and to the F1 site. However, in *Drosophila* cell extracts, binding was almost exclusively to the F1 site. It was proposed that the F1 pseudo 5' splice site sequence, although not used as a functional 5' splice site, competes with the wild-type 5'ss for binding to U1 snRNP and prevents the binding of U1 snRNP to the wild-type 5'ss in somatic cells.
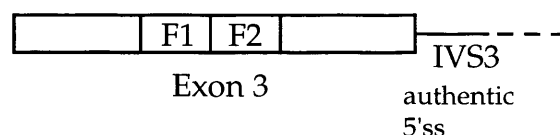
Figure 1.11 Arrangement of exon pseudo 5' splice sites (F1 and F2) upstream of IVS3.

Two proteins have been identified that bind specifically to exon 3 RNA: a 50Kd protein known as *Drosophila* hrp48 and a 97Kd protein now known as PSI. The role of PSI in inhibition of IVS3 splicing was confirmed when anti-PSI antibodies were demonstrated to specifically activate IVS3 splicing in *Drosophila* somatic cell extracts. Inhibition of IVS3 splicing was restored by the addition of recombinant PSI to the splicing extracts (Siebel *et al*, 1994; Siebel *et al*, 1995). Similar experiments to identify the role of hrp48 in inhibition of IVS3 splicing were not informative because an anti-hrp48 antibody inhibited general splicing. However, the protein was demonstrated to bind specifically to the F2 pseudo 5'ss sequence (Siebel *et al*, 1994).

PSI is specifically expressed in somatic cells and is localised to the nucleus. In contrast, hrp48 is expressed in both the soma and germline (Siebel *et al*, 1995). The role of PSI as a tissue-specific regulator of IVS3 splicing *in vivo* was confirmed when a maternally expressed promoter used to express a PSI transgene in germline cells during oogenesis was able to repress splicing of an IVS3-containing reporter RNA. Additionally, targetted disruption of PSI mRNA in somatic cells resulted in a reduction in PSI expression in these cells and activation of somatic splicing of an IVS3-containing reporter transgene (Adams *et al*, 1997).

It is proposed that hrp48 and PSI function together in the somatic inhibition of IVS3 splicing and that this interaction results in the misdirection of U1 snRNP to the F1 pseudo 5'ss sequence upstream of the wild-type 5'ss. Several details in the somatic inhibition of IVS3 splicing, however, remain to be determined. In particular, the involvement of hrp48 is not defined. The mechanism by which hrp48 and PSI are able to redirect the binding of U1 snRNP away from the wild-type site to the F1 site, and why other U1 snRNPs are not able to locate and bind the wild-type site, is not known. Also, the reason that the F1 site, after recognition by U1 snRNP, is not selected as a functional 5'ss is not understood. The control of PSI expression is an important event because tissue-specific transposase expression is ultimately controlled at this step.

Exon skipping

Role of specific factors in exon skipping

Regulated exon skipping of the c-src gene is thought to represent an example of positive stimulation of spliceosome assembly at the 5'ss in a cell-type specific manner. It is the most extensively studied example of exon skipping and has provided the first identification of a factor acting at a specific site in mammalian splicing regulation. The gene encodes a 60Kd tyrosine kinase. In some neurones an 18 nucleotide exon (N1) is inserted between the constitutive

exons 3 and 4 to produce a neural-specific protein n-src (figure 1.12). How insertion of the neural-specific exon affects the function of src is not known (Black, 1991).
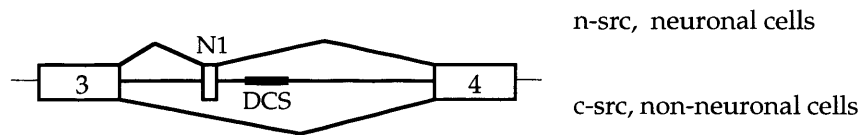
Figure 1.12 Regulated exon skipping in c-src

Mutation studies identified a ~30 nucleotide conserved region (the downstream control sequence, DCS in figure 1.12) located downstream of the N1 exon that is required for expression of that exon in neuronal cells (Black, 1992). Neuronal cell splicing extracts made from a retinoblastoma cell line (Weri-1 cells) were used to purify components of neuronal cells that bind specifically to the DCS. So far, two of these components have been identified and shown to be involved in activation of splicing to the 5'ss sequence of exon N1. One is a protein known as hnRNP F. The other was not previously known and has been cloned, sequenced and named KSRP (Min et al, 1997). KSRP has significant sequence identity with the Drosophila PSI protein. Although PSI is a repressor of splicing it is proposed that it is involved in recruiting U1 snRNP to the F1 pseudo 5'ss sequence of P element pre-mRNA (see above). Thus, KSRP may function similarly and recruit U1 snRNP to the 5'ss sequence of exon N1.

Both hnRNP F and KSRP are expressed in neuronal and non neuronal cells. The inactivity of the N1 exon in non neuronal cells cannot therefore be explained by absence of one of these factors from non neuronal cells. Other components of Weri-1 splicing extracts that are able to bind specifically to the DCS must be identified before activation of splicing to exon N1 in neuronal cells can be explained.

Role of constitutive factors in exon skipping

A mechanism by which expression of short internal exons in other pre-mRNAs could be regulated without the use of cell-type specific factors has also been proposed. It relies on a change in the concentration of constitutively expressed factors. In vitro, three-exon substrates derived from the β-globin gene in which the central exon is 33 nucleotides or less in length are predominantly spliced to give the skipped isoform (Dominski and Kole, 1991). Skipping is overcome if the length of the internal exon is increased or if the strength of the splicing signals is increased (Dominski and Kole, 1991). It was also discovered that skipping of the short internal exon was reduced in vitro and in vivo by the addition of SF2/ASF to the splicing reaction (Mayeda et al, 1992; Caceres et al, 1994). In contrast, exon skipping was promoted in vitro by addition of a protein

1.24

known as hnRNP A1 (Mayeda *et al*, 1992, see below). However, if the length of the short internal exon, or if the strength of the upstream PPT was increased, hnRNP A1 no longer had an effect. It was proposed that exons have evolved to be of a particular length in combination with particular splicing signals in order to reduce the likelihood of skipping occurring in the case of constitutive exons, and to allow regulation in the case of alternative exons. This regulation could be achieved by changes in the relative levels of constitutively expressed factors such as SF2/ASF and hnRNP A1.

It has been suggested that the inherent weakness of short internal exons compared with longer internal exons arises because the interactions across exons involved in exon definition are less efficient across short exons (Dominski and Kole, 1992). Presumably, this inefficiency is overcome by adjusting the strength of the splicing signals or by increasing the concentration of the factors involved, for example, SF2/ASF.

Alternative 5'ss use

Several genes give rise to more than one mature mRNA through alternative 5'ss use. Some of the more extensively studied examples are shown in figure 1.13.

1) SV40 pre-mRNA
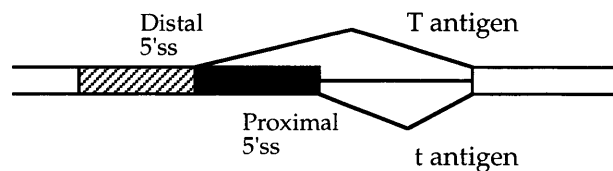
2) Human β-thalassaemia pre-mRNA
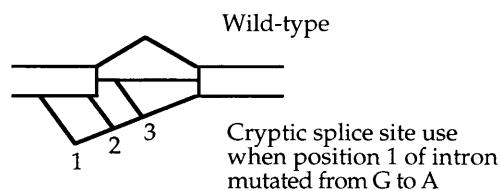
3) Adenovirus E1a pre-mRNA
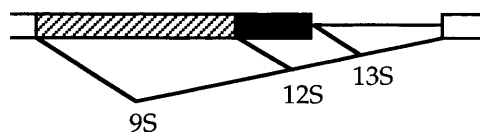
Figure 1.13 Examples of alternative 5'ss use
See text for discussion

1.25

It was noted that the ratio of small t antigen to large T antigen mRNAs produced from SV40 pre-mRNA in 293 cells was over 10-fold higher than in many other mammalian cells (Fu and Manley, 1987). This was used as an assay to purify ASF from 293 extracts. Highly purified ASF was able to shift 5'ss use away from the large T site and towards the small t site when added to HeLa cell splicing extracts (Ge and Manley, 1990). At around the same time SF2, which had been shown to complement S100 extracts for splicing activity (Krainer et al, 1990), was demonstrated to affect cryptic 5'ss use in mRNA produced from a β-thalassaemic allele of human β-globin. The mutant allele contains a G to A mutation at position 1 of the intron which results in the activation of 3 cryptic 5' splice sites. In vivo and in HeLa splicing extracts, cryptic site 2 use is greater than use of sites 1 and 3. Titration of SF2 into HeLa extracts resulted in a shift towards use of site 3 (Krainer et al, 1990). Subsequent cloning of ASF and SF2 revealed that the two proteins were identical (Krainer et al, 1991; Ge et al, 1991). The importance of these in vitro observations on alternative splice site selection in vivo was confirmed by effects of ASF/SF2 on reporter gene expression in HeLa cells. Transient overexpression of ASF/SF2 promoted selection of proximal 5' splice sites in β-thalassemia and adenovirus E1A pre-mRNAs (Caceres et al, 1994).

Other members of the SR protein family have also been shown to affect alternative 5' splice site use. Their reported effects, in vivo and in vitro, on the use of alternative 5' splice sites in SV40, adenovirus E1a and human β-thalassaemia pre-mRNAs are given in table 1.4. Where effects have been tested in vivo and in vitro the results are generally in close agreement, except for the effects of SRp30a and SRp30b on SV40 expression. The reason for these differences are not clear but emphasise that the in vivo and in vitro systems are not identical. In all situations where SRp30a (ASF/SF2) and SRp30b (SC35) have a positive influence on splice site selection the proximal site is stimulated. In contrast, the other SR proteins have all been shown to be capable of activating sites other than the proximal site. Some effects are substrate-specific. For example, SRp20 stimulated use of the 12S site in E1a pre-mRNA but the proximal site in β-thalassaemia pre-mRNA. In several cases, the promotion of splicing to one site is accompanied by a reduction in use of another site. This suggests that SR proteins act by adjusting the relative use of sites rather than by activation of use of a site in a previously unreactive transcript.

Proximal 5'ss activation by ASF/SF2 in HeLa cell nuclear cell extracts required much more SF2 than was needed to reach the same level of activation in S100 extracts (Krainer et al, 1990). It was proposed that a factor(s) present in nuclear extract and absent in S100 extract may have been able to inhibit the effect of ASF/SF2 on splice site selection. This was used as an assay to purify the heterogeneous nuclear ribonucleoprotein hnRNP A1 from HeLa cell nuclear

1.26

Table 1.4 Effect of SR proteins on 5' splice site selection

Down and up refer to the effect of an individual SR protein on the relative use of alternative 5' splice sites. Results of *in vivo* studies are shown in bold

Pre-mRNA	E1a			SV40		ß-thalassaemia			Cells/Extract	References
5'ss	9S	12S	13S	T	t	1	2	3		
SRp20	down	up	down						HeLa cells	Screaton et al, 1995
						-	down	up	HeLa cells	Screaton et al, 1995
SRp30a	down	down	up						HeLa cells	Caceres et al, 1994
	down	down	up						HeLa cells	Screaton et al, 1995
	same	down	up						HeLa cells	Wang and Manley, 1995
	down	down	up						HeLa extract	Harper and Manley, 1992
				down	down				293 cells	Wang and Manley, 1995
				down	up				HeLa extract	Ge and Manley, 1990
						down	down	up	HeLa cells	Caceres et al, 1994
						-	down	up	HeLa cells	Screaton et al, 1995
						down	same	up	HeLa extract	Krainer et al, 1990
SRp30b	down	down	same						HeLa cells	Wang and Manley, 1995
	-	down	up						S100	Zahler et al, 1993
				down	down				293 cells	Wang and Manley, 1995
				same	up				S100	Zahler et al, 1993
SRp30c	down	down	up						HeLa cells	Screaton et al, 1995
						up	same	up	HeLa cells	Screaton et al, 1995
SRp40	down	down	up						HeLa cells	Screaton et al, 1995
	-	down	up						S100	Zahler et al, 1993
				up	same				S100	Zahler et al, 1993
						same	same	same	HeLa cells	Screaton et al, 1995
SRp54	up	same	down						HeLa cells	Zhang and Wu, 1996
SRp55	down	same	up						HeLa cells	Screaton et al, 1995
	-	same	up						S100	Zahler et al, 1993
				up	same				S100	Zahler et al, 1993
						same	same	same	HeLa cells	Screaton et al, 1995
SRp70	-	down	up						S100	Zahler et al, 1993
				up	same				S100	Zahler et al, 1993

extract (Mayeda and Krainer 1992). In substrate pre-mRNAs containing either duplicated 5'ss or duplicated 5' and 3'ss, and in adenovirus E1a pre-mRNA, addition of hnRNP A1 promoted distal 5'ss use *in vitro* (Mayeda and Krainer 1992) and, with E1a pre-mRNA, *in vivo* (Caceres *et al*, 1994). These observations led to the proposal that the relative levels of SR proteins and hnRNP A1 could determine alternative splice site choice and that tissue-specific or developmental stage-specific regulation of the level or activity of these proteins was a way in which the regulated expression of genes could be achieved (Zahler *et al*, 1993; Caceres *et al*, 1994). Consistent with these proposals, the relative levels of these proteins has been shown to vary between tissue types (Zahler *et al*, 1993).

The mechanisms by which SR proteins and hnRNP A1 are able to switch alternative 5'ss selection are largely unknown although some models have been proposed. It was first suggested that the selection of a particular 5'ss could be explained by the affinity of U1 snRNP for the alternative sequences. The rate of dissociation of U1 snRNP from the stronger site would be slower than the rate of dissociation from the weaker site(s). The stronger site would therefore be bound for a longer period of time by U1 snRNP than the weaker site(s) and the probability of interaction of with components of the splicing machinery would be higher (Eperon *et al*, 1986). However, it has been shown that the affinity of U1 snRNP for different 5'ss sequences does not correlate with their use as splice sites (Chabot and Steitz, 1987; Nelson and Green, 1988). The identification of SR proteins as factors able to switch 5'ss choice (Ge and Manley, 1990; Krainer et al, 1990) and to recruit U1 snRNP to the 5'ss (Eperon *et al*, 1993; Kohtz *et al*, 1994; Jamison *et al*, 1995; Zahler and Roth, 1995) has allowed new models for alternative 5'ss selection to be tested.

It has been demonstrated that two alternative consensus 5'ss sequences were bound indiscriminately by U1 snRNP and that only the downstream site was used (Eperon *et al*, 1993). It was proposed that the downstream site was chosen because all the RNA molecules containing the two alternative sites were doubly occupied by U1 snRNP and that the probability of interaction of splicing components bound at a single 3'ss with U1 snRNP is higher for U1 snRNP bound to a proximal 5'ss than for U1 snRNP bound to a more distal one. It was predicted that weakening the interaction of U1 snRNP with the 5'ss sequences, for example by reducing the concentration of U1 snRNP, would result in use of the distal site because in some of these molecules the distal site only would be occupied by U1 snRNP. An experiment to test this prediction showed that it was correct. It was suggested that factors such as ASF/SF2 and SC35, which activate proximal 5'ss selection (table 1.4), could exert their effect by strengthening the interaction of U1 snRNP with 5' splice site sequences, so that the proportion of pre-mRNA molecules occupied at all competing sites is increased. Similarly, factors which activate distal 5'ss selection (hnRNP A1 and members of the SR protein family,

table 1.4) do so by weakening the interaction of U1 snRNP with 5'ss sequences. The distal site will then be used provided that the proximal site is not fully occupied by U1 snRNP. In support of the model, ASF/SF2 was shown to increase the occupancy of two competing 5'ss sequences by U1 snRNP. It was concluded that ASF/SF2 is able to act non-specifically at 5'ss sequences to stabilise the interaction of U1 snRNP with these sequences.

In another study, the effect of SRp40 and SRp30b on the interaction of U1 snRNP with two alternative 5'ss sequences in SV40 pre-mRNA was examined by crosslinking (Zahler and Roth, 1995). SRp40 stimulates use of the distal site, whereas proximal site use is stimulated by SRp30b. Consistent with the proximity effect proposed by Eperon et al (1993), U1 snRNP was observed to crosslink only to the distal site in the presence of SRp40, but to both sites in the presence of SRp30b. Further evidence consistent with the model of Eperon et al (1993) is that the position, rather than the sequence, of alternative 5' splice sites has been shown to be important in determining their relative use. ASF/SF2 promotes 13S and inhibits 12S splicing in wild-type E1a pre-mRNA (figure 1.13 and table 1.4). However, when the 13S site was deleted, splicing to the 12S site was promoted (Wang and Manley, 1995). Similarly, hnRNP A1 was able to activate or inhibit use of the human β-globin exon 1 5'ss depending on whether it was located distally (activated) or proximally (inhibited) with respect to other sites (Mayeda and Krainer, 1992).

As discussed above, the RS domains of SF2/ASF and the U1 70K protein are known to interact (Wu and Maniatis, 1993) and the U1 70K protein has been shown to be required for the formation of a stable ternary complex comprising U1 snRNP, SF2/ASF and pre-mRNA providing that the pre-mRNA was pre-incubated with SF2/ASF before the addition of U1 snRNP (Kohtz et al, 1994). Thus, it might be expected that the action of SF2/ASF in alternative 5'ss selection, in which the binding of U1 snRNP to 5'ss sequences is stimulated (Eperon et al, 1993), would involve interaction between the pre-mRNA substrate and the RRM and/or RRMH domains of SF2/ASF and between the RS domains of SF2/ASF and U1 70K. The RRM and RRMH domains of SF2/ASF are both required for the alternative splicing activity of SF2/ASF, but, unexpectedly, the RS domain is not required (Zuo and Manley, 1993; Caceres and Krainer, 1993). Thus, if any protein-protein interactions are involved they must occur away from the RS region of SF2/ASF.

Alternative models for 5'ss selection involve U6 snRNP. U6 snRNA is known to form contacts with the 5'ss sequence in the spliceosome after it has been recognised by U1 snRNA (see above). It has been shown that splicing can occur *in vitro* in the absence of U1 snRNP if the concentration of SR proteins is increased (Crispino et al, 1994; Tarn and Steitz, 1994) and that mutations in the vicinity of the 5'ss sequence that increase its complementarity with U6 snRNA

enhance the efficiency of U1-independent splicing (Crispino and Sharp, 1995). It is possible that SR proteins could direct the choice of 5'ss sequence by altering the binding of U6 snRNA to this sequence and that this involves interaction with regions of SF2/ASF or other SR proteins away from the RS domain. The direct involvement of SR proteins in recognition of 5'ss sequences by U6 snRNP has yet to be demonstrated.

The exact mechanism(s) by which certain SR proteins and hnRNP A1 are able to promote distal 5'ss selection is also not known. It was suggested that hnRNP A1 could weaken the binding of U1 snRNP to 5'ss sequences (Eperon et al, 1993). HnRNP A1 contains two RNA-binding domains and a C-terminal glycine-rich domain. Mutation studies have shown that amino acid residues within the RNA binding domains of hnRNP A1 are required for the activity of this protein in alternative splicing (Mayeda et al, 1994). Burd and Dreyfuss (1994) used an iterative selection procedure to determine the sequence of RNAs which have high affinity for hnRNP A1. RNA containing two copies of the motif UAGGGU/A was identified. This sequence has similarity to both the 5'ss and 3'ss consensus sequences. Thus, hnRNP A1 could directly block binding of U1 snRNP to 5'ss sequences. HnRNP A1 is known to possess RNA and DNA annealing activity (Kumar and Wilson, 1990; Pomtius and Berg, 1990; Munroe and Dong, 1992). Interaction of U1 snRNP with 5'ss sequences could be affected by this activity.

Like SF2/ASF, the activity of hnRNP A1 may be regulated by the presence of alternatively spliced isoforms. A number of alternatively spliced isoforms of hnRNP A1 have been reported (reviewed in Dreyfuss et al, 1993). HnRNP A1(B) contains a longer glycine-rich domain than hnRNP A1 and binds more strongly to pre-mRNA. Its alternative splicing activity is limited. In contrast, hnRNP A2 and B1 bind more weakly to pre- mRNA and have stronger splice site switching activities than hnRNP A1 (Mayeda et al, 1994). It seems likely that alternative splicing could be regulated by alterations in the relative levels of hnRNP A1, SR proteins and alternative isoforms of these proteins.

The sequence binding specificities of the SR protein family (other than SF2/ASF, SC35 and SRp40), and the requirement of domains within these proteins for their activities in constitutive and regulated splicing are yet to be investigated fully. Such information should help to design models for their mechanism(s) of action in alternative splice site selection. Some may act in a similar way to that proposed for the action of SF2/ASF in alternative 5'ss selection (Eperon et al, 1993). However, for those that are able to promote proximal and distal splice site use, depending on the substrate (table 1.4), such a model, in which these proteins would either strengthen or weaken the interaction of U1 snRNP with all competing 5'ss sequences in a substrate-specific manner, seems unlikely. It is possible that strong sequence preferences in

particular pre-mRNAs are able to override an effect on all competing 5'ss sequences.

SR proteins and redundancy of function

The early experiments involving complementation of S100 extracts and alternative splice site selection described above suggested that individual SR proteins might be functionally redundant (Ge and Manley, 1990; Krainer *et al*, 1990; Fu *et al*, 1992; Zahler *et al*, 1992). However, it is clear that there are differences in the splice site switching activity of these proteins (table 1.4) and that SF2/ASF and SC35 have different RNA binding specificities *in vitro* (Tacke and Manley, 1995). It has also been shown that a β-globin pre-mRNA was committed to splicing by SC35 and not by SF2/ASF or other SR proteins. In contrast, HIV tat pre-mRNA was committed by SF2/ASF and not by SC35. The non-redundancy of SR protein function was recently confirmed when it was demonstrated that SF2/ASF is an essential gene (Wang *et al*, 1996). Clones of a chicken B cell line in which both alleles of the SF2/ASF gene were disrupted could only be obtained when SF2/ASF was expressed from an exogenous plasmid. The SF2/ASF expressed from the exogenous plasmid was then placed under the control of a tetracycline-repressible promoter system. 96 hours after addition of tetracycline to the cell culture medium, the cells were nearly all dead.

Alternative 3'ss use

The first event in definition of the 3'ss is binding of U2AF to the polypyrimidine tract (PPT). Thus, regulation of 3'ss use is expected to involve factors that can either inhibit or stimulate the interaction of U2AF with the PPT. After it was reported that SF2/ASF was able to promote proximal 5'ss use (Ge *et al*, 1991; Krainer *et al*, 1991) and that this effect could be counteracted by hnRNP A1 (Mayeda and Krainer 1992; Caceres *et al*, 1994) the effect of these proteins on alternative 3'ss selection was examined. Addition of increasing amounts of SC35 or SF2/ASF to splicing extract was observed to shift the splicing of pre-mRNA containing duplicated 3'ss sequences to use of the proximal site. This effect was not counteracted by hnRNP A1 (Fu *et al*, 1992). Again, the mechanism by which SR proteins are able to achieve this is unknown.

Alternative 3'ss use has been extensively studied in *Drosophila*. An example of both regulated inhibition and stimulation of the U2AF-PPT interaction is found in the sex determination pathway of *Drosophila*. This involves Sxl, Tra and Dsx pre-mRNAs in a regulated cascade of splicing events (figure 1.14) (Baker 1989; Hodgkin 1989) and is initiated by female-specific transcription from the early promoter of the Sxl gene. Transcription from this promoter begins approximately two hours after egg laying and lasts for a further

Figure 1.14 The regulatory cascade involved in Drosophila sex determination

The alternatively spliced regions of the Sxl, Tra, and Dsx pre-mRNAs are shown. The splicing pattern in male and female flies is represented by the angled lines. See text for further discussion.

Male, no Sxl
protein

Sxl pre-mRNA [2   3   4]

Female, Sxl protein
Sxl and Snf proteins required.

Male, no Tra
protein

Tra pre-mRNA [1   2   3   4]

Female, Tra protein.
Sxl protein required.

Female, Dsx protein.
Tra, Tra2 proteins required.
Dsx protein blocks
male differentiation.

Dsx pre-mRNA [3   4   5]

Male, Dsx protein blocks
female differentiation
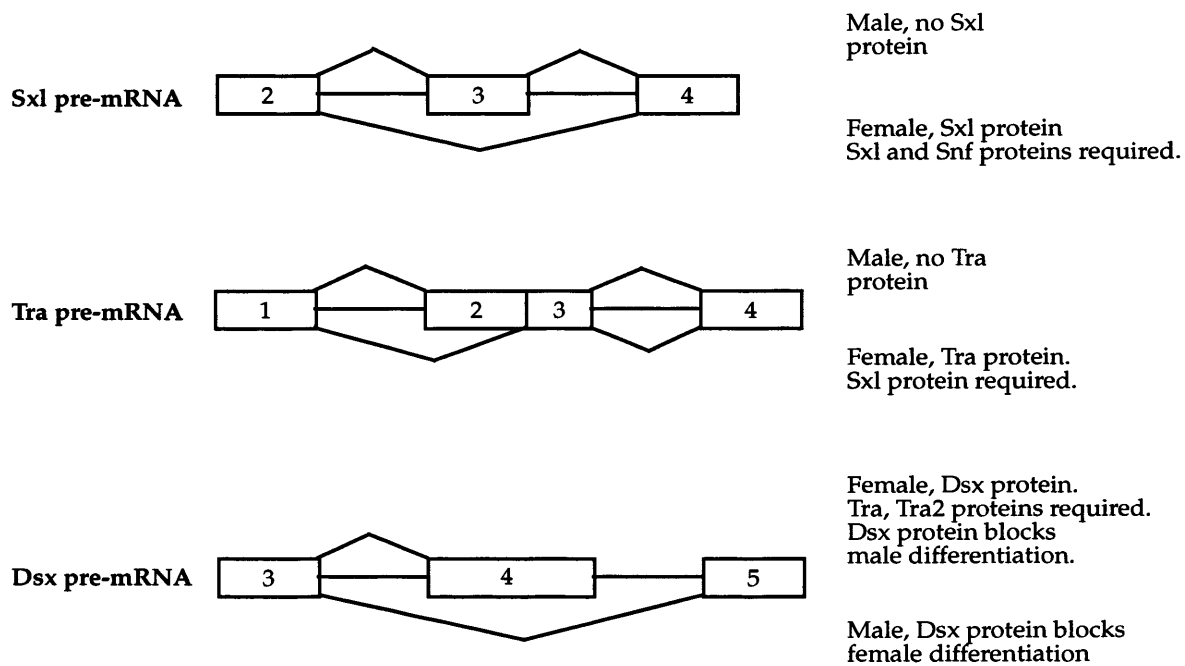
two hours. Three hours after egg laying, transcription from the Sxl late promoter is initiated in both sexes (reviewed in Burtis 1993). Pre-mRNA expressed from the late promoter is alternatively spliced in a sex-specific manner. In males, splicing to exon 3 produces Sxl mRNA which contains two in-frame stop codons. Translation of this mRNA results in production of a truncated protein which is non-functional. In females, exon 3 is skipped and the mRNA produced codes for functional Sxl protein. The regulation of exon 3 skipping in females is dependent on the Sxl and Snf gene products (Bell et al, 1991; Oliver et al, 1993; Bopp et al, 1993), but the mechanism by which this is achieved is unclear (Deshpande et al, 1996).

Inhibition of the U2AF/polypyrimide tract interaction

The functional Sxl protein produced in females is also used to regulate sex-specific alternative splicing of tra pre-mRNA. In males flies, use of the tra exon 2 3'ss leads to the production of mRNAs containing in frame stop codons (Boggs et al, 1987; Butler et al, 1986). In females, the Sxl protein promotes use of the tra exon 3 3'ss which is downstream of the stop codons. The Tra protein produced then directs female-specific splicing of dsx pre-mRNAs (Burtis and Baker, 1989). The Sxl protein promotes female-specific tra pre-mRNA splicing by binding to the polypyrimide tract upstream of the male-specific exon 2 and blocking access to that sequence by U2AF. In male flies the Sxl protein is not expressed. U2AF has higher affinity for the polypyrimide tract which forms part of the male-specific 3'ss sequence than for the polypyrimide tract upstream of exon 3, so splicing to exon 2 is observed (Valcarcel et al, 1993).

Another factor thought to inhibit binding of U2AF to certain mammalian PPT sequences is polypyrimidine tract binding protein, or PTB (Valcarcel et al, 1993; Lin and Patton, 1995). The proposed action of this protein in 3'ss selection is discussed in the section on tropomyosin genes below.

Stimulation of the U2AF/polypyrimide tract interaction

The final step in the Drosophila sex determination regulatory cascade is alternative splicing of dsx pre-mRNA (figure 1.14). The 3'ss sequence upstream of exon 4 is weak (Hoshijima et al, 1991) and is not recognised in males. The mRNA produced contains exons 1, 2, 3, 5, and 6 and codes for a protein which blocks female sexual differentiation. In females, the weak 3'ss sequence is recognised and a protein that represses expression of male-specific genes is produced (Baker and Ridge 1980; Nothiger et al, 1987). Female-specific splicing of dsx pre-mRNA requires the products of the tra and tra-2 genes, the Tra and Tra2 proteins, respectively (Nagoshi et al, 1988; McKeown et al, 1988). Male flies produce functional Tra2 but not Tra protein. The sex-specific splicing of dsx pre-mRNA

can therefore be explained by the presence of Tra2 in female flies (Nagoshi *et al*, 1988; Amrein *et al*, 1990; Mattox *et al*, 1990).

Tra and Tra2 are required for activation of the female-specific 3'ss (Nagoshi and Baker, 1990; Hedley and Maniatis, 1991; Ryner and Baker, 1991; Tian and Maniatis 1992). Regulation of female-specific splicing by these proteins is dependent on a sequence located within exon 4, the dsxRE (Ryner and Baker, 1991; Tian and Maniatis, 1992). This sequence contains two distinct elements (figure 1.15). One comprises six repeats of a 13 nucleotide sequence (Burtis and Baker, 1989) and the other is an 18 nucleotide purine-rich sequence known as the purine-rich enhancer, or PRE (Lynch and Maniatis, 1995). Repeats 1-5 together with the PRE are required for efficient Tra and Tra2-dependent activation of dsx female-specific splicing (Lynch and Maniatis, 1995).

1-6=13 nucleotide repeat sequences
PRE= purine-rich element
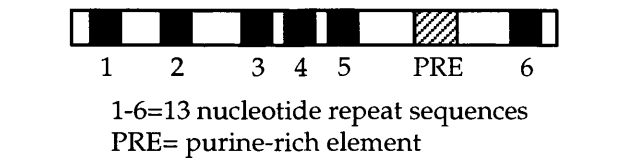
Figure 1.15 Sequence motifs in the dsxRE

Tra, Tra2, and SR proteins are sufficient to commit dsx pre-mRNA containing the dsxRE to splicing (Tian and Maniatis, 1993). These proteins bind cooperatively to the dsxRE and with high specificity (Tian and Maniatis 1993, 1994; Lynch and Maniatis 1995). Both the repeat sequences and the PRE are required for high specificity binding (Lynch and Maniatis 1995). The exact binding sites of Tra, Tra2 and SR proteins within the dsxRE were identified by Lynch and Maniatis (1996). Their results also demonstrated that specific SR proteins are involved. A schematic representation of the results is shown in figure 1.16.

Figure 1.16 Interaction of Tra, Tra-2 and SR proteins with the *Drosophila* dsxRE in *Drosophila* and human cell extracts.

9G8 and SF2/ASF are human SR proteins. RBP1 and dSRp30 are *Drosophila* SR proteins

Thus, the cooperative interaction of Tra and Tra2 with a specific SR protein allows these proteins to bind with high affinity to completely different RNA sequences (the repeats are rich in U and C nucleotides, and the PRE is rich in purines). Individually, none of the proteins have high affinity for the repeat sequences (Tian and Maniatis 1992; Lynch and Maniatis 1996) and only Tra2 alone binds with a high degree of specificity to the PRE (Lynch and Maniatis 1995). This suggests that protein-protein interactions between Tra, Tra2 and the SR proteins determine the exact sequences recognised by them.

The protein-protein interactions of Tra, Tra2, and SR proteins have been examined in a number of experiments. Tra and Tra2 were shown to interact with themselves, with each other, and with SR proteins (Wu and Maniatis 1993; Amrein *et al* 1994). These interactions are likely to be mediated by the RS domains of these proteins. The RS domain of SC35 was shown to be required for its interaction with Tra and Tra2 (Wu and Maniatis 1993) and one of the RS domains in Tra2 is required for interaction with itself and with SF2 (Amrein *et al* 1994). Tra does not contain any recognised RNA binding domains and was previously shown to require components of nuclear extract for association with dsx RNA (Tian and Maniatis 1992). The presence of this protein in the complex appears to be due solely to interaction with the Tra2 and SR proteins.

Experiments using the yeast two-hybrid system showed that the 35Kd subunit of U2AF could simultaneously interact with the 65Kd subunit of U2AF and either SF2/ASF or SC35. It was proposed therefore, that the cooperative binding of Tra, Tra2, and SR proteins functions to recruit SR proteins to the dsxRE. SR proteins are then able to stimulate the binding of U2AF to the polypyrimidine tract upstream of dsx exon 4 via interactions with the 35Kd subunit of U2AF and enhance E complex formation immediately upstream of the female-specific exon (Wu and Maniatis 1993). In this way, the dsxRE replaces the role of the 5'ss sequence in exon definition.

The involvement of U2AF in dsx alternative pre-mRNA splicing was demonstrated when splicing extracts supplemented with Tra and Tra2 but depleted of U2AF were shown to be unable to splice dsx pre-mRNA. Addition of recombinant U2AF completely restored splicing (Zuo and Maniatis 1996). Assembly of E complex on the same substrate pre-mRNA was also examined in this study. Efficient E complex assembly only occurred if Tra, Tra2 and SR proteins were added to the splicing extracts. Both subunits of U2AF were present in these E complexes. Association of U2AF was dependent on Tra and Tra2 and was greatly enhanced in the presence of SR proteins. Purified Tra, Tra2, SC35, U2AF^{65} and U2AF^{35} components were used to demonstrate the minimal requirements for stimulation of U2AF binding to the polypyrimidine tract upstream of dsx exon 4. Association of U2AF^{65} was observed only when all of these components were included in the binding reaction (Zuo and Maniatis 1996).

The exact nature of the interactions between U2AF and components bound to the dsxRE was also investigated by Zuo and Maniatis (1996). The RS domain of U2AF^{35} which is required for interactions with Tra, Tra2 and SR proteins was also shown to be required for dsx enhancer-dependent splicing as was the region of U2AF^{65} involved in interaction with U2AF^{35}.

A model for the mechanism of action of the dsxRE in activation of female-specific dsx splicing shown in figure 1.17. The recruitment of specific SR proteins to the dsxRE by Tra and Tra2 allows the formation of 7 tandemly arranged stable and specific heterotrimeric complexes on the dsxRE. These complexes interact with the small subunit of U2AF and stimulate and stabilise the interaction of U2AF with the PPT. All the repeat sequences are required for efficient Tra/Tra2-dependent splicing but the affinity of the heterotrimeric complex for single repeats is about the same as for the entire dsxRE (Lynch and Maniatis 1996), so cooperative interactions between heterotrimeric complexes bound at different repeat sequences are not thought to occur. The distance between the female specific 3'ss and the dsxRE is large (300 nucleotide), so several repeats may be required in order to increase the probability that components assembled on them will interact with U2AF. Stable association of U2AF with the PPT upstream of the exon 4 3'ss is then predicted to result in an increased probability that E complex formation will involve the female-specific 3'ss.

Figure 1.17 Model for the action of the dsxRE in female-specific dsx pre-mRNA splicing.

It is not known what causes specific SR proteins to be recruited to the heterotrimeric complexes, but it seems likely that differences between these proteins determine their ability to form protein-protein interactions with Tra and Tra2 and to bind with high specificity to the sequences within the dsxRE.

Exon splicing enhancers

A growing number of exon sequences have been shown to be required for splicing to the 3'ss of those exons. These sequences are known collectively as exon splicing enhancers (ESEs). A number of examples are given in table 1.5. Most of the sequences identified are purine-rich and contain repeats of a consensus GAR motif (where R is purine). Tanaka *et al* (1994) examined the ability of 24 nucleotide purine sequences to enhance splicing to an upstream 3'ss. They found that alternating purine sequences were effective ESEs, but that poly(A) or poly(G) sequences did not stimulate splicing. In addition, interruptions of purine sequence by pyrimidine residues greatly reduced the ESE activity. When the orientation of a number of ESEs was reversed their activity was disrupted (Watakabe *et al*, 1993). In some studies, increasing the length of the purine-rich region or including repetitions of the enhancing sequence has been shown to result in an increase in splicing activity (Tanaka *et al*, 1993; Dirksen *et al*, 1994). Exon splicing enhancers must be active within the confines of the coding sequence. It is perhaps not surprising, therefore, that the number of repetitions of the GAR motif varies widely between different ESEs and that they can function as discrete repetitions of small sequences.

ESEs from a number of pre-mRNAs, when placed within a different exon environment, have been shown to stimulate splicing to that exon (Lavigueur *et al*, 1993; Yeakley *et al*, 1993; Xu *et al*, 1993; Tanaka *et al*, 1994). Their activity, however, is dependent on their position within the exon to some extent. A 9 nucleotide ESE within the human fibronectin ED1 exon stimulated splicing to the 3'ss of that exon when it was placed in a variety of positions within the exon, but when the distance between the ESE and the 3'ss of the ED1 exon was increased to over 294 nucleotides, stimulation was lost (Lavigueur *et al*, 1993).

A number of studies have shown that the activity of ESEs is mediated by *trans*-acting factors (Sun *et al*, 1993; Lavigueur *et al*, 1993; Tanaka *et al*, 1994). In some of these studies SR proteins have been demonstrated to bind specifically to the ESE sequence (table 1.5). The functional significance of the interaction with SR proteins has only been demonstrated in a few cases. SF2/ASF (SRp30a) was shown to bind specifically to a 115 nucleotide ESE in the terminal exon of bovine growth hormone pre-mRNA and then to stimulate splicing to that exon *in vitro* only if the ESE was present (Sun *et al*, 1993). In another study, SRp40 and SRp55 bound specifically to an ESE in exon 5 of cardiac troponin T pre-mRNA and were also able to complement S100 extract (cell extract that is inactive for splicing in the absence of SR proteins) for splicing to that exon. Another SR protein (SRp30b) which did not bind specifically to the ESE could not complement S100 extract for splicing to that exon (Ramchatesingh *et al*, 1995).

Purified SR proteins did not bind to a synthetic ESE containing a repetitive GAA sequence. Instead, it was recognised by a complex including SRp40 and a

Table 1.5 Exon enhancer sequences

Exon	Sequence	Binding proteins	Reference
Bovine growth hormone exon 5	GGAAG and other elements within 115n.t. sequence	SRp30a	Sun et al, 1993; Dirksen et al 1994
Human fibronectin ED1 exon	GAAGAAGA	SRs	Lavigueur et al 1993
Chicken cardiac troponin T exon 5	GAGGAAGAA	SRp30a, SRp40, SRp55, SRp75	Ramchatesingh et al 1995
HIV-1 terminal tat-rev exon	GAAGAAGAAG		Staffa and Cochrane 1995
Rat CT/CGRP exon 4*	GAAGAAGAAGCTC		Yeakley et al 1993
Chicken cardiac troponin T exon 16	GGCCACCAGAAGGUAUG	SRp30b	Wang et al 1995
Equine infectious anemia virus exon 3	GAAGAAAAAGAAGAAAUGAC	SRp20, SRp30a/b, SRp40, SRp55	Gontarek and Derse 1996
Mouse IgM exon M2	GGAAGGACAGCAGAGACCAAGAG	U1 snRNP	Watakabe et al 1993
Human caldesmon exon 5	GAGGAAGAGAAAGGGCAGCAGAGGAGAGGCA		Humphreys et al 1995
ASLV env 3' exon	CTCCAAGAAGAAGCCGCCAGCAACAAGCAAGAAGGA	SRp20, SRp30, SRp40	Fu et al 1991; Staknis and Reed 1994
Synthetic	GAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA	SRp40, p37	Yeakley et al, 1996

37Kd GAA-binding protein. It was proposed that the 37Kd protein specifically recruits SRp40 to the RNA (Yeakley *et al*, 1996). In another study U1 snRNP was shown to bind specifically to an ESE located in exon M2 of the mouse IgM pre-mRNA (Watakabe *et al*, 1993). However, the functional significance of this binding was not determined.

The observation that ESEs are bound specifically by SR proteins suggested that these elements may stimulate splicing of upstream introns in the same way that the dsxRE in *Drosophila* dsx pre-mRNA is able to enhance female-specific dsx pre-mRNA splicing, except that there is no requirement for Tra and Tra-2 because SR proteins bind directly to the ESE. Thus, once bound to the ESE, SR proteins are able to interact with U2AF and enhance its interaction with the polypyrimidine tract immediately upstream. The efficiency of E complex formation is thereby increased. The evidence for this model is not yet as compelling as that for the involvement of the dsxRE in regulated splicing of *Drosophila* dsx pre-mRNA. However, several studies have provided evidence consistent with such a mechanism.

Lavigueur *et al* (1993) demonstrated that binding of U2 snRNP to the branch point sequence is enhanced by the presence of the ED1 ESE (table 1.5) in the exon immediately downstream. A possible involvement of SR proteins in the activity of the ED1 ESE was demonstrated when SR proteins were shown to be present in a complex formed specifically on RNA containing the ED1 ESE. This study did not however, provide any evidence for a direct involvement of U2AF in stimulation of splicing by the ED1 ESE, although U2AF is known to stimulate binding of U2 snRNP to the branch point sequence. Evidence that ESEs stimulate E complex formation was first obtained by analysis of complexes assembled in splicing extract on RNA containing an ESE (Staknis and Reed 1994). RNA containing the bovine growth hormone ESE or the ASLV ESE (table 1.5) assembled complexes that resembled E complex. This resemblance was confirmed when it was shown that each complex contains SR proteins and U1 snRNP. The functional relevance of these complexes was demonstrated by a correlation between the reduced ability of a mutant ESE to function in splicing activation and to assemble E complex.

Direct biochemical evidence for the involvement of U2AF in exon enhancer function has been obtained from another study (Wang *et al*, 1995). Preprotachykinin pre-mRNA containing exons 3, 4, and 5 and the intervening intron sequence splices predominantly by skipping exon 4 in HeLa cell splicing extract. Skipping is due to the presence of suboptimal 5' and 3' splice site sequences associated with exon 4 (Kuo *et al*, 1991). Replacement of exon 4 by cardiac troponin T exon 16 or the L2 exon of the adenovirus major late transcription unit promoted almost exclusive use of the central exon. Interaction of purified U2AF with the polypyrimidine tract upstream of the middle exon was

independent of the identity of the middle exon. However, interaction of U2AF as a component of HeLa cell splicing extract with this poylpyrimidine tract was dependent on the identity of the middle exon and correlated with the ability of that exon to be recognised by the splicing machinery. It was concluded that binding of U2AF to the polypyrimidine tract is assisted by components of splicing extract that recognise exon sequences. Whilst this evidence is consistent with a role for U2AF in enhancer function, it remains to be demonstrated that there is a direct interaction between components bound to an ESE and U2AF and that such an interaction stimulates splicing.

The Drosophila dsxRE and ESEs share a number of common features. They are both located within exons, stimulate splicing to upstream 3' splice sites and are involved in an interaction with SR proteins. However, the dsxRE is located 300 nucleotides downstream from the female-specific 3'ss and relies on the presence of Tra and Tra-2 proteins for activity. In contrast, the human fibronectin ED1 ESE was shown to lose its activity when positioned further than 293 nucleotides from the upstream 3'ss (Lavigueur *et al*, 1993) and has no requirement for Tra and Tra-2. The similarities between the dsxRE and ESEs prompted Tian and Maniatis (1994) to investigate the effect of placing the dsxRE closer to the dsx female-specific 3'ss. When the distance between the female-specific 3'ss and the dsxRE was reduced to 150 nucleotides, a significant level of splicing to the upstream 3'ss was observed in the absence of Tra and Tra-2. An ESE was able to substitute for the function of this constitutive activity but not for the Tra/Tra-2-dependent activity of the dsxRE. The dsxRE was shown to bind SR proteins in the absence of Tra and Tra-2 and this interaction was proposed to mediate the effect of the constitutive dsxRE. The work of Tian and Maniatis (1994) demonstrates that the dsxRE and ESEs are, under certain conditions, functionally interchangeable. It is highly likely therefore that the ESEs operate in a similar way to the *Drosophila* dsxRE, although direct biochemical evidence supporting all of the proposed interactions is currently lacking.

Recently, a further class of ESE has been discovered (Coulter *et al*, 1997). An iterative procedure (SELEX) was used to select for exon sequences that enhance exon inclusion *in vivo*. Two classes of enhancing sequence were identified: purine-rich and A/C-rich sequences. The A/C-rich sequence is very similar to the sequence of the *Drosophila* dsxRE repeats. Sequence found in three of the repeats was tested for mammalian enhancer activity and shown to be highly active. Interestingly, two human homologues of Tra2 have been identified (Dauwalder *et al*, 1996), suggesting that human Tra2 may bind to the A/C-rich sequences.

Exon enhancers represent an ideal way to control the pattern of expression in a cell type-specific, or developmental stage-specific manner. In the non-expressing cell type/developmental stage, the SR protein(s) required for the effect of the exon enhancer is inactive or not expressed and the other splicing signals

associated with the regulated exon are too weak to drive its expression. In the permissive cell type/developmental stage, all that is required for expression of the exon is expression or activation of an SR protein which was not expressed or active in the non-permissive cell type/developmental stage.

ESEs do not appear to be limited to stimulation of 3'ss use. A purine-rich ESE is involved in the regulation of competing 5'ss sequences in human caldesmon pre-mRNA (Humphrey *et al*, 1995). The mechanism by which this is achieved is not yet known.

Exon inhibitory elements

In contrast to ESEs, a number of sequences located within exons (in addition to the inhibitory element in exon 3 of *Drosophila* P element pre-mRNA) have been demonstrated to inhibit splicing to those exons. Some of those identified are listed in table 1.6.

Exon	Sequence	Reference
Human fibronectin,	CAAGG	Caputi et al, 1994
Human FGFR-2, K-SAM exon	UAGG	Del et al, 1996
Human TMnm, exon Vsk	UAAGUGUUCUGAGCU	Graham et al, 1993
HIV-1 tat exon 2, tat-rev exon 3	AGAUCC	Amendt et al, 1995;
		Staffa and Cochrane, 1995
EIAV exon 3	Exon bound by Rev protein, but sequence not defined	Gontarek and Derse, 1996

Table 1.6 Exon inhibitory elements

FGFR=fibroblast growth factor receptor. EIAV=Equine infectious anaemia virus.

Exon inhibitory elements are located adjacent to ESEs in some cases (Caputi *et al*, 1994; Amendt *et al*, 1995; Staffa and Cochrane, 1995; Gontarek and Derse, 1996). This has led to suggestions that they may function to recruit factors which sterically interfere with the interaction of SR proteins with the adjacent ESE. Thus, exon inhibitory elements may provide a way of regulating the activity of an ESE in a cell-type or developmental stage specific manner if the activity of the SR protein which recognises that ESE is unaltered in the cell types or developmental stages in which regulation occurs.

Alternative polyadenylation site use

Most eukaryotic pre-mRNAs undergo cleavage at their 3'-end concomitant with addition of a poly(A) sequence. Addition of the poly(A) sequence is known

as polyadenylation and causes RNA polymerase II to terminate transcription, often a significant distance downstream from the 3'-end of the mature mRNA, and also confers RNA stability and translatability on the mRNA (reviewed in Proudfoot, 1988, 1991). Polyadenylation is signalled by two elements: an AAUAAA sequence located approximately 20-30 nucleotides upstream of the polyadenylation site (poly(A) site), and a diffuse GU-rich sequence located immediately downstream from the 3'-end of the mRNA. Other elements upstream of the poly(A) site may influence polyadenylation efficiency (reviewed in Proudfoot, 1991). The elements which signal polyadenlyation serve as nucleation sites for the formation of a multicomponent complex which assembles prior to cleavage at the 3'-end of the pre-mRNA and polyadenylation (reviewed in Keller, 1995).

Some pre-mRNAs contain alternative 3'-end exons with associated polyadenylation signals. An example which has been extensively studied is pre-mRNA expressed from the CT/CGRP gene. In thyroid C cells, use of exon 4 and its associated poly(A) site gives rise to calcitonin (CT). However, in neural tissue, exon 4 is skipped and the polyA site associated with exon 6 is used, resulting in production of calcitonin gene-related peptide (CGRP) (figure 1.18). Subsequent proteolytic processing removes the first three exons, so the mature peptides are encoded entirely by the variable sequences (reviewed in Smith *et al*, 1989). The two proteins have completely different function: CT is involved in calcium and phosphate homeostasis and CGRP is a neurotransmitter.
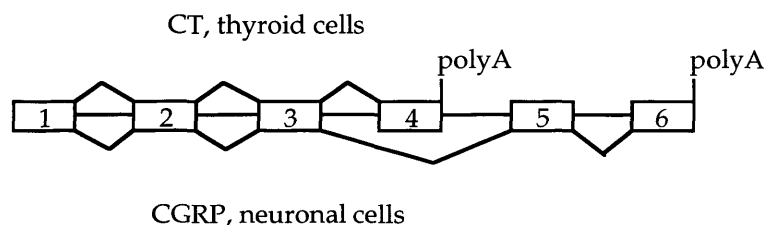
Figure 1.18 Alternative splicing of CT/CGRP pre-mRNA

Cis-acting sequences required for expression of exon 4 *in vivo* have been identified both within exon 4 (Yeakley *et al*, 1993) and within the intron immediately downstream of exon 4 (Lou *et al*, 1995). The exon contains two enhancer sequences, one is purine-rich (table 1.4) and the other resembles the *Drosophila* dsxRE repeat elements. The intronic enhancer sequence is 127 nucleotides in length beginning 168 nucleotides downstream of the exon 4 poly(A) site and contains a pseudo 5'ss (p5'ss) sequence preceded by a pyrimidine-rich tract. Mutation of the p5'ss sequence, or conversion of some of the pyrimidine nucleotides to purines, inhibited exon 4 expression from a minigene *in vivo*, and polyadenylation cleavage *in vitro* (Lou *et al*, 1996). Other experiments showed that the p5'ss sequence is bound by U1 snRNP and that the

1.39

pyrimidine-rich tract is bound by a protein called polypyrimidine tract binding protein (PTB). When interaction of U1 snRNP with the p5'ss sequence or of PTB with the pyrimidine-rich tract was inhibited, polyadenylation immediately downstream of exon 4 was also inhibited. It was proposed that U1 snRNP binds to the p5'ss sequence within the intronic enhancer element and stimulates polyadenylation at the exon 4 poly(A) site. Interaction of PTB with the pyrimidine-rich region of the intron enhancer may function to prevent the productive recognition of the p5'ss as an authentic 5'ss sequence so that U1 snRNP bound to this site is instead involved in the activation of polyadenylation.

It was also proposed that stimulation of polyadenylation immediately downstream of exon 4 enhances recognition of the exon 4 3'ss by the splicing machinery. Terminal exons lack 5'ss sequences, so the exon bridging interactions proposed to occur across internal exons involving the 5'ss sequence cannot take place. Instead, at the 3'-end of pre-mRNAs, components of the polyadenylation machinery could interact with the splicing machinery and stimulate recognition by the splicing machinery of the 3'ss sequence of the terminal exon (Robberson *et al*, 1991). The requirement for an intron enhancer element for both expression of exon 4 and polyadenylation immediately downstream of exon 4 is consistent with interaction of the polyadenylation and splicing machinery. An alternative explanation for the activity of this intron enhancer element is that it is able to enhance both polyadenylation and exon 4 3'ss use and that no interaction of the poladenylation machinery with the splicing machinery is required.

Mutually exclusive alternative splicing

Several of the tropomyosin genes contain pairs of exons which are expressed in a tissue-specific mutually exclusive manner. These genes and mechanisms by which the regulation of mutual exclusion is thought to occur are discussed in detail below.

The tropomyosin family

The tropomyosins (TMs) are a highly conserved family of structural proteins (reviewed in Pittenger *et al*, 1994) that have been identified in many organisms, including yeast, nematodes, *Drosophila*, avians and mammals. They have a dimeric α-helical coiled coil structure along their entire length. This structure is a reflection of their amino acid sequence which comprises seven-residue repeats with hydrophobic residues at positions one and four. There are a large number of tropomyosin (TM) isoforms and these are expressed in a tissue-specific manner (see table 1.7). TM isoform diversity is greater in non muscle cells

TM gene	Isoforms in different cell-types					Number of isoforms
	Fibroblasts	Skeletal muscle	Smooth muscle	Brain	Cochlea	
α	TM-2 TM-3 TM-5a TM-5b	Skeletal muscle α-TM ($\alpha_{\text{fast twitch}}$-TM)	Smooth muscle α-TM	TMBr-1 TMBr-2 TMBr-3		9
β	TM-1	Skeletal muscle β-TM	Smooth muscle β-TM (=TM-1)			2
TM-4	TM-4					1
TM$_{nm}$	TM30 (=TM-5)	Skeletal muscle ($\alpha_{\text{slow twitch}}$-TM)			NM-1 (=TM30) NM-2 NM-3 NM-4	5

Table 1.7 Mammalian tropomyosin isoform diversity

The α-TM, β-TM genes have been characterised in rats (Ruiz-Opazo and Nadal, 1987; Helfman et al, 1986; Lees-Miller et al, 1990; Goodwin et al, 1991). The TMnm gene has been characterised in humans and rats (Clayton et al, 1988; Beisel and Kennedy, 1994; Constance et al, 1996). The TM-4 gene has been described in rats (Lees-Miller et al, 1990). The fibroblast isoforms are also expressed in different combinations in various nonmuscle cell-types.

than in muscle cells. Tropomyosin isoform diversity is achieved in a number of ways; through the use of several different tropomyosin genes (in mammals there are four genes: the α-, β-, TM-4, and TM_{nm} genes), the use of alternative transcription promoters and through alternative RNA processing. The proteins are also able to form homo- or heterodimers, so a large number of different combinations of dimerised proteins are possible.

Explanations for the existence of so many tropomyosin isoforms are only just beginning to emerge (reviewed in Pittenger *et al*, 1994). These proteins all bind to actin filaments and, in skeletal and cardiac muscle, tropomyosin functions in association with a complex of proteins known as the troponin complex (troponin-I, -T, and -C) to regulate the calcium-sensitive interaction of actin and myosin. Binding of tropomyosin to actin is a cooperative process involving head-to-tail interaction between adjacent tropomyosin molecules and is affected by differences in amino- and carboxy-terminal domains between isoforms (Cho *et al*, 1990; Matsumura and Yamashiro, 1985). Thus, one explanation for TM isoform diversity is that it allows differences in the interaction with actin.

In non muscle and smooth muscle cells the troponin complex is absent and the calcium-dependent interaction of actin and myosin is instead regulated by myosin light chain kinase and by caldesmon in cooperation with calmodulin. In non muscle cells, actin also functions in intracellular transport, secretion, motility and cell structure. The organisation of actin monomers into different types of polymeric structure is associated with this variety of function. Very little is understood about the function(s) of tropomyosin in non muscle cells, but the different isoforms are expected to be important because they have been highly conserved. It seems likely that the wider variety of tropomyosin isoform expression in non muscle cells compared with muscle cells reflects differences in the way the contractile apparatus is regulated, and the greater diversity of roles played by actin in non muscle cells.

Antibody studies *in vivo* have shown differences in the localisation of some tropomyosin isoforms (Lin *et al*, 1988), suggesting that they may have distinct functions *in vivo*. It may be that the different TM isoforms have specific functions that arise from localisation to particular actin structures. Experiments on bacterially expressed cDNAs encoding TM isoforms have shown that the different isoforms do not have the same affinity for actin (Pittenger and Helfman, 1992) and differ in their ability to form dimers (Matsumura and Yamashiro, 1985; Lin *et al*, 1985; Lehrer and Stafford, 1991; Jancso and Graceffa, 1991). Fibroblast isoforms exist mainly as homodimers, whereas smooth and skeletal muscle TMs exist almost entirely as α/β heterodimers. Homodimers of smooth muscle α- and β-TMs do not bind as well to actin as heterodimers (Jancso and Graceffa, 1991). The ability of different TM isoforms to form dimers was investigated *in vivo* by

immunoprecipitating epitope-tagged isoforms, followed by analysis of dimer composition (Gimona *et al*, 1995). When muscle-specific isoforms were introduced into non muscle cells they were still able to form heterodimers. Similarly, when non muscle isoforms were introduced into muscle cells homodimers were formed. It was concluded that the information required for dimer formation was contained within the isoform, and was not dependent on non muscle-, or muscle-specific factors.

The regulation of actin stability is likely to be tightly controlled. The entire actin network must be disassembled before mitosis and reassembled again after division. Assembly and disassembly of actin microfilaments is also required for cell movement. It has been proposed that tropomyosin promotes the stability of actin in non muscle cells, suggesting that regulation of the actin-tropomyosin interaction could contribute to control of actin network dynamics in these cells. The protein has been shown to protect actin from the severing action of gelsolin (Ishikawa *et al*, 1989) and villin (Burgess *et al*, 1987) and differences have been observed in the ability of individual isoforms to do this (Ishikawa *et al*, 1989). TMs are known to interact with a number of other proteins including tropomodulin, caldesmon and calponin. Caldesmon has been shown to enhance the binding of the TM-1 isoform (see table 1.7) to actin (Pittenger *et al*, 1995). The interaction of caldesmon with tropomyosin is regulated by calcium/calmodulin and by phosphorylation (Yamashiro et al, 1990). It has been proposed that regulation of the binding of caldesmon to tropomyosin by phosphorylation or changes in the intracellular calcium level could be a mechanism of regulating the interaction of tropomyosin with actin, and therefore the stability of actin, in non muscle cells. Other factors shown to be important in the actin-tropomyosin interaction are acetylation of the amino terminus of TM-1, or the presence of other TM isoforms (Pittenger et al, 1995).

At the RNA processing level, tropomyosin isoform diversity is achieved by mutually exclusive alternative pre-mRNA splicing in a tissue-specific manner. Here, one exon of a pair of mutually exclusive exons is expressed in one tissue type. The other exon of the pair is expressed in another tissue type. In all cases of mutually exclusive tissue-specific alternative splicing involving tissue A and tissue B, three questions arise:

      1) What determines the splicing pattern in tissue A?
      2) What determines the splicing pattern in tissue B?
      3) What determines mutual exclusion?

Several of the tropomyosin genes have been studied in order to gain an understanding of this process. The most extensively studied examples are shown in figure 1.19. The introns that separate these mutually exclusive exon pairs are unusual in that the branchpoint is located abnormally far from the 3'ss (the usual
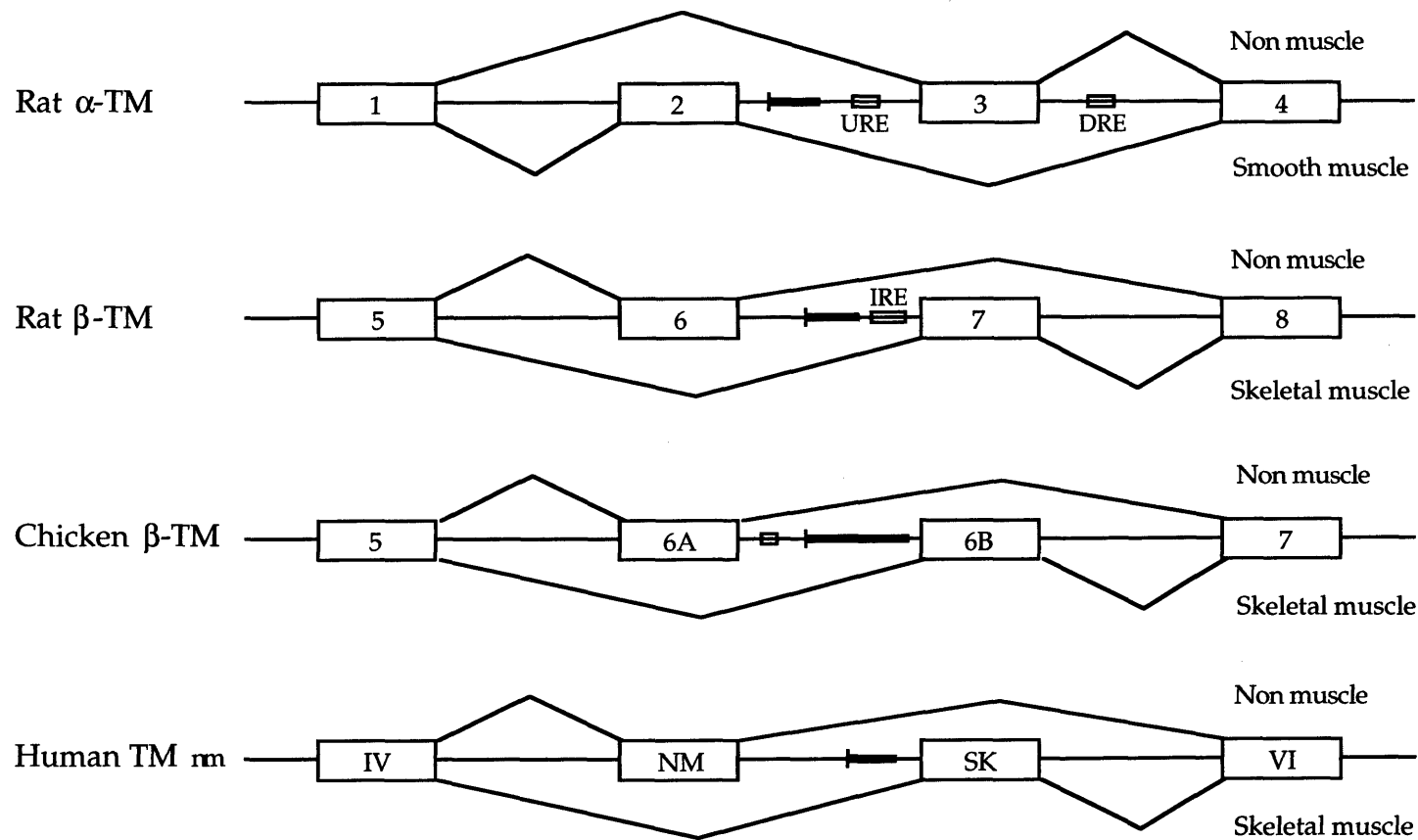
Figure 1.19 Mutually exclusive exon use in different tropomyosin genes.

The exon organisation of the 5'-end of the rat α-TM gene and of the central region of the rat β-TM, chicken β-TM and human TMnm genes is shown. Exon sequences are represented as boxes, introns as thin horizontal lines. The dominant splicing pattern in nonmuscle and muscle tissues is shown for each gene by the angled thin lines. The approximate position of the BPS and PPT is shown in the intron separating the mutually exclusive exon pair for each gene (BPS=short vertical line, PPT=thick horizontal line). The length of the PPT is not drawn to scale. Known intron regulatory elements are shown in their approximate positions but are not drawn to scale. The URE of rat α-TM is ~50 n.t. upstream of exon 3, the DRE is ~130 n.t. downstream from exon 3 (Gooding et al, 1994). The IRE of rat β-TM is located between the PPT of intron 6-7 and the 3'-end of the intron (Helfman et al, 1990). The intron element of chicken β-TM is ~30 nucleotides long and located 37 n.t. downstream of exon 6A (Balvay et al, 1992).

distance is 20-40 nucleotides) and is adjacent to a very long polypyrimidine tract (figure 1.20).

Rat α-TM | 2 | 42 ─┤▬▬▬ 177 ──── | 3 |
                        50

Rat β-TM | 6 | ~140 ────┤▬▬▬ ~150 ──── | 7 |
                            55

Chicken β-TM | 6A | 264 ────────┤▬▬▬ 105 ── | 6B |
                                  80

Human TM nm | Vnm | 362 ─────────┤▬▬▬ 79 ── | Vsk |
                                   46
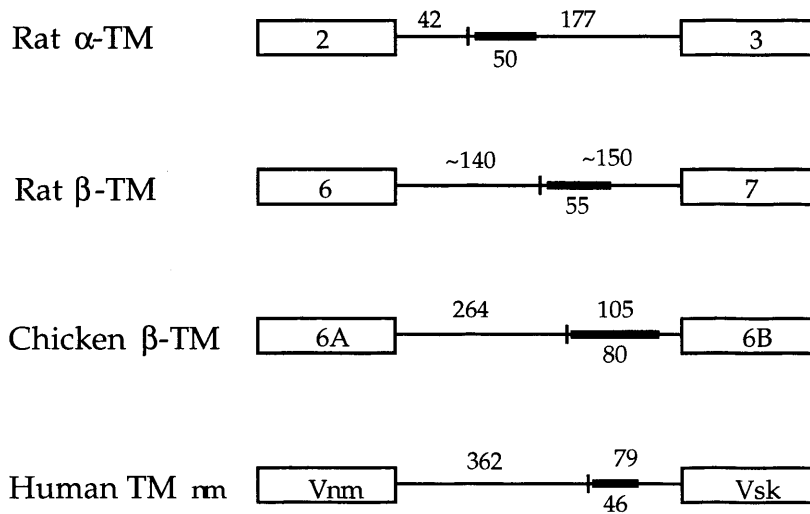
Figure 1.20 Comparison of rat α-TM, rat β-TM, chicken β-TM, and human TMnm intron structure.

The branch point is represented by a thin vertical line, and the polypyrimidine tract by a thick horizontal line. Numbers are in nucleotides and above the intron line give the intron length upstream and downstream of the branch point. Below the polypyrimidine tract the numbers refer to the length of the polypyrimidine tract. Approximate values are given for the rat β-TM intron because three branch points are used in close proximity to each other (Helfman and Ricci, 1989).

Rat α-TM gene

The rat α-TM gene contains a pair of mutually exclusive exons towards the 5'-end of the gene (figure 1.19). Exon 3 is expressed in non muscle cells, and exon 2 in smooth muscle cells. It was noted that the branch point upstream of exon 3 is abnormally close to the 5'ss of exon 2 (42 nucleotides). It was proposed that exons 2 and 3 are used mutually exclusively because the proximity of these two elements presents a steric block to removal of the intron that separates them. This was confirmed when spacer sequence was inserted between these two elements and removal of the intron was observed (Smith and Nadal-Ginard, 1989).

In non muscle cells, deletion of exon 3 resulted in efficient use of exon 2. Thus, normally exon 2 is not expressed in non muscle cells because the activity of exon 3 inhibits exon 2 use. It is proposed that exons 2 and 3 are in competition with each other in non muscle cells and that exon 3 use dominates because the polypyrimidine tract and branch point sequence which form part of the exon 3 3'ss sequence are stronger than the corresponding sequences associated with the exon 2 3'ss sequence (Mullen et al, 1991).

When expression of exon 2 was disrupted in smooth muscle cells, no expression of exon 3 was observed. It was concluded that exon 3 is specifically

inhibited in smooth muscle cells. Conserved sequence located in the introns flanking exon 3 was shown to be required for inhibition of exon 3 expression in these cells (Gooding *et al*, 1994). The sequence located in the upstream intron was termed the URE and the sequence in the downstream intron, the DRE (figure 1.19). The URE and DRE are similar to sequence shown to have high affinity for the polypyrimidine tract binding protein (PTB) (Singh *et al*, 1995; Perez *et al*, 1997)). The URE also overlaps with the polypyrimidine tract (PPT) sequence required for splicing to exon 3. It was proposed that interaction of PTB with the URE and DRE could prevent binding of U2AF to the polypyrimidine tract upstream of exon 3 and inhibit splicing to exon 3 in smooth muscle cells. Consistent with this proposal, PTB was demonstrated to bind with high affinity to the intron sequences either side of exon 3. Further, *in vitro* PTB was shown to inhibit use of the 3'ss of rat α-TM exon 3 and to compete with U2AF for binding to the polypyrimidine tract associated with that exon (Singh *et al*, 1995).

Evidence for the model was also obtained from another study (Lin and Patton, 1995). Pre-mRNA containing the first three exons of the rat α-TM gene is usually spliced almost exclusively in non muscle cell splicing extract from exon 1 to exon 3 (1-3 splicing). However, when the ionic conditions and concentration of the splicing extract were altered, a significant level of splicing from exon 1 to exon 2 (1-2 splicing) was also observed. Addition of recombinant PTB protein to these reactions specifically inhibited 1-3 splicing and had no effect on 1-2 splicing. This inhibition was antagonised to some extent when the wild-type α-TM pre-mRNA was pre-incubated with U2AF before addition to extracts supplemented with PTB. However, 1-2 splicing did not increase, suggesting that other factors are involved in the normal regulation of α-TM pre-mRNA splicing. When the pre-mRNA was pre-incubated with SR proteins before being added to PTB-supplemented splicing extract, the ratio of 1-2 to 1-3 splicing was reversed and was similar to that observed in smooth muscle cells.

It was proposed that α-TM pre-mRNA splicing could be regulated in smooth muscle cells by an alteration in the levels or activity of PTB and one or more SR proteins. An increase in active PTB levels in these cells would prevent the association of U2AF with the PPT of intron 2 by binding to that sequence, and an increase in the level of an active SR protein(s) would activate splicing to exon 2. This model provides another example of the proposal that alteration in the levels of constitutively-acting factors is central to tissue-specific regulation of alternative splicing.

Rat and chicken β-TM genes

The rat β-TM and chicken β-TM genes both contain a centrally located mutually exclusive exon pair (figure 1.19). In both genes, the muscle specific exon remains inactive in non muscle cells when the non muscle exon is inactivated by

mutation. Thus, the muscle specific exon appears to be repressed in non muscle cells. Mutation of sequence at the 5'-end of the muscle specific exon, or of sequence between the branch point and the 3'ss of the muscle specific exon (the IRE in rat β-TM, figure 1.19) activated expression of the muscle specific exon in non muscle cells (Helfman *et al*, 1990; Guo *et al*, 1991; Libri *et al*, 1990; Libri *et al*, 1992; Gallego *et al*, 1992). It has been shown *in vitro* that PTB is able to bind specifically to the polypyrimidine tract and to the IRE located between exons 6 and 7 of rat β-TM pre-mRNA (Mulligan *et al*, 1992). PTB was able to compete with U2AF for binding to the polypyrimidine tract of this intron (Singh *et al*, 1995). Mutations of the IRE sequence which allow expression of exon 7 in non muscle cells also disrupt PTB binding (Mulligan *et al*, 1992). It is proposed, that PTB is able to inhibit splicing of the rat β-TM muscle-specific exon 7 in non muscle cells by preventing the interaction of U2AF with the PPT of the intron upstream of exon 7. The role of the IRE sequence in this repression is yet to be established, but it may act to recruit PTB to the vicinity of the PPT.

Based on computer secondary structure predictions using chicken β-TM pre-mRNA, it was proposed that the muscle specific exon 6B is inactive in non muscle cells because it is part of a secondary structure that makes it inaccessible to the splicing apparatus (Clouet d'Orval *et al*, 1991). Evidence for such a model was obtained from experiments *in vitro* . Mutations were made in a region of the pre-mRNA predicted to be involved in base pairing with another region of the pre-mRNA. These mutations activated splicing to exon 6B. Compensatory mutations in the region predicited to base pair with the wild-type version of the mutated region were made. These mutations restored the complementarity of the two regions and restored inhibition of splicing to exon 6B (Clouet-d'Orval *et al*, 1991). However, subsequent experiments have shown that the extent of this secondary structure *in vivo* is much less than was originally proposed and that its role in repression is very limited (Libri *et al*, 1992; Gallego *et al*, 1992).

Other mutations of chicken β-TM pre-mRNA have identified a pyrimidine-rich intron sequence just downstream of exon 6A that is required for repression of the skeletal muscle exon 6B and for use of the non muscle exon 6A in non muscle cells (Balvay *et al*, 1992). Mutation of this element did not repress use of exon 6A in muscle cells, suggesting that exon 6A is specifically activated in non muscle cells (Gallego *et al*, 1992). Interestingly, this sequence which is not purine-rich, could be replaced by a purine-rich ESE sequence without effect on exon 6A expression (Gallego *et al*, 1997). This prompted Gallego *et al* (1997) to test the effect of SR proteins on expression of exon 6A in non muscle cell splicing extracts. SF2/ASF was demonstrated to stimulate splicing to exon 6A.

In muscle cells, exon 6A and 6B are active and in competition with each other, unlike in non muscle cells where exon 6B is inactive. The 5'ss sequence and branch point sequence of exons 6A and 6B are suboptimal, but those of exon

6B are stronger relative to those of exon 6A. It was proposed that exon 6B is used in preference to exon 6A in muscle cells because its 5'ss sequence and branch point sequence are able to compete for recognition by the splicing apparatus more effectively than those of exon 6A (Libri *et al*, 1992). It was also discovered that the effect of SF2/ASF on exon 6A expression in non muscle cell splicing extracts was counteracted by SC35 (Gallego *et al*, 1997) and that SR proteins isolated from skeletal muscle tissue were not able to enhance splicing to exon 6A. Thus, the relative strengths of splicing signals associated with each exon and the activity of individual SR proteins are predicted to determine the outcome of chicken β-TM pre-mRNA splicing in muscle cells.

Some experiments have provided information to explain the mutual exclusivity of exons 6A and 6B. Mutation of the 5'ss of exon 6A to the consensus sequence and removal of the inhibitory sequence at the 5'-end of exon 6B resulted in expression of an isoform containing exons 6A and 6B (Libri *et al*, 1992). There is therefore no intrinsic block to removal of the intron separating exons 6A and 6B as there is for the intron separating the rat α-TM exons 2 and 3. When the distance of the branch point of the intron separating exons 6A and 6B from the 3'ss of exon 6B was reduced, exons 6A and 6B could be spliced together. Spacer sequence was used to restore the original distance between the branch point and the exon 6B 3'ss, but exons 6A and 6B could still be spliced together. Thus, the distance of the branch point from the exon 6B 3'ss, as well as the actual sequence separating these elements, appears to be important. It was proposed that these factors, together with the suboptimal exon 6A 5'ss and exon 6B inhibitory sequence ensure relatively inefficient removal of the intron separating the two exons (Goux-Pelletan *et al*, 1990).

Human TMnm gene

As with the rat and chicken β-TM genes, the human TMnm skeletal muscle specific exon SK is not in competition with the non muscle exon NM in non muscle cells, but is instead inactivated (Graham *et al*, 1992; Hamshere, PhD thesis). Sequence at the 5'- and 3'-end of exon SK is required for inhibition of splicing to that exon in non muscle cells. Also important is the branch point sequence upstream of exon SK; when this sequence was mutated to the consensus sequence, expression of exon SK was observed in non muscle cells (Graham *et al*, 1992; Hamshere, PhD thesis).

When the NM and SK exons were precisely swapped and expression of the resulting mutant gene determined in non muscle cells, no expression of exon SK was observed (Graham *et al*, 1992). It was concluded that the sequence of the SK exon, and not its intronic environment was a primary determinant of exon SK inactivity in non muscle cells. Thus, the mechanism of hTMnm alternative pre-mRNA splicing regulation is likely to be different to that of the rat and chicken β-

TM pre-mRNAs where intron sequences have been shown to play key roles in the expression pattern observed in non muscle cells (see above).

The splicing pattern of the exon-swap mutant gene described above was also determined in muscle cells. The only mRNA isoform detected was the NM exon-containing isoform. In comparison, the wild-type gene expressed predominantly the SK exon-containing isoform (Hamshere, PhD thesis). It was concluded that the intron sequence flanking the mutually exclusive exon pair is a major determinant of the splicing pattern in muscle cells.

As with chicken β-TM pre-mRNA, there is no intrinsic block to removal of the intron separating the NM and SK exons; a mutant gene in which the only intron present was that separating the NM and SK exons, expressed mRNA containing both exons (Graham *et al*, 1992).

Chapter 2
Materials and methods

Solutions

TE.1

10mM Tris-HCl, pH 7.5

0.1mM EDTA

TES

10mM Tris, pH 7.6

1mM EDTA

0.5% SDS

TY medium

18g/l peptone

10g/l yeast extract

10g/l NaCl

Ampicillin plates

18g/l peptone

10g/l yeast extract

10g/l NaCl

15g/l agarose

50μg/ml ampicillin

DMEM

As powder from Gibco BRL

Add 2g/l $NaHCO_3$

2x HBS

280mM NaCl

10mM KCl

1.5mM $Na_2HPO_4.2H_2O$

12mM D-glucose

50mM HEPES, pH 7

1x PBS
        8g/l NaCl
        0.2g/l KCl
        1.44g/l Na_2HPO_4
        0.24g/l KH_2PO_4
        Adjust pH to 7.4 with HCl

Trypsin-EDTA (Gibco-BRL)
        0.5g/l trypsin
        0.2g/l EDTA
        In modified Puck's Saline A

10x TAE
        0.4M Tris-acetate
        10mM EDTA

10x TBE
        0.9M Tris-borate
        20mM EDTA

Formamide dyes
        90% (v/v) Formamide
        50mM EDTA, pH 8.0
        0.025% (w/v) bromophenol blue
        0.025% (w/v) xylene cyanol

Native gel loading dye
        6x TAE or TBE
        30% (v/v) glycerol
        0.25% (w/v) bromophenol blue
        0.25% (w/v) xylene cyanol

2x SDS gel loading dye
        100mM Tris-HCl, pH 6.8
        200mM dithiothreitol
        4% SDS
        0.2% bromophenol blue
        20% glycerol
Stored at room temperature in absence of dithiothreitol. Dithiothreitol added just before use

Fixing solution
    10% glacial acetic acid
    12% ethanol

10x kinase buffer
    0.5M Tris-HCl, pH 7.5
    0.1M Magnesium chloride
    10mM DTT

1x Barnes buffer pH 8.55
    20mM Tris-HCl (pH 8.55)
    Bovine serum albumin (150µg/ml)
    16mM ammonium sulphate
    3.5mM magnesium chloride
    250µM each dNTP

10x PCR buffer
    0.5M KCl
    100mM Tris-HCl, pH 7.5
    25mM magnesium chloride
    2mM each dNTP
    2µg BSA

1x binding buffer
    1M lithium chloride
    0.2M Tris-HCl, pH 7.5
    2mM EDTA

Buffer A
    50ml buffer A base (10mM KCl, 1.5mM MgCl$_2$)
    0.5ml 1M TEA
    25µl 1M DTT

Buffer C
    10ml Buffer C base (20ml glycerol, 4.5g KCl, 40µl 0.5M EDTA,water to
    97.5ml)
    0.2ml 1M TEA
    50µl PMSF stock solution
    5µl 1M DTT

Buffer D
   500µl 1M DTT
   20ml 1M TEA
   5ml PMSF stock solution
   Buffer D base (80mM potassium glutamate, 5% glycerol, 0.2mM Na⁺-free
   EDTA, pH 8.0) to 1 litre

PMSF stock solution
   0.13g PMSF per 7.5ml propan-2-ol
   Stored at -20°C

Proteinase K mixture
   1 volume proteinase K stock (10mg/ml in water) added to 25 volumes
   proteinase K buffer (2ml 1M Tris-HCl, pH 7.5, 0.5ml 0.5M EDTA, 0.6ml 5M
   sodium chloride, 2ml of 10% SDS, 14.9ml water)

5x TMS
   250mM Tris, pH 7.5
   50mM magnesium chloride
   10mM spermidine

Low G NTPs
   2.5mM rCTP
   2.5mM rATP
   2.5mM rUTP
   0.25mM rGTP

RNA elution buffer
   16.5ml 3M NaAc
   0.2ml 0.5M EDTA
   2ml 10% SDS
   Water to 100ml

RNAse mix
   1x RNAse dilution buffer
   37.5 µg RNAseA
   100 units RNAse T1
   4.5µg yeast tRNA
   water to 37.5µl

1x RNAse dilution buffer

      10mM Tris-HCl pH 7.5

      1mM magnesium chloride

      0.1M KCl

Blot buffer

      48mM Tris

      39 mM glycine

      10% methanol

      0.0375% SDS

Extraction buffer

      Buffer D (20% glycerol) with 0.4M potassium glutamate

Dilution buffer

      Buffer D (20% glycerol) without potassium glutamate

FSP buffer

      60mM KCl

      2.5mM EDTA

      20mM Tris-HCl, pH 7.5

      0.1% Triton-x-100

Oligonucleotide primers

<u>Oligonucleotide primers used for inverse long PCR mutagenesis</u>
The SK16-30 mutant TMnm minigene plasmids:
Random primer 1:
TAAGTGTTCTGAGCTNNNNNNNNNNNNNNNGAATGTCACCAACAACC
TCAAGTCTCTTG (59-mer)
Random primer 2:
TTGTTGGTGACATTCNNNNNNNNNNNNNNNAGCTCAGAACACTTACTG
TGAATATTTTA (59-mer)
N=any nucleotide.
The SK2-4CGC TMnm minigene plasmid:
SK2-4CGC 1 TAAAATATTCACAGTCGCTGTTCTGAGCTGGAGGAGGAG
SK2-4CGC 2 CTCCAGCTCAGAACAGCGACTGTGAATATTTTAAATACC
The αCGC U1 snRNA-encoding plasmids:
U1 αCGC 1 CCAAGATCTCATAGCGACCTGGCAGGGGAGATACCATGA
U1 αCGC 2 CTGCCAGGTCGCTATGAGATCTTGGGCCTCTGCCCCGAC

<u>Other oligonucleotide primers used to generate recombinant plasmids</u>
The ß-G/TM1 plasmid:
ClaI SKBP primer   GGGGATCGATAATTCACGTCTATGT
ClaI SK 3' primer   CCCCATCGATCCCTACCTTCTCCGC
The ß-G/TM2 plasmid:
ClaI SKBP primer   GGGGATCGATAATTCACGTCTATGT
ClaI SK ds primer   CCCATCGATCCTGTCACTTTCACA
The pcDNA3-based U1WT plasmid:
U1us primer   GGCGAAGCTTCTTTGGGAGA
U1ds primer   GGCGTCTAGACACCACTGTAGGAT

<u>Oligonucleotide primers used to determine expression from the TMnm
minigene and the ß-globin, ßG/TM1, and ßG/TM2 genes:</u>
TMnm minigene expression:
RT-PSVR   CACTGCATTCTAGTTGTGGT
Exon VI/VII   ACGGGTCTCTGCCTCCTT
CAT5015   GTGTGCACCTCCAAGCTTCCGTCGACTTGGTCACCCTTGG
Exon IV   GAGGAACGAGCTGAGCT
ß-globin, ßG/TM1, ßG/TM2 expression:
ß-globin 3'   TGCAATGAAAATAAATTTCCTTTAT
PCR1 exon 1   AGGTGAATGTGGAAGAAGTT
PCR1 exon 3   GAGTGAATTCTTTGCCAAAA
PCR2 exon 1   GGTGAGGCCCTGGGCA

Oligonucleotide primers used to generate DNA templates for *in vitro* transcription

T7NM	AAATTAATACGACTCACTATAGGGCCGTTGCCGAGAGAT
T7SKus	AAATTAATACGACTCACTATAGGGCCTCTCCTGTGGTAT
T7SK	AAATTAATACGACTCACTATAGGGTAAGTGTTCTGAGCT
SK3'	CTTCTCCGCCTGAGC
NM3'	CTTTTCTTCAGCAGC

2.7

Plasmids

The TMnm minigene plasmids

Construction of the plasmid containing the wild-type human TMnm minigene (referred to as SKWT in the text) has been described in Graham *et al* (1992). Briefly, the minigene was produced by replacing a portion corresponding to exons IV through to VI (inclusive) of the skeletal form of TMnm cDNA with genomic sequence from exons IV through to VI (*i.e.* including the three introns flanking and separating exons NM and SK). This cDNA-genomic DNA hybrid was cloned into an expression vector under the control of the SV40 early promoter and upstream of an SV40 polyadenylation signal, and is represented diagrammatically in figure 2.1.

Mutants SK1-15NM and SK61-75NM are the same as SKWT except that nucleotides 1-15 and 61-75, respectively, of the SK exon were substituted with the corresponding sequence from the NM exon. Similarly, SK16-30NM is the same as SKWT except that nucleotides 16-30 of exon SK have been substituted with the corresponding sequence from exon NM. The sequence of the other mutations at SK nucleotides 16-30 are as follows:

      SK16-30R1 TCGCGCTACACGCTT
      SK16-30R2 ATTTATAAGCTCCAA
      SK16-30R3 CAAACCCAGAAAGAA
      SK16-30R4 TGTGTCATTTTATGA
      SK16-30R5 CCCTTTCTCACCAGC

The wild-type branch point sequence upstream of exon SK (the SK BPS) was mutated from CAAUUCACGUCU to CAUACUAACGUGU (a consensus branch point sequence) to generate the bpup minigene. The SK16-30R1 mutation was made of the bpup minigene so the TMnm minigene plasmid containing this mutation is referred to as SK16-30R1bpup. The SK16-30R2, SK16-30R3, SK16-30R4, and SK16-30R5 mutations were all made using the SKWT TMnm minigene and are referred to as SK16-30R2, SK16-30R3, SK16-30R4, and SK16-30R5, respectively. The ΔNM5'ss mutant contains a deletion of the first two nucleotides (GT) of the NM-SK intron. The exon SK BPS is mutated to CAAGUUAGUA in the ΔSKBP mutant and in the NM5'ssC construct, the 5'ss sequence of exon NM is changed to AAG|GTAAGT (where "|" represents the exon-intron boundary). The SK2-4CGC minigene is identical to SKWT except that exon SK nucleotides 2 to 4 were altered from AAG to CGC.

The SKWT, SK1-15NM, SK16-30NM, SK61-75NM, bpup, ΔNM5'ss, ΔSKBP, and NM5'ssC were made previously (Graham *et al*, 1992, Ayres, unpublished).

The plasmids SK16-30R1bpup, SK16-30R2, SK16-30R3, SK16-30R4, SK16-30R5, SK2-4CGC were generated by inverse long PCR mutagenesis (see below) and were sequenced across the SK exon upto the SK branch point sequence.

Figure 2.1 Schematic diagram of plasmids
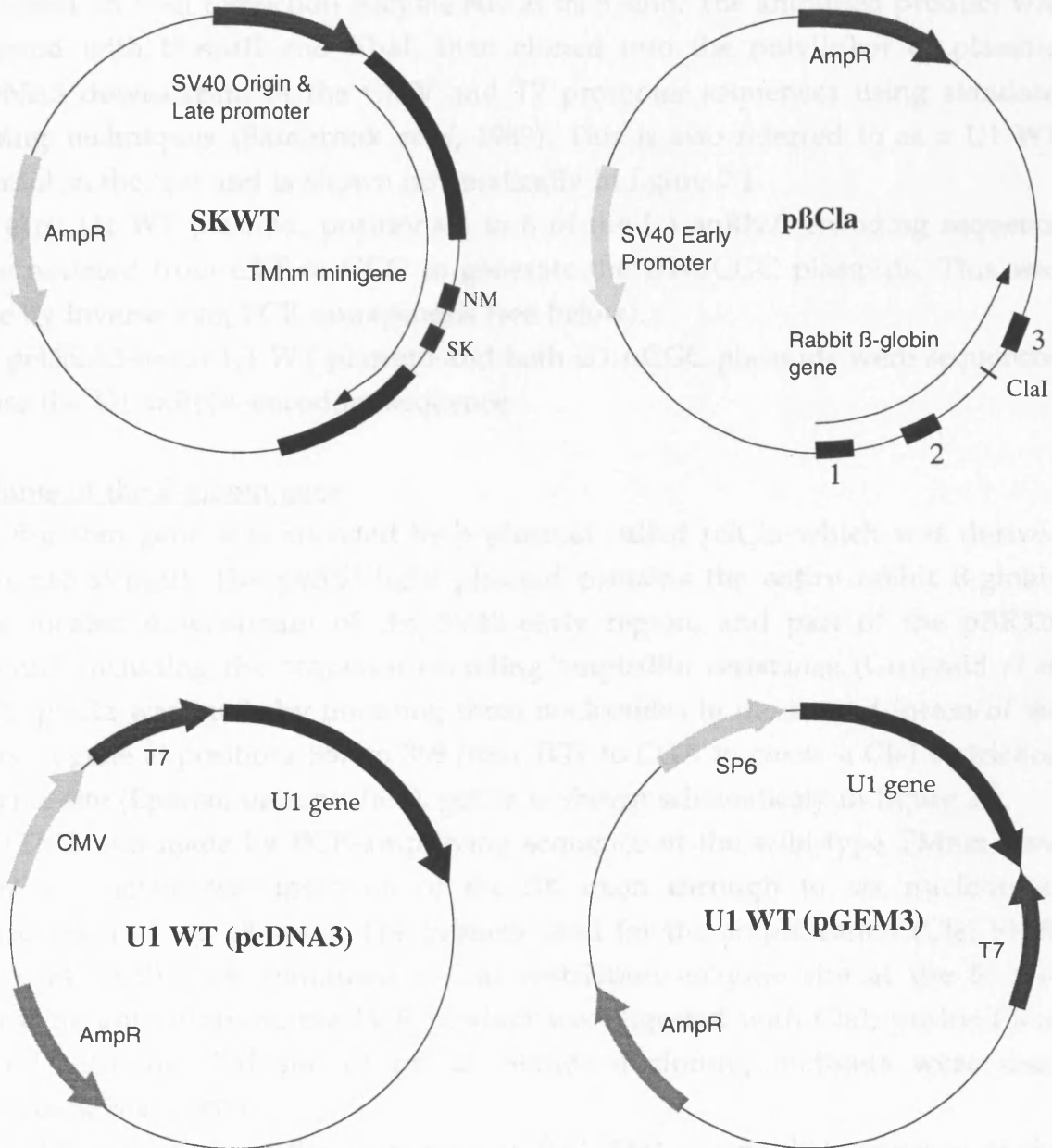
T7=T7 promoter sequence, SP6=SP6 promoter sequence, CMV=CMV promoter sequence.
AmpR=Sequence encoding ampicillin resistance, ClaI=ClaI resitriction enzyme site.
The arrows show the direction of transcription.

U1 snRNA-encoding plasmids

A fragment (-390 to +200, Lund and Dahlberg, 1984) of the human U1 snRNA gene was BamHI linkered and then cloned into the polylinker of plasmid pGEM3 downstream of the SP6 promoter sequence (Smith, unpublished). This is referred to as a U1 WT plasmid in the text and is shown schematically in figure 2.1.

A fragment (-384 to +295) of the human U1 snRNA pseudogene HSD2 (Manser and Gesteland, 1982) was amplified by PCR. One PCR primer, U1us, contained a HindIII restriction enzyme site at its 5'-end. The other PCR primer, U1ds contained an XbaI restriction enzyme site at its 5'-end. The amplified product was digested with HindIII and XbaI, then cloned into the polylinker of plasmid pcDNA3 downstream of the CMV and T7 promoter sequences using standard cloning techniques (Sambrook et al, 1989). This is also referred to as a U1 WT plasmid in the text and is shown schematically in figure 2.1.

For each U1 WT plasmid, positions 4 to 6 of the U1 snRNA encoding sequence were mutated from CTT to GCG to generate the U1-αCGC plasmids. This was done by inverse long PCR mutagenesis (see below).

The pcDNA3-based U1 WT plasmid and both U1-αCGC plasmids were sequenced across the U1 snRNA-encoding sequence.

Mutants of the ß-globin gene

The ß-globin gene was encoded by a plasmid called pßCla which was derived from pß5'SVBglII. The pß5'SVBglII plasmid contains the entire rabbit ß-globin gene located downstream of the SV40 early region, and part of the pBR328 plasmid, including the sequence encoding ampicillin resistance (Grosveld et al, 1982). pßCla was made by mutating three nucleotides in the second intron of the ß-globin gene at positions 357 to 359 from TGT to CGA to create a ClaI restriction enzyme site (Eperon, unpublished). pßCla is shown schematicaly in figure 2.1.

ß-G/TM1 was made by PCR-amplifying sequence of the wild-type TMnm gene from 85 nucleotides upstream of the SK exon through to six nucleotides downstream of the SK exon. The primers used for the amplification (ClaI SKBP and ClaI SK3') each contained a ClaI restriction enzyme site at the 5'-end. Following amplification, the PCR product was digested with ClaI, purified and cloned into the ClaI site of pßCla. Standard cloning methods were used (Sambrook et al, 1989).

ß-G/TM2 was made in the same way as ß-G/TM1 except that sequence of the wild-type TMnm gene from 85 nucleotides upstream of the SK exon through to 20 nucleotides downstream of the SK exon were PCR-amplified. The primers used for the amplification were ClaI SKBP and ClaI SKds.

Both the ß-G/TM1 and ß-G/TM2 plasmids were sequenced across the inserted TMnm sequence.

Site directed mutagenesis

<u>Inverse long PCR mutagenesis</u> was used to generate the TMnm minigene plasmids SK16-30R1, SK16-30R2, SK16-30R3, SK16-30R4, SK16-30R5, SK2-4CGC, and both U1-αCGC plasmids. The method used was an adapatation of the method described by Barnes (1993). For each mutant, two oligonucleotides that are complementary to one another at their 5'-ends were used to prime inverse synthesis of the entire double stranded template plasmid (the wild-type TMnm minigene plasmid or a wild-type U1 snRNA-encoding plasmid). The desired mutation is contained in the overlap region of each primer. Each reaction (100μl) contained the two primers (20pmol each), template DNA (25ng), 1x Barnes buffer pH 8.55, Taq polymerase mixed with Pfu polymerase (160 units to 1 unit). The reactions were subjected to 30 amplification cycles each comprising 94°C, 20secs; 68°C, 30secs; 72°C, 1 minute per kilobase of template DNA. The products of the reaction were then precipitated and digested with DpnI (per 20μl reaction: 1 unit DpnI, 1x DpnI restriction enzyme buffer, 0.5x total precipitated PCR reaction). DpnI is specific to methylated DNA, and therefore only digests the template DNA, not the linear products of the PCR reaction. The DpnI-digested reactions were then precipitated, resuspended in distilled water, and used to electroporate homologous recombination-competent DH5α cells. The electroporated DH5α cells were then selected on ampicillin plates and screened for the presence of mutant plasmids. Once high quality plasmids were made, the sequence of the mutant region of each plasmid was checked.

Electroporation

Homologous recomination-competent DH5α cells were grown in 1 litre of TY medium to an OD A_{600} of 0.55-0.7. The cells were then pelleted and resuspended in 1 litre of distilled water. The cells were re-pelleted and suspended in 0.5 litre distilled water. The cells were pelleted again and resuspended in 20ml 10% glycerol, and then were pelleted for a final time and resuspended in 2ml 10% glycerol. Cells not for immediate use were frozen and stored at -80°C. 1-5μl of DNA (dissolved in distilled water) to be transformed was mixed in an electroporation cuvette on ice with 40μl of the washed, concentrated cells. The cells were then electroporated at 2.48 KV, 25μF, 400Ω using a BioRad electroporator. Immediately after electroporation, the cells were diluted in 1ml TY medium and incubated at 37°C for 1 hour before being plated onto ampicillin plates (50μg/ml).

Preparation of plasmid DNA

For preparation of high quality plasmid DNA, 100ml TY medium containing ampicillin (25µg/ml) was innoculated with the transfected bacterial colony of interest and incubated for ~8 hours at 37°C. After chloramphenicol was added (170µg/ml) the innoculated medium was mixed overnight at 37°C. The cultures were then centrifuged for 10mins at 2,800g and the pelleted cells were resuspended in 4ml GTE (50mM glucose, 25mM Tris-HCl, pH8, 10mM EDTA) and placed on ice. 8ml 0.2M sodium hydroxide, 1%SDS were added, mixed in and the mixture was left on ice. After 5 minutes, 6ml potassium acetate (29.4g, plus 11.5ml glacial acetic acid in 100ml water) was added. The mixture was left on ice for 10 minutes then centrifuged at 2,800g for 15mins at 4°C. The supernatant was filtered through a 0.4µm Acrodisc into another tube and 22ml ethanol was added. The tube was placed on dry ice for 15 minutes, then centrifuged for 15 minutes at 2,800g. The pellet was dried, then resuspended in 2ml TE.1. 2.5ml 4.4M lithium chloride was added and the mixture was left on ice. After ~1hr the mixture was centrifuged as before, then 10ml absolute ethanol was added to the supernatant before a further centrifugation step. The resulting pellet was washed in 70% ethanol, dried and resuspended in 400µl TE.1. 100µg RNaseA was added and left to digest contaminating RNA for 15 minutes at 37°C. The RNAseA was inactivated with 20µl 10% SDS followed by a 70°C incubation for 10 minutes. The plasmid DNA was then cleaned by two standard phenol/chloroform extractions, precipitated in absolute ethanol (2.5 volumes) and 3M sodium acetate (0.1 volumes) and resuspended in TE.1.

Poorer quality plasmid DNA prepared for screening purposes was purified according to the method of Stephen *et al* (1990). A single transfected bacterial colony was innoculated into 3ml TY medium containing ampicillin (25µg/ml) and incubated at 37°C overnight. The overnight culture was then centrifuged at 13000 rpm in a benchtop centrifuge for 1 minute. The pellet was resuspended in 200µl GTE (50mM glucose, 25mM Tris-HCl, pH8, 10mM EDTA) and 400µl 0.2M sodium hydroxide/1% SDS was added. The preparation was then mixed and left on ice for 5 minutes. 300µl 3M potassium acetate (pH 4.8) was added and after mixing, the preparation was left on ice for a further 5 minutes before centrifugation at 13,000 rpm for 5 minutes. 1 volume absolute ethanol was added to the supernatant and the contents of the tube were mixed then centrifuged as before. The pellet was washed in 70% ethanol, dried and resuspended in 40µl TE.1.

Double-stranded DNA sequencing

Sequencing of high quality plasmid DNA was done with sequenase version 2.0 T7 DNA polymerase according to the United States Biochemical Corporation (USB) protocol. Sequencing of poorer quality plasmid DNA was also carried out according to the USB protocol. The poorer quality plasmid DNA was alkali denatured and contaminating RNA hydrolysed by incubation with 2M sodium hydroxide (9µl plasmid preparation, 1µl sodium hydroxide) for 10 minutes at 37°C. Then 10pmol of sequencing primer was added followed by addition of 3µl 3M potassium acetate to neutralise the sodium hydroxide. The DNA was then precipitated with 75µl ethanol, washed in 75% ethanol and resuspended in distilled water and sequenase reaction buffer. After incubation at 37°C for 30 minutes, the labelling reaction and subsequent steps of the USB protocol was followed.

Culture of mammalian cells

Growth of mouse myoblast cells

Mouse C2-C12 cells (ECACC) cells were cultured in 10cm diameter Nunc tissue culture dishes pre-treated with collagen (see below). DMEM culture medium was supplemented with 10% foetal calf serum and 0.1% gentamycin. Cells were grown at 37°C in an atmosphere of 5% CO_2 and high humidity. When above 50% confluence, cells were removed from the surface of the culture dish by treatment with trypsin-EDTA for 3 minutes, before resuspension and dilution in fresh culture medium and application to fresh culture dishes.

Collagen treatment involved application of collagen solution (0.4mg/ml collagen in 150mM glacial acetic acid mixed with 1ml 3.6% sodium chloride immediately before use) to the cell surface of each dish at 4°C overnight. Before use, each dish was rinsed several times with sterile water.

Growth of mouse myotube cells

When myotubes were required, mouse C2-C12 cells (ECACC) were grown to confluence using the same conditions for culture of myoblasts described above. Approximately 24 hours after confluence had been reached, the medium was changed to 5% horse serum in DMEM containing 0.1% gentamycin. The medium was changed every day until differentiation was complete (typically five days).

Growth of COS cells

COS-1 cells were cultured in 10cm diameter Nunc tissue culture dishes. DMEM culture medium was supplemented with 10% newborn calf serum and 0.1% gentamycin. Cells were grown at 37°C in an atmosphere of 5% CO_2 and high humidity. When above 50% confluence, cells were removed from the surface of the culture dish by treatment with trypsin-EDTA for 3 minutes, before

resuspension and dilution in fresh culture medium and application to fresh culture dishes.

Transfection of mammalian cell cultures

Mouse C2-C12 cells at ~75% confluence were transfected using a calcium phosphate-DNA coprecipitation method. 5µg of plasmid DNA was diluted in 450µl TE.1 and 500µl of 2x HBS. The DNA/calcium phosphate precipitate was formed by the addition of 50µl of 2.5M calcium chloride and incubation on ice for 15 minutes. The suspended precipitate was added, dropwise, over the surface of the culture dishes. After incubation at 37°C in 5% carbon dioxide for 5 hours the cells were washed twice with warm 1x PBS. Uptake of DNA was promoted by a 90 second glycerol shock (1ml of 20% glycerol in PBS) followed by two washes with 1x PBS and the addition of fresh culture medium. After 20 hours, total RNA was harvested from the cells as myoblasts, or the cells were induced to differentiate by the addition of differentiation medium (DMEM supplemented with 5% horse serum and 0.1% gentamycin) as described above.

The cotransfection protocol was identical to the protocol above except that 15µg of U1 plasmid DNA together with 5µg of TMnm plasmid DNA were diluted in 450µl TE.1 and 500µl 2x HBS.

COS-1 cells were transfected using the same procedure as the transfection procedure described above, except that 5µg of plasmid was diluted to 60µl in TE.1 and 0.6 ml calcium chloride (0.25M) were added. Precipitation was achieved by the addition of 0.66 ml 2x HBS.

Isolation of total RNA from transfected cell cultures

Total RNA was isolated by one of two lithium chloride/urea extraction procedures:
1) Adapted from Walther et al (1982).
Medium was removed from each 10cm tissue culture dish and the cells were washed with 2ml 0.9% NaCl. After removal of the wash solution and addition of 2ml of a LiCl (3M)/urea (6M) solution each dish was left at room temperature for 5 minutes. Addition of 0.1 volumes of 3M sodium acetate and 2.5 volumes of absolute ethanol was followed by a room temperature spin at 2,800g for 15 minutes. The pellet was resuspended in 400µl TES and then extracted with an equal volume of phenol/chloroform. The aqueous phase was precipitated with 0.1 volumes of 3M sodium acetate and 2.5 volumes of ethanol at -70°C for 15 minutes before a spin at 13,000g for 10 minutes. The pellet was washed in 70% ethanol and resuspended in 40µl TE.1.

2) Adapted from Eperon and Hamshere (1993).

Medium was removed from each 10cm culture dish and the cells were washed with PBS. After removal of the wash solution 2ml of ice cold lithium chloride (3M)/urea (6M) solution was added to each plate and transferred, after repeated pipetting, to a Corex tube. This step was repeated and the samples were left overnight at 4°C. The next day samples were centrifuged at 12,000g for 15 minutes. The pellet was resuspended in 400μl TES then extracted with an equal volume of chloroform/isoamylalcohol (24:1). The aqueous phase was precipitated as in method 1.

Selection of poly (A) RNA from total RNA

Scigen magnetised cellulose oligo dT$_{30}$ beads were used for the selection of poly(A) RNA from total RNA isolated from transfected cells (Scigen protocol, Scigen Ltd, Sittingbourne, UK). For each isolation, 10μl total RNA (diluted to 50μl in TE.1) was used with 50μl beads. The beads were washed twice in 1x binding buffer and mixed for 10 minutes at room temperature with the total RNA that had previously been heated to 65°C for 2 minutes and placed on ice. The beads were sedimented using a magnet and washed twice with 0.5x binding buffer, then once with 0.1x binding buffer. Poly(A) RNA was eluted from the beads by mixing them at 65°C in 10μl distilled water for 10 minutes. This step was repeated with a further 10μl distilled water.

RT-PCR of hTMnm mRNA

Reverse transcription
1/20th volume total RNA sample or 1/5th volume polyA RNA sample plus 10pmol RT-PSVR primer were mixed in 1x PCR buffer in a total volume of 10μl. Each sample was overlaid with mineral oil then heated to 95°C for 5 minutes and placed on ice. 5 units each of Murine-MLV reverse transcriptase and RNasin in a total volume of 5μl 1x PCR buffer were then added to each tube. All samples were incubated at 37°C for 1 hour followed by 5 minutes at 94°C.
First round PCR
10pmol CAT 5015 primer plus 2.5 units of *Taq* polymerase (RedHot) were mixed in 1x PCR buffer in a total volume of 10μl and added to each completed reverse transcription reaction. Each sample was then heated to 94°C for two minutes followed by 30 cycles of: 94°C, 1 minute; 55°C, 30 seconds; 72°C, 1 minute.
Second round (nested PCR)
1/25th volume of the first round PCR sample, 10pmol exon IV primer (labelled or cold, see below), 10pmol exon VI/VII primer (labelled or cold, see below) and 2 units of *Taq* polymerase were mixed in 1x PCR buffer in a total volume of 20μl.

2.14

Each sample was overlaid with mineral oil then heated to 94°C for two minutes followed by 20 cycles of: 94°C, 1 minute; 55°C, 30 seconds; 72°C, 1 minute for amplification of first round products generated from myoblast/myotube RNA, or by 8 cycles of: 94°C, 1 minute; 55°C, 30 seconds; 72°C, 1 minute for amplification of first round products generated from COS RNA.

Radioactive labelling of PCR primer

50µCi γ^{32}P-ATP (10µCi/µl), 5 units of T4 PNK, and 500pmol exon IV or exon VI/VII primer were mixed in a total volume of 50µl 1x kinase buffer and incubated at 37°C for 30 minutes. The choice of primer for kinase labelling depended on the restriction enzyme analysis to be carried out on nested PCR products. Digestion with MboII required that exon IV primer was labelled, digestion with AluI required that exon VI/VII primer was labelled.

Restriction enzyme digestion of nested PCR products

0.5µl nested PCR product and 0.25 units MboII or AluI were incubated in 1x restriction enzyme digestion buffer in a total volume of 5µl for 2-4 hours. Reactions were then stopped by addition of an equal volume of formamide dyes. Digested and undigested samples (both in formamide dyes) were loaded onto a denaturing polyacrylamide gel (6-8%) containing 20% formamide.

RT-PCR of ß-globin, ß-G/TM1, and ß-G/TM2 mRNA

A similar protocol to that for RT-PCR of human TMnm mRNA was followed except that different primers were used:

Reverse transcription primer: ß-globin 3'

First round PCR primer: PCR exon 1

Second round PCR primers: PCR 1 exon 3, PCR 2 exon 1.

The products of nested PCR were radioactively labelled but were not digested with a restriction enzyme as the amplified products derived from different mRNA isoforms were of different lengths.

Generation of transcription templates by PCR

DNA templates for transcription were amplified from the plasmid containing the desired sequence by PCR. For each template the following reaction was set up:

     1x PCR buffer
     10pmol 5'-primer
     10pmol 3'-primer
     1-10ng plasmid DNA
     1-5 units Taq polymerase

The reaction was then overlaid with mineral oil and heated to 94°C for two minutes before undergoing 30 cycles of amplification in a PCR machine:

94°C, 1 minutes;

55°C, 30 seconds;

72°C, 1 minute.

A portion of the products of the reaction were checked by separation on an agarose gel (0.8-2% agarose, depending on the length of the product) containing ethidium bromide (1-10µg/ml). The transcription template was then either purified away from the other reaction components using an appropriate Clontech Chroma Spin Column or was purified by electroeluting the DNA after electrophoresis on an agarose gel.

The PCR primers used to generate transcription templates are listed below:

NM-SKbpup pre-mRNA	T7NM, SK3'
SK exon RNA	T7SK, SK3'
NM exon RNA	T7NM, NM3'
85 nucleotide RNA containing the SK exon	T7SKus, SK3'

Gel purification of PCR generated transcription templates

Approximately half of the PCR reaction was loaded on an appropriate percentage agarose gel containing ethidium bromide (1-10µg/ml) and electrophoresed in TAE or TBE electrophoresis buffer. After electrophoresis, the template DNA was detected by exposing the gel to UV light and was excised from the gel. The resulting gel slice was placed in a section of dialysis tubing containing 0.4ml electrophoresis buffer. The tubing containing the gel slice was placed in the electrophoresis buffer used for the previous gel electrophoresis. The current was turned on once more until the DNA had migrated out of the gel slice into the electrophoresis buffer within the tubing. This buffer was then removed and added to 1ml absolute ethanol. The DNA was pelleted by centrifugation for 10 minutes at 13,000 rpm in a benchtop microcentrifuge. The pellet was washed in 70% ethanol and resuspended in TE.1.

In vitro transcription of PCR generated templates

For a typical transcription to generate radioactively labelled RNA, the following were mixed, then incubated at 37°C:

> 2μl 5XTMS
>
> 2μl low G NTPs
>
> 0.5μl 0.1M DTT
>
> 1μl RNasin
>
> 0.25μl T7 RNA pol
>
> 1μl α^{32}P GTP (10μCi/μl)
>
> 1/50 PCR reaction product
>
> 1μl 10mM Cap ()
>
> Water to 10μl

After approximately 3 hours, an equal volume of formamide dyes were added to stop the reaction. The transcript was then purified after electrophoresis on a denaturing polyacrylamide gel containing 20% formamide. The transcripts used in the *in vitro* splicing experiment described in figure 5.1 were labelled with α^{32}P UTP (10μCi/μl). The NTP mix was adjusted to low U NTPs accordingly.

Gel purification of radioactively labelled RNAs

The radioactively labelled transcript was first visualised on the denaturing polyacrylamide gel by autoradiography (1 to 5 minute exposure). The band corresponding to the transcript was then excised from the gel and eluted overnight at 4°C in 0.4ml RNA elution buffer. The next day, the RNA elution buffer containing the eluted transcript was transferred to a 1.5ml eppendorf tube containing 1ml absolute ethanol, incubated at -70°C for 10 minutes, then centrifuged at 13000rpm in a benchtop microcentrifuge for 10 minutes. The pellet was washed with 70% ethanol, dried and resuspended in upto 10μl TE.1. The purified transcript was then stored at -70°C.

Preparation of whole cell extracts

Whole cell extracts were prepared by a modification of the Jiang and Eberhardt method (1995). C2-C12 myoblasts or myotubes grown in 750ml culture flasks were washed twice in 1x PBS and were then harvested in 1.5ml 1x PBS by scraping using a rubber policeman. After a low speed spin for one minute using a benchtop microfuge at 4°C, the supernatant was removed and the pellet was resuspended in three packed cell volumes of extraction buffer by vortexing. Cell lysis was achieved by freezing on dry ice and thawing at 37°C in a water bath. After a high speed spin for 30 seconds using a benchtop microfuge the

supernatant was removed and diluted with four volumes of dilution buffer. The extracts were then aliquoted and stored at -70°C.

Preparation of HeLa cell splicing extracts (based on the Dignam method)

HeLa cells were grown in suspension in DMEM or similar medium, containing 10% newborn calf serum and antibiotics (1% penicillin-streptamycin), to a density of approximately $5x10^5$-$1x10^6$ cells per ml. $2x10^9$ cells were used for each extract preparation. All the subsequent extract preparation steps were carried out at 4°C. First, the cells were centrifuged at 75-300g for 10 minutes and the pellet was resuspended in 80ml 1x PBS. The cells were then centrifuged at 450g for 5 minutes, resuspended in 30ml buffer A, centrifuged again at 450g for 5 minutes and resuspended in 14ml buffer A. The resuspended cells were homogenised in a Dounce homogeniser (tight fitting pestle). The nuclei were collected by centrifuging the homogenate at 4300g for 15 minutes. After the supernatant was removed, the tube was respun at 26000g for 15 minutes. The nuclear pellet was then resuspended in a total of 4ml buffer C and the nuclei were homogenised using the Dounce homogeniser. The nuclear homogenate was mixed gently for 30 minutes then centrifgued for 30 minutes at 26000g. The supernatant was transferred to dialysis tubing (prepared in advance by autoclaving in water) and dialysed for several hours against 2-3 changes of buffer D (1 litre in total). Finally, the extract was centrifuged at 26000g and the supernatant was aliquoted and stored at -70°C.

Splicing reactions

For the experiment described in figure 5.1, the following components were mixed on ice:

HeLa splicing extract	125µl
RNAse inhibitor	300 units
80mM magnesium chloride	12µl
0.5M phosphocreatine	12µl
0.1M ATP	5µl
13% polyvinyl alcohol	60µl
Buffer D	to 295µl

Approximately 3 fmol of each radioactively labelled RNA were mixed with 60µl of the above reaction mixture on ice. The mixtures were then aliquoted into 5x 10µl. Two aliquots of each reaction were incubated for 3 hours at 30°C, two were kept at 4°C for 1.5 hours and were then placed at 30°C for 1.5 hours, and one was kept on ice for 3 hours. After the incubations were complete, 50µl proteinase K mixture was added to each aliquot before incubation at 37°C for 5 minutes. 150µl

absolute ethanol were then added and the samples were centrifuged at 2800g for 10 minutes. The pellets were washed in 200μl 70% ethanol then centrifuged at 2800g for 5 minutes. The supernatants were removed, then the pellets were air dried and resuspended in 10μl formamide dyes. To analyse the splicing patterns, each sample was electrophoresed on a 6% denaturing polyacrylamide gel containing 20% formamide.

For the experiment described in figure 5.5, radioactively labelled adenovirus E1a 2,1 pre-mRNA (described in Eperon et al, 1993) was added to a 70μl splicing reaction. This pre-mix was then split into five equal sized aliquots. One aliquot was kept on ice for one hour. To the other aliquots 0, 1.5, 3, or 7.5μg SR proteins (purified from HeLa cells, supplied by A.Krainer) were added and then these aliquots were incubated for 1 hour at 30°C. After the incubations were complete, the samples were processed and analysed as above.

Gel retardation assays

For the experiment described in figure 5.2, ~0.5 fmol radioactively labelled NM-SKWTbpup or NM-SK16-30R1bpup pre-mRNA was added to a 10μl splicing reaction in the presence or absence of 5μg yeast tRNA and incubated at 30°C for 30 minutes. Native gel loading dye was added to each of the reactions before they were run on a 0.8% agarose gel. The gel was run overnight at 4°C, then fixed, dried and exposed to a PhosphorImager cassette before analysis using the PhosphorImager.

For the experiment described in figure 5.7, 5pmol radioactively labelled SKWT (85) or SK16-30R3 (85) RNA was added to 5μl reactions containing 6μg SR proteins, 5μg myoblast or myotube whole cell extract or 5μg HeLa cell splicing extract, or a combination of these (see figure 5.7 for combinations), magnesium chloride (3.5mM), and D buffer. The samples were incubated at 30°C for 30 minutes, then 0.25μg heparin was added. After 5 minutes, the samples were placed on ice, added to native gel loading dye and run on a 0.5% agarose/3% acrylamide gel at 4°C overnight. The gel was then fixed, dried and exposed to a PhosphorImager cassette before analysis using the PhosphorImager.

Shortwave UV crosslinking

In the experiment described in figure 5.4, ~0.5fmol radioactively labelled NM-SKWTbpup or NM-SK16-30R1bpup RNA was incubated at 30°C for 30 minutes with 3x 10μl splicing reactions containing one of three different HeLa splicing extracts, in the absence of ATP, creatine phosphate and polyvinyl alcohol. The samples were then placed on ice and irradiated for 30 seconds with shortwave UV light using a Spot-cure (UVP). 5μl of RNAse mix was then added to each

incubation and the mixtures were placed at 37°C for 15 minutes. The RNAse digestions were stopped by addition of an equal volume of SDS-polyacrylamide gel loading buffer to each sample. The products of crosslinking were separated on a 10% SDS polyacrylamide gel and the proteins in the gel were electroblotted onto a nitrocellulose filter. The filter was then dried and exposed to a PhosphorImager cassette before analysis by the PhosphorImager.

In the experiment described in figure 5.6A, radioactively labelled wild-type SK exon RNA was incubated with SR proteins (purified from HeLa cells, supplied by A.Krainer) at 30°C for 30 minutes under splicing conditions in the absence of ATP, creatine phosphate and polyvinyl alcohol. The reaction was then split into seven equal sized aliquots, each containing ~0.25pmol labelled RNA and ~0.2μg SR proeins. 0.25pmol or 2.5pmol unlabelled competitor RNA (wild-type SK exon RNA, SK16-30R3 exon RNA, or wild-type NM exon RNA) was then added to each aliquot. No competitor was added to one control aliquot. After 5 minutes at 30°C, each aliquot was placed on ice and irradiated with shortwave UV light for 30 seconds. The samples were then processed and analysed as above.

For the experiment described in figure 5.9, 5pmol radioactively labelled SKWT (85) or SK16-30R3 (85) RNA was added to 5μl reactions containing 6μg SR proteins, 5μg myoblast or myotube whole cell extract or 5μg HeLa cell splicing extract, or a combination of these (see figure 5.9 for combinations), magnesium chloride (3.5mM), and D buffer. The samples were incubated at 30°C for 30 minutes, then 0.25μg heparin was added. After 5 minutes, the samples were placed on ice and irradiated, processed and analysed as before.

Electroblotting onto nitrocellulose filter

Proteins crosslinked to RNA that had been separated on SDS polyacrylamide gels were electroblotted onto Hybond-C Super nitrocellulose in blot buffer at 5mA/cm² for 40 minutes. The Hybond-C Super nitrocellulose was separated from the positive electrode, and the SDS polyacrylamide gel was separated from the negative electrode of the electroblotter by nine sheets of 3MM paper.

Affinity Purification (Adapted from Reed, 1990)

200pmol each biotinylated RNA were added to separate 5x splicing reactions (see above), incubated at 30°C for 30 minutes, then placed on ice. The reactions were diluted to 20ml with ice cold FSP buffer, 200μl avidin-agarose beads were added and were then mixed overnight at 4°C. After the overnight step the avidin-agarose beads were pelleted and most of the supernatant was removed. The pelleted avidin-agarose beads were resuspended in the remaining supernatant and separated from the supernatant using a BioRad disposable spin column. The

beads were washed in 5ml FSP buffer containing 100mM sodium chloride. Components bound to the beads were eluted several times in 25-50µl SDS-polyacrylamide gel loading buffer. Components of the purification were separated on a 12% SDS-polyacrylamide gel. The gel was then stained using the silver staining method (Sambrook *et al*, 1989) to show the proteins and nucleic acids that had been purified.

Preparation of biotinylated RNA

Standard transcription reaction conditions containing 2.5mM each NTP (the UTP contained 15% biotin-11-UTP) were used to transcribe RNA from a gel-purified PCR generated template.

Radioactive markers

Radioactively labelled markers for use on denaturing polyacrylamide gels were made according to the following protocol:

    2µl pBR322 (0.5µg/µl)

    2µl 10x kinase buffer

    0.2µl HpaII

    15.8µl water

The components were mixed then incubated at 37°C for 30 minutes, then heated to 95°C for 5 minutes. Then:

    1µl 10x kinase buffer

    1µl ^{32}P dCTP (10µCi/µl)

    1µl 10mM dGTP

    0.5µl T7 DNA polymerase, Klenow fragment (5 units/µl)

    6.5µl water

were added before incubation at 37°C for 30 minutes. 200µl formamide dyes were added to stop the reactions. Typically, 1µl was loaded on each gel.

Gel electrophoresis

Denaturing polyacrylamide gels were made up as 38% (w/v) acrylamide, 2% (w/v) N,N'-methylene-bis-acrylamide in water. An X% gel was made up by mixing 16.8g urea, Xml of the above acrylamide mix, 4ml 10x TBE, 8ml formamide, then adding water to 40ml. Samples were denatured in the presence of an equal volume of formamide dyes by incubation at 80°C for 8 minutes. Gels were 6, 7, or 8% and were run in 1x TBE electrophoresis buffer, typically at a constant voltage of 1100-1500V for 2-4 hours.

The agarose/acrylamide gel used in the experiment described in figure 5.7 contained 0.5% agarose, 3% acrylamide, 4.25% glycerol, 0.5x TBE and was run at 4°C overnight in 0.5x TBE.

SDS polyacrylamide gel electrophoresis and agarose gel electrophoresis was as standard (Sambrook *et al*, 1989).

Analysis using the PhosphorImager

Gels or nitrocellulose filters containing radioactively labelled RNA for analysis by the PhosphorImager were exposed to a PhosphorImager cassette overnight. The next day, the cassette was scanned and the image was analysed using Molecular Dynamics software.

Chapter 3
Role of *cis*-acting sequences in the regulation of human TMnm pre-mRNA splicing

Introduction

This study involves the human TMnm gene. The gene is over 42Kb long and contains 13 exons. It gives rise to two different protein isoforms which are expressed in a tissue-specific manner. The non muscle isoform, TM30, is 240 amino acids in length and the skeletal muscle isoform, skαTM.2, is 285 amino acids long. Five exons are common to both isoforms. The differences between the two isoforms arise from the use of two alternative promoters and from three regions of mutually exclusive alternative splicing (Clayton *et al*, 1988) (figure 3.1A). In particular, this study is concerned with the tissue specific and mutually exclusive use of exons NM and SK. and the mechanisms involved. Three questions arise:

    1) What determines the splicing pattern in non muscle cells?

    2) What determines the splicig pattern in muscle cells?

    3) Why are exons NM and SK not spliced together?

As discussed above, Graham *et al* (1992) and Hamshere (PhD thesis) identified *cis*-acting sequences involved in the tissue specific regulation of TMnm pre-mRNA splicing. The experiments described in the present study were aimed at further investigating these sequences and trying to elucidate some of the mechanisms by which they influence the TMnm splicing pattern in non muscle and muscle cells.

In their earlier work, Graham *et al* and Hamshere, used a TMnm minigene in which the only intron sequences present were those separating and flanking exons NM and SK. Four mRNA isoforms can potentially be expressed from this minigene. These are shown in figure 3.1B. The TMnm minigene was transfected into non muscle cells (COS-1 cells and mouse myoblast cells) and shown to express only the non muscle isoform (NM isoform) and an isoform in which both the NM and SK exons were skipped (skipped isoform). In skeletal muscle cells (mouse myotube cells), expression of the SK isoform was dominant to expression of the NM isoform. The skipped isoform was also observed (Graham et al, 1992; Hamshere, PhD thesis). Thus, all the pre-mRNA sequence required for regulated expression of the muscle and non muscle specific mRNA isoforms is encoded by the minigene.

Expression of the SK isoform was observed in non muscle cells from a mutant TMnm minigene in which the branch point sequence upstream of exon SK (the SK BPS) was altered to the consensus branch point sequence (this is referred to as the branch point upregulated, bpup, mutation below). Other mutant minigenes which contained the bpup mutation were tested which also contained mutations of the SK exon sequence itself. Two of these were shown to express
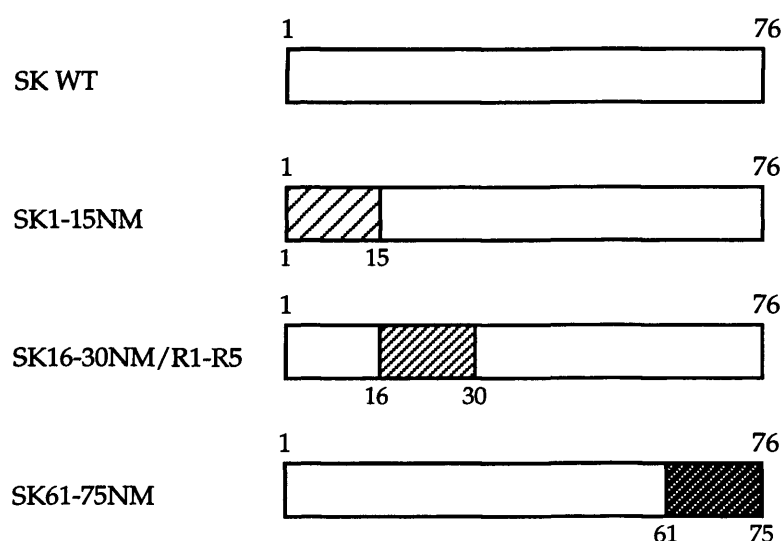
C) Mutations of the SK exon

The position of mutations of the SK exon is shown schematically. The wild-type SK exon (SK WT) is represented as an open box. Mutated regions are shown by the cross-hatching. In each case, the numbers above the exon represent the first nucleotide (1) and the last nucleotide (76), respectively of the wild-type or mutant exon. The numbers below the exon represent the nucleotides that were mutated. In the SK1-15NM mutant, the first fifteen nucleotides of the wild-type exon were substituted with nucleotides corresponding in sequence to the first fifteen nucleotides of the NM exon. In the SK16-30NM mutant, nucleotides 16-30 of the wild-type exon were substituted with nucleotides corresponding in sequence with nucleotides 16-30 of the NM exon. The sequence of mutants SK16-30R1, SK16-30R2, SK16-30R3, SK16-30R4, SK16-30R5 (R1-R5 in the figure) are given in table 3.2 below. In the SK61-75NM mutant, nucleotides 61-75 of the wild-type exon were substituted with nucleotides corresponding in sequence to nucleotides 61-75 of the NM exon (the NM exon is also 76 nucleotides long). The name given to each mutant in the figure refers only to the sequence of the SK exon. See text for details of whether the branch point sequence upstream of the mutant SK exon was the wild-type sequence or the consensus sequence.

Figure 3.1 The human TMnm gene

i)

ii)

TM30

I_{nm} II   III   IV   NM   VI   VII $VIII_{nm}$ 3'-UTR

skαTM.2

3'-UTR

I_{sk} II_{sk} II   III   IV   SK   VI   VII $VIII_{sk}$ IX_{sk}
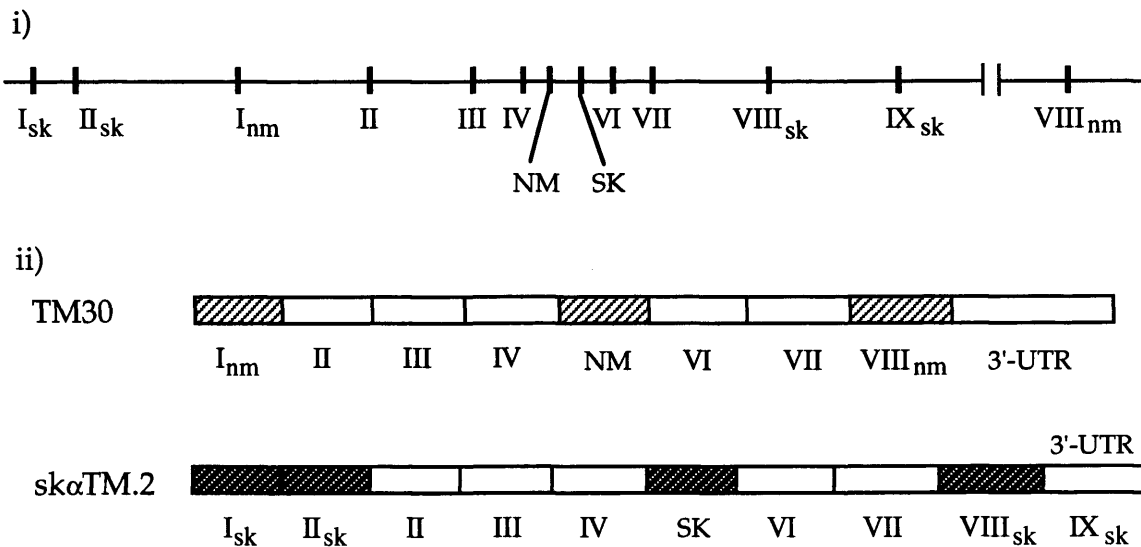
A) The human TMnm gene and the mRNA isoforms expressed from it.

i) Structure of the TMnm gene drawn approximately to scale. Intron sequence is shown as a horizontal thin line. The exons are shown as vertical lines. The IXsk-VIIInm intron is shown with a double vertical line because it is much longer than the other introns.

ii) TMnm non muscle (TM30) and skeletal muscle (skαTM.2) mRNA isoforms. Exons unique to each isoform are shaded. The 3' untranslated sequences are indicated as 3'-UTR.

i)

ii)

NM/SK isoform

NM isoform

SK isoform

Skipped isoform
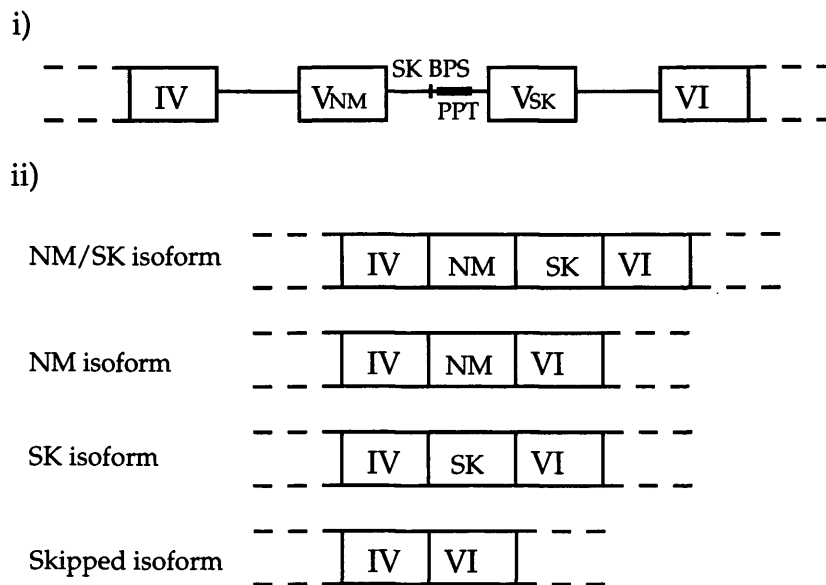
B) The TMnm minigene and potential mRNA isoforms

i) The central region of the TMnm minigene. Exon sequences shown as open boxes, intron sequence as thin lines. The only intron sequences present in the minigene are those separating and flanking the NM and SK exons. The branch point sequence upstream of exon SK (SK BPS) is shown as a short vertical line, and the polypyrimidine tract (PPT) is shown as a thick horizontal line. The dashed sequences represent exon sequence either side of exons IV and VI (Isk, IIsk, II, III, VII, VIIIsk, IXsk). For full details of the minigene, see chapter 2 and appendix 1.

ii) The four possible mRNA isoforms which can be expressed from the minigene are shown.

significantly higher levels of the SK isoform than the minigene containing the bpup mutation alone (Graham *et al*, 1992; Hamshere, PhD thesis). In one mutant, the first fifteen nucleotides of the SK exon were replaced with fifteen nucleotides identical in sequence to the first fifteen nucleotides of the NM exon. In the other mutant, SK exon nucleotides 61 to 75 were replaced with fifteen nucleotides identical in sequence to NM exon nucleotides 61 to 75 (see Figure 3.1C). Since both the NM and the SK exons are 76 nucleotides in length, this mutation involved substitution of 15 nucleotides at the 3'-end of exon SK with 15 nucleotides identical in sequence to the 3'-end of exon NM. It was concluded that the SK isoform is not expressed from the wild-type minigene in non muscle cells because the SK BPS is suboptimal and because sequence located at the 5'-, and the 3'-end of exon SK is inhibitory to its expression.

Repression of exon SK in non muscle cells

In the studies described above, no minigenes containing mutant SK exon sequence were tested that also retained a wild-type SK BPS. Thus, it was not known if the increase in the relative level of SK isoform expression observed when SK exon sequence was mutated was dependent on the presence of a consensus SK BPS. Two TMnm minigenes containing the mutations within exon SK that had been shown to enhance expression of the SK isoform were made that retained the wild-type SK BPS (K.Ayres). The minigene in which the first fifteen nucleotides of the SK exon were substituted with the corresponding sequence from the NM exon is referred to as "SK1-15NM" below. Similarly, the other mutant in which nucleotides 61 to 75 of the SK exon were substituted with the corresponding sequence of the NM exon is referred to as "SK61-75NM" below.

The SK1-15NM and SK61-75NM mutant minigenes were expressed in non muscle cells (mouse myoblast cells) and the relative level of each TMnm mRNA isoform expressed from these mutants was determined by reverse transcription-polymerase chain reaction (RT-PCR)/restriction enzyme analysis of RNA prepared from the cells. Details of this procedure are given in chapter 2. It involves amplification of mRNA species expressed from the TMnm minigenes followed by restriction enzyme digestion of the amplified products to determine the mRNA isoforms that they originated from. Appendix 1 explains the procedure and describes a control experiment showing that the relative levels of TMnm mRNA isoforms are not distorted during amplification. Figure 3.2 shows the PhosphorImager scan of the gel obtained after radioactively labelled nested PCR products (before and after digestion with MboII) had been separated by denaturing polyacrylamide gel electrophoresis. The relative levels of each mRNA isoform expressed were calculated using the PhosphorImager and are shown in table 3.1. The complete data and details of the calculations are given in appendix 2.

Figure 3.2. Effect of mutation of SK exon sequence on TMnm mRNA isoform expression in myoblasts.

Total RNA isolated after transfection of myoblast cells with the TMnm wild-type (SK WT), SK1-15NM, and SK61-75NM minigenes was analysed by RT-PCR. A proportion of the resulting radioactively labelled nested PCR products were then digested with the restriction enzyme MboII. Samples before and after digestion with MboII were electrophoresed on a denaturing polyacrylamide gel. The gel was then fixed, dried, and analysed using the PhosphorImager. The figure shows the three PhosphorImager scans that were obtained. Bands resulting from amplification of particular TMnm mRNA isoforms are labelled on either side of each scan.
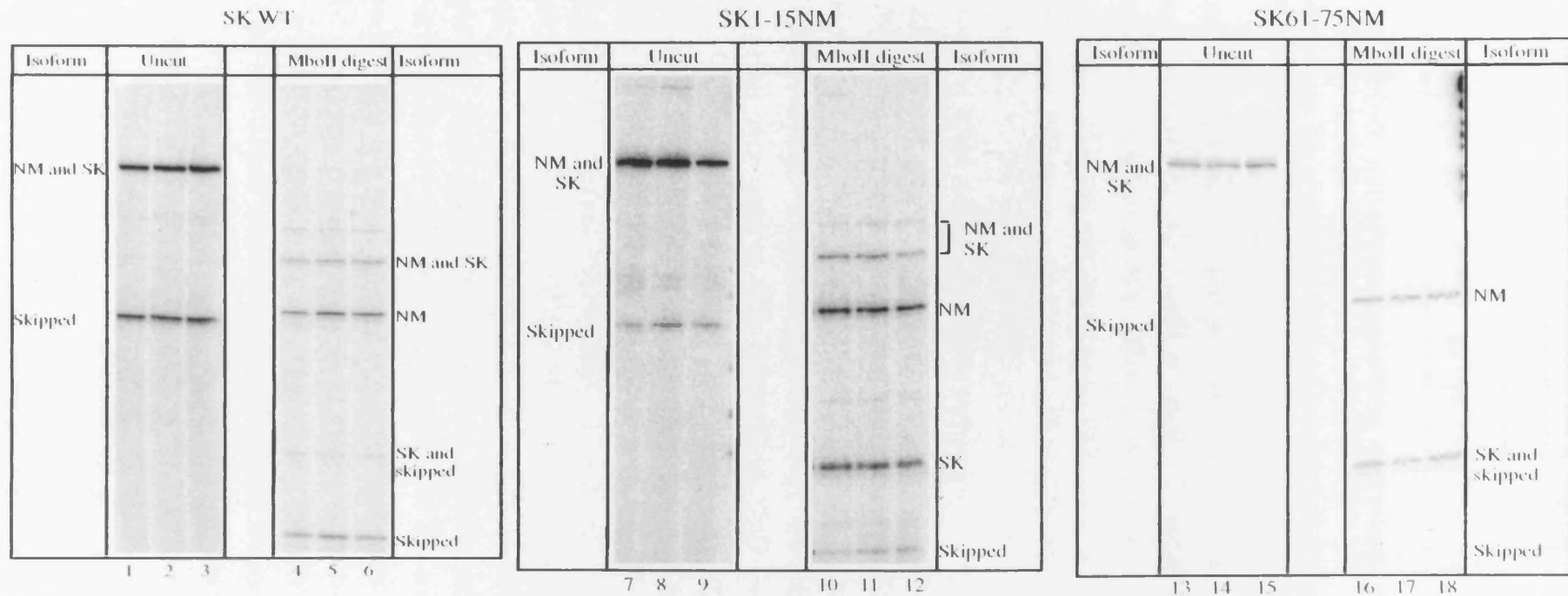
TMnm plasmid	Isoform			
	NM/SK	NM	SK	Skipped
SKWT	0 (0)	59 (2)	0 (0)	41 (2)
SK1-15NM	1 (1)	48 (3)	38 (2)	13 (3)
SK61-75NM	0 (0)	47 (1)	48 (2)	5 (1)

Table 3.1 Expression of TMnm minigenes containing mutations within exon SK in non muscle cells (mouse myoblasts).
Values are shown in percent and are averages of triplicate transfections. Standard deviations are shown in brackets. SKWT refers to the wild-type TMnm minigene.

Table 3.1 shows that SK isoform expression was activated in non muscle cells when 15 nucleotides at the 5'-, or the 3'-end of exon SK were substituted with the corresponding sequence from exon NM. When nucleotides 1-15 of exon SK were substituted (SK1-15NM), SK isoform expression increased to 38 (2)% and when nucleotides 61-75 of exon SK were substituted (SK61-75NM), SK isoform expression increased to 48 (2)%. In comparison, results from earlier work demonstrated that the relative level of SK isoform expression from a TMnm minigene in which the SK branch point sequence had been mutated to the consensus sequence (the bpup TMnm minigene) was 58%. When SK1-15NM or SK61-75NM exon mutations were also made in the bpup minigene (the SK1-15NMbpup, and the SK61-75NMbpup minigenes, respectively), the relative level of SK isoform expression further increased to 97% (Hamshere, PhD thesis). Thus, mutation of sequence at the 5'-, or 3'-end of exon SK caused the relative level of SK isoform expression to increase by 39% over that from the bpup minigene. This increase approximates to the relative SK isoform expression level observed from the SK1-15NM and SK61-75NM mutant minigenes. It was concluded that substitution of sequence at the 5' and 3' ends of the SK exon with sequence from the corresponding regions of the NM exon significantly increased the relative level of SK isoform expression from the TMnm minigene in non muscle cells, and that this effect did not depend on whether the sequence of the branch point sequence upstream of exon SK was the wild-type or the consensus sequence.

It was noted that the effect on the relative level of SK isoform expression of substitution of sequence within exon SK with sequence from the corresponding region of exon NM could be explained in two ways; either the deleted SK exon nucleotides contain sequence which is inhibitory to SK isoform expression, or the inserted NM exon nucleotides contain sequence which is activatory to SK isoform expression.

SK exon use in non muscle cells

Previous experiments in non muscle cells demonstrated that SK isoform expression observed from the bpup TMnm minigene was lost if exon SK nucleotides 16 to 30 were replaced by sequence identical to exon NM nucleotides 16 to 30 (Graham *et al*, 1992. This mutant is referred to as SK16-30NMbpup below, see table 3.2). It was concluded that SK exon nucleotides 16-30 were required for expression of the exon once it had been activated in non muscle cells by mutation of the SK BPS to the consensus sequence. However, it is possible that this effect was due, not to a loss of inherently activatory SK exon sequences, but instead, to the introduction of inhibitory NM exon sequence. To rule out this possibility, the SK exon sequence at nucleotides 16-30 was replaced by a sequence different to that of NM exon nucleotides 16-30 (see figure 3.1C). The minigene containing this mutation is referred to as SK16-30R1bpup below and the sequence of the mutant SK exon nucleotides at positions 16-30 is given in table 3.2 (the SK16-30R1 sequence). If the wild-type SK exon contains inherently activatory sequence at nucleotides 16-30 it was predicted that the SK16-30R1bpup minigene would not express the SK isoform in non muscle cells. However, if NM exon nucleotides 16-30 are inhibitory to expression of the SK isoform when located within the SK exon (as in the SK16-30NMbpup minigene), it was predicted that the SK16-30R1bpup minigene would express the SK isoform in non muscle cells.

SK exon mutant	SK BPS	SK 16-34 sequence						Pu
SKWT	WT or C	GGA	GGA	GGA	GCT	GAA	GAA T	16
SK16-30NM	WT or C	GGA	TGA	GCA	GAT	TAG	GAA T	14
SK16-30R1	C	TCG	CGC	TAC	ACG	CTT	GAA T	8
SK16-30R2	W T	ATT	TAT	AAG	CTC	CAA	GAA T	10
SK16-30R3	W T	CAA	ACC	CAG	AAA	GAA	GAA T	14
SK16-30R4	W T	TGT	GTC	ATT	TTA	TGA	GAA T	9
SK16-30R5	W T	CCC	TTT	CTC	ACC	AGC	GAA T	6

Table 3.2 Sequence of mutations at SK exon nucleotides 16-34.

The name given to each mutant in the table refers only to the sequence of the SK exon. Mutant TMnm minigene plasmids containing these sequences are referred to below by the same name as that given in the table if the SK branch popint sequence (SK BPS) of the minigene is the wild-type sequence. If the SK BPS is the consensus sequence then the name of the mutant in the table is followed by "bpup". For example, two plasmids containing the SK16-30NM mutation were made. The mutant containing a wild-type SK BPS is referred to as SK16-30NM, and the mutant containing the consensus SK BPS is referred to as SK16-30NMbpup. WT and C in the SK BPS column denote the sequence of the SK BPS in the minigene plasmid (WT represents a wild-type SK BPS and C represents a consensus SK BPS). WT or C shows that two plasmids were made with the same mutant SK exon sequence. One containing the wild-type SK BPS, the other containing the consensus SK BPS. The number of purine nucleotides for each mutant sequence between SK nucleotides 16 and 34 is given in the Pu column.

The SK16-30R1bpup mutant minigene was transfected into non muscle cells (COS cells) at the same time as the bpup and SK16-30NMbpup mutant TMnm minigenes. Expression of the relative levels of TMnm mRNA isoforms was determined, again by RT-PCR/restriction enzyme analysis of RNA purified from the cells after transfection. Figure 3.3 shows the PhosphorImager scan obtained. Products derived from the SK mRNA isoform are visible in the lanes representing expression from the bpup minigene (lanes 2, 4, and 6). However, these products are not visible in the lanes representing expression from the SK16-30R1bpup and SK16-30NMbpup minigenes (lanes 8, 10, 12, 15, and 17). Thus, neither the SK16-30R1bpup or the SK16-30NMbpup minigene was able to express the SK isoform. It was concluded that wild-type SK exon sequence at nucleotides 16-30 is required for expression of the SK exon in non muscle cells.

SK exon use in muscle cells

The requirement of SK exon nucleotides 16-30 for expression of the SK isoform in non muscle cells suggested that these nucleotides might also be required for expression of that exon in muscle cells, where SK isoform expression is dominant to NM isoform expression (Hamshere, PhD thesis). To examine this possibility, the sequence at SK nucleotides 16-30 was replaced with four different mutant sequences. If the wild-type sequence is required for expression of the SK exon in muscle cells it was predicted that these mutants would all demonstrate a reduced ability to express the SK isoform compared to the wild-type minigene. The sequence of each mutant replacing the wild-type SK exon sequence at nucleotides 16-30 is shown in table 3.2. The TMnm minigenes containing these mutations are referred to as SK16-30R2, SK16-30R3, SK16-30R4, and SK16-30R5 below. The wild-type SK BPS was retained in these mutant minigenes because the relative level of SK isoform expression from the wild-type TMnm minigene in muscle cells is high and, therefore, does not need to be boosted by conversion of the SK BPS to the consensus sequence in order to see an effect on reduction of SK isoform expression.

The mutant minigenes were transfected into mouse myoblast cells at the same time as the SK16-30NM (containing a wild-type SK BPS) and wild-type TMnm (SK WT) minigenes. The cells were caused to differentiate into myotubes (i.e. muscle cells), then the relative levels of TMnm mRNA isoforms expressed from the minigenes were determined by RT-PCR/restriction enzyme analysis of RNA purified from the myotubes. The products resulting from the analysis are shown in figure 3.4A.

The products amplified from the NM and SK isoforms are identical in size and are digested with the restriction enzyme MboII to enable their relative levels of expression to be calculated. However, the MboII restriction enzyme site critical for distinguishing nested PCR products amplified from NM and SK isoform

Figure 3.3. Effect of mutation of the purine-rich sequence within exon SK on TMnm mRNA isoform expression in COS cells

COS cells were transfected with three mutant TMnm minigenes. After transfection total RNA was isolated and TMnm mRNA was amplified by RT-PCR. A proportion of the labelled nested PCR products were were then digested with the restriction enzyme AluI. Samples before and after digestion with AluI were electrophoresed on a denturing polyacrylamide gel. The gel was then fixed, dried and analysed using the PhosphorImager. The figure shows the PhosphorImager scan of the dried gel. Lanes 1, 3, 5, 7, 9, 11, 14, 16 are undigested samples. Lanes 2, 4, 6, 8, 10, 12, 15, 17 are digested samples. Radioactive markers were loaded in lane 13. Products representing different TMnm mRNA isoforms are labelled.

The TMnm minigene mutants tested all contained a mutation of the branch point sequence (BPS) located upstream of exon SK. The mutation converted the wild-type BPS to the consensus BPS. This mutation is known as branch point upregulated (bpup). Two of the mutants also contained a mutation of the SK exon sequence at nucleotides 16-30 (see text for details).
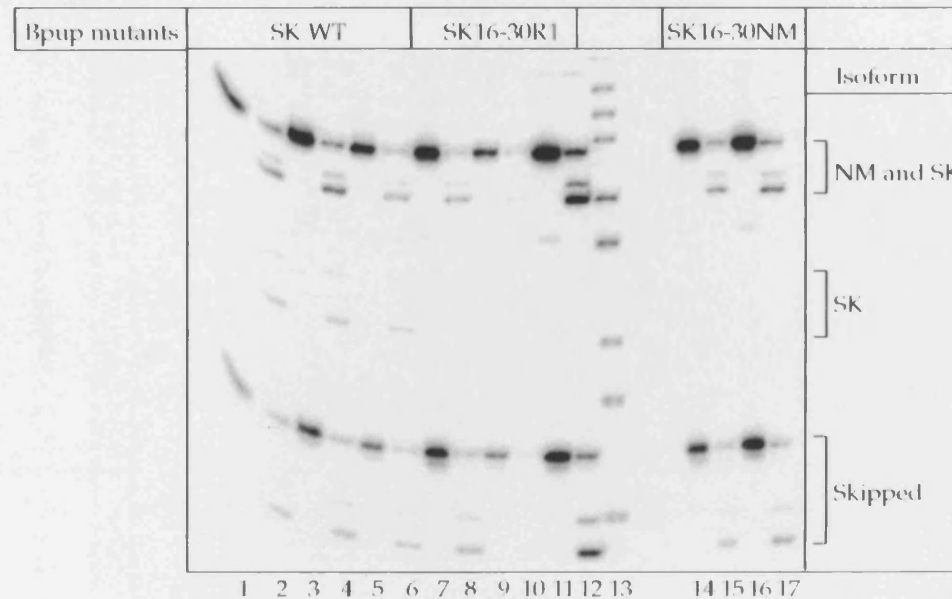
Figure 3.4 Effect of mutation of the purine-rich sequence within exon SK on TMnm mRNA isoform expression in myotube cells

Mouse myoblast cells were transfected with the wild-type TMnm minigene and five other TMnm minigenes containing mutations of the purine-rich sequence in exon SK. The cells were caused to differentiate into myotubes, then RNA was purified from the cells and TMnm mRNA was amplified by RT-PCR. A proportion of the resulting labelled nested PCR products were then digested with the restriction enzyme MboII or AluI. Samples before and after digestion with MboII or AluI were electrophoresed on denaturing polyacrylamide gels. The gels were then fixed, dried and analysed using the PhosphorImager.
A) ThePhosphorImager scan of the MboII digested products. Products representing different isoforms are labelled for the SK WT and SK16-30R3 lanes (1-3, and 8-10, respectively).
B) The PhosphorImager scan of the AluI digested products, only the products corresponding to amplification of the NM and SK isoforms are shown.
The undigested products of amplification are not shown. M=marker lane.

A)
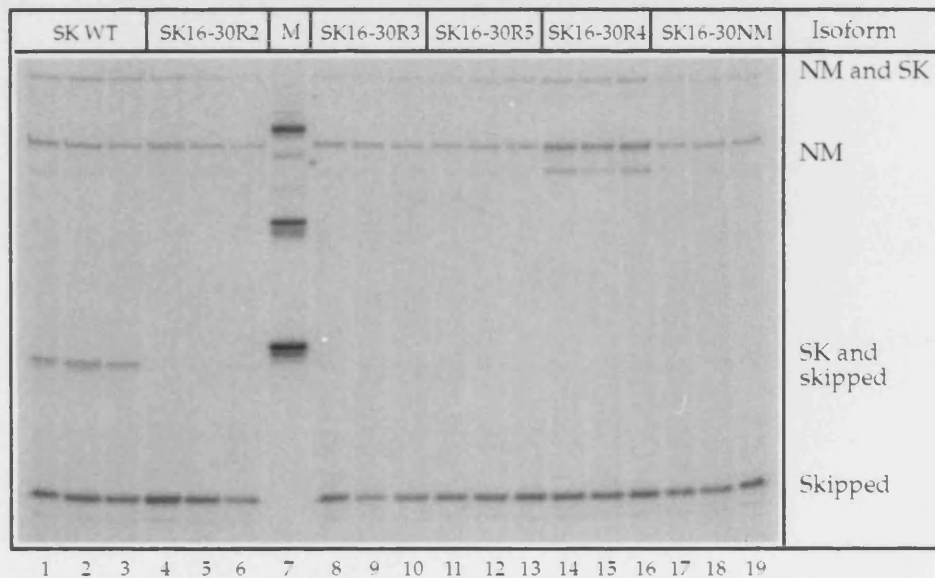
B)
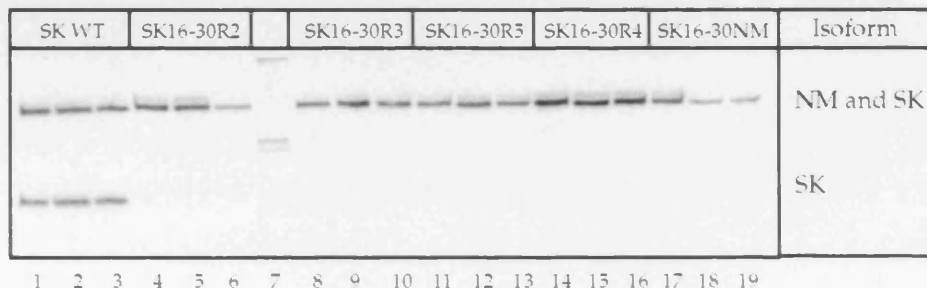
expressed from the SK16-30R2, SK16-30R4, and SK16-30R5 mutant minigenes is not retained by the mutant sequence. Consequently, the expected size of the SK isoform-derived product after MboII analysis was different to that resulting from SK isoform expression from the wild-type and SK16-30R3 minigenes and the relative level of SK isoform expression from the SK16-30R2, SK16-30R4, and SK16-30R5 mutant minigenes was not calculated. The relative levels of isoforms expressed from the wild-type and SK16-30R3 minigenes are shown in table 3.3. Again, appendix 2 gives details of the complete data and calculations. The data shows that the relative level of SK isoform expression was reduced by a factor of ~9 when the wild-type SK exon sequence was mutated to the SK16-30R3 sequence.

TMnm plasmid	Isoform		
	NM	SK	Skipped
SKWT	24 (2)	26 (1)	49 (1)
SK16-30R3	39 (5)	3 (1)	59 (4)

Table 3.3 Effect of mutation of SK nucleotides 16-30 on TMnm mRNA isoform expression in muscle cells. Values are shown in percent and are averages of triplicate transfections. Standard deviations are shown in brackets.

In order to determine if SK isoform expression from the SK16-30R2, SK16-30R4, and SK16-30R5 mutant minigenes was much reduced compared to that from the wild-type minigene, the products of nested PCR amplification were analysed using AluI restriction enzyme. Figure 3.4B shows the products of AluI digestion resulting from amplification of the NM and SK isoforms. The wild-type lanes (1-3) clearly contain a product derived from the SK mRNA isoform. As with the MboII analysis, the expected size of the SK isoform-derived product after amplification/AluI restriction enzyme analysis of SK isoform expression from the SK16-30R2, SK16-30R4, and SK16-30R5 mutant minigenes is different to that resulting from SK isoform expression from the wild-type minigene. However, unlike the MboII analysis, a product specific for expression of the SK isoform is produced. The size of any AluI digestion product derived from amplification of the SK isoform expressed from the SK16-30R3, SK16-30R4, SK16-30R5, and SK16-30NM mutant minigenes is expected to be 12 nucleotides longer than the product derived from SK isoform expression from the wild-type minigene. Figure 3.4B shows that there is no such product visible in lanes 8-19. The product derived from the SK16-30R2 minigene is expected to be only one nucleotide shorter than that from the wild-type minigene. This product is not visible in the lanes representing expression from this minigene (lanes 4-6).

Thus, none of the minigenes in which the wild-type SK exon nucleotides at positions 16-30 were mutated were able to express a significant level of the SK isoform in muscle cells. It was concluded that wild-type SK exon nucleotides 16-30 contain sequence that is necessary for efficient expression of that exon in

muscle cells and, for expression of the SK isoform from the bpup minigene in non muscle cells. This sequence of the SK exon from positions 16-30 is referred to below as the SK exon splicing enhancer, or SK ESE.

All six mutants of SK exon nucleotides 16-30 had a dramatic effect on SK expression. The wild-type sequence in this region is very purine-rich (see table 3.2). It might be expected that substitution of the wild-type sequence for other purine-rich sequences would not seriously affect the ability of this region to act as an enhancer of SK exon expression. However, the mutants which retain a high purine content (the SK16-30NM and SK16-30R3 mutants contain 11 purines out of 15 nucleotides) still do not support significant expression of the SK isoform. Therefore the exact sequence of the SK ESE is important to its function.

The wild-type SK exon sequence at positions 16-33 can be divided into two purine-rich sections: a $(GGA)_3$ motif (SK exon nucleotides 16-24) and a $(GAA)_2$ motif (SK nucleotides 27-33), separated by CT. The smallest change in sequence needed to disrupt efficient SK exon expression was mutation of only 12 nucleotides (the SK16-30R3 mutation). In this mutant the $(GAA)_2$ motif remains intact. It is possible, therefore that this sequence is not required for efficient recognition of the SK exon and that the $(GGA)_3$ CT sequence is sufficient. Mutation of SK exon nucleotides 31-45 to sequence corresponding to NM exon nucleotides 31-45 (SK31-45NM) had no effect on SK expression in non muscle or muscle cells (Graham et al, 1992; Hamshere, PhD thesis). Thus, the 3'-half of the $(GAA)_2$ motif (SK exon nucleotides 31-33) is not required for efficient recognition of exon SK by the splicing apparatus. No mutation has yet been made of the 5'-half of the $(GAA)_2$ motif (SK exon nucleotides 28-30) that does not also disrupt the $(GGA)_3$ motif. A role for the 5'-half of the $(GAA)_2$ motif in SK exon expression cannot, therefore be ruled out.

Switch to use of exon SK in muscle cells

In muscle cells, SK isoform expression is dominant to NM isoform expression (Hamshere, PhD thesis). This switch in expression pattern from that in non muscle cells could be explained by loss of the repression of exon SK specifically in muscle cells. The NM and SK exons would both then be available to the splicing apparatus and the relative level of use of these exons would be determined by the strength of the splicing signals associated with each exon. Alternatively, more complex models would involve specific inactivity of the NM exon and/or specific enhancement of exon SK use in muscle cells. In order to distinguish these types of model, either the NM or the SK exon was inactivated by mutation. If the NM and SK exons are directly in competition with each other for recognition by the splicing apparatus, then inactivation of the SK exon is expected to result in an increase in the level of NM isoform expression. Similarly,

inactivation of the NM exon is expected to allow the level of SK isoform expression to increase. If the NM exon is specifically inactive in muscle cells, then inactivation of exon SK should not alter the level of NM isoform expression.

The NM exon was inactivated by deleting nucleotides at its 5'ss (the corresponding mutant TMnm minigene plasmid is referred to as ΔNM5'ss and is identical to the wild-type TMnm minigene except for the nucleotides deleted at the NM exon 5'ss), and the SK exon was inactivated by deleting the branch point sequence upstream of the exon (the corresponding mutant TMnm minigene plasmid is referred to as ΔSKBP and is identical to the wild-type TMnm minigene except for the nucleotides deleted at the SK BPS). These mutants have been described previously (Graham *et al*, 1992). The ΔNM5'ss and ΔSKBP mutant minigenes were expressed in myoblast and myotube cells. The results of the RT-PCR/restriction enzyme analysis of RNA purified from myoblast cells are shown in figure 3.5. RNA purified from the myotube cells was analysed in another study (I. Eperon, unpublished) and the gel is not shown here. The quantified results of TMnm mRNA isoform expression in myoblasts and myotubes are shown in table 3.4. Data from the myoblast transfections and details of the calculations are given in appendix 2.

TMnm plasmid	Myoblasts				Myotubes			
	Isoform				Isoform			
	NM/SK	NM	SK	Skipped	NM/SK	NM	SK	Skipped
SKWT	0 (0)	56 (2)	0 (0)	44 (2)	0 ±0	24 ±2	20 ±2	54 ±1
ΔNM5'ss	0 (0)	0 (0)	9 (1)	91 (1)	0 ±0	0 ±0	25 ±3	75 ±3
ΔSK BP	0 (0)	52 (2)	0 (0)	48 (2)	0 (0)	16 (1)	18 (4)	65 (4)
NM5'ssC	0 (0)	89 (1)	0 (0)	11 (1)	1 ±0	79 ±3	18 ±1	3 ±1

Table 3.4 Effect of mutation of NM and SK exon splicing signals on TMnm mRNA isoform expression in myoblast and myotube cells.

Values are shown in percent and are averages of duplicate or triplicate transfections. Standard deviations for triplcate transfections are shown in brackets. For the duplicate transfections difference of individual values from the mean is shown.

There was no SK isoform expressed from the ΔSKBP minigene in myoblasts (figure 3.5B, lanes 10-12; table 3.4), as expected. Unfortunately, in myotubes, this mutant minigene did support expression of the SK isoform (table 3.4). It is likely that an alternative branch point upstream of exon SK was recognised that allowed incorporation of the SK exon into the mature mRNA. This was not investigated further. SK isoform expression is, however, disrupted in myotubes by mutation of SK exon nucleotides 16-30 (figure 3.4). The relative levels of mRNA isoforms expressed from the SK16-30R3 minigene in muscle cells are shown in table 3.3.

Figure 3.5 Effect of mutation of exon NM or SK splice site signals on TMnm mRNA isoform expression in myoblasts

Mouse myoblast cells were transfected with the wild-type, NM5'ssC, ΔNM5'ss, and ΔSKBP TMnm minigenes. After RNA had been purified from the cells, TMnm mRNA was amplified by RT-PCR. A proportion of the resulting labelled nested PCR products were then digested with the restriction enzyme MboII. Products before and after digestion with MboII were electrophoresed on a denaturing polyacrylamide gel. The gel was then fixed, dried and analysed using the PhosphorImager.
A) PhosphorImager scan of the undigested products.
B) PhosphorImager scan of the MboII digested products.
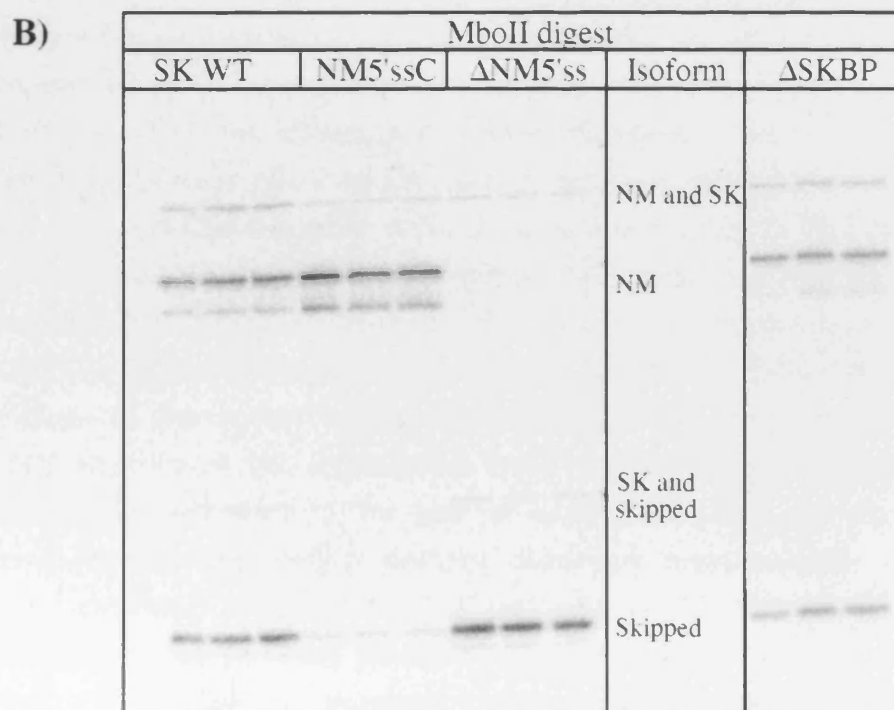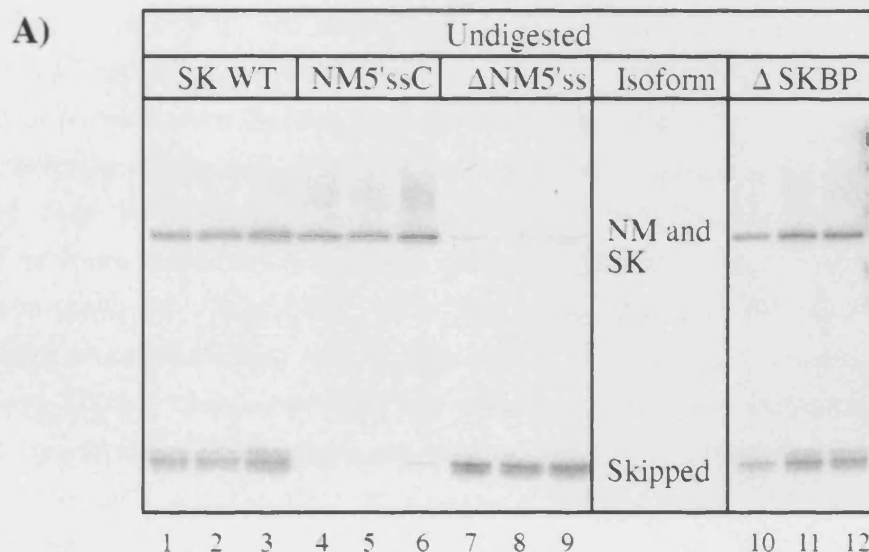
The relative level of the NM and skipped isoforms observed in the SK16-30R3 transfection was higher than in the wild-type transfection. However, care must be taken over the interpretation of these results. Because the level of SK isoform expression is reduced to 3% by the SK16-30R3 mutation, the relative levels of the NM and skipped isoforms will increase even if actual expression of these isoforms did not alter at all. If this is the case, the relative levels of these isoforms, in percentage terms, are expected to be 32% NM (i.e. [(24/73) x97]), 3% SK, and 65% skipped (i.e. [(49/73) x97]). Comparison of these relative levels with those calculated for expression from the SK16-30R3 minigene (39% NM, 3% SK, 59% skipped), shows that the relative level of NM isoform expression is slightly higher than expected than if the actual level of expression of that isoform was unaffected by a significant reduction in SK isoform expression. It was concluded that, under wild-type conditions in muscle cells, NM isoform expression is limited to some extent by the activity of the SK exon.

Activity of the SK exon does not appear to be the only limitation on use of the NM exon in muscle cells. When exon SK was inactivated, the relative level of NM isoform expression did not return to the level observed in myoblast cells (compare table 3.3: SK16-30R3, 39% NM with table 3.4: WT SK, 56% NM). It was concluded that NM exon use is reduced in muscle cells because of competition with exon SK for components of the splicing apparatus, but also because the NM exon is specifically inactivated, or because factors which enhance its expression in non muscle cells are inactive or absent in muscle cells. This is discussed further in chapter 6.

Analysis of the relative levels of isoform expression in myoblast cells from the ΔNM5'ss minigene (figure 3.5B, lanes 7-9, table 3.4) showed that relative use of the SK exon did increase slightly in comparison with SK isoform expression from the wild-type TMnm minigene (compare WT SK, 0% SK with ΔNM5'ss, 9% SK). Therefore use of exon SK in non muscle cells does appear to be limited by activity of the NM exon, although the effect of sequence within the SK exon and a suboptimal SK branch point sequence is of primary importance in SK repression (Graham *et al*, 1992). In muscle cells, if inactivation of exon NM had no effect on actual expression of the SK and skipped isoforms, the relative level of these isoforms calculated from the analysis of wild-type expression levels is expected to be 27% SK (i.e. [((20/(20+54)) x100], and 73% skipped (i.e. [((54/(20+54)) x100]. This is very close to the actual values observed (25% SK, and 75% skipped). It was concluded that exon SK expression from the wild-type TMnm minigene in muscle cells is not limited by use of exon NM. This conclusion was also confirmed by the expression pattern observed from another mutant TMnm minigene (NM5'ssC) in which the 5'ss sequence of the NM exon was mutated to the consensus sequence. This minigene was expressed in myoblast and myotube cells (figure 3.5; I.Eperon, personal communication) and the relative levels of mRNA isoforms produced were determined (table 3.4). The data shows that the

relative expression level of the NM isoform from the NM5'ssC minigene was much enhanced in both cell types compared to that from the wild-type minigene. However, despite the increase in the relative NM isoform expression level, no reduction in the relative level of SK isoform expression was observed.

It was concluded from the results described above that use of exon NM in is limited in muscle cells by the activity of exon SK and because the NM exon is specifically inactivated, or because factors which enhance its expression in non muscle cells are inactive or absent in muscle cells.

Residual repression of exon SK in muscle cells

Expression of the SK isoform from the TMnm wild-type minigene in muscle cells is dominant to expression of the NM isoform, but is still not 100%. One limitation on SK isoform expression in myotubes is the sequence of the branch point sequence upstream of exon SK; the only isoform expressed from the TMnm bpup minigene in muscle cells was the SK isoform (Hamshere, PhD thesis). Thus, the SK BPS is suboptimal in non muscle and in muscle cells. In order to determine whether the sequence of the 5'- or 3'-end of exon SK is also inhibitory to SK isoform expression in muscle cells, the SK1-15NM and SK61-75NM mutant minigenes were transfected into myoblast cells. These cells were caused to differentiate into myotubes and RNA purified from the cells was analysed by RT-PCR/restriction enzyme analysis as before. The resultsare shown in figure 3.6 and the quantified results are given in table 3.5. Appendix 2 contains the complete data.

TM plasmid	NM/SK	NM	SK	Skipped
SKWT	0 (0)	11 (1)	32 (2)	56 (0)
SK1-15NM	0 (0)	0 (0)	100 (0)	0 (0)
SK61-75NM	0 (0)	0 (0)	100 (0)	0 (0)

Table 3.5 Effect of mutation of SK exon sequence on TMnm mRNA
isoform expression in myotube cells.
Values are shown in percent and are averages of triplicate
transfections. Standard deviations are shown in brackets.

Table 3.5 shows that both mutants of the SK exon allowed 100% use of that exon in muscle cells. It was concluded that SK isoform expression is limited in muscle cells by a suboptimal branch point sequence and by sequences at the 5'-, and 3'-end of exon SK (assuming that NM exon sequence which replaced exon SK sequence in the SK1-15NM and SK61-75NM minigenes did not contain sequence activatory to SK isoform expression, see below). Together the above results demonstrate that repression of exon SK is only partially lifted in muscle cells, but

Figure 3.6 Effect of mutation of SK exon sequence on TMnm mRNA isoform expression in myotubes

Mouse myoblast cells were transfected with the wild-type, SK1-15NM, and SK61-75NM TMnm minigenes and caused to differentiate. After RNA had been purified from the cells, TMnm mRNA was amplified by RT-PCR. A proportion of the resulting labelled nested PCR products were then digested with the restriction enzyme MboII. Products before and after digestion with MboII were electrophoresed on a denaturing polyacrylamide gel. The gel was then fixed, dried and analysed using the PhosphorImager.
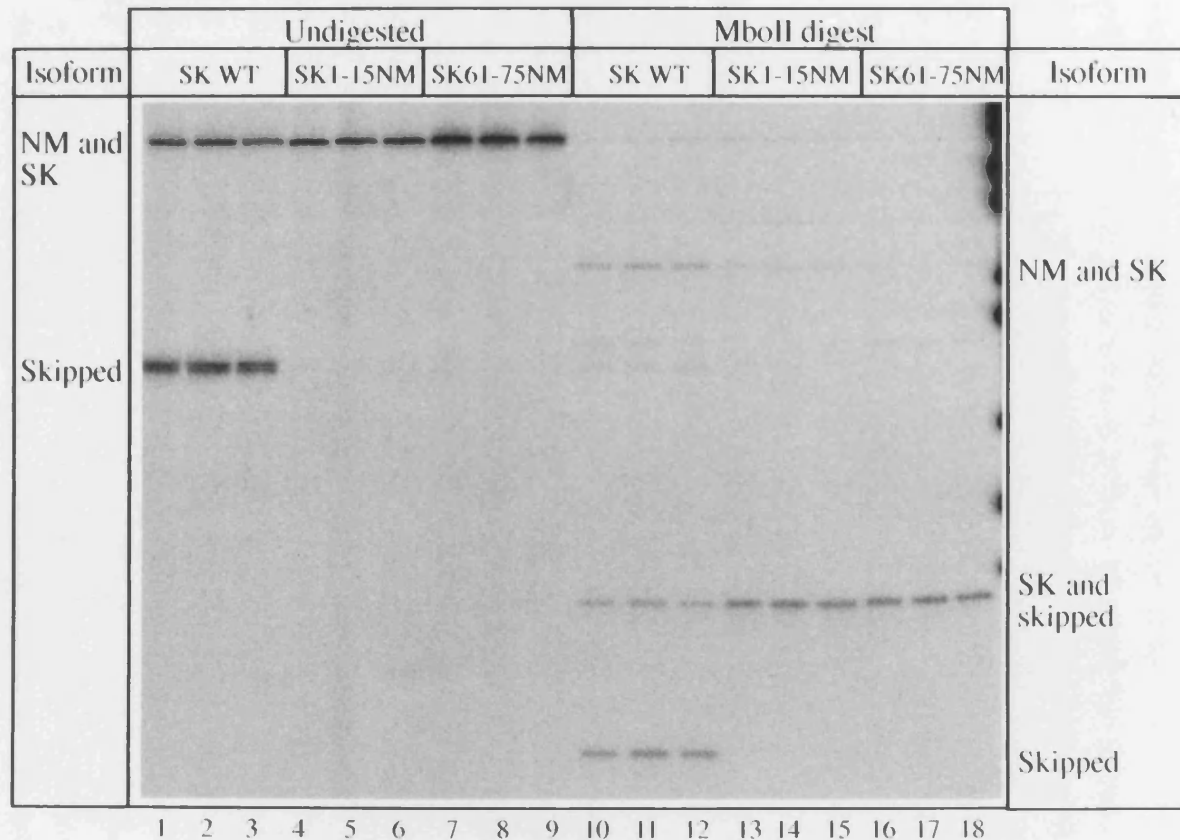
that the activity of the SK exon in these cells and a reduction in the activity of exon NM is sufficient to allow the SK isoform to dominate the expression pattern.

Analysis of TMnm pre-mRNA splicing patterns in myotube cells is potentially complicated by the possibility that myotube cells may be contaminated with undifferentiated myoblast cells. There is, however, no evidence that this has occurred in the results presented here. The SK1-15NM and SK61-75NM mutant minigenes supported high levels of NM isoform expression in myoblast cells (48 (3)% and 47 (1)%, respectively) and no NM isoform expression in myotube cells. If the myotubes were contaminated with significant numbers of myoblast cells then the percent NM isoform expression in myotube cells would have been greater than zero. The influence of undifferentiated myoblasts on these results can therefore be ruled out. The significant differences in isoform levels observed when the other mutant minigenes were expressed in myoblasts and myotubes suggests that contamination effects in these transfections are also minimal.

Sequence spanning the SK exon and the SK branch point sequence is not sufficient to allow tissue-specific regulation of exon SK splicing in a heterologous environment

The work described above, together with previous experiments, identified sequences required for the regulation of SK isoform expression in non muscle and muscle cells. If these sequences are sufficient for that regulation they should be able to confer tissue specific splicing activity on the SK exon when it is placed in a different sequence environment. It was decided to insert TMnm sequence spanning the SK exon and the SK BPS, which contains all the cis-acting sequences identified so far, into the β-globin gene. Two mutants were created by insertion of different lengths of TMnm sequence into β-globin intron 2. The inserted sequences differ in the length of TMnm sequence immediately downstream of exon SK. One sequence contains only 6 nucleotides (ßG/TM1) and the other contains 28 nucleotides (ßG/TM2). The extra sequence was included because it contains a number of sequences which resemble 5'ss sequences. It was thought that these sequences may have a functional role in the regulation of SK exon expression. The basic exon arrangement of the heterologous constructs is shown in figure 3.7 and is very similar to that of the TMnm gene itself.

Figure 3.7 The exon arrangement of the TMnm minigene and the β-globin/TMnm heterologous genes

A) The central part of the TMnm minigene and the heterologous genes (represented as one gene here) are shown. Two heterologous genes were made which contained different lengths of TMnm sequence immediately downstream of exon SK (see text). Exon sequence is represented by open boxes, intron sequence by horizontal lines. The region of the heterologous genes that contains TMnm sequence is shown as the thick black line and by the filled box. The dominant splicing pattern in non muscle and muscle cells is shown for the TMnm gene by the diagonal lines. The predicted dominant splicing pattern is shown for the heterologous genes.

B) Possible mRNA isoforms that can be expressed from the heterologous genes. Only the regions of the TMnm and ß-globin genes that are of interest are shown.
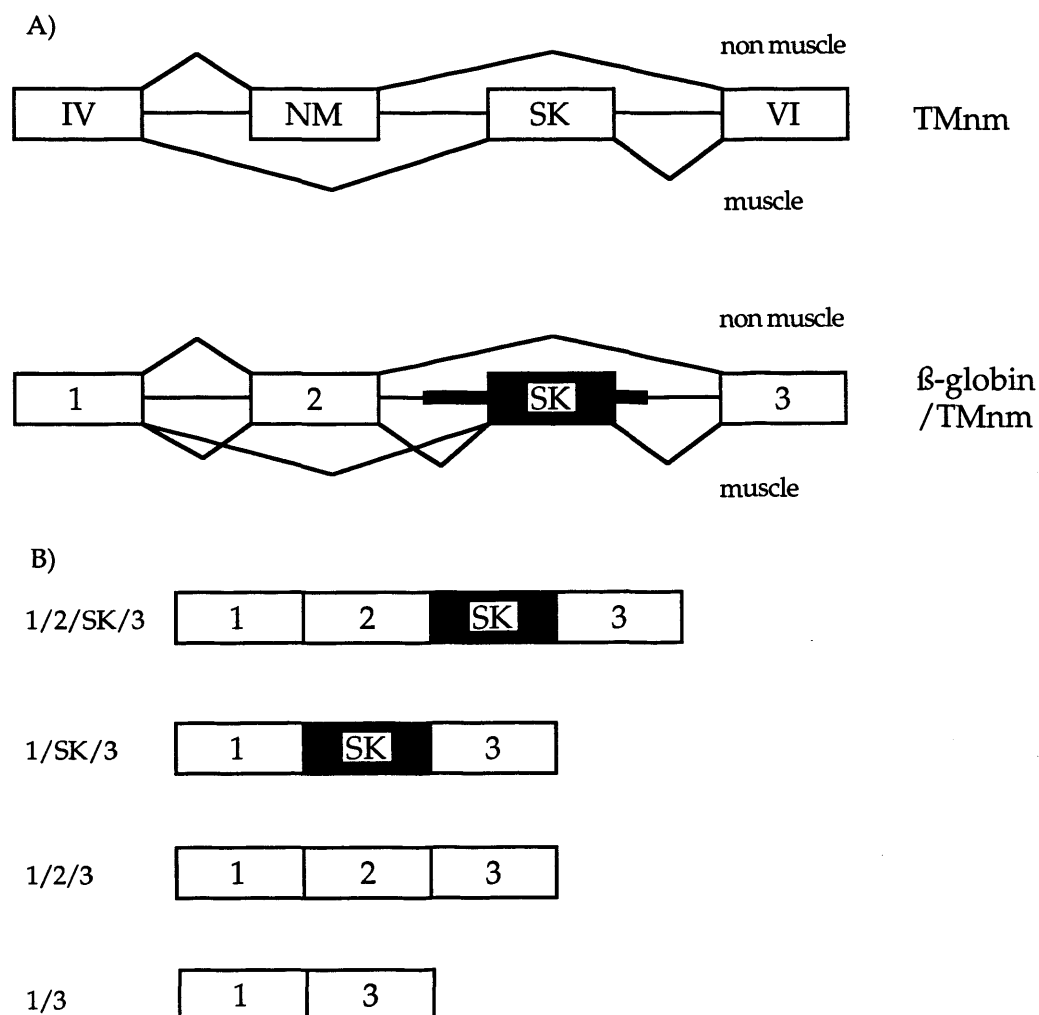
If all the sequence necessary for regulation of TMnm SK exon expression is contained within the TMnm sequence inserted into the β-globin gene, then the splicing pattern of the heterologous genes should be similar to that of the TMnm minigene. The mRNA isoforms which can be expressed from the heterologous genes are shown schematically in figure 3.7B. In non muscle cells, expression of the 1/3 and 1/2/3 isoforms is expected. The predicted pattern is more complicated

in muscle cells. The mechanism responsible for the mutually exclusive expression of exons NM and SK has not yet been elucidated. It is possible that sequences involved in the mutually exclusive incorporation of these two exons are located within the inserted TMnm sequence. If mutual exclusion is preserved, then 1/3, 1/SK/3 and 1/2/3 isoforms are expected. If it is not, the 1/2/SK/3 isoform is expected to be expressed in addition to these isoforms. In either case, the relative levels of SK exon-containing isoforms should be dominant to the relative level of the 1/2/3 isoform just as expression of the SK isoform is dominant to NM isoform expression from the TMnm minigene in muscle cells.

Messenger RNA isoform expression from the wild-type β-globin gene and the two heterologous genes was determined in non muscle and muscle cells. This was done by RT-PCR analysis of RNA purified from cells transfected with these genes (the exact procedure is described in chapter 2). Figure 3.8 shows the results. Bands corresponding to products derived from different mRNA isoforms are labelled. It can be seen that the wild-type β-globin gene predominantly expressed an isoform containing all three exons in both non muscle and muscle cells (the isoform is labelled as 1/2/3 in the figure, lanes 3, 4, 9, 10). A degree of skipping of exon 2 was also observed in each cell type (the 1/3 isoform, lanes 3, 4, 9, 10). 1/2/3 isoform expression was also the dominant isoform expressed from the heterologous genes (lanes 5-8, 11-14). However, bands corresponding in size to those expected from 1/SK/3 and 1/2/SK/3 isoform expression are also visible for each hybrid gene in non muscle and muscle cells. There is no obvious increase in the relative level of isoforms containing exon SK in muscle cells compared to myoblast cells for either of the heterologous genes (compare lanes 4, 5 with 11, 12 and 6, 7 with 13, 14).

It was concluded that TMnm sequence spanning the SK BPS and SK exon does not contain all the *cis*-acting elements required for the inhibition of exon SK use in non muscle cells, for the muscle-specific expression of exon SK, or for the mutually exclusive incorporation of the NM and SK exons. Alternatively, it is possible that the β-globin sequences surrounding the TMnm sequence had an inhibitory effect on the regulatory activity of the TMnm sequence.

Figure 3.8. Sequence of the TMnm gene required for regulated expression of the SK exon in non muscle and muscle cells

Mouse myoblast cells were transfected with the wild-type ß-globin gene (ßG) and the two mutant genes containing TMnm sequence spanning the SK branchpoint sequence and the SK exon (ßG/TM1 and ßG/TM1). 24 hours later, RNA was purified from the cells or, in a parallel experiment, the cells were caused to differentiate into myotubes before RNA was purified. mRNA expressed from the genes was amplified by RT-PCR. The products were electrophoresed on a denaturing polyacrylamide gel. The gel was then fixed, dried and analysed using the PhosphorImager.
The identity of the 1/2/SK/3, 1/2/3, 1/SK/3 and 1/3 isoforms are given in figure 3.7. 1-2/3 represents amplified product thought to be derived from expression of an RNA in which exons 2 and 3 have been spliced together, but in which the intron separating exons 1 and 2 remains.
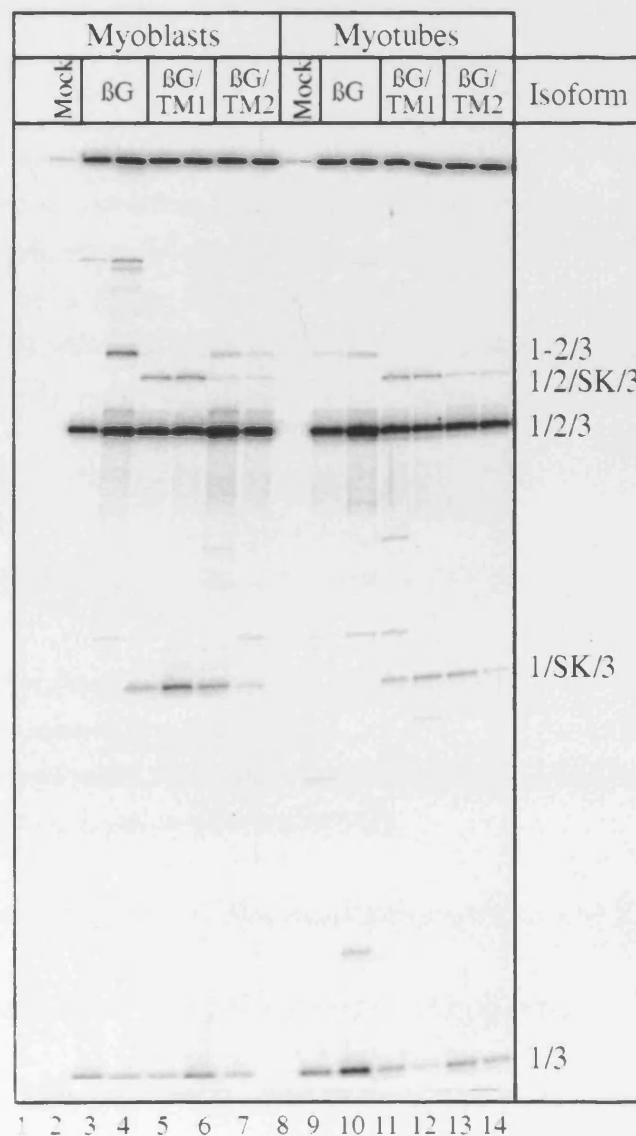
Chapter 4
Repression of exon SK in non muscle cells

The 5' end of exon SK contains a pseudo 5'ss sequence

The results described in the previous chapter, as well as earlier work (Graham *et al*, 1992; Hamshere, PhD thesis), established that mutation of sequence at the 5'end of exon SK activates expression of that exon in non muscle cells. When the pre-mRNA sequence of the mutated region was examined it was noted (I.Eperon) that the three nucleotides immediately upstream of the SK exon, together with the first five nucleotides of the exon strongly resemble a 5'ss sequence:

TMnm SK exon	$_{-3}$CAGUAAGU$_5$
Consensus 5'ss	(A/C)AGGURAGU

Thus, the TMnm sequence contains an exact match to the consensus 5'ss except that the G nucleotide at position 3 of the consensus sequence is not present. One of the first steps in spliceosome assembly is recognition of the 5'ss sequence by U1 snRNP. U1 snRNP is a large RNA-protein complex. The 5'-end of the RNA component of U1 snRNP (U1 snRNA) mediates binding of this complex to 5'ss sequences and is exactly complementary to the consensus 5'ss sequence:

Consensus 5'ss	CAGGUAAGU$_{3'}$
	• • • • • • • • •
5'-end of U1snRNA	GUCCA$\psi\psi$CAUApppm$_3$G$_{5'}$

• represents Watson-Crick base-pairing

ψ represents pseudo-uridine

The 5'-end of U1 snRNA ends with adenosine triphosphate (Appp) and a tri-methlyguanosine cap (m$_3$G)

The TMnm sequence SK (-3 to +5) also base pairs well to the 5'-end of U1 snRNA:

5'-end of U1snRNA	GU^{C}CA$\psi\psi$CAUApppm$_3$G
	• • • • • • • • • o •
Tropomyosin SK$_{-3\ to\ +7}$	CA GUAAGUGU$_{3'}$

There are six continuous Watson-Crick base pairs predicted to form between the TMnm pre-mRNA and U1snRNA. There is also a potential G-U base pair (marked"o") at position +6 of the TMnm SK exon sequence. A further two base

pairs can be formed with positions -3 and -2 relative to the SK exon if a one nucleotide bulge occurs in U1 snRNA. Thus, in total 10 out of 11 nucleotides of U1 snRNA could pair with the TM pre-mRNA sequence. The TM SK$_{-3 \text{ to } +7}$ sequence is referred to as the SK pseudo 5'ss (p5'ss) sequence below.

It was proposed (I.Eperon) that the inhibitory effect of sequences at the 5'-end of exon SK, on expression of that exon in non muscle cells, could be mediated by the binding of U1 snRNP to the SK p5'ss. In muscle cells, U1 snRNP would in some way be prevented from binding to the TMnm pseudo 5'ss sequence and the repression would be lifted. The mechanism by which U1 snRNP could inhibit SK splicing in non muscle cells might, in principle, involve steric inhibition of access of splicing factors to the SK 3'ss which is contained within the SK p5'ss sequence, or of access of a factor(s) binding to the adjacent SK ESE:
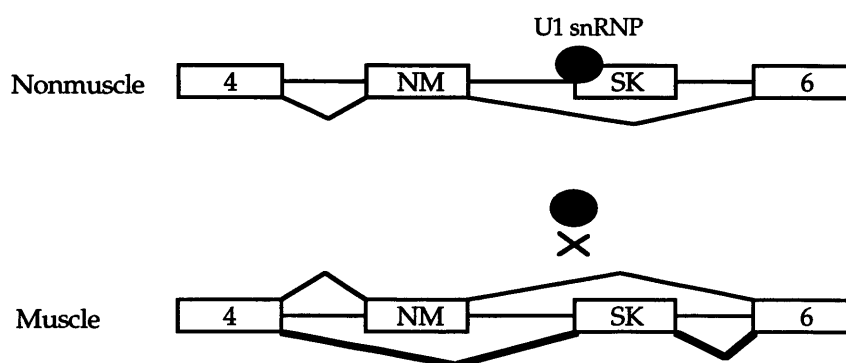
Figure 4.1 Model for the regulation of TM expression by U1 snRNP

There are some attractive features to this model:

1) the factor proposed to mediate repression of SK isoform expression in non muscle cells is constitutively expressed. It seems very unlikely that all non muscle cells would express a specific factor just to inhibit SK isoform expression. Instead it is more likely that an event specific only to muscle cells would lift the repression;

2) it is well known that many sequences which contain a strong match to a 5'ss consensus sequence are not used as splice sites, yet sequences with a weaker match are faithfully used. Thus, the machinery of the cell is well adapted for the selective use of sequences as 5'splice sites. In the model proposed here, the cell would exploit this ability for selecting 5'ss sequences in a tissue-specific way: in non muscle the SK pseudo 5'ss is recognised by U1 snRNP and in muscle it is not.

The Role of U1snRNP in repression of exon SK expression in non muscle cells

If U1 snRNP does mediate repression of exon SK expression in non muscle cells, a TMnm minigene containing mutations within the SK p5'ss that disrupt the proposed base-pairing interactions with U1 snRNA is predicted to express the

4.2

SK isoform in non muscle cells. It should then be possible to suppress this expression *in vivo* by expressing a U1 snRNA containing compensatory mutations that restore the complementarity of base pairing with the mutant TMnm sequence (assuming that the mutant U1 snRNA can assemble with U1 snRNP proteins within the cell). A similar experiment was used to demonstrate a functional interaction of U1snRNA with 5' splice sites in adenovirus E1a pre-mRNA (Zhuang and Weiner, 1986).

The interaction between the SK pseudo 5'ss and U1 snRNA is predicted to be in the ratio 1:1. Any effect of a U1 snRNA carrying a suppressing mutation may not be observed if the levels of mutant TMnm RNA within the cell are significantly higher than the mutant U1 snRNA levels. This affected the choice of cell-type used for the suppression experiment. Two cell-types have been used in the laboratory for analysis of TMnm minigene splicing, COS cells and mouse myoblast cells. COS cells constitutively express T antigen. This protein is known to significantly enhance expression from SV40 promoters. Expression of the TMnm minigene is dependent on an SV40 promoter (chapter 2), whereas the U1 snRNA gene was cloned with its own promoter. Therefore COS cells are likely to express TMnm minigene RNA to much higher levels than U1 snRNA. However, myoblast cells, which do not express T antigen, are expected to express much lower levels of TMnm minigene RNA (the difference in amplification conditions required for visualisation of radioactively labelled products amplified from TMnm mRNA isolated from myoblast and COS cells (20 compared with 8 cycles of nested PCR, respectively) confirms this). It was decided therefore, to use myoblast cells for experiments aimed at investigating the role of U1 snRNP in repression of exon SK splicing in non muscle cells.

If U1 snRNP does mediate repression of exon SK it is predicted that:
1) mutations within the SK p5'ss sequence that disrupt the proposed base-pairing interactions with U1 snRNA will activate expression of exon SK in non muscle cells;
2) the relative level of SK isoform expression observed in non muscle cells from a TMnm minigene containing a mutation within the SK p5'ss will be the same as the relative level of SK isoform expression observed from the SK1-15NM minigene;
3) no SK isoform expression will be observed from the TMnm minigene containing a mutation of the SK p5'ss if U1 snRNA containing compensatory mutations that restore base pairing with the mutant SK p5'ss is present and is assembled as part of U1 snRNP in equimolar amounts with the mutant TMnm pre-mRNA;
4) expression of U1 snRNA that does not contain compensatory mutations that restore base pairing with the mutant SK p5'ss should have no effect on the levels of SK isoform expression from the mutant TMnm minigene;

5) expression of U1 snRNA that contains compensatory mutations that restore base pairing with the mutant SK p5'ss used in (3) above should have no effect on the relative level of SK isoform expression observed from the SK1-15NM minigene.

These predictions are summarised in table 4.1.

TMnm plasmid	U1 snRNA plasmid	SK exon expression
SK p5'ss mutant	None orWT	Yes
SK p5'ss mutant	α-SK p5'ss mutant	No
SK1-15NM	None or WT or α-SK p5'ss mutant	Yes

Table 4.1 Predicted SK expression patterns from TMnm minigene mutants in the presence of wild-type or mutant exogenous U1 snRNA.

SK p5'ss mutant represents a TMnm plasmid which expresses TMnm pre-mRNA containing a mutation of the SK p5'ss sequence. α-SK p5'ss mutant represents a plasmid expressing mutant U1 snRNA in which complementarity is restored with pre-mRNA expressed from the SK p5'ss mutant minigene in the region of the SK p5'ss.

A mutation of the SK p5'ss at positions 2-4 of the SK exon was made in the TMnm minigene (the resulting minigene is referred to as SK2-4CGC below). The 5'-end of wild-type U1 snRNA is predicted to form only two continuous Watson-Crick base pairs with mutant TMnm pre-mRNA expressed from the SK2-4CGC minigene in the region of the mutant SK p5'ss, compared to six continuous Watson-Crick base pairs which are predicted to form with pre-mRNA expressed from the wild-type TMnm minigene. A compensatory mutation was made in the U1 snRNA gene that restores the proposed base pairing of U1 snRNA expressed from this gene with pre-mRNA expressed from the SK2-4CGC minigene (the resulting plasmid carrying this mutant U1 gene is referred to as U1 α-CGC below):

Wild-type TMnm SK$_{-3 \text{ to } +7}$ pre-mRNA

    CA GUAAGUGU$_{3'}$
    • • • • • • • • o •

5'-end of wild-type U1snRNA

    GUCCAψψCAUApppm3G$_{5'}$

SK2-4CGC TMnm SK$_{-3 \text{ to } +7}$ pre-mRNA

    CA GUCGCUGU$_{3'}$
    • • • • x x x • o •

5'-end of wild-type U1snRNA

    GUCCAψψ CAUApppm3G$_{5'}$

x represents loss of base-pairing

SK2-4CGC TMnm SK$_{-3 \text{ to } +7}$ pre-mRNA

    CA GUCGCUGU$_{3'}$
    • • • • • • • • o •

5'-end of U1 α-CGC snRNA

    GUCCAGCGAUApppm3G$_{5'}$

Figure 4.2 Proposed base pairing interactions between wild-type and mutant TMnm RNA and U1snRNA

The compensatory U1 snRNA mutation was made from a wild-type human U1 gene cloned into a pGEM3 plasmid which was kindly supplied by Dr C.J.W. Smith. An identical plasmid carrying a point mutation in the U1 snRNA sequence had been made from this construct and it was known that the mutant U1 snRNA could be expressed to detectable levels (CJW Smith, personal communication).

An experiment was carried out to see if the SK2-4CGC minigene expressed the SK isoform in myoblast cells and if any such expression could be suppressed by expression of α-CGC U1 snRNA. The SK2-4CGC mutant minigene was cotransfected into mouse myoblast cells with the pGEM3-based plasmid coding for human wild-type U1 snRNA (the U1 WT plasmid) or for α–CGC U1 snRNA (the U1 α–CGC plasmid). Any U1 snRNA expressed from the transfected U1 genes was expected to assemble with mouse U1 snRNP proteins to form functional U1 snRNPs. U1 snRNPs comprising expressed transfected wild-type U1 snRNA and endogenous wild-type myoblast cell U1 snRNA were not expected to bind with high affinity to pre-mRNA expressed from the SK2-4CGC minigene in the region of the mutated SK p5'ss sequence. The relative level of SK isoform expression observed from the SK2-4CGC minigene when this minigene was cotransfected with a plasmid expressing the wild-type U1 snRNA gene was therefore predicted to be the same as that observed previously from the SK1-15NM transfection (figure 3.2; table 3.1), i.e. 38 (2)%. U1 snRNPs assembled with expressed transfected α–CGC U1 snRNA were expected to bind to pre-

mRNA expressed from the SK2-4CGC minigene in the region of the mutated SK p5'ss sequence and to inhibit SK exon expression. The relative level of SK isoform expression observed from the SK2-4CGC minigene when this minigene was cotransfected with a plasmid expressing α–CGC U1 snRNA was therefore predicted to be less than 38%. The extent of repression would depend on the number of functional recombinant U1 snRNPs containing the suppressing U1 snRNA sequence relative to the number of mutant TMnm pre-mRNA species.

The SK2-4CGC TMnm minigene was also transfected alone, as a control, so that a comparison could be made between the relative level of SK isoform expressed from the SK2-4CGC mingene in the absence and presence of exogenous wild-type U1 snRNA. It was predicted that the presence of exogenous wild-type U1 snRNA would have no effect on the relative level of SK exon expression. Thus, the relative SK isoform expression level from the SK2-4CGC minigene alone and from a SK2-4CGC/U1 WT cotransfection was expected to be the same as that from the SK1-15NM minigene, i.e. 38 (2)% (table 3.1).

It was important to show that any suppression effect was specific to the restoration of base pairing predicted to occur between the TMnm SK2-4CGC pre-mRNA and the α–CGC U1 snRNA in the region of the mutant SK p5'ss. Therefore a minigene expressing a different mutation of the SK p5'ss (SK1-15NM) was cotransfected with the pGEM3-based plasmid coding for α-CGC U1 snRNA. The relative level of SK isoform expressed from the SK1-15NM minigene in the absence of exogenous mutant U1 snRNA was determined by transfection of the SK1-15NM minigene alone. It was predicted that the relative level of SK isoform expression observed from the cotransfection would be the same as that from the transfection of the SK1-15NM minigene alone, i.e. 38%, the same as the level observed when the SK1-15NM minigene was transfected alone previously (table 3.1).

Figure 4.3 shows a PhosphorImager scan of the gel obtained after RT-PCR/restriction enzyme analysis of RNA purified from the transfections. The levels of each isoform calculated after quantification of the data are shown in table 4.2. The complete data and details of the calculations are given in appendix 2.

Figure 4.3. Analysis of TMnm mRNA isoform expression in myoblasts. The SK2-4CGC and SK1-15NM TMnm minigenes were transfected alone or cotransfected with a plasmid encoding wild-type or anti-CGC U1 snRNA.

Poly(A)-selected mRNA isolated after the transfections was analysed by RT-PCR. A proportion of the radioactively labelled nested PCR products were then digested with the MboII restriction enzyme. Samples before and after digestion with MboII were subjected to electrophoresis on a denaturing polyacrylamide gel. The gel was then fixed, dried, and analysed using the PhosphorImager. The figure shows the PhosphorImager scan obtained. Bands resulting from amplification of particular TMnm mRNA isoforms are labelled on either side of the scan. Wild-type is abbreviated to WT.
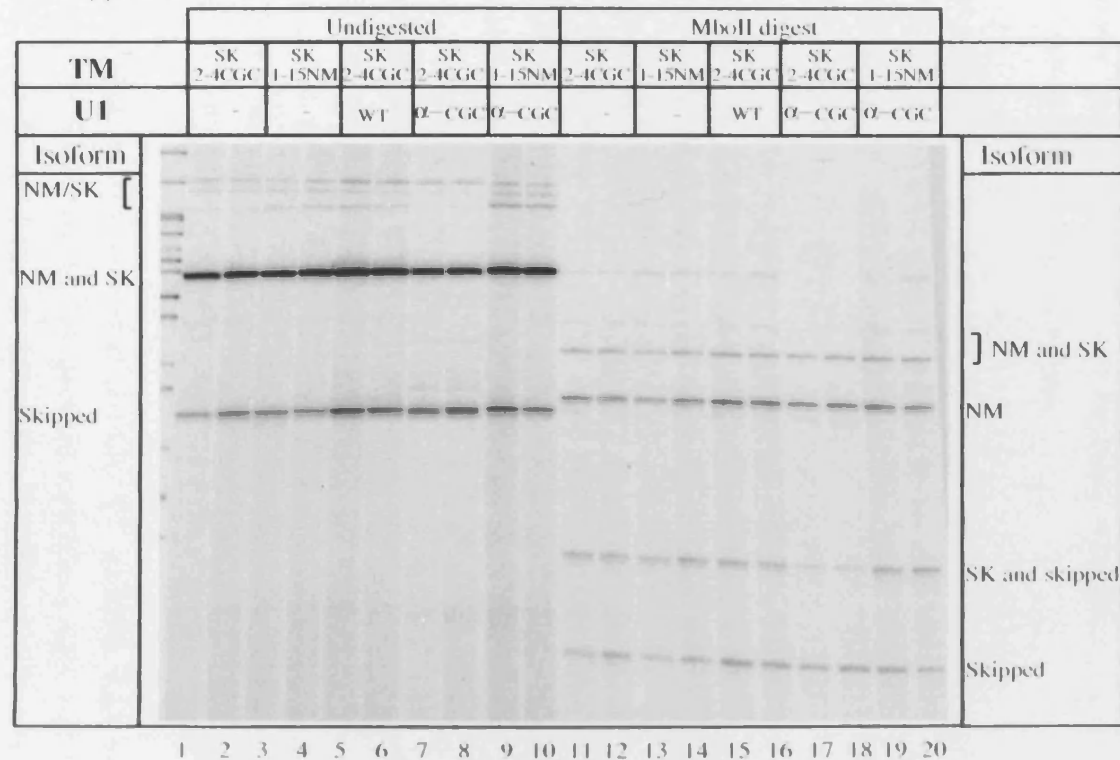
		Isoform			
TMnm plasmid	U1 plasmid	NM/SK	NM	SK	Skipped
SK2-24CGC	-	5.0 ±0.7	49 ±3	25 ±0.5	21 ±4
SK1-15NM	-	8.6 ±0.5	39 ±0	32 ±2	20 ±2
SK2-4CGC	U1 WT	7.9 ±2	46 ±3	22 ±2	24 ±1
SK2-4CGC	U1α-CGC	3.8 ±0.2	51 ±0.5	13.3 ±0.5	32 ±1
SK1-15NM	U1α-CGC	10 ±0.3	35 ±0.5	29 ±2	26 ±2

Table 4.2 Effect of transfection of wild-type and mutant U1 snRNA-expressing plasmids on exon SK expression from the SK2-4CGC and SK1-15NM minigenes.

Values are shown in percent and are averages from duplicate cotransfections. The difference of the individual values from the mean is shown.

It can be seen from the results that the relative level of SK isoform expression from the SK2-4CGC minigene was comparable to that from the SK1-15NM minigene (25±0.5% and 32 ±2%, respectively). Thus, two different mutations at the 5'-end of exon SK were able to activate expression of that exon in non muscle cells. The effect of substitution of the first fifteen nucleotides of exon SK with sequence identical to the first fifteen nucleotides of exon NM on SK isoform expression was due, therefore, to loss of inhibitory exon SK sequences rather than the introduction of activatory NM sequences.

The relative level of SK isoform expression from the SK2-4CGC minigene was not significantly affected by the presence of a pGEM3-based plasmid coding for wild-type U1 snRNA (compare SK2-4CGC, 25 ±0.5% SK and SK2-4CGC/U1 WT, 22 ±2% SK). However, the relative level of SK isoform expression from the SK2-4CGC minigene was much reduced when the pGEM3-based plasmid coding for α-CGC U1 snRNA was present (compare SK2-4CGC/U1wt, 22 ±2% SK and SK2-4CGC/U1α-CGC, 13.3 ±0.5% SK). In contrast, the relative level of SK isoform expression from the SK1-15NM minigene was unaffected by the presence of the plasmid coding for α-CGC U1 snRNA (compare SK1-15NM, 32 ±2% SK; SK1-15NM/U1α-CGC, 29 ±2% SK).

It was concluded from this experiment that α-CGC U1 snRNA had been expressed from the transfected mutant human U1 gene in myoblast cells and had assembled with mouse U1 snRNP proteins to produce functional U1 snRNP particles. These recombinant U1 snRNPs had bound to the SK2-4CGC mutant TMnm pre-mRNA in the region of the SK p5'ss and had inhibited expression of exon SK. The suppression of exon SK expression was not complete, possibly because the number of functional recombinant U1 snRNP particles containing α-CGC U1 snRNA was less than the number of mutant TMnm pre-mRNA species expressed from the SK2-4CGC minigene. The effect, however, was specific to the SK2-4CGC/U1α-CGC cotransfection: the relative level of SK isoform expression from the SK2-4CGC minigene and the SK1-15NM minigene appeared to be

4.7

unaffected by the presence of a plasmid coding for wild-type and α-CGC U1 snRNA, respectively. It was assumed that the pGEM3-based plasmids encoding U1 snRNA were all able to express U1 snRNA (wild-type or mutant), although direct proof of this was not obtained.

As expected, the relative level of SK isoform expression from the SK1-15NM minigene was similar to that obtained previously, although it was not identical (compare table 3.1: SK1-15NM, 38 (2)% SK; table 4.3: SK1-15NM, 32 ±2% SK). Comparison of the relative level of SK isoform expressed from the SK2-4CGC and SK1-15NM TMnm minigenes shows that the relative level was higher when the first 15 nucleotides of exon SK were mutated to the same sequence as the first 15 nucleotides of the NM exon than when three nucleotides at positions 2-4 of exon SK were mutated (compare table 4.3 SK2-4CGC, 25 ±0.5% SK; SK1-15NM 32 ±2% SK). The three nucleotide change at positions 2-4 of exon SK was predicted to have the same effect on SK isoform expression as mutation of the first fifteen nucleotides of the SK exon if binding of U1 snRNP to the SK p5'ss is responsible for repression of SK exon splicing in non muscle cells. However, this was not observed. The difference could be explained in a number of ways:

1) The SK1-15NM and SK2-4CGC mutations may have had exactly the same effect on disrupting binding of U1 snRNP to the SK p5'ss but the SK1-15NM exon sequence may have been spliced with higher efficiency than the mutant SK2-4CGC exon sequence. Consistent with this proposal is the fact that the first 15 nucleotides of the NM exon contain a high proportion of purine nucleotides towards the 3'-end of the sequence: CCGTTGCCGAGAGAT (NM exon nucleotides 1-15). The purine-rich region of exon SK at positions 16-33 is therefore considerably extended in length when the first 15 nucleotides of the NM exon form the first 15 nucleotides of the SK1-15NM exon. This lengthened purine-rich stretch is not present in the SK2-4CGC mutant. Given the effect of purine-rich exon sequences on splicing efficiency (see introduction) it is possible that the SK1-15NM mutant exon is spliced more efficiently than the SK2-4CGC mutant exon;

2) Not all the proposed base pairing interactions between TMnm pre-mRNA and U1 snRNA have been disrupted by the 2-4CGC mutation. Some residual binding of wild-type U1 snRNP to the mutant SK2-4CGC exon could therefore occur and allow partial repression of exon SK splicing;

3) U1 snRNP has no role in the repression of exon SK splicing in non muscle cells and the three nucleotide change has partially disturbed binding of some other factor.

It is very difficult to explain the specific effect of the plasmid encoding the α-CGC U1 snRNA mutation on SK isoform expression from the SK2-4CGC minigene if explanation (3) above is true. It is also difficult to see how U1 snRNP could bind with any specificity to the mutated SKp5'ss when only two contiguous

Watson-Crick base-pairs can be formed with wild-type U1 snRNA. The most likely explanation is that the two mutant SK exons (SK1-15NM and SK2-4CGC) are spliced with differing efficiencies. This could be confirmed if a different non purine-rich mutant of SK nucleotides 1-15 was made. A TMnm minigene containing such a mutant would be expected to express exon SK to a similar relative level as the SK2-4CGC mutant minigene.

The cotransfection experiment was repeated in order to confirm the results. Again, the SK2-4CGC minigene was cotransfected with a wild-type or αCGC U1 snRNA-expressing plasmid. To test the specificity of any supression effect the SK1-15NM minigene was cotransfected with the αCGC U1 snRNA-expressing plasmid as before. As a control to monitor the level of SK isoform expression from the SK1-15NM minigene in the absence of a plasmid coding for αCGC U1 snRNA, the SK1-15NM minigene was also cotransfected with the pGEM3-based plasmid coding for wild-type U1 snRNA. The results of RT-PCR analysis of RNA purified from the transfections are shown in figure 4.4. In the figure, the names of the SK2-4CGC and SK1-15NM TMnm minigene plasmids are abbrieviated to 2-4CGC and 1-15NM. The quantified results are given in table 4.3. The complete data and details of the calculations are given in appendix 2.

TM plasmid	U1 plasmid	NM/SK	NM	SK	Skipped
SK2-4CGC	U1 WT	1.4 ±0	59 ±10	27 ±4	12 ±6
SK2-4CGC	U1 α-CGC	1.2 ±0.6	58 ±0.5	12.5 ±1.5	29 ±2
SK1-15NM	U1 WT	0 ±0	52 ±2	38 ±2	10 ±0.2
SK1-15NM	U1 α-CGC	2.3 ±0.2	40 ±2	42 ±2	16 ±0.5

Table 4.3 Effect of transfection of wild-type and mutant U1 snRNA-expressing plasmids on exon SK isoform expression from the SK2-4CGC and SK1-15NM minigenes.

Values are shown in percent and are averages from duplicate cotransfections. The difference of the individual values from the mean is shown.

The results show that the relative level of SK isoform expression from the SK2-4CGC minigene was again lower in the presence of the αCGC U1 snRNA encoding plasmid than in the presence of the wild-type U1 snRNA-encoding plasmid (compare SK2-4CGC/U1 α-CGC, 12.5 ±1.5% SK with SK2-4CGC/U1 WT, 27 ±4% SK). This suppression effect was specific to the SK2-4CGC minigene: the relative level of SK isoform expression from the SK1-15NM minigene was not significantly affected by cotransfection with a plasmid coding for αCGC U1 snRNA (compare SK1-15NM/αCGC U1, 42 ±2% SK with SK1-15NM/U1 WT, 38 ±2% SK). It was concluded that the reduction in relative level of SK isoform expression from the SK2-4CGC minigene in the presence of the U1 α-CGC

Figure 4.4. Analysis of TMnm mRNA isoform expression in myoblasts. The SK2-4CGC and SK1-15NM TMnm minigenes were cotransfected with a plasmid encoding wild-type or mutant U1 snRNA.

Poly(A)-selected mRNA isolated after the cotransfections was analysed by RT-PCR. A proportion of the radioactively labelled nested PCR products were then digested with MboII. Samples before and after MboII digestion were subjected to electrophoresis on a denaturing polyacrylamide gel. The gel was then fixed, dried, and analysed using the PhosphorImager. The figure shows two PhosphorImager scans obtained. The samples in lanes 13-16 were electrophoresed at a different time to those in lanes 1-12. Bands resulting from amplification of particular TMnm mRNA isoforms are labelled on either side of each scan.
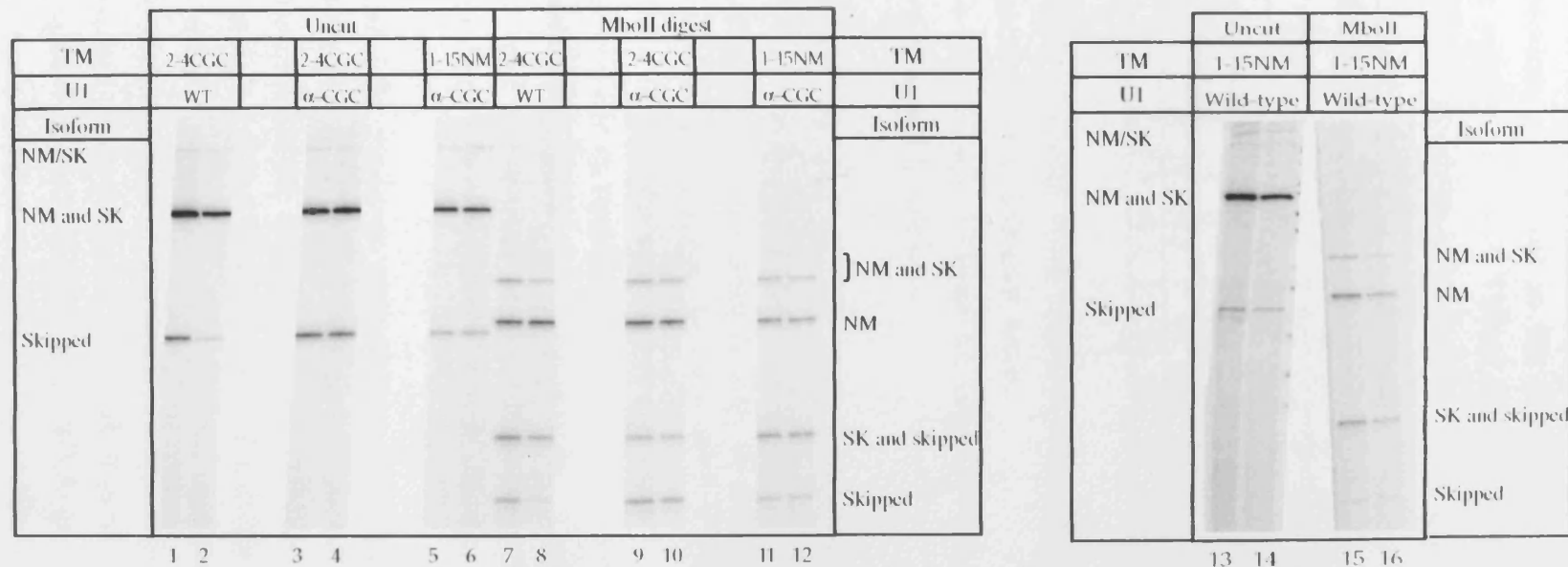
plasmid compared with that from the same minigene in the presence of the U1 WT plasmid observed in the previous experiment (figure 4.3, table 4.3) was also observed in this experiment and that repression of exon SK in non muscle cells by sequences at the 5'-end of the exon is mediated, at least in part, by U1 snRNP.

Some of the cotransfections performed in the experiments of figures 4.3 and 4.4 were identical. The relative levels of TMnm mRNA isoform expression observed in these cotransfections are given in table 4.4:

TM plasmid	U1 plasmid	Experiment	NM/SK	NM	SK	Skipped
SK2-4CGC	U1 WT	A	7.9 ±0.2	46 ±3	22 ±2	24 ±1
SK2-4CGC	U1 WT	B	1.4 ±0	59 ±10	27 ±4	12 ±6
SK2-4CGC	αCGC	A	3.8 ±0.2	51 ±0.5	13.3 ±0.5	32 ±1
SK2-4CGC	αCGC	B	1.2 ±0.6	58 ±0.5	12.5 ±1.5	29 ±2
SK1-15NM	αCGC	A	10 ±0.3	35 ±0.5	29 ±2	26 ±2
SK1-15NM	αCGC	B	2.3 ±0.2	40 ±2	42 ±2	16 ±0.5

Table 4.4 Consistency of results from identical cotransfections performed at different times.

Experiment A refers to the experiment of figure 4.3, table 4.2. Experiment refers to the experiment of figure 4.4, table 4.3.

The relative levels of NM/SK and skipped isoform expression are higher in experiment A than in experiment B. In contrast, the relative levels of NM and SK isoform expression are higher in experiment B than in experiment A (with the exception of SK isoform expression in the SK2-4CGC/U1α-CGC cotransfection). The reasons for these differences have not been investigated, but it seems likely that slight differences in transfection efficiency, conditions of RNA isolation and RT-PCR analysis could be responsible. This comparison demonstrates that it is important to compare results within the same experiment rather than between experiments when relatively subtle changes in expression pattern are being examined.

The relative level of NM isoform expression observed in the SK1-15NM/U1 WT cotransfection (52 ±2%) is anomalous compared to that in the other transfections using the SK1-15NM minigene (compare with table 4.3 SK1-15NM/U1α-CGC, 40 ±2%; table 4.4 SK1-15NM, 39 ±0%; SK1-15NM/U1α-CGC, 35 ±0.5%). Cotransfection of a plasmid coding for wild-type U1 snRNA with the SK1-15NM minigene appears to have enhanced the relative expression of exon NM. The reason for this is not known. It may be that the level of wild-type U1 snRNP in the myoblast cells was highest in the SK1-15NM/U1 WT cotransfection: in the other transfections either no, or mutant exogenous U1 snRNA is present. An increase in U1 snRNP concentration may favour use of the NM 5'ss over other 5' splice sites. Alternatively, the fact that the RT-PCR analysis of the SK1-15NM/U1 wild-type cotransfection was done at a different time to that of the other

transfections in the experiment may be important, as suggested by the above comparison of results from cotransfections repeated in separate experiments (table 4.5). The explanation for this anomoly was not pursued further.

Chapter 5
Identification of SK ESE binding proteins (SEBPs)

Effect of mutation of the SK ESE on *in vitro* splicing efficiency

The *in vivo* transfection experiments discussed in chapter 3 identified a sequence (the SK ESE) located adjacent to the SK p5'ss sequence that is required for SK isoform expression in muscle cells. In order to begin to understand the mechanism of activation of SK isoform expression in muscle cells, it was first important to identify any factors binding specifically to the SK ESE and then to show that the binding of such factors to this sequence correlated with enhanced splicing to the SK exon. The SK enhancer sequence is purine-rich. As discussed in the introduction, several other examples of purine-rich exon splicing enhancers (ESEs) have been reported. In some cases these sequences have been shown to bind directly to SR proteins and for this binding to correlate with enhancement of splicing (Sun et al, 1993; Ramchatesingh et al, 1995). Therefore, a member of the SR protein family was a likely candidate to bind to the SK ESE and to stimulate splicing to the SK exon in muscle cells.

A number of *in vitro* binding techniques are used to study the interaction of proteins with RNA. The proteins are obtained by making extracts from cultured cells. The first choice of cell-type to study an exon enhancer active in skeletal muscle cells was the mouse myotube cell-type used in the laboratory for the *in vivo* transfection experiments discussed above. Unfortunately it is very difficult to make extracts from these cells that retain activity in splicing assays *in vitro* . In addition, these cells do not grow in suspension, so a vast culture dish surface area is required to grow enough cells from which to make sufficient quantities of extract. The *in vivo* transfection experiments (chapter 3) established that the SK ESE was active in non muscle cells once the SK exon had been activated by mutation of the branch point sequence upstream of this exon (the SK BPS). It was assumed that the SK ESE is recognised by the same factor(s) in non muscle and muscle cells and that this factor(s) is expressed in all cell types. Therefore, in principle, extracts from any cell-type could be used to study the proposed SK ESE-protein interactions.

HeLa cells are the most common cell-type chosen to make splicing extracts. These cells are non muscle in origin and can be grown to a high density in suspension to produce large quantities of active splicing extract. Before any *in vitro* binding experiments were carried out to examine potential SK ESE-protein interactions it was first necessary to show that the SK ESE is functional in HeLa cell splicing extracts. This was tested by comparing the splicing efficiency in such extracts of a TMnm pre-mRNA containing the SK ESE with that of an identical pre-mRNA lacking this sequence. If the extracts contain an active component(s)

that binds to the SK ESE and stimulates splicing to the SK exon, the efficiency of splicing of the wild-type pre-mRNA should be greater than splicing of an identical pre-mRNA lacking the SK ESE. It was assumed that any factor(s) shown to bind to the SK ESE and stimulate splicing to exon SK would be the same as that involved in stimulating SK isoform expression in muscle cells and in other types of non muscle cell.

Any reduction in splicing efficiency caused by the absence of the SK ESE would be more evident if the splicing efficiency of the wild-type sequence was high. In general, the efficiency of pre-mRNA splicing *in vitro* decreases with increasing intron length and number. The entire TMnm minigene pre-mRNA contains three introns and is ~3 Kb long and was not suitable for use as an *in vitro* pre-mRNA. To maximise splicing efficiency, it was decided to use single-intron pre-mRNA and, specifically, to use pre-mRNA spanning the NM and SK exons and intervening intron sequence (referred to as NM-SK pre-mRNA below). This pre-mRNA is about 150 nucleotides shorter than pre-mRNA spanning the SK and VI exons and intervening intron sequence (referred to as SK-VI pre-mRNA below). Other studies *in vitro* had revealed that the splicing efficiency of SK-VI pre-mRNA was very low (data not shown). Pre-mRNA spanning the IV and SK exons and intervening sequence was not used because it contains two introns: the exon IV-exon NM intervening intron, and the exon NM-exon SK intervening intron.

It was known from earlier experiments using NM-SK pre-mRNA that the NM and SK exons could be spliced together *in vitro* despite their mutually expression use *in vivo*. When the branchpoint sequence (the SK BPS) within this pre-mRNA was mutated to the consensus sequence, the efficiency of splicing was much increased (Ayres, unpublished). Because any effect of mutation of the SK ESE on splicing efficiency would be more obvious if the splicing efficiency of the wild-type pre-mRNA was high, it was decided to use NM-SK pre-mRNA containing a consensus SK BPS (this pre-mRNA is referred to as NM-SKWTbpup pre-mRNA below) in these studies. Another reason for using transcripts containing the consensus SK BPS was that the SK ESE was only active in non muscle cells after the SK BPS had been changed to the consensus sequence (Graham et al, 1992; Hamshere, PhD thesis; and chapter 3 of this study). If HeLa cells regulate the expression of TMnm pre-mRNA in the same way as the cell types used in the previous experiments, the SK ESE is expected to be active only when the SK BPS is converted to the consensus sequence.

Two mutants of the SK ESE were used to examine the effect of the wild-type sequence on splicing efficiency of NM-SKWTbpup pre-mRNA in HeLa cell nuclear extracts. The sequence of these mutants was the same as the sequence of two mutants previously described in chapter 3 and tested *in vivo*. One mutant was referred to as SK16-30NM and the other as SK16-30R1. TMnm minigenes

containing these mutant SK exon sequences did not express significant levels of the SK isoform in non muscle cells (the SK16-30NMbpup and SK16-30R1bpup minigenes). Similarly, a TMnm minigene containing the SK16-30NM mutant SK exon sequence did not express a significant level of the SK isoform in muscle cells (the SK16-30NM minigene). Pre-mRNA containing these mutant sequences is referred to below as NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNA. It was hoped that the efficiency of splicing of NM-SKWTbpup pre-mRNA containing a wild-type SK exon would be higher than that of the NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNAs if the SK ESE is active in HeLa cell nuclear extracts.

Radioactively labelled NM-SKWTbpup, NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNAs were incubated with HeLa cell nuclear extract under splicing conditions. The products of splicing were analysed by denaturing polyacrylamide gel electrophoresis. The results are shown in figure 5.1. The pre-mRNAs are visible at the bottom of the figure and the lariats (intermediates and products in the splicing reaction), which run anomalously compared to their linear equivalents, are visible above them. Other intermediates and products of the splicing reaction are much shorter than the pre-mRNAs and lariat RNAs. They run much faster on the gel and were too faint to quantify accurately. They are omitted from the figure.

It will be noted that there are a total of four different lariat RNAs visible for each pre-mRNA instead of the two expected lariat RNAs (the lariat intermediate and lariat product). The extra lariats are likely to have arisen from use of two different branch point nucleotides in the splicing reactions or from 3'-exonuclease activity, resulting in lariats of different mobility or length, respectively. The exact explanation was not determined because the aim of the experiment was to test only the activity of the SK ESE sequence in HeLa cell nuclear extracts; use of the correct branch point nucleotide was of secondary importance.

The relative levels of lariats as a percentage of the total of lariats and pre-mRNA for each pre-mRNA was calculated using the PhosphorImager. Table 5.1 shows the averaged results for each pre-mRNA.

Figure 5.1 Effect of mutation of the SK exon purine-rich sequence on single intron TM pre-mRNA splicing in vitro

Radioactively labelled single intron pre-mRNA containing the NM exon, SK exon (wild-type or SK16-30R1 mutation) and the intervening intron sequence (including a consensus branch point sequence, bpup) was incubated with HeLa cell nuclear extract, under splicing conditions for 0, 1.5, or 3 hours. The RNA was then precipitated and analysed by denaturing polyacrylamide gel electrophoresis. The gel was fixed, dried and exposed to a PhosphorImager cassette. The figure shows the PhosphorImager scan obtained. Only the pre-mRNA and the lariats are shown.
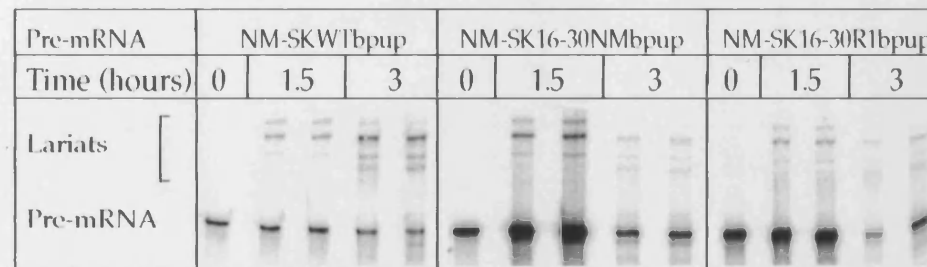
	Incubation time	
Pre-mRNA	1.5 hrs	3 hrs
NM-SKWTbpup	19	49
NM-SK16-30NMbpup	9	10
NM-SK16-30R1bpup	5	16

Table 5.1 Splicing efficiency of different pre-mRNAs in HeLa cell nuclear extract. The level of lariat is shown as a percentage of the total of pre-mRNA and lariats. Values are averages from duplicate experiments.

The results show that the NM-SKWTbpup pre-mRNA was spliced much more efficiently than the NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNAs. 19% lariat formation had occurred with the NM-SKWTbpup pre-mRNA after 1.5 hours compared with only 9% and 5% with the NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNAs. After 3 hours, the percentage lariat formation with NM-SKWTbpup pre-mRNA had increased to 49% compared with an increase to only 10% and 16% with the NM-SK16-30NMbpup and NM-SK16-30R1bpup pre-mRNAs, respectively (approximately 20% that of the NM-SKWTbpup pre-mRNA level). It was concluded that the HeLa cell splicing extract used in this experiment contained components that could recognise the SK ESE and that recognition of the SK ESE by these components stimulated splicing of the NM exon to the SK exon.

Splicing of exon NM to exon SK is not normally observed *in vivo*. However, the relevance of the above results to splicing *in vivo* is suggested by the fact that mutation of the SK ESE had a similar effect on splicing to the SK exon *in vitro* as was observed *in vivo*. It was assumed that the factor(s) involved in recognising the SK ESE and stimulating splicing to exon SK in HeLa cell splicing extract was the same as the factor(s) that recognises the SK ESE when the SK exon is expressed in muscle cells and in other types of non muscle cell.

Binding of components of HeLa cell splicing extract to RNA containing the SK ESE

Having shown that the SK ESE was able to function in splicing *in vitro* as part of NM-SKWTbpup pre-mRNA, it was decided to carry out experiments using HeLa cell splicing extracts in order to try to identify any components which specifically recognise the SK ESE. The differences in splicing efficiency of the NM-SKWTbpup transcript containing wild-type SK exon and the equivalent transcript containing mutant exon suggested that the two types of pre-mRNA would differ in their ability to assemble spliceosome complexes in HeLa cell splicing extract. Purification of any complexes specific to SK ESE-containing pre-mRNA would

then provide a way of identifying factors that bound directly to the SK ESE or were stimulated to assemble onto the NM-SKWTbpup RNA as a result of the presence of the SK ESE.

Complexes formed when radioactively labelled RNAs are added to splicing extract can be visualised after electrophoresis under non denaturing conditions. In these experiments labelled RNA is run on agarose gels at the same time as similar labelled RNA that has been incubated with splicing extract. The mobility of any RNA which has components stably associated with it is retarded relative to the free RNA. This technique is known as a gel retardation, or gel mobility shift assay. Any complexes specific to the SK ESE-containing RNA could then be identified and potentially purified from the gel. The NM-SKWTbpup and NM-SK16-30R1bpup pre-mRNAs used in the *in vitro* splicing assays described above were incubated under splicing conditions, in the presence or absence of yeast tRNA. The yeast tRNA acts as a non specific competitor RNA and binds any proteins that would otherwise bind non specifically to the labelled RNAs and may obscure the results. Complexes formed on each RNA were then analysed by non denaturing gel electrophoresis. The result is shown in figure 5.2.

The distribution of radioactivity in lanes 2, 3, 5, and 6 of figure 5.2 is shown in figure 5.3. This provides a representation of the complexes formed on the labelled pre-mRNAs. The profile from left to right represents complexes with low to high mobility on the non dentauring agarose gel, and therefore, with high to low molecular weight. Comparison of the distribution of radioactivity in lanes 2 and 5 shows that there is possibly a high molecular weight complex (marked by the arrow) which is formed more efficiently on NM-SKWTbpup pre-mRNA than on NM-SK16-30R1bpup pre-mRNA in the absence of yeast tRNA. This complex was apparently not stable to the presence of 5 μg yeast tRNA (lanes 3, 6 figures 5.2 and 5.3). However, because the resolution of the native gel was so poor, it was concluded that this experiment gave no strong evidence that the SK ESE was able to stimulate the formation of specific complexes on NM-SKWTbpup pre-mRNA.

It is clear from figure 5.2 (lane 2) that the resolution obtained on the native gel was not good enough to allow purification of the complex that was possibly specific to SK ESE-containing RNA. An alternative *in vitro* binding technique was used therefore to look at the differences in components binding to the NM-SKWTbpup wild-type and mutant SK exon pre-mRNAs. Proteins in direct contact with RNA can be crosslinked to that RNA using shortwave UV light. In this way proteins binding specifically and directly to the SK ESE could potentially be identified. Components that are part of complexes assembled onto the RNA but which are not in direct contact with the RNA are invisible using this technique. It is possible that proteins binding specifically to the SK ESE would be more easily identified by a crosslinking experiment than by a gel retardation experiment. Therefore, radioactively labelled NM-SKWTbpup or NM-SK16-

Figure 5.2 Effect of mutation of the SK ESE on the ability of NM-SKbpup pre-mRNA to form complexes in HeLa cell nulcear extract

Continuously labelled NM-SKWTbpup and NM-SK16-30R1bpup pre-mRNAs were incubated under splicing conditions, in the presence or absence of yeast tRNA, for 30 minutes at 30°C. The samples were then placed on ice and loaded onto a 0.8% agarose gel. The gel was run overnight at 4°C then fixed, dried and analysed using the PhosphorImager. The figure shows the PhosphorImager scan of the dried gel.

SK WT			SK16-30R1			NM-SKbpup RNA
–	+	+	–	+	+	Extract
–	0	5	–	0	5	Yeast tRNA (µg)
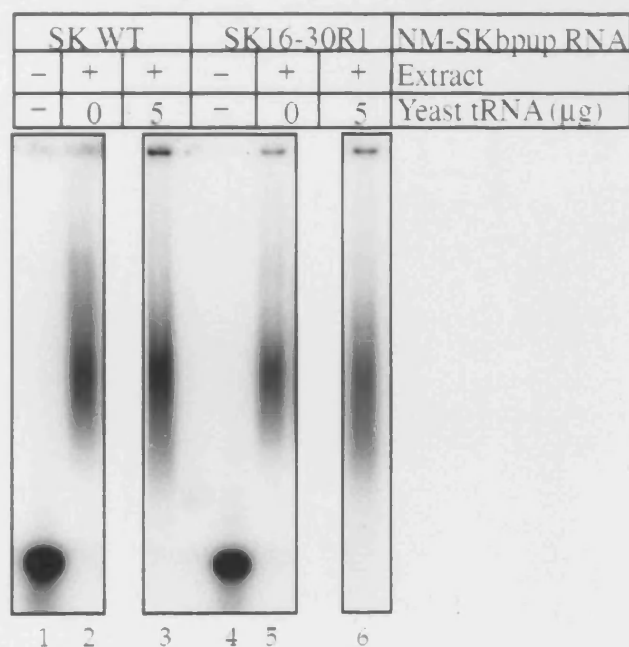
Figure 5.3 PhosphorImager profile for distribution of radioactivity in lanes 2, 3, 5, and 6 of figure 5.2

The level of radioactivity is shown on an arbitrary scale to facilitate comparison between profiles. The significance of the complex labelled by the arrow is discussed in the text.
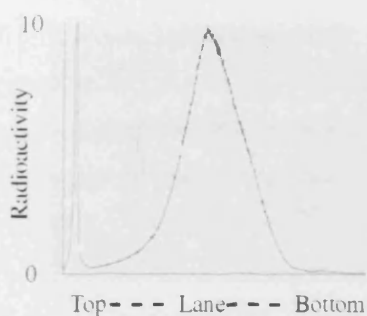
Lane 2

NM-SKWTbpup
Extract
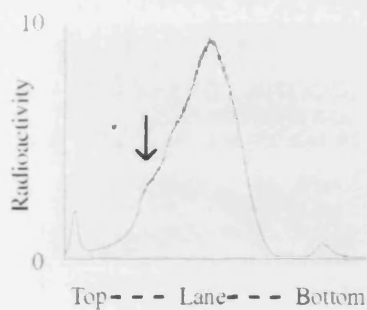
Lane 3

NM-SKWTbpup
Extract
5µg yeast tRNA

Lane 5

NM-SK16-30R1bp
up
Extract
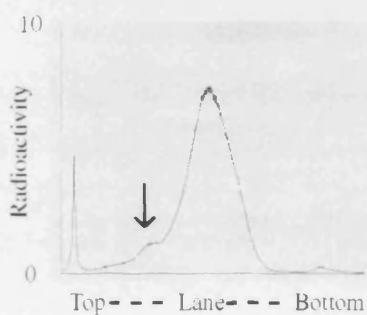
Lane 6

NM-SK16-30R1bp
up
Extract
5µg yeast tRNA
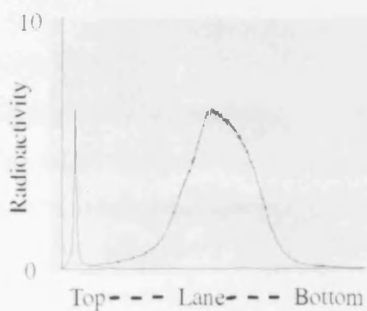
30R1bpup pre-mRNA was incubated under splicing conditions with HeLa cell splicing extract. The incubations were then irradiated on ice with shortwave UV light, digested with RNase to remove RNA not protected by bound protein, and run on an SDS-polyacrylamide gel. Proteins present on the gel were blotted onto nitrocellulose. Figure 5.4 shows the PhosphorImager scan of the nitrocellulose filter obtained.

Three different HeLa cell nuclear extracts were tested. There are no clear differences between the proteins cross-linked to the NM-SKWTbpup or NM-SK16-30R1bpup pre-mRNAs in any of the extracts. Thus, the UV cross-linking technique gave no information as to the identity of a factor(s) binding specifically to the SK ESE. However, the efficiency of shortwave UV crosslinking of proteins to RNA is known to be well below 100% and crosslinking is thought to occur specifically to uridine residues in the RNA. A possible explanation for the lack of any visible differences in proteins crosslinking to the NM-SKWTbpup or NM-SK16-30R1bpup pre-mRNAs in the experiment described above is that the factor(s) that binds to the SK ESE is not abundant enough for it to be observably crosslinked to the SK ESE under these conditions and that any signal specific to binding of proteins to the SK ESE is swamped by signals from proteins binding to uridine-rich sequence elsewhere in the RNA. Thus, an SK ESE binding protein(s) is expected to be best visualised using extract preparations which contain enhanced amounts of proteins thought to bind to the SK ESE (i.e. SR proteins) and shorter RNA than was used above.

Binding of components of an SR protein preparation to RNA containing the SK ESE

SR proteins can be purified from HeLa cells using a two-step magnesium sulphate precipitation procedure (Zahler et al, 1992). These preparations contain much higher concentrations of SR proteins than HeLa splicing extract. Given that the SK ESE is likely to be recognised by an SR protein (see introduction and above), specific binding of a component to the SK ESE was much more likely to be visible in a crosslinking experiment using a preparation of purified SR proteins than HeLa splicing extract.

A preparation of SR proteins purified from HeLa cells was kindly supplied by A.Mayeda. The activity of these proteins was tested by their effect on the splicing pattern of a mutant of the adenovirus E1a pre-mRNA transcript (E1a 2,1). The splicing pattern of this substrate had previously been shown to be affected by the concentration of active SF2/ASF (Eperon et al, 1993). Thus, E1a 2, 1 pre-mRNA was incubated under splicing conditions with HeLa cell splicing extract and increasing amounts of the SR protein preparation. After one hour the reactions were stopped and the spliced products were analysed by denaturing

Figure 5.4 Effect of mutation of the SK ESE on shortwave UV crosslinks formed between radioactively labelled NM-SKbpup pre-mRNA and components of HeLa cell nuclear extract

Continuously labelled NM-SKWTbpup or NM-SK16-30R1bpup pre-mRNA was incubated in HeLa cell nuclear extract under splicing conditions for 30 minutes at 30°C. The incubations were then irradiated for 30 seconds on ice using shortwave UV light and digested with RNAse. The samples were then run on a 10% SDS-polacrylamide gel and blotted onto a nitrocellulose filter. The filter was dried and analysed using the PhosphorImager. The figure shows the PhosphorImager scan of the dried filter.

Three different preparations of HeLa cell nuclear extract were used. They were prepared using the same method, but at different times and from different batches of HeLa cells.
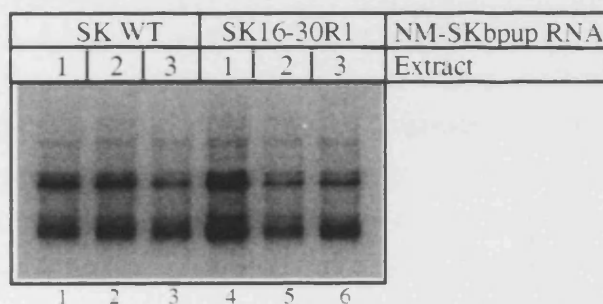
polyacrylamide gel electrophoresis. The result is shown in figure 5.5. The figure shows that increasing the amount of the SR protein preparation in the splicing reaction caused a change in the pattern of bands obtained. In particular, bands B and C are present in higher relative amounts in lane 5 than in the other lanes. This change could be explained by the addition to the splicing extract of SR proteins active in splice site switching ability, or by the presence of a nuclease in the SR protein preparation. The identity of each band was not confirmed, but they are consistent with their identification as linear intermediates and products of the splicing reaction. It was concluded that the SR protein preparation possibly had an active ability to switch relative use by the splicing apparatus of two alternative splice sites. It was assumed that if the SR proteins were active, any ability of these proteins to interact with exon enhancer sequences, such as the SK ESE, was also intact. This SR protein preparation was used, therefore to determine whether any SR proteins could be observed to bind specifically to the SK ESE.

The specificity of the interaction between proteins and a radioactively labelled RNA can be demonstrated by the ability of unlabelled RNA to compete with the labelled RNA for binding to those proteins. If binding to the labelled RNA is specific, unlabelled RNA of the same sequence as the labelled RNA will be a more effective competitor than unlabelled RNA of similar length that lacks the sequences required for specific binding. The SR protein preparation was incubated with radioactively labelled wild-type SK exon RNA under splicing conditions. The incubation was then split into seven equal-sized aliquots and either water, or unlabelled wild-type SK exon SK16-30R1 exon, or NM exon RNA at a low or high concentration was added to each aliquot. Each sample was irradiated with shortwave UV light, digested with RNase then run on an SDS-polyacrylamide gel and blotted onto nitrocellulose as before. Figure 5.6A shows the PhosphorImager scan of the dried nitrocellulose filter.

Any proteins binding specifically to the SK ESE were expected to be competed off the labelled RNA more effectively by unlabelled wild-type SK exon RNA than by unlabelled SK16-30R1 mutant SK exon RNA or NM exon RNA. Proteins binding specifically to SK exon sequences other than the SK ESE sequence were expected to be competed off the labelled wild-type SK exon RNA more effectively by the unlabelled wild-type or SK16-30R1 SK exon RNA than by the unlabelled NM exon RNA. The effect of the competitor RNAs was therefore predicted to distinguish proteins binding specifically to the SK ESE from those binding specifically to other sequences in the SK exon.

Three radioactively labelled bands (A, B, and C) are visible in lane 1 of figure 5.6A (no competitor RNA). These bands were interpreted to represent proteins that crosslinked to the labelled wild-type SK exon RNA. The relative levels of the three bands, as determined using the PhosphorImager, are shown in table 5.2 and figure 5.6B.

Figure 5.5 Effect of increasing amounts of an SR protein preparation on splicing of adenovirus E1a 2,1 pre-mRNA

Radioactively labelled E1a 2,1 pre-mRNA was incubated with HeLa cell nuclear extract, and increasing amounts of an SR protein preparation, under splicing conditions for one hour. The RNA was then precipitated and analysed by denaturing polyacrylamide gel electrophoresis. The gel was fixed, dried and exposed to a PhosphorImager cassette. The figure shows the PhosphorImager scan obtained. Products and intermediates of the splicing reaction are labelled A-G.
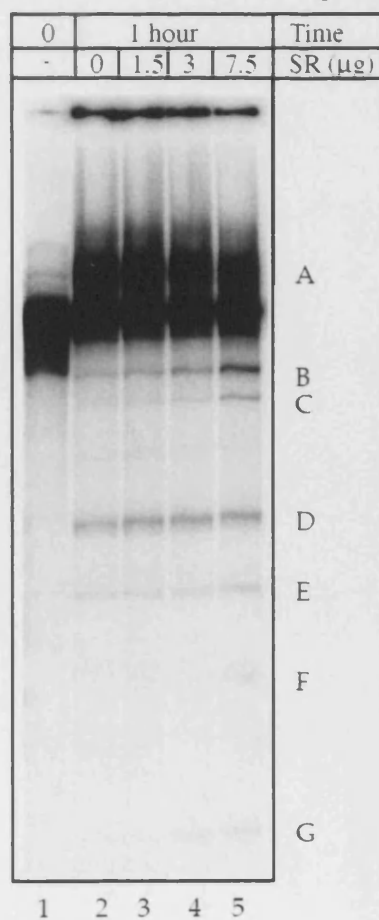
Figure 5.6. Effect of unlabelled competitor RNA on crosslinking of components of an SR protein preparation to radioactively labelled wild-type SK exon RNA.

A) An SR protein preparation was incubated with continuously labelled wild-type SK exon RNA under splicing conditions for 30 minutes at 30°C. The incubation was then split into seven equal aliquots and either water or unlabelled competitor RNA at a low or high concentration was added to each aliqout. Each sample was irradiated with shortwave UV light for 30 seconds, digested with RNAse, then run on an SDS-polacrylamide gel and blotted onto nitrocellulose. The filter was dried and analysed using the PhosphorImager. The figure shows the PhosphorImager scan of the dried filter. A, B, and C refer to the bands produced. The numbers are the sizes (in kilodaltons, Kd) of radioactively labelled marker proteins loaded in lane 8.
WT=wild-type SK exon RNA. R1=SK16-30R1 SK exon RNA. NM=wild-type NM exon RNA. The amount of each competitor used was either the same as the labelled wild-type SK exon RNA (1x), or a 10-fold increase (10x).

B) Graph of relative levels of bands A, B, and C. Values were calculated using the PhosphorImager (see text).
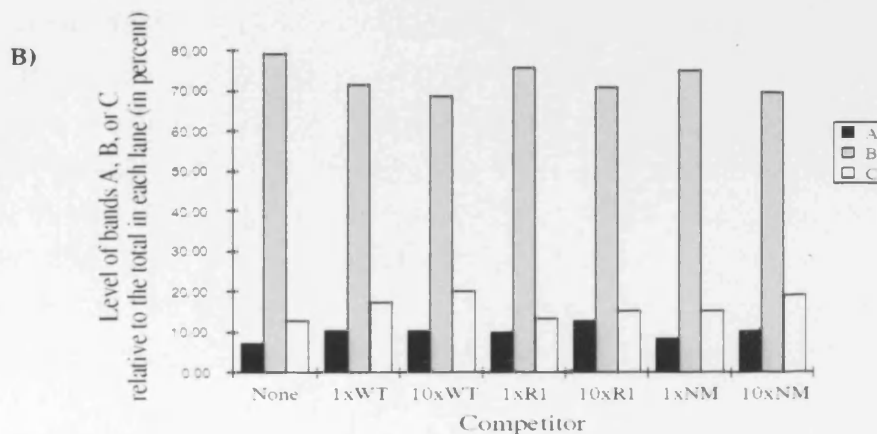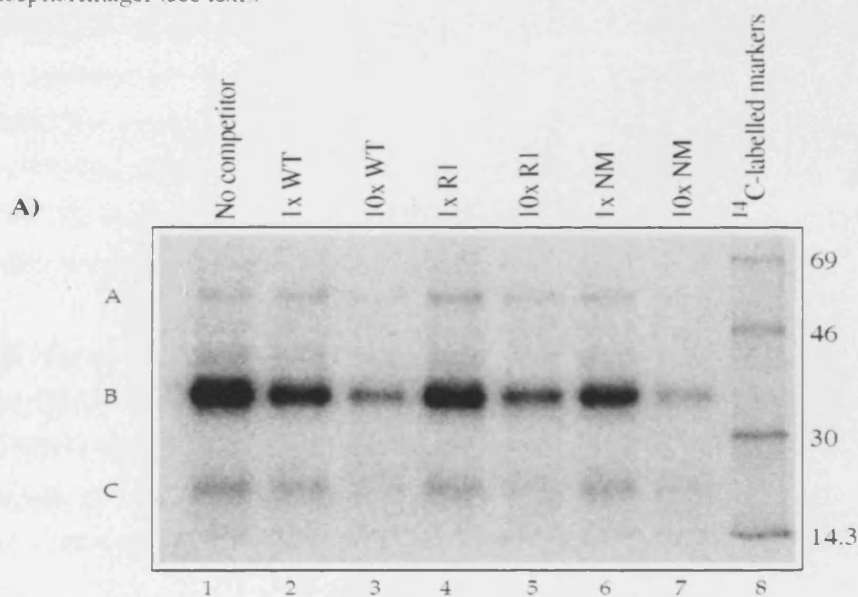
Competitor RNA	A (% of A+B+C)	B (% of A+B+C)	C (% of A+B+C)
None	7	79	13
1xWT SK	11	72	18
10xWT SK	11	69	20
1xSK16-30R1	10	76	14
10xSK16-30R1	13	71	16
1xWT NM	9	76	16
10xWT NM	11	70	19

Table 5.2 Effect of unlabelled competitor RNAs on the crosslinking of components of an SR protein preparation to labelled wild-type SK exon RNA

Figures represent the levels of each band (A, B, or C) relative to the total and are shown in percent. The amount of unlabelled competitor RNA added, relative to the amount of labelled SK exon RNA, is shown as 1x, and 10x. Wild-type SK exon RNA is referred to as WT SK. SK exon RNA in which nucleotides 16-30 have been mutated is referred to as SK16-30R1. Wild-type NM exon RNA is referred to as WT NM.

It is clear from figure 5.6A that as the concentration of unlabelled competitor RNA was increased, the level of crosslinking to the labelled RNA decreased for each competitor (compare lanes 2 and 3, 4 and 5, 6 and 7). The relative levels of the three labelled bands are largely unaffected by any of the unlabelled competitor RNAs. It was concluded that the proteins in this preparation of SR proteins do not bind specifically to wild-type SK exon RNA, to the SK ESE or to wild-type NM exon RNA. Specific binding of all three proteins to the labelled RNA would result in the relative levels of the three proteins remaining the same when specific competitor RNA was added. However, in this situation, the amount of each protein crosslinked to labelled RNA would be expected to be significantly reduced by the presence of specific competitor compared to the amount of each protein in the presence of non-specific competitor. The levels of each protein crosslinked to the labelled RNA are similar in all the 1x competitor lanes (lanes 2, 4, 6) and also in all the 10x competitor lanes (lanes 3, 5, 7), so such an explanation is unlikely.

Effect of combining an SR protein preparation with cell or splicing extract on binding of components to SK ESE-containing RNA

There are a number of possible explanations for the lack of specific binding by SR proteins to the SK ESE, some of which are given below:

1) SR proteins do not interact with the SK ESE to stimulate splicing to exon SK;

2) The SR protein(s) required for binding to the SK ESE is not active/present in the SR protein preparation that was tested;

3) SK exon RNA was folded in a conformation that did not allow recognition of the SK ESE by any of the SR proteins in the preparation;

4) The purified SR proteins tested have lost a modification (such as a phosphorylation) during the purification procedure that has affected their ability to bind specifically to the SK ESE;

5) SR proteins do not bind directly to the SK ESE but interact with a non SR protein that binds specifically to the SK ESE;

6) RNA used for the crosslinking experiment did not contain the branch point sequence upstream of the SK exon (the SK BPS). The SK ESE was only active in non muscle cells in the presence of a consensus SK BPS (chapter 3) so it is possible that the consensus sequence is required before any specific binding to the SK ESE can occur.

The addition of a small amount of splicing extract to the SR protein preparation could, in theory, overcome the potential problems outlined by explanations (3), (4) and (5) above. Helicase enzymes within the splicing extract would unwind any RNA secondary structure which blocked binding to the ESE, kinase enzymes would re-phosphorylate de-phosphorylated SR proteins, and if an adaptor protein is required it must be present in the extracts otherwise the SK ESE would not have been able to stimulate splicing of NM-SKbpup pre-mRNA. If any of these three explanations are true, specific binding of SR proteins to the SK ESE could potentially be observed if the *in vitro* binding experiments were repeated using SR proteins combined with a small amount of splicing extract. An SR protein(s) bound to the SK ESE via an adaptor protein may not directly contact the RNA. Such interactions would not be identified by crosslinking of proteins to RNA but should be detected by gel retardation of complexes associated with labelled RNA. The gel retardation assay was used therefore to examine the binding of SR proteins in the presence of HeLa cell splicing extract to the SK ESE.

A technical difficulty associated with the previous gel retardation experiment (figure 5.2) was that the resolution was too poor to allow accurate purification of any SK ESE-specific complexes. Resolution may be improved if the length of the labelled RNA is kept to a minimum. For this reason, a shortened substrate containing only a few nucleotides upstream of the SK exon and the exon itself (i.e. the SK ESE and its immediate sequence environment, 106 nucleotides in total length) was used for the experiment. Unfortunately, the activity of the SK ESE in this shortened sequence could not be tested in the same way that the effect of the SK ESE on NM-SKbpup splicing activity was determined, because the truncated sequence does not contain a complete intron.

Therefore it could only be assumed that the SK ESE was active as part of this RNA.

The gel retardation assay was also used to determine whether HeLa cell splicing extract, myoblast or myotube extracts contained components that bind specifically to the SK ESE. This was feasible because only small quantities of extract are required for such experiments. Whole cell myoblast and myotube extracts were used. These are not active in splicing but were made using a very rapid procedure designed to minimise denaturation of proteins and breakdown of components by RNAses and proteases released after cell lysis (see chapter 2). It was hoped that the myotube extracts may have retained an ability to assemble complexes specifically on, or to stimulate SR protein binding specifically to, the 106 nucleotide SK ESE-containing RNA described above, in comparison with complexes formed in myoblast extracts on identical RNA, or in myotube extracts on similar RNA lacking the SK ESE.

The 106 nucleotide RNA that was used contained either wild-type SK exon sequence (referred to as SK WT (85) RNA below) or mutant SK exon sequence lacking the SK ESE and identical to the mutant SK exon sequence in the SK16-30R3 TMnm minigene (referred to as SK16-30R3 (85) RNA below). Labelled wild-type or SK16-30R3 85 nucleotide RNA was incubated under splicing conditions with SR proteins, HeLa cell splicing extract, myoblast cell extract, or myotube cell extract, either alone or in combination. The combinations used are shown in table 5.3. Assembled complexes were analysed by native gel electrophoresis (the gel used contained agarose and acrylamide because it was thought that the resolution might be improved by the presence of acrylamide). The result is shown in figure 5.7.

RNA	SRs	HeLa	Mb	Mt	RNA	SRs	HeLa	Mb	Mt
SK WT (85)	-	-	-	-	SK16-30R3 (85)	-	-	-	-
"	+	-	-	-	"	+	-	-	-
"	-	+	-	-	"	-	+	-	-
"	-	-	+	-	"	-	-	+	-
"	-	-	-	+	"	-	-	-	+
"	+	+	-	-	"	+	+	-	-
"	+	-	+	-	"	+	-	+	-
"	+	-	-	+	"	+	-	-	+

Table 5.3 Combinations of SR proteins and extracts used in the gel retardation experiment of figure 5.7.

The resolution obtained by the agarose/acrylamide gel was again poor, but despite this several interesting observations were made.

Figure 5.7. Effect of combining SR proteins with extracts on assembly of complexes onto labelled RNA containing wild-type or mutant SK exon.

Labelled RNA comprising wild-type or SK16-30R3 mutant SK exon RNA and ~10 nucleotides of the preceding intron (SKWT (85) and SK16-30R3 (85) RNA) was incubated under splicing conditions with various extract combinations for 30 minutes at 30°C. The samples were then incubated for a further five minutes at 30°C in the presence of heparin, placed on ice, and loaded onto an agarose/acrylamide gel. The gel was run at 4°C overnight then fixed, dried and analysed using the PhosphorImager. The figure shows the PhosphorImager scan obtained.
The dashed boxes show regions of the scan in which PhosphorImager radioactivity profiles were taken (figure 5.8).
SR=SR protein preparation; Mb=myoblast whole cell extract; Mt=myotube whole cell extract; HeLa=HeLa cell splicing extract.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
SK (85) RNA	SK WT	+	+	+	+	+	+	+	+								
	SK16-30R3									+	+	+	+	+	+	+	+
Extract	SR		+				+	+	+		+				+	+	+
	Mb			+			+					+			+		
	Mt				+			+					+			+	
	HeLa					+			+					+			+
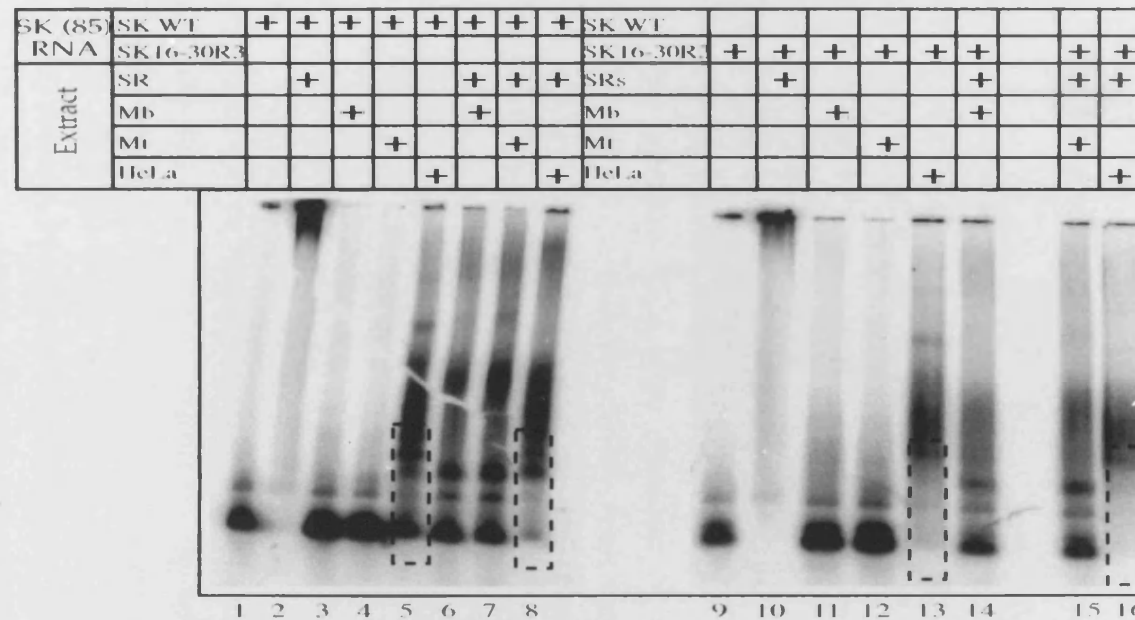
1) The figure shows that a relatively large amount of radioactivity is present in the region of the wells of lanes 2 and 10. These lanes represent incubation of SR proteins with wild-type and mutant labelled RNAs, respectively. The location of this radioactivity was interpreted to mean that complexes containing labelled RNA had formed that were too large to enter the gel. It is likely that these complexes are the result of coagulation of positively charged denatured proteins with the negatively charged labelled RNA and not binding of functional proteins to the RNA. Unfortunately some degree of denaturation is inevitable in preparations of proteins made from cells. For this reason it is essential to check that these preparations contain some active protein before use. It was concluded that proteins in the SR preparation do not form stable native complexes with the labelled wild-type or mutant RNAs under the conditions of the experiment.

2) Lanes 3, 4, 11, and 12, representing incubation of the wild-type or mutant labelled RNAs with either myoblast or myotube extracts alone, do not differ significantly from the RNA only lanes (1 and 9). Thus, components of these extracts were not capable of binding stably to either of the labelled RNAs under these conditions. The low amount of radioactivity in the wells of these lanes suggests that the level of denatured protein in the myoblast and myotube extracts was very low compared to that in the SR protein preparation.

3) The mobility of the labelled RNA was retarded in the lanes containing wild-type or mutant RNA and HeLa splicing extract alone (lanes 5 and 13). The PhosphorImager radioactivity profiles of the boxed regions of these lanes are shown in figure 5.8 and reveal differences in the complexes assembled on the two RNAs. As with the previous gel retardation experiment (figure 5.2) wild-type RNA was able to support more efficient assembly of higher molecular weight complexes than RNA lacking the SK ESE. Again, the resolution was not good enough to allow purification of the complexes specific to the SK ESE-containing RNA from the gel.

4) Addition of myoblast or myotube extracts to the SR protein preparation had a dramatic effect on the assembly of complexes on the labelled RNAs (lanes 6, 7, 14, 15). SR proteins alone or myoblast or myotube extracts alone were not able to support formation of stable native complexes on either of the labelled RNAs (lanes 3, 4, 11, 12). However, combination of either of the extracts with the SR protein preparation stimulated complex assembly onto the wild-type and mutant RNAs. This confirms that the SR protein preparation contains some activity, although it does not confirm that the preparation is active for splicing. It was not possible to determine from the gel whether the wild-type and mutant RNAs or the myoblast and myotube extracts differed in their ability to assemble complexes.

Figure 5.8 PhosphorImager profile for distribution of radioactivity in lanes 5, 8, 13, and 16 of figure 5.7

The level of radioactivity is shown using an arbitrary scale to facilitate comparison between profiles. The complex labelled by the arrow is discussed in the text.

Lane 5
Wild-type RNA
HeLa extract

Lane 13
Mutant RNA
HeLa extract

Lane 8
Wild-type RNA
HeLa extract
SR proteins

Lane 16
Mutant RNA
HeLa extract
SR proteins
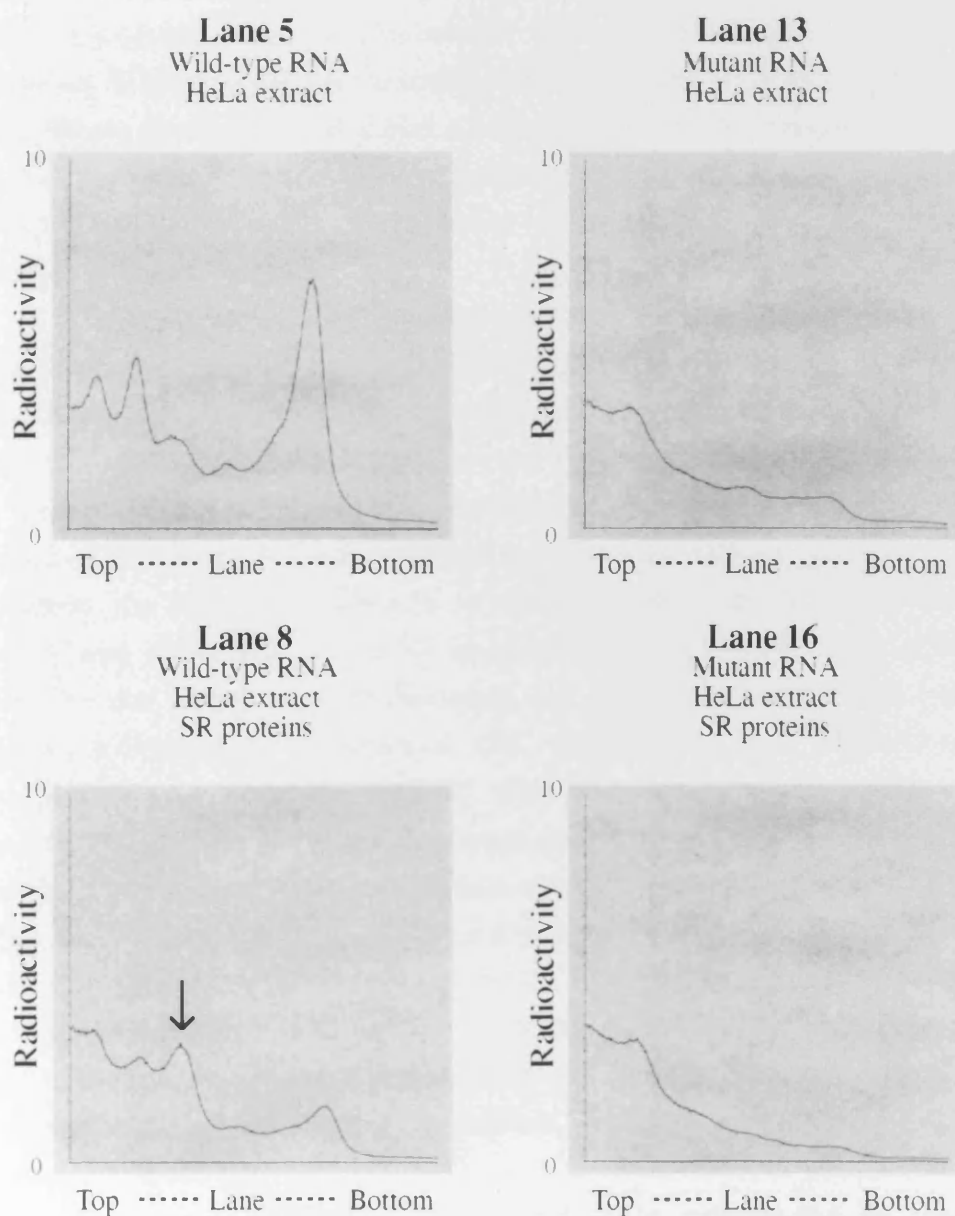
5) The radioactivity profile of the boxed regions of lanes 8 and 16 is shown in figure 5.8. The profile corresponding to lane 8 (containing wild-type RNA, HeLa cell nuclear extract and SR proteins) has a peak marked with an arrow. This peak is specific to the wild-type RNA (compare the profile of lane 8 with that of lanes 13 and 16). The amount of radioactivity in this peak is specifically enhanced by the presence of SR proteins (compare the profile of lane 8 with that of lane 5).

It was concluded that combination of SR proteins with HeLa cell splicing extract facilitated the formation of a complex specific to RNA containing the SK ESE. This effect is consistent with the proposal that a component(s) of HeLa cell splicing extract allowed the specific interaction of an SR protein(s) with the SK ESE. However, other explanations are also consistent with this result. For example, it is possible that a component(s) of the SR protein preparation was able to facilitate the binding of a HeLa cell splicing extract component(s) to the SK ESE, or that an extract component(s) had stimulated the binding of a non SR protein component(s) of the SR preparation to the SK ESE. This is discussed further in chapter 6.

The resolution of the agarose/acrylamide gel was too poor to allow purification of the SK ESE-specific complex formed in the incubation with wild-type RNA, SR proteins and HeLa cell splicing extract (figure 5.7, lane 8; marked with the arrow in figure 5.8). It was not clear whether there were any differences in the complexes formed on wild-type and mutant RNA in the lanes that included SR proteins and myoblast or myotube extract (lanes 6, 7, 14, and 15) or whether the myotube extracts in comparison with the myoblast extracts had shown any difference in ability to assemble complexes on the wild-type RNA. In order to see if any such differences did exist, the incubations were repeated as before, followed by shortwave UV crosslinking and SDS-polyacrylamide gel electrophoresis of each sample. The results obtained using shortwave UV crosslinking of proteins to RNA were expected to be clearer than the results from the gel retardation experiment because only components bound directly to the RNA are visible. Disadvantages of this method are that components of complexes that are not directly associated with the RNA are invisible and complexes assembled on the RNA cannot be purified. However, the approximate sizes of proteins crosslinked to the RNA can be deduced from their mobility on SDS-polyacrylamide gels relative to marker proteins. The actual results obtained are shown in figure 5.9.

The bands visible in figure 5.9 were interpreted to represent protein crosslinked to radioactively labelled RNA. The amount of the band which correlates to a protein running with a mobility of ~35Kd is shown in table 5.4A for lanes 1-3, 5-7, 9, 10, 12, and 13. The relative levels of the bands in lanes 4, 8, 11, and 14 are shown in table 5.4B. Tables 5.4A and B show that the pattern of

Figure 5.9 Specific crosslinking of components of an SR protein preparation and extracts to radioactively labelled TMnm RNA containing the SK ESE.

An SR protein preparation and/or extract were incubated under splicing conditions with continuously labelled RNA containing the SK exon (wild-type or SK16-30R3 mutant) and ~10 nucleotides of the preceding intron (SK WT (85) andSK16-30R3 (85) RNA). After 30 minutes, heparin was added to remove non specifically bound proteins and the samples were incubated for a further five minutes. The samples were then irradiated with shortwave UV light, digested with RNAse, and the crosslinked products were analysed as described before. The figure shows the PhosphorImager scans of the nitrocellulose filter obtained.

Sizes, in Kd, of radioactively labelled marker proteins are shown. The labels A-E refer to bands present in lans 4, 8, 11, and 14 and are discussed in the text.

SR=SR protein preparation; Mb=myoblast whole cell extract; Mt=myotube whole cell extract; HeLa=HeLa cell splicing extract.

RNA	SK WT (85)				SK16-30R3 (85)			
SR	+				+			
Mb		+				+		
Mt			+				+	
HeLa				+				+

RNA	SK WT (85)			SK16-30R3 (85)		
SR	+	+	+	+	+	+
Mb	+			+		
Mt		+			+	
HeLa			+			+
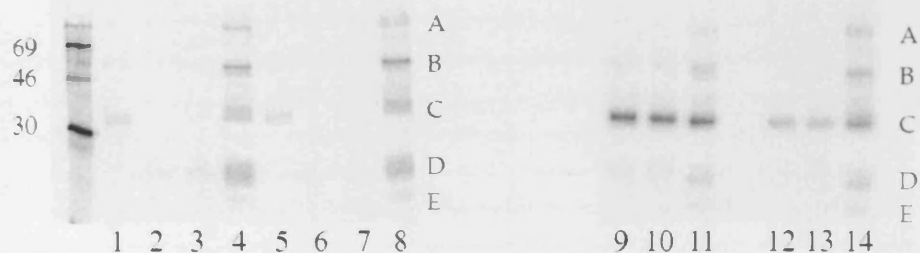
The relative levels of bands A-E formed after incubation of wild-type RNA with SR proteins and HeLa cell nuclear extract were different to those formed in similar incubations using mutant RNA (table 5.4B). In particular, the relative level of the ~35Kd band (band C, figure 5.9A) was 71% in the wild-type RNA lane (11) compared with only 52% in the mutant RNA lane (14). This is in contrast to the relative levels of bands A-E formed after HeLa cell nuclear extract alone was incubated with wild-type or mutant RNA (lanes 4, 8). The relative levels of these bands are very similar in the wild-type and mutant lanes (table 5.4B). It was concluded that addition of myoblast or myotube cell extract or HeLa cell splicing extract to the SR protein preparation had allowed a ~35Kd protein to crosslink specifically to SK ESE-containing RNA. The presence of a band of similar mobility to the ~35Kd band in the SR protein preparation only lanes (1, 5) and its absence from the myoblast or myotube extract only lanes (2, 3, 6, 7) suggests that a ~35Kd protein present in the SR protein preparation was able to interact specifically with RNA containing the SK ESE only in the presence of extract. The apparent mobility of this protein is consistent with its identity as an SRp30 protein. There are, however a number of other possible interpretations of the data. These are discussed further in chapter 6.

Two different methods were used in the experiments described above to detect proteins which bind specifically to SK ESE-containing RNA (figures 5.7 and 5.9A). Similar conclusions can be drawn from each experiment:

1) components of the myoblast and myotube extracts did not form stable complexes with either of the labelled RNAs used (figure 5.7, lanes 3, 4, 11, 12; figure 5.9A, lanes 2, 3, 6, 7);

2) components of the SR preparation did not bind specifically to RNA containing the SK ESE (figure 5.7, lanes 2, 10; figure 5.9A, lanes 1, 5);

3) combination of myoblast or myotube cell extract with the SR protein preparation stimulated binding of complexes to both labelled RNAs (figure 5.7, lanes 6, 7, 14, 15; figure 5.9A, lanes 9, 10, 12, 13);

4) addition of SR proteins to HeLa cell splicing extract stimulated a change in the relative levels of complexes formed on labelled RNA containing the SK ESE (figure 5.7, lanes 8, 16; figure 5.9A, lanes 11, 14).

The crosslinking experiment also identified a difference in the level of components crosslinked to wild-type and mutant RNAs which was not easily detectable from the results of the gel retardation experiment (table 5.4).

Affinity purification of HeLa cell splicing extracts using wild-type and mutant SK exon RNA

One interpretation of the results described above is that a component of HeLa cell splicing extracts present also in myoblast and myotube cell extracts

facilitates binding of an SRp30 protein to the SK ESE. If this component was only present in small quantities in these extracts it would not have been detected by the crosslinking experiments. A method by which components of extracts that recognise specific RNA sequences can be purified is known as affinity purification. The method is suited to purification using large quantities of extract and can therefore be used to purify components of extracts that are present only in small amounts.

According to the affinity purification method, RNA containing the specific RNA sequence of interest is labelled with biotin. Biotin is not thought to interfere with the recognition of RNA sequences by proteins and binds with high affinity to avidin. The biotinylated RNA is incubated with splicing extract under splicing conditions. Agarose beads that are covalently linked to avidin are then mixed with the RNA and splicing extract. The biotin-RNA binds to the avidin-agarose beads. This biotin-RNA will include biotin-RNA that has been bound by components of the splicing extract during the initial incubation. The avidin-agarose beads are far larger than any of the other components of the incubation and are separated from the incubation by a simple filtration step. Biotin-RNA bound to the avidin-agarose beads with associated components of the splicing extract can then be eluted from the beads. The components of the splicing extract that have been eluted off the beads along with the biotin-RNA are visualised by SDS-polyacrylamide gel electrophoresis followed by silver staining of the resulting gel (see chapter 2).

Affinity purification was performed using HeLa cell splicing extract and biotinylated wild-type SK exon RNA. The components of HeLa cell splicing extract that were purified were compared with those of two parallel purifications using mutant biotinylated SK exon RNAs that contained the SK16-30NM or SK16-30R1 mutations. The results are shown in figure 5.10. The arrow in figure 5.10 is directed to a protein running with a mobility of ~24Kd (p24) that is specific to lane 4. This lane represents components purified using biotinylated wild-type SK exon RNA. Lanes 2 and 3 represent components purified using biotinylated mutant SK exon RNA lacking the SK ESE sequence. Thus, p24 was only purified using biotinylated wild-type SK exon RNA and not RNA containing two different mutations of the SK ESE. It was concluded that p24 binds to SK exon RNA containing the SK ESE sequence. Control reactions in which parallel purifications were done in the absence of biotinylated RNA in other similar experiments did not purify any components that were visible after silver staining (not shown). Thus, there appears to be no background binding to the avidin-agarose beads using this method. Some of the other bands visible in the affinity purifications shown in figure 5.10 were assigned in another study (Taylor, PhD thesis) to components of U1 snRNP.

Figure 5.10. Affinity purification of components of HeLa cell splicing extract using biotinylated wild-type and mutant SK exon RNA

Wild-type or mutant biotinylated RNA was incubated with HeLa cell splicing extract under splicing conditions. Then avidin-agarose beads were added and the mixture was incubated overnight at 4°C. The avidin-agarose beads were separated from the mixture, washed, and extract components that remained associated with the biotinylated RNA were eluted. The eluted components were separated on an SDS-polyacrylamide gel. The protein and RNA components were then visualised by silver staining.
Sizes of marker proteins, in kd, are shown to the left of lane 1. The arrow marks the position of p24 in lane 4.
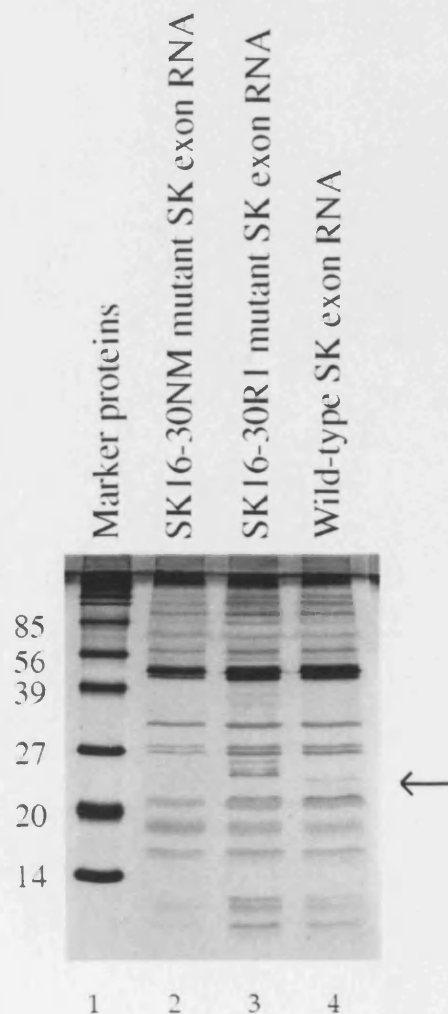
SR proteins do not remain associated with biotinylated RNA during affinity purification (Staknis and Reed 1994). None of the components purified in the above experiment were expected therefore to be SR proteins. This was confirmed by western blot analysis with an anti-SR protein antibody of the components purified using biotinylated wild-type SK exon RNA (data not shown). Thus, p24 is not an SR protein.

The results of the affinity purification experiment are consistent with the identity of p24 as a protein that recruits an SRp30 protein to the SK ESE. Other interpretations of the results are discussed in chapter 6.

Chapter 6
Discussion

TM pre-mRNA splicing regulation

Three important questions must be answered in order to understand the regulation of hTM_{nm} alternative pre-mRNA splicing:

    1) What determines the splicing pattern in non muscle cells?

    2) What determines the splicing pattern in skeletal muscle cells?

    3) Why are exons NM and SK not spliced together?

In theory, the absence of SK isoform expression in non muscle cells could be explained by specific inihibtion of use of the SK exon or by a model in which both the NM and SK exons are active but splicing to the NM exon is dominant because it is associated with stronger splicing signals. These two explanations can be differentiated if splicing to the NM exon is inactivated (either by removing the exon or inactivating splicing signals associated with it). If the SK exon is in competition with the NM exon then inactivation of the NM exon should allow splicing to the weaker SK exon to occur. If, however, the SK exon is specifically repressed in non muscle cells, then inactivation of the NM exon should have no effect on splicing to exon SK. The results of these experiments were obtained before this study using a TMnm minigene (Graham *et al*, 1992, Hamshere, PhD thesis) and established that the SK exon is specifically repressed in non muscle cells.

Role of exon sequences

Graham *et al* (1992) and Hamshere (PhD thesis) also determined the *cis*-sequence requirements for regulation of TMnm minigene expression in non muscle cells. Expression of the SK isoform was observed in non muscle cells when the branch point sequence upstream of exon SK (the SK BPS) was changed to the consensus sequence. SK isoform expression was further increased in these cells when both the branch point sequence upstream of exon SK (the SK BPS) was changed to the consensus sequence and sequence within the SK exon was mutated. In these studies, all the mutants tested which contained mutant SK exon sequence also contained a consensus SK BPS instead of the wild-type SK BPS. An aim of this study was to determine whether SK isoform expression occured in non muscle cells if the SK exon sequence was mutated as before, but the wild-type SK BPS was retained. Thus, mutations of the TMnm minigene were made in which 15 nucleotides at the 5'- or 3'-end of the SK exon were substituted with 15 nucleotides of sequence from the NM exon.

Mutation of 15 nucleotides at the 5'- or 3'-end of exon SK activated expression of the SK isoform to significant levels in non muscle cells. These mutations could either have removed wild-type sequence from the SK exon that is inhibitory to SK exon expression, or they could have introduced sequence into the SK exon which is able to enhance its expression. Another mutant in which three nucleotides at the 5'-end of exon SK were mutated (SK2-4CGC) also expressed the SK isoform in non muscle cells. This result confirmed that wild-type sequence at the 5'-end of exon SK is inhibitory to SK isoform expression. No other mutations of sequence at the 3'-end of the SK exon have yet been tested, so it is not known whether the wild-type sequence of this region of the exon contains inhibitory sequence or whether the only mutant sequence tested contained activatory sequence. Consistent with the latter possibilty, it was noted that the mutant sequence which replaced the wild-type sequence at the 3'-end of exon SK was purine-rich. The requirement of purine-rich sequences for the expression of some exons is well documented (see introduction).

It was concluded that repression of exon SK in non muscle cells is mediated, at least in part, by inhibitory sequences at the 5'-end, and possibly at the 3'-end, of the exon itself and by a sub-optimal branch point sequence upstream of the SK exon. The effect of the SK exon sequence mutations is not dependent on whether the SK branch point sequence is the wild-type or the consensus sequence.

The earlier experiments of Graham *et al* (1992) had also indicated that SK exon nucleotides 16-30 may be required for expression of the SK exon. In the present study, a series of mutant mingenes each containing a different mutant sequence at SK exon positions 16-30 were tested in non muscle and muscle cells. The mutants tested in non muscle cells also contained a consensus SK BPS (mutation of SK exon nucleotides 16-30 would not be expected to disrupt SK exon expression if the wild-type SK BPS was retained because no SK isoform expression is observed under these conditions). Several conclusions were drawn from the effect of these mutants on SK isoform expression:

1) Substitution of the wild-type SK exon sequence at positions 16-30 with a number of other sequences disrupted SK isoform expression in non muscle and in muscle cells. Therefore wild-type sequence in this region of the SK exon (the SK exon enhancer sequence, or SK ESE) is required for expression of that exon.

2) The activity of the SK ESE in non muscle and muscle cells suggests that it is recognised by constitutively active components;

3) The nucleotide sequence, rather than just the purine content, of the SK 16-30 region is important;

4) Mutation of sequence either side of the 16-30 region either activated (SK1-15NM) or had no effect (SK31-45NM) on SK isoform expression (this study; Graham et al, 1992; Hamshere, PhD thesis). The location of the nucleotides

required for expression of the SK exon is therefore limited to the 16-30 sequence.

5) Mutation of SK nucleotides at positions 16-27 was able to inactivate SK isoform expression in myotubes. Thus, it is possible that the enhancer element is only 12 nucleotides long or less.

Switch to use of exon SK in muscle cells

In muscle cells, SK isoform expression is dominant to NM isoform expression (Hamshere, PhD thesis). This switch in expression pattern from that in non muscle cells could be explained if the repression of exon SK is relieved specifically in muscle cells. The NM and SK exons would both then be available to the splicing apparatus and the relative level of use of these exons would be determined by the strength of the splicing signals associated with each exon. Alternatively, more complex models would involve specific inactivity of the NM exon and/or specific enhancement of exon SK use in muscle cells. In order to distinguish these types of model, either the NM or the SK exon was inactivated by mutation of one of their respective splicing signals. If the NM and SK exons are directly in competition with each other for recognition by the splicing apparatus, then inactivation of the SK exon is expected to result in an increase in the level of NM isoform expression. Similarly, inactivation of the NM exon is expected to allow the level of SK isoform expression to increase. If the NM exon is specifically inactive in muscle cells, then inactivation of exon SK should not alter the level of NM isoform expression.

When the SK exon was inactivated in muscle cells, the relative level of NM isoform expression observed was slightly higher than expected if the actual level of expression of that isoform was unaffected by a significant reduction in SK isoform expression. It was concluded that, under wild-type conditions in muscle cells, NM isoform expression is limited to some extent by the activity of the SK exon. However, activity of the SK exon in muscle cells is not the only limitation on NM isoform expression in these cells. When the SK exon was inactivated in muscle cells, the relative level of NM isoform expression was not restored to the level seen in non muscle cells where the SK isoform is not expressed (compare table 3.3: SK16-30R3, 39% NM with table 3.4: WT SK, 56% NM). Thus, there are additional reasons for the decline of NM expression in muscle cells. One possible explanation is that muscle cells lack a factor(s) present in non muscle cells which enhances expression of the NM isoform. Alternatively, NM isoform expression may be specifically inhibited in muscle cells. Consistent with the former explanation, NM expression was demonstrated to be dependent in non muscle cells on the wild-type exon NM sequence at position 62 (data not shown). If a non muscle-specific factor(s) enhances NM isoform expression and recognises sequence in this region of the NM exon, this mutant is not expected to have any

effect on expression of exon NM in muscle. The dependence of NM isoform expression on this sequence in muscle cells has not yet been tested.

When the NM exon was inactivated in muscle cells (ΔNM5'ss) the relative level of splicing to exon SK remained approximately constant. Thus SK exon use is not limited by NM exon use. Even high relative levels of NM isoform expression did not reduce the relative level of SK exon use in muscle cells. SK exon use in muscle cells is, however, limited by sequence within exon SK and by a suboptimal SK BPS, as in non muscle cells. It was concluded that repression of exon SK is only partially lifted in muscle cells, but that the levels of muscle cell SK isoform expression are sufficient to compete with NM isoform expression and to dominate the splicing pattern. This is aided by a reduction in the activity of the NM exon in muscle cells compared to non muscle cells.

Sequences sufficient for the tissue-specific regulation of exon SK splicing

The studies described here, and others (Graham et al, 1992; Hamshere, PhD thesis) have identified *cis*-acting sequences that are necessary for mediating repression or use of the SK exon in non muscle and muscle cells, respectively. Having identified such sequences it was important to determine whether they were sufficient for tissue-specific regulation of TMnm pre-mRNA splicing. This was done by placing these sequence elements in a heterologous environment. Thus, TMnm sequence containing a region spanning just upstream of the SK BPS to beyond the 3'-end of the SK exon was inserted into the second intron of the β-globin gene. The four-exon structure of this heterologous gene was then similar to that of the TMnm minigene. Analysis of the splicing patterns obtained after transfection of this TM/β-globin hybrid construct in non muscle and muscle cells demonstrated that no regulation of SK exon splicing had occurred. This contrasts with data obtained using rat β-TM sequences. Here. the corresponding RNA sequence from the rat β-TM gene (sequence immediately upstream of the BPS through 25 nucleotides downstream of the muscle-specific exon) when flanked by exons derived from adenovirus pre-mRNA was sufficient to regulate the suppression of the muscle-specific exon in non muscle cells (Guo and Helfman, 1993).

Two conclusions can be drawn from this experiment. Either:

1) sequences sufficient for the tissue-specific regulation of SK exon expression are not present within the region that was inserted into the β-globin gene; or,

2) sequences in β-globin pre-mRNA were able to suppress the effect of the TMnm sequences on the tissue-specific regulation of SK exon splicing.

The first conclusion would be strengthened if the same region of TM sequence was placed in a different heterologous context and was still unable to confer tissue-specific splicing.

Mechanism of action of sequence at the 5'-end of exon SK on SK isoform expression

It was of interest that two adjacent sequence elements located at the 5'-end of the SK exon (the first 15 nucleotides of the SK exon and the SK ESE) had opposite effects on SK isoform expression. Examination of the sequences of these two elements provided clues as to their possible mechanisms of action. SK nucleotides 1-15 contain part of a sequence that strongly resembles a 5'ss. This pseudo-5'ss sequence (the SK p5'ss sequence) begins three nucleotides upstream of the SK exon and continues into the first six nucleotides of the exon. Thus, a potential U1 snRNP binding site overlaps with a sequence known to inhibit the expression of exon SK in non muscle cells. It was proposed that U1 snRNP mediates the repression of exon SK via these sequences. SK nucleotides 16-30 are purine-rich and contain three repetitions of the consensus GAR motif common to other mammalian exon enhancer sequences (see introduction). It was proposed that the SK enhancer element could function in the same way as these sequences.

Role of U1 snRNP in repression of splicing to exon SK

Immunoprecipitation studies in HeLa cell nuclear extracts using RNA containing the SK exon and surrounding intron sequence established that U1 snRNP does in fact bind to the 5' end of the SK exon *in vitro* (Taylor, PhD thesis). The aim of experiments described here was to show, *in vivo* that repression of splicing to exon SK is mediated by U1 snRNP. Consistent with this model, a TMnm minigene containing three mutations within the SK p5'ss which disrupted proposed base-pairing interactions with the 5'-end of wild-type U1 snRNA (the SK2-4CGC minigene) expressed the SK isoform in non muscle cells. It was proposed that wild-type U1 snRNA, and consequently U1 snRNP, was unable to bind to the mutant SK p5'ss and that repression of splicing to the SK exon was lifted. A mutant plasmid encoding U1 snRNA was made in which the three nucleotides which no longer base-pair with the mutant SK p5'ss sequence were altered to restore complementarity (the U1 α-CGC plasmid). Expression of the SK isoform from the SK2-4CGC TMnm minigene in non muscle cells was significantly lower in the presence of the U1 α-CGC plasmid than in the presence of a plasmid encoding wild-type U1 snRNA. The U1 α-CGC plasmid had no effect on the relative levels of SK isoform expression observed from the SK1-15NM

TMnm mingene. It was concluded that mutant U1 snRNA was expressed from the U1 α-CGC plasmid and that U1 snRNP proteins had been able to assemble on this RNA and to bind to the mutant SK p5'ss. Interaction of the recombinant U1 snRNP with the mutant SK p5'ss was then able to mediate repression of splicing to the SK exon. The specificity of this effect was demonstrated by the fact that SK isoform expression from the SK1-15NM minigene was unaffected by the presence of the U1 α-CGC plasmid.

The results of this study would be confirmed if the effect of another mutation in the SK p5'ss sequence was suppressed by expression of a mutant U1 snRNA in which the complementarity of base pairing with the mutant SKp5'ss was restored.

The role of SR proteins in SK isoform expression

SR proteins have been shown to bind specifically to a number of mammalian purine-rich splicing enhancers and in some cases binding has been demonstrated to correlate with enhancer activity (see introduction). Involvement of a member of the SR protein family in activity of the SK ESE was therefore a likely possibility.

It was decided to look for components of cell extracts which bound specifically to the SK ESE. Ideally, active muscle cell splicing extracts would have been used because the activity of the SK ESE, within wild-type pre-mRNA, is confined to muscle cells. However, these extracts were not chosen initially because such extracts have very low splicing activity, if any, and can only be produced in small quantities. It was known from the effect of mutations of the SK ESE in non muscle cells (Graham *et al*, 1992; this study) that this element can be recognised by factors present in these cells. Large quantities of active splicing extract are routinely made from HeLa cells which are non muscle in origin. Extracts from such cells were chosen therefore for use in *in vitro* binding experiments aimed at identifying a specific factor(s) binding to the SK ESE. Before any binding experiments were carried out the activity of the SK ESE in these extracts was first tested. The splicing efficiency of a single intron pre-mRNA substrate containing the SK exon was significantly reduced when the SK ESE was mutated. It was concluded that the SK ESE was effective in these extracts and that they could be used to look for a factor(s) binding specifically to the ESE.

Binding experiments using the gel retardation and shortwave UV crosslinking techniques were unable to identify any components of HeLa cell nuclear extract binding specifically to the SK ESE. Similarly, only non specific binding of purified HeLa SR proteins to RNA containing SK exon sequence was observed using the gel retardation and shortwave UV crosslinking assays. However, when SR proteins were combined with HeLa cell nuclear extracts, or

myoblast or myotube whole cell extracts (the myoblast and myotube extracts were inactive for splicing but sufficient quantities were available for limited crosslinking and gel retardation experiments) a protein running with a mobility of ~35Kd was observed to crosslink specifically to RNA containing the SK ESE. This specificity was dependent on components of the SR protein preparation and extract components. It was proposed that the ~35Kd protein was an SRp30 protein which was able to bind specifically to the SK ESE in the presence of extract components. Several questions remain to be answered concerning this interpretation of the results:

1) Based on its apparent molecular weight and presence when SR proteins alone were crosslinked to RNA containing SK exon sequences, it is possible that the ~35Kd protein is an SRp30 protein. However, this interpretation is difficult to confirm experimentally. Unfortunately the nitrocellulose filter containing the crosslinked ~35Kd protein cannot be probed with an anti-SR antibody to show that this protein is a member of the SR family because other SR proteins are also present on the filter. It would therefore be very difficult to distinguish binding of an anti-SR antibody at the position of the ~35Kd labelled signal from binding to unlabelled SR proteins. Attempts to immunoprecipitate complexes bound to SK ESE-containing RNA in the presence of splicing extract using an anti-SR protein antibody were unsuccessful. The presence of an SR protein in the complex visible on native gels when SR proteins and extracts were combined could, in theory, be demonstrated using a supershift assay. If an SR protein is present then addition of an anti-SR antibody to the incubation prior to electrophoresis would be expected to inhibit the mobility of the complex relative to parallel incubations without the anti-SR antibody (a supershift). However, only the presence of an SR protein would be confirmed. The exact identity of the ~35Kd protein would remain uncertain. The supershift approach would also be associated with practical problems because the resolution of complexes on native gels was poor. A possible way to confirm that the ~35Kd protein is an SR protein could be to repeat the crosslinking experiment using radioactively labelled RNA that was also labelled with biotin. The RNA, some of which would be covalently crosslinked to the ~35Kd protein, could then be purified away from other components of the incubation using beads coupled to streptavidin. The biotin-strepatvidin interaction is extremely strong and would allow very stringent washing of the RNA and consequent removal of all other components not covalently associated with the RNA. The presence of an SR protein crosslinked to the RNA would then be determined by SDS-PAGE analysis and immunoblotting after the RNA had been eluted off the beads.

2) If it is assumed that the same ~35Kd protein was responsible for the labelled crosslinked product visible when SR proteins alone or in combination with extract components were crosslinked to RNA, an important question

concerns the source of extra protein visible when SR proteins were combined with extract components. Above it was proposed that extract components could stimulate the specific binding of a protein present in the SR preparation (possibly an SR protein) to the SK ESE. However, it is possible that the extra protein came instead from the extracts. These possibilities could, in theory, be distinguished using an SR protein preparation in which the proteins had been radioactively labelled (for example with ^{35}S). If extract components were able to stimulate binding to the SK ESE of a protein present in the SR preparation then the level of radioactive signal should be higher when crosslinks are formed between SR proteins and RNA in the presence of extract compared with those formed with SR proteins alone and RNA. In order to avoid masking of this potential signal by other labelled proteins present in the SR preparation, the RNA would have to be purified away from other components after crosslinking. This could be achieved using beads coupled to streptavidin and biotinylated RNA.

3) The exact binding site of the ~35Kd protein to RNA containing SK exon sequence has not been determined. The increase in amount of crosslinked product in the presence of RNA containing the SK ESE suggests that the ESE itself is the binding site, but this is not necessarily so. For example, RNA secondary structure effects may be important. The SK ESE may be required to destroy or create a secondary structure that either inhibits or allows binding, respectively, of the ~35Kd protein. Specific binding to the SK ESE could be determined using RNA which has been radioactively labelled in a site-specific manner at one or more nucleotides within the SK ESE. Formation of a visible ~35Kd crosslinked product only to RNA labelled in this region would demonstrate a specific interaction with the SK ESE.

4) It is not known if the crosslinking efficiency of the ~35Kd protein to SK ESE-containing RNA is representative of the interaction of this protein with RNA. The effect of unlabelled wild-type and mutant competitor RNA on crosslinking of the ~35Kd protein to SK ESE-containing RNA in the presence and absence of extract should determine whether actual binding of the ~35Kd protein to SK ESE-containing RNA is specific.

5) If binding of an SRp30 protein to SK ESE-containing RNA is facilitated by components of extract this could potentially be achieved in a number of ways.

a) The extract components may contain a protein(s) which binds specifically to the SK ESE and also interacts with the ~35Kd protein (similar to the GAA-binding protein identified by Yeakley et al, 1996). In this way the extract components could serve to recruit the ~35Kd protein to the SK ESE.

b) The RNA containing SK exon sequence used for the binding experiments could form a secondary structure *in vitro* that made the ESE inaccessible to specific binding proteins. The extracts are likely to contain active

helicase enzymes which would unravel any secondary structure and could then allow binding of the ~35Kd protein to the ESE.

c) SR proteins are known to be phosphorylated at residues present in their RS domains. The degree of phosphorylation could affect RNA binding specificity. If this was the case, and the purified SR proteins were under-phosphorylated, then addition of extract containing active kinase enzymes could restore the specificity of binding.

Identification of the extract and SR protein components required for formation of the specific ~35Kd crosslinked product could potentially be achieved by fractionation of these preparations using formation of the specific crosslinked product as an assay for purification.

In vivo, under wild-type conditions, the SK ESE is active in muscle but not non muscle cells, although the factors involved in recognition of the SK ESE *in vivo* are thought to be active in non muscle cells. The *in vitro* binding experiments, in which SR proteins were combined with myoblast or myotube extracts, revealed that both extracts were able to stimulate specific crosslinking of the ~35Kd protein to SK ESE-containing RNA. It is possible that whatever prevents the activity of the SK ESE in non muscle cells is no longer able to function under the conditions used for the *in vitro* binding experiments. For example, RNA sequences other than those present in the binding experiments may be required. The proposal that the SK exon and surrounding sequence is not sufficient to maintain tissue-specific regulation in an heterologous environment (chapter 3) is consistent with this. Alternatively, an activity may have been lost in the preparation of the extracts. One attractive explanation is a possible involvement of U1 snRNP. If binding of U1 snRNP to the SK p5'ss was able to inhibit specific recognition of the adjacent SK ESE, for example by steric interference, this would explain the absence of SK ESE activity under normal conditions in non muscle cells. If U1 snRNP was unable to bind to the SK p5'ss *in vitro* under the conditions used in the binding experiments this could account for the ability of both myoblast and myotube extracts to stimulate specific binding to RNA containing the SK ESE. Consistent with this proposal it was noted that the binding experiments were all carried out in the presence of heparin. This is known to prevent recognition of 5'ss sequences by U1 snRNP *in vitro* (Bindereif and Green, 1987).

Four bands were observed after analysis of the products formed when RNA in the presence of HeLa cell nuclear extract alone, or in combination with SR proteins, was exposed to shortwave UV light. However, when myoblast or myotube whole cell extracts alone were exposed under similar conditions, no bands were visible after the products were analysed. The reasons for this difference are unknown but may be connected to the inactivity of the myoblast or myotube extracts in splicing or the fact that these extracts were made from whole

cells rather than from nuclei. Also unexplained is the apparent reduction in intensity of three of the four bands visible after HeLa cell nuclear extracts were exposed to shortwave UV light in the presence of SR proteins and wild-type RNA compared with the intensity of the bands visible under similar conditions using mutant RNA (table 5.4B). It is possible that the SK ESE acts as a recruitment site for SR proteins and that these proteins are then able to bind along the length of the RNA and remove non specifically bound proteins.

Significance of p24 identified by affinity purification

The requirement of an SR protein preparation component(s) and an extract component(s) for specific crosslinking of an ~35Kd protein to RNA containing the SK ESE suggested that an extract component(s) could facilitate binding of an SRp30 protein to the SK exon. The extract component(s) would be expected to bind specifically to the SK ESE or to RNA containing the SK ESE. Wild-type SK exon RNA was used therefore to affinity purify components of HeLa cell splicing extract that bind to SK exon RNA. One component running with a mobility of ~24Kd (p24) on an SDS-polyacrylamide gel was purified that was not purified using two different mutant SK exon RNAs, each lacking the SK ESE sequence. Attempts to sequence p24 were unsuccessful, but it was demonstrated that p24 is not recognised by an antibody directed against SR proteins.

It was proposed that p24 could be a non-SR protein that binds specifically to the SK ESE sequence and facilitates binding of an SRp30 protein to the SK exon. However, it was not shown that p24 was a protein. The silver staining technique that was used to visualise the affinity purified components also stains RNA. Thus, p24 could be an RNA resulting from partial digestion of the biotinylated RNA used for affinity purification. Its uniqueness to the affinity purification using wild-type SK exon RNA could be explained by the fact that the mutant SK exon RNAs do not form the same partial digestion products. This explanation does, however, seem unlikely as purification of p24 was repeated in a different affinity purification experiment. The activity of nucleases in different experiments would not be expected to be repeatable.

P24 could be the same protein that was identified as the ~35Kd protein in the crosslinking experiments. The mobility of proteins that have been crosslinked to RNA is known to be anomslously high. Alternatively, p24 was purified under conditions in which it is known that SR proteins do not remain associated with RNA. If an SR protein does bind directly to the SK ESE, p24 could merely represent a protein that is only able to bind to the SK ESE when that sequence is not bound by an SR protein and that is not involved in stimulating splicing to exon SK. Clearly, sequencing and cloning of p24 would help to resolve some of

these issues. Unfortunately, an attempt to obtain enough p24 for sequencing was unsuccessful.

Models for tissue-specific regulation of SK isoform expression

Very little is known about the regulatory mechanisms involved in mammalian alternative pre-mRNA splicing. However, it seems likely that early steps in spliceosome assembly will be the target of regulation rather than later steps. The mechanism by which U1 snRNP is able to mediate repression of exon SK in non muscle cells via sequences at the 5'-end of the exon is unknown. Figure 6.1 illustrates two models for the regulation of TM pre-mRNA splicing. In the first model, SR proteins bound directly, or recruited to the SK ESE in muscle cells are able to stimulate binding of U2AF to the upstream polypyrimidine tract (PPT) by interacting with U2AF^{35} (see introduction). However, when U1 snRNP is bound to the SK p5'ss in non muscle cells recognition of the SK ESE is prevented, interaction of U2AF with the PPT upstream of exon SK is much weaker, and E complex formation is inhibited. This model predicts that recognition of the SK p5'ss sequence and SK ESE sequences is mutually exclusive. Thus the SK ESE binding protein (SEBP) may be expressed and active in non muscle cells but is unable to interact with the SK ESE because U1 snRNP is bound to the adjacent SK p5'ss sequence. Consistent with this explanation, SK isoform expression from the bpup mutant TMnm minigene is dependent on the SK ESE, suggesting that the SEBP is active in non muscle cells.

In order to establish whether formation of the SK ESE-specific crosslinked product is dependent on the absence of an interaction of U1 snRNP with the SK p5'ss these crosslinking experiments could be repeated, but in the presence and absence of heparin. An alternative approach would be to compare the crosslinked products formed (in the absence of heparin) using extracts in which the 5'-end of U1 snRNA had been cleaved or mock cleaved. U1 snRNA lacking 5'-end sequences does not bind to 5' splice site sequences and, therefore, U1 snRNPs assembled on such RNA would not be expected to interact with the SK p5'ss. If the SEBP is only able to interact with the SK ESE when U1 snRNP is not bound to the SK p5'ss, it is predicted that the ~35Kd specific crosslinked product will only be able to form when the experiment is carried out in the presence of heparin or of extract in which the 5'-end of U1 snRNA has been cleaved. If, however, binding of the SEBP is not mutually exclusive with binding of U1 snRNP to the SK p5'ss sequence, the presence of heparin or of extract in which the 5'-end of U1 snRNP has been cleaved, is not expected to affect formation of the ~35Kd crosslinked product (assuming that this crosslinked product is representative of the SEBP-SK ESE interaction).

Figure 6.1 Models for regulation of TM pre-mRNA splicing

1) Mutually exclusive binding of U1 snRNP and the SEBP (shown here as an SR protein)

A) Nonmuscle

BPS   PPT   ESE
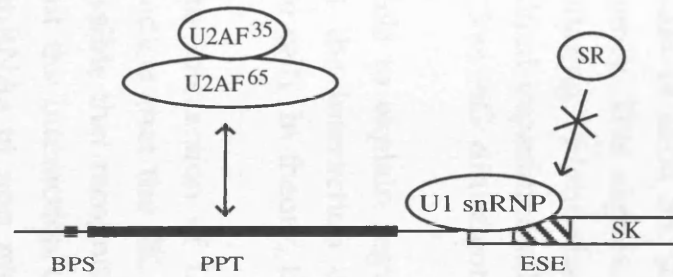
B) Muscle

BPS   PPT   ESE

2) Inhibition of E complex assembly by U1 snRNP

A) Nonmuscle

Binding of U1 snRNP to region of 3'ss inhibits efficient E complex assembly

B) Muscle

E complex assembles efficiently
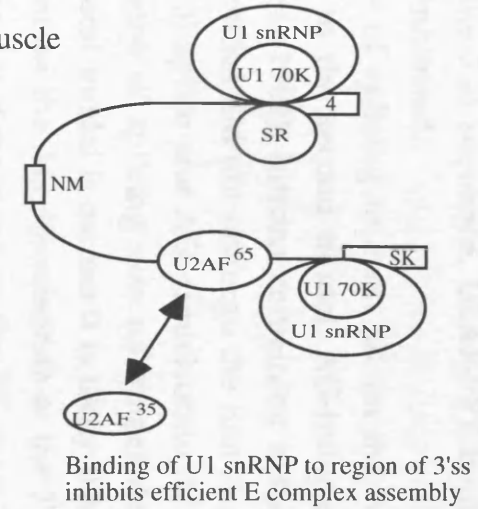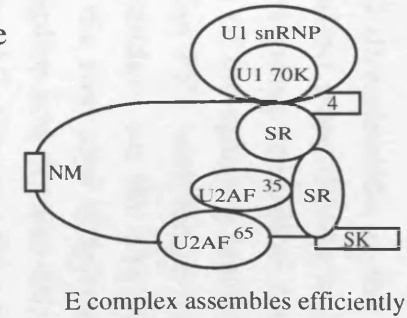
In the second model, SR proteins are able to bind directly to the 3'ss sequence of exon SK in muscle cells and stimulate the interaction of U2AF with the PPT. Recognition of the 3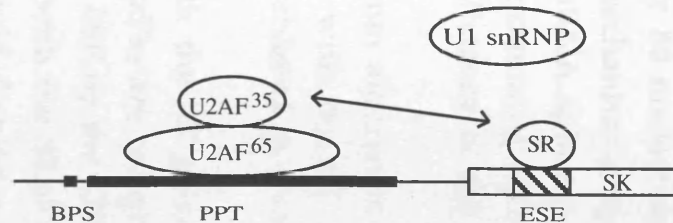'ss sequence by SR proteins in adenovirus pre-mRNA has been proposed by Staknis and Reed (1994) (see introduction). When U1 snRNP is bound to the SK p5'ss in non muscle cells, SR proteins are no longer able to interact with the 3'ss sequence, U2AF-PPT binding is weakened, and E complex formation is inhibited.

The dependence of splicing to exon SK on the AG dinucleotide at the 3' splice site is relevant to the second model. AG-independent pre-mRNAs were first identified by Reed (1989). Introns containing a short polypyrimidine tract (PPT) of only 14 nucleotides did not undergo the first step of the splicing reaction in the absence of the 3' splice site AG dinucleotide. If the PPT was longer (26 nucleotides) the first step of splicing was no longer dependent on the 3'ss AG dinucleotide. If the second model is correct it is likely that the first step in splicing to exon SK is dependent on the AG dinucleotide at the 3' splice site of exon SK. If splicing to exon SK was not dependent on the SK exon 3'ss AG dinucleotide, the early stages of spliceosome assembly would not be expected to be affected by interaction of U1 snRNP with the SK p5'ss. The second step of splicing would, however, be inhibited by such an interaction but control of alternative splicing at later stages of the splicing process does not seem likely.

The AG-dependence of splicing to exon SK has not been investigated but the length of the PPT upstream of exon SK (approxiamately 80 nucleotides) is consistent with AG-independence. This argues against the mechanism proposed in the second model for regulating TMnm alternative pre-mRNA splicing. AG-dependence could be determined experimentally *in vitro* by examining the effect of mutation of the SK exon 3'ss AG dinucleotide on the efficiency of the first cleavage step.

All the proposed models to explain regulation of TMnm alternative pre-mRNA splicing require that the interaction of U1 snRNP with the SK p5'ss sequence is reduced in muscle cells. In theory, this could be achieved in a variety of ways:

1) Steric block to interaction of U1 snRNP with the SK p5'ss. As described above, model 1 predicts that the SK ESE and SK p5'ss are recognised mutually exclusively. It is possible that recognition of the SK ESE by the SEBP in muscle cells is able to prevent the interaction of U1 snRNP with the SKp5'ss. If this is the case, TMnm pre-mRNAs in non muscle cells should display a high occupancy by U1snRNP at the SKp5'ss and a low occupancy by the SEBP at the SK ESE. In contrast, TMnm pre-mRNAs in muscle cells should display a low occupancy by U1 snRNP at the SKp5'ss and a high occupancy by the SEBP at the SK ESE. Experiments described in chapter 5 suggested that the SEBP may be an SR protein. The levels of these proteins are known to vary between non muscle and

muscle tissues (Zahler et al, 1993a and b). Thus, expression of an SR protein to higher levels in muscle tissue than non muscle tissue could stimulate increased occupancy at the SK ESE and a reduction in the interaction of U1 snRNP with the SKp5'ss;

2) Reduction in the levels of U1 snRNP in muscle cells. Potentially the splicing of all muscle pre-mRNAs would be affected by a reduction in the levels of functional U1 snRNP. The wide-ranging effects of such a reduction make this an unlikely mechanism. However, the level of U1 snRNA expressed in non muscle cells compared with muscle cells could be determined by Northern blotting. The level of U1 snRNP proteins in the different cell types could be determined by Western blotting;

3) The presence of U1-specific snRNP proteins has been shown to affect the affinity of U1 snRNP for 5'ss sequences (Heinrichs et al, 1990). Thus, interaction of U1 snRNP with the SK p5'ss could be regulated by the removal/inactivation of a U1-specific snRNP protein in muscle cells or the expression/activation of such a factor in non muscle cells. Activation-inactivation of U1 snRNP proteins could potentially be regulated by post-translational modifications such as phopshorylation by a kinase(s) or dephosphorylation by a phosphatase(s). Again, it is more likely that any factor involved has a specific effect on the U1 snRNP-SK p5'ss interaction than a more general effect on interaction of U1 snRNP with all 5'ss sequences in all transcripts;

4) The affinity of U1 snRNP for 5'ss sequences has been shown to be enhanced by ASF/SF2 (see introduction). Thus, the affinity of U1 snRNP for the SK p5'ss could be regulated by a reduction in the expression level/activity of SF2/ASF in muscle compared with non muscle tissue. Interestingly, the sequence at SK exon nucleotides 27-34 strong resembles a sequence shown *in vitro* to have high affinity for ASF/SF2 (Tacke and Manley, 1995):

SK 27-34   TGAAGAAT
High affinity sequence   (A/G)GAAGAAC

The SK exon sequence at positions 27-34 could act as a recruitment site for SF2/ASF in non muscle cells. In muscle cells, a reduction in the level/activity of SF2/ASF would decrease the affinity of U1 snRNP for the SK p5'ss. If recognition of the SK p5'ss by U1 snRNP was able to prevent SK ESE activity, then weakening of the U1 snRNP SK-p5'ss interaction would allow the SEBP to bind the SK ESE and activate splicing to exon SK (figure 6.2).
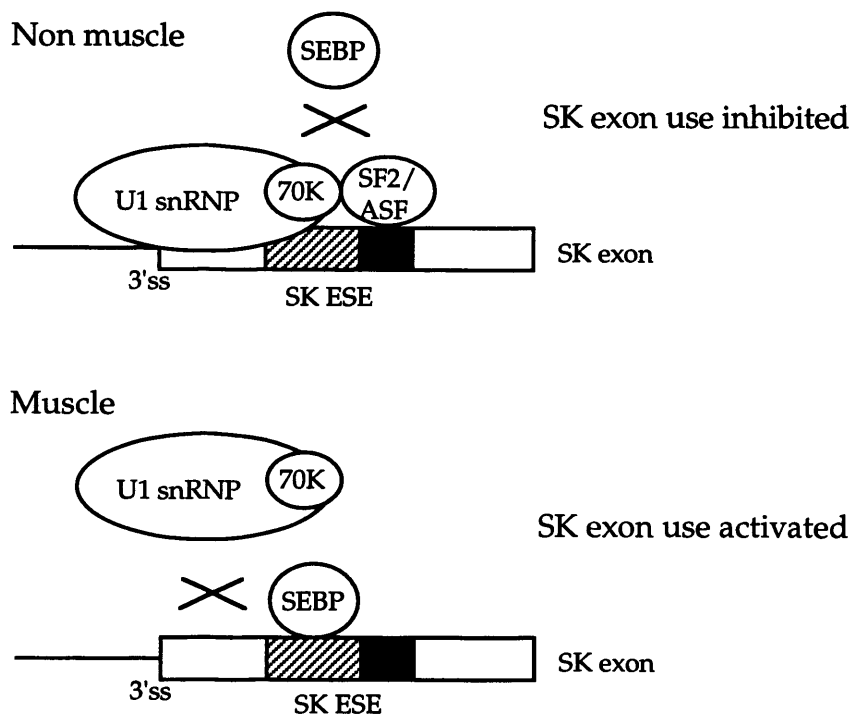
Figure 6.2 Possible role for an SF2/ASF-T(GAA)$_2$T interaction in the tissue-specific regulation of human TMnm pre-mRNA splicing

The SK exon and some upstream intron is shown. The SK ESE is represented by the striped box and the 5'-end of the black box. The TGAAGAAT sequence is represented by the black box. SEBP= SK ESE binding protein.

This model could be tested by examining the splicing activity of a mutant in which only SK nucleotides 27-34 were mutated. If the model is correct then mutation of these nucleotides would be expected to activate expression of exon SK in non muscle cells.

Mutually exclusive use of exons NM and SK

Mutually exclusive alternative pre-mRNA splicing is involved in the regulated expression of several of the tropomyosin genes (see introduction). The mutually exclusive behaviour of rat α-TM exons 2 and 3 is enforced by the close proximity of the exon 3 branch point to exon 2 (42 nucleotides). Insertion of spacer sequence to increase the distance between the 3'-end of exon 2 and the branch point adenosine upstream of exon 3 to greater than 51 nucleotides allowed splicing of exon 2 to exon 3 (Smith and Nadal-Ginard, 1989). This steric block to splicing of mutually exclusive exon pairs does not appear to be a general mechanism, at least in the tropomyosin family: the distance between the upstream exon and the branch point within intron sequence separating mutually exclusive exon pairs has been established in the chicken and rat β-TM, and

6.14

hTM_{nm} genes and is much greater than 51 nucleotides (Helfman *et al*, 1989, 1990; Goux-Pelletan *et al*, 1990; Graham *et al*, 1992).

In non muscle cells, under wild-type conditions, the human TMnm SK exon is inactive, so mutual exclusion with the NM exon is easily explained. However, under some conditions in non muscle cells and in muscle cells, both exons are active yet mutual exclusion is maintained. There is no intrinsic block to removal of the intron separating the NM and SK exons: the NM and SK exons were spliced together *in vivo* when the introns separating exon IV and exon NM and exon SK and exon VI were removed leaving no other splicing choice (Graham *et al*, 1992). Some evidence suggests that splicing of exon NM to exon SK is very inefficient. When the strength of the splicing signals associated with exon SK were enhanced, some splicing of exon NM to exon SK was observed (Hamshere, PhD thesis). The equivalent intron separating the central mutually exclusive exon pair in the chicken β-TM gene is also spliced inefficiently. The position of the branch point sequence within this intron is important: splicing of the intron was much enhanced *in vitro* when it was moved from over 100 nucleotides upstream of the skeletal muscle exon to within 40 nucleotides of that exon (Goux-Pelletan *et al*, 1990). The BPS within the intron separating human TMnm exons NM and SK is also positioned far upstream of the skeletal muscle exon. This positioning may function to ensure that removal of the intron separating exons NM and SK does not occur efficiently.

The activity of the NM exon is lower in muscle cells than in non muscle cells; when the SK exon was inactivated in muscle cells, the relative level of NM isoform expression observed in muscle cells was lower than in non muscle cells where the SK exon is naturally inactive (table 3.3). This reduction in exon NM activity in muscle cells combined with a low efficiency of removal of the intron separating the NM and SK exons may ensure that there is a low probability that exon NM will splice to exon SK.

No *cis*-acting sequences, other than known splicing signals, have been shown to affect the mutually exclusive pattern of splicing in human TMnm pre-mRNA. A role for such sequences, however, cannot be ruled out. The splicing pattern obtained from expression of the heterologous TM/β-globin gene, described in chapter 3, demonstrated that mutual exclusion was not preserved when the human TMnm SK BPS, its position relative to the SK exon 3' splice site, and the splicing signals associated with the SK exon were retained in the heterologous environment. Therefore, either the NM exon 5'ss sequence is also required or sequences in addition to those already identified must be involved.

Future work

Several questions remain unanswered concerning the regulation of human TMnm pre-mRNA splicing and will need to be addressed in the future:

6.15

1) The experiments described in chapter 4 demonstrate the involvement of U1 snRNP in the inhibition of SK isoform expression in non muscle cells. Models to explain the mechanism by which U1 snRNP is able to achieve this and experiments to test these models have been discussed above;

2) The specific factor(s) required for the change from repression of exon SK in non muscle cells to its activation in muscle cells has not been identified. The approach to identification of such a factor(s) will depend on the mechanism by which the U1 snRNP-SK p5'ss interaction is reduced. For example, a reduction in the level of U1 snRNA or of one or more U1 snRNP proteins will involve factors concerned with the expression of these components, whereas a reduction in the activity of U1 snRNA or of any of the U1 snRNP proteins will involve a factor(s) acting at the post-translational level. Experiments to differentiate between these different mechanisms were discussed above. The levels or activity of one or more of the SR proteins could also play a role in the activation of exon SK. The levels and activity of all the known SR proteins in non muscle and skeletal muscle cells have not yet been determined. Such information may be useful in identifying possible candidate regulatory factors in human TMnm expression;

3) The experiments described in chapter 3 suggest that all the *cis*-acting sequences required for the tissue-specific regulation of exon SK splicing have not yet been identified. Other sequence requirements could be investigated by expanding the region of the human TMnm gene which is inserted into the β-globin gene;

4) The identity of the ~35Kd protein is not known. A possible way to determine whether this protein is an SR protein was discussed above. The amount of the ~35Kd crosslinked product obtained was dependent on SR proteins, extract components and RNA containing the SK ESE. The reason for this dependence is not known, but identification of the factor(s) required using specific formation of the crosslinked product as an assay for purification would be an important step to answering this question;

5) The functional significance of the ~35Kd protein has not been established. The results described here have shown a correlation between the *in vitro* splicing efficiency of pre-mRNA containing mutant sequence at SK exon nucleotides 16-30 and the ability of a shorter RNA containing the mutant sequence to interact with a ~35Kd protein (assuming that the crosslinking results are representative of interaction of the ~35Kd protein with RNA). However, this does not demonstrate that the ~35Kd protein is functional in splicing or that it is responsible for the activity of the SK exon in muscle cells. Once the identity of the ~35Kd protein is known it should be possible to obtain it in pure form and to investigate its ability to enhance SK splicing and to bind to the SK ESE *in vitro*.

6) Attempts to sequence the p24 protein purifed by affinity purification of HeLa splicing extracts using biotinylated wild-type SK exon RNA were

unsuccessful because not enough protein could be purified. More recent protein analysis techniques require only very small amounts of protein to obtain sequence information. Thus, re-purification of p24 should enable the protein to be sequenced and eventually cloned. The possible role of p24 in splicing to exon SK could then be investigated.

Relevance of mechanisms proposed to be involved in regulation of human TMnm pre-mRNA splicing to other alternatively spliced pre-mRNAs

The chicken α-TM and rat and chicken β-TM pre-mRNAs share extensive sequence homology with human TMnm pre-mRNA. They also contain a centrally located mutually exclusive exon pair. These exons are expressed in a similar tissue-specific manner to the human TMnm NM and SK exons (see introduction). The sequence of each pre-mRNA at the 5'-end of the skeletal muscle-specific exon is shown in figure 6.3.

Human TMnm:	CAG I UAAGUGUUCUGAGCUGGAGGAGGAGCUGAAGAAU
Chicken α-TM:	CAG I UAA̲AUGUUC̲C̲GAGCUGGAGGAGGAGCUGAAGAAU
Chicken β-TM:	U̲AG I UAA̲AUGU̲G̲G̲UGA̲C̲CU̲AGAGGAGGAGCUGAA̲A̲AU̲U
Rat β-TM:	CAG I UAA̲AUGU̲G̲G̲G̲GA̲C̲CU̲AGAGGAGGAGCUGAA̲A̲AU̲U

Figure 6.3 Comparison of human TMnm pre-mRNA sequence from positions -3 to +34 with equivalent sequence from chicken α-TM and chicken and rat β-TM pre-mRNAs

Differences in each sequence from the human TMnm pre-mRNA sequence are shown underlined. The vertical line represents the intron-exon boundary.

The sequence of the chicken α-TM and the chicken and rat β-TM pre-mRNAs is almost identical to the human TMnm exon V_{SK} sequence in the region corresponding to the SK p5'ss of human TMnm exon SK (-3 to +6). This suggests that U1 snRNP may also play a role in the inhibition of splicing to the skeletal muscle-specific exon in these pre-mRNAs. Consistent with this proposal, mutation of the pre-mRNA sequence at the 5'-end of the skeletal muscle-specific exon has been shown to activate use of that exon in non muscle cells for both the rat and chicken β-TM pre-mRNAs (Guo et al, 1991; Gallego et al, 1992). The sequence at positions 16 to 30 of the muscle-specific exon of the chicken α-TM and the chicken and rat β-TM pre-mRNAs is also almost identical to the human TMnm sequence (figure 6.3). Thus, it is likely that an exon splicing enhancer is also located in this region of these pre-mRNAs.

Although the *cis*-acting sequences involved in the tissue-specific regulation of the rat and chicken β-TM pre-mRNAs appear to be very similar,

there is some evidence that the mechanisms involved may be quite different. Differences in the requirements for regulated tissue-specific splicing have been demonstrated for the chicken and rat β-TM pre-mRNAs. When a chicken β-TM minigene was transfected into quail myoblasts and myotubes, regulated tissue-specific splicing was observed. When the same minigene was transfected into mouse myoblasts and myotubes the mutually exclusive behaviour of the regulated exon pair was lost and the same splicing pattern was observed in both cell types. The rat β-TM pre-mRNA was spliced as expected in mouse myoblast and myotube cells, but in quail myoblasts and myotubes neither of the mutually exclusive exons was recognised (Balvay *et al*, 1994). Further work is needed to understand the reasons for these differences.

The chicken α-TM and the chicken and rat β-TM pre-mRNAs all contain an 'A' nucleotide at position four of the skeletal muscle-specific exon. This nucleotide disrupts the run of six continuous base pairs predicted to form between this region of the pre-mRNA and the 5'-end of U1 snRNA. Instead, only four continuous base pairs are predicted to form. This difference may well be critical to the involvement of U1 snRNP in the inhibition of skeletal muscle exon splicing in non muscle cells with these pre-mRNAs. Similar experiments to those described in chapter 4 should determine whether U1 snRNP is involved in regulation of the non muscle splicing pattern of the chicken α-TM and the chicken and rat β-TM pre-mRNAs.

Some other pre-mRNAs have been described in which positive and negative cis-acting exon elements are located adjacent to each other. The equine infectious anemia virus (EIAV) pre-mRNA undergoes regulated alternative splicing. Exon 3 has been shown to contain an exon splicing enhancer (ESE) which interacts specifically with SR proteins. In the absence of EIAV Rev protein exon 3 is expressed. However, when Rev is expressed, exon 3 is skipped. Exon 3 is known to bind Rev so it has been proposed that EIAV Rev promotes exon skipping by interfering with SR protein binding to the exon 3 ESE (Gontarek and Derse, 1996). Similarly, the HIV-1 tat-rev terminal exon contains an ESE and an exon splicing silencer (ESS), separated by 20 nucleotides. This proximity has been shown to have an effect on the levels of mature mRNA produced. It has been proposed that the level of splicing to this exon could be modulated by altering the abundance of ESE and ESS binding factors (Staffa and Cochrane, 1995).

The location of positively and negatively acting exon splicing elements within close proximity of each other may turn out to be a relatively common mechanism by which the level of mature mRNAs is regulated. The involvement of U1 snRNP in the negative regulation of splicing of other pre-mRNAs remains to be established.

Appendix 1
Analysis of TMnm splicing patterns by RT-PCR

In the first step of this analysis, mRNA expressed from the TMnm minigene was reverse transcribed using an oligodeoxyribonucleotide primer (RTPSVR, figure A1.1A) that annealed to SV40 mRNA sequence derived from the expression vector. In this way only mRNA produced from the expression vector, and not endogenous TMnm mRNA, was reverse transcribed. A DNA strand complementary to the reverse transcribed strand was then synthesised by Taq polymerase, primed from another oligodeoxyribonucleotide primer (CAT5015, figure A1.1A). This primer annealed to sequences that are complementary to sequences in TMnm exon IV. The double stranded product from these reactions was then amplified by Taq polymerase using a further pair of oligodeoxyribonucleotide primers (nested primers) that annealed internally to the first pair (exon IV and exon VI/VII, figure A1.1A). This pair of nested primers were chosen so that they would allow amplification of all four possible TMnm minigene isoforms (figure A1.1B).

The non muscle and skeletal muscle exons NM and SK are both 76 nucleotides long. Therefore, PCR products derived from the NM and SK isoforms are indistinguishable from each other when they are run on a denaturing polyacrylamide gel. The relative incorporation of exon NM and exon SK was determined using a restriction enzyme that cleaves exon SK sequences but not exon NM sequences. Two different restriction enzymes were used: AluI and MboII. The approximate location of the restriction enzyme recognition sites within the products of the nested PCR reactions are shown in figure A1.2. In order to visualise the products of the nested PCR reactions one of the nested PCR primers was radioactively labelled at its 5'-end. It can be seen from figure A1.2. that the restriction enzymes which cut within exon SK, but not within exon NM, also cut sequences outside of exon SK (AluI within exon IV, MboII within exon VI). This affected which primer was radioactively labelled. If the exon VI/VII primer was end-labelled, only AluI could be used to analyse the splicing patterns. Similarly, if the exon IV primer was end-labelled only MboII could be used, otherwise identical length digestion products would be obtained for each isoform. Figure A1.2 illustrates the expected nested PCR products resulting from amplification of spliced TMnm isoforms before and after digestion with AluI or MboII. Table A1.1 gives the expected lengths, in nucleotides, of these products.

A)

ExonVI/VII   RTPSVR

| I_SK | II_SK | II | III | IV | | V_NM | | V_SK | | VI | VII | VIII_SK | 3'-UTR |

SV40
O&LP   AUG

CAT   ExonIV
5015

AAUAAA

B)

NM/SK isoform

| I_SK | II_SK | II | III | IV | V_NM | V_SK | VI | VII | VIII_SK | 3'-UTR | ——AAAA.........

NM isoform

| I_SK | II_SK | II | III | IV | V_NM | VI | VII | VIII_SK | 3'-UTR | ——AAAA.........

SK isoform

| I_SK | II_SK | II | III | IV | V_SK | VI | VII | VIII_SK | 3'-UTR | ——AAAA.........

Skipped isoform

| I_SK | II_SK | II | III | IV | VI | VII | VIII_SK | 3'-UTR | ——AAAA.........
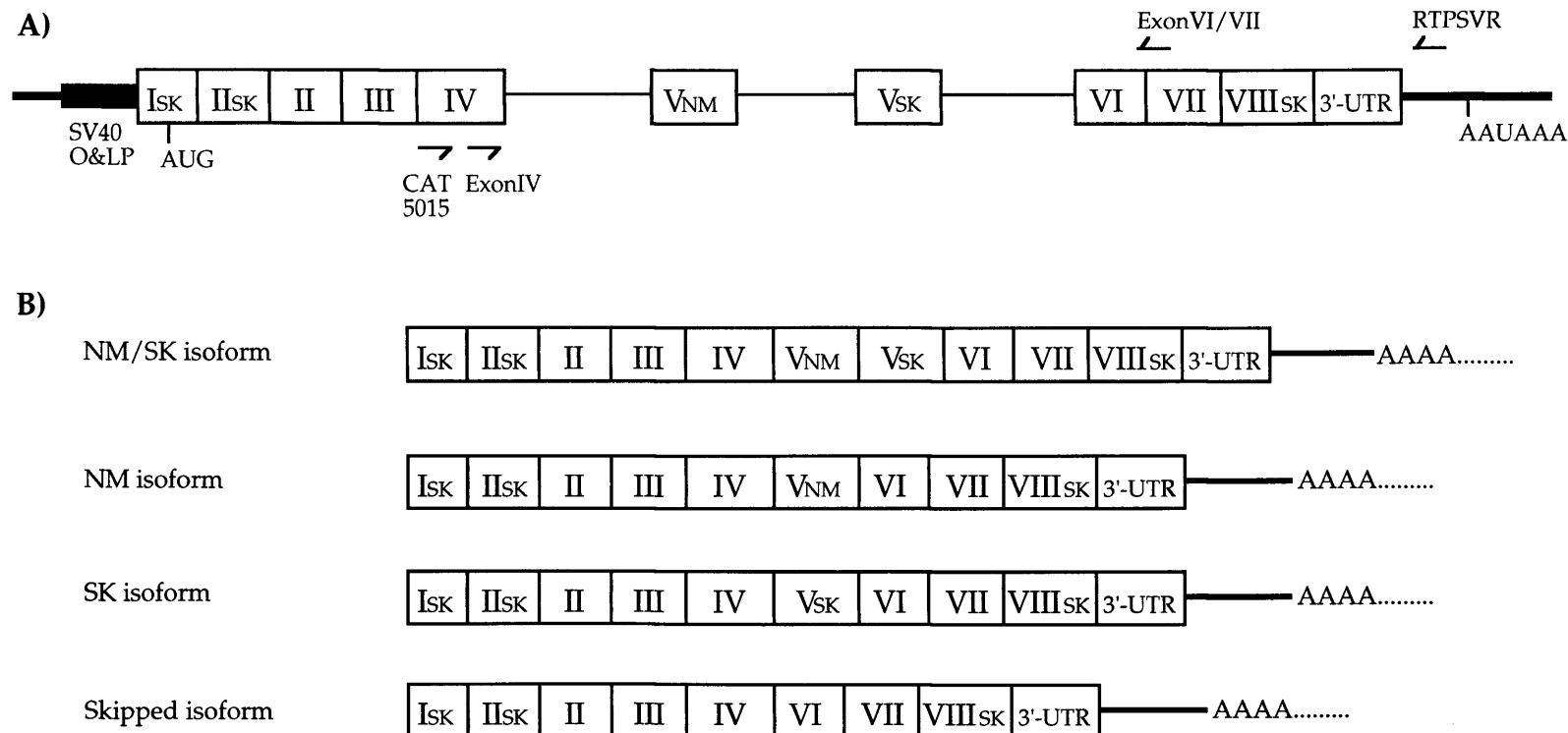
Figure A1.1 The human TMnm minigene

A) Schematic representation of the human TMnm minigene cloned into a eukaryotic expression vector. Intron sequences are represented by a thin line, exon sequences by open boxes. Transcription is driven by the SV40 late promoter (shown with SV40 origin sequences as a black rectangle). The thick black line represents further SV40 vector sequences. The position of the AAUAAA polyA signal is shown. The AUG initiation codon is located within exon I_SK, the 5' sequences of that exon are therefore untranslated. At the 3'-end of the minigene untranslated sequence is shown as 3'-UTR (3'-untranslated region).

The position of binding of DNA oligo sequences used for RT-PCR analysis of spliced isoforms is shown by arrows above and below the minigene. RT-PCR analysis, with reference to these oligos, is described in detail in the text.

B) The potential variety of mRNA isoforms expressed from the TMnm minigene. Mature mRNA is shown schematically. The open boxes represent TMnm sequence, and the thick horizontal line represents vector sequence. The poly(A) tail added during the polyadenylation reaction is represented as "AAAA..........".

Figure A1.2 RT-PCR Products before and after digestion with MboII or AluI

Undigested products resulting from reverse transcription, PCR and nested PCR of spliced isoforms are represented as boxes. ^{32}P end-labelled nested RT PCR products before and after digestion with MboII or AluI are represented as lines. The labelled end is shown with an asterix. The positions of the MboII (M) and AluI (A) restriction enzyme sites are represented by the arrows. The length in, nucleotides, of each labelled nested product before and after digestion is shown.
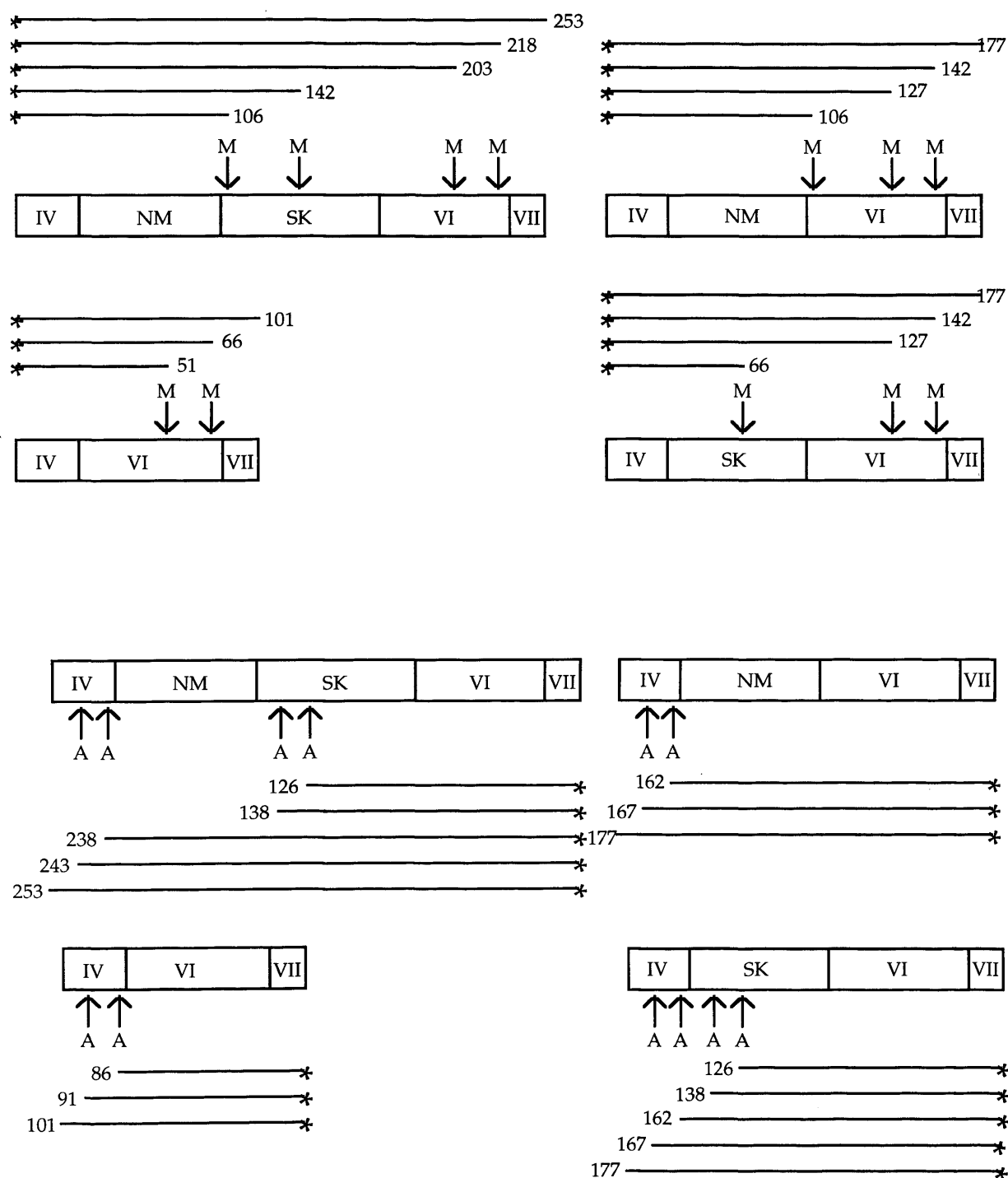
Nested product	Undigested	AluI digest	AluI partial	MboII digest	MboII partial
IV/NM/SK/VI	253	126	138, 238, 243	106	142, 203, 218
IV/NM/VI	177	162	167	106	127, 142
IV/SK/VI	177	126	138, 162, 167	66	127, 142
IV/VI	101	86	91	51	66

Table A1.1 Sizes, in nucleotides, of products of nested RT-PCR analysis before and after digestion with MboII and AluI (sizes of partial digestion products are also given)

Quantification of radioactive nested PCR products

It was important to show that the RT-PCR analysis did not distort the relative levels calculated for the spliced isoforms from their respective levels in the RNA preparation. To see how the levels of the different isoforms varied relative to each other during the course of the PCR analysis an aliquot of a nested reaction was removed after successive cycles of amplification of first round products generated from myoblast RNA. Each aliquot was then loaded onto a denaturing polyacrylamide gel. After electrophoresis the gel was fixed, dried and analysed using a PhosphorImager. The scan of the relevant part of the gel is shown in figure A1.3A. The amount of radioactivity and the relative levels, in percent, of the NM and SK and skipped isoforms as determined by the PhosphorImager are shown in table A1.2. The change in amount of radioactivity for each isoform in the successive cycles is shown graphically in figure A1.3B.

	Amount of radioactivity		Relative level (%)	
Cycle number	NM and SK	Skipped	NM and SK	Skipped
x	910	950	49	51
x+1	1644	1895	46	54
x+2	3592	3324	52	48
x+3	6503	6247	51	49
x+4	13507	14258	49	51
x+5	24114	24336	50	50
x+6	60239	58859	51	49

Table A1.2 Relative levels of the NM and SK and skipped isoforms in successive nested PCR cycles (x=14)

It can be seen that the relative levels of each isoform are maintained over seven cycles of nested PCR. The amount of radioactivity increased by over 60 times for each isoform during this amplification. Other experiments confirmed

Figure A1.3 Relative levels of TMnm mRNA isoforms
in successive cycles of nested PCR

A) Aliquots from successive cycles of the PCR reaction were taken and then loaded on a denaturing polacrylamide gel. The figure shows the PhosphorImager scan obtained after the gel had been fixed, dried and exposed to a PhosphorImager cassette. The NM and SK isoforms are not distinguishable on the gel because they are the same length.
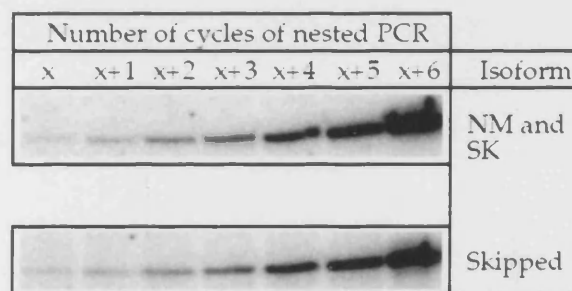
B) Graph showing the amount of radioactivity for the skipped and NM and SK isoforms in successive cycles of PCR as determined using the PhosphorImager.
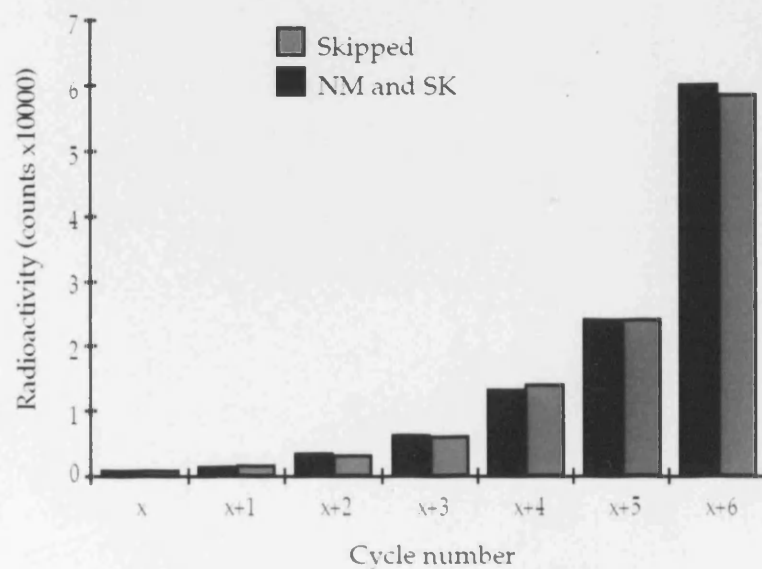
that the relative levels of isoforms were also not distorted during the first PCR reaction (data not shown: Raymond, unpublished).

In all the reverse-transcription/PCR amplification experiments described in chapters 3 and 4, controls were performed in which RNA from mock transfections was purified and subsequently amplified in parallel with the experimental samples. For each control no amplified products were detected, confirming that the amplified products visible after analysis of the experimental samples were derived from expression of transfected DNA and not from endogenous cellular mRNA.

Appendix 2
PhosphorImager data

PhosphorImager data for chapters 3 and 4

The PhosphorImager was used to calculate the relative level of radioactivity resulting from nested PCR/restriction enzyme analysis of TMnm minigene mRNA isoform expression in each lane of figures 3.2, 3.5, 3.6, 4.2, 4.3, and 4.4 and for lanes 1-3, 8-10 of figure 3.4A. The relative levels (shown as percent) are given in tables A2.2-A2.8 below. Table A2.1 gives the expected sizes of products resulting from the nested PCR/restriction enzyme analysis before and after digestion with AluI or MboII.

Isoform	Undigested	AluI digest	AluI partial	MboII digest	MboII partial
NM/SK	253	126	138, 238, 243	106	142, 203, 218
NM	177	162	167	106	127, 142
SK	177	126	138, 162, 167	66	127, 142
Skipped	101	86	91	51	66

Table A2.1 Expected sizes, in nucleotides, of nested PCR products from nested PCR/restriction enzyme analysis of TMnm minigene isoform expression *in vivo* before and after digestion with restriction enzyme.

Table A2.1 shows that some of the digested and partially digested nested PCR products derived from restriction enzyme analysis of different amplified isoforms are the same size. Despite this fact, the relative levels of each amplified isoform can still be calculated even if partial digestion occurred:

For each transfection, the contribution to the relative isoform level by the skipped isoform and by the NM/SK isoform (if any) can be determined from the undigested nested PCR products. Table A2.1 shows that the skipped isoform is amplified as a 101 nucleotide band by the nested PCR procedure. Thus, the relative level of this band in each lane was taken as the relative level of the skipped isoform expressed in the corresponding transfection. The NM/SK isoform is amplified as a 253 nucleotide band by the nested PCR procedure. Thus, the relative level of this band in each lane was taken as the relative level of the NM/SK isoform expressed in the corresponding transfection. The NM and SK isoforms are each amplified as 177 nucleotide products. The relative levels of these isoforms can, therefore, only be determined from the digested nested PCR products.

Table A2.1 shows that the NM isoform-derived 177 nucleotide nested PCR product is digested to a 106 nucleotide product and partially digested to 127 and 142 nucleotide products by MboII. The SK isoform-derived 177 nucleotide nested

PCR product is digested to a 66 nucleotide product and partially digested to 127 and 142 nucleotide products by MboII. Thus, the 106 and 66 nucleotide products are distinctive for expression of the NM and SK isoforms. However, the 106 nucleotide product can also arise from MboII digestion of the 253 nucleotide NM/SK isoform-derived nested PCR product and the 66 nucleotide product can also arise from partial MboII digestion of the 101 nucleotide skipped isoform-derived nested PCR product.

In order to calculate the amount of the 66 nucleotide product that results from complete digestion of the SK isoform-derived 177 nucleotide nested PCR product, the relative level of the 101 nucleotide band determined from the undigested samples was subtracted from the total relative level of the 66 nucleotide band and the 51 nucleotide band (this band results from complete digestion of the 101 nucleotide skipped isoform-derived nested PCR product by MboII). For those experiments where no 253 nucleotide band representing expression of the NM/SK isoform was visible, the 106 nucleotide band was due solely to MboII digestion of the NM isoform-derived nested PCR product. When the 253 nucleotide band was visible, the amount of the 106 nucleotide band that resulted from MboII digestion of the 177 nucleotide NM isoform-derived nested PCR product was calculated by subtracting the relative level of the 253 nucleotide band in the undigested samples from the relative level of the 106 nucleotide band in the MboII digested samples. In this calculation it was assumed that complete digestion of the 253 nucleotide band had occurred and that any 142 nucleotide band visible after analysis of the digested samples was derived only from partial digestion of nested PCR products derived from the NM and SK isoforms. This assumption is unlikely to have significantly affected the results because the relative level of the NM/SK isoform-derived nested PCR product and of the 142 nucleotide partial digestion product was always low.

Once a value for the 106 and 66 nucleotide products resulting from complete digestion of the NM and SK isoform-derived nested PCR products had been calculated, the ratio of NM isoform to SK isoform expression was determined. This ratio was then used to calculate the relative contribution of partially MboII-digested NM and SK isoform-derived 177 nucleotide products to the 142 and 127 partially digested products. The relative levels of the NM and SK isoforms expressed in the transfections were then determined by adding the values obtained from this calculation to those obtained for the contribution of the NM and SK isoform-derived 177 nested PCR products to the 106 and 66 nucleotide digestion products, respectively.

The above assumption that the relative contribution to the 142 and 127 partially digested products by digestion of the NM and SK isoform-derived 177 nucleotide products was directly proportional to the calculated ratio of NM to SK isoform expression was based on the observation that incomplete digestion may

be due to heteroduplex formation between the undigested products before they are digested with MboII restriction enzyme. It is proposed that heteroduplexes form between the NM and SK isoform-derived products, but also between the skipped isoform-derived product and the NM or SK isoform-derived products; partial digestion is evident in results where NM and skipped isoforms only were expressed (for example, figure 3.2, lanes 4-6. There are no examples in the gels shown where SK and skipped isoforms only were expressed). Thus, the heteroduplex populations are likely to contain NM and SK isoform-derived products proportional to the total amounts of those products in the undigested samples. The 142 and 127 nucleotide products of partial MboII digestion will then contain amounts of the NM and SK isoform-derived products proportional to the total amount of those products. In the samples analysed the proportion of the partial digestion products relative to the total is low. Thus, any error in assignment of the partial digestion products is likely to be small.

In the section of the figures showing undigested samples, the 253 nucleotide band is indicated as the "NM/SK" isoform, the 177 nucleotdie band is indicated as the "NM and SK" isoforms and the 101 nucleotide band is indicated as the "skipped" isoform. In the section of the figures showing MboII digested samples, the 142 and 127 nucleotide bands are indicated as the "NM and SK" isoforms, the 106 nucleotide band is indicated as the "NM" isoform, the 66 nucleotide band is indicated as the "SK and skipped" isoforms and the 51 nucleotide band is indicated as the "skipped" isoform.

The calulations for each experiment can be represented by the following equations:

A=142+127 (partial digestion of nested PCR products derived from NM/SK, NM and SK isoforms. The contribution to the 142 nucleotide band from partial digestion of NM/SK isoform-derived PCR product was ignored in the calculations because it was considered to be negligible);

B=106 (complete digestion of NM and complete digestion NM/SK isoform-derived PCR products);

C=66 (complete digestion of SK and partial digestion of skipped isoform-derived PCR products);

D=51 (complete digestion of skipped isoform-derived PCR product);

E=106-253 (complete digestion of NM isoform-derived PCR product);

F=[66+51]-101 (complete digestion of SK isoform-derived PCR product);

G=253 (derived from NM/SK isoform);

H=101 (derived from skipped isoform);

where each number refers to the relative level of the corresponding nested product or MboII digestion product. The relative levels for expression of each mRNA isoform are then given by:

NM/SK=G

NM=E+[A x (E/(E+F))]

SK=F+[A x (F/(F+E))]

Skipped=H

For each experiment, the results were calculated for each lane. If duplicate transfections were done, averages were calculated and the difference of each value from the mean given. If triplicate transfections were done, averages were again calculated and standard deviations were given according to the formula:

$$\sqrt{\{[\Sigma(x-\underline{x})^2]/n\}}$$

where x represents each value, $\underline{x}$ represents the average of the values, n represents the number of values.

For some of the mutations tested, the restriction enzyme sites critical for distinguishing the NM and SK isoform-derived nested PCR products are no longer present in the mutant sequence. These are referred to specifically in the text relevant to those experiments.

Values for the relative expression of TMnm minigene mRNA isoforms in the experiments illustrated in figures 3.3 and 3.4B were not calculated. In these experiments, the samples were analysed using AluI restriction enzyme. If partial digestion occurs using this method of analysis (as in both experiments), there is no digestion product specifically derived from the NM isoform (see table A2.1). Consequently, the relative contribution of the NM and SK isoforms to the 177 nucleotide nested PCR product cannot be determined accurately.

Three bands are visible at the top of lanes 1-10 of figure 4.3. It was assumed, for the purpose of quantification, that these three bands represent the 253 nucleotide nested PCR product derived from the NM/SK spliced isoform. The triple banding pattern is likely to be due to incomplete denaturation of the sampes before they entered the denaturing polyacrylamide gel or to renaturation of previously denatured samples. When this occurs, the samples exist as a mixture of single and double stranded DNA populations and consequently do not run as a uniform band on the gel. This interpretation is consistent with the mobility of the bands visible here. The fastest migrating band has a mobility equivalent to a single stranded product 253 nucleotides long. The other bands have slower mobility, consistent with their identity as double stranded products.

TM plasmid	Lane number, figure 3.2	Undigested PCR products			Lane number, figure 3.2	MboII-digested PCR products				
		NM/SK (253)	NM and SK (177)	Skipped (101)		142	127	106	66	51
SK WT	1	-	56.6	43.4	10	5.5	14.4	40.8	6.9	32.4
	2	-	59.2	40.8	11	1.7	15	41.8	5.6	35.9
	3	-	62.4	37.6	12	4.1	15.5	45.5	4.3	30.6
SK1-15NM	4	0.7	90.3	9	13	1.4	12.2	44.8	36.8	4.8
	5	2.6	83	14.4	14	2.6	12.1	41.9	34.3	9.1
	6	0.0	83.3	16.7	15	2.1	8.6	38.7	39.6	11
SK61-75NM	7	-	94.3	5.7	16	0	2.4	47.2	44.7	5.7
	8	-	95.3	4.7	17	0	3.7	46	45.5	4.8
	9	-	95.6	4.4	18	0	3.6	43.7	48.7	4

TM plasmid	NM/SK	NM	SK	Skipped
SK WT	-	56.7	-	43.3
	-	59.2	-	40.8
	-	62.4	-	37.6
1-15NM	0.7	51.92	38.38	9
	2.6	47.76	35.24	14.4
	0.0	44.40	38.90	16.7
61-75	-	48.43	45.87	5.7
	-	47.86	47.44	4.7
	-	45.41	50.19	4.4

Table A2.2 A) Data (as percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 3.2. The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products B) Relative levels (as percent) of each TMnm isoform calculated from the PhosphorImager data

A)

TM plasmid	NM and SK (177)	Skipped (101)	Lane number, figure 3.4	MboII-digested PCR products			
				127	106	66	51
SK WT	62.0	38.0	1	9.3	22.2	20.0	48.5
	59.2	40.8	2	10.6	19.3	21.8	48.4
	56.6	43.4	3	11.2	16.3	21.0	51.5
SK16-30R3	40.1	59.9	8	6.4	31.5	3.5	58.6
	46.7	53.4	9	7.0	38.4	3.1	51.6
	37.7	62.3	10	9.0	25.1	3.7	62.2

B)

TM plasmid	NM	SK	Skipped
SK WT	27.1	24.4	48.5
	24.3	27.3	48.4
	21.2	27.3	51.5
SK16-30R3	37.4	2.7	59.9
	45.1	1.6	53.4
	33.0	4.7	62.3

Table A2.3   **A)** Data (in percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 3.4A
The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products
B) Relative levels (in percent) of each TMnm isoform calculated from the PhosphorImager data

A)

TM Plasmid	Lane number, figure 3.5A	Undigested PCR products		Lane number, figure 3.5B	MboII-digested PCR products			
		NM and SK (177)	Skipped (101)		127	106	66	51
SK WT	1	58.8	41.2	1	9.4	50.1	0	40.5
	2	55.6	44.4	2	9.1	49.5	2.7	38.7
	3	54.3	45.7	3	8.2	46.2	1.9	43.7
NM5'ssC	4	89.3	10.7	4	4	89.2	-	6.8
	5	87.6	12.4	5	4.6	88.4	-	7
	6	89.5	10.5	6	5.1	87.4	-	7.5
ΔNM5'ss	7	9.2	90.8	7	5.1	-	6	88.9
	8	7.6	92.4	8	4.5	-	4.7	90.8
	9	9.6	90.4	9	5.6	-	4.8	89.6
ΔSKBP	10	54.7	45.3	10	11.4	53.7	-	34.9
	11	49	51	11	10.1	47.4	-	42.5
	12	52.5	47.5	12	10.7	47.5	-	41.8

Table A2.4   A) Data (in percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 3.5
The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products

B)

TM plasmid	NM/SK	NM	SK	Skipped
SKWT	0	58.8	0	41.2
	0	55.6	0	44.4
	0	54.3	0	45.7
NM5'ssC	0	89.3	0	10.7
	0	87.6	0	12.4
	0	89.5	0	10.5
ΔNM5'ss	0	0	9.2	90.8
	0	0	7.6	92.4
	0	0	9.6	90.4
ΔSK BP	0	54.7	0	45.3
	0	49	0	51
	0	52.5	0	47.5

Table A2.4 **B)** Relative levels (in percent)of each TM isoform calculated from the PhosphorImager data

A)

TM plasmid	Lane number, figure 3.6	Undigested PCR products		Lane number, figure 3.6	MboII-digested PCR products					
		NM and SK (177)	Skipped (101)		177	127	106	101	66	51
WT	1	43.2	56.8	10	5.3	16.2	6.5	7	22.4	42.6
	2	43.5	56.5	11	3.7	15.1	5.4	5.4	23.7	46.7
	3	43.9	56.1	12	4.1	16.1	6.5	8.1	22.7	42.5
1-15	4	100	-	13	-	-	-	-	100	-
	5	100	-	14	-	-	-	-	100	-
	6	100	-	15	-	-	-	-	100	-
61-75	7	100	-	16	-	-	-	-	100	-
	8	100	-	17	-	-	-	-	100	-
	9	100	-	18	-	-	-	-	100	-

B)

TM plasmid	NM	SK	Skipped
SKWT	12.9	30.3	56.8
	9.5	34.0	56.5
	12.0	31.9	56.1
SK1-15NM	-	100.0	-
	-	100.0	-
	-	100.0	-
SK61-75NM	-	100.0	-
	-	100.0	-
	-	100.0	-

Table A2.5 **A)** Data (in percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 3.6 The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products **B)** Relative levels (in percent) of each TM isoform calculated from the PhosphorImager data

A)

TM plasmid	U1 gene	Lane number, figure 4.3	Undigested PCR products			Lane number, figure 4.3	MboII-digested PCR products					
			NM/SK	NM and SK	Skipped		177	142	127	106	66	51
SK2-4CGC	-	1	5.7	77.3	17	11	2.2	2.4	14.5	45.1	20.4	15.4
		2	4.3	71.4	24.3	12	3.4	1.4	13.5	39.6	20.2	21.9
SK1-15NM	-	3	8.1	69.8	22.1	13	7.8	1.5	13.5	35.2	22	20
		4	9	73.2	17.8	14	3.6	1.3	12.9	39.2	23.2	19.8
SK2-4CGC	WT	5	8	67.2	24.8	15	1.8	1.6	11.6	41.5	18.7	24.8
		6	7.7	68.6	23.7	16	2	1.3	11.8	45.9	17.2	21.8
SK2-4CGC	α-CGC	7	3.9	65.3	30.8	17	0	0.7	14.9	43.1	10.5	30.8
		8	3.6	63.9	32.5	18	0	0	13.5	43.9	9.3	33.3
SK1-15NM	α-CGC	9	10.2	62.7	27.1	19	3.3	2	11.3	36.1	25.3	22
		10	9.7	66.3	24	20	1.2	2.1	12.8	36.9	23	24

B)

TM plasmid	U1 gene	NM/SK	NM	SK	Skipped
SK2-4CGC	-	5.7	52.1	25.3	17.0
		4.3	46.8	24.6	24.3
SK1-15NM	-	8.1	39.2	30.6	22.1
		9	39.6	33.6	17.8
SK2-4CGC	WT	8	42.8	24.4	24.8
		7.7	48.6	20.0	23.7
SK2-4CGC	α-CGC	3.9	51.5	13.8	30.8
		3.6	51.1	12.8	32.5
SK1-15NM	α-CGC	10.2	34.8	27.9	27.1
		9.7	35.8	30.5	24.0

Table A2.7 **A)** Data (in percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 4.3
The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products
B) Relative levels (in percent) of each TM isoform calculated from the PhosphorImager data

A)

TM plasmid	U1 gene	Lane number, figure 4.4	Undigested PCR products			Lane number, figure 4.4	MboII-digested PCR products				
			NM/SK (253)	NM and SK (177)	Skipped (101)		142	127	106	66	51
SK2-4CGC	WT	1	1.4	80.1	18.6	7	0.6	13.5	42.3	26.4	17.2
		2	1.4	92.3	6.3	8	0.3	9.1	63.4	22.8	4.4
SK2-4CGC	α-CGC	3	1.7	68.3	29.9	9	0.6	13.7	47	15.2	23.5
		4	0.6	72.1	27.3	10	0.3	12.7	48.4	14.5	24.1
SK1-15NM	α-CGC	5	2.1	81.4	16.5	11	1	12.7	36.4	35	14.9
		6	2.4	81.9	15.7	12	1	10.9	34.9	36.6	16.6
SK1-15NM	WT	13	0	90	10	15	1.3	14	45.1	32.9	6.7
		14	0	90.2	9.8	16	1.8	13.5	42	33.7	9

B)

TM plasmid	U1 gene	NM/SK	NM	SK	Skipped
SK2-4CGC	WT	1.4	49.65	30.35	18.6
		1.4	69.03	23.27	6.3
SK2-4CGC	α-CGC	1.7	57.27	11.13	29.9
		0.6	58.31	13.79	27.3
SK1-15NM	α-CGC	2.1	41.24	40.16	16.5
		2.4	38.03	43.88	15.7
SK1-15NM	WT	0	54.34	35.66	10
		0	50.58	39.62	9.8

Table A2.8   **A)** Data (in percent) from PhosphorImager quantification of RT-PCR analysis shown in figure 4.4
The numbers heading the columns refer to the size in nucleotides of the undigested and digested nested PCR products
B) Relative levels (in percent) of each TM isoform calculated from the PhosphorImager data

PhosphorImager data for figure 5.1

The data obtained from the PhosphorImager for lanes 1-12 of figure 5.1 is tabulated below. The relative values for the lariats and pre-mRNA in each lane are shown in percent. The lariat intermediates were quantified as the top two lariat bands, and the lariat products as the lower two lariat bands.

Pre-mRNA	NM-SKWTbpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	16.3	16.2	29.1	28.1
Lariat product	1.2	1.2	18.1	17.8
Pre-mRNA	82.5	82.7	52.9	54.1

Pre-mRNA	NM-SK16-30NMbpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	6.3	8.3	4.4	4.3
Lariat product	1.1	1.3	4.3	4.4
Pre-mRNA	92.6	90.5	91.3	91.3

Pre-mRNA	NM-SK16-30R1bpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	4.1	4.5	5.2	5.4
Lariat product	0.4	0.5	12	7.4
Pre-mRNA	95.4	95	82.8	87.2

Table A2.9 Data for the relative levels (in percent) of lariat products, lariat intermediates, and pre-mRNA in figure 5.1.

The transcripts used for the *in vitro* splicing experiment were continuously labelled with ^{32}P-UTP. The lariat products, lariat intermediates, and pre-mRNA contain different numbers of uridine nucleotides to each other because they are of different lengths. The lariat intermediates should contain 228 uridines, the lariat products 213, and the pre-mRNA 243. Thus, the amount of radioactive label incorporated into these respective RNAs is different. Consequently, the relative level of radioactivity for these RNAs in each lane is not directly representative of their relative molar levels. For each mole of pre-mRNA, the amount of radioactivity recorded will be higher than for each mole of lariat product or intermediate RNA. The relative molar values of each RNA were determined by correcting the values given in table A2.9. The corrected values are shown in table A2.10. The corrected values were calculated by scaling up the relative amount of radioactivity for the lariat intermediates and products by a factor of 243/228 and 243/213, respectively. The relative molar values, in percent, for the pre-mRNA, lariat intermediate and lariat product were then determined using the original value for the pre-mRNA given in table A2.9 and the corrected values for the lariat intermediate and product RNAs. Table A2.11 gives the averages of the values shown in table A2.10. Table 5.1 shows the lariat intermediates and

products as a percentage of the total of lariats and pre-mRNA calculated from the average values given in table A2.11.

Pre-mRNA	NM-SKWTbpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	17.1	17.0	29.7	28.7
Lariat product	1.4	1.4	19.7	19.5
Pre-mRNA	81.4	81.6	50.6	51.8

Pre-mRNA	NM-SK16-30NMbpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	6.7	8.8	4.6	4.5
Lariat product	1.2	1.5	4.9	5.0
Pre-mRNA	92.0	89.8	90.5	90.5

Pre-mRNA	NM-SK16-30R1bpup			
Time (hours)	1.5	1.5	3	3
Lariat intermediate	4.4	4.8	5.4	5.7
Lariat product	0.5	0.6	13.4	8.3
Pre-mRNA	95.2	94.7	81.2	86.0

Table A2.10 The relative molar values of pre-mRNA, lariat intermediate and lariat product for the experiment described in figure 5.1

Pre-mRNA	NM-SKWTbpup	
Time (hours)	1.5	3
Lariat intermediate	17.1	29.2
Lariat product	1.4	19.6
Pre-mRNA	81.5	51.2

Pre-mRNA	NM-SK16-30NMbpup	
Time (hours)	1.5	3
Lariat intermediate	7.8	4.6
Lariat product	1.3	4.9
Pre-mRNA	90.9	90.5

Pre-mRNA	NM-SK16-30R1bpup	
Time (hours)	1.5	3
Lariat intermediate	4.6	5.6
Lariat product	0.5	10.8
Pre-mRNA	94.9	83.6

Table A2.11 Averages of the values given in table A2.10.

References

AbdulManan, N., S. M. OMalley and K. R. Williams (1996) Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry* **35** (11) 3545-3554.

Adams M.D., Tarng R.S., and Rio D.C. (1997) The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes and Development* **11** 129-138.

Amendt, B.A., Si, Z-H., Stoltzfus, C.M. (1995) Presence of exon splicing silencers within HIV type 1 tat exon 2 and tat-rev exon 3: evidence for inhibition mediated by cellular factors. *Molecular and Cellular Biology* **15** (8) 4606-4615.

Amrein, H., M. Gorman and R. Nothiger (1988) The sex-determining gene tra-2 of Drosophila encodes a putative RNA binding protein. *Cell* **55** (6) 1025-1035.

Amrein, H., Maniatis T., and Nothiger R. (1990) Alternatively spliced transcripts of the sex determining gene tra-2 of Drosophila encode functional proteins of different size. *EMBO J* **9** 3619-3629.

Amrein, H., Hedley M.L., and Maniatis T. (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** 735-748.

Ayane M., Preuss U., Kohler G., Nielson P.J. (1991) A differentially expressed murine RNA encoding a protein with similarities to two types of nucleic acid binding motifs. *Nucleic Acids Res* **19** 1273-1278.

Baker, B. S. and Ridge K. (1980) *Genetics* **94** 383-423.

Baker, B. S. (1989) Sex in flies: The splice of life. *Nature* **340** (6234) 521-524.

Balvay, L., D. Libri, M. Gallego and M. Y. Fiszman (1992) Intronic sequence with both negative and positive effects on the regulation of alternative transcripts of the chicken beta-tropomyosin transcripts. *Nucleic Acids Research* **20** (15) 3987-3992.

Barnes W.M. (1993) PCR amplification of upto 35Kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc Natl Acad Sci* **91** 2216-2220.

Beisel, K. W. and J. E. Kennedy (1994) Identification of novel alternatively spliced isoforms of the tropomyosin-encoding gene, TMnm, in the rat cochlea. *Gene* **143** (2) 251-256.

Bell, L. R., J. I. Horabin, P. Schedl and T. W. Cline (1991) Positive autoregulation of Sex-lethal by alternative splicing maintains the female determined state in Drosophila. *Cell* **65** (2) 229-239.

Bennett, M., S. Michaud, J. Kingston and R. Reed (1992) Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes and Development* **6** (10) 1986-2000.

Bennett, M., S. PinolRoma, D. Staknis, G. Dreyfuss and R. Reed (1992) Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Molecular and Cellular Biology* **12** (7) 3165-3175.

Bernards A., Rubin C. M., Westbrook C. A., Paskind M., and Baltimore D. (1987) The first intron in the human c-abl gene is at least 200Kb long and is a target for translocations in chronic myelogenous leukemia. *Molecular and Cellular Biology* **7** (9) 3231-3236.

Black, D. L. (1991) Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes and Development* **5** (3) 389-402.

Black, D. L. (1992) Activation of c-src neuron-specific splicing by an unusual RNA element in vivo and in vitro. *Cell* **69** (5) 795-807.

Blencowe B.J., Issner R., Kim J., McCaw P., and Sharp P.A. (1995) New proteins related to the Ser-Arg family of splicing factors. *RNA* **1** 852-865.

Boggs R.T., Gregor P., Idriss S., Belote J.M., and McKeown M. (1987) Regulation of sexual differentiation in D.Melanogaster via alternative splicing of RNA from the transformer gene. *Cell* **50** 739-747.

Bopp, D., J. I. Horabin, R. A. Lersch, T. W. Cline and P. Schedl (1993) Expression of the sex-lethal gene is controlled at multiple levels during Drosophila oogenesis. *Development* **118** (3) 797-812.

Burd, C.G., and Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO Journal* **13** (5) 1197-1204.

Burtis, K. C. and B. S. Baker (1989) Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56** (6) 997-1010.

Burtis, K. C. (1993) The regulation of sex determination and sexually dimorphic differentiation in Drosophila. *Current Opinion in Cell Biology* **5** (6) 1006-1014.

Buvoli, M., F. Cobianchi, G. Biamonti and S. Riva (1990) Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Research* **18** (15) 6595-6600.

Caceres, J. F. and A. R. Krainer (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *Embo Journal* **12** (12) 4715-4726.

Caceres, J. F., S. Stamm, D. M. Helfman and A. R. Krainer (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265** (5179) 1706-1709.

Caputi, M, Casari G, Guenzi S, Tagliabue R, Sidoli A, et al, 1994. A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *NAR* **22**(6) 1018-1022.

Cartegni, L., M. Maconi, E. Morandi, F. Cobianchi, S. Riva and G. Biamonti (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *Journal of Molecular Biology* **259** (3) 337-348.

Cavaloc, Y., M. Popielarz, J. P. Fuchs, R. Gattoni and J. Stevenin (1994) Characterization and cloning of the human splicing factor 9G8: A novel 35 kDa factor of the serine/arginine protein family. *Embo Journal* **13** (11) 2639-2649.

Chabot, B. and J. A. Steitz (1987) Recognition of mutant and cryptic 5' splice sites by the U1 small nuclear ribonucleoprotein in vitro. *Molecular and Cellular Biology* **7** (2) 698-707.

Chan, R. C. and D. L. Black (1995) Conserved intron elements repress splicing of a neuron-specific c-src exon in vitro. *Molecular and Cellular Biology* **15** (11) 6377-6385.

Chaudhary, N., C. McMahon and G. Blobel (1991) Primary structure of a human arginine-rich nuclear protein that colocalizes with spliceosome components. *Proceedings of the National Academy of Sciences of the United States of America* **88** (18) 8189-8193.

Cho Y., Liu J., and Hitchcock-DeGregori S.E. (1990) The amino terminus of muscle tropomyosin is a major determinant for function. *J Biol Chem* **265** 538-545.

Clayton L., Reinach F.C, Chumbley G.M. and MacLeod A.R. (1988) Organisation of the hTMnm gene. *J Mol Biol* **201** 507-515.

Clouet d'Orval B., Yves d'Aubenton-Carafa, Marie J., and Brodie E. (1991). RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science* **252** 1823-1828.

Clouet d'Orval B., Yves d'Aubenton-Carafa, Sirand-Pugnet P., Gallego M., Brodie E. and Marie J. (1991). Determination of an RNA secondary structure involved in splicing inhibition of a muscle-specific exon. *J Mol Biol* **221** 837-856.

Colwill K., Pawson T., Andrews B., Prasad J., Manley J.L., Bell J.C., and Duncan P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intracellular distribution. *EMBO J.* **15** 265-275.

Coulter L.R., Landree M.A., and Cooper T.A. (1997) Identification of a new class of exonic splicing enhancers by in vivo selection. *Molecular and Cellular Biology* **17** (4) 2143-2150.

Crispino, J. D., B. J. Blencowe and P. A. Sharp (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science* **265** (5180) 1866-1869.

Crispino, J. D., and Sharp P. A. (1995) A U6 snRNA:pre-mRNA interaction can be rate-limiting for U1-independent splicing. *Genes and Development* **9** 2314-2323.

Dauwalder B.F., Amayamanzanares F., and Mattox W. (1996) A human homologue of the Drosophila sex determination factor tra-2 has conserved splicing regulatory functions. *Proc Natl Acad Sci* **93** 9004-9009.

Deshpande, G., M. E. Samuels and P. D. Schedl (1996) Sex-lethal interacts with splicing factors in vitro and in vivo. *Molecular and Cellular Biology* **16** (9) 5036-5047.

Dirksen, W. P., R. K. Hampson, Q. Sun and F. M. Rottman (1994) A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *Journal of Biological Chemistry* **269** (9) 6431-6436.

Dominski, Z. and R. Kole (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Molecular and Cellular Biology* **11** (12) 6075-6083.

Dominski, Z. and R. Kole (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Molecular and Cellular Biology* **12** (5) 2108-2114.

Dreyfuss G., Matunis M.J., Pinol-Roma S., and Burd C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62** 289-321.

Eperon I.C. and Hamshere M. (1993) Applications of gene transfer to study RNA splicing in mammalian cell lines. *Journal of Tissue Culture Methods* **15** 99-107.

Eperon, I. C., D. C. Ireland, R. A. Smith, A. Mayeda and A. R. Krainer (1993) Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *Embo Journal* **12** (9) 3607-3617.

Eperon, L. P., J. P. Estibeiro and I. C. Eperon (1986) The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. *Nature* **324** (6094) 280-282.

Feener, C. A., Koenig, M., an Kunkel L. M. (1989) Alternative splicing of human dystrophinmRNA generates isoforms at the carboxy terminus. *Nature* **338** (6215) 509-511.

Frendewey, D. and W. Keller (1985) Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42** (1) 355-367.

Fu, X. Y. and J. L. Manley (1987) Factors influencing alternative splice site utilization in vivo. *Molecular and Cellular Biology* **7** (2) 738-748.

Fu, X. D. and T. Maniatis (1990) Factor required for mammalian spliceosome assmebly is localised to discrete regions in the nucleus. *Nature* **343** 437-441.

Fu, X. D., R. A. Katz, A. M. Skalka and T. Maniatis (1991) The role of branchpoint and 3'-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes and Development* **5** (2) 211-220.

Fu, X. D. and T. Maniatis (1992) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* **256** (5056) 535-538.

Fu, X. D. and T. Maniatis (1992) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. *Proceedings of the National Academy of Sciences of the United States of America* **89** (5) 1725-1729.

Fu, X. D., A. Mayeda, T. Maniatis and A. R. Krainer (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) 11224-11228.

Fu, X. D. (1993) Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature* **365** (6441) 82-85.

Fu X. Y. (1995) The superfamily of arginine/serine-rich splicing factors. *RNA* **1** 663-680.

Gallego, M. E., L. Balvay and E. Brody (1992) cis-Acting sequences involved in exon selection in the chicken beta- tropomyosin gene. *Molecular and Cellular Biology* **12** (12) 5415-5425.

Gallego M.E., Gattoni R., Stevenin J., Marie J., and Expert-Bezncon A. (1997) The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splcing of the beta-tropomyosin alternative exon 6A. *EMBO Journal* **16** (7) 1772-1784.

Gamberi, Izaurralde, Beisel, and Mattaj (1997) Interaction between the human nuclear cap-binding protein complex and hnRNP F. *Molecular and Cellular Biology* **17** (5) 2587-2597.

Gatto, F.D. and Breathnach, R. (1995) Exon and intron sequences, respectively, repress and activate splicing of a fibroblast growth factor receptor 2 alternative exon. *Molecular and Cellular Biology* **15**(9) 4825-4834.

Ge, H. and J. L. Manley (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* **62** (1) 25-34.

Ge, H., P. Zuo and J. L. Manley (1991) Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell* **66** (2) 373-382.

Gimona, M., A. Watakabe and D. M. Helfman (1995) Specificity of dimer formation in tropomyosins: Influence of alternatively spliced exons on homodimer and heterodimer assembly. *Proceedings of the National Academy of Sciences of the United States of America* **92** (21) 9776-9780.

Gontarek, R. R. and D. Derse (1996) Interactions among SR proteins, an exonic splicing enhancer, and a lentivirus Rev protein regulate alternative splicing. *Molecular and Cellular Biology* **16** (5) 2325-2331.

Gooding, C., G. C. Roberts, G. Moreau, B. NadalGinard and C. Smith (1994) Smooth muscle-specific switching of alpha-tropomyosin mutually exclusive exon selection by specific inhibition of the strong default exon. *Embo Journal* **13** (16) 3861-3872.

Goodwin L.O., Lees M.J., Leonard M.A., Cheley S.B., Helfman D.M. (1991) Four fibroblast tropomyosin isoforms are expressed from the rat alpha tropomyosin gene via alternative RNA splicing and the use of two promoters. *J Biol Chem* **266** 8408-8415.

GouxPelletan, M., D. Libri, Y. DAubentonCarafa, M. Fiszman, E. Brody and J. Marie (1990) In vitro splicing of mutually exclusive exons from the chicken beta-tropomyosin gene: Role of the branch point location and very long pyrimidine stretch. *Embo Journal* **9** (1) 241-249.

Grabowski, P. J., S. R. Seiler and P. A. Sharp (1985) A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42** (1) 345-353.

Graham, I. R., M. Hamshere and I. C. Eperon (1992) Alternative splicing of a human alpha-tropomyosin muscle-specific exon: Identification of determining sequences. *Molecular and Cellular Biology* **12** (9) 3872-3882.

Green, M. R. (1991) Biochemical mechanisms of constitutive and regulated Pre-mRNA splicing. *Annual Review of Cell Biology* **7** 559-599.

Gui J.F., Tronchere H., Chandler S.D., and Fu X.D. (1994) Purification and chartacerisation of a kinase specific for the serine- and arginine-rich pre-mRMA splicing factors. *Proc Natl Acad Sci* **91** 10824-10828.

Guo, W., G. J. Mulligan, S. Wormsley and D. M. Helfman (1991) Alternative splicing of beta-tropomyosin pre-mRNA: cis-acting elements and cellular factors that block the use of a skeletal muscle exon in nonmuscle cells. *Genes and Development* **5** (11) 2096-2107.

Harper, D. S., L. D. Fresco and J. D. Keene (1992) RNA binding capacity of a Drosophila snRNP protein that shares sequence homology with mammalian U1-A and U2-B' proteins. *Nucleic Acids Research* **20** (14) 3645-3650.

Hawkins JD, 1988 A survey on intron and exon lengths *NAR* **16**(21) 9893-9908.

Hedley, M. L. and T. Maniatis (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) 579-586.

Heinrichs, V., M. Bach, G. Winkelmann and R. Luhrmann (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247** (4938) 69-72.

Helfman D.M., Cheley S., Kuismanan E., Fin L.A., Yamawaki-Kataoka Y. (1986) Non muscle and muscle tropomyosin are expressed from a single gene by alternative splicing and polyadenylation. *Mol Cell Biol* **6** 3582-3595.

Helfman, D. M. and W. M. Ricci (1989) Branch point selection in alternative splicing of tropomyosin pre- mRNAs. *Nucleic Acids Research* **17** (14) 5633-5650.

Helfman, D. M., R. F. Roscigno, G. J. Mulligan, L. A. Finn and K. S. Weber (1990) Identification of two distinct intron elements involved in alternative splicing of beta-tropomyosin pre-mRNA. *Genes and Development* **4** (1) 98-110.

Hodgkin, J. (1989) Drosophila sex determination: A cascade of regulated splicing. *Cell* **56** (6) 905-906.

Hoffman BE and Grabowski PJ 1992 U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes and Dev* **6**(12B) 2554-2568.

Hong, Bennett, Xiao, Kramer, Wang, and Reed (1997) Association of U2 snRNP with the spliceosomal complex E. *NAR* **25** (2) 354-361.

Hoshijima, K., K. Inoue, I. Higuchi, H. Sakamoto and Y. Shimura (1991) Control of doublesex alternative splicing by transformer and transformer-2 in Drosophila. *Science* **252** (5007) 833-836.

Humphrey MB, Bryan J, Cooper TA, Berget SM 1995. A 32 nucleotide exon-splicing enhancer regulates usage of competing 5' splice sites in a differential internal exon *MCB* **15**(8) 3979-3988

Inoue, K., K. Hoshijima, H. Sakamoto and Y. Shimura (1990) Binding of the Drosophila Sex-lethal gene product to the alternative splice site of transformer primary transcript. *Nature* **344** (6265) 461-463.

Ishikawa R., Yamashiro S., and Matsumura F. (1989) Annealing of gelsolin-severed actin fragments by tropomyosin in the presence of Ca^{2+}. *J Biol Chem* **264** 16764-16770.

Ishikawa, F., M. J. Matunis, G. Dreyfuss and T. R. Cech (1993) Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)(n). *Molecular and Cellular Biology* **13** (7) 4301-4310.

Izaurralde, E., J. Stepinski, E. Darzynkiewicz and I. W. Mattaj (1992) A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs. *Journal of Cell Biology* **118** (6) 1287-1295.

Izaurralde, E., J. Lewis, C. McGulgan, M. Jankowska, E. Darzynkiewicz and I. W. Mattaj (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78** (4) 657-668.

Jamison, S. F., A. Crow and M. A. Garcia-Blanco (1992) The spliceosome assembly pathway in mammalian extracts. *Molecular and Cellular Biology* **12** (10) 4279-4287.

Jamison, S. F., Pasman Z., Wang J., Will C., Luhrmann R., Manley J.L., Garcia-Blanco M.A. (1995) U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterisation of required elements *Nucleic Acids Res* **23** 3260-3267.

Jancso A., and Graceffa P. (1991) Smooth muscle tropomyosin coiled-coil dimers. Subunit composition, assembly, and end to end interaction. *J Biol Chem* **266** 5891-5897.

Jiang S., and Eberhardt N.L. (1995) A microscale method to isolate DNA-binding proteins suitable for quantitative comparison of expression levels from transfected cells.*Nucleic Acids Res* **23** (17) 3607-3608.

Kanopka, A., O. Muhlemann and G. Akusjarvi (1996) Inhibition by SR proteins splicing of a regulated adenovirus pre-mRNA. *Nature* **381** (6582) 535-538.

Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. GarciaBlanco and J. L. Manley (1994) Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368** (6467) 119-124.

Konarska, M.M. and Sharpe, P.A. (1987) Interactions between snRNPs in formation of spliceosomes. *Cell* **49** 763-774.

Krainer, A. R., G. C. Conway and D. Kozak (1990) The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62** (1) 35-42.

Krainer, A. R., G. C. Conway and D. Kozak (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes and Development* **4** (7) 1158-1171.

Krainer, A. R., A. Mayeda, D. Kozak and G. Binns (1991) Functional expression of cloned human splicing factor SF2: Homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell* **66** (2) 383-394.

Kramer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annual Review of Biochemistry* **65** 367-409.

Kumar, A. and S. H. Wilson (1990) Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. *Biochemistry (Usa)* **29** (48) 10717-10722.

Kuo HC, Nasim FH, Grabowski PJ 1991. Control of alternative splicing by the differential binding of U1 snRNP. *Science* **251** (4997) 1045-1050.

Lamond, A. I. (1993) Splicing in stereo. *Nature* **365** (6444) 294-295.

Laski F.A. and Rubin G.M. (1989) Analysis of the cis-acting requirements for germ-line-specific splicing of the P-element ORF2-ORF3 intron. *Genes and Development* **3** 720-728.

Lavigueur A, La BH, Kornblihtt AR, Chabot B (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes and Dev* **7**(12A) 2405-2417

Lees-Miller J.P., Yan A., Helfman D.M. (1990) Structure and complete nucleotide sequence of the gene encoding rat fibroblast tropomyosin 4. *J Mol Biol* **213** 339-405.

Lees-Miller J.P., Goodwin L.O. and Helfman D.M. (1990) Three novel brain tropomyosin isoforms are expressed from the rat alpha tropomyosin gene through the use of alternative promoters and the use of laternative RNA processing. *Mol Cell Biol* **10** 1729-1742.

Lehrer S.S., and Stafford W.F.I. (1991) Preferential assembly of the tropomyosin heterodimer: equilibrium studies. *Biochemistry* **30** 5682-5688.

Lewis, J. D., E. Izaurralde, A. Jarmolowski, C. McGuigan and I. W. Mattaj (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes and Development* **10** (13) 1683-1698.

Lewis, J. D., D. Gorlich and I. W. Mattaj (1996) A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing. *Nucleic Acids Research* **24** (17) 3332-3336.

Libri, D., M. GouxPelletan, E. Brody and M. Y. Fiszman (1990) Exon as well as intron sequences are cis-regulating elements for the mutually exclusive alternative splicing of the beta tropomyosin gene. *Molecular and Cellular Biology* **10** (10) 5036-5046.

Libri, D., L. Balvay and M. Y. Fiszman (1992) In vivo splicing of the beta tropomyosin pre-mRNA: A role for branch point and donor site competition. *Molecular and Cellular Biology* **12** (7) 3204-3215.

Lin, R. J., A. J. Newman, S. C. Chen and J. Abelson (1985) Yeast mRNA splicing in vitro. *Journal of Biological Chemistry* **260** (27) 14780-14792.

Lin J.J., Hegmann T.E., Lin J.L. (1988) Differential localisation of tropomyosin isoforms in cultured non muscle cells. *J Cell Biol* **107** 563-572.

Lin CH and Patton JG, (1995). Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA* **1** 234-245.

Lou, H., Y. Yang, G. J. Cote, S. M. Berget and R. F. Gagel (1995) An intron enhancer containing a 5' splice site sequence in the human calcitonin/calcitonin gene-related peptide gene. *Molecular and Cellular Biology* **15** (12) 7135-7142.

Lou, H., R. F. Gagel and S. M. Berget (1996) An intron enhancer recognized by splicing factors activates polyadenylation. *Genes and Development* **10** (2) 208-219.

Lynch KW and Maniatis T, 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer *Genes and Dev* **9**(3) 284-293

Lynch, K. W. and T. Maniatis (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes and Development* **10** (16) 2089-2101.

Matsumura F., and Yamashiro M.S. (1985) Purification and characterisation of multiple isoforms of tropomyosin from rat cultured cells. *J Biol Chem* **260** 13851-13859.

Mattox, W., M. J. Palmer and B. S. Baker (1990) Alternative splicing of the sex determination gene transformer-2 is sex-specific in the germ line but not in the soma. *Genes and Development* **4** (5) 789-805.

Mayeda, A., A. M. Zahler, A. R. Krainer and M. B. Roth (1992) Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America* **89** (4) 1301-1304.

Mayeda, A. and A. R. Krainer (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68** (2) 365-375.

Mayeda, A., D. M. Helfman and A. R. Krainer (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) 2993-3001.

Mayeda, A., S. H. Munroe, J. F. Caceres and A. R. Krainer (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *Embo Journal* **13** (22) 5483-5495.

McKeown M., Belote J.M., Boggs R.T. (1988) Ectopic expression of the female transformer gene product leads to femlae differentiation of chromosomally male Drosophila. *Cell* **53** 887-895.

Mermoud, J. E., P. Cohen and A. I. Lamond (1992) Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Research* **20** (20) 5263-5269.

Mermoud, J. E., P. Cohen and A. I. Lamond (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo Journal* **13** (23) 5679-5688.

Michaud, S. and R. Reed (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes and Development* **5** 2534-2546.

Michaud, S. and R. Reed (1993) A functional association between the 5' and 3' splice sites is established in the earliest prespliceosome complex (E) in mammals. *Genes and Development* **7** (6) 1008-1020.

Min, H., R. C. Chan and D. L. Black (1995) The generally expressed hnRNP F is involved in a neural-specific pre- mRNA splicing event. *Genes and Development* **9** (21) 2659-2671.

Min H., Turck C.W., Nikolic J.M., and Black D.L. (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes and Development* **11** 1023-1036.

Misra S. and Rio D.C. (1990) Cytotype control of Drosophila P element transposition. *Cell* **62** 269-284.

Mount, S.M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Research* **10** 459-472.

Mount, S. M. (1983) RNA processing. Sequences that signal where to splice. *Nature* **304** (5924) 309-310.

Mullen, M. P., C. Smith, J. G. Patton and B. NadalGinard (1991) Alpha-Tropomyosin mutually exclusive exon selection: Competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes and Development* **5** (11) 642-655.

Mulligan, G. J., W. Guo, S. Wormsley and D. M. Helfman (1992) Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. *Journal of Biological Chemistry* **267** (35) 25480-25487.

Munroe, S. H. and X. Dong (1992) Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA-RNA annealing. *Proceedings of the National Academy of Sciences of the United States of America* **89** (3) 895-899.

Nagoshi R.N., McKeown M., Burtis K.C., Belote J.M., Baker B.S. (1988) The control of alternative splicing at genes regulating sexual differentiation on Drosophila Melanogaster *Cell* **53** 229-236.

Nagoshi R.N., and Baker B.S. (1990) Regulation of sex-specific RNA splicing of the at the Drosophila dsx gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes and Development* **4** 89-97.

Nasim, F. H., P. a. Spears, H. M. Hoffmann, H. Kuo and P. J. Grabowski (1990) A sequential splicing mechanism promotes selection of an optional exon by repositioning a downstream 5' splice site in preprotachykinin pre-mRNA. *Genes and Development* **4** (7) 1172-1184.

Nelson K.K. and Green M.R. (1988). *Genes and Development* **2** 319-329.

Niwa, M., C. C. MacDonald and S. M. Berget (1992) Are vertebrate exons scanned during splice-site selection? *Nature* **360** (6401) 277-280.

Nothiger R.M., Leuthold N., Andersen, Gerschwiler, Gruter, Keller, Leist, Roost, Schmid (1987) Genetic and developmental analysis of the sex-determining gene dsx of drosophila melanogaster. *Genet. Research* **50** 113-123.

Oliver, B., Y. J. Kim and B. S. Baker (1993) Sex-lethal, master and slave: A hierarchy of germ-line sex determination in Drosophila. *Development* **119** (3) 897-908.

Perez I., Lin C., McAfee J., and Patton J.G. (1997) Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA* **3** 764-778.

Pittenger, M. F., and D. M. Helfman (1994) In vitro and in vivo characterisation of four fibroblast tropomyosins produced in bacteria. *J Cell Biol* **118** 841-858.

Pittenger, M. F., A. Kazzaz, J.A., and D. M. Helfman (1994) Functional properties of non-muscle tropomyosin isoforms. *Current Opinion in Cell Biology* **6** 96-104.

Pittenger, M. F., A. Kistler and D. M. Helfman (1995) Alternatively spliced exons of the beta tropomyosin gene exhibit different affinities for F-actin and effects with nonmuscle caldesmon. *Journal of Cell Science* **108** (10) 3253-3265.

Pontius B.W. and Berg P. (1990) Renaturation of complementary DNA strands mediated by purified mammalian hnRNP A1: implications for rapid molecular assembly. *Proc Natl Acad Sci* **87** 8403-8407.

Proudfoot, N. (1988). *Trends Biochem Sciences* **14** 105-110.

Proudfoot, N. (1991) Poly(A) signals. *Cell* **64** 671-674.

Proudfoot, N. (1996) Ending the message is not so simple. *Cell* **87** 779-781.

Query, C. C., R. C. Bentley and J. D. Keene (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57** (1) 89-101.

Ramchatesingh, J., Zahler, A.M., Neugebauer, K.M., Roth, M.B., Cooper, T.A. (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) 4898-4907.

Reed, R. (1990) Protein composition of mammalian spliceosomes assembled in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **87** (10) 8031-8035.

Rio, D. C. (1990) Molecular mechanisms regulating Drosophila P element transposition. *Annual Review of Genetics* **24** 543-578.

Rio, D. C. (1991) Regulation of Drosophila P element transposition. *Trends in Genetics* **7** (9) 282-287.

Robberson, B. L., G. J. Cote and S. M. Berget (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Molecular and Cellular Biology* **10** (1) 84-94.

Robertson H.M. and Engels W.R. (1989) Modified P elements that that mimic the P cytotype in Drosophila melanogaster. *Genetics* **123** 815-824.

Rossi F. et al (1996) *Nature* **381** 80-83.

Roth, M. B., A. M. Zahler and J. A. Stolk (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *Journal of Cell Biology* **115** (3) 587-596.

RuizOpazo, N. and B. NadalGinard (1987) alpha-Tropomyosin gene organization. Alternative splicing of duplicated isotype-specific exons accounts for the production of smooth and striated muscle isoforms. *Journal of Biological Chemistry* **262** (10) 4755-4765.

Ruskin, B., P. D. Zamore and M. R. Green (1988) A factor U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52** (2) 207-219.

Ryner, L. C. and B. S. Baker (1991) Regulation of doublesex pre-mRNA processing occurs by 3'-splice site activation. *Genes and Development* **5** (11) 2071-2085.

Sakashita, E. and H. Sakamoto (1994) Characterization of RNA binding specificity of the Drosophila sex- lethal protein by in vitro ligand selection. *Nucleic Acids Research* **22** (20) 4082-4086.

Sambrook J., Fritsch E.F., and Maniatis T. (1989) Molecular cloning: a laboratory manual, 2nd Edition. *Cold Spring Harbour Laboratory Press*.

Samuels, M. E., D. Bopp, R. A. Colvin, R. F. Roscigno, M. A. GarciaBlanco and P. Schedl (1994) RNA binding by Sxl proteins in vitro and in vivo. *Molecular and Cellular Biology* **14** (7) 4975-4990.

Screaton G.R., Caceres, J.F., Mayeda, A., Bell, M.V., Plebanski, M., Jackson, D.G., Bell, J.I., Krainer, A.R. (1995) Identification and characterisation of three members of the human SR family of pre-mRNA splicing factors. *EMBO Journal* **14** (17) 4336-4349.

Sharp, P. A. (1994) *Split genes and RNA splicing. *Cell* **77** (6) 805-815.

Siebel, C. W. and D. C. Rio (1990) Regulated splicing of the Drosophila P transposable element third intron in vitro: Somatic repression. *Science* **248** (4960) 1200-1208.

Siebel, C. W., L. D. Fresco and D. C. Rio (1992) The mechanism of somatic inhibition of Drosophila P-element pre-mRNA splicing: Multiprotein complexes at an exon pseudo-5' splice site control U1 snRNP binding. *Genes and Development* **6** (8) 1386-1401.

Siebel, C. W., R. Kanaar and D. C. Rio (1994) Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes and Development* **8** (14) 1713-1725.

Siebel, C. W., Admon A., and D. C. Rio (1995) Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes and Development* **9** 269-283.

Singh R., Valcarcel J., and Green M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract binding proteins. *Science* **268** 1173-1176.

Smith, C. and B. NadalGinard (1989) Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell* **56** (5) 749-758.

Smith, C., J. G. Patton and B. NadalGinard (1989) Alternative splicing in the control of gene expression. *Annual Review of Genetics* **23** 527-577.

Sosnowski, B. A., J. M. Belote and M. McKeown (1989) Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell* **58** (3) 449-459.

Staffa, A. and Cochrane, A. (1995) Identification of positive and negative splicing regulatory elements within the terminal tat-rev exon of Human Immunodeficiency Virus type 1. *Molecular and Cellular Biology* **15** (8) 4597-4605.

Staknis, D. and R. Reed (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) 7670-7682.

Stephen D., Jones C., and Schofield J.P. (1990) A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acid Res* **18** (24) 7463-7464.

Sun, Q., R. K. Hampson and F. M. Rottman (1993) In vitro analysis of bovine growth hormone pre-mRNA alternative splicing. Involvement of exon sequences and trans-acting factor(s). *Journal of Biological Chemistry* **268** (21) 15659-15666.

Tacke, R., and Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO Journal* **14** (14) 3540-3551.

Tacke, R., Chen Y., and Manley, J.L. (1997) Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. *Proceedings of the National Academy of Sciences of the United States of America* **94** 1148-1153.

Talerico, M. and S. M. Berget (1990) Effect of 5' splice site mutations on splicing of the preceding intron. *Molecular and Cellular Biology* **10** (12) 6299-6305.

Tanaka K., Watakabe A., and Shimura Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. *Molecular and Cellular Biology* **14** (2) 1347-1354.

Tarn, W. Y. and J. A. Steitz (1994) SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes and Development* **8** (22) 2704-2717.

Tarn, W. Y. and J. A. Steitz (1996) A novel slpiceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84** (5) 801-811.

Temm-Grove C.T., Guo W., and Helfman D.M. (1996) Low molecular weight rat fibroblast TM-5: cDNA cloning, actin binding, localisation, and coiled-coil interactions. *Cell motility and the cytoskeleton* **33** 223-240.

Tian, M. and T. Maniatis (1992) Positive control of pre-mRNA splicing in vitro. *Science* **256** (5054) 237-240.

Tian, M. and T. Maniatis (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74** (1) 105-114.

Tian, M. and T. Maniatis (1994) A splicing enhancer exhibits both constitutive and regulated activities. *Genes and Development* **8** (14) 1703-1712.

Traut T.W. (1988) *Proc Natl Acad Sci* **85** 2944-2948.

Valcarcel, J., R. Singh, P. D. Zamore and M. R. Green (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362** (6416) 171-175.

Walther W., Stein U., and Uckert W. (1993). Rapid method of total RNA mini-preparation from eucaryotic cells*Nucleic Acids Res* **21** (7) 1682.

Wang, J., and Bell L. (1994) The sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes and Development* **8** 2072-2085.

Wang, Z., Hoffman, H.M., Grabowski, P.J., (1995) Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA* **1** 21-35.

Wang, J., and Manley, J.L., (1995) Overexpression of the SR proteins ASF/SF2 and SC35 influences alternative splicing in vivo in diverse ways. *RNA* **1** 335-346.

Wang J., Takagaki Y., and Manley J.L. (1996) Targetted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes and Development* **10** 2588-2599.

Watakabe, A., K. Tanaka and Y. Shimura (1993) The role of exon sequences in splice site selection. *Genes and Development* **7** (3) 407-418.

Wise, J.A. (1993). Guides to the heart of the spliceosome. *Science* **262** 1978-1979.

Woppmann A., et al (1993). *Nucleic Acids Res* **21** 2815-2821.

Wu J.Y. and Maniatis T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75** 1061-1070.

Xiao S., and 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 and Development* **11** 334-344.

Xu, R., J. Teng and T. A. Cooper (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Molecular and Cellular Biology* **13** (6) 3660-3674.

Yamashiro S., Yamakita Y., Ishikawa R., and Matsumura F. (1990) Mitosis specific phosphorylation causes 83K nonmuscle caldesmon to dissociate from microfilaments. *Nature* **344** 675-678.

Yeakley, J. M., F. Hedjran, J. P. Morfin, N. Merillat, M. G. Rosenfeld and R. B. Emeson (1993) Control of calcitonin/calcitonin gene-related peptide pre-mRNA processing by constitutive intron and exon elements. *Molecular and Cellular Biology* **13** (10) 5999-6011.

Yeakley, J. M., J. P. Morfin, M. G. Rosenfeld and X. D. Fu (1996) A complex of nuclear proteins mediates SR protein binding to a purine- rich splicing enhancer. *Proceedings of the National Academy of Sciences of the United States of America* **93** (15) 7582-7587.

Zahler, A. M., W. S. Lane, J. A. Stolk and M. B. Roth (1992) SR proteins: A conserved family of pre-mRNA splicing factors. *Genes and Development* **6** (5) 837-847.

Zahler, A. M., K. M. Neugebauer, W. S. Lane and M. B. Roth (1993) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260** (5105) 219-222.

Zahler, A. M. and M. B. Roth (1995) Distinct functions of SR proteins in recruitment of U1 snRNP to alternative 5' splice sites. *Proc Natl Acad Sci* **92** 2642-2646.

Zamore, P. D. and M. R. Green (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proceedings of the National Academy of Sciences of the United States of America* **86** (23) 9243-9247.

Zamore, P. D. and M. R. Green (1991) Biochemical characterization of U2 snRNP auxiliary factor: An essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo Journal* **10** (1) 207-214.

Zamore, P. D., J. G. Patton and M. R. Green (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355** (6361) 609-614.

Zhang, M., P. D. Zamore, M. CarmoFonseca, A. I. Lamond and M. R. Green (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proceedings of the National Academy of Sciences of the United States of America* **89** (18) 8769-8773.

Zhang, W. J. and J. Y. Wu (1996) Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. *Molecular and Cellular Biology* **16** (10) 5400-5408.

Zuo, P. and J. L. Manley (1993) Functional domains of the human splicing factor ASF/SF2. *Embo Journal* **12** (12) 4727-4737.

Zuo, P. and J. L. Manley (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proceedings of the National Academy of Sciences of the United States of America* **91** (8) 3363-3367.

Zuo, P. and T. Maniatis (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes and Development* **10** (11) 1356-1368.