

Alternative Pre-Messenger RNA Splicing
of Murine N-CAM and Human Tropomyosin
in Non-Muscle and Muscle cells.

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

by

Marion Hamshere BSc (Nottingham)
Department of Biochemistry
University of Leicester

June 1993

UMI Number: U541872

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U541872

Published by ProQuest LLC 2015. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



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



7501799351

To S. Edith Barber

Acknowledgements:

I would like to thank Nigel Hamshire for his encouragement and love, my son Thomas who could bring a smile to my face even after the most disastrous of days and to Sian Daire for looking after Thomas whilst I finished this thesis. I would also like to thank my supervisor, Ian Eperon, for his help and support in the laboratory and during preparation of this manuscript, and other members of the laboratory, particularly Deborah Ireland for her friendship and encouragement, Ian Graham for the gifts of α -tropomyosin mini-gene constructs, Andy Smith for sequence analysis of the mini-gene constructs, and John Keyte and Debra Langford for the synthesis of oligonucleotide primers. I would also like to thank Frank Walsh and George Dickson for their gifts of murine RNA samples. This work was funded by the Medical Research Council.

ABSTRACT

A quantitative RT-PCR method for analysis of alternative isoforms of RNA, and a method for the transient expression of mini-genes in differentiated muscle cells have been developed. This has enabled the analysis of endogenous RNA transcribed from rodent N-CAM and from human α_s -tropomyosin mini-genes which were expressed in COS cells and mouse C-2 myoblasts and myotubes.

The sequences of the previously unreported mouse homologues of human exons MSD1b and MSD1c of N-CAM have been determined, and deposited in the EMBL data base. The tissue- and stage-specific alternative splicing patterns of exons within the muscle-specific domain (MSD) of N-CAM have also been established; the exons were normally incorporated as a unit in muscle cells, but were not included in transcripts derived from non-muscle myoblasts and neural cells. The triplet AAG exon was also included in a stage- and tissue-specific manner, but independently of inclusion of other exons of the MSD.

Transfection of C-2 myoblasts with mutant mini-gene constructs of human α_s -tropomyosin determined the *cis*-acting elements which regulate the mutually exclusive alternative splicing of the central exons (NM and SK) in both non-muscle and muscle cells. In non-muscle, these were found to be due either to *cis*-acting repressor sequences within the SK exon or *cis*-acting activator sequences within the NM exon. In differentiated cells, exclusion of the NM exon is not via *cis*-acting repressor sequences within the NM exon, but because the upstream (NM) exon site is dormant and is therefore skipped by the splicing machinery.

The evolution of alternative pre-mRNA splicing has also been discussed, and on this basis and from analysis of the data presented here, I conclude that regulation of alternative pre-mRNA splicing of transcripts from different genes may be founded upon a common mechanism which is largely dependent upon the presence of sub-optimal splice-signals and the potential for variation in the relative concentrations of certain splicing factors.

Chapter 1: Introduction.....pages 1-47

1.1 Evolution of split genes

- 1.1.1 Evolution of introns.
- 1.1.2 Exon shuffling
- 1.1.3 Gene duplication and gene conversion
- 1.1.4 Use of alternative exons
- 1.1.5 Evolution of the tropomyosin gene family.

1.2 Mechanisms of pre-mRNA splicing

- 1.2.1 Constitutive splicing
- 1.2.2 Alternative splicing

1.3 Control of alternative splicing

- 1.3.1 Cis-acting elements
- 1.3.2 Trans-acting factors
- 1.3.3 The tropomyosin gene family
- 1.3.4 Drosophila genes

1.4 Evolution of alternative splicing.

- 1.4.1 Concerted evolution of trans-acting factors
- 1.4.2 Constraints on exon duplication
- 1.4.3 A role for diploidy and sex

1.5 Muscle as a model system

- 1.5.1 Muscle as an in-vivo test system
- 1.5.2 N-CAM: a model for cassette-type alternative splicing
- 1.5.3 Human α -tropomyosin: a model for mutually-exclusive alternative splicing

1.6 Identification and quantification of RNA

- 1.6.1 S1 mapping
- 1.6.2 Primer extension
- 1.6.3 Northern and north-western blotting
- 1.6.4 cDNA libraries
- 1.6.5 Reverse-transcriptase/PCR: cloning and sequencing
- 1.6.6 Reverse-transcriptase/PCR: direct analysis

Chapter 2: Materials and methods.....pages 48-68

- 2.1 Stock solutions
- 2.2 Cell culture
 - 2.2.1 Stock solutions
 - 2.2.2 Preparation of collagenised flasks
 - 2.2.3 Growth conditions for myoblasts
 - 2.2.4 Induction of myogenesis
 - 2.2.5 Transfection of cells
- 2.3 RT-PCR
 - 2.3.1 Isolation of RNA from cells
 - 2.3.2 Poly(A) selection of mRNA
 - 2.3.3 RT-PCR of endogenous RNA
 - 2.3.4 RT-PCR of exogeneous RNA
 - 2.3.5 Cloning PCR products
 - 2.3.6 Direct analysis of PCR products
 - 2.3.7 AluI digestion of PCR products
- 2.4 Sequence analysis
 - 2.4.1 Use of m13 templates
 - 2.4.2 PCR sequencing
 - 2.4.3 Double-stranded sequencing
- 2.5 Ethanol precipitation
- 2.6 Polyacrilamide gel electrophoresis
- 2.7 Preparation of plasmid DNA
- 2.8 Isolation of genomic DNA
- 2.9 PCR amplification of genomic DNA
- 2.10 Southern blotting and hybridisation
- 2.11 End-labelling of oligo-nucleotides
- 2.12 Oligonucleotide sequences
- 2.13 Touchdown PCR programs
- 2.14 Standard "second-round" PCR reaction
- 2.15 Human α_s -tropomyosin mini-gene constructs.

Chapter 3: Alternative splicing in mouse N-CAM.....pages 69-100

- 3.1 Cloning of a mouse homologue to the human MSD region
 - 3.1.1 Development of a combined RT-PCR method
 - 3.1.2 Myoblasts and myotubes as a source of RNA
 - 3.1.3 Results of cloning and sequencing
- 3.2 Direct analysis of RT-PCR products
 - 3.2.1 Poly(A)⁺ selected RNA
 - 3.2.2 Using a nested primer
 - 3.2.3 Quantification of PCR
 - 3.2.4 Identification of the bands
- 3.3 Exon usage in developing muscle
 - 3.3.1 Developmental profile of muscle
 - 3.3.2 Effect of denervating muscle
- 3.4 Exon usage in neural cells
 - 3.4.1 Developmental profile in brain
 - 3.4.2 Exon use in neural cell lines
- 3.5 Sequence similarities between exons
 - 3.5.1 Similar because of common descent?
 - 3.5.2 Could the sequence similarity have arisen by chance?
- 3.6 Genomic PCR
 - 3.6.1 Exon 12 to MSD1a
 - 3.6.2 MSD1a to MSD1b
 - 3.6.3 MSD1b to MSD1c
 - 3.6.4 MSD1c to exon 13
 - 3.6.5 MSD1c to SEC

Chapter 4: Regulation of alternative splicing in α_s -tropomyosin... 101-141

- 4.1 Development of RT-PCR method
 - 4.1.1 Vector specific oligonucleotide
 - 4.1.2 The identification of alternative isoforms
 - 4.1.3 Linearity check of PCR
- 4.2 Optimisation of transient expression in myoblasts and myotubes
 - 4.2.1 Endogenous expression of α -tropomyosin
 - 4.2.2 Transfection of C-2 cells
 - 4.2.3 Effect of vector concentration

- 4.2.4 Transient expression with α -tropomyosin
- 4.3 Regulation of alternative splicing in COS cells
 - 4.3.1. Does splicing of the exons follow an ordered pathway?
 - 4.3.2. Does repression lead to exon skipping if the SK exon is in both sites?
 - 4.3.3 Does uprating the SK branch-point lead to double incorporation?
 - 4.3.4. Does de-repression of the SK site lead to double incorporation?
- 4.4 Regulation of alternative splicing in C-2/C-12 cells by cis-acting elements.
 - 4.4.1 Is the cis-acting sequence within exons?
 - 4.4.2 What prevents NM exon use in myotubes?
 - 4.4.3 Is the cis-acting sequence within the introns?
 - 4.4.4 Identification of elements within SK exon
- 4.5 A role for competition in alternative splicing in C-2/C-12 cells
 - 4.5.1 Are the splice-sites sub-optimal?
 - 4.5.2 Do exons compete for a limited splicing component?
 - 4.5.3 Does the potential stem-loop play a role?

Chapter 5: Discussion.....pages 142-179

- 5.1 Development of methods
 - 5.1.1 Development of RT-PCR methods.
 - 5.1.2 Development of a system for analysis of sequences determining alternative splicing in muscle cells.
- 5.2 Analysis of splicing in the MSD region of mouse N-CAM.
 - 5.2.1 Discrete and coordinated regulation of the MSD: in muscle cells
 - 5.2.2 Effect of denervating muscle
 - 5.2.3 Discrete and coordinated regulation of the MSD: in neural cells
 - 5.2.4 Identification of putative cis-acting elements
 - 5.2.5 Possible role for MSD exons
 - 5.2.6 Genomic structure of the MSD region of N-CAM

- 5.3 Alternative splicing in human α_s -tropomyosin
 - 5.3.1 Identification of signals which regulate expression in non-muscle (COS cells and myoblasts)
 - 5.3.2 Identification of signals which regulate expression in muscle (myotubes)
 - 5.3.3 Proposed mechanism which regulates the switch in α_s -tropomyosin isoform use.
- 5.4 Relevance of results to mechanisms involved in constitutive splicing
- 5.5 Evolutionary considerations for the analysis of alternative splicing
 - 5.5.1 Evolution of the tropomyosin family
 - 5.5.2 Evolution of the NM and SK exons
 - 5.5.3 Linkage and co-evolution of cis- and trans-acting elements
 - 5.5.4 Evidence for common regulatory mechanisms.
- 5.6 Conclusions

Appendix 1: Sequences of mini-gene constructs.....pages 180-201

Appendix 2: Absorbance/Emission datapages 202-203

References.....pages 204-230

Table of Figures

- 1.1 Evolution of tropomyosin genes
- 1.2 The mechanism of constitutive splicing
- 1.3.4 Sex determination in *Drosophila*
- 2.1 Cartoons of mini-gene constructs
- 3.1.1a Diagram of RT-PCR method
- 3.1.1b Optimum annealing temp for PCR
- 3.1.3a Myoblast isoform "T" track sequences
- 3.1.3b Myotube isoform "T" track sequences
- 3.1.3c Myoblast isoform sequences
- 3.1.3d Myotube isoform sequences
- 3.1.3e Diagram of isoform use
- 3.1.3f Sequence of MSD in mouse, human and chicken
- 3.2.3 Quantitative PCR
- 3.2.4 PCR sequencing to verify bands
- 3.3.1a Developmental profile in muscle
- 3.3.1b Inclusion/exclusion of AAG
- 3.3.1c Histograms of % isoform use
- 3.4.1 Developmental profile in brain
- 3.6.1 Genomic PCR 12-a
- 3.6.2a Genomic PCR a-b; c-sec
- 3.6.2b Genomic PCR a-b; c-sec (Southern blot)
- 3.6.3 Genomic PCR b-c (agarose gel and southern blot)
- 3.6.4 Genomic PCR c-13 (agarose gel and southern blot)
- 3.6.5 Proposed genomic organisation of MSD
- 4.1.1 Vector specific oligonucleotide
- 4.1.2 Restriction sites in cDNAs
- 4.1.3 Linearity of PCR amplifications
- 4.2.1 Endogenous mouse α -tropomyosin
- 4.2.3 Effect of vector concentration
- 4.2.4 Developmental profile of transfected α -tropomyosin
- 4.3 Transfections into COS cells
- 4.4 Transfections into myoblasts and myotubes
- 4.5 Transfections into myoblasts and myotubes

Chapter 1

Introduction

1.1 Evolution of split genes

Introns, a term coined to describe tracts of non-coding DNA sequences within genes (Gilbert 1978), were first discovered in 1977 in the genes of yeast tRNA (Goodman et al 1977), and have since been identified in the genomes of eukaryotes, eukaryotic viruses, mt DNA, chloroplast DNA, and in archaeobacteria, but not in the genomes of prokaryotes. Their discovery, only 16 years ago has produced an area in molecular biology which has become the centre for intense research.

1.1.1 Evolution of introns

Two main theories for the evolution of introns have been proposed (reviewed by Lambowitz 1989; Seidel 1992). The "intron gain" theory proposes that the original coding regions for all organisms were infiltrated by introns, which in their ancestral form had the capacity to self-splice from the RNA encoded by the interrupted gene. The "intron loss" theory proposes that all coding regions were separated by regions of non-coding nucleic acid (possibly RNA, if RNA was the original genetic template (Gilbert 1986)).

The "intron gain" theory requires that ancestral introns acted as transposable elements, and that the different classes of introns arose by their evolution from such transposable ancestors (Gilbert 1986; Cavalier-Smith 1985; Cavalier-Smith 1991). The group I introns are proposed to most closely resemble the ancestral transposable element as they are capable of self-splicing (see Cech, 1990, for a review of the mechanisms of splicing in group I introns), and because at least one (omega intron in Mt 21S rRNA of *S.cerevisiae*) still retains the capacity for transposition (Dujon 1980). This intron encodes a sequence-specific endonuclease, a site for which is within the Mt 21S rRNA gene (Colleaux et al 1988). The potential for similar intron mobility has now been demonstrated for introns of T-even phages (Quirk et al 1989), primitive

eukaryote (Muscarella et al 1989) and other yeast mitochondrial introns (Delahodde et al; Wenzlau et al 1989). The horizontal transfer of introns across species barriers, mediated by "intron homing", has also been proposed recently (Johansen et al 1992).

A second route through which introns could be integrated into the genome is via their incorporation into RNA in a reverse-splicing type reaction, and thence into the genome via reverse transcription and site-specific integration or recombination. Original experiments to ascertain the feasibility of such a reaction required that the reaction conditions were significantly non-physiological (high temperature and/or high magnesium concentration) (Woodson and Cech 1989). However, recent experiments (Mohr and Lambowitz 1991) indicate that these requirements can be circumvented when tyrosyl-tRNA synthetase (a protein required for the splicing of Mt introns in *Neurospora* (Akins et al 1987)) is included in the reaction.

The "intron gain" theory predicts that group II introns evolved from group I introns, and pre-mRNA introns from group II introns (Hickey 1989). Flaws in this evolutionary progression are first, the lack of "missing links", although the mechanism for *trans*-splicing of spliced leader RNA (SL RNA) has been proposed as an example of the link between group II and pre-mRNA introns (Bruzik et al 1988), and second, the analysis of the sequences of the structurally similar U2/U6 snRNP complex and domain 5 of group II introns (Mahdani and Guthrie 1992). Although these snRNPs may be the analogues of the 5th domain of the group II introns, the sequences do not appear to have been conserved. It has been proposed, therefore, that the similarity is a consequence of convergent evolution, and not because of divergence from a common ancestor (Weiner 1993). Very little work has concentrated on the evolution of the pre-mRNA introns *per se* to support an "intron gain"

model. However, the work of Dibb and Newman on comparisons of the intron positions of the tubulin and actin gene-families (Dibb and Newman 1989) does support an intron gain rather than loss mode of evolution, and is based upon the integration of introns at specific sites dictated by the presence of "proto-splice sites" which have the consensus C/AAGR (Dibb 1991).

The "intron loss" theory proposes that the original coding sequence arose from the random association of nucleic acid into purely random (probably RNA) sequence. The presence of stop codons (and presumably initiation codons) dictated the length of the open reading frames (ORF). Selective pressure drove the ORFs toward linkage disequilibrium for peptides that were active in a concerted manner. A mechanism that co-encoded the different peptides on a single transcript was then favoured. Pre-mRNA splicing emerged as such a mechanism. Proponents of an intron loss theory cite the finding that exons normally encode functionally distinct domains (Blake et al 1978; Go et al 1981) as evidence in favour of such a mechanism. Thus, the possibility that exons arose from discrete functional units has been investigated in two ways. First, under this mechanism it should be possible to predict the number of introns within a gene by reference to the protein structure. This was undertaken for the ancient triosephosphate isomerase (TPI) gene (Gilbert et al 1986) in which a gene arrangement containing 11 exons and 10 introns was predicted. However, until recently only 9 of the introns had been identified in extant genes. The discovery of the "missing" 10th intron in the mosquito (Tittinger et al 1993) appeared to provide evidence in support of the intron loss theory (Doolittle and Stoltzfus 1993) although it could not exclude a mechanism based on intron gain (Doolittle and Stoltzfus 1993). Second, the theoretical distribution of stop codons in simulated random sequence RNA was compared to the observed distribution in sequenced genes (Senapathy 1986; Senapathy 1988). The similarity is

striking for nuclear genes: the predicted maximum distance between randomly produced stop codons (at about 600 nucleotides) matches the upper limit for observed exon length, and the frequencies of exon length, which exhibit a negative exponential distribution, resembled the frequencies of the lengths in the simulated random sequence (Senapathy 1986). The "intron loss" theory was recently rekindled (Seidel et al 1992a) because of sequence analysis of a gene encoding a mutase protein from *Tetrahymena*. In this gene there are two in-frame amber codons (TAG), which in most organisms is one of the three "stop" codons. *Tetrahymena* reads this as a glutamine codon. Seidel cites this as evidence for the existence of ancient "stop" codons in the original intron/exon borders (Seidel 1992b), which in other organisms have been selected against once the splicing machinery is in place, but which have been retained in *Tetrahymena* because of the evolution of suppressor amber mutations.

The question "intron gain or intron loss?" still remains to be answered. However, the arguments for "intron gain" appear to me to be stronger, particularly for the group I and group II introns, although evidence of an evolutionary progression from group II to pre-mRNA introns (a pre-requisite for the intron gain theory) remains elusive. However, the problems which have been encountered in predicting the evolutionary origin of introns probably indicates that they are ancient, and as such have almost certainly played an important role in evolution of the extant eukaryote genome.

1.1.2 Exon shuffling

The evolution of exons can be attributed to three main types of event: exon shuffling, duplication and gene conversion. The possibility that genes evolved by the shuffling and re-assortment of exons was first proposed shortly after the finding of non-coding regions within genes of higher

eukaryotes (Gilbert 1978). This exon shuffling was proposed to be due to mutational hot-spots within the introns leading to recombination events (Gilbert 1978; Hall et al 1989). It is interesting to postulate that the highly pyrimidine rich region associated with the 3'splice-site is such a sequence. Evidence for this is circumstantial, but secondary structures can be taken up by regions of DNA, dependent on their sequence composition, which are not normal duplex DNA (Wells 1988). Pyrimidine rich regions have also been shown to form into triplet DNA which is sensitive to S1 nuclease and exhibits aberrant migration on gels (Htun and Dahlberg 1989; Bernues et al 1989; Casasnovas et al 1989). The possibility that such sequences are mutational hot-spots has been demonstrated by the identification of an unequal sister chromatid exchange site (USCE) in a mouse myeloma cell line which is a $(TC)_n$ rich region (Katzenberg et al 1989), and by finding of the $(CT)_3CG(CT)_9$ sequence at the recombination point for the polymorphic duplication of the GDPH locus in *D.melanogaster* populations (Kusakabe et al 1990). Further evidence for exon shuffling is provided by analysis of sequence similarity in exons from different genes. By this method the number of putative original exon sequences was calculated to be in the order of 1000 to 7000 (Dorit et al 1990). These ancestral exons would have evolved into extant genes by duplication and shuffling. Recently, more evidence for exon shuffling has been provided by analysing the position of introns in relation to the reading frame of their boundary exons. It is proposed that the non-random distribution found could not be attributed solely on the basis of mobile introns, and therefore is the consequence of exon shuffling (Fedorov et al 1992).

1.1.3 Gene duplication and gene conversion

The duplication of a gene arises either because it was selected for as an advantageous adaptation, or because it was fixed in a population because of

random genetic drift.

There are examples, in prokaryotes, in which the duplication of a gene is advantageous, however, in eukaryotes the advantage of duplication *per se* is more difficult to assess. The use of computer simulation of the evolution of the multi-gene families predicts that they could have arisen as a consequence of random genetic drift, if random gene duplication and deletion can occur and if the size of the genome is not critical to the survival of the organism (Loomis and Gilpin 1986).

Whether the fixation of a duplication event is the consequence of selection or of random genetic drift, it is clear that gene duplication has probably played a major role in the formation of the eukaryote genome.

One important effect of gene duplication is the potential for non-reciprocal recombination or gene-conversion. This has certainly played an important role in the evolution of multi-gene families (Liskay et al 1987; Wysocki et al 1989; Hickey et al 1991). In the calcium dependent cell-adhesion molecules K-CAM and L-CAM, which have arisen as a duplication (Sorkin et al 1991), gene conversion is proposed as the mechanism which caused the 3' terminal exons to be almost identical whilst the 5'exons have diverged (Gally and Edelman 1992). This may be an important mechanism for the conservation of specific sequences within a gene in families whose members have both shared and individual functions.

1.1.4 Use of alternative exons

The potential for isoform diversity is realised when alternative exons are available for incorporation into mature transcripts. These exons impart an alteration in the primary sequence of the mRNA, and hence often in expression of the protein product by a variety of mechanisms including: presence of alternative promoters (e.g. myosin; tropomyosin); provision of

alternative polyadenylation sites (e.g. exon 15 of N-CAM (Walsh and Dickson 1989); IgM (Paterson and Perry 1989); CGRP (Leff et al 1987) and the incorporation of stop codons (e.g. sec exon in N-CAM (Walsh and Dickson 1989). The function of the protein can also be moderated: the incorporation of additional sequence(s) (e.g. leucocyte common antigen (LCA/CD45) (Streuli and Saito 1989); N-CAM (Dickson et al 1987); or the use of mutually exclusive alternative exons (e.g. tropomyosin (Helfman et al 1988), will perturb the primary protein sequence which may cause an alteration in its physiological role.

1.1.5 Evolution of the tropomyosin gene family

Tropomyosins are a heterogeneous family of actin binding proteins, members of which have been found in all higher eukaryote cell types characterised to date (see Helfman et al 1988 for review).

Two genes have been identified and characterised which encode the tropomyosins in *Drosophila melanogaster*. The TmII gene is a complex gene with 17 exons, many of which are alternatively spliced in different tissues, which also contains a tissue specific internal promoter (Hanke and Storti 1988). The second gene, TmI (Basi and Storti 1986), is closely linked to the TmII gene and is thought to have evolved from a duplication of an ancestral Tm gene. The shorter and less complex TmI gene is thought to have arisen by the recombination of a cDNA, processed from mRNA, into one of the duplicated genes (Hanke and Storti 1988).

The chicken tropomyosins are encoded by two genes: β -tropomyosin, for which amino acid (Sanders and Smillie 1985), mRNA (Bradac et al 1989) and genomic sequences (Libri et al 1989a, Libri et al 1989b, Forry-Schaudies et al 1990) have been identified, and α -tropomyosin (Takenaga et al 1988, Pearson-White et al 1987). Rat tropomyosins are encoded on at least three genes: β -tropomyosin (Helfman 1986), TM-4 (Yamausaki and Helfman 1987, Lees-

Miller et al 1990a), and α -tropomyosin (Ruiz-Opazo et al 1985; Ruiz-Opazo and Nadal-Ginard 1987; Wieczorek et al 1988; Lees-Miller et al 1990b; Goodwin et al 1991). However, the recently identified TM-4 gene does not encode alternative exons, although vestigial exon sequences are present in similar positions to those of the alternatively spliced exons of the β -tropomyosin gene (Lees-Miller et al 1990a). The human tropomyosins are encoded on four genes: hTM_a (MacLeod and Gooding 1988); hTM_b (MacLeod et al 1985); hTM_{pl} (MacLeod et al 1987); and hTM_{nm} (MacLeod et al 1986; Clayton et al 1988).

Analysis of exon sequences and exon/intron arrangements of the different tropomyosin genes indicates that the genes arose by gene duplication of an ancestral sequence which contained exons which were alternatively spliced (Wieczorek et al 1988). A maximum parsimony tree can be derived from the exon sequences (Sri-Widada et al 1989; Lees-Miller et al 1990). This predicts the duplication events that led to the formation of four tropomyosin loci, thought to exist in higher eukaryotes (Forry-Schaudies 1990). This tree is shown in Figure 1.1. The absence of a homologue to human hTM_{nm} in the rat may be accounted for by an increased complexity on the rat α -tropomyosin gene. From these data it seems apparent that the duplication of the α -tropomyosin is either a recent event and occurred since the divergence of rodents and man, or the homologue has been inactivated either by deletion, or is present as a pseudogene in the rat and chicken. The original progenitor gene has been proposed to have contained more exons than the descendant tropomyosin genes (Lees-Miller et al 1990a). The complex *Drosophila* Tm11 gene may be the extant descendant of such a gene (Hanke and Storti 1988).

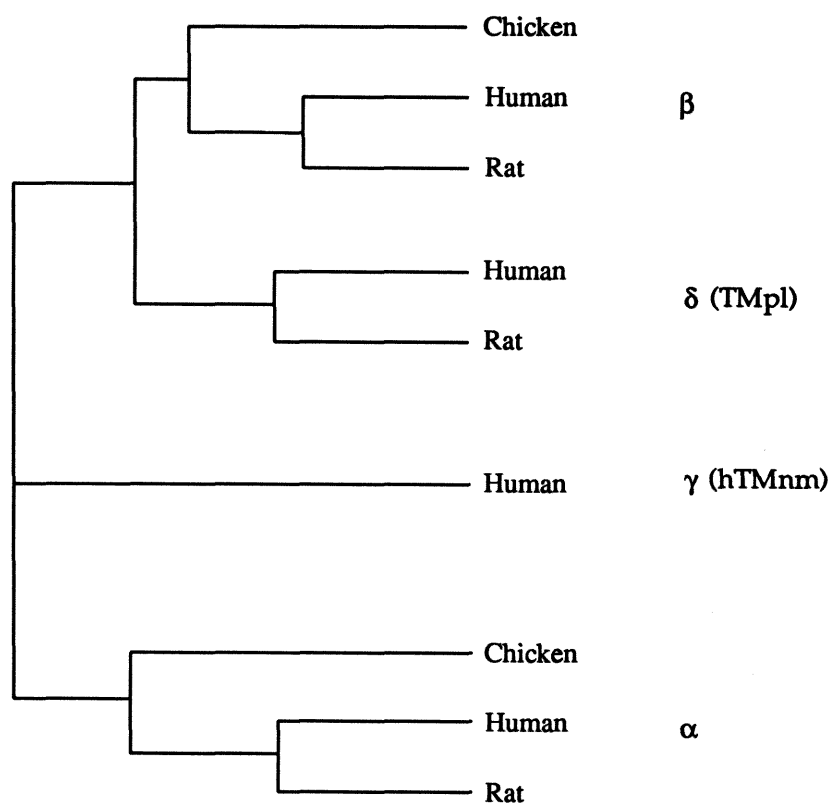


Figure 1.1
 An evolutionary tree, indicating the proposed points for gene duplication, which shows the relationship between the four tropomyosin loci. These encode the α - and β -tropomyosins in human, chicken and the rat.

1.2 Mechanisms of pre-mRNA splicing

The development of *in-vitro* splicing reactions (Hernandez et al 1983; Padgett et al 1983; Krainer et al 1984; Noble 1986) has enabled a dissection of the splicing reaction. This has increased our understanding of both the role played by the *cis*-acting elements encoded by the transcript, and the splicing factors which bind to them.

1.2.1 Constitutive splicing

A catalogue of splice-site sequences which form the junctions between exons and introns was compiled (Mount 1982). The 5' splice-site (CAG/GUAAGU) and 3' splice-site (YAG) consensus sequences were derived from this catalogue.

The role of 5'splice-site sequence has been investigated both *in vitro* and *in vivo*, sometimes with different results. Deletion of the 5'splice-site led to the use of cryptic sites *in vivo* (Wieringa et al 1983; Aeby et al 1986) or the arrest of splicing at stage 1 *in vitro* (Aeby et al 1986), although mutation of the bases within the splice-site resulted in one gene in aberrant positional cleavage of the intron/exon border (Aeby et al 1987). The insertion of 5'splice-sites in an intron (Mayeda and Ohshima 1988) or the duplication of sites within an exon (Lang and Spritz 1983; Wieringa et al 1983; Nelson and Green 1988; Eperon et al 1986; Eperon et al 1988; Cunningham et al 1991), and subsequent use of that splice-site, indicated that the sequence itself contained sufficient signals to bring about cleavage, and splicing at the new site.

The discovery of lariat intermediates (Padgett et al 1984; Grabowski et al 1984; Ruskin et al 1984) indicated the possibility that a conserved sequence, analogous to the yeast branch-point sequence (Keller and Noon 1984), was

also present in pre-mRNA. The mapping of the branch for lariat intermediate formation indicated that a loosely conserved sequence was present at the branch-point (Konarska et al 1985; Reed and Maniatis 1985). Mutation of this sequence resulted in cryptic site use (Ruskin et al 1985; Padgett et al 1985), insertion into a new site, or duplication led to switch of branch-point use (Rautmann and Breathnach 1985; Reed and Maniatis 1988; Wu and Manley 1989). The branch-point has been implicated as the signal which defines the 3'splice-site, by a simple scanning mechanism (Smith et al 1989).

The consistent finding of (Py)_n rich tracts between the branch-point and 3'splice site (Mount et al 1982) indicated that these sequences may also play a role in 3'splice-site selection. Their importance has now been demonstrated (Reed 1989), although they are thought to play a role in branch-point, rather than 3'splice-site selection (Smith et al 1989; Reed 1989) particularly when the branch-point is at a distance from the 3'splice site.

Analysis of the 3'splice-site reveals an absolutely conserved AG motif. The requirement for the AG is absolute (Ruskin and Green 1985a), mutation to AU or UG results in abortive splicing *in-vitro* (Aebi et al 1987; Lamond et al 1987) although with transcripts containing long (Py)_n tracts, the lariat and 5'splice-site cleaved intermediates of splicing can form (Reed 1989). Cryptic site use has been demonstrated *in vivo* (Aebi et al 1986).

The possibility that snRNPs were involved in splicing (Lerner 1980) was indicated by the finding that the free 5'end of the RNA molecule in U1 snRNP was complementary to the consensus 5'splice-site (Lerner et al 1980). Removal of this sequence from U1 abolishes splicing *in vitro* (Kramer 1984; Krainer and Maniatis 1985; Black et al 1985; Berget and Robberson 1986). U2 snRNP has been shown to be associated with the branch-point (Krainer et al

1985a; Black et al 1985; Berget and Robberson 1986; Chabot and Steitz 1987a; Zillmann et al 1987; Kramer et al 1987) even in the absence of a 3'splice-site (Nelson and Green 1989). There is evidence that U5 snRNP interacts with the 3'splice-site (Chabot et al 1985; Gerke and Steitz 1986; Tazi et al 1986), and is associated with the pre-formed U4/U6 complex (Black and Steitz 1986; Berget and Robberson 1986) and with U6 in the spliceosome (Grabowski and Sharp 1986; Konarska and Sharp 1987 Bindereif and Green 1987;). The association of U5 with U4/U6 can also form de-novo under appropriate conditions in the absence of mRNA (Konarska and Sharp 1988).

Recent experiments have now shed light on the interaction between U2 and U6 (Datta and Weiner 1991; Madhani and Guthrie 1992) and the interaction of U5 with both the 3' and 5'splice-sites (Newman and Norman 1992). This has provided an insight into the possible mechanisms involved (Steitz 1992). In this model, the interaction of U1 with the 3' and 5'splice-sites, coupled with the involvement of U5 provides a mechanism for intron excision based on resolution of a Holliday junction (reviewed by Steitz 1992).

Several non-snRNP protein factors have been identified, which include U2AF, a protein required for U2 binding at the branch-point (Ruskin et al 1988; Zamore and Green 1989; Zamore and Green 1991; Zamore et al 1992), SF-2 (also now shown to be the same as ASF), a protein which appears to promote proximal 5'splice-site use in splicing (Krainer et al 1990a; Krainer 1990b; Zuo and Manley 1991; Krainer et al 1991), and pPTB, a polypyrimidine tract binding protein (Garcia-Blanco et al 1989; Gil et al 1991; Patton et al 1991).

The interaction of these snRNP and non-snRNP factors forms the mechanism of constitutive splicing reviewed extensively elsewhere (see Green 1991; Steitz 1992). The main interactions, and the proposed pathway for

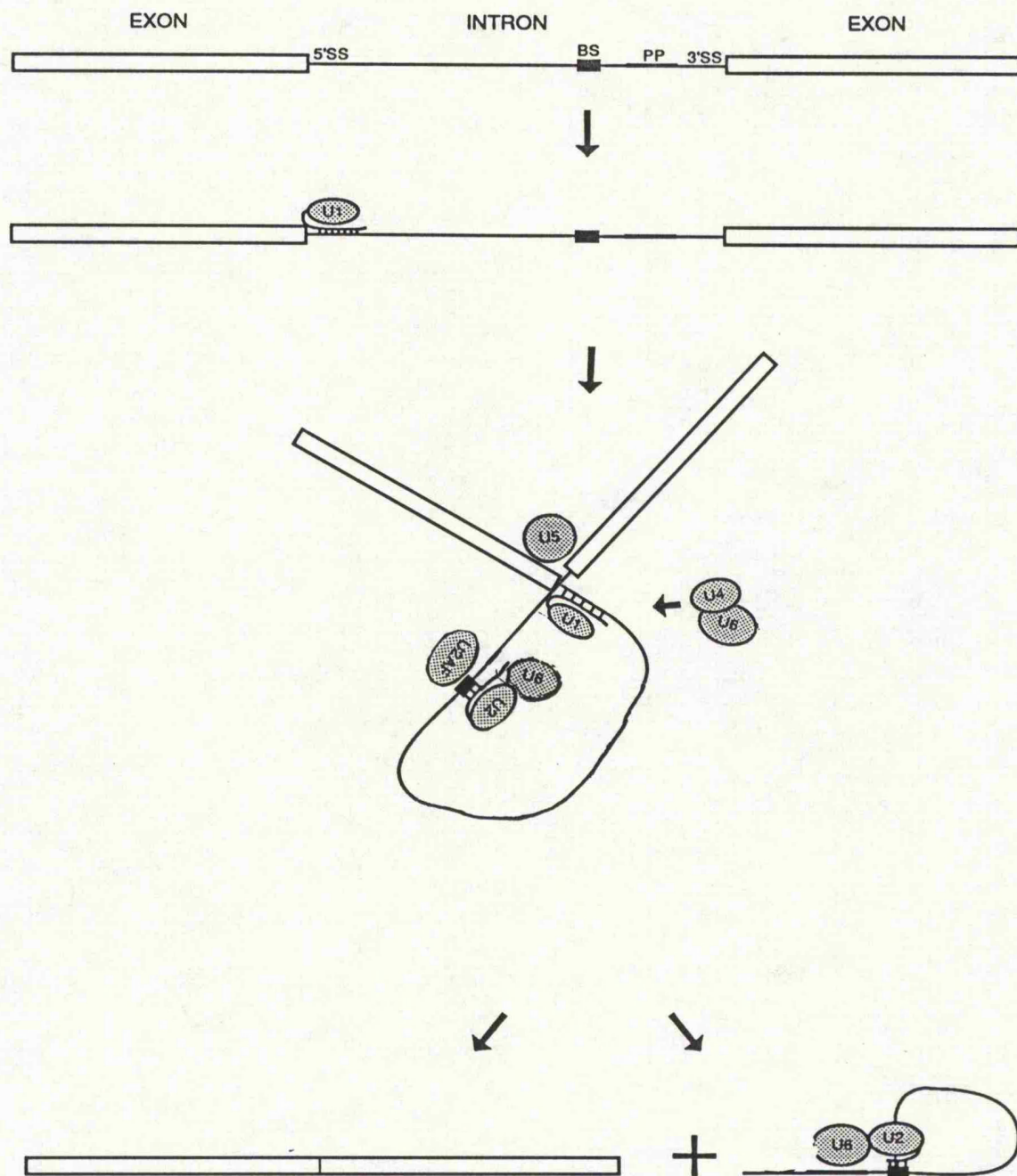


Figure 1.2

The mechanism of constitutive splicing. Exon sequences are indicated by open boxes, the position of the 5' splice-site (5' SS), branch-point (BP), polypyrimidine tract (PP), and 3' splice-site (3' SS) are shown on the first diagram. The points interaction of RNA with snRNPs U1, U2, U4, U5, and U6 and the auxiliary splicing factor U2AF are indicated.

constitutive splicing are outlined in Figure 1.2.

However, the mechanism of splice-site selection (i.e. the selection of appropriate splice-site amongst the many possible "cryptic" sites) is poorly understood. Simple intron scanning models (Sharp 1981; Lang and Spritz 1983; Wieringa et al 1983) do not fit with experimental data (Wieringa et al 1983; Kuhne et al 1983; Reed and Maniatis 1986; Konarska et al 1985a; Solnick 1985). Exon scanning or "exon definition" has been proposed as an alternative mechanism (Robberson et al 1990 ; Niwa et al 1992): the initial recognition by snRNP and other splicing factors occurs at the 3'end of the intron, an exon scanning device is then used by U1 to locate the downstream 5'splice-site. However, the presence of cryptic splice-sites within exons and the use of alternative exons indicates that simple exon scanning cannot be the sole mechanism for splice-site selection.

A role for hnRNP has also been proposed (Choi et al 1986; Dreyfuss et al 1988; Bennett et al 1992) based upon the apparent sequence dependence for binding of different hnRNP (Swanson and Dreyfuss 1988; Bennett et al 1992). This may indicate a packaging role for the hnRNP which may be active *in-vivo*, but less apparent *in vitro*. Splice-site selection experiments may therefore require the use of an *in vivo* test mechanism, rather than the *in-vitro* reactions widely used to date.

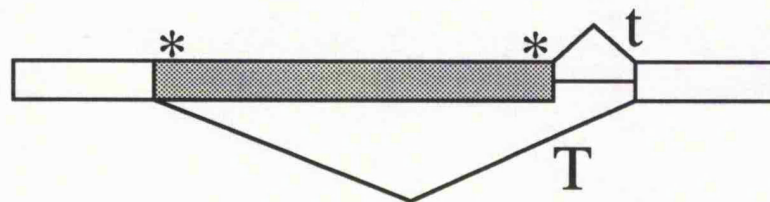
1.2.2 Alternative splicing

Alternative splicing provides a natural system in which to study splice-site selection. The analysis of the mechanisms which control regulated alternative splicing may lead to an insight into constitutive splice-site selection, particularly if heterologous transcripts share common mechanisms for alternative splice-site selection in different tissues.

Several types of alternative splicing have been identified. These have been classified into seven or eight different types (Smith et al 1989; Andreadis et al 1987). These classifications are shown below, although I have included a separate category for the inclusion of multiple exons as a unit, and have indicated that two of the categories should be regarded as special cases. The number of examples in each class are now too numerous to cite within this text. Therefore, only a few examples in each group, which are particularly relevant to this project, have been included. A comprehensive list of examples is reviewed elsewhere (Andreadis et al 1987; Smith et al 1989).

a) Alternative 5'splice-sites

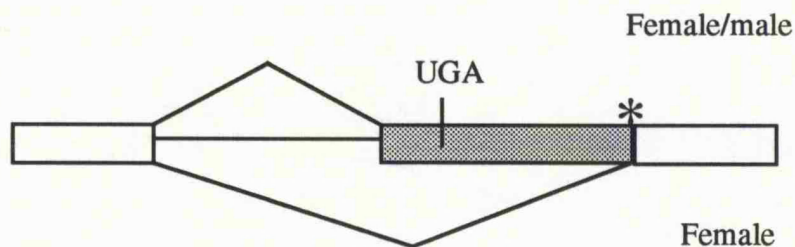
The classic example for differential use of one of two adjacent 5'splice-sites is the switch to use of the downstream 5'site of the T/t transcription unit in later stages of infection in the virus SV40 (Arrigo et al 1988). The organisation of the adjacent alternative 5'splice-sites is illustrated below.



Open boxes represent constitutive exons, the shaded box is the alternatively spliced region. * indicates that splice-site choice is influenced by the *trans*-acting factor SF2/ASF. This regulation is reviewed in a later section (1.3.2. *Trans*-acting factors).

b) *Alternative 3'splice-sites.*

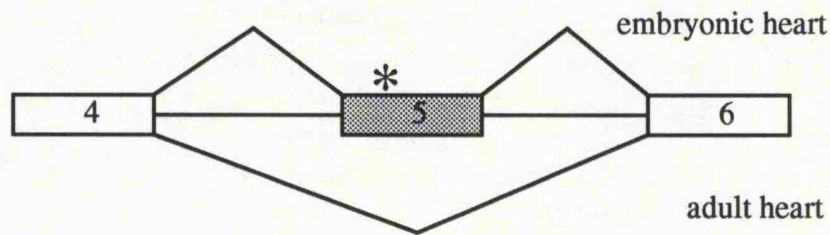
The use of alternative 3'splice-sites is illustrated by the *tra* gene (Boggs et al 1987), one of a series of genes which take part in a hierarchical system for the regulation of sex determination in *Drosophila*. The organisation and usage of the two adjacent 3'splice sites is illustrated below.



Open boxes represent constitutive exons, the shaded box is the alternatively spliced region.* indicates point of control of alternative 3'splice-site choice via the repressor protein Sxl. The details of regulation are reviewed in a later section (1.3.4 *Drosophila* genes).

c) *Incorporation/skipping of a single exon*

This cassette-type of alternative splicing is exhibited by the central exon 5 of troponin T. Troponin T is one of many genes which encode proteins for the contractile apparatus of muscle cells which utilise alternative splicing as a mechanism for increasing their isoform diversity. The patterns of splicing are illustrated below.

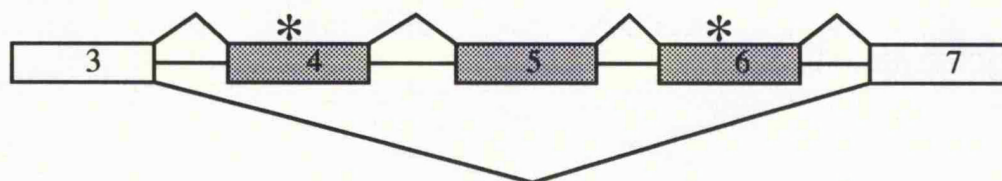


Open boxes represent constitutive exons, the shaded box is the alternatively spliced region.* indicates point of regulation is via *cis*-acting activator sequences. The details of regulation are reviewed in a later section (1.3.1 *Cis*-acting elements).

Exon 18 in N-CAM and the *Sxl* gene in *Drosophila* also exhibit this type of splicing. The identification of the control mechanisms will be discussed in a later sections (1.3.1. *Cis*-acting elements; 1.3.4 *Drosophila* genes).

d) *Incorporation/skipping of multiple exons as a unit*

The leukocyte common antigen (LCA) / CD45 gene and the exons of the muscle-specific domain (MSD) of the neural cell adhesion molecule (N-CAM) exhibit this class of splicing. The pattern of splicing for LCA/CD45 is shown below.

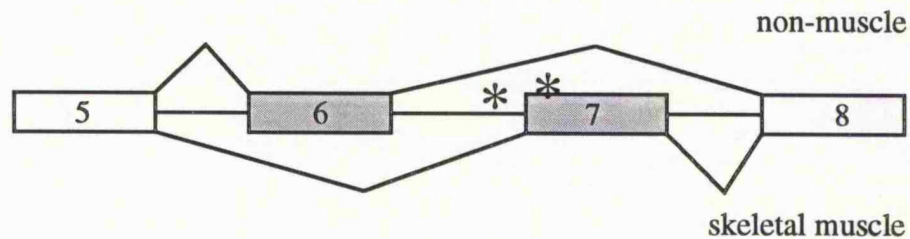


Open boxes represent constitutive exons, the shaded box is the alternatively spliced region. * indicates the positions of *cis*-acting repressor elements. Details and references are reviewed in a later section (1.3.1. *Cis*-acting elements). The patterns and control points for the alternative splicing of the MSD of N-CAM are indicated in this thesis.

e) *Mutually exclusive use of two exons*

This type of alternative splicing is common in the transcripts from contractile protein genes. The most extensively studied are the α - and β -tropomyosins.

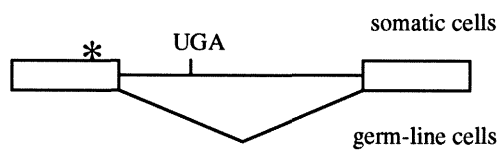
The organisation of the central exons of rat β -tropomyosin is shown below.



Open boxes represent constitutive exons, the shaded box is the alternatively spliced region. * indicates the *cis*-acting repressor elements involved in the regulation of alternative splicing of the rat β -tropomyosin. Details and references are given in a later section (1.3.4 Tropomyosin genes)

f) *Retained intron*

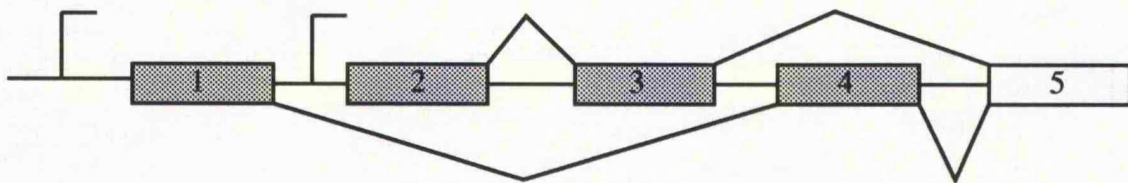
The *Drosophila* P-element transcript is alternatively spliced in the germ-line and somatic cells. In somatic cells, a non-functional truncated transposase is produced by the retention of the third IVS which encodes a stop codon. The pattern of splicing is illustrated below.



Open boxes represent constitutive exons. * indicates the position of a repressor *cis*-element. Details of this regulation are reviewed in a later section (1.3.4 *Drosophila* genes).

g) *Alternative promoters*

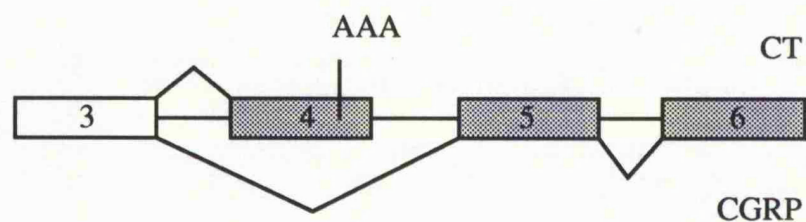
This is the first of two special examples in which alternative splicing is interconnected with another process in the formation of mature transcript. As it seems likely that transcription and splicing may be linked *in-vivo* (Jimenez-Garcia and Spector 1993), these cases may be governed by different control mechanisms to the alternative splicing of internal exons, as a competition between initiation of transcription and splicing may also be involved. An example for the use of alternative exons which contain alternative promoters is given by myosin light chain, and illustrated below.



The open box represents the constitutive exon, the shaded boxes are the alternatively spliced regions. The mechanism which regulates this alternative exon/promoter use is unknown.

h) Alternative polyadenylation signals

Again the competition, this time between splicing and 3'end termination (by polyadenylation), may introduce a different regulatory mechanism from the ones which control alternative splicing of internal exons. This type of alternative splicing is exhibited by CT/CGRP (Amara et al 1982) and IgM transcripts. The pattern of alternative splicing for CT/CGRP alternative splicing is illustrated below.



Open boxes represent constitutive exons, shaded boxes represent the alternatively spliced region.

In this gene, the 3'acceptor region of the calcitonin specific exon in the rat (Emeson et al 1989), or the unusual branch-point sequence (a uridine rather than an adenosine at the branch-point) in the human (Adema et al 1988; Adema et al 1990) are proposed as the *cis*-regulatory elements. The controlled manipulation of competition between polyadenylation and splicing, as found for the control of alternative splicing of IgM, may not be the case for calcitonin/CGRP expression (Leff et al 1987; Cote et al 1992).

The switch to production of the secreted isoform of IgM depends upon the relative efficiencies of polyadenylation at the upstream site, resulting in formation of a truncated transcript; and splicing of C μ 4 to the membrane specific exon which splices out the upstream poly(A) site (Paterson and Perry 1989). This regulation is linked to transcription: reducing the spacing between the alternative poly(A) sites causes a switch in splicing (Tsurushita and Korn

1987; Galli et al 1988).

N-CAM contains a similar mechanism for the production of membrane bound isoforms, produced at later stages of muscle development, by the incorporation of exon 15.

1.3 Control of alternative splicing

The mechanisms which govern the regulation of alternative splicing can be subdivided conveniently into elements that operate in *cis* and those that operate in *trans*. *Cis*-acting elements are sequence dependent. They may be contained within introns, splice-sites, or exons and function either by the formation of a specific secondary structure, or by RNA/RNA or RNA/protein interactions with other (*trans*-acting) elements. *Trans*-acting elements (or factors) usually describe the products of other genes (either RNA, protein or multicomponent complex), although some experimental models could not exclude the possibility that RNA sequence from a distal site of the transcript (and therefore a *cis*-element) could function as a *trans*-acting factor.

1.3.1 *Cis*-acting elements

The formation of secondary structure, by the folding of RNA into stem-loop structures, has been proposed as a mechanism whereby an exon (Solnick 1985), or an exon splice-signal (Eperon et al 1988) could be sequestered from the general splicing machinery. Formation of the structure could be influenced by the binding either of sequence-specific *trans*-acting factors, or by the binding of non-specific factors such as hnRNPs. In the later case, the time interval in which small structures could form, before their repression by the putative binding of hnRNPs, was investigated and shown to involve a critical loop length (Eperon et al 1988).

The roles for secondary structures, in alternative splicing, have been investigated in vitro (Solnick 1985; Solnick and Lee 1987; Eperon et al 1988; Watakabe et al 1989;) and in vivo (Solnick 1985; Solnick and Lee 1987; Eperon et al 1988; Libri et al 1989a; Libri et al 1990b; Libri et al 1991; Goux-Pelletan et al 1990; D'Orval et al 1991). Regulation of alternative splicing has been

investigated extensively in chicken β -tropomyosin and has indicated that secondary structure formation may play a role in the alternative splicing of the central exons. This regulation is reviewed in a later section (Chapter 1, section 1.3.3).

The role for *cis*-acting elements, which act to repress or activate the alternatively spliced exon (probably via their interaction with a *trans*-acting factor), has been investigated for a number of different genes (see Smith et al 1989 for review). Most pertinent to this project are the alternatively spliced genes which exhibit either cassette-type alternative splicing, or mutually exclusive alternative splicing. Mutually exclusive alternative splicing has been studied extensively in the tropomyosin gene family. Tropomyosin *cis*- and *trans*-acting elements are outlined in a later section.

The leukocyte common antigen (LCA/CD45) expresses five isoforms in different tissues and stages of thymic development. In T cells the central three exons (4, 5 and 6) are skipped whereas they are included in transcripts from mature B cells. The *cis*-acting elements involved in the alternative splicing of exon 4 (Streuli and Saito 1989) and exon 6 (Tsai et al 1989) have been identified by linker scanning analysis, and found to reside within the exons themselves. Recently, as a result of transient fusion studies of mouse and human B and T cells, the mechanism by which the alternative splice could be achieved was proposed as a negative *trans*-acting factor present in T cells (Rothstein et al 1992). Inclusion of the central exons in B cells was interpreted as a default choice (Rothstein 1992).

Cardiac troponin T (cTNT) alternatively splices exon 5 in a tissue- and developmental-specific manner. Sequence analysis of the region surrounding the exon 5 indicated that large, and potentially stable, secondary structures could form which would sequester the exon (Cooper et al 1988). However,

deletions which removed complementary intron regions had no effect on alternative splicing patterns (Cooper et al 1988). The *cis*-acting elements have now been localised to within the alternatively spliced exon (Cooper and Ordahl 1989). Recently, *in-vitro* analysis indicated the regulation to be via the "activation" of the exon which allows the exon 4-exon 5 splice (Cooper 1992). This could be in a manner analagous to the 3'splice-site activation of drosophila *dsx* transcripts (Hedley and Maniatis 1991), or via induction of exon recognition via exon definition (Robberson et al 1990).

The penultimate exon of N-CAM (exon 18) is alternatively spliced in a stage specific manner, such that it is included in transcripts derived from mature neural cells, but is skipped in non-neural cells. The point of control for this alternative splice has been localised to the 5' splice-site of exon 18 (Tacke and Goridis 1991). The regulation was sequence specific such that a mutation of β -globin exon 2 in which only the 5' splice-site was altered to the exon 18 sequence induced a repression of downstream intron excision in non-neural cells, which was partially relieved on differentiation. A role for the alternative splicing factor ASF/SF-2 was proposed as one mechanism by which a 5' splice-site was recognised and used in one tissue, but was dormant in another (Tacke and Goridis 1991).

1.3.2 *Trans-acting factors*

Trans-acting factors for the regulation of alternative splicing have been identified by two routes: biochemical characterisation of components of splicing extracts; and the molecular analysis of mutations in *Drosophila* (these are outlined in a later section).

Four alternative *trans-acting* factors have now been identified by the former route, although the sequence analysis of cDNA encoding two of these factors (ASF and SF2) has recently indicated that they are the same factor (Ge et al 1991; Krainer et al 1991).

SF2, although originally thought to be a necessary component of constitutive splicing in *in-vitro* splicing reactions (Krainer et al 1990a), has now been shown to act to enhance the use of proximal 5'splice-sites when duplicate sites are available (Krainer et al 1990b). This result has now been corroborated by the independent finding that ASF promotes a switch in 5'splice-site use from the distal large T site to the proximal small t site in an *in-vitro* splicing reaction of SV40 transcripts (Ge and Manley 1990). Analysis of the biochemical function of ASF/SF2 indicated an RNA binding activity (Krainer 1990a; Ge and Manley 1990) although ASF did not contain an associated RNA moiety, as is found in snRNPs (Ge and Manley 1990). Sequence analysis indicates that ASF/SF2 contains RNA recognition and binding domains (Krainer et al 1991) which share a homology with the RNP-1 and RNP-2 elements identified in U1 70K, the *Drosophila* regulatory proteins Sxl and Tra-2 (reviewed in Mattaj 1989). In common with the sex determining regulatory proteins Sxl and Tra, ASF/SF2 is also alternatively spliced (Ge et al 1991), the significance of this remains to be elucidated although it raises interesting questions about the potential mechanisms that regulate

alternative splicing in higher eukaryotes, and the possibility that these mechanisms may be conserved between flies and man.

A distal splicing factor (DSF), capable of promoting a switch from the proximal 12S and 9S 5'splice sites to the distal 13S 5'splice site of E1A transcripts, has recently been purified and characterised (Harper and Manley 1991). The splice-site selection activity of DSF has also been shown to be substrate independent (Harper and Manley 1991) as the switch is also observed with SV40 transcripts. The primary sequence of cDNA encoding DSF is, at present, unknown.

1.3.3 The tropomyosin gene family

The tropomyosin gene family has provided a suitable model for the analysis of alternative splicing mechanisms. To date, four different genes have been studied: the β -tropomyosin gene from chicken (Libri et al 1989a; Libri et al 1989b; Libri et al 1990a; Libri et al 1990b; Goux-Pelletan et al 1990; D'Orval et al 1991; Libri et al 1991; Libri et al 1992; Balvay et al 1992); the homologous β -tropomyosin gene from rat (Helfman et al 1988; Helfman and Ricci 1989; Helfman et al 1990; Guo et al 1991); the α -tropomyosin gene of rat (Smith and Nadal-Ginard 1989; Mullen et al 1991); and the α_s -tropomyosin gene (hTMnm) from human (Graham 1992; Graham et al 1992; this project). The human α_s -tropomyosin and the rat α -tropomyosin are not true homologues: the isoforms produced from the complex rat α -tropomyosin gene appear to be encoded by three separate, and less complex, genes in the human.

The chicken β -tropomyosin contains two central exons which are alternatively spliced, in a mutually exclusive manner, in different tissues and stages of development. These two exons, 6A and 6B, and their flanking introns contain all the information required for appropriate alternative splicing in both muscle and non-muscle cells (Libri et al 1989b). Until very recently, the correct splicing of transcripts in differentiated muscle cells (myotubes) could only be achieved in stably transfected cell lines (Libri et al 1989b). This technical point has resulted in an emphasis on investigations into the *in vivo* mechanisms which govern the exclusion of the skeletal muscle exon (6B) in transcripts produced in non-muscle cells, and the *in-vitro* analysis of regulation using non-muscle (HeLa) nuclear extracts. *In vivo* transfection of mini-gene constructs into quail myoblasts identified two regions in which mutations induced de-repression of exon 6B in myoblasts: a deletion of 34 nucleotides from the polypyrimidine tract combined with the introduction of a "consensus" branch-point sequence at position -22 (the

native branch-point is 105 nucleotides upstream of the 3'splice-site), and mutation of the first 16 nucleotides of the 6B exon (Libri et al 1990b). These were identified as part of a potential stem-loop structure which had already been predicted from the genomic DNA sequence (Libri et al 1989a). A model based on the involvement for large secondary structures was proposed (Libri et al 1990b). The reversal of exon choice by compensatory suppressor mutations in the proposed apposing portions of the regulatory stem region (Libri et al 1991), produced further evidence in favour of a secondary structure model.

Using *in-vitro* splicing experiments, the deletion of a 65 nucleotide region between the branch-point and the 3'splice-site was shown to result in the derepression of exon 6B use. Insertion of a heterologous 65 nucleotide sequence (pBR322) into the deletion mutation still resulted in derepression of 6B, indicating that the sequence and not the distance between the branch-point and the 3'splice-site was the negative *cis*-acting element (Goux-Pelletan et al 1990). *In-vitro* splicing of several other mutations, which would disrupt a larger secondary structure around the 6B exon (D'Orval et al 1991), also resulted in a derepression of splicing to the 6B exon. These sets of data were both proposed as further evidence for the role of secondary structures (Libri et al 1989a) in regulation of alternative splicing (Goux-Pelletan et al 1990; D'Orval et al 1991). The involvement of other, non-secondary structure, *cis*-acting elements has recently been identified (Libri et al 1992; Balvay et al 1992). Mutation of the branch-point and 5'splice-site of exon 6B to consensus sequence induces use of the repressed 6B exon in myotubes (Libri et al 1992), and mutation of a 30nt sequence 37 nucleotides downstream of the 6A exon causes a profound switch toward use of the downstream 6B exon (Balvay et al 1992). This element appears to be both sequence and position dependent (Balvay et al 1992).

Shortly before completion of this project, Libri and co-workers (Libri et al

1992) published details of a transient transfection system for analysis of alternative splicing in quail myotubes, by using a temperature sensitive mutant cell line (although the cells must be maintained at the non-permissive temperature for at least 72 hours for the muscle-specific exon switch to occur).

Four mutations were tested in myotubes: a null mutation of the 3' and 5'splice-sites of exon 6B led to the use of exon 6A in myoblasts (wt pattern) and in myotubes; deletion of exon 6A resulted in the skipping of exon 6B in myoblasts and the use of exon 6B in myotubes (wt pattern); mutation of the 5'splice-site of 6A to the consensus sequence resulted in exclusive exon 6A use in myoblasts (wt pattern) and an increase in exon 6A use in myotubes (both in conjunction with exon 6A (double exon inclusion) or on its own); and mutation of a 15 nucleotide sequence just upstream of branch-point of exon 6B resulted in an increase in exon 6B use in myoblasts and myotubes (wt pattern). These data indicated that inactivation (by deletion or mutation of the splice-sites) of one exon enables the other exon to be incorporated in transcripts produced in myotubes. The authors cite this as evidence that a competition for splicing of the exons and their flanking constitutive splice sites regulates alternative splicing of b-tropomyosin in myotubes. The ability for exon 6A to be incorporated when the 5'splice-site is uprated to the consensus sequence is evidence in favour of this model. The role for the branch-point and polypyrimidine tract signals in the competition is more circumstantial, because the mutations which alter the branch-point and the polypyrimidine tract were not tested in myotubes, but were analysed by transfection of myoblasts using a background mutation which caused the derepression of exon 6B use in myoblasts. The authors predict that the branch-point of exon 6B is in competition with the weaker branch-point of exon 6A and that splicing components assemble more rapidly around the 6B branch-point. Splicing of exon 5 to 6B then proceeds as a result of a

competition between the exon 5 5'splice site and the poor 5'splice-site of exon 6A for the 6B acceptor site.

The rat β -tropomyosin contains two central exons designated (6 and 7) which are alternatively spliced. Exon 6 is found in transcripts from non-muscle cells, exon 7 is skeletal muscle specific. The mechanisms regulating alternative splicing of these exons have been investigated by *in-vitro* splicing experiments (Helfman et al 1988; Helfman and Ricci 1989; Helfman et al 1990), and transient transfection of non-muscle (HeLa) cells (Helfman et al 1988; Helfman et al 1990; Guo et al 1991). Regulation is achieved by an ordered pathway of splicing in which the critical point is exon 6/8 or exon 7/8 splicing (Helfman et al 1988). Once this downstream splice is made, the intermediate form then splices exon 5 (Helfman et al 1988). Thus, only 5/6/8 or 5/7/8 transcripts are produced. A role in the regulation of alternative splicing by the unusual branch-point location (Helfman and Ricci 1989), and the high polypyrimidine content of the region between the branch points and the 3'splice-site of exon 7, has been suggested (Helfman and Ricci 1989; Helfman et al 1990). Also, two distinct *cis*-acting elements have also been identified: mutation of a short sequence within exon 7 (5 nucleotides near the 5'end of the exon) causes the exon to be incorporated (Guo et al 1991), and mutations in the intron, just upstream of the 3'splice-site of exon 7, also induce incorporation of exon 7 (Helfman et al 1990; Guo et al 1991). The intron *cis*-acting element contains seven overlapping copies of an imperfect repeat (YRYYYRYYYRY where Y=pyrimidine and R=purine) and is proposed as a binding site for a *trans*-acting repressor present in non-muscle (Guo et al 1991). The switch to incorporation of exon 7 in transcripts from HeLa cells which have been transfected with high concentrations of vector (Guo et al 1991), implies that a regulatory factor is limited in HeLa cells and can be titrated out. A gel-mobility shift assay indicated the presence of a factor in

HeLa nuclear extract which appeared to bind the intron *cis*-acting element (Guo et al 1991). This factor has not been characterised. A role for secondary structure has not been formally excluded for the control of alternative splicing of the rat β -tropomyosin (Helfman et al 1990). Tissue specific *trans*-acting factors would bind to sequence specific *cis*-acting elements to either stabilise or disrupt the formation of such structures (Guo et al 1991).

Investigation into the alternative splicing of two exons (2 and 3) situated at the 5' end of the rat α -tropomyosin gene revealed a novel mechanism which caused the absolute mutual exclusive behaviour of these exons. The branch-point of exon 3 is only 42 nucleotides downstream of exon 2, too short a distance for lariat formation to occur (Smith and Nadal-Ginard 1989). This block on exon 2 to exon 3 splicing was relieved by the insertion of spacer sequences (Smith and Nadal-Ginard 1989). The use of exon 3 is proposed as the default pathway, in the absence of exon 3 exon 2 is incorporated (Mullen et al 1991). The mechanism by which exon 3 is selected as the default is a competition between the branch-point/polypyrimidine tracts of the two exons. A hierarchy, deduced from mutations which swap either or both the branch-points and the polypyrimidine tracts of exon 2 and exon 3 *in-vitro* and *in-vivo*, indicated that the branch-point/polypyrimidine tract combination for exon 3 was stronger than that for exon 2 (Mullen et al 1991). In the search for *trans*-acting factors which may influence exon selection, UV cross-linking identified an approximately 60kDa protein which bound the polypyrimidine tract of exon 3 (Mullen et al 1991). This protein remains to be characterised, but may prove to be the same as the 62kDa polypyrimidine tract binding protein (PTB) identified by Garcia-Blanco and co-workers (Garcia-Blanco 1989).

The human α_5 -tropomyosin gene also contains two alternatively spliced

central exons (NM and SK, located between the constitutive exons 4 and 6), which are alternatively spliced in a mutually exclusive manner. Recently the mechanism by which the SK exon is excluded from transcripts produced in non-muscle cells (COS) has been identified (Graham 1992; Graham et al 1992; this project). Two important *cis*-acting elements were identified: the first 15 nucleotides (and to a lesser extent the last 15 nucleotides) of exon SK contained *cis*-acting repressor elements, and the weak branch-point of exon SK were both involved in the effective sequestration of the SK exon in non-muscle environments. The exons were not in competition in non-muscle cells, as deletion of the "default" exon NM did not result in incorporation of the SK exon (Graham et al 1992; this project). The *cis*-acting element within the exon was thought to be a possible site for binding of a non-muscle specific repressor (Graham et al 1992), although this factor has not been identified.

1.3.4 *Drosophila* genes

Drosophila has provided a useful model for the study of alternative splicing because of the extensive information available on the effects of mutations of one gene upon another. In *Drosophila*, both the *cis*-acting elements and their *trans*-acting factors can be investigated, which enables analysis of the molecular basis for the interaction by sequence and mutational analysis of the genes. Three such gene interactions have been studied.

The regulatory gene *suppressor-of-white-apricot* (*su(w^a)*) which exhibits alternative splicing by a retaining intron mechanism under developmental regulation (Chou et al 1987) has been shown to autoregulate the splicing of its transcript (Zachar et al 1987).

The transposable P-element is also alternatively spliced by a retained intron mechanism (Laski et al 1986): complete splicing and thence production of full

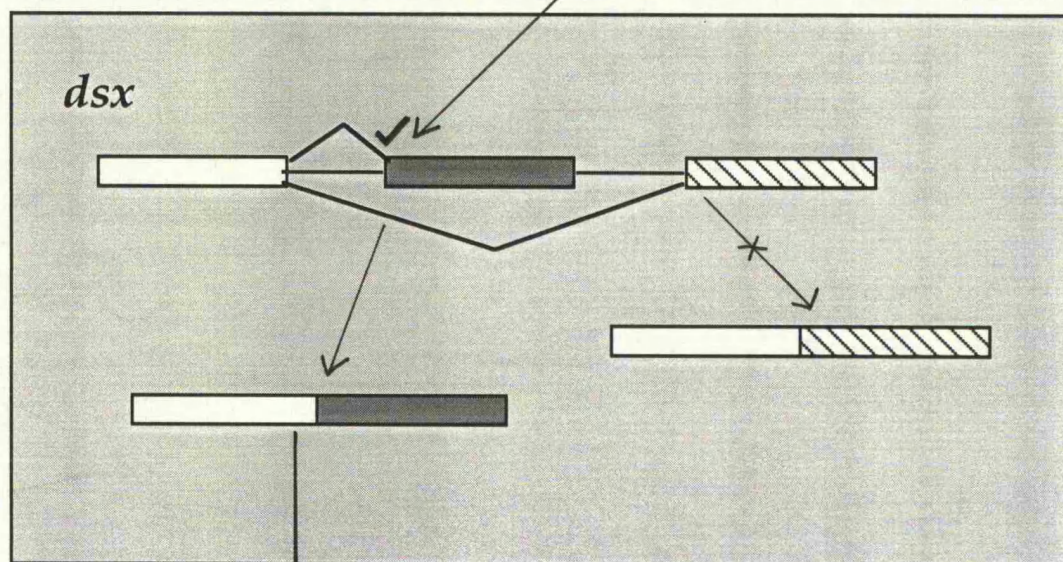
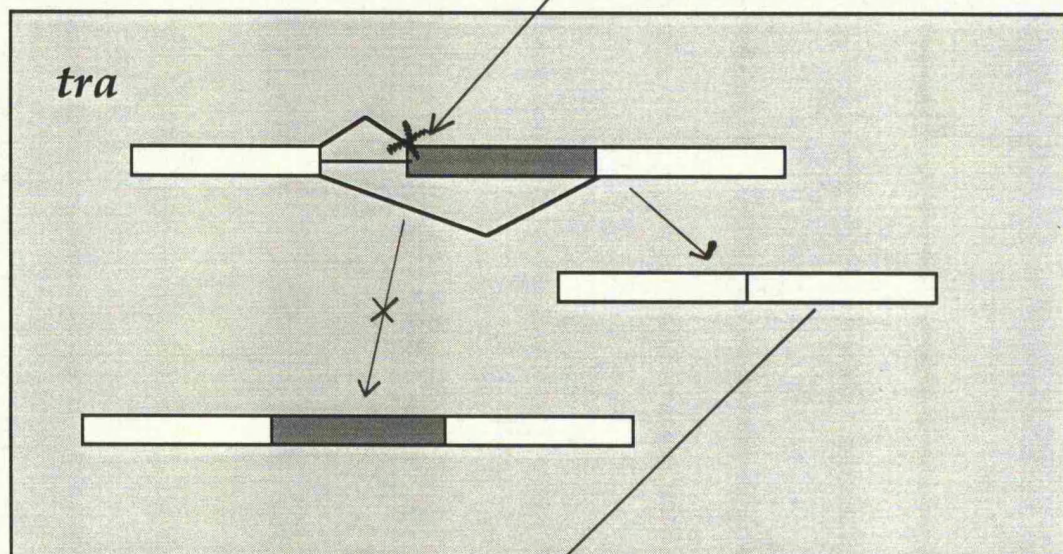
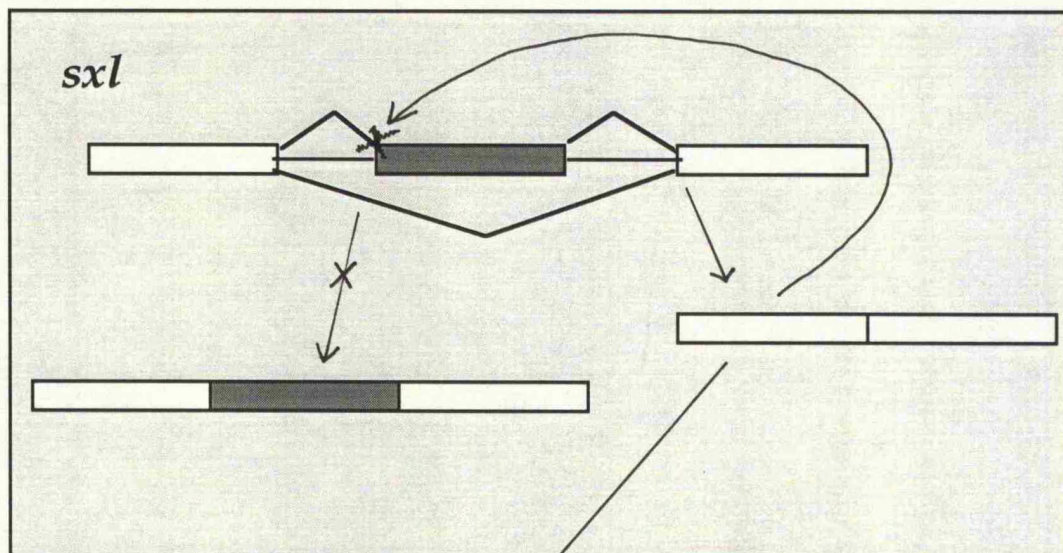
length product only occurring in the germ-line. The *cis*-acting elements have been identified genetically (Laski and Rubin 1989) and biochemically (Siebel and Rio 1990; Tseng et al 1991) as a region within the second open reading frame (ORF). Putative *trans*-acting factors have been identified (Tseng et al 1991; Siebel and Rio 1990) but not characterised.

The sex determining genes in *Drosophila* are involved in a complex "cascade" of interactions (see Maniatis 1991; Smith et al 1989; Hodgkin 1989; Baker 1989; Wolfner 1988; Bingham et al 1988 for reviews). *Sex-lethal* (*sxl*) is at the top of the hierarchy. The female-specific alternative splice of transcripts from *sxl* is initiated by the X:A ratio of chromosomes and requires both maternal and zygotic supplied gene products (see Baker 1989 for review), although the actual mechanism is unknown (Baker 1989). The female-specific gene product (Sxl) maintains the female specific alternative splicing of itself by repression of the male specific 3'splice site (Bell et al 1988; Bell et al 1991) in conjunction with the product of *female-lethal-2-d* (*fl(2)d*) (Grandino et al 1990)); and of *transformer* (*tra*) transcripts by repression of the non-sex specific 3'splice site (Sosnowski et al 1989; Inoue et al 1990). The female specific product of *tra*, in conjunction with the product of *transformer-2* (*tra-2*), determines the splicing pattern of the transcripts of the *double-sex* gene (*dsx*) (Nagoshi et al 1988; Burtis and Baker 1989; ; Nagoshi and Baker 1990; Hedley and Maniatis 1991; Hoshijima et al 1991) by activation of the female specific 3'splice-site (Ryner and Baker 1991; Tian and Maniatis 1992).

Sequence analysis of the gene products Tra and Su(w^a) reveals a homology between these proteins and U1 70k (a protein which forms part of U1snRNP) (Mattaj 1989). The gene products Sxl and Tra-2 contain RNA binding domains which include the RNP consensus (Amrein et al 1988; Bell et al 1988; Mattaj 1989).

Figure 1.3.4. The sex determining cascade of *Drosophila*. The major points of interaction between the products of alternative splicing and their target regions on pre-messenger RNA transcripts are indicated. Open boxes represent constitutively spliced exons and shaded boxes regions of alternative splicing. Bracketed arrows indicate the alternative pathway taken by non-female specific splicing.

X:A ratio



Female fly

1.4 Evolution of alternative splicing.

Alternative splicing may have evolved as a mechanism which allowed a single gene to produce different protein isoforms by using different combinations of its constituent exons. Production of the different isoforms may have been selectively advantageous if their expression could be regulated. Alternative splicing could therefore be said to have evolved secondary to the evolution of the gene. Alternatively, alternative splicing may have evolved first: any gene (and hence organism) which had the capacity to utilise this apparatus may then have gained an advantage. Alternative splicing as a mechanism would be the driving force, the evolution of the gene would be secondary. In the first scenario, in which the mechanism of alternative splicing may be specific to each gene, common mechanisms would not be expected. However, in the alternative splicing first model, the mechanisms used by different genes would be predicted to share many features. Thus, the identification of common features in different systems of alternative splicing may provide an insight to the evolution of alternative splicing as a strategy for the regulation of gene expression.

1.4.1 Concerted evolution of *trans-acting* factors

In order for an alternative splicing strategy to be maintained, *trans-acting* factors and the *cis*-elements involved in regulation must co-evolve in a concerted manner. As these elements are encoded on different genes there would appear to be a significant advantage for the evolution of close linkage between the gene encoding the specific *trans-acting* factor and its target gene. The most extensive information about the position of *cis-acting* elements and the genes that encode the *trans-acting* factors that interact with those elements is found for the *Drosophila* sex determining genes. The linkage of these genes can be estimated by their relative position within the genome. The *sxl* gene is located on chromosome 1 at position 1-19, *tra* on chromosome

3 at 3-45, *tra-2* on chromosome 2 at 2-70, *ix* on chromosome 2 at 2-60.5, and *dsx* on chromosome 3 at 3-48.1. Of these, *tra* and *dsx* are closely linked (only 3.1 map units apart). This may be a significant finding, as the product of the *tra* gene has been shown to be the activator of female specific splicing of *dsx* RNA. Two other genes, *ix* and *tra-2* are also closely linked. It is interesting that two pairs of the 5 genes indicated are closely linked. If linkage disequilibrium is important for the co-segregation of an alternatively spliced gene and its *trans*-acting factor, an alternative approach for the identification of *trans*-acting factors could involve a chromosome walk or jump from the gene in question. However, this may be more feasible when more sequence data is available for eukaryote genomes (e.g. via the human genome project).

1.4.2 Constraints on exon duplication

The major constraint on exon rather than gene duplication is the potential that a frame-shift will be introduced into the processed transcript. This provides a conundrum for the evolution of the mutually exclusive exons of tropomyosin. The mutually exclusive exons NM and SK of hTMnm are both 76 nucleotides long, incorporation or skipping of both would lead to a frame shift. Two mechanisms can be envisaged for the evolution of the extant gene. First: if these exons had evolved by slippage during replication, or unequal crossover, expression of this gene would be defective, unless the complex alternative splicing machinery were also present, or had evolved at exactly the same time. If an alternative splicing mechanism has already been established in another gene (which must beg the question as to how the first one arose?), this evolutionary scenario is plausible. If this is the case, common mechanisms would be predicted for the regulation of splicing of disparate genes. Second: the NM and SK exons evolved separately on different genes, subject to their own evolutionary pressures. The differential expression of the "NM" gene was controlled by sequences contained between

the flanking exons to the extant NM exon (see below). Similarly, "SK" gene expression was controlled by the sequences contained between the flanking exons to the extant SK exon. These two genes recombined (because of their common flanking exons) such that both NM and SK were incorporated into a single gene. The mechanism which shut down expression of one gene, via interaction with sequences surrounding the NM or SK exons depending on the tissue-type, could now operate as an alternative splicing mechanism. This model predicts that repressors but not activators for each exon will be produced by cells that exhibit alternative splicing of these transcripts. This model also predicts that genes which include mutually exclusive exons were derived from a larger family of related genes, some of which may now be pseudogenes. The tropomyosins are found as such a family (Sri-Widada et al 1989).

1.4.3 A role for diploidy and sex?

The presence of two sets of chromosomes allows a unique combination of genetic material to be produced (at meiosis) in the gametes of diploid organisms. The fusion of gametes from two individuals in the production of offspring (sex) produces individuals with a unique genetic identity. Diploidy also allows for recombination between sister chromatids which leads to either homologous exchange (and hence further diversity of genetic constitution) or non-homologous exchange (which causes a mutation). Genetic recombination is proposed as one mechanism for exon duplication and gene conversion, two mechanisms which have both been implicated in the evolution of alternative splicing (see section 1.1.3 Gene duplication and gene conversion). Without the availability of two copies of the gene, these mechanisms for the evolution of alternative splicing cannot be invoked.

1.5 Muscle as a model system.

Numerous examples of alternative splicing have been identified (reviewed in Smith et al 1989; Andreadis et al 1987; Breitbart et al 1987; this thesis). Interestingly, many of the examples are from genes which exhibit differential expression in muscle and non-muscle tissues. This raises the question: are these transcripts regulated by a common mechanism? Analysis of the regulation of alternative splicing of transcripts from heterologous genes explains not only the mechanism for the gene under study but addresses the wider question of common or specific regulation. To date, insufficient information has been collated about the regulation of alternative splicing, particularly in skeletal muscle, to ascertain whether common mechanisms are operating. The aims of this project are to study the alternative splicing of N-CAM and hTMnm in muscle and non-muscle cells, in order to widen our knowledge on the mechanisms involved in the alternative splicing of unrelated transcripts from a common environment (muscle). The findings may have significant implications for our understanding of exon evolution, and the evolution of the constitutive and alternative splicing apparatus.

1.5.1 C-2/C-12 muscle cells as an *in-vivo* test system.

Much of the work on the regulation of non-muscle/muscle-specific exon splicing has centred on the regulation of non-muscle splicing patterns using non-muscle cells (COS and HeLa) and non-muscle cell extracts (HeLa) as the test systems. This lack of investigation into the regulation of skeletal muscle exon selection is in part due to the difficulties associated with transient expression of exogenous genes in differentiated cells (Libri et al 1998b). Libri and co-workers have only recently reported the use of a temperature sensitive mutant of quail cells which could be used to study transient transfection in differentiated cells (Libri et al 1992).

Mouse C-2/C-12 cells, which exhibit a non-muscle physiology as myoblasts can be induced into a myogenic pathway in which they exhibit a muscle-type physiology. These cells have been cultured routinely in this laboratory and appeared to be an ideal candidate for the analysis of regulation of non-muscle/muscle-specific exon use. However, the ability for these cells to express the appropriate alternatively spliced transcript from transient transfections was unknown. One of the aims of this project was to establish whether transient transfection of C-2/C-12 differentiated muscle cells was possible.

If transient transfection in C-2/C-12 cells could be established this would provide a model system in which the alternative splicing of exogenous transcript could be assessed at different stages of muscle cell development. Also, if transient transfection of differentiated cells could be produced by their transfection as undifferentiated cells, the switch in alternative splicing could be "followed" through the development of the muscle cell.

1.5.2 N-CAM: a model for cassette-type alternative splicing

Neural cell adhesion molecules (N-CAMs) are a group of cell surface sialoglycoproteins (see Edelman and Crossin 1991 for a review) which belong to the immunoglobulin (Ig) super-family (Hunkapillar and Hood 1986). N-CAM mediates calcium-independent interactions between cells; homophilic binding (Cunningham et al 1987) of N-CAMs on apposing cells results in enhanced or extended interactions between cells (Dickson et al 1990; Kadmon et al 1990a; Kadmon et al 1990b; Doherty et al 1989; Doherty et al 1990; Rutishauser et al 1988).

A number of isoforms of N-CAM have been reported which differ in their mode of membrane attachment and in cytoplasmic domain structure (Owens

et al 1987, Barthels et al 1987, Small et al 1987, Barbas 1988; Barton et al 1988) and in their extracellular domains (Santoni et al 1989, Small et al 1990; Reyes et al 1991, Dickson et al 1987, Gower et al 1988), although only a single N-CAM gene has been demonstrated in a variety of species (Walsh and Dickson 1989). These isoforms appear to arise via alternative splicing that is regulated in a developmental and tissue-specific manner (Moore et al 1987, Reyes 1991; Hamshire et al 1991).

One region of human N-CAM gene, the MSD (Dickson et al 1987), was found to contain at least three or four exons (MSD1a, MSD1b, MSD1c and a triplet AAG) which appeared to be incorporated or skipped in a concerted manner (Thompson et al 1989). MSD1b, MSD1c and SEC had not been identified in the mouse, either in a genomic clone or in cDNA derived from mRNA. An MSD like insert had been identified in cDNA from the chicken (Prediger et al 1988), but was 15 bases shorter than the human MSD.

These findings indicated the need for a comprehensive analysis of exon usage in the mouse: first, to determine if homologues to MSD1b, MSD1c and SEC were present; second, to deduce the sequence and length of the exons to predict whether the difference in MSD1b length pre-dates the chicken/man divergence; and, third, to elucidate the alternative splicing patterns exhibited by these exons.

The proposed splicing patterns for these exons suggested a complex regulation for the tissue-specific incorporation of exons of the MSD. The analysis of the regulation of alternative splicing in a complex system such as N-CAM may provide insights into the co-ordinated regulation of different genes, and perhaps indicate a common mechanism for the incorporation of alternative exons in a tissue-specific manner.

1.5.3 Tropomyosin: a model for mutually exclusive alternative splicing

Human α_S -tropomyosin or hTMnm (Clayton et al 1988) is a member of the tropomyosin multi-gene family. The tropomyosins are a group of proteins shown to be important for calcium-mediated contraction in skeletal muscle. Tropomyosins have a conserved α -helical coiled-coil structure which binds to actin. Tropomyosin isoforms are produced in a tissue dependent manner by alternative splicing and the use of alternative promoters (Helfman et al 1988).

Tropomyosin genes from several species have been cloned and sequenced. The manageable size of the exons and introns in the regions of alternative splicing has allowed the production of mini-gene constructs in which the entire exon and flanking intron sequences could be included. The control of alternative splicing in these constructs has been investigated both by in vitro splicing experiments and in vivo by transfection of mammalian cells.

Analysis of the control of alternative splicing of the rat, and chicken α - and β -tropomyosins has given some indications to the mechanisms which govern the use of the non-muscle exon in non-muscle tissue. However, only very recently have there been any indications to the nature of the regulation of alternative splicing in muscle cells (Libri et al 1992).

The aims of this part of the project were three-fold. First, to extend our present knowledge of the control of non-muscle splicing of human α_S -tropomyosin (hTMnm) in true non-muscle cells (kidney epithelial COS cells). Second, to identify the control of alternative splicing in pre-muscle myoblasts (which also exhibit a non-muscle tropomyosin phenotype). Third, to elucidate the mechanism of control of alternative splicing in myotubes (which exhibit a muscle-type tropomyosin phenotype).

The *cis*-acting elements which control this regulation were investigated by the analysis of RNA from COS cells, pre-muscle myoblasts and myotubes, which had been transiently transfected with wild-type and mutant α -tropomyosin mini-gene constructs. Exogenous RNA was analysed by a quantitative reverse-transcriptase-PCR.

An investigation into the mechanism of regulation of alternative splicing of human α _S-tropomyosin in muscle cells would not only elucidate the control of expression of that particular human gene, but give an indication as to whether mechanisms were common for different genes from different species. The finding of a common mechanism for different tropomyosin genes in diverse species, such as chicken, rat and man, would have implications for our view on the evolution of exons and alternative splicing.

1.6 Identification and quantification of RNA

Several methods have been developed for the identification of specific RNA sequences. The analysis of RNA from *in vivo* experiments relies largely upon the exploitation of specific sequences within the RNA by the use of radiolabelled probes. In *in vitro* analysis, such as in *in-vitro* splicing using cell-free nuclear extracts, the RNA is visualised directly, and the length and conformation of the RNA products, determined by polyacrylamide gel electrophoresis, leads to the identification of RNA products.

1.6.1 S1 mapping

S1 mapping involves the hybridisation of total RNA from a cell extract to a radioactive end-labelled probe. RNA/DNA duplex forms where the probe and target RNA are complementary. An endonucleolytic nuclease (DNase S1) which cleaves single stranded DNA and RNA is then incubated with the duplex. The length of protected fragment (deduced by polyacrylamide gel electrophoresis) can be related to the probe, and thence the region of RNA protection assigned. This method has proved useful for the analysis of mutually exclusive alternative splicing where only two isoforms are produced. Parallel S1 mapping experiments with both probes allows for the accurate assessment and quantification of alternative exon use. However, S1 mapping strategies are unsuitable for the assessment of alternative splicing where more than two isoforms can be produced. This is almost certainly the case for the MSD region of N-CAM, and may be important in the analysis of the effect of mutations in a-tropomyosin mini-gene constructs. In the latter case the level of inclusion of both exons as double incorporation in a single transcript, or the skipping of both, (thereby producing two more potential isoforms) would be difficult to quantify by an S1 strategy. It has also been noted that when an extra sequence is inserted into the RNA (as is found in double incorporation), the DNA probe can be fully protected even though a

region of the RNA has been looped out (Sisodia et al 1987; Graham 1992) .

1.6.2 Primer extension

Primer extension of an oligonucleotide directed against the RNA by reverse transcription has also been utilised. This method can be combined with an RNase H directed cleavage of the transcript prior to extension (Streuli and Saito 1989). If alternative exons are cleaved in different positions, relative to their flanking constitutive exon, then the length of product can be used to assess alternative exon use. Alternatively, a chain terminating dideoxynucleotide can be incorporated in the extension reaction. The relative position of a particular nucleotide residue in the alternative transcripts will yield products of different length which can be separated by polyacrylamide gel electrophoresis. These methods, based on primer extension, would not be suitable for the analysis of individual exon use when the skipping/incorporation of several exons between two constitutive exons was involved. (Streuli and Saito 1989).

1.6.3 Northern and North-western blotting

RNA can be identified by separation on agarose gels according to its length. The RNA can then be transferred to a membrane support (blotted). The immobilised RNA can then be identified by hybridisation to either labelled DNA or RNA probe (Northern blotting) which contains a sequence complementary to the test RNA, or by specific binding of labelled RNA to a protein ligand (North-western blotting). Northern blotting has been used to identify alternative isoforms which share common sequences, as is noted in alternative splicing, but its limitation is that only isoforms with relatively large differences in size can be assigned. The analysis of the alternative splicing patterns exhibited by exons which are as short as three nucleotides (triplet AAG in N-CAM) would require other methods.

1.6.4 cDNA libraries

The reverse-transcription of total mRNA isolated from specific tissues, followed by second strand synthesis, produces cDNA which can be cloned into suitable vectors to produce tissue-specific cDNA libraries. Such libraries have been screened by the hybridisation of a specific probe to the inserted cDNA fragments. Clones which are positively identified can be characterised by standard sequence analysis. This method enables libraries (and hence RNA) from different tissues to be analysed and compared for the presence of isoform diversity produced by alternative splicing. This method is invaluable for the isolation of full-length cDNA encoding alternative isoforms, but is impractical for a detailed quantitative analysis of alternative isoform use, or for the analysis of transient expression experiments.

1.6.5 Reverse-transcription/PCR: cloning and sequencing

The development of the polymerase chain reaction (Saiki et al 1985) for the amplification of specific DNA sequences, when combined with a reverse-

transcription reaction (Erich 1989), and used in conjunction with a method in which large numbers (96) of clones can be sequenced (Eperon 1986), has produced a novel mechanism for the analysis of alternatively spliced RNA (and particularly of rare RNA isoforms) (Hamshire et al 1991; Graham et al 1992; Libri et al 1992; this thesis). Oligonucleotide primers directed against the flanking constitutive exons will amplify any sequence in between. By incorporating restriction endonuclease sites into the primer sequence, RT-PCR products can be cut, cloned into an m13 derived vector and sequenced. The development of such a method, for the analysis of endogeneously produced N-CAM, forms part of this thesis.

1.6.6 Reverse-transcription/PCR: direct analysis

The RT-PCR method described above provides semi-quantitative data on the use of alternative exons, but would be impracticable for the analysis of large numbers of different samples, and may not be quantitative for the analysis of RNA of different lengths. One of the aims of this project was to develop a quantitative method for the direct analysis of PCR products derived from combined RT-PCR of alternatively spliced RNA (Hamshire et al 1991; Graham et al 1992; Eperon and Hamshire 1993; this thesis).

Chapter 2

Materials and Methods

2.1 Stock solutions

10 x C

0.5M Tris-Cl pH 7.5

0.1M MgCl_2

10mM DTT

50 x Denhardt's reagent

1% (w/v) Ficoll

1% (w/v) polyvinylpyrrolidone

1% (w/v) BSA (DNase free)

DMEM

As powder from Gibco BRL

Add 2g/l NaHCO_3

50 x E buffer

2M Tris-acetate

50mM EDTA

E-loading dyes

6 x E buffer

30% (v/v) glycerol

0.25% (w/v) bromophenol blue (BPB)

0.25% (w/v) xylene cynol (XC)

E/G-loading dyes

2 x E buffer

30% (v/v) glycerol

0.25% (w/v) bromophenol blue (BPB)

0.25% (w/v) xylene cynol (XC)

F-loading dyes

90% (v/v) Formamide

50mM EDTA

0.025% (w/v) bromophenol blue (BPB)

0.025% (w/v) xylene cynol (XC)

GTE

50mM glucose

25mM Tris-Cl pH 8

10mM EDTA

2 x HBS

280mM NaCl

1.5mM Na₂HPO₄·2H₂O

5mM HEPES pH 7.1

(stored at -20°C)

ISO-B/NP40

10mM Tris pH 7.9

0.15M NaCl

1.5mM MgCl₂

0.65% NP-40

KOAc

29.4% (w/v) potassium acetate

11.5% (w/v) glacial acetic acid

NaOH/SDS

0.2M NaOH

1% (w/v) SDS

PBS

8g /l NaCl
0.2/l KCl
1.44g /l Na₂HPO₄
0.24g /l H₂PO₄
adjust pH to 7.4

10 x PCR200

0.5M KCl
0.1M Tris-Cl pH 8
15mM MgCl₂
2mM each dNTP

10 x PCR 0

0.5M KCl
0.1M Tris-Cl pH 8
15mM MgCl₂

10 x PCR2002.5

0.5M KCl
0.1M Tris-Cl pH 8
25mM MgCl₂
2mM each dNTP

10 x PCR2002.5+

0.5M KCl
0.1M Tris-Cl pH 8
25mM MgCl₂
2mM each dNTP
100 µg/ml BSA (DNase free)

PEG/NaCl

20% (w/v) polyethylene glycol (6000 grade)

2.5M NaCl

Pre-Hybridisation solution

6 x SSC

0.5% SDS

5 x Denhardt's reagent

100µg/ml sonicated salmon sperm

1 x PK buffer

0.1M Tris pH 7.5

12.5mM EDTA

0.15M NaCl

1% SDS

RNA elution buffer

0.5M NaOAc pH 4.8

1mM EDTA

0.2% (w/v) SDS

RPS

0.5M Tris pH 9.0

50mM EDTA

2.5% SDS

SB agar

1.0% (w/v) Bacto-tryptone

0.5% (w/v) yeast extract

0.5% (w/v) NaCl

2.0% (w/v) agar

10 x SEQ

0.1M Tris-Cl pH 7.5

50mM MgCl₂

20 x SSC

175.3 g/L NaCl

88.2 g/L sodium citrate

Adjust pH to 7.0 with NaOH

Top agar

1.0% (w/v) Bacto-tryptone

0.5% (w/v) yeast extract

0.5% (w/v) NaCl

0.7% (w/v) agar

TY broth

1.6% (w/v) Bacto-tryptone

1.0% (w/v) yeast extract

0.5% (w/v) NaCl

10 x TBE buffer

0.9M Tris-borate

20mM EDTA

TE.1

10mM Tris-Cl pH 7.5

0.1mM EDTA

TE/SDS

10mM Tris-Cl pH 7.5

0.1mM EDTA

0.5% (w/v) SDS

Trypsin (obtained as 1 x solution form Gibco-BRL)

0.5g/L Trypsin

0.2g/L EDTA

In Modified Puck's Saline A.

2.2 Cell culture

2.2.1 Preparation of collagenised flasks

20ml of a stock solution of 0.4mg/ml collagen in 150mM glacial acetic acid was mixed with 1ml of 3.6% sodium chloride immediately before use. 5ml of this solution was used for each 80ml Nunc tissue culture flask, or 2ml for a 9cm diameter plate. Flasks or plates were incubated at room temperature in a sealed bag (to prevent evaporation of glacial acetic acid), then washed with 4 changes of sterile distilled water. Collagenised plates were kept at 4 °C.

2.2.2 Growth conditions for myoblasts

C-2 myoblasts were maintained in 10% Foetal calf serum (FCS) (from Seralab) in DMEM containing 100 units/ml penicillin and 100µg/ml streptomycin. Cells were grown on collagenised Nunc tissue culture flasks, or collagenised plates at 37 °C, in an atmosphere of 5% CO₂ and high humidity. Cells were subcultured every 2 days by detachment with 1 ml trypsin-EDTA and incubation at 37 °C for 2 minutes, and subsequent dilution (usually 1/10) in fresh medium onto new flasks. Transfected cells were grown in media supplemented with 7mM sodium butyrate to induce expression of exogenous transcripts (Sambrooke et al 1989).

2.2.3 Induction of myogenesis

C-2 myoblasts were grown to confluence (3 days post passage) or near confluence (2 days post passage) in 10% FCS in DMEM containing 100units/ml penicillin and 100µg/ml streptomycin. The medium was then switched to 5% horse serum in DMEM containing 100units/ml penicillin and 100µg/ml streptomycin. The medium was changed daily until cells were harvested. Transfected cells were grown in media supplemented with 7mM sodium butyrate to induce expression of exogenous transcripts (Sambrooke et al 1989).

2.2.4 Transfection of cells

Cells were transfected by a modified calcium phosphate-DNA co-precipitation method (Graham and van der Eb, 1973). 5µg of vector DNA was diluted in 450µl TE.1 and 500µl of 2 x HBS. A DNA/CaPO₄ precipitate was formed by the addition of 50µl 2.5M CaCl₂ and incubation on ice for 15 minutes. The suspension was carefully added to the culture dishes by dropwise addition. Cells were incubated at 37 °C in 5% CO₂ for 6 hours, and then washed twice with warm PBS. Uptake of DNA was promoted by a 90 second glycerol shock (by the addition of 1ml of 20% glycerol in PBS followed by two washes with PBS). Fresh culture medium containing 7mM sodium butyrate was added to each flask; cells were harvested 20 hours later (as myoblasts) or induced into a myogenic pathway and harvested after 6 days (as myotubes).

It was found that transfections were only successful when fresh calcium chloride and HBS (stored at -20 °C) were used.

2.3 RT-PCR

2.3.1 *Isolation of RNA from cells*

RNA, for the analysis of N-CAM mRNA isoforms, was isolated by the use of LiCl/urea (Auffray and Rougeon 1980) followed by high speed centrifugation. Cells were removed from dishes or flasks by the addition of trypsin, and pelleted in sterile glass tubes. RNA was isolated by the addition of 4ml 3M LiCl/6M urea and incubation on ice for at least 2 hours, followed by centrifugation at 10,000 rpm for 30 minutes. The pellet was washed with cold LiCl/urea and resuspended in 300 μ l TE.1/0.5% SDS. Three chloroform/isoamyl alcohol extractions were performed and the RNA finally precipitated with ethanol/sodium acetate (20/1) and washed with 75% ethanol. Samples were stored, in ethanol, at -70 °C.

RNA, for the analysis of α -tropomyosin mRNA isoforms, was isolated by the use of ISO-B/NP40. In this method, RNA was isolated directly from cells which were still attached to plates. First, cells were washed twice with cold PBS. The plasma membranes were lysed by the addition of 700 μ l ISO-B/NP40 and incubation on ice for 5 minutes. The solution was transferred to an Eppendorf, vortexed for 1 minute and then centrifuged at 3000 \times g for 10 minutes to remove cells debris and nuclei. 400 μ l aliquots of supernatant were transferred to fresh eppendorf tubes and 100 μ l of RPS added. Protein was removed by two phenol/chloroform/iso-amyl alcohol extractions. RNA was precipitated by the addition of 2.5 volumes of ethanol/sodium acetate, and washed with 75% ethanol. Samples were stored, in ethanol, at -70 °C.

2.3.2 *Poly(A) selection of mRNA*

Poly(A)⁺ RNA was isolated from total RNA using either the spun columns available from Pharmacia, or with mAPTM paper from Amersham, following the protocol for small sample volumes. Both methods gave comparable results for isoform variation in adult mouse brain and adult muscle when analysed by RT-PCR. Subsequent poly(A)⁺ isolations were undertaken using mAP paper. Poly(A)⁺ enriched RNA was concentrated by ethanol precipitation.

2.3.3 *RT-PCR of endogenous RNA*

For the analysis of N-CAM isoforms, 25pmol of oligonucleotide MH2b was added to poly(A)⁺ RNA purified from 5ug total RNA in a final 1 x PCR200 buffer. The RNA was denatured at 90 °C for 5 minutes then cooled on ice. 20 units of RNAGuardTM was added and the primer allowed to anneal at 65 °C for 5 minutes. 10 units of MMuLV Reverse Transcriptase (Pharmacia) was added and the reaction incubated at 37 °C for one hour. The first strand cDNA was amplified by the addition of 20pmol MH1b2 oligonucleotide and a further 10pmol Mh2b, and 2 units of Taq polymerase in a final volume of 100µl in 1 x PCR200 buffer. Samples were overlaid with paraffin and amplified by 30 cycles of: 92 °C denaturation for 1.5 minutes, 65 °C annealing for 1.5 minutes, and 72 °C extension for 2 minutes. Final steps comprising annealing at 65 °C for 1.5 minutes and extension at 72 °C for 5 minutes to ensure that full length double stranded PCR products were formed.

For the analysis of endogenous α_s -tropomyosin RNA from myoblasts and myotubes the method developed for the analysis of exogeneous RNA was employed, except that a human sequence oligonucleotide (No1/8824) was used in place of the vector-specific primer RTPSVR.

2.3.4 RT-PCR of exogenous RNA

2 μ l of RNA was annealed to 10pmol (1 μ l of 10pmol/ μ l) of downstream primer (RT-PSVR or No 1/8824) in a buffer containing 1 μ l PCR2002.5+ and 6 μ l water; filter sterilised mineral oil was overlaid to prevent evaporation. The reaction was incubated at 95 °C for 5 minutes then snap-cooled on ice. After each subsequent addition, samples were centrifuged briefly to mix the reagents. Reverse transcription was performed by the addition of 5 μ l of a mixture containing: 5 units MuMLV reverse transcriptase (supplied by Pharmacia), 5 units of RNAGuard, 0.5 μ l PCR2002.5+, 4.5 μ l water followed by incubation at 37 °C for 60 minutes. Second strand synthesis and PCR amplification were performed by addition of 10 μ l of a mixture containing: 0.5 units of Taq polymerase (supplied by Cambio), 10pmol (1 μ l of 10pmol/ μ l) upstream primer (CAT5015+), 1 μ l PCR2002.5+, in 10 μ l water followed by a 1 minute touchdown PCR (Don et al 1991; section 2.13 for details).

2.3.5 Cloning PCR products

cDNA products were cut with restriction enzymes, corresponding to those engineered into the oligonucleotide primers, in standard reaction buffers and cloned into mICE 18 (Eperon 1986a) or pBlueScript. Ligations were transfected into competent JM109 (M13 vector) or HB101 (pBlueScript). 96 plaques from each M13 cDNA cloning were selected and grown in 250 μ l TY broth in microtitre plates with constant shaking for 9 hours. Single stranded template was prepared as described by Eperon (1986b), except that the cells were removed by centrifugation rather than filtration. 150 μ l of supernatant was transferred into a second microtitre plate. ssDNA was precipitated by the addition of 30 μ l PEG/NaCl. After 10 minutes the plates were centrifuged at 3,000 \times g, and the supernatant discarded. Pellets were resuspended in 50 μ l TE.1/SDS and incubated at 70 °C for 10 minutes. DNA was finally precipitated by the addition of ethanol/sodium acetate (20/1), washed with 70% ethanol

and resuspended in TE.1

2.3.6 Direct analysis of PCR products

An aliquot (typically 1/50th) of the initial RT-PCR was taken into a second PCR. 10pmol upstream and downstream primers (internal to those used in the RT-PCR) were added in a standard 25µl PCR reaction (see section 2.14). PCR products were labelled either by the incorporation of α -³²PdATP into the PCR products by the addition of 1-2 µCi to each 50ul reaction; or 10pmol of end-labelled primer mix (see section 2.16) was used as downstream PCR primer.

15 cycles of PCR were used for the analysis of RNA from C-2 myoblasts; 8 cycles, for the analysis of RNA from COS cells. Products were analysed by removal of 10µl of reaction mixture (from below the mineral oil) into a 1.5ml eppendorf tube containing 200µl TE.1. The PCR products were washed twice by the addition of 400µl diethyl ether, and then precipitated rapidly with ethanol (see section 2.5). Samples were resuspended in 20µl of either TE.1 or in One-Phor-All™ buffer (supplied by Pharmacia). Samples were analysed by separation on native polyacrylamide gels either directly, or following digestion with AluI (see section 2.3.7)

2.3.7 AluI digestion of PCR products

5ul of resuspended PCR product was incubated with either 15µl of One-Phor-All buffer (uncut sample), or 15µl of One-Phor-All which contained 1 unit of AluI (cut sample) for 90 minutes at 37 °C. A 5µl aliquot was mixed with 5µl E/G dye and an aliquot loaded onto a native polyacrylamide gel.

2.4 Sequence analysis

2.4.1 Use of M13 templates

2µl of each template (in batches of up to 96) was transferred into a microtitre plate and sequenced using 2',3'-dideoxynucleotide chain terminators (Sanger et al 1977) using the adaptations described in Eperon et al (1988). Primer concentrations of 2ng were employed, and usually 0.5µl of 0.25mM dNTP and 1.5µl of 0.25mM ddNTP were used, although the ratio dNTP/ddNTP was varied depending on the distance from the primer to the required sequence. The products were separated on 8% denaturing polyacrylamide gels.

2.4.2 PCR sequencing

15µCi of γ -³²PATP was used to end-label 32 pmol of oligonucleotide MH13. The reaction was performed in 1 x C and included 2 units of T4 polynucleotide kinase. The sample was incubated at 37 °C for 60 minutes. Four separate PCR reactions were set-up for each sequence by the addition of 2 pmol of labelled MH13 to 1µl PCR 0 buffer, 1.5µl of 0.25mM ddN, 0.5µl N^O (a mixture containing 0.25mM each dNTP except for the one corresponding to the chain terminating ddNTP, which was at 8.3µM), 0.3 units of Taq polymerase and 2µl of acrylamide gel purified PCR product, to a final 10µl with water and overlaid with mineral oil. 30 cycles of a PCR with denaturation temperature 92 °C for 1 minute, annealing temperature 59 °C for 1 minute and extension temperature of 72 °C for 1 minute, was then performed. PCR products were analysed by separation on a 8% denaturing polyacrylamide gel.

2.4.3 *Double-stranded sequencing*

9µl of a typical plasmid preparation, which had been isolated using a rapid isolation method (Jones and Schofield 1990), was alkali-denatured and then annealed to 10pmol sequence-specific primer prior to rapid ethanol precipitation and resuspension in sequencing buffer (Jones and Schofield 1990). Standard Sequenase reactions (but omitting the annealing step) were then performed. The products were then separated on 8% denaturing polyacrylamide gels.

2.5 Ethanol precipitation

Ethanol precipitation of DNA or RNA was typically performed by addition of 10% (v/v) 5M sodium acetate and then 2.5 volumes of absolute ethanol. RNA was normally precipitated by incubation at -70 °C, DNA was precipitated by incubation on ice, or at room temperature (for labelled PCR products) for 5 minutes. RNA or DNA pellets were washed with 75% ethanol and air dried prior to resuspension in TE.1 or reaction buffer.

2.6 Polyacrylamide gel electrophoresis

8% denaturing polyacrylamide gels were made by dissolving 42g urea in 20ml of 40% acrylamide (w/v) and 10ml 10 × TBE to a final volume of 100ml. Samples were denatured in the presence of an equal volume of F-dyes by incubation at 80C for 5-8 minutes. Denaturing gels were typically run at constant voltage of 1200-1500v.

10% native gels contained 25ml/100ml of 40% acrylamide in a final 1 × TBE buffer. Samples were loaded onto the gel after the addition of an equal volume of E/G loading dyes. Native gels were typically run at a constant voltage of 300v overnight.

2.7 Preparation of plasmid DNA

Plasmids required for the transformation of myoblasts were produced by an alkaline lysis method. Transformed colonies were picked individually into 20ml TY broth containing ampicillin and grown with shaking at 37 °C for 6-8 hours. Plasmid DNA was then amplified by the addition of chloramphenicol. Cells were pelleted by gentle centrifugation, then resuspended in 2ml GTE by incubation on ice for 5 minutes. Cells were lysed by the addition of 4ml NaOH/SDS and incubation on ice for 5 minutes. 3ml of KOAc was added, the sample shaken well and left on ice for 10 minutes before centrifugation at 3800 x g for 15 minutes in a cooled rotor. Plasmid from the resultant supernatant was precipitated by the addition of an equal volume of propan-2-ol. RNA was removed from the plasmid preparation by incubation of resuspended pellet (in 0.4ml TE.1) with 100ug Ribonuclease A (Pharmacia) for 15 minutes at 37 °C. Proteins were removed from the plasmid preparation by the addition of SDS to a final 0.5% and incubation at 70 °C for 10 minutes, followed by three extractions with phenol/chloroform/iso-amyl alcohol and precipitation with ethanol.

Plasmids required for sequence analysis were produced by the rapid method described by Jones and Schofield (1990).

2.8 Isolation of genomic DNA

5×10^3 C-2 myoblast cells were pelleted and washed with 1 x PBS. 500 μ l of PK buffer was added to the pellet with 150 μ g proteinase K and incubated overnight at 55 °C. Three phenol extractions and one phenol/chloroform/iso-amyl alcohol extraction were performed. The genomic DNA was precipitated by addition of ethanol/sodium acetate and washed with 75% ethanol. DNA was resuspended in TE.1 to a final concentration of approximately 1 μ g/ μ l.

2.9 PCR amplification of genomic DNA

Approximately 0.01 μ g of purified genomic DNA was amplified using 10pmol each of upstream and downstream oligonucleotide primer in 25 μ l of 1 x PCR2002 buffer with 1 unit Taq polymerase (Cambio). A touchdown protocol adapted from Don et al (1991) was used for the amplification program (given in section 2.13). Both the touchdown 3' and touchdown 1' programs were used.

2.10 Southern blotting and hybridisation

Following agarose gel electrophoresis (typically with a gel containing 1% agarose in E buffer), the gel was denatured by soaking in 0.25M HCl for 20 minutes, followed by neutralisation in 1.5M NaCl / 0.5M NaOH for 35 minutes and two washes with 3M NaCl / 0.5M Tris (pH7-8). DNA was transferred onto Hybond N using 10 x SSC by overnight blotting. DNA was cross-linked with the filter by a 20 second exposure to U.V. using a transilluminator. Filters were pre-hybridised in pre-hybridisation solution at 68 °C for 2-4 hours. Labelled oligonucleotide probe was hybridised to the DNA by overnight incubation at room temperature, in pre-hybridisation solution containing 0.01M EDTA. Excess and non-specifically bound probe was removed by three washing steps: once with 2 x SSC / 0.5% SDS for 15 minutes at room temperature, once with 2 x SSC / 0.1% SDS for 15 minutes at room temperature, and twice with 0.1 x SSC / 0.5% SDS for 15 minutes at 42 °C. Filters were then exposed to X-ray film, typically for 3 days using pre-flashed film and an intensifying screen.

2.11 End-labelling of oligo-nucleotides

600 pmols of oligonucleotide (E6-7) were labelled at the 5'end by the addition of 5 units of T4 polynucleotide kinase in the presence of 100 μ Ci γ -³²PATP or γ -³³PATP in a total of 20 μ l in 1 x C buffer. The reaction was incubated at 37 °C for 60 minutes.

2.12 Oligodeoxynucleotide sequences (5' to 3')

MH1 (*human N-CAM, exon 12*)

TGTACGATCAGAATTCAGGGCTGGGTGAGATCAGCGCGGCC

MH2 (*human N-CAM, exon 13*)

CCATGTACTGAAGCTTGAGGAGAGCGCTCGGTACCTGACC

MH1b2 (*mouse N-CAM, exon 12*)

CCCCCGCCCCGAATTCCTGAGTTCAAGACACAG

MH2b (*mouse N-CAM, exon 13*)

GCCGGCGCGGAAGCTTTCTGCCCTTCCAGCTTGGGT

MH13a (*mouse N-CAM, exon MSD1a (downstream)*)

CGCCCGCCGCCTCGAGACTTACGTGGAGGAGGGCTAT

MH14a (*mouse N-CAM, exon MSD1a (upstream)*)

CGCCCGCCGCAGATCTATAGCCCTCCTCCACGTAAGT

MH15b (*mouse N-CAM, exon MSD1b (downstream)*)

CGCCCGCCGGGACTTACTAGCTCGAGGAGACAA

MH16b (*mouse N-CAM, exon MSD1b (upstream)*)

CGCCCGCCGCAGCTCAGAGGCTCCTGCTAACTCTTC

MH17c (*mouse N-CAM, exon MSD1c (downstream)*)

CGCCCGCCGCCATCGATACTTACTAGACAGGTCTGTGGTT

MH18c (*mouse N-CAM, exon MSD1c (upstream)*)

CAACCTGGCCTCTTCCTGCTT

MHSECd (*human N-CAM, exon SEC (downstream)*)

CCTTCTTGTTCTGCRITGATGG

MH ex13 (*mouse N-CAM, exon 13*)

CTTGGGTGCACTGGGT

RT-PSVR (*SV40, vector specific primer*)

CACTGCATTCTAGTTGTGGT

No1/8824 (*human tropomyosin, exon 6*)

TTCCAGGTCATCGCGGCCGCTTTCCAGCTTGG

E6/7 (*human tropomyosin, exon 6*)

ACGGGTCTCTGCCTCCTT

CAT5O15 (*human tropomyosin, exon 4*)

GTGTGCACCTCCAAGCTTCCGTCGACTTGGTCACCCTTGG

Exon 4 PCR(*human tropomyosin, exon 4*)

GAGGAACGAGCTGAGCT

2.13 Touchdown PCR program

Touchdown 1'

(94 for 1 minute; 70 for 1 minute; 72 for 1 minute) x 2;
(94 for 1 minute; 68 for 1 minute; 72 for 1 minute) x 2;
(94 for 1 minute; 66 for 1 minute; 72 for 1 minute) x 2;
(94 for 1 minute; 64 for 1 minute; 72 for 1 minute) x 2;
(94 for 1 minute; 62 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 60 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 58 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 56 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 54 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 52 for 1 minute; 72 for 1 minute) x 2;
(90 for 1 minute; 50 for 1 minute; 72 for 1 minute) x 26.

Touchdown 3'

As above, except that extension times at 72 C were extended to 3 minutes.

2.14 Standard "second-round" PCR reaction.

1/50th of initial RT-PCR reaction

10pmol upstream primer

10pmol downstream primer (usually end-labelled)

2.5µl PCR 2002.5+

0.5 units Taq polymerase

to a final 25µl with water, and overlaid with mineral oil.

8 or 15 cycles comprising: 94 °C for 1 minute

55 °C for 1 minute

72 °C for 1 minute

then one cycle of 72 °C for 5 minutes.

2.15 Human α_s -tropomyosin mini-gene constructs

Costruction of the α_s -tropomyosin mini-gene constructs is given elsewhere (Graham 1992) but briefly the mini-gene was produced by replacing a portion (corresponding to exons 4-6 inclusive) of a skeletal form of cDNA with genomic sequence from exons 4 to 6 (thereby including the three introns which flank the two alternatively spliced exons NM and SK). This full length cDNA was cloned into an expression vector which was under the control of the SV40 early promoter, and contained an SV40 polyadenylation signal. Mutations to this mini-gene were made via oligonucleotide-directed (Zoller and Smith 1983) or "sticky-feet"-directed (Clackson and Winter 1989) mutagenesis of a fragment cloned into an m13 derived vector. A cartoon of the mutations to the mini-gene is given in Figure 2.1. Sequences of the wild-type and mutant constructs are given in Appendix 1.

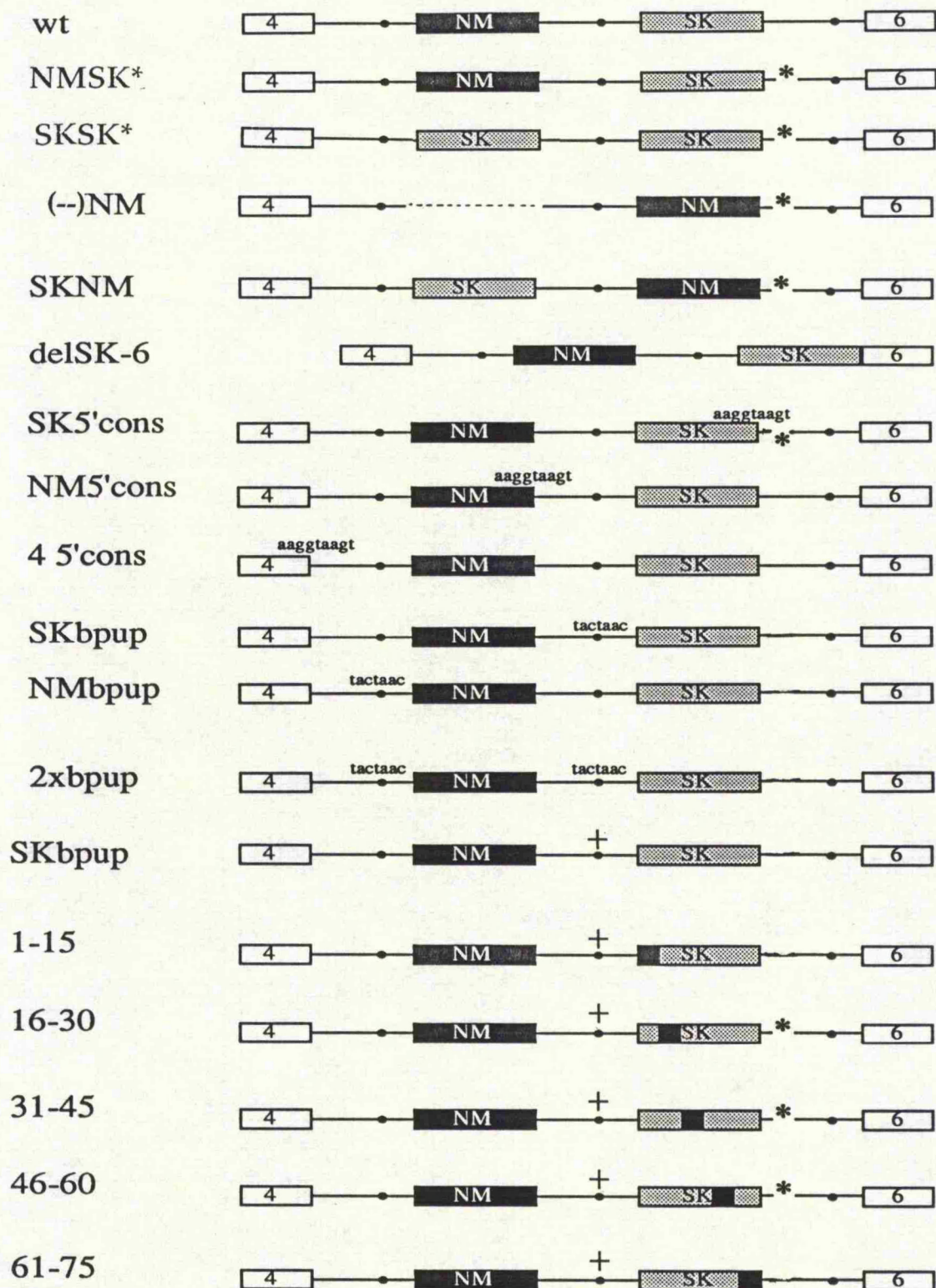


Figure 2.1.a

The nature of the mutations used for transient transfection of COS cells, C-2 myoblasts and C-2 myotubes. Open boxes represent the constitutive exons 4 and 6, dark shaded boxes the alternatively spliced NM exon, and pale shaded boxes the alternatively spliced SK exon. The position of the stem deletion is indicated by an asterisk (*). + indicates those mutations which include SKbpup.

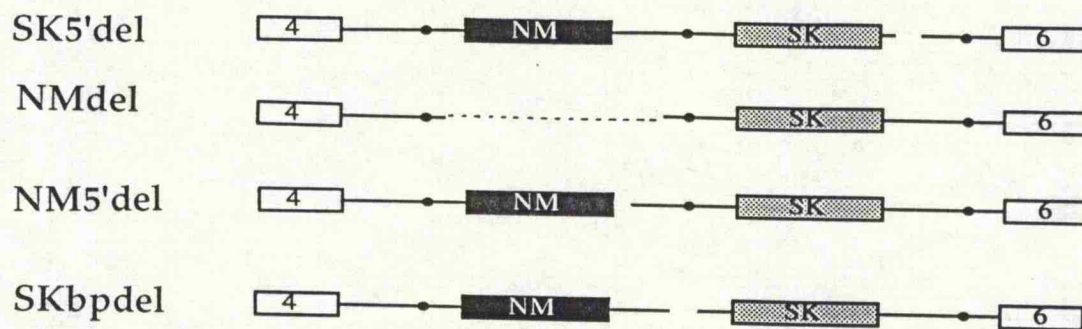


Figure 2.1.b
The nature of the mutations used for transient transfection of COS cells, C-2 myoblasts and C-2 myotubes. Open boxes represent the constitutive exons 4 and 6, dark shaded boxes the alternatively spliced NM exon, and pale shaded boxes the alternatively spliced SK exon. The position of the stem deletion is indicated by an asterisc (*).

Chapter 3

Alternative splicing in mouse N-CAM

3.1 Cloning of the mouse homologue to human MSD region

A region of N-CAM, that was alternatively spliced in muscle (Dickson et al 1987), contained four or five individual exons (Thompson et al 1989). Only one of these, and a putative sequence for a triplet exon had been identified in the mouse (Santoni et al 1989). A combined RT-PCR was developed to identify and characterize the previously unreported exons of the mouse MSD.

3.1.1 *Development of a combined RT-PCR method.*

The region of murine N-CAM equivalent to that between exons 12 and 13 (containing the possible MSD homologue) was to be amplified from cDNA by a combined RT-PCR (Figure 3.1.1a). Initial attempts, using classical methods for the production of initial cDNA (Maniatis et al 1982) followed by amplification of second strand cDNA by PCR, resulted in the cloning of non-N-CAM sequences (as determined by a search of the EMBL data-base) and were probably the result of non-specific reverse transcription and amplification (data not shown). A variety of reasons for inappropriate amplification were likely; N-CAM RNA was present at low concentrations, the constituents of the reaction buffers were affecting the efficiency of reverse-transcription and amplification, the concentration of magnesium in the buffer was allowing inaccurate annealing of the primers, or the annealing temperature of the primers was too low.

In order to concentrate the N-CAM mRNA, poly(A)⁺ selected RNA was used. This was made using Pharmacia spun-columns which use oligo dT beads to separate RNA containing tracts of adenosine residues, notably those with poly(A)⁺ tails.

The reaction conditions were substantially modified such that the same buffer was used in both the reverse-transcription step and the PCR, and the second strand cDNA synthesis was done in the first cycle of the PCR. A series of test

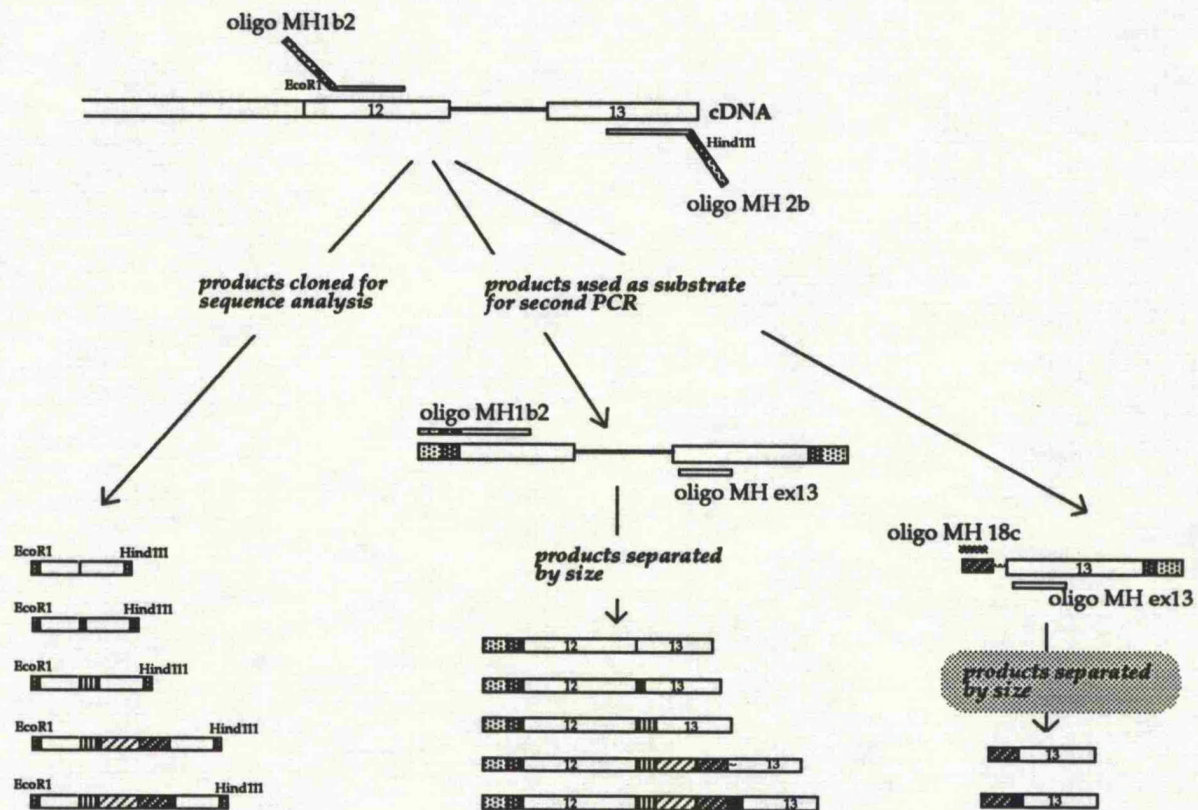


Figure 3.1.1a
 Strategies for the analysis of N-CAM splicing between exons 12 and 13. Oligonucleotide sequences are represented as clear for regions of sequence complementarity; solid boxes for mismatched restriction sites; and shaded boxes for flanking GC-rich "clamps". Exon sequences are shown as clear for constitutive exons 12 and 13, vertical hatching for MSD1a, pale hatching for MSD1b, dark hatching for MSD1c and solid boxes for the triplet AAG.

amplifications of the human cDNAs indicated the optimum conditions for PCR amplification. The resulting buffer used was based on that of Erlich (1989) although a variety of nucleotide concentrations, magnesium concentrations and pH were tested (data not shown).

The theoretical optimum annealing temperature for the oligonucleotides used can be calculated in a variety of ways. Two general formulae are given below:

$$T_m = [69.3 + 41(\%G+C/100) + 14\log I] - 400/L \quad (1)$$

where I = ionic strength of the reaction

L = number of complementary bases

$$T_m = 4(G+C) + 2(A+T) \quad (2)$$

For oligonucleotides MH1b2 and MH2b (sequences given in section 2.16) two annealing temperatures can be calculated. First, the temperature of the early cycles when only the N-CAM specific sequence has homology to its target DNA (T_m (early)), and second, the temperatures later in the reaction when the sequences of the restriction site and the spacer sequences are now also complementary to the amplification products of previous cycles (T_m late). The theoretical annealing temperatures for amplification of the human cDNAs are given in Table 3.1.1.

Table 3.1.1

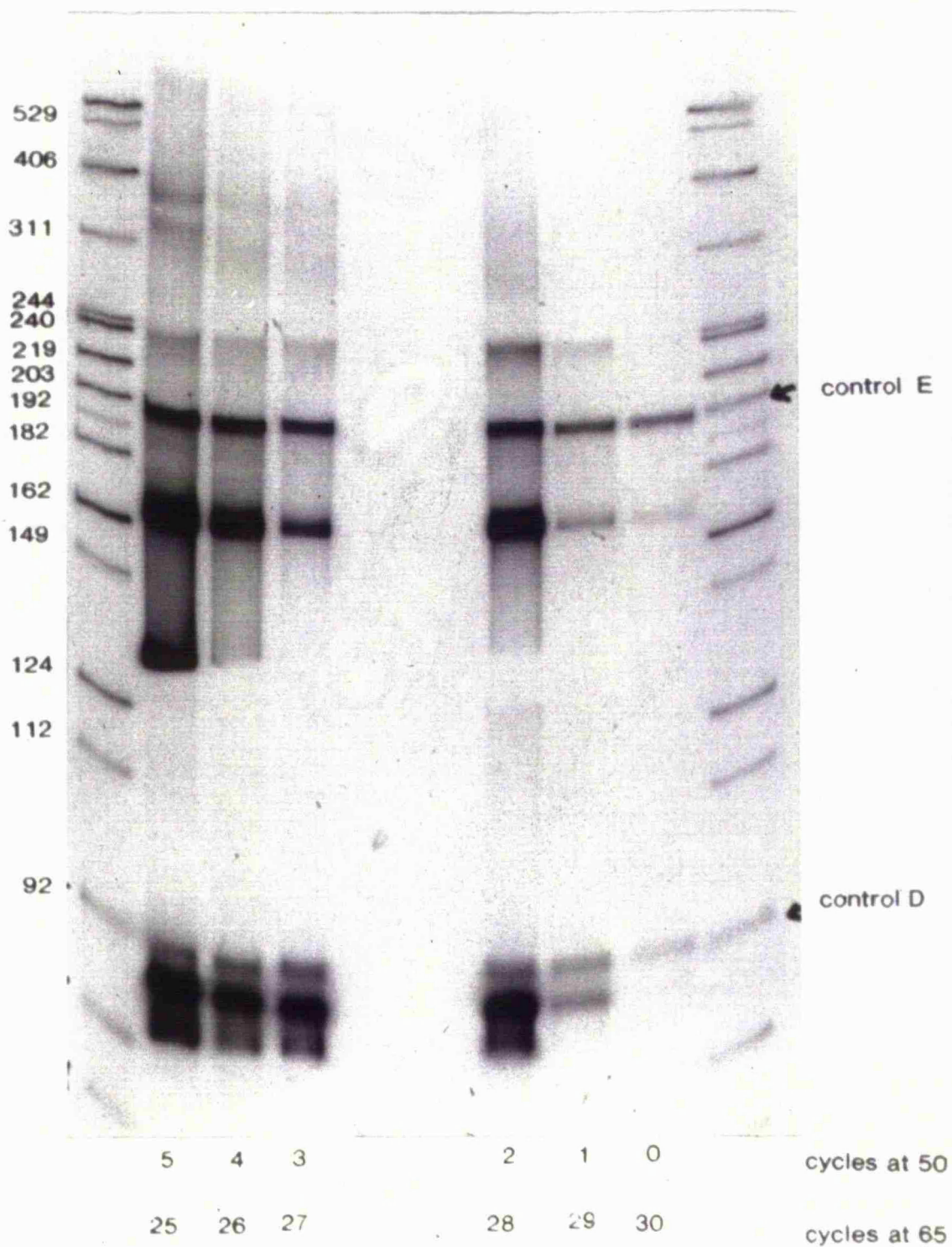
	Oligonucleotide	T _m (early)	T _m (late)
equation (1)	MH1b2	55.9	69.4
	MH2b	55.9	69.4
equation (2)	MH1b2	60.0	104.0
	MH2b	60.0	104.0

Table 3.1.1 The theoretical annealing temperatures for formation of stable DNA/DNA duplex between the oligonucleotides MH1b2 and MH2b and their target sequence in a human cDNA template.

Clearly the simplified equation (2) is not appropriate for the calculation of annealing temperatures for oligonucleotides longer than about 20 bases. These theoretical temperatures were tested empirically by the amplification of human test cDNA in a two step PCR in which the number of cycles at the higher temperature was increased at the expense of numbers of cycles at the lower temperature. The results are shown in Figure 3.1.1b Contrary to predictions, the oligonucleotides could anneal at 65 °C even though the calculations predicted that this should only have been possible in later cycles, after production of longer amplified product. These results led to a modification of the RT-PCR protocol to incorporate annealing temperatures 65 °C throughout.

3.1.1b The effect of increasing the annealing temperature during a PCR. Two human cDNAs (control D contained the isoform 12-13, control E the isoform 12-MSD1a-MSD1b-MSD1c-AAG-13) were co-amplified in a series of PCRs in which the number of cycles at the higher (65) annealing temperature were increased at the expense of the number of cycles at the lower (55) annealing temperature. A total of 30 cycles were used in each case. PCR products were labelled by the universal incorporation of α -³²PdATP, and separated on a 8% denaturing polyacrylamide gel.

EFFECT OF INCREASING ANNEALING TEMPERATURE DURING PCR



3.1.2. Myoblasts and myotubes as a source of RNA

A source for RNA from non-muscle and muscle was required, and for two reasons the mouse cell culture line C2 was chosen. First, as skeletal muscle contains more than one cell type (nerve, lymph, blood-vessel etc.) the direct comparison of the splicing patterns from this, with a non-muscle source (brain, kidney etc) may not reflect the true difference between individual cell types because the heterogeneous nature of these tissues: the difference may result from a combination of splicing patterns from different cell types within the complex tissue. By choosing to study cells in culture, a direct comparison of any observed switch in splicing patterns of N-CAM in myoblasts and myotubes could then be attributed to a regulation mechanism operating within this single cell type. Secondly, C-2 myoblasts are relatively easy to grow and can be induced to the terminally differentiated myogenic pathway. Cells which were still capable of division (myoblasts) were chosen as a source of non-muscle cells, and cells which had been induced to differentiate (myotubes) as a source of muscle cells. The conditions for cell culture are given in Chapter 2 but it is pertinent to add that cells that had undergone few divisions, were growing rapidly, and were supplied with fresh medium daily would fuse and form "twitching" myotubes more readily than older and slower growing cells. Cells also required an adherent surface in order to grow well, have an appropriate morphology, and fuse. This was provided by a layer of collagen. The RNA was isolated from the cells as described elsewhere.

3.1.3. Results of cloning and sequencing.

The RT-PCR products were cloned and the sequences determined initially by ddTTP screening for 78 clones, 34 of which were derived from myoblast RNA (Figure 3.1.3a) and 44 from myotube RNA (Figure 3.1.3b). An example of each of the different sequences from myoblasts (Figure 3.1.3c) and myotubes (Figure 3.1.3d) were then analysed by four-track sequencing. Ten of the clones incorporated all of the exons homologous to the human MSD1a, MSD1b, MSD1c and triplet AAG exons. The other cDNA clones contained various combinations of these exons, and are illustrated in Figure 3.1.3e.

The exon use in myoblasts and myotubes can be compared and subjected to statistical analysis. In order to prevent a bias in the comparison because of the problems associated with the analysis of small numbers only the major forms are compared. Table 3.1.3 gives the numbers of each clone produced in myoblasts and myotubes.

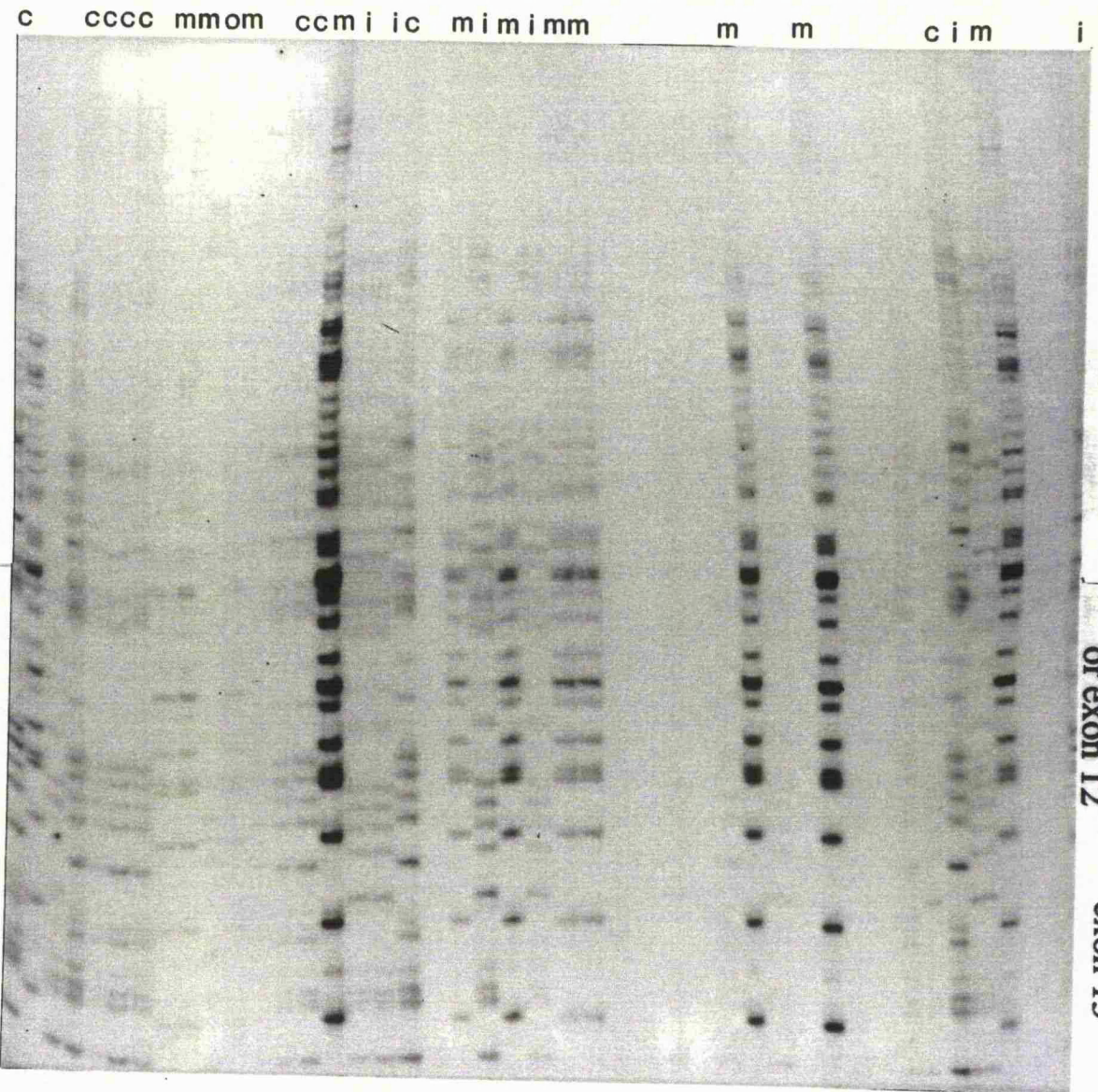
Table 3.1.3

Isoform	myoblasts	myotubes
12-13	15	0
12-AAG-13	14	23
12-a-b-c-13	2	9
12-a-b-c-AAG-13	0	10

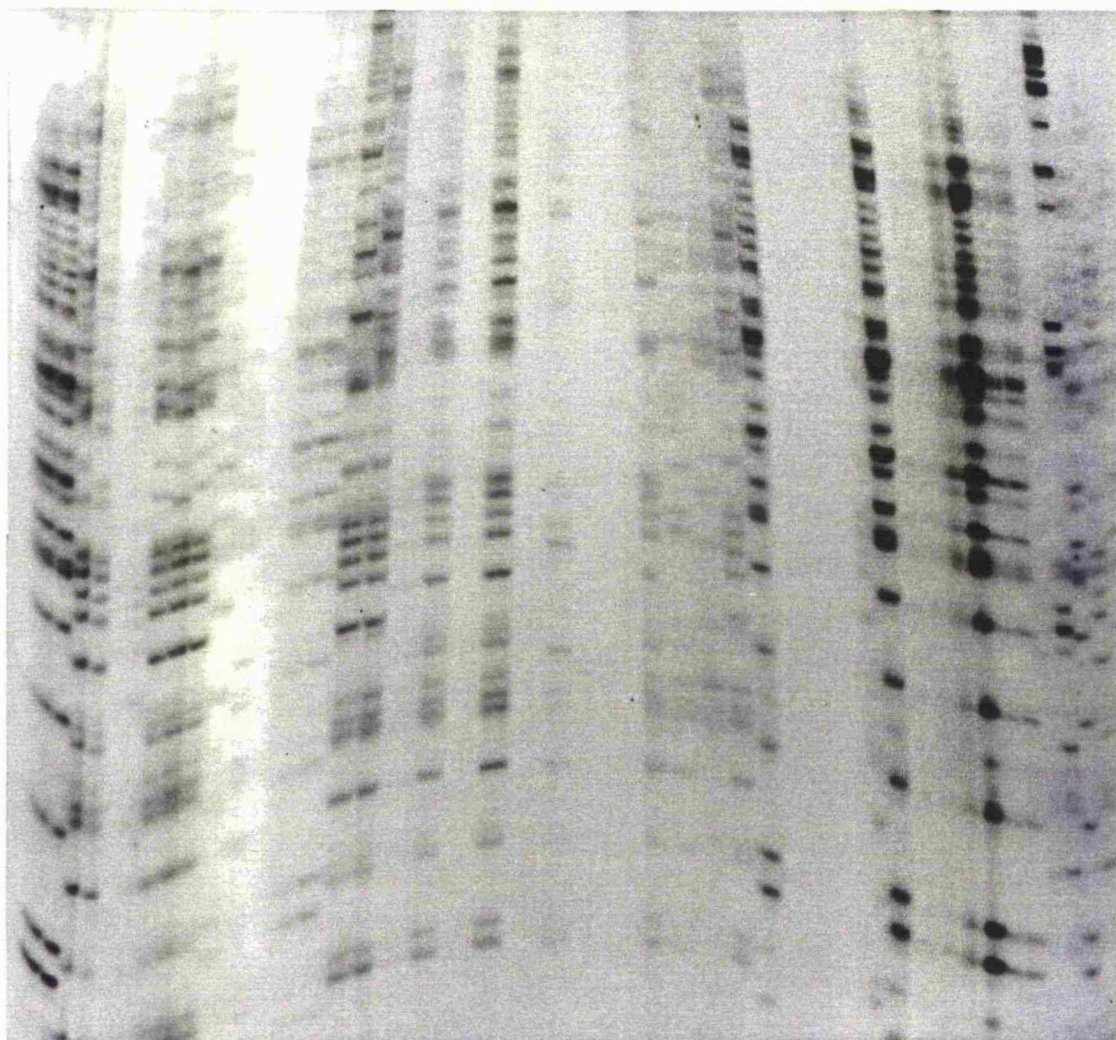
Table 3.1.3. The distribution of major isoform use in myoblasts and myotubes. Statistical analysis reveals that the distribution is significantly non-random ($\chi^2 = 30.7$).

3.1.3a Initial ddTTP sequencing of 96 m13 clones derived from the ligation of the RT-PCR product from myoblast mRNA into EcoRI/HindIII cut mICE 18 (Eperon 1986). MH ex13 oligonucleotide was used as the sequencing primer. Seven different patterns were identified, with 14 examples of pattern C; 15 examples of pattern I; 1 example of pattern J; 1 example of pattern K; 1 example of pattern L; 24 examples of pattern M (wild-type mICE 18) and 2 examples of pattern O.

MSD exons,
or exon 12 exon 13



mmll - - lllmJ - mmllCC - C - K - - CLOlm - - Cm - - mmmm - AmIm

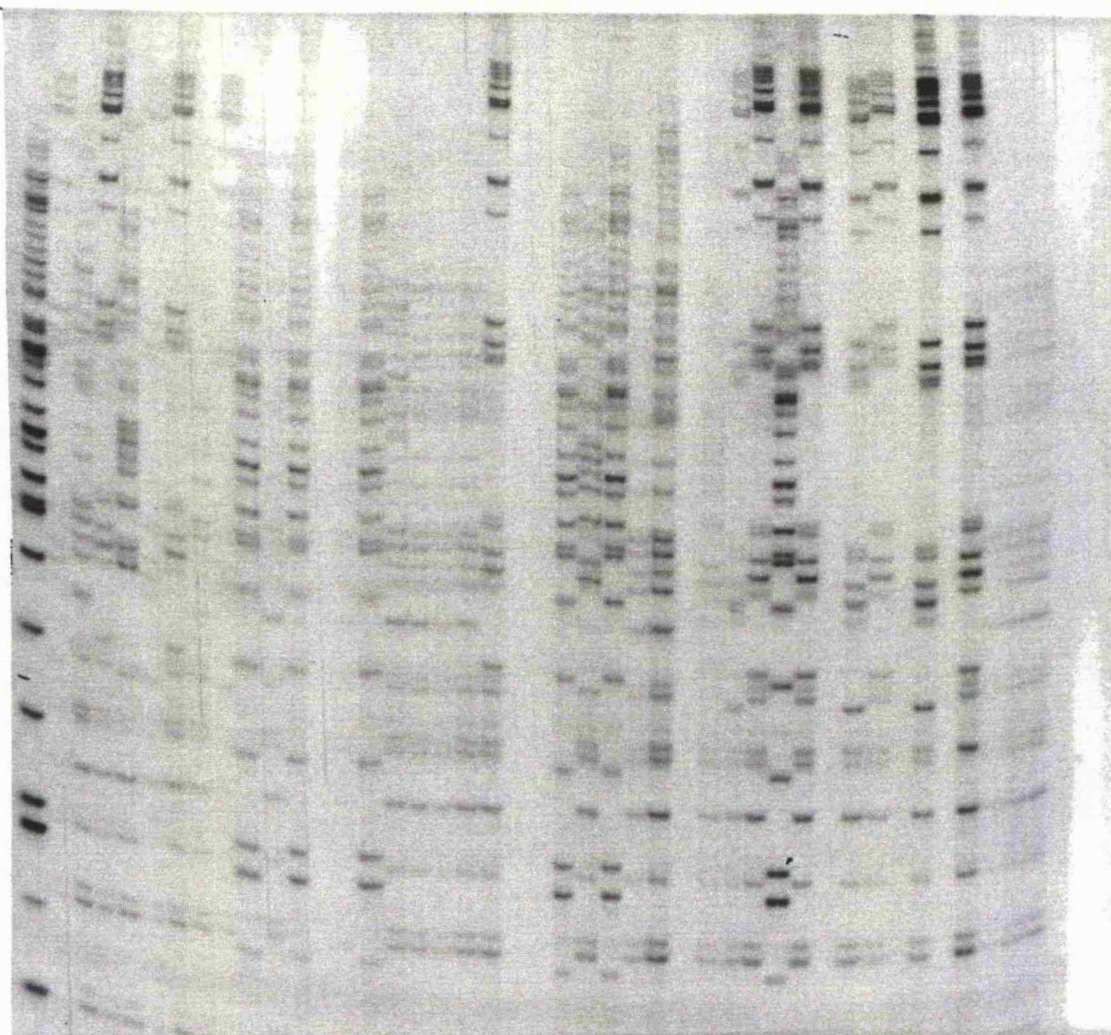


MSD exons,
or exon 12

exon 13

3.1.3b Initial ddTTP sequencing of 96 m13 clones derived from the ligation of the RT-PCR product from myotube mRNA into EcoRI/HindIII cut mICE 18 (Eperon 1986). MH ex13 oligonucleotide was used as the sequencing primer. Seven different sequences were initially identified, with 9 examples of pattern A; 10 examples of pattern B; 20 examples of pattern C; 1 example of pattern D; 1 example of pattern E; 1 example of pattern F; and 1 example of pattern G. Patterns F and G were later shown to be the same.

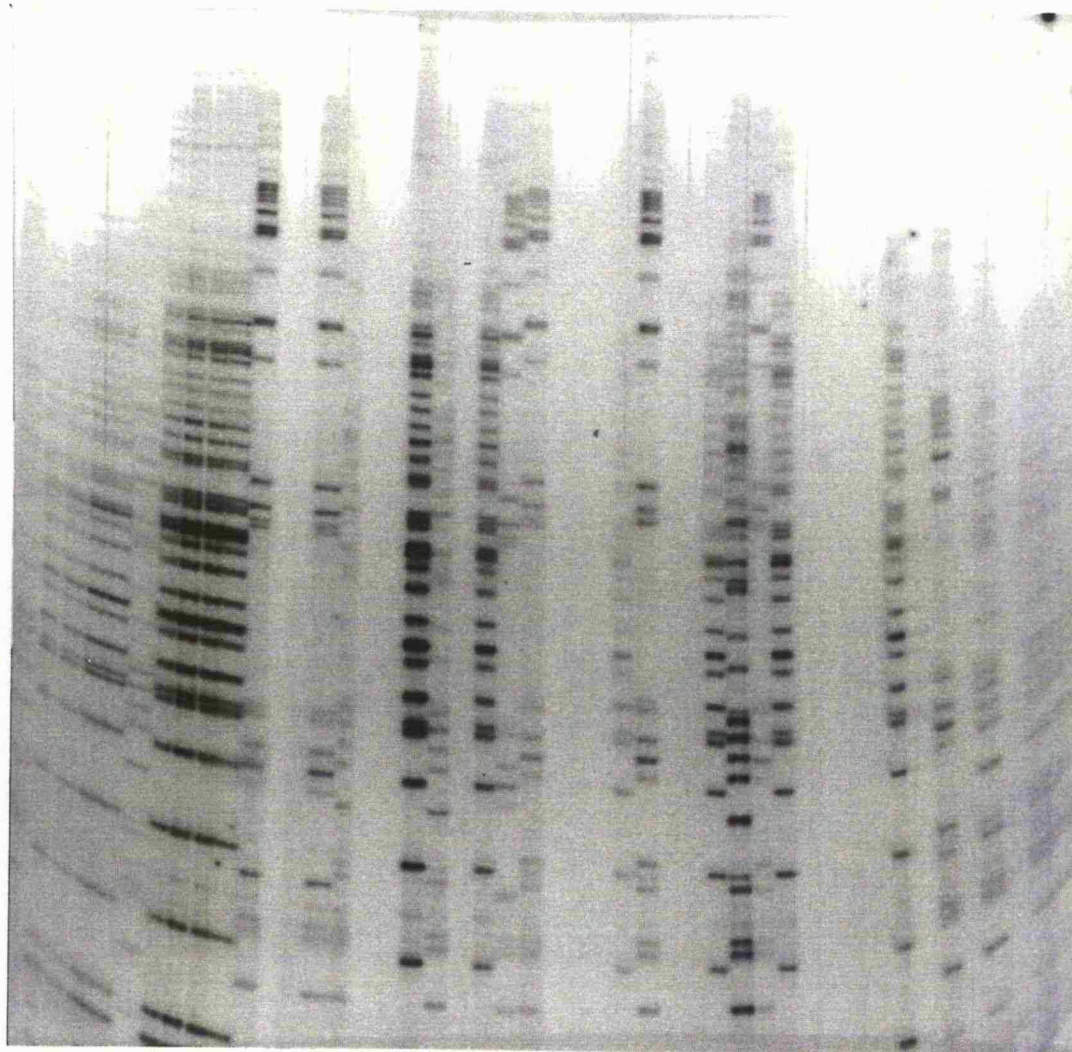
C m A C B E A B F A m G m - C m C C C C B - - m F m C C - H A B m B - A B - A - B C C C



MSD exons,
or exon 12

exon 13

C m m m m C m m m m A - - A C - - m C - B A B - - - B - - m C B m - - - m - D - C - C C



MSD exons,
or exon 12 exon 13

3.1.3c The dideoxynucleotide sequences of the alternative N-CAM isoforms. These isoforms were initially identified by ddTTP screening of m13 clones derived from myoblast RNA, using MHex13 as the sequencing primer. The sequence of each pattern is given below. The bases shown in bold type are those of the alternatively spliced exons of the MSD between the constitutive exons 12 and 13.

Pattern C: 12-AAG-13:

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/**AAG**/GGGAACCCAGTGCACCC

Pattern I: 12-13

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/GGGAACCCAGTGCACCCAA

Pattern J: 12-MSD1a-AAG-13

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/**ATAGCCCTCCTCCAC/AA**G/GGGAACCCAGTGCACCCAA

Pattern K: 12-MSD1a-MSD1c-13

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/**ATAGCCCTCCTCCAC/CA**ACCTGGCCTCTTCCTGTCCTTCCAACACAGACCTGTCTA/GGGAACCCAGTGCACCCAA

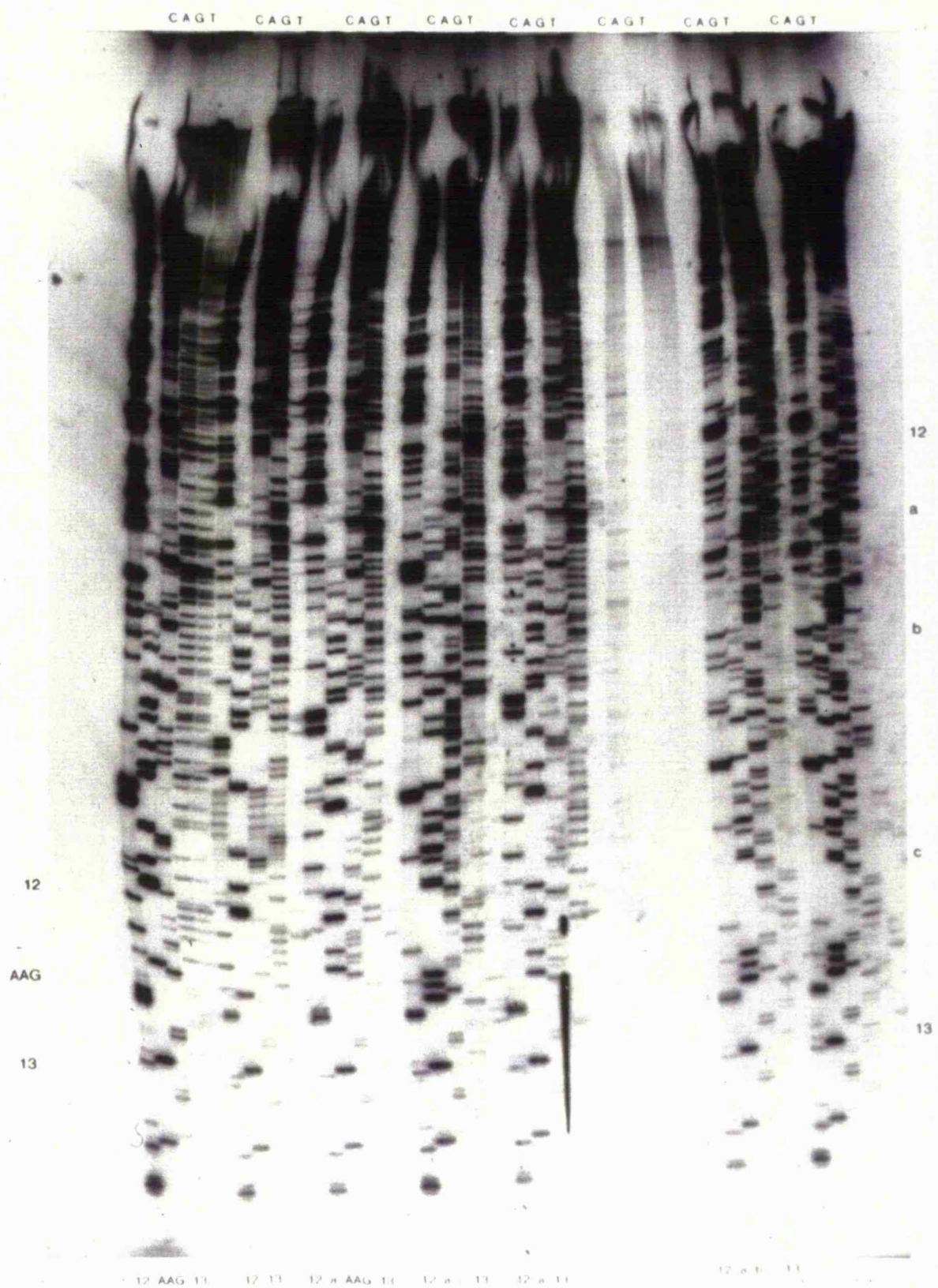
Pattern L: 12-MSD1a-13

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/**ATAGCCCTCCTCCAC/GG**GAACCCAGTGCACCCAA

Patterns A and O: 12-MSD1a-MSD1b-MSD1c-13

GAATTCCCACTGAGTTCAAGACACAGCCAGTCC/**ATAGCCCTCCTCCAC/AG**GCTCCTGCTAACTCTCACCCCTGTTCCGTTGTCTCCTCGAGCTA/CAACCTGGCCTCTTCCTGTCCTTCCAACACAGACCTGTCTA/GGGAACCCAGTGCACCCAA
A

SEQUENCES OF cDNAs PRODUCED FROM MYOBLAST mRNAs



3.1.3d The dideoxynucleotide sequences of the alternative N-CAM isoforms, initially identified by ddTTP screening of m13 clones derived from myotube RNA, using MHex13 as the sequencing primer. The sequence of each pattern is given below. The bases shown in bold type are those of the alternatively spliced exons of the MSD between the constitutive exons 12 and 13.

Pattern H: 12-MSD1a-13

GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/ATAGCCCTCCTCCAC/GG
GAACCCAGTGCACCCAA

Pattern D: 12-MSD1c-AAG-13

GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/CAACCTGGCCTCTTCCTG
TCCTTCCAACCAACAGACCTGTCTA/AAG/GGGAACCCAGTGCACCCAA

Patterns G and F: 12-AAG-13

GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/AAG/GGGAACCCAGTGC
ACCC

Pattern A: 12-MSD1a-MSD1b-MSD1c-13

GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/ATAGCCCTCCTCCAC/AG
GCTCCTGCTAACTCTCACCCGTGTTCCGTTGTCTCCTCGAGCTA/CAACCTGGC
CTCTTCCTGTCTTCCAACCAACAGACCTGTCTA/GGGAACCCAGTGCACCCA
A

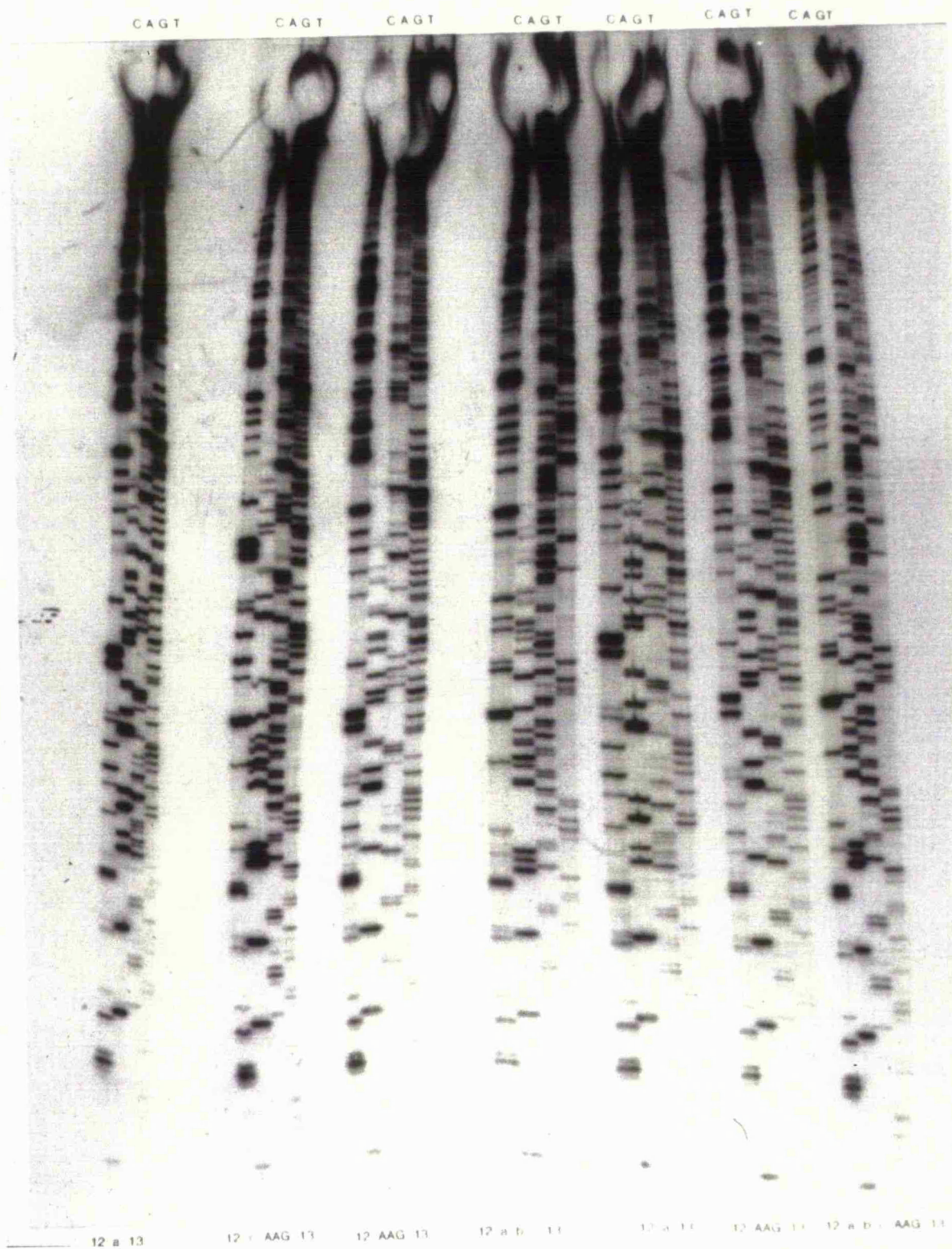
Pattern E: 12-MSD1a-13

GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/ATAGCCCTCCTCCAC/GG
GAACCCAGTGCACCCAA

Pattern B: 12-MSD1a-MSD1b-MSD1c-AAG-13

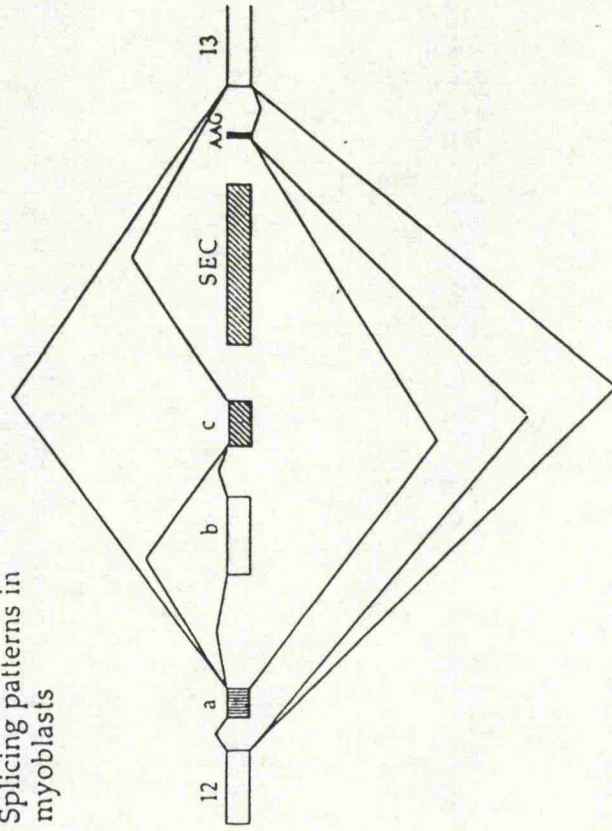
GAATTCCCCTGAGTTCAAGACACAGCCAGTCC/ATAGCCCTCCTCCAC/AG
GCTCCTGCTAACTCTCACCCGTGTTCCGTTGTCTCCTCGAGCTA/CAACCTGGC
CTCTTCCTGTCTTCCAACCAACAGACCTGTCTA/AAG/GGGAACCCAGTGCA
CCCAA

SEQUENCES OF cDNAs PRODUCED FROM MYOTUBE mRNAs

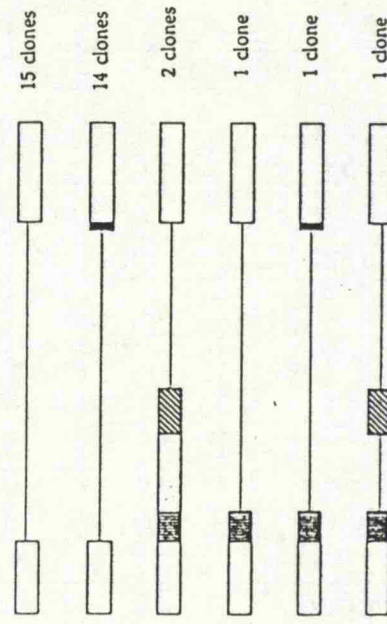


3.1.3e Patterns of exon usage in C-2 myoblasts and C-2 myotubes, assigned by the number of clones isolated and sequenced for each isoform. Exon sequences are shown as clear for the constitutive exons 12 and 13, vertical hatching for MSD1a, pale hatching for MSD1b, dark hatching for MSD1c and solid boxes for the triplet exon AAG. The putative splicing patterns required to produce these combinations are also shown.

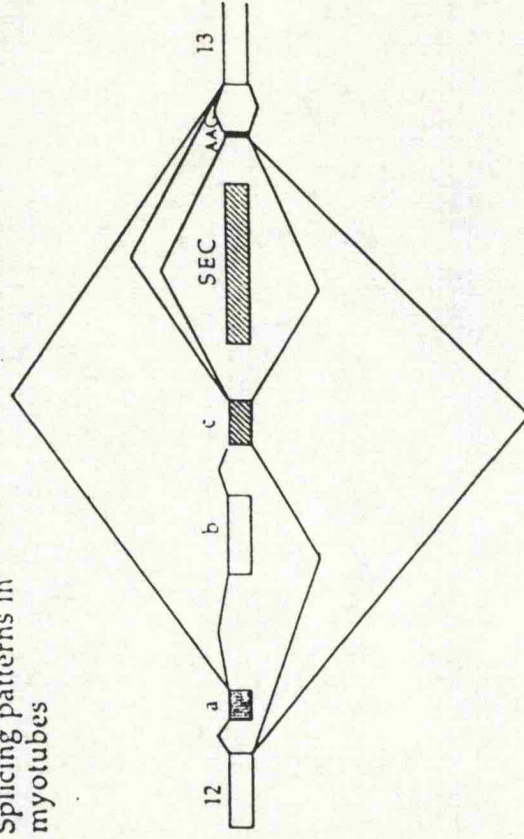
Splicing patterns in myoblasts



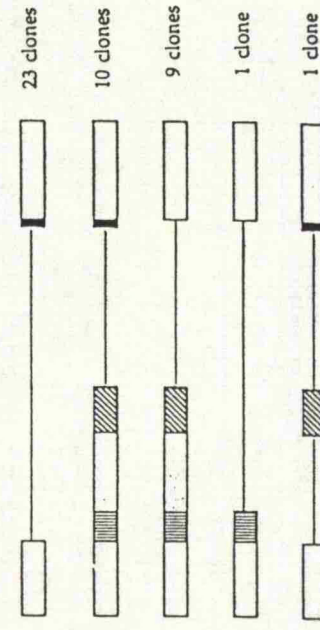
cDNA patterns



Splicing patterns in myotubes



cDNA patterns



Thus, significant differences were noted for the exon use in undifferentiated myoblasts and differentiated myotubes. The less abundant clones also included two unreported combinations of exons between the constitutive exons 12 and 13, in the region homologous to human MSD (12-a-c-13, 12-c-AAG-13)

These data provided the first evidence for mouse homologues to the human MSD exons MSD1b and MSD1c. The homologue to MSD1a and the triplet AAG exon had already been identified. Figure 3.1.3f shows the sequence of the mouse homologues to MSD1a, MSD1b, MSD1c and AAG flanked by regions of the constitutive exons 12 and 13, aligned with the MSD exons identified in the human and chicken.

Analysis of the sequence similarity indicates a high level of identity for all of the exons with 100% identity to MSD1a, 77% to MSD1b and 83% to MSD1c. The 11 base-pair differences between mouse 1b and human MSD1b are not clumped within the exon nor do they significantly favour third-base substitutions. Of the 7 base-pair differences between mouse 1c and human MSD1c four are within the last 8 bases. This distribution is not random; its significance will be discussed later.

3.1.3f The cDNA sequence of the mouse exons homologous to human MSD1a, MSD1b MSD1c and the triplet AAG. The sequence is aligned with the human, mouse (MSD1a) and chicken sequences already elucidated.

EXON 12	..G	TTC	AAG	ACA	CAG	CCA	GTC	C
mouse	..G	TTC	AAG	ACG	CAG	CCA	GTC	C
human	ACA	CAG	CCA	GTT	C
chicken
EXON MSD1a	AT	AGC	CCT	CCT	CCA	C
mouse	AT	AGC	CCT	CCT	CCA	C
human	GT	ATC	CCT	CAC	TCA	C
chicken
EXON MSD1b	AG	GCT	CCT	GCT	AAC	TCT	TCC	ACC
mouse	CG	GCA	TCT	GCT	AGC	TCG	TCT	ACC
human	CA	AGT	GCT
chicken
EXON MSD1c	CA	ACC	TGG	CCT	CTT	CCT	GTC	CTT
mouse	CA	ACT	TGG	CCT	CTT	CCT	GCC	CTT
human	CA	ACT	CAG	CCT	CTC	CCT	GCA	TCC
chicken
EXON AAG	AA	G
mouse	AA	G
human	AA	G
chicken
EXON 13	GG	GAA	CCC	AGT	GCA	CCC	AAG	C..
mouse	GG	GAA	CCC	AGT	GCA	CCT	AAG	C..
human	GG	GAA	CCC	AGT
chicken	GG	GAA	CCC	AGT

3.2 Direct analysis of RT-PCR products

The data obtained from the cloning and sequencing strategy for the identification of alternative isoforms indicated that there was tissue-specific alternative splicing of the MSD exons and that rare and previously uncharacterised isoforms may also be present. This led to extending the characterisation of the alternative patterns of expression to determine a more complete picture of the alternative splicing patterns exhibited at different stages of development and in different tissues.

However, the cloning and sequencing strategy employed for initial characterisation of myoblasts and myotubes would have been unsuitable for two reasons. Firstly, there were too many samples for this method to be practical. Secondly, and more importantly, this method contained certain inherent problems in the quantification of the data: there was no guarantee that the ratios of different mRNA isoforms would be maintained during the RT-PCR and the cloning strategy. Analysis of the inherent error of the RT-PCR could be investigated, but an analysis of the efficiencies of cloning of different length products would require a very large-scale experiment (in the order of 1000 sequencing reactions).

For these reasons an alternative method was derived. This method was based on direct analysis of RT-PCR products. No cloning or sequencing of the products was required once initial experiments had been undertaken.

3.2.1 *Poly(A)⁺ selected RNA*

As stated earlier, it appeared that the concentration of N-CAM RNA was low in myoblasts and myotubes and that the selection of poly(A)⁺ RNA effectively concentrated the mRNA. The method used for poly(A)⁺ selection, in the initial investigations, used spun-columns from Pharmacia. The cost of these columns

was prohibitive for the number of RNA samples that I wished to analyse. I therefore tested two other methods: oligo dT beads and the commercially available mAP paper. Total RNA from adult mouse muscle and adult mouse brain were poly(A)⁺ selected by all three methods. Although these were only preliminary tests, it was clear that there is no appreciable difference in the final results obtained. The method based on small scale preparation of poly(A)⁺ by the mAP paper was chosen as the most practical for analysis of large numbers of samples.

3.2.2 Using a nested PCR primer

Initial analysis of the products of RT-PCR, by their re-amplification in the presence of α -³²PdATP and subsequent separation on polyacrylamide gels, resulted in non-specific background amplifications which made interpretation of the gels difficult. To circumvent this problem, the second round PCR was re-designed such that only an internal portion of the first RT-PCR product was amplified. This use of a "nested primer" was achieved by directing one of the primers of the second round PCR to a sequence 10 bases 5' to the site used by primer MH2b. The upstream primer was the same as that used in the first RT-PCR. The use of the nested primer reduced the non-specific background amplifications and acted as a sequence-specific probe. Thus, only DNA substrates which contained a contiguous 30 bases of exon 13 sequence were amplified.

3.2.3 Quantification of PCR

It was important to validate the PCR amplification reaction, both because of the potential for gross distortions of the proportions of isoforms and because I wanted to use a second amplification reaction, with the products of the first amplification as substrates in order to visualise the ratios of the products directly.

In order to test whether the proportions of substrate cDNA were maintained during amplification, independent of the starting concentrations of the substrate, two cDNA sequences were amplified for 15 cycles in a series of reactions where their relative starting concentrations were varied systematically. A third substrate (cDNA 12-a-13) was included at a constant concentration as a control. The resultant PCR products were separated on a native polyacrylamide gel. The incorporation of α - $^{32}\text{PdATP}$ during the PCR enabled quantification of the different length products by laser densitometry of an autoradiograph. The result is shown in Figure 3.2.3a. The data from the densitometer are shown below in Table 3.2.3a.

3.2.3a cDNAs were co-amplified in a series of PCRs to assess the linearity of PCRs. A constant concentration of isoform 12-a-13 was incorporated in all reactions, the amount of 12-a-b-c-13 and 12-13 cDNAs (in fmols) was varied. The values are given in the table below.

cDNA	Track number									
	1	2	3	4	5	6	7	8	9	10
[12-13] fmol	242	233	220	196	147	98.0	49.0	24.5	12.25	2.45
[12-a-b-c-13] fmol	3.95	19.8	39.5	79.0	158	237	316	355	375	391

QUANTIFICATION OF PCR PRODUCTS

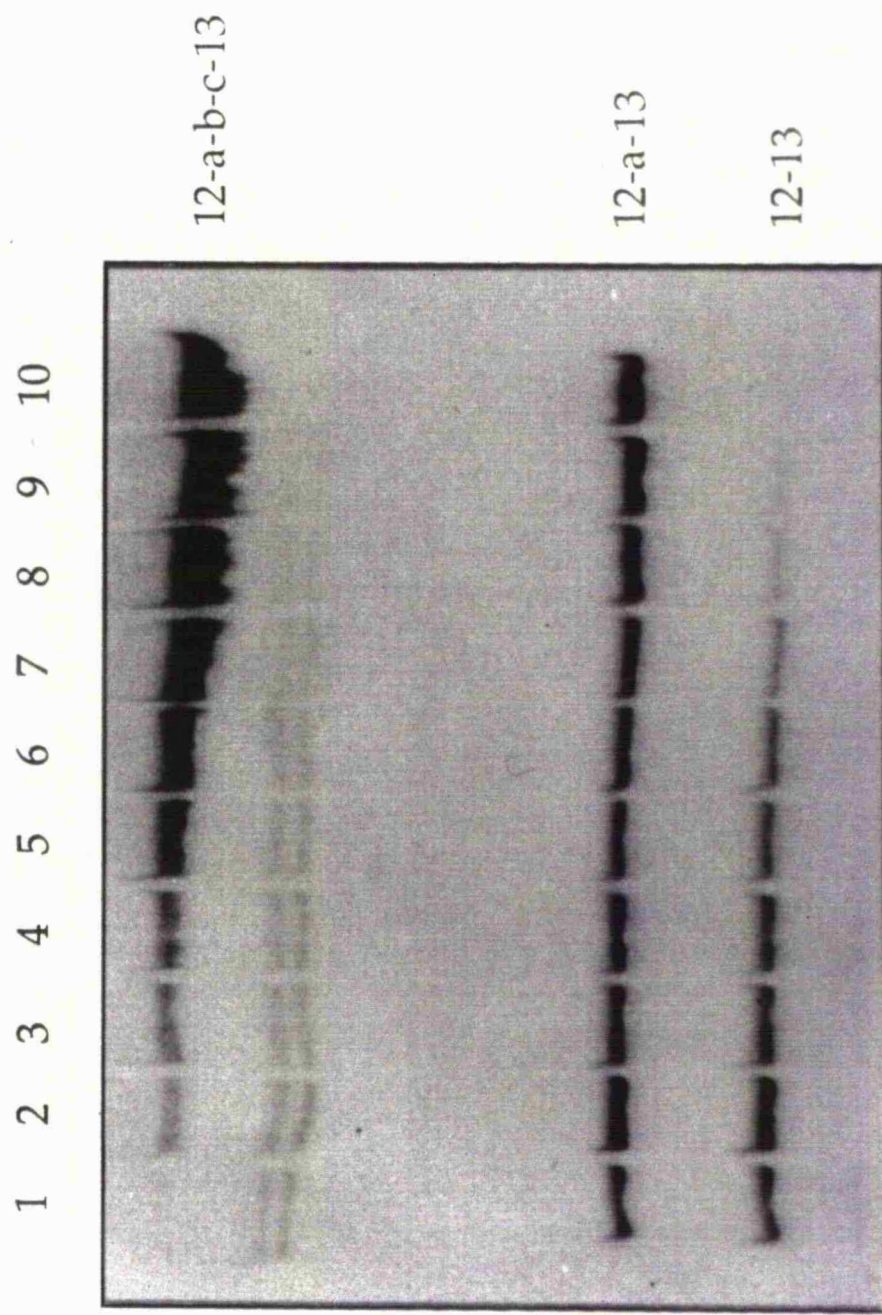


Table 3.2.3a

Track No.	Isoform		
	12-13	12-a-13	12-a-b-c-13
1	273	291	0
2	450	491	126
3	296	362	154
4	203	333	295
5	164	304	503
6	166	365	755
7	120	449	1104
8	37	503	1268
9	25	469	887
10	11	622	1781

Table 3.2.3a. The autoradiograph shown in Figure 3.2.3a was scanned by laser densitometry. The absorbance value for each band (and hence each isoform) is given in arbitrary absorbance units. The initial substrate concentration, for each track, is given in Figure 3.2.3a.

However, corrections must be applied to this data; first, to account for the number of A+T residues (as α - ^{32}P dATP was used to label the products), and, second, to correct for loading variation by reference to the absorbance obtained from cDNA 12-a-13 which was at the same concentration in all the reactions.

The corrected absorbance was calculated by the following equations:

$$\text{Corrected absorbance (Abs.cor)} = \frac{(\text{Abs} \times C)}{\text{number of A+T residues}}$$

$$\text{where } C = \frac{\text{average absorbance reading for 12-a-13}}{\text{actual absorbance reading for 12-a-13}}$$

The corrected data is given in Table 3.2.3b. shown below.

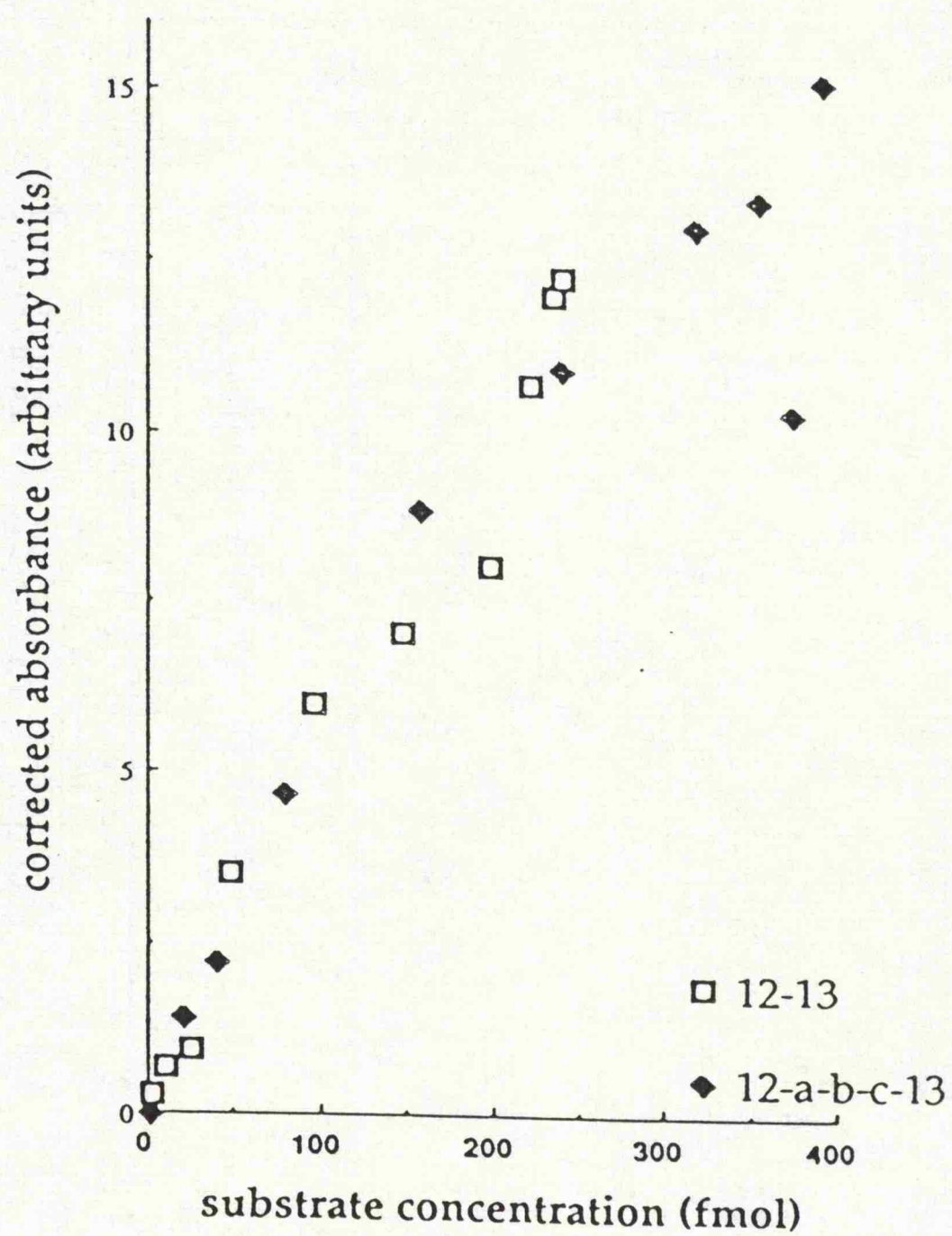
Table 3.2.3b

Track No.	12-13		12-a-b-c-13	
	fmol	Abs.cor	fmol	Abs.cor
1	242	12.3	3.95	0
2	233	12.0	19.8	1.36
3	220	10.7	39.5	2.25
4	196	8.00	79.0	4.70
5	147	7.05	158	8.77
6	98.0	5.95	237	11.0
7	49	3.5	316	13.0
8	24.5	0.96	355	13.4
9	12.25	0.70	375	10.3
10	2.45	0.23	391	15.2

Table 3.2.3b. The absorbance readings for the relative intensities of the bands shown in Figure 3.2.3a, and given in Table 3.2.3a were corrected, using equation 1, to account for total A + T residues and for loading differences by reference to the internal standard isoform 12-a-13. The corrected absorbance reading (Abs_{cor}) is given for each concentration of substrate DNA.

Figure 3.2.3b shows the graph of product yield (corrected absorbance) versus substrate concentration. For both substrates the relationship was linear and the substrates were amplified with equivalent efficiencies. The product yields obtained with these control cDNA were much higher (in the order of 50 fold) than those obtained when test RT-PCR cDNA was used as substrate for the labelled second round PCR. From these data I concluded that the second amplification step would maintain the input ratios of cDNA isoforms, and inferred that as the starting concentrations were very low in the first round PCR,

Figure 3.2.3b. Graph showing the corrected absorbance reading plotted against the input concentrations of cDNAs 12-13 (open boxes) and 12-a-b-c-13 (black diamonds).



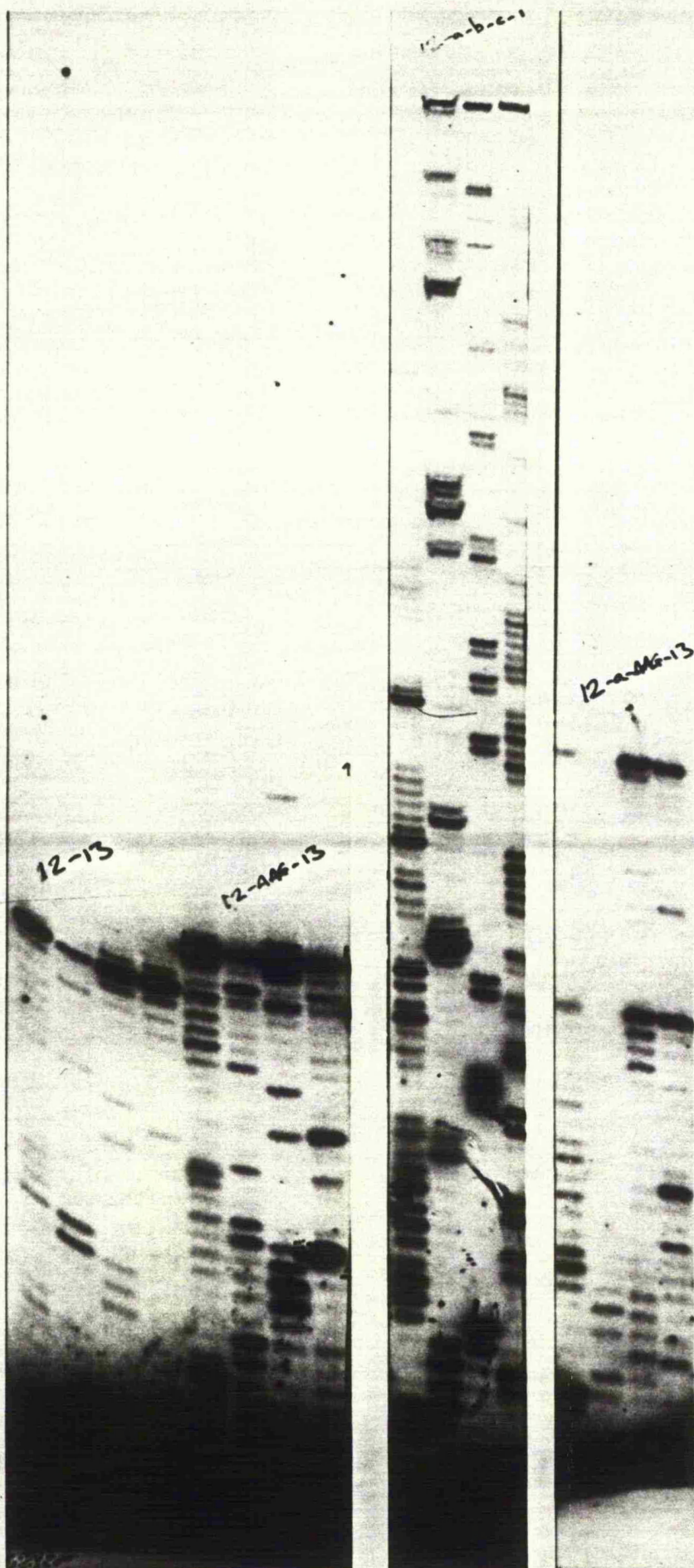
and that even after 30 cycles of amplification the products were scarce (they could not be seen on agarose gels), that the first round PCR also maintained the isoform ratios. However, these experiments only established that the PCR was quantitative; it was possible still that the isoform ratios could have been perturbed by the reverse transcription step. This was thought unlikely, although the possibility for sequence specific (and hence isoform specific) inhibition of reverse transcription (perhaps via secondary structure formation) could not be discounted. Recent investigations have indicated that reverse transcription does maintain isoform ratios (Graham 1992)

3.2.4 Identification of the bands.

I had identified and sequenced the homologues to the human MSD exons in the mouse and found them to be of identical length. All the possible combinations of exons would give a different length products which could be predicted from the knowledge of oligonucleotide priming site, and the length of each exon. However, although an internal primer was used for the second round PCR I thought it important to verify the exact nature of each of the bands separated on the gel. It was possible that different, and as yet unidentified, exons could also produce PCR products of identical lengths. The verification of the bands was achieved by modified PCR amplifications of excised bands. The results are shown in Figure 3.2.4a. Although the sequences were difficult to interpret, they were informative and could verify the nature of each band. No new exon sequences were identified using this method.

3.2.4a PCR sequencing to verify the bands. PCR products, labelled and separated on a native polyacrylamide gel, were cut out of the gel, and used as substrates for a PCR based sequence analysis using MH ex13 as the sequencing primer. Band 1 is the sequence derived from a band migrating at 63bp (the expected length of PCR product of the 12-13 isoform); band 2, from a band migrating at 66bp (12-AAG-13); band 3, from a band migrating at 168 (12-a-b-c-13); and band 4, from a band migrating at 81 (12-a-AAG-13).

band 1 band 2 band 3 band 4
C A G T C A G T C A G T C A G T



3.3 Exon usage in developing muscle

Three developmental progressions were studied in order to provide a comprehensive analysis of the changing pattern of expression of the exons of the MSD during muscle development. First, the development of progenitor muscle cells in culture (C2 myoblasts, mid-fusion and myotubes) was analysed; although these may not represent the exact patterns of expression of similar cells within muscle because they lack the complex interactions with surrounding milieu (notably neuromuscular junctions). Secondly, I analysed development of muscle cells from different stages of development of laboratory mice (embryonic (E16), newborn (3 day post-natal) and adult (10 week)). In the last series I analysed the effect of removing nerve/muscle interaction in adult mice by the surgical denervation of the hind-limb (by surgical section of the sciatic nerve).

3.3.1 *Developmental profile of muscle.*

The results of direct identification of RT-PCR products are shown in Figure 3.3.1a. Eight cDNA clones, derived from N-CAM mRNA from C-2 myoblasts and myotubes as previously described, were co-amplified as markers: C included the variant forms 12-13, 12-a-13, 12-a-c-13, 12-a-b-c-13; CA included 12-AAG-13, 12-a-AAG-13, 12-c-AAG-13, 12-a-b-c-AAG-13. Two bands were observed (at 69bp and 149bp) which did not correlate with any of the cDNAs used as size markers. These bands were due either to previously uncharacterised patterns of splicing or the inclusion of novel exon sequence, or were gel artefacts. As detailed in a later section these bands were shown to be due to aberrant gel migrations. These rogue bands could have been derived by either of two mechanisms: the partial denaturation of the RT-PCR product leading to aberrant migration on a native gel, or the production of single-stranded DNA (ssDNA) during the PCR reaction. The production of ssDNA had not been seen in the amplification of cDNA

markers, derived from a mixture of cloned mouse MSD isoforms, in which much higher concentrations of final product were made using identical PCR conditions. The production of ssDNA was therefore thought to be the less likely explanation for the aberrant migration. The rogue bands were identified by PCR sequencing following their elution from the gel (see section 3.2.4), they were then assigned to the appropriate category and included in the calculations as such. Recently, heteroduplex formation during the last cycle of a PCR reaction has been proposed as a reason for aberrant band migration on native polyacrylamide gels (Zorn and Kreig 1991). It is possible that this could have produced the rogue bands that were observed. However, as already stated, these bands were not seen with the marker cDNA amplifications which were normally co-amplified at much higher initial concentrations, and would therefore be expected to form heteroduplex if the PCR conditions had been favourable.

Laser densitometry of the autoradiograph shown in Figure 3.3.1 led to the quantification of the cDNA bands. The intensities had to be corrected to account for the number of A+T residues because of the universal incorporation of radio-labelled dATP in the PCR amplification. The quantification of exon AAG use in conjunction with the exons of the MSD was undertaken by performing a separate second PCR with two internal oligonucleotides, directed at exon c and exon 13. PCR products produced in this reaction were short enough for the inclusion of the triplet AAG exon to be determined on a native polyacrylamide gel. Figure 3.3.1b illustrates the inclusion of the triplet AAG exon in combination with exon c of the MSD. The percentage isoform use derived from the corrected intensities are shown in Tables 3.3.1a and 3.3.1b.

Table 3.3.1a

Isoform	C-2 myoblasts	C-2 mid-fusion	C-2 myotubes
12-13	43%	65%	11%
12-AAG-13	57%	31%	57%
12-a-13	<0.5%	3%	2%
12-a-AAG-13	<0.5%	1%	3%
12-a-b-c-13	<0.5%	<0.5%	22%
12-a-b-c-AAG-13	<0.5%	<0.5%	5%

Table 3.3.1a. The alternative isoforms of the MSD region of N-CAM, which were expressed in C-2 cells at different stages of development, were quantified by laser densitometry of the autoradiograph shown in Figures 3.3.1a and 3.3.1b. The initial absorbance readings were corrected for number of A+T residues as described earlier. The corrected absorbance values (in arbitrary units) are given.

Table 3.3.1b

Isoform	Embryonic muscle	Newborn muscle	Adult muscle
12-13	18%	12%	55%
12-AAG-13	39%	25%	13%
12-a-13	<0.5%	<0.5%	<0.5%
12-a-AAG-13	<0.5%	<0.5%	<0.5%
12-a-b-c-13	31%	60%	31%
12-a-b-c-AAG-13	12%	3%	1%

Table 3.3.1b. The alternative isoforms of the MSD region of N-CAM, which were expressed in Balb c mouse skeletal muscle at different stages of development, were quantified by laser densitometry of the autoradiograph shown in Figures 3.3.1a and 3.3.1b. The initial absorbance readings were corrected for number of A+T residues as described earlier. The corrected absorbance values (in arbitrary units) are given.

If the profile from myoblasts, mid-fusion cells and myotubes is taken as one developmental progression, and from embryonic, newborn and adult as another, clear shifts in exon usage can be seen. These are illustrated in Figure 3.3.1c in which the % isoform use in each cell/tissue type is compared.

Several conclusions can be made from these data:

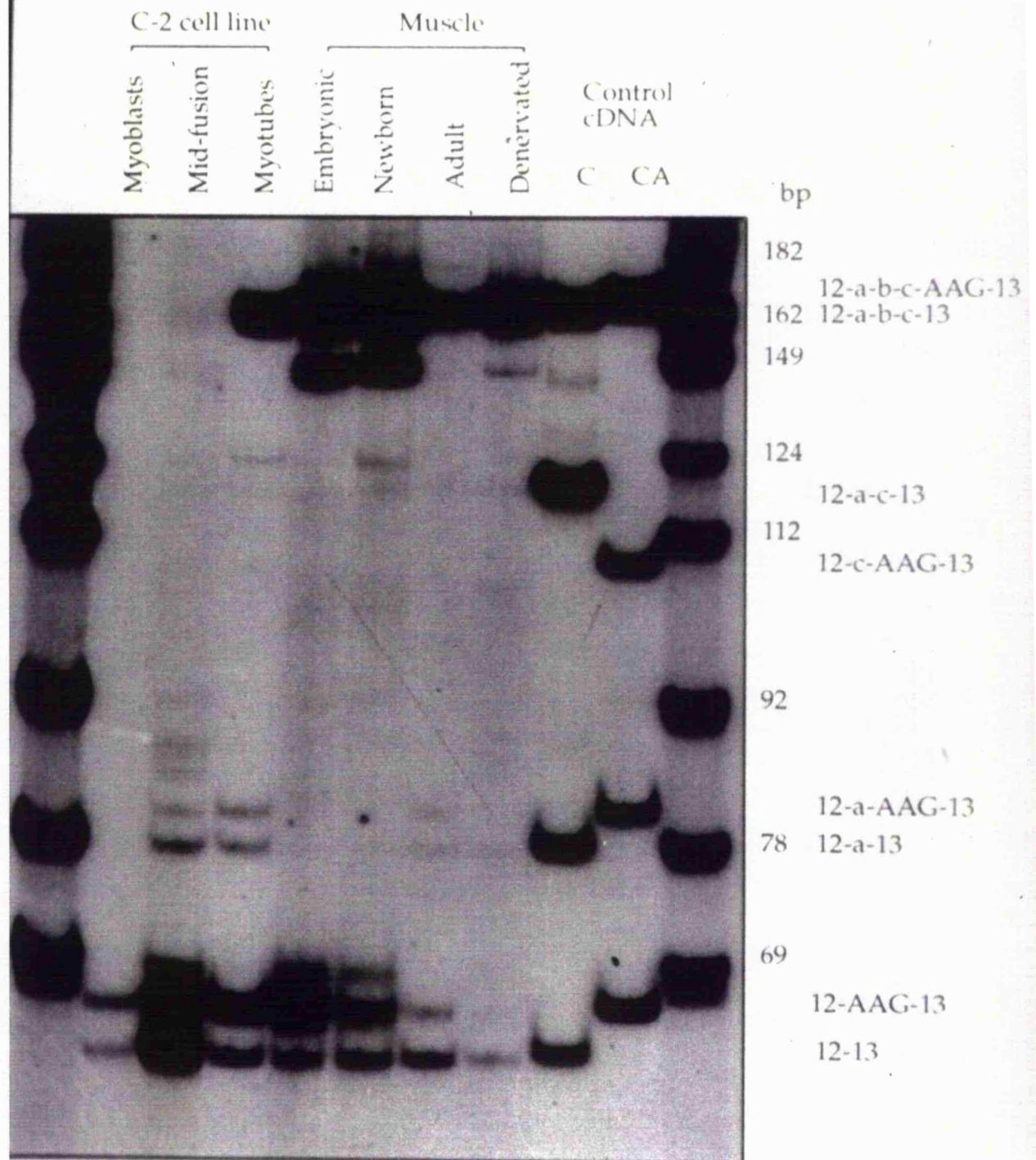
- 1). There was a shift toward inclusion of exons a-b-c as a unit during the differentiation of C-2 cells, and during development of skeletal muscle in Balb c mice.
- 2). There was a corresponding reduction in triplet AAG incorporation, in both developmental progressions, either alone or in combination with the MSD exons.
- 3). A transient reduction in triplet AAG usage at mid-fusion (2 days post-induction) was superimposed on the general trend toward reduced incorporation, and is reproducible.

3.3.2 *Effect of denervation of muscle.*

RNA from adult skeletal muscle before and after denervation (by surgical section of the sciatic nerve) was analysed for MSD isoform usage. The role for innervation in the control of gene expression of N-CAM had already been established (Walsh 1989). I thought it interesting to identify any switch in alternative splicing which may also accompany removal and reformation of neuromuscular junctions. The results of a direct RT-PCR are shown in Figure 3.3.1a and 3.3.1b. The quantitative data for exon use given in Table 3.3.2a.

3.3.1a RT-PCR products (using oligonucleotides complimentary to sequences in exons 12 and 13) were labelled and separated on a 10% native polyacrylamide gel. Three developmental progressions are shown, for the analysis of MSD exon incorporation in mRNA from developing muscle: C-2 cells at mononucleate myoblast stage, to mid-fusion (2 days post induction into myogenesis), to multinucleate myotubes (4 days post induction into myogenesis); embryonic, to newborn to adult mouse; pre- and post-denervation of adult skeletal muscle. Two markers include cDNAs previously isolated by cloning and sequencing. cDNA marker C includes the isoforms 12-13,12-a-13, 12-a-c-13, 12-a-b-c-13. cDNA marker CA includes the isoforms 12-AAG-13, 12-a-AAG-13, 12-c-AAG-13, 12-a-b-c-AAG-13.

SPLICING PATTERNS IN MUSCLE



3.3.1b RT-PCR products (using oligonucleotides complementary to sequences in exons MSD1c and 13) were labelled and separated on a 10% native polyacrylamide gel. Two developmental progressions are shown, for the analysis of MSD exon incorporation in mRNA from developing muscle: embryonic, to newborn to adult mouse; and pre- (labelled as adult mouse) and post-denervation of adult skeletal muscle. Two markers include cDNAs previously isolated by cloning and sequencing. cDNA marker C includes the isoforms 12-13,12-a-13, 12-a-c-13, 12-a-b-c-13. cDNA marker CA includes the isoforms 12-AAG-13, 12-a-AAG-13, 12-c-AAG-13, 12-a-b-c-AAG-13. The bands obtained from PCR of RT-PCR product derived from adult muscle required a longer exposure than the other PCR products. The band marked with an asterisk is thought to be due to partial denaturation of a PCR product.

ALTERNATIVE SPLICING OF THE AAG EXON IN TRANSCRIPTS CONTAINING THE MSD

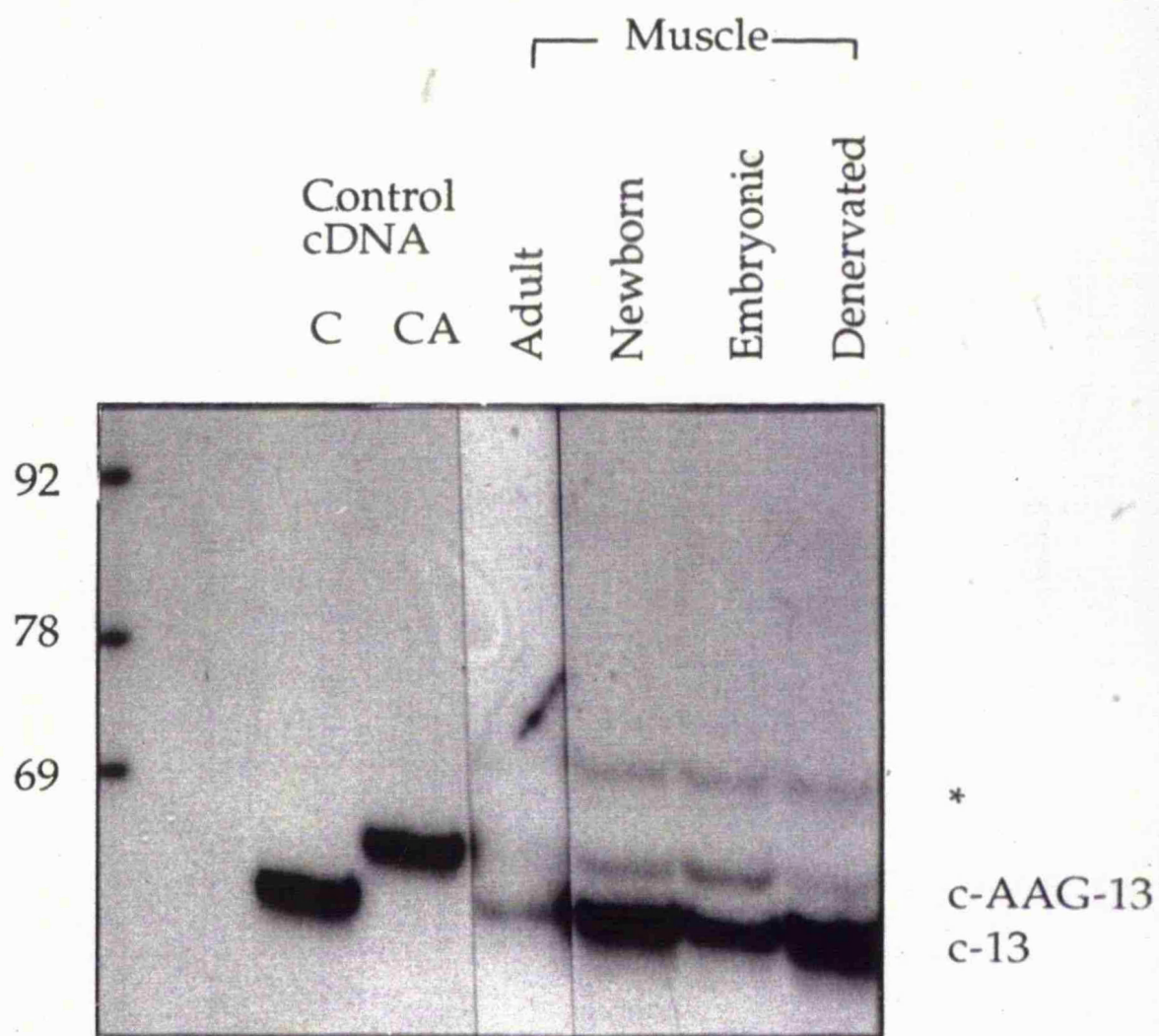


Table 3.3.1a

Isoform	C-2 myoblasts	C-2 mid-fusion	C-2 myotubes
12-13	43%	65%	11%
12-AAG-13	57%	31%	57%
12-a-13	<0.5%	3%	2%
12-a-AAG-13	<0.5%	1%	3%
12-a-b-c-13	<0.5%	<0.5%	22%
12-a-b-c-AAG-13	<0.5%	<0.5%	5%

Table 3.3.1a. The alternative isoforms of the MSD region of N-CAM, which were expressed in C-2 cells at different stages of development, were quantified by laser densitometry of the autoradiograph shown in Figures 3.3.1a and 3.3.1b. The initial absorbance readings were corrected for number of A+T residues as described earlier. The corrected absorbance values (in arbitrary units) are given.

Table 3.3.1b

Isoform	Embryonic muscle	Newborn muscle	Adult muscle
12-13	18%	12%	55%
12-AAG-13	39%	25%	13%
12-a-13	<0.5%	<0.5%	<0.5%
12-a-AAG-13	<0.5%	<0.5%	<0.5%
12-a-b-c-13	31%	60%	31%
12-a-b-c-AAG-13	12%	3%	1%

Table 3.3.1b. The alternative isoforms of the MSD region of N-CAM, which were expressed in Balb c mouse skeletal muscle at different stages of development, were quantified by laser densitometry of the autoradiograph shown in Figures 3.3.1a and 3.3.1b. The initial absorbance readings were corrected for number of A+T residues as described earlier. The corrected absorbance values (in arbitrary units) are given.

If the profile of myoblasts, mid-fusion cells, myotubes, embryonic, newborn and adult is taken as representative of a developmental profile, clear shifts in exon usage can be seen. These are illustrated in Figure 3.3.1c in which the % isoform use in each cell/tissue type is compared.

Several conclusions can be made from these data:

- 1). There was a shift toward inclusion of exons a-b-c as a unit during development.
- 2). There was a corresponding reduction in triplet AAG incorporation either alone or in combination with the MSD exons.
- 3). A transient reduction in triplet AAG usage at mid-fusion (2 days post-induction) was superimposed on the general trend toward reduced incorporation, and is reproducible.

3.3.2 Effect of denervation of muscle.

RNA from adult skeletal muscle before and after denervation (by surgical section of the sciatic nerve) was analysed for MSD isoform usage. The role for innervation in the control of gene expression of N-CAM had already been established (Walsh 1989). I thought it interesting to identify any switch in alternative splicing which may also accompany removal and reformation of neuromuscular junctions. The results of a direct RT-PCR are shown in Figure 3.3.1a and 3.3.1b. The quantitative data for exon use given in Table 3.3.2a.

3.3.1c The proportion of each isoform (analysed by a combined RT-PCR) is shown as a percentage of total N-CAM mRNA isolated from C-2 myoblasts, C-2 mid-fusion (2 day post induction of myogenesis), C-2 myotubes (4 days post induction of myogenesis) from cell culture, embryonic, newborn adult and post-surgically-denervated skeletal muscle.

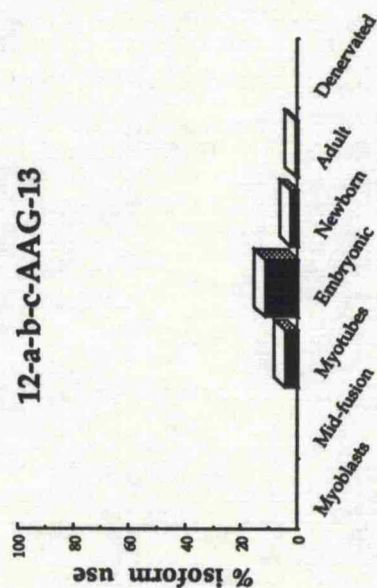
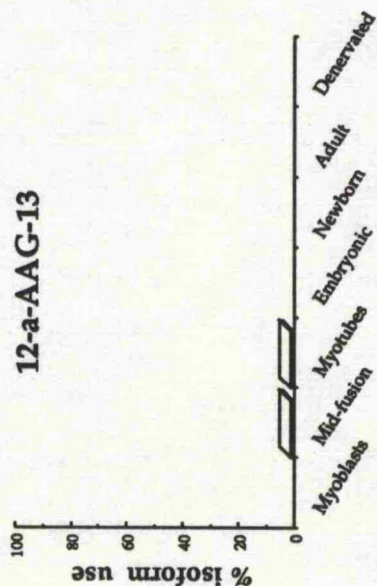
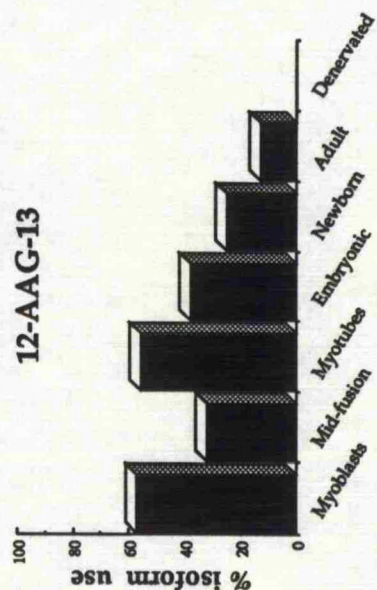
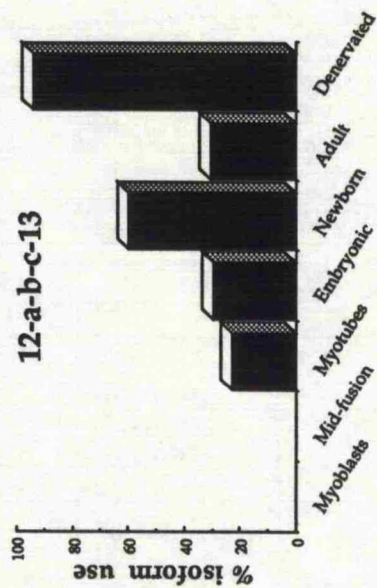
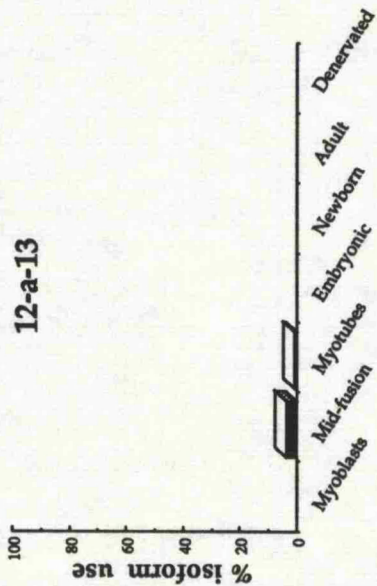
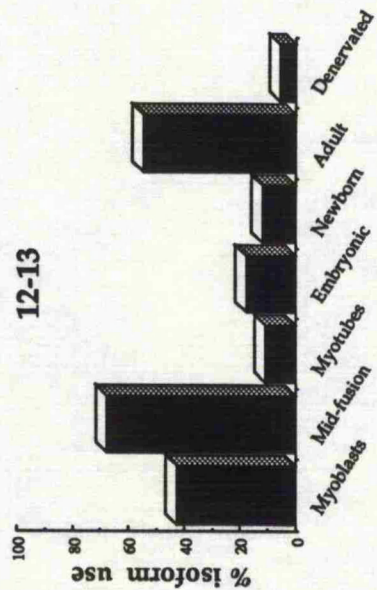


Table 3.3.2a

Isoform	Adult muscle	Denervated muscle
12-13	55%	6%
12-AAG-13	13%	<0.5%
12-a-13	<0.5%	<0.5%
12-a-AAG-13	<0.5%	<0.5%
12-a-b-c-13	31%	94%
12-a-b-c-AAG-13	1%	<0.5%

Table 3.3.2a. The alternative isoforms of the MSD region of N-CAM, which were expressed in Balb c mouse skeletal muscle before and after surgical section of the sciatic nerve, were quantified by laser densitometry of the autoradiograph shown in Figures 3.3.1a and 3.3.1b. The initial absorbance readings were corrected for number of A+T residues as described earlier. The corrected absorbance values (in arbitrary units) are given.

Surgical denervation of adult muscle cells caused a shift towards the exon combination 12-a-b-c-13 with reduced levels of inclusion of the triplet AAG. However, this could be a consequence of reduced contribution of RNA from neural tissue, and is consistent with the finding of an absence of isoform 12-AAG-13 in RNA isolated from brain (see later).

3.4 Exon usage in neural cells

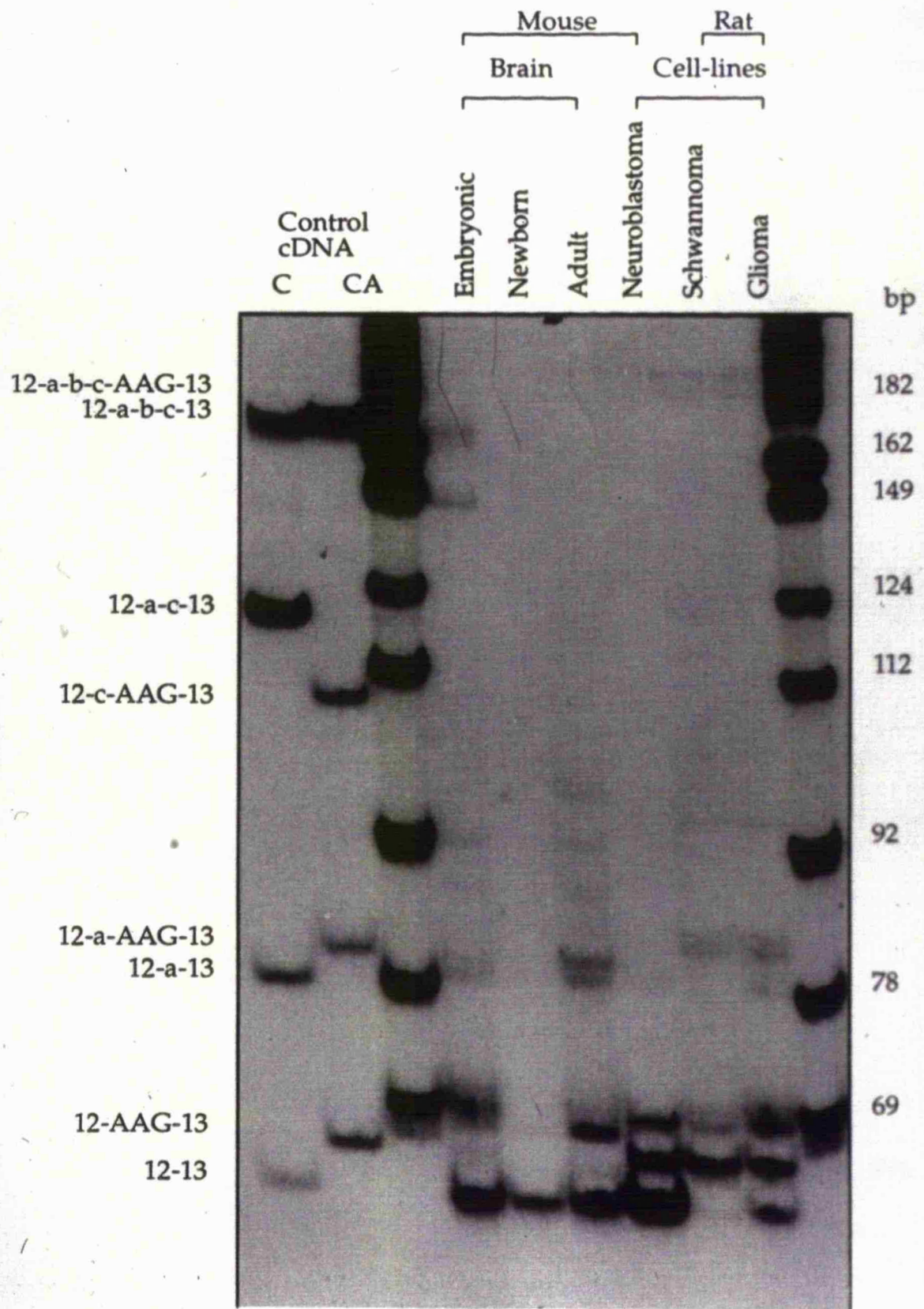
3.4.1 *Developmental profile in brain.*

The patterns of expression of the splicing within the 12-13 region of N-CAM in mouse, had in part been elucidated by the analysis of cDNA libraries (Santoni et al 1987). These data indicated the presence of one of the exons of the MSD exon MSD1a but not of the other exons. I wanted to confirm these findings, and to extend the investigation to see if there were also developmental differences in the expression of the MSD exons in brain tissue. RNA was analysed by a combined RT-PCR already described. Three stages in the development of mouse brain were investigated: embryonic (E16), newborn (post-natal 3 day) and adult (6 week). The results are shown in Figure 3.4.1. No laser densitometry was undertaken, as little evidence for a shift in exon use was seen. The identification of a rogue band at 69bp has already been discussed and was assumed to be due to aberrant migration also of the isoform 12-13 (as discussed in section 3.3.1). From Figure 3.4.1 several conclusions can be made:

- 1). Little or no incorporation of the MSD exons a, b and c was detected in RNA from the brains of mice at different stages of development. The very low levels of apparent incorporation of these exons in mRNA isolated from embryonic brain may have been derived from muscle due to contamination during dissection.
- 2). Low levels of the previously identified isoforms 12-a-13 and 12-a-AAG-13 were present in embryonic and adult tissues.
- 3). The isoform 12-AAG-13 was not found in normal brain tissue at any stage during development.

3.4.1. Analysis of N-CAM exons incorporated in mRNA from brain tissues and neural cell lines. RT-PCR products of RNA from embryonic, newborn and adult mouse is compared with RNA isolated from mouse N2A neuroblastoma, rat C6 glioma, and rat SWA schwannoma. Two control cDNAs were run as size markers. cDNA marker C includes the isoforms 12-13, 12-a-13, 12-a-c-13, 12-a-b-c-13. cDNA marker CA includes the isoforms 12-AAG-13, 12-a-AAG-13, 12-c-AAG-13, 12-a-b-c-AAG-13.

SPLICING PATTERNS IN NEURAL TISSUE



3.4.2 Profile in individual neural cell types.

The analysis of the pattern of expression of some of the cells which make up a complex tissue such as brain, was provided by the analysis of neural cell cultures. In a situation analogous to that for the switch in muscle cells, I thought it would be interesting to investigate the patterns of splicing of the different cells that make up the complex tissue. Three cell lines were investigated, one from the mouse and two from the rat. Although cross species comparisons are less reliable, the rat and mouse are closely related genetically. Figure 3.4.1 shows the results of RT-PCR of RNA isolated from these cells. A striking difference is seen between these RNA samples and those of mouse brain: they all include substantial proportions of the isoform 12-AAG-13. In the schwannoma this isoform predominates. These data indicate either that the immortalisation of these cells causes a switch in isoform use, or that cells are influenced by their neighbours such that taking them out of context can alter the expression of an alternatively spliced gene.

3.5 Sequence similarities between exons

The evidence presented earlier (Hamshire et al 1991; this thesis) indicates that the three exons of the MSD function to a high degree as a unit. This is particularly true for exons MSD1b and MSD1c. A possible mechanism for the coordinated exclusion or inclusion of the three exons of the MSD in different tissues involves the recognition of all three exons by a single *trans*-acting factor. In this situation it would be predicted that similar exon sequences would be bound by that factor. A search through the sequences of the three exons revealed one such sequence based upon a repeat of 12 nucleotides. These are aligned below.

```

a   a  T  A  G  C  C  C  T  C  C  T  C  C  a  c
b1  a  G  G  C  T  C  C  T  G  C  T  A  A  c  t  c  t  t
b2  c  C  A  C  C  C  A  T  G  T  T  C  C  g  t  t  g  t
b3      C  T  C  C  T  C  G  A  G  C  T  A
c1      C  A  A  C  C  T  G  G  C  C  T  C  t  t  c
c2      C  T  G  T  C  C  T  T  C  C  A  A  c  c  a
c3      C  A  G  A  C  C  T  G  T  C  T  A

```

To give a consensus sequence of:

```

C  AC/G  C  C  C  T  G  C  C  T  A
5  4  3  4  6  5  5  4  4  4  3  4  total = 51 matches

```

Two types of analysis of these sequences was undertaken: the first to determine if the sequences were the result of duplication events from a common ancestor; the second to assess the probability that the apparently similar sequences were not so purely by chance.

5.5.1. Similar because of common descent?

Two analyses were used to try to find the maximum parsimony tree: a simple cluster analysis (Hedrick 1985) and a more complex unweighted pair group method with arithmetic mean (with thanks to J.Brookfield).

Cluster analysis can deduce the maximum parsimony tree by establishing those pairs of sequences which show most similarity as follows:

	b1	b2	b3	c1	c2	c3
a	5	7	2	5	5	5
b1		4	3	3	7	5
b2			3	6	3	6
b3				5	5	5
c1					4	6
c2						7

The numbers of bases which are identical for each pair are indicated, those in bold define the pairs that have the highest level of identity which are grouped in the next table.

	b1c2	c2c3	b3	c1
ab2	8	8	4	7
b1c2		12	6	5
c2c3			6	7
b3				5

As before, the pair with the highest match are grouped and used to produce a third table:

	b1c2c3	b3	c1
ab2	9	4	7
b1c2c3		7	7
b3			5

Again, the highest pair are grouped and used to produce a fourth table:

	b3	c1
ab2b1c2c3	9	8
b3		5

The maximum parsimony tree is therefore:

a b2 b1 c2 c3 b3 c1

The tree derived by the more complex UPGMA method (shown below) differs slightly as there were certain points in which an arbitrary choice between two arrangements was made. However, the tree was essentially very similar.

a b2 c3 b1 c2 c1 b3

Having deduced the tree it is important to consider the mutations which occurred to produce the tree. In order to arrive at the final sequence, 31 mutations would have been needed. However, many of these mutations seemed to be back-mutations, in which the mutation generated the same base at the position as is found at another place on the tree. The more back-mutations the less likely it is that the tree is derived from a common ancestor. The nature of the mutations was also considered. These can be described as transition or transversion mutations. The number of transitions is usually higher or about the same as the number of transversions (J. Brookfield personal communication), even though there are twice the number of types of transversions as transitions. In analysis of the mutations which must have

occurred to produce this tree, only 12 out of 31 are transitions. Both factors argue against a hypothesis that the similarity of the sequences is due to a common descent.

3.5.2 Could the sequence similarity have arisen by chance?

To test this, we proposed the null hypothesis that "the sequence is the result of a random distribution".

In order to test the null hypothesis a probability that these sequences are similar by chance must be calculated. The calculation will determine the probability that a collection of 7 random 12 base-pair sequences would show a consensus as strong as 51 matches by chance. The basis of the calculation is the probability that at any position the most frequent base is found n times where $n=2$ to 7.

$$\begin{aligned}\text{Therefore } P(7) &= (0.25)^6 &= 0.00024414 \\ P(6) &= 4 \times (0.25)^6 \times 0.75 \times 7! / 6!1! &= 0.00512695 \\ P(5) &= 4 \times (0.25)^5 \times (0.75)^2 \times 7! / 5!2! &= 0.046140 \\ P(4) &= 4 \times (0.25)^4 \times (0.75)^3 \times 7! / 4!3! &= 0.230713\end{aligned}$$

When the base is common to only three of the sequences it is no longer the most abundant base, the calculation therefore includes all the combinations which can give three bases.

$$\begin{aligned}P(3:3:1:0) &= 12 \times (0.25)^7 \times 7! / 3!3!1!0! = 0.102539 \\ P(3:2:1:1) &= 12 \times (0.25)^7 \times 7! / 3!2!1!1! = 0.307617 \\ P(3:0:2:2) &= 12 \times (0.25)^7 \times 7! / 3!2!2!0! = 0.153809 \\ P(2) &= 4 \times (0.25)^7 \times 7! / 2!2!2!1! = 0.153809\end{aligned}$$

The sum of these probabilities is 1.

A calculation of the total probability that there would be at least 51 instances of identity to the consensus is calculated by multiplying the probabilities for all the possible combinations that produce 51. For example the probability that in the first two positions of the sequence there are 9 (5+4) matches to the consensus is the sum of:

$$\begin{aligned}
 &P(7) \times P(2) + \\
 &P(6) \times P(3) + \\
 &P(5) \times P(4) + \\
 &P(4) \times P(5) + \\
 &P(3) \times P(6) + \\
 &P(2) \times P(7) +
 \end{aligned}$$

this can be written as a general equation $P^2(i)$

where i =number of matches observed

Thus the probability at $P^n(i)$ can be written as

$$P^{n+1}(i) = \sum_{j=4}^{i-4} P^n(i-j)P^1(j)$$

A short program was written (J.Brookfield) to calculate the $P_{12}(i)$ values for the relevant i , which in this case was 51. This gave a final probability that 51 matches had arisen by chance of 5.857×10^{-5} , or 1 in 17,074.

However, a problem arises in that the alignment of the sequences was only one of many possible combinations. It is in fact possible to produce 40,080 different sequence alignments by multiplying together all of the possible non-overlapping 12bp sequences in each exon (i.e.3 for a, 334 for b and 40 for c). This would clearly indicate that at a purely statistical level, the null hypothesis that these sequences

were the consequence of a random distribution could not be discounted. Therefore, it seems likely that the apparently high degree of similarity between these sequences is coincidental.

3.6 Genomic PCR

Characterisation of the tissue and stage-specific patterns of expression of the MSD region of N-CAM has been undertaken with human (Dickson et al 1987; Gower et al 1988), mouse (Santoni et al 1988; Hamshire et al 1991; this thesis), rat (Small et al 1987), and chicken (Prediger et al 1988) derived RNA. However, there was little evidence of the genomic organisation of these exons other than in the human (Thompson et al 1989), and for one of the exons MSD1a and a putative site for the triplet AAG in the mouse (Santoni et al 1988). It was therefore thought prudent to characterise the genomic arrangement of these exons in the mouse. First, because of the discrepancies noted between the proposed genomic arrangement in the mouse and the human, and the finding of a shorter MSD1b exon in the chicken; and second, in order to isolate the exons and their surrounding introns such that a mutational analysis of the *cis*-acting control mechanisms regulating their alternative splicing may be established.

In order to map the exons a PCR approach was used. The rationale behind using a PCR approach rather than screening of genomic library was that with the very short exons involved (particularly to isolate the 15bp MSD1a) the use of exon specific probes may be problematic. Amplification by Taq polymerase is error prone as the enzyme exhibits no proof-reading activity. However, if amplifications are done in duplicate the chance of the same mutation occurring at the same early point in the amplification process is low. By the direct sequencing of PCR reactions, or the sequencing of a large number of clones, the sequence of the original genomic fragment can be determined.

Five different PCR amplifications were undertaken using DNA from mouse C-2 cells.

3.6.1 Exon 12 to MSD1a

The results obtained for this amplification are shown in Figure 3.6.1. The size of PCR product is approximately that expected from the sequence analysis of Santoni (1989). Unfortunately the identity of product has not been confirmed as genomic exon 12 to MSD1a as there was no internal exon position which could be hybridised to a nested oligonucleotide in a southern hybridisation. A direct sequencing of the PCR product was undertaken but failed to give any sequence data (data not shown). Initial attempts to clone the product were unsuccessful. Further attempts at cloning, and the possibility of using an intron directed oligonucleotide (for Southern hybridisation) based on the sequence data of Santoni awaited the results of the other PCR reactions.

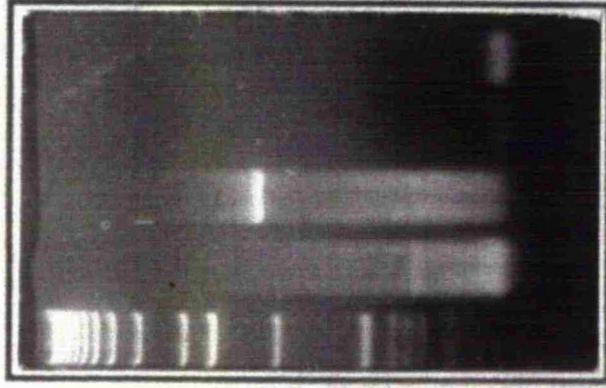
3.6.2. MSD1a to MSD1b

The intron between MSD1a and MSD1b was thought to be large as no evidence of the MSD1a exon had been demonstrated on the human genomic clone even though this clone contained about 5kb of DNA upstream of the MSD1b exon. Long PCR products have been amplified (Jeffreys et al 1988) although they are often only detectable by Southern hybridisation. Figure 3.6.2a shows the results of product separated on an agarose gel; Figure 3.6.2b shows the results of a southern blot and hybridisation from that gel using a probe directed against an internal portion of the MSD 1c exon. Clearly no single product was formed.

3.6.1 The genomic PCR of a region of N-CAM, using oligonucleotides complementary to exons 12 and MSD1a, was separated on a 1% agarose gel. Tracks 1 and 3 show the result of PCRs in the absence of genomic DNA, tracks 2 and 4 with the addition of mouse C-2 genomic DNA. Tracks 1 and 2 used PCR2002.5 as the buffer, tracks 3 and 4 used an alternative buffer system developed for the amplification of mini-satellite DNA (Jeffreys et al 1990).

12-a

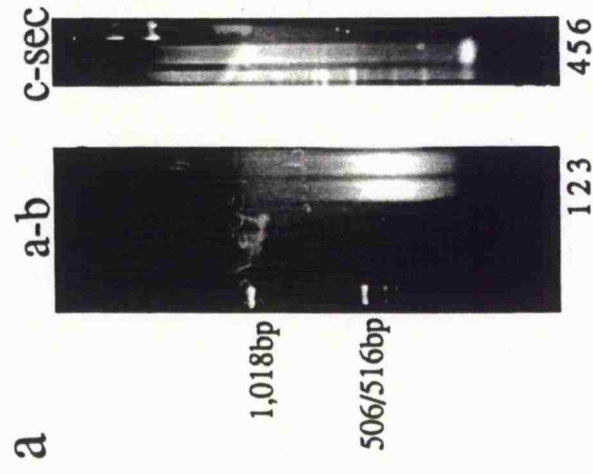
1,635
1,018
506/516



1 2 3 4

3.6.2a The genomic PCRs of two regions of N-CAM (a-b and c-SEC), using oligonucleotides complementary to exons MSD1a and MSD1b (tracks 1-3) and to exons MSD1c and SEC (tracks 4 and 5), were separated on a 1% agarose gel. Tracks 1 and 4 show the result of PCRs in the absence of genomic DNA; Tracks 2, 3 and 5 show the result of PCRs in the presence of mouse C-2 genomic DNA. Track 6 contains a positive control of the human genomic clone pUC4.8 which includes MSD1b, MSD1c and SEC exons.

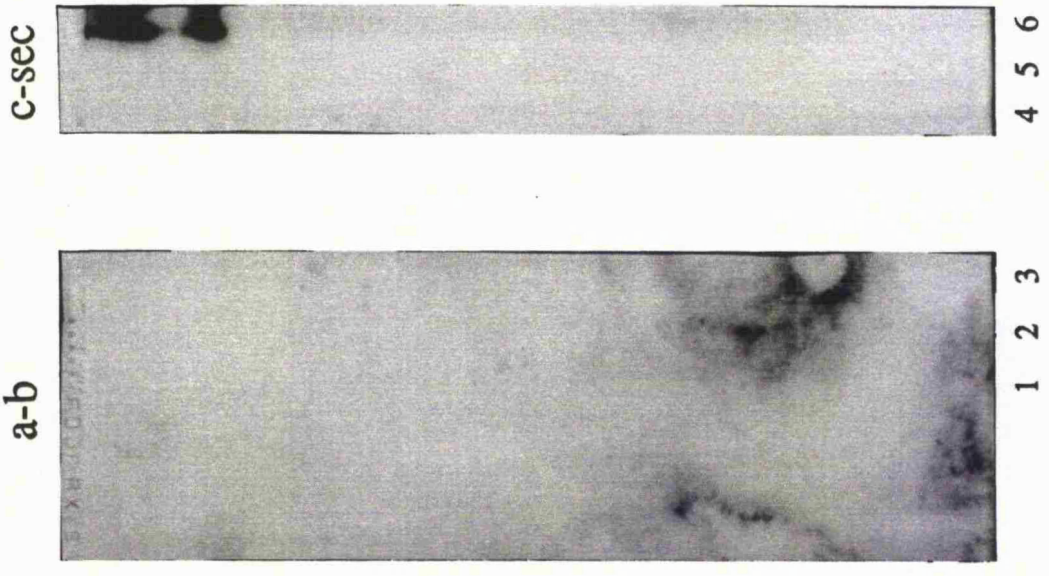
3.6.2b A southern blot of genomic PCR products, separated on an agarose gel (shown in Figure 3.6.2b) and probed with labelled oligonucleotides which are complementary to exons MSD1b and MSD1c. Tracks 1 and 4 show the result of PCRs in the absence of genomic DNA; Tracks 2, 3 and 5 show the result of PCRs in the presence of mouse C-2 genomic DNA. Track 6 contains a positive control of the human genomic clone pUC4.8 which includes MSD1b, MSD1c and SEC exons.



b

1,018bp

506/516bp



3.6.3. *MSD1b to MSD1c*

The results for the amplification of genomic DNA between exons MSD1b and MSD1c are shown in Figure 3.6.3. An oligonucleotide complementary to part of the MSD1c exon (but internal to the product of PCR) was used as a probe in a southern blot. This confirmed that one of the longer PCR products of 1.8kb was probably amplified genomic DNA which included MSD1b and MSD1c. The high level of hybridisation of probe to a much shorter product was probably due to amplification of mRNA by a reverse-transcriptase activity of the Taq polymerase. The reverse transcriptase activity of Taq polymerase was confirmed by the amplification of RNA using both N-CAM and human α _S-tropomyosin primers. These reactions were inconsistent and prone to very high background amplifications (data not shown).

The heterogenous PCR products were separated by size by elution through a ChromaSpin 400 column. This should have removed any unincorporated primers, and more importantly, the short products of amplification of mRNA. Following elution from the column the products were cut with the enzymes SstI and ClaI, sites for which had been incorporated into the sequence of the PCR primers. The products were cloned into pBlueScript. 40 positive colonies were isolated from the transformed JM101. These were sequenced by initial ddTTP screening followed by full four track sequencing of any unknown inserts. Although 40 clones were sequenced (10 of which included 8 different, but unknown sequences) the 4 clones which contained MSD1b and MSD1c did not contain any intervening intron and so must have been derived by reverse transcription and PCR of contaminating mRNA, even though these products would only have been 100 bases long and should have been removed by the ChromaSpin 400 column.

Figure 3.6.3: Southern blot of N-CAM genomic DNA.

3.6.3 The genomic PCRs of a region of N-CAM (b-c) using oligonucleotides complementary to exons MSD1b and MSD1c was separated on a 1% agarose gel (shown on the left), transferred onto a solid support (by southern blotting) and probed with labelled oligonucleotides which are complementary to exon MSD1c (shown on the right). Track 1 shows the result of a PCR in the absence of genomic DNA; Tracks 2 and 3 show the result of PCRs in the presence of mouse C-2 genomic DNA. Size markers are shown alongside, and the lengths given in base pairs.

b-c

Agarose gel

Southern blot

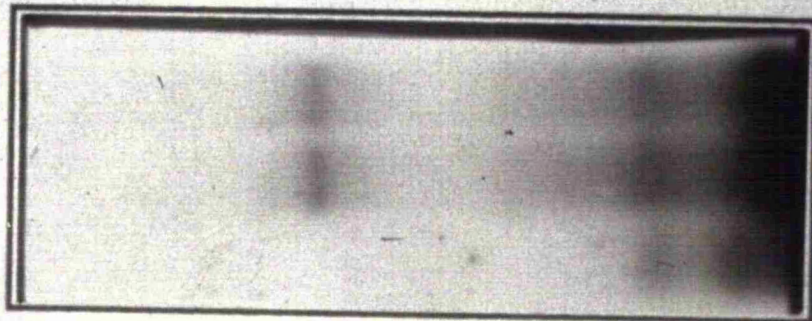


2,036

1,635

1,018

506/516



2,036

1,635

1,018

506/516

3.6.4 *MSD1c to exon 13*

There was evidence from preliminary studies that the distance between exons 12 and 13 was in the order of 10kb in the human (G. Dickson personal communication), and that MSD1c was in the middle of this region. Therefore, amplification from MSD1c to exon 13 would require amplification of a long product. The results of PCR amplification using oligonucleotides MH18c and MH2b are shown in Figure 3.6.4. Both the agarose gel and southern blot using a probe internal to the product gave no clear band although the smudge seen at about 3kb on the southern blot may have been produced by some binding of amplified product, but the background amplification is too high to give a definitive result.

3.6.5 *MSD1c to SEC*

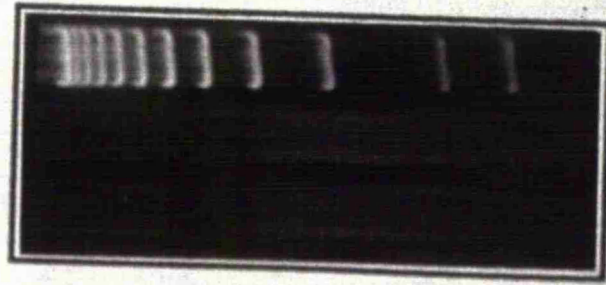
A fifth alternative exon, SEC, had been identified in the region of the MSD in human cDNA (Gower et al 1989) and had been mapped to a position approximately 1kb downstream of exon MSD1c. No evidence for this exon had been found in any of the analyses of the patterns of expression of the MSD exon of the mouse (see sections 3.3 and 3.4). For this reason a genomic amplification of the region between MSD1c and SEC was undertaken. An oligonucleotide complementary to the human SEC sequence (MHSECd) was used for the attempted amplification between MSD1c and the mouse homologue to the human SEC exon in the hope that sufficient conservation of sequence had been maintained between the human and a mouse homologue. A southern blot of a PCR amplification using oligonucleotides MH18c and MHSECd is shown in Figure 3.6.2. No product was identified.

3.6.4 The genomic PCRs of a region of N-CAM (c-13) using oligonucleotides complementary to exons MSD1c and 13 was separated on a 1% agarose gel (shown on the left), transferred onto a solid support (by southern blotting) and probed with labelled oligonucleotides which are complementary to exon MSD1c (shown on the right). Track 1 shows the result of a PCR in the absence of genomic DNA; Track 2 shows the result of a PCR in the presence of mouse C-2 genomic DNA. Size markers are shown alongside, and the lengths given in base pairs.

c-13

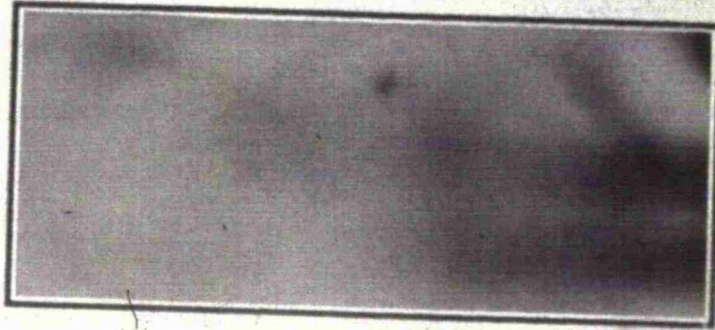
Agarose gel

Southern blot



5,090
4,072
3,054
2,036
1,635

1 2



5,090
4,072
3,054
2,036
1,635

1 2

The data suggested that the region of alternative splicing between exons 12 and 13 might cover a large stretch of DNA (shown in Figure 3.6.5). The shortest possible distance between exons 12 and 13 can be calculated by summing the distances deduced by PCR (this thesis) and genomic mapping (Santoni et al 1988). From exon 12 to MSD1a is 1.2kb (Santoni et al 1989; this thesis), MSD1b to MSD1c is 1.8kb (this thesis), and MSD1c to exon 13 is at least 0.5kb (determined by sequencing part of the intron upstream of exon 13, and not finding a homologous MSD1c sequence (Santoni et al 1989), although is probably at least 1kb if the SEC exon is also present in the mouse. Therefore, at a minimum, the region of alternative splicing covers 3.5 to 4kb, and from the genomic mapping possibly covered up to 20kb (Santoni et al 1988). It therefore appeared likely that a mini-gene construct to identify the important *cis*-acting elements would have to include substantial deletions in at least two of the introns in this region (MSD1a to MSD1b; MSD1c to exon 13). This tentative mapping of the gene, combined with the inability to isolate some of the intron sequences by a PCR technique led to the decision to switch the study of regulated alternative muscle-specific splicing to another gene system.

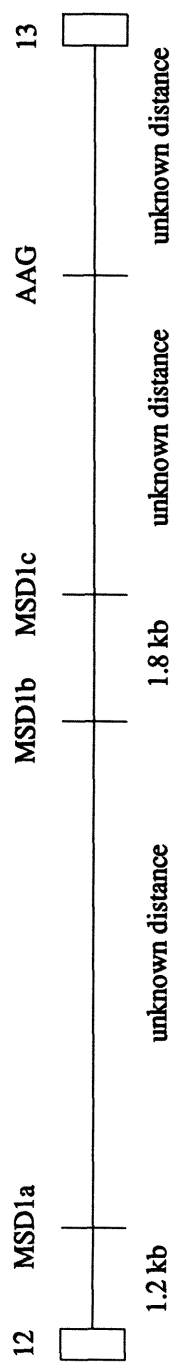


Figure 3.6.5
The genomic structure of the MSD region of murine N-CAM, deduced by PCR amplification of genomic DNA. The distance between exons 12 and MSD1a, and MSD1b and MSD1c is shown. The distances between other exons remain unknown.

Chapter 4

Regulation of alternative splicing in α_s -tropomyosin

As indicated earlier, the human α_s -tropomyosin gene contains a central region which encodes two exons which are incorporated into mature mRNA in a mutually exclusive manner. Inclusion of these exons, termed NM and SK, is stage- and tissue-specific. The aims of part of this project were the elucidation of the mechanisms which regulated the alternative splicing of these exons in both non-muscle and muscle cells. In order to achieve this, a method was required for transient expression of mini-genes in muscle cells which were maintained in culture. Transfection of mutant mini-genes could then be used to establish the important *cis*-acting elements involved in the regulation of alternative splicing.

4.1 Development of RT-PCR method

4.1.1 Vector specific oligonucleotide

During RT-PCR of RNA isolated from transfected cells it was possible that endogenous transcripts would also be amplified. In a system with both a high transfection efficiency and highly expressing vectors, the "signal to noise" ratio is high. In mouse C-2 cells where efficiency of both transfection and expression were expected to be low, this ratio would be critical. To circumvent this problem, the use of an oligonucleotide directed against the vector derived sequence, downstream of the inserted mini-gene and 85 nucleotides upstream of the SV40 poly-adenylation sequence, was tested. Amplification of RNA from untransfected cells (mock) was compared to the amplification of RNA from cells which had been transfected with the α_s -tropomyosin mini-gene construct containing the mutation delSK-6. In this initial experiment the RNA was derived from transfected COS cells. Figure 4.1.1 shows the result of amplification using a human α_s -tropomyosin-directed oligonucleotide (No1/8824) and the vector specific oligonucleotide (RT-PSVR). Using the α_s -tropomyosin oligonucleotide, endogenous transcripts also appear to have been amplified after

4.1.1 Two oligonucleotide primers (No1/8824 and RT-PSVR) are compared for their ability to differentiate between endogenous and transfected α -tropomyosin. RNA isolated from mock transfected cells and cells transfected with mutant construct D1 was analysed by RT-PCR. A second PCR, using the products of the initial RT-PCR as substrate, involved 8 cycles of PCR and end-labelled oligonucleotide. The products are separated by size on a native polyacrylamide gel. The expected positions of the isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

No1/8824
mock delSK-6



RT-PSVR
mock delSK-6



8 cycles only the exogenous transcripts (D1) were amplified when the vector-specific oligonucleotide was used.

4.1.2 The identification of alternative isoforms

Heterogeneous products of combined RT-PCR can be analysed in a variety of ways, including cloning and sequencing, or, if their lengths differ, separation in polyacrylamide gels. Neither of these two applications was satisfactory for the identification of the NM versus the SK isoforms of tropomyosin. Cloning and sequencing would have been impractical because of the large numbers of different mutations to be tested, and because of the difficulty in ensuring that products of different lengths (for example incorporation of both NM and SK in the same transcript compared to incorporation of NM alone) were cloned with the same efficiency. Direct separation by polyacrylamide gel electrophoresis was not feasible because of the identical lengths of the NM and SK exons.

The cDNA sequences for the final PCR products of the SK and NM isoforms were analysed for restriction endonuclease sites. Figure 4.1.2 shows the restriction sites unique to one of the isoforms. The problem with using one of these enzymes for the identification and subsequent quantification of alternative isoform use would be in ensuring that the enzyme had cut to completion. For this reason the restriction patterns of more common cutters was scrutinised. Alu1 restriction sites were found within exons 4 and exon SK, but were absent from exon NM (shown in Figure 4.1.2). By cutting the final PCR product with Alu1, not only could the different isoforms be separated by their size, but an ideal internal control for complete cutting by the enzyme was incorporated by the shortening of all products because of the cleavage sites within exon 4.

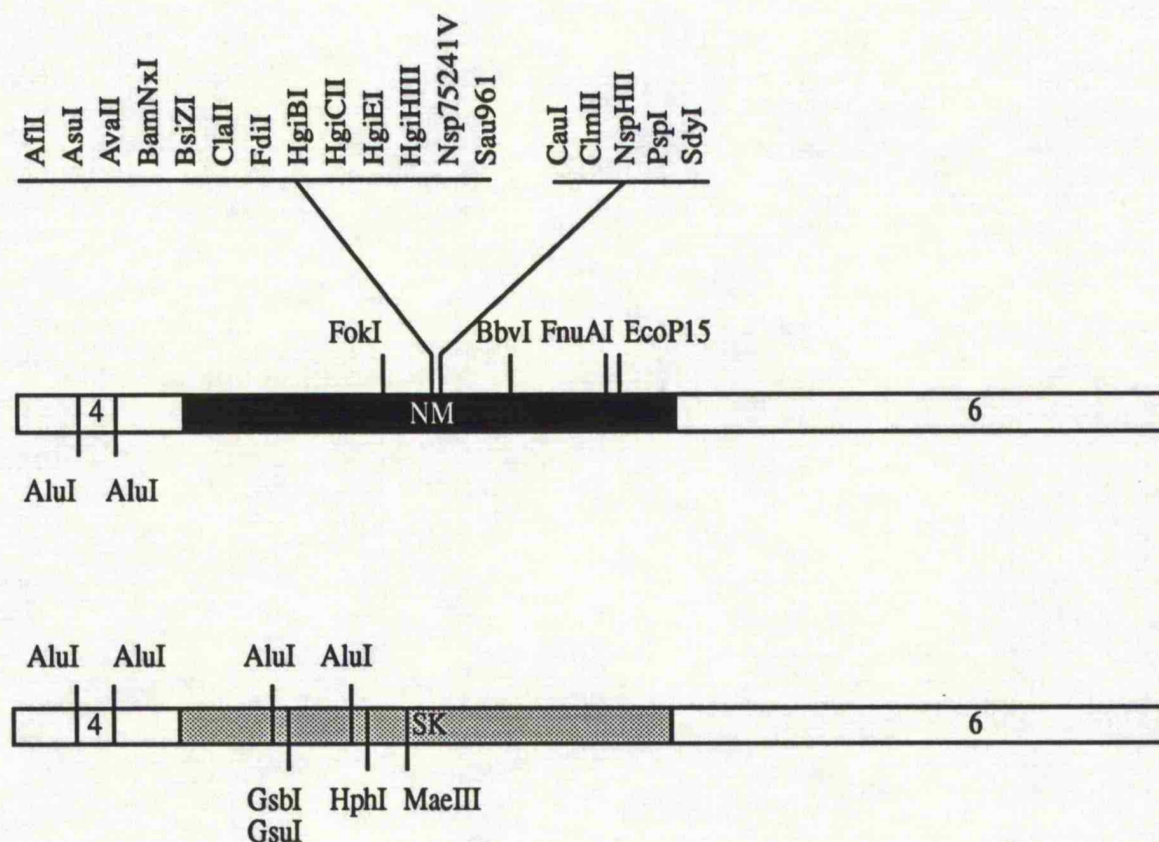


Figure 4.1.2

Map of the restriction sites of the cDNA isoforms of SK and NM. Open boxes represent the constitutive exons 4 and 6, dark shaded boxes represent NM sequences, and pale shaded boxes represent SK sequences. All enzymes cut only once within each cDNA at the position shown, except AluI which cuts in exon 4 and the SK exon.

Alu1 digestion of the products gives three different length products for NM, SK or jump-spliced isoforms. However, isoforms containing NM plus SK as a double incorporation would also give a product of identical length to the SK isoform. For this reason the uncut bands were also quantified such that the apparent use of the SK exon could be distributed between single and double incorporation of the SK exon.

A general equation to calculate the proportion of transcripts that have single SK exon incorporation can be derived as follows:

$$\% \text{ SK incorporation} = (1-D) \times sk / (nm+sk+4/6) \times 100 \quad (1)$$

$$\text{where } D = d / (d+s+j) \quad (2)$$

D= proportion of transcripts with double incorporation

d= absorbance reading for band migrating at 251 bp (double incorporation)

s= absorbance reading for band migrating at 176 bp (single incorporation)

j= absorbance reading for band migrating at 101 bp (jump-splicing)

nm= absorbance reading for band migrating at 152 bp

sk= absorbance reading for band migrating at 125 bp

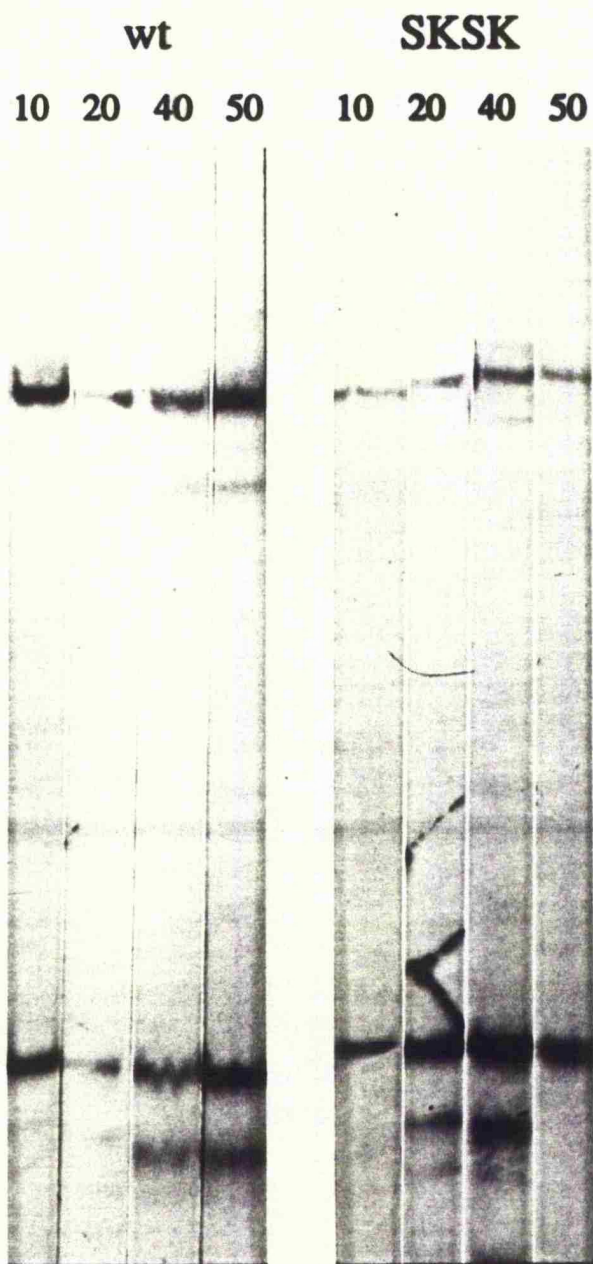
4/6=absorbance reading for band migrating at 87 bp

4.1.3 Linearity check of PCR

During a PCR amplification it is possible that substrates of different lengths are amplified with different efficiencies. This is a potential source of error if a mixture of substrates is to be co-amplified. In these studies it was important that longer substrates (due to the inclusion of both NM and SK) and shorter substrates (due to jump-splicing) were amplified with the same efficiency as the standard length 4-NM-6 or 4-SK-6 isoforms. If the co-amplification could not be shown to be equally efficient then it would be difficult to interpret any quantitative data derived with respect to comparisons between different length products. To test the linearity of amplification, cDNA sequences derived by RT-PCR were amplified for 10, 20, 40 and 50 cycles of PCR. Two substrates were amplified: SK,SK contained a high proportion of jump-splice; NM,SK(wt) contained a lower proportion of jump-splice. Figure 4.1.3 shows the results obtained. The autoradiograph shown is that used for the quantitative analysis and is a chimera of different exposures. With a constant exposure of the autoradiograph the intensity of the bands ranges from barely visible (10cycles) to grossly over-exposed (40 and 50 cycles).

4.1.3 Linearity of PCR amplifications

The linearity of co-amplification of DNAs of different lengths is assessed by analysis of PCR products after successive cycles. Two sources of RNA were used to produce PCR product, by RT-PCR, which was then amplified by 10-50 cycles of second round PCR. The wild-type construct produces isoforms 4-NM-6, 4-SK-6 (both migrating at the same position) and 4-6. The exon duplication construct (SKSK) produces the same isoforms, but with an increased proportion of isoform 4-6. The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon. Different exposures are shown for 10, 20, 40 and 50 cycles.



The table below (Table 4.1.3) shows the results of quantification of the products.

Table 4.1.3

Mutation	Isoform	Number of cycles			
		10	20	40	50
wt	4-NM-6 or 4-SK-6	1400	700	975	1798
	4-6	860	283	477	1154
	% 4-6	62%	71%	67%	61%
SKSK*	4-NM-6 or 4-SK-6	165	147	75	1003
	4-6	901	2540	1849	4268
	% 4-6	15%	5%	4%	19%

Table 4.1.3. Absorbance readings, obtained by laser densitometry, of autoradiograph shown in Figure 4.1.3. Percentage use of each isoform is also indicated.

No significant difference ($\chi^2 = 0.4$) was noted between the amount of product detected after 10 and after 50 cycles for the amplification of the different length products when the substrate was derived from nmsk(wt) derived DNA.

4.2 Optimisation of transient expression in myoblasts and myotubes

4.2.1 Endogenous expression of α -tropomyosin

It has been established that COS cells express a gene which contains an identical sequence to the human α_s -tropomyosin cDNA spanning exons 4, NM and exon 6 (Graham 1992b). The homologue of human α_s -tropomyosin has not been reported in the mouse, although the central exons (7 and 8) of the single α -tropomyosin gene in the rat, show a high degree of sequence similarity to the human α_s -tropomyosin SK and NM exons.

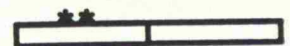
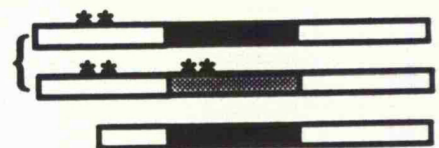
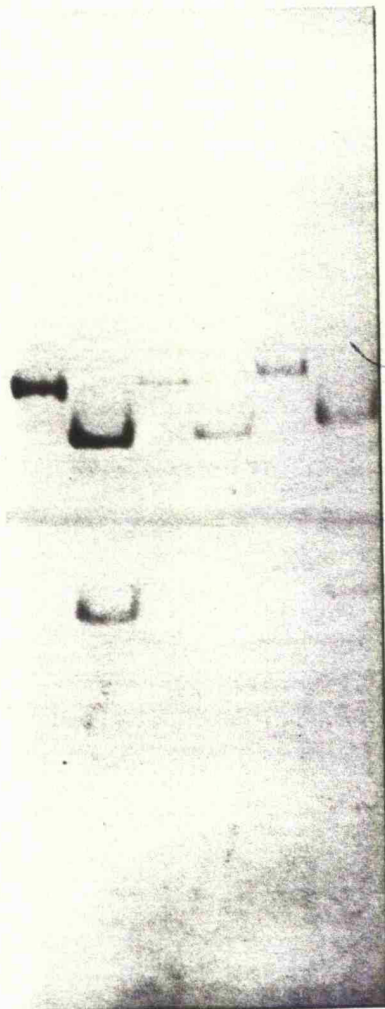
In order to study the regulation of alternative splicing of human α_s -tropomyosin in a mouse cell, it was important to establish that homologues to the alternatively spliced SK and NM exons were also expressed in the mouse. This was necessary to ensure that sequence-specific *trans*-acting factors would also be present. These could then regulate the splicing of the RNA encoded by transfected human mini-gene. Using an α_s -tropomyosin rather than a vector-specific oligonucleotide for the initial RT-PCR a homologue which had maintained a high level of sequence identity might be amplified from RNA isolated from non-transfected C-2 myoblasts and myotubes. In Figure 4.2.1 a product of identical length to that expected for the isoform 4-NM-6 was amplified from RNA isolated from 2-day myoblasts. In 6-day myotubes a second isoform was also present. This isoform contains an Alu1 recognition sequence in the same position as the Alu1 sequence in human α_s -tropomyosin SK isoform. These products are strong evidence for the presence of a mouse gene which contains homologous exons to the central exons of human α_s -tropomyosin, and that the sequence identity of these exons is high. Unlike exogenous tropomyosin expressed in mouse cells, the endogenous RNA from both the mouse and human cells does not appear to contain significant amounts of jump-

4.2.1 Endogenous RNA was analysed by RT-PCR using a non-vector specific oligonucleotide (No1/8824). RNA from untransfected 2 day myoblasts and 6 day myotubes is compared to RNA from HeLa cells. Alternative isoforms are identified by the presence of AluI restriction sites in exons 4 and SK. Uncut PCR product (u) and AluI cut product (c) are shown for each sample. The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

C-2 Mb/Mt HeLa

6 day 2 day

u c u c u c



spliced RNA.

4.2.2 Transfection of C-2 cells

Studies had already been undertaken to determine the optimum conditions for transfection of cultured C-2 cells (A.Lear, T.Parr personal communication). The technique already described was adopted but with some important modifications which were discovered during initial experiments into the transfection of cells with the tropomyosin constructs.

1. The media on the cells had to be changed daily.
2. Cells could be transfected as myoblasts and then induced to differentiate into myotubes. These myotubes would express the exogenous gene for at least 5 days post-transfection.
3. Expression of the transfected mini-gene was improved by the use of a glycerol shock to promote uptake of DNA (rather than a DMSO shock), and by the presence of 7mM sodium butyrate in the culture medium.
4. Cells which were living but either slow to grow or "sick" did not express the exogenous gene.
5. Cells that had been repeatedly passaged (more than 25 passages) lost the ability to express the exogenous gene when differentiated into myotubes, although could express the gene when harvested as myoblasts.
6. Success, in transfection, depended upon formation of good precipitates (recently, this has also been noted by others (Xu et al 1992). Fresh or frozen solutions gave the best results.

4.2.3 Effect of vector concentration

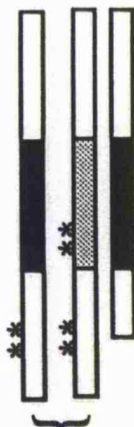
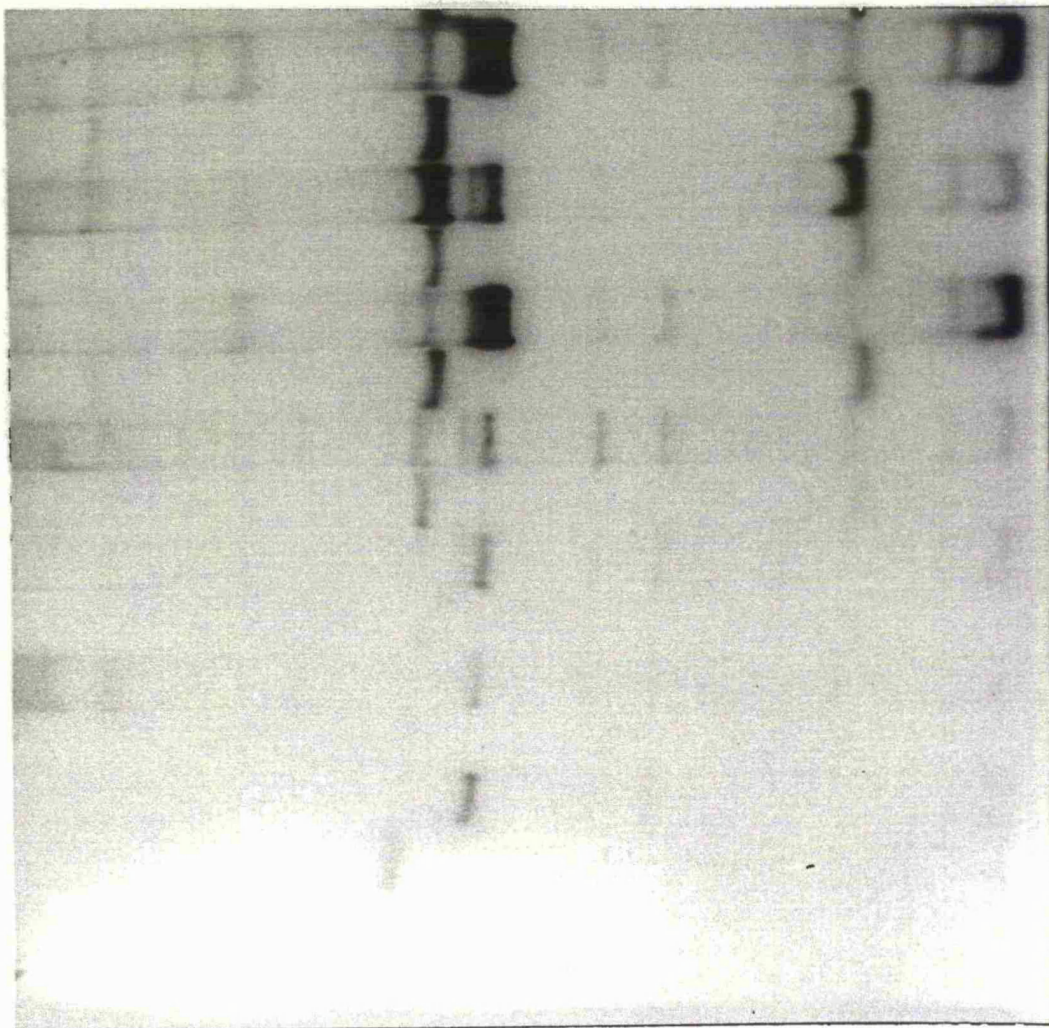
It has been noted by others (Guo et al 1991) that high concentrations of DNA led to inappropriate expression of an alternatively spliced gene (i.e increased incorporation of the muscle specific exon). In order to check that appropriate expression was produced by the exogenous gene, a series of transfections were undertaken using varying concentrations of vector. Figure 4.2.3 shows that even at the highest input concentration of vector the NM to SK ratio did not fall (as noted by others), but in fact appeared to rise. This result indicated that regulation of human α -tropomyosin alternative splicing is significantly different to the mechanism utilised by rat β -tropomyosin. The significance of the high proportion of jump-spliced transcripts is discussed in a latter section. At concentrations of 20 μ g and 40 μ g a longer product is also seen (*). The nature of this product is unknown, but could be the product of partially spliced RNA. The 5 μ g concentration used for all the transient expression experiments does not yield this errant product.

4.2.4 Transient expression with human α_s -tropomyosin

C-2 cells were chosen as a model system for the identification of the *cis*-elements required for the control of regulated alternative splicing in human α_s -tropomyosin. These cells retain the ability to differentiate into muscle cells and it was expected that a mini-gene construct transfected into these cells would be under the control of muscle-specific factors such that a switch in splicing would occur. This was tested by the transfection of C-2 myoblasts with a construct containing the "wild-type" mini-gene (wt). RNA from the transfected and mock-transfected cells was analysed either a day later whilst still undifferentiated myoblasts, or 5 days later after the cells had been induced to differentiate into

4.2.3 The effect of input concentration of vector is assessed by analysis of RNA, by RT-PCR, from myoblasts which were transfected with 0ug, 1ug, 5ug, 20ug and 40ug of wild-type α -tropomyosin mini-gene construct. The effect of each concentration is shown by the relative proportions of the alternative isoforms. Alternative isoforms are identified by the presence of AluI restriction sites in exons 4 and SK. Uncut PCR product (u) and AluI cut product (c) are shown for each sample. The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

0μg	1μg	1μg	5μg	5μg	20μg	20μg	40μg
u	c	u	c	u	c	u	c
c	u	c	u	c	u	c	u



4.2.4 Inclusion of the SK exon in myoblast and myotubes. RNA was isolated from mock transfected myoblasts (mock) and myoblasts transfected with wild-type construct from which the RNA was harvested after 20 hours (2 days) and 92 hours (5 days). The switch in the relative proportions of the alternative isoforms is identified by the length of AluI cut product (c) (AluI recognition sequences are present in exons 4 and SK and are indicated by *, but are absent from the NM exon). The uncut product (u) is also shown for each transfection in order to identify presence of double incorporation of exons. The expected positions of each isoform are illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

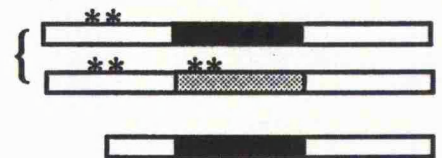
2 day

u c



5 day mock

u c c u



The table below (Table 4.2.4) gives the data as percentage use.

Table 4.2.4

Cell type	Isoform		
	4-NM-6	4-SK-6	4-6
2 day myoblasts	35%	18%	47%
5 day myotubes	20%	32%	48%

Table 4.2.4. A wild-type human α_s -tropomyosin mini-gene construct was transfected into myoblasts. RNA was isolated from cells before (as myoblasts) and after induction into myogenesis (as myotubes). Percentage isoform use was calculated from laser densitometry of the autoradiograph shown in Figure 4.2.4. Absorbance readings are given in Appendix 2.

Analysis of the raw data indicates a significant shift toward inclusion of the SK exon ($\chi^2_{2df}=14.2$) in differentiated myotubes. The high (although constant) level of jump-splice was an unexpected result and will be discussed later.

4.3 Regulation of alternative splicing in COS cells

In COS cells the regulation of alternative splicing was proposed to be via an interaction of a *trans*-acting repressor on the SK exon combined with the presence of a sub-optimal branch-point sequence for that exon (Graham 1992b). However, the analysis of spliced products was made by S1 nuclease digestion or by the cloning of RT-PCR products. Neither method could produce quantitative data for the level of double incorporation (NM plus SK) or jump-splicing. S1 mapping could not definitively identify double incorporation or jump splicing, and the RT-PCR conditions used to produce substrate for cloning were later found to be non-linear such that shorter products were overamplified (Graham 1992b). The cloning procedure may also have favoured incorporation of shorter products. For many of the constructs tested these factors were unimportant (as the relative level of identical length NM vs SK incorporation was important). However, the quantification of the relative levels of double to single incorporation was important for some of the constructs tested.

In order to clarify the effect of these mutations, I adapted the quantitative RT-PCR method which I had developed for the analysis of alternative splicing of α_s -tropomyosin in myotubes and myoblasts. Using the RT-PCR strategy, the extent to which NM-SK splicing was permitted could be assessed.

4.3.1. Does splicing of the exons follow an ordered pathway?

The splicing of the central exons of the rat β -tropomyosin had been proposed to follow an ordered pathway (Helfman et al 1988) with the splicing of the downstream alternatively spliced exon to its flanking constitutive exon forming the regulatory splice. If an analagous pathway were operating for SK inclusion into mature transcript, the fusion of SK to exon 6 should lead to rapid splicing of this fusion to the upstream constitutive exon 4, producing 4-SK-6 transcripts, but not double incorporation of transcripts. Figure 4.3.1 and Table 4.3.1a show the results of the quantitative RT-PCR of RNA isolated from COS cells following transfection with SK-6 construct.

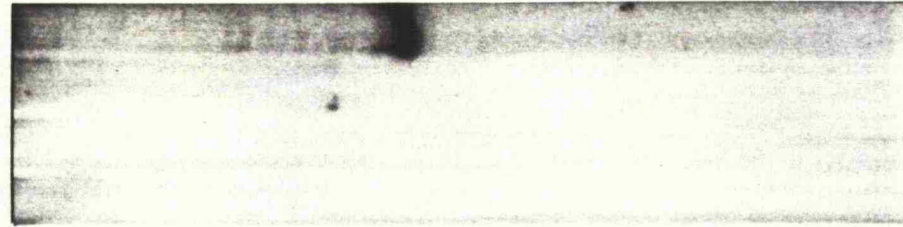
Table 4.3.1a

Mutation	Absorbance				
	d	s	n m	sk	j
wild-type	0	100**	100**	0	0
delSK-6	19.8	14.9	0	100**	0

Table 4.3.1a. The absorbance readings obtained by laser densitometry of the autoradiograph shown in Figure 4.3.1. Terminology is as given in section 4.1.2, but briefly: d and s refer to double and single incorporation of NM+SK, and NM or SK respectively in the uncut sample; nm, sk and j correspond to 4-NM-6, 4-SK-6 and 4/6 isoforms respectively, as identified by AluI digestion. **Indicates arbitrary absorbance values, where only a single isoform was detected.

4.3.1 Analysis of α_s -tropomyosin mRNA derived from transient transfection of COS cells with mutant α_s -tropomyosin mini-genes. RNA was analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon. Panel A shows the results from cells transfected with no vector (mock) and with wild-type construct (wt). Panel B shows the results from cells transfected with a duplication of the SK exon (SKSK*); and constructs which swapped the sequence of exons (SKNM* and (--)NM*). Panel C shows the results from mini-gene in which the SK exon and exon 6 were fused (delSK-6). Panel D shows the results from mini-gene in which the branch-points of both the SK and NM exons were uprated (2 x bpup).

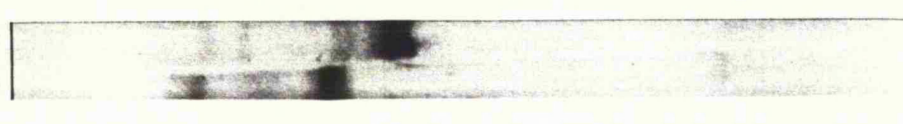
mock wt
u c u c



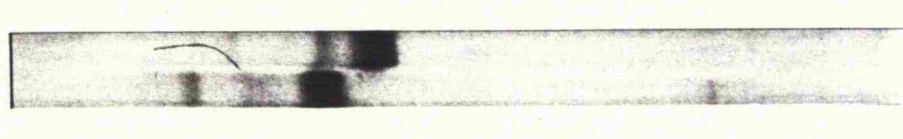
SKSK
u c



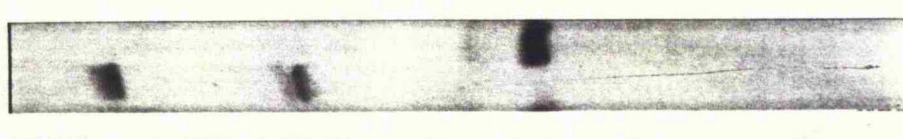
SKNM
u c



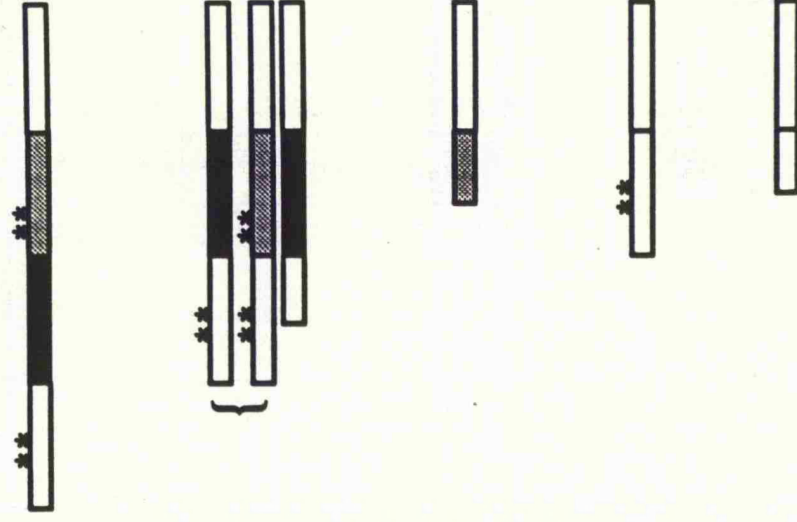
NMNM
u c



delSK-6
u c



2xbpup
u c



The percentage use for each isoform was calculated using equation 1 (section 4.1.2) and is given below in Table 4.3.1b.

Table 4.3.1b

	Percentage isoform use			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
wild-type	0%	100%	0%	0%
delSK-6	57%	0%	43%	0%

Table 4.3.1b. The percentage isoform use from wild-type and the mutation delSK-6.

Clearly, an ordered pathway analogous to the mechanism used by β -tropomyosin did not operate in human α -tropomyosin, and there was no absolute bar to NM-SK splicing.

4.3.2. Does repression lead to exon skipping if the SK exon is in both sites?

Three constructs had been made in which an entire exon sequence had been altered, but which retained the sequence of flanking intron (SKSK*, (--)NM*, SKNM*). The mechanism proposed by Graham (1992) that the SK exon contained *cis*-acting repressor sequences which could be bound by *trans*-acting factor(s) expressed in non-muscle cells, indicated that the transfection with SKSK* should result in expression of jump-spliced (4-6) products, (--)NM* should lead to expression of 4-NM-6 products, and SKNM* to incorporation of the NM (but now in the downstream site) to form 4-NM-6. Figure 4.3.1 shows the results of the quantitative RT-PCR of RNA isolated from COS cells following transfection with these constructs. These data were quantified by laser densitometry of the bands, the result are given in Tables 4.3.2a and 4.3.2b.

Table 4.3.2a

Mutation	Absorbance				
	d	s	nm	sk	j
SKSK*	0	4.0	0	4.0	91.0
(--)NM*	0	100**	100**	0	0
SKNM*	0	100**	100**	0	0

Table 4.3.1a(i). The absorbance readings obtained by laser densitometry of the autoradiograph shown in Figure 4.3.1. Terminology is as given in section 4.1.2, but briefly: d and s refer to double and single incorporation of NM+SK, and NM or SK respectively in the uncut sample; nm, sk and j correspond to 4-NM-6, 4-SK-6 and 4/6 isoforms respectively, as identified by AluI digestion. **Indicates arbitrary absorbance values, where only a single isoform was detected.

The percentage use for each isoform was calculated using equation 1 (section 4.1.2) and is given below in Table 4.3.2b.

Table 4.3.2b

	Percentage isoform use			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
SKSK*	0%	0%	4%	96%
(--)NM*	0%	100%**	0%	0%
SKNM*	0%	100%**	0%	0%

Table 4.3.2b. The percentage isoform use from mutations SKSK*, (--)NM*, and SKNM*. delSK-6. **N.B these reflect use of the NM exon in the native SK exon site.

Thus the SK exon and not the bounding splice-signals or the context of the site, contains sufficient information to cause its exclusion from mature mRNA from non-muscle COS cells. Alternatively, the NM exon contains an essential *cis*-acting activator sequence (which is absent from the SK exon) which allows for its inclusion in transcripts from non-muscle cells.

4.3.3 Does uprating the SK branch-point lead to double incorporation?

Two constructs had been made which altered the branch-points of the NM and SK exons. Uprating the NM branch-point would not be expected to alter isoform expression in COS cells (as NM is included in all transcripts), however alteration of the SK branch-point to the consensus sequence had been shown to have an affect on the level of SK incorporation (Graham 1992). A simple SK repressor mechanism in which uprating of the branch-point could cause a partial depression of the exon, predicts that incorporation of the SK exon would be within double incorporation (4-NM-SK-6) rather than 4-SK-6, as the NM exon is still active. Figure 4.3.1 and Tables 4.3.3a and 4.3.3b show the results of the quantitative RT-PCR of RNA isolated from COS cells following transfection with these constructs.

Table 4.3.3a

Mutation	Absorbance				
	d	s	n m	sk	j
SKbpup	0	77.7	52.2	25.5	0
2 x bpup	0	109.2	58.0	51.2	26.5

Table 4.3.1a(i). The absorbance readings obtained by laser densitometry of the autoradiograph shown in Figure 4.3.1. Terminology is as given in section 4.1.2, but briefly: d and s refer to double and single incorporation of NM+SK, and NM or SK respectively in the uncut sample; nm, sk and j correspond to 4-NM-6, 4-SK-6 and 4/6 isoforms respectively, as identified by AluI digestion.

The percentage use for each isoform was calculated using equation 1 (section 4.1.2) and is given below in Table 4.3.2b.

Table 4.3.3b

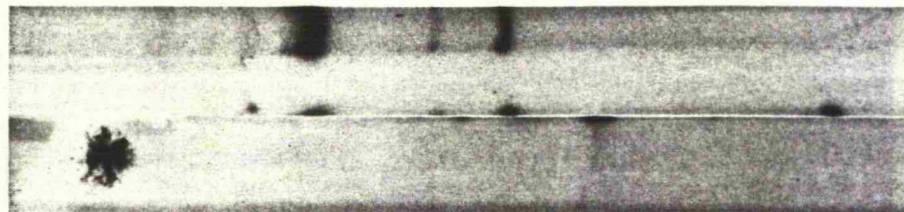
	Percentage isoform use			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
SKbpup	0%	67%	33%	0%
2 x bpup	0%	43%	38%	19%

Table 4.3.3b. The percentage isoform use from mutations SKbpup and 2 x bpup.

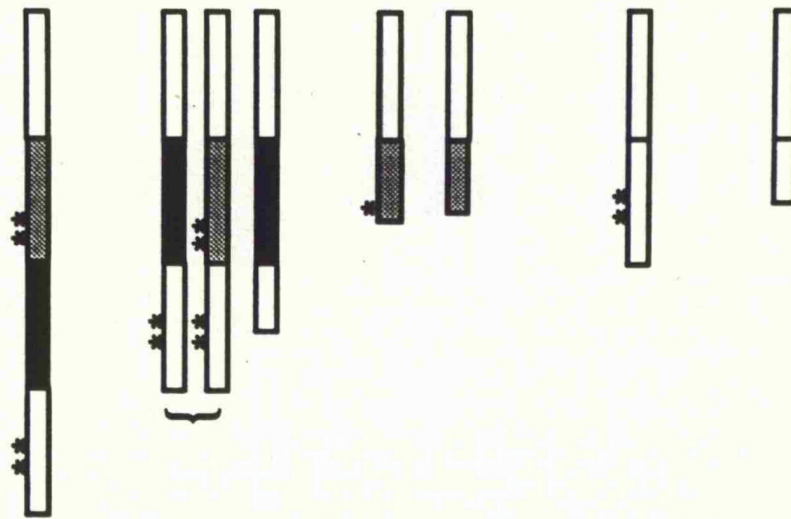
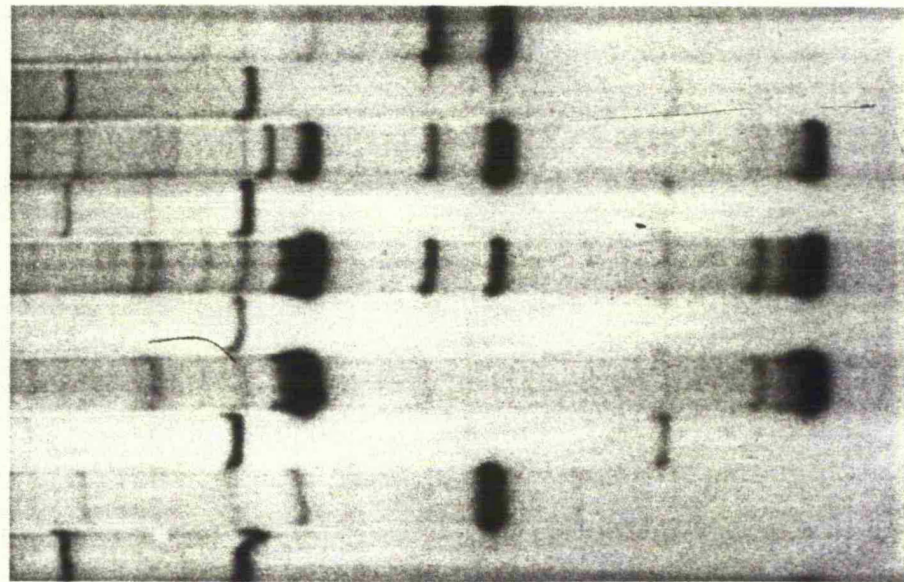
Therefore, both types of construct in which the branch-point of the SK exon was uprated (SKbpup, 2 x bpup) produced only single incorporation of exon SK.

4.3.2 Analysis of α_s -tropomyosin mRNA derived from transient transfection of COS cells with mutant α_s -tropomyosin mini-genes. RNA was analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon. Panel A shows the results from cells transfected with no vector (mock) and with the background mini-gene construct containing an uprated SK branch-point (SKbpup). Panel B shows the results from cells transfected with a series of mini-genes in which 15 base-pair sequences of the SK exons were swapped to the corresponding NM exon sequence.

mock SKbpup
u c u c



1-15 16-30 31-45 46-60 61-75
u c u c u c u c u c



4.3.4. Does de-repression of the SK site lead to double incorporation?

A series of constructs had been made in order to identify the putative *cis*-acting repressor sequences within the SK exon. Mutation of a repressor sequence was predicted to cause an increase in SK usage. De-repression of the SK exon should lead to double incorporation.

Figure 4.3.2 and Tables 4.3.4a and 4.3.4b show the results of the quantitative RT-PCR of RNA isolated from COS cells following transfection with these constructs.

Table 4.3.4a

Mutation	Absorbance				
	d	s	nm	sk	j
1-15*	17.4	20.9	8.2	58.6	0
16-30**	0	na	75.3	0	77.8
31-45**	0	na	55.3	33.5	72.6
46-60**	6.1	13.9	46.8	69.8	34.2
61-75*	7.4	15.8	2.0	69.8	0

Table 4.3.1a(i). The absorbance readings obtained by laser densitometry of the autoradiograph shown in Figure 4.3.1. Terminology is as given in section 4.1.2, but briefly: d and s refer to double and single incorporation of NM+SK, and NM or SK respectively in the uncut sample; nm, sk and j correspond to 4-NM-6, 4-SK-6 and 4/6 isoforms respectively, as identified by AluI digestion.

The percentage use for each isoform was calculated using equation 1 (section 4.1.2) and is given below in Table 4.3.4b.

Table 4.3.4b

Mutation	Percentage isoform use			
	4-NM-SK-6	4-NM-6	4-SK-6	4-6
SKbpup	0%	67%	33%	0%
1-15 ⁺	40%	12%	48%	0%
16-30 ⁺⁺	0%	49%	0%	51%
31-45 ⁺⁺	0%	34%	21%	45%
46-60 ⁺⁺	14%	31%	32%	23%
61-75 ⁺	31%	3%	66%	0%

Table 4.3.4b. The percentage isoform use from mutations within the SK exon.

From these data it is apparent that replacing bases 1-15 and the last 15 bases of the SK exon with NM sequence caused a dramatic switch toward inclusion of the mutant SK exon. This could be either because of removal of a *cis*-acting repressor element within the SK exon, or because an NM encoded activator element was introduced. Interestingly, one mutation (16-30) appeared to cause the re-repression of the SK exon to usage below that of the background mutation (SKbpup). I propose that this sequence could represent a repressor sequence of NM or an activator sequence of SK which was active in non-muscle cells.

4.4 Regulation of alternative splicing in C-2/C-12 cells by *cis*-acting elements.

4.4.1 Is the cis-acting sequence within exons?

Two constructs were tested, in which the exon sequences had been swapped, but in which the flanking intron sequence was retained. By removing the exon from its native splice-site signals (and other exon-specific intron elements), the roles for exon versus intron sequences could be dissected. In the (--)NM* construct, the downstream SK exon had been replaced with NM sequence and the upstream NM exon removed by deletion. In the SKSK* construct, the NM exon had been replaced with SK sequence. For technical reasons these constructs also incorporated a deletion in the stem region of the proposed stem-loop (see section 4.5.3) and the results are therefore compared to that background mutation.

If low SK usage in myoblasts is due to repression of SK via its exon sequences (as shown in COS cells), then replacing the NM exon with SK sequence (SKSK*) would cause repression of both sites (leading to jump-splicing), and replacing the SK exon with NM sequence ((--)NM*) would de-repress the downstream site (leading to 4-NM-6 splicing). Similarly, if low NM use in myotubes is due to repression of NM via its exon sequence, then replacing the NM exon with SK sequence (SKSK*) would de-repress the upstream site (leading to double incorporation 4-SK-SK-6), and replacing the SK exon with NM sequence ((--)NM*) would repress the downstream site (leading to jump-splicing). An alternative mechanism (to repression) which could regulate alternative splicing is tissue-specific activation of quiescent exon(s). If both exon sites were quiescent as a default pattern, but triggered by the binding of appropriate factor to the exon sequence, then in myoblasts the construct (--)NM* would produce 4-NM-6 isoforms, whereas the construct SKSK* would yield jump-spliced products. Similarly, in myotubes the construct (--)NM* would give jump-spliced products

whereas SKSK* would give double incorporation. Figure 4.4A. shows the results obtained when these constructs, and the background construct containing only a deletion of the stem loop (SKNM*) were transfected into myoblasts. Quantification of isoform use was achieved by phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and it's associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.1a

Table 4.4.1a

Mutation	Percentage exon use in myoblasts		
	Double (d)	Single (s)	Jump (j)
NMSK*	0%	97%, 69%**	3%, 31%**
SKSK*	0%	25% ₍₂₎	75% ₍₂₎
(--)NM*	0%	100% ₍₂₎	0% ₍₂₎

Table 4.4.1a. The percentage isoform use in transcripts derived from mutations NMSK*, SKSK* and (--)NM* following transfection into myoblasts. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. d, s and j refer to double (NM+SK), single (NM or SK) and jump-spliced (4/6) isoforms respectively. Numbers in brackets indicate that the percentage given is the average of duplicate transfections. ** The apparent irreproducibility of these results is analysed in a later section (4.5.3).

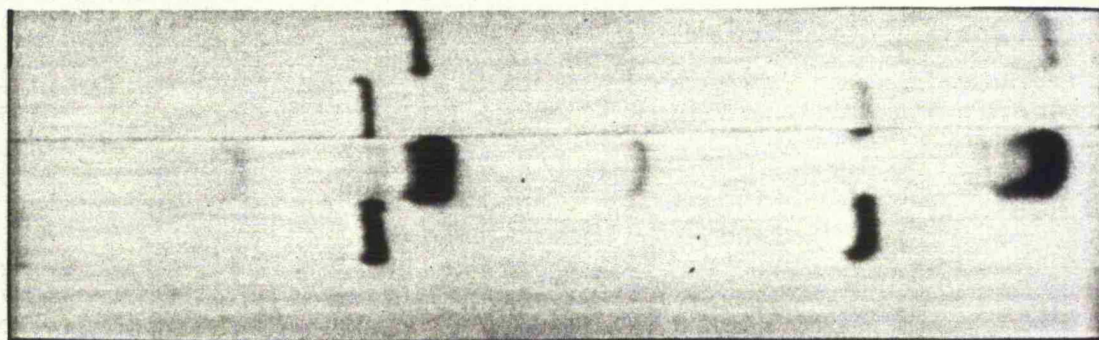
Figure 4.4B. shows the results obtained when these constructs, and the background construct containing only a deletion of the stem loop (SKNM*) were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video

4.4A Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 6 hours and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), with wild-type construct (wt) and with a background mutation construct (NMSK*) are compared to the results obtained from cells transfected with constructs containing: an SK exon duplication (SKSK), an exon swap (--)NM*, an inverted order of exons (SKNM*), uprated 5' splice-site of NM (NM5'cons) and an uprated branch-point of NM (NMbpup).

mock wt NMSK*

u c u c u c



SKSK

u c



(--)/NM

u c



SKNM

u c



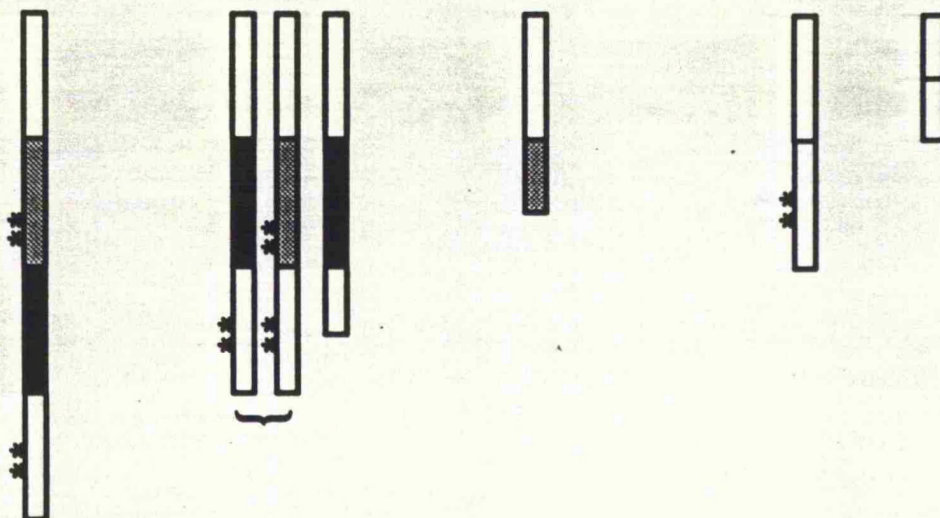
NM5'cons

u c



NMbpup

u c



4.4B. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 92 hours (when the cells had formed myotubes) and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), with wild-type construct (wt) and with a background mutation construct (delstem) are compared to the results obtained from cells transfected with constructs containing: an NM exon duplication (NMNM), SK exon duplication (SKSK), uprated 5' splice-site of NM (NM5'cons), and inverted order of exons (SKNM).

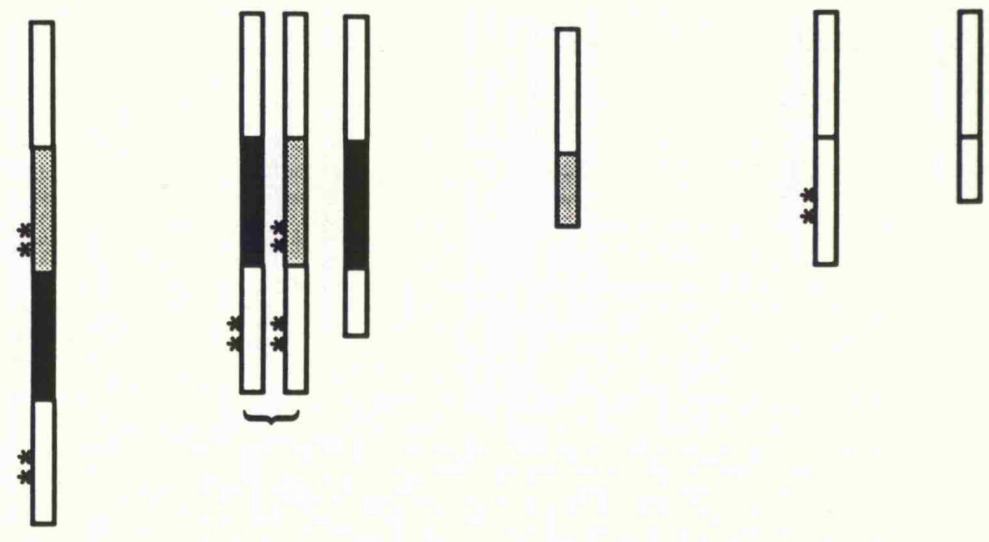
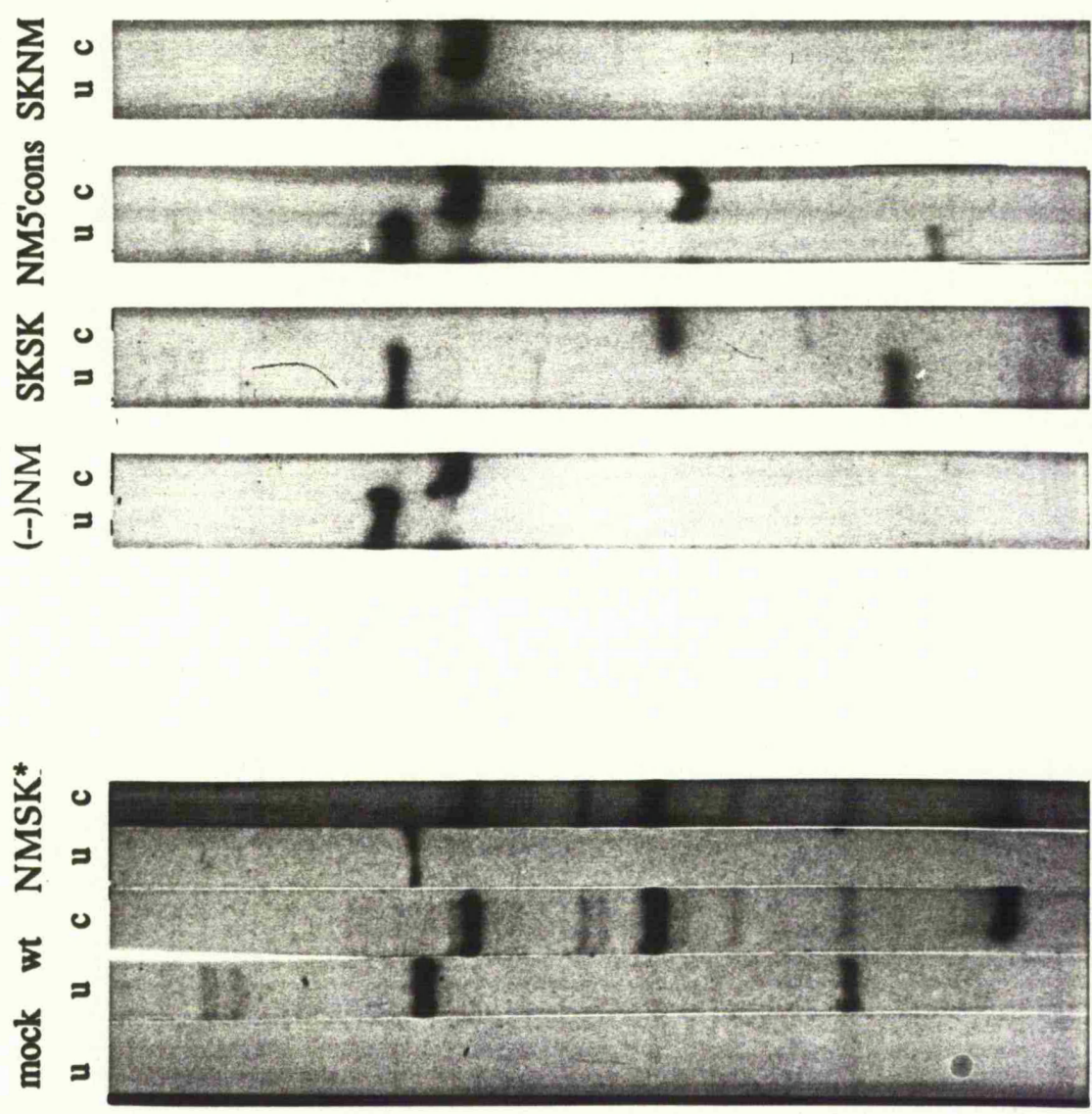


image of the autoradiograph using the ImageTM system (from UVP) and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.1b.

Table 4.4.1b

Mutation	Percentage exon use in myotubes		
	Double (d)	Single (s)	Jump (j)
NMSK*	0%	78% ₍₂₎	22% ₍₂₎
SKSK*	0%	50% ₍₂₎	50% ₍₂₎
(--)NM*	0%	100% ₍₂₎	0% ₍₂₎

Table 4.4.1b. The percentage isoform use in transcripts derived from mutations NMSK*, SKSK* and (--)NM* following transfection into myoblasts. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. d, s and j refer to double (NM+SK), single (NM or SK) and jump-spliced (4/6) isoforms respectively. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

These results show that in myoblasts either there are repressor elements within exon SK which inhibit the use of that exon in myoblasts, or activator elements in NM which enable use of an otherwise quiescent SK exon site. In myotubes, reduced incorporation of the NM exon must be regulated by a mechanism other than NM exon encoded repressor sequences.

4.4.2 What prevents NM exon use in myotubes?

If there is no repressor element within the NM exon, how is splicing of this exon prevented? Is the exon effectively "dormant" in muscle cells, because it is surrounded by incompetent splicing signals? To test this two mutations were made. The first changed the 5' splice-site to the consensus (NM 5'cons); the second changed the branch-point to the consensus (NM bpup). Figure 4.4A. shows the results obtained when these constructs were transfected into myoblasts. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and its associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.2a

Table 4.4.2a

	Percentage exon use in myoblasts			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
wild-type	0%	55% ₍₂₎	6% ₍₂₎	39% ₍₂₎
NM5'cons	0%	91% ₍₃₎	4% ₍₃₎	5% ₍₃₎
NMbpup	0%	47%,49%	4%,14%	49%,37%

Table 4.4.2a. The percentage isoform use in transcripts derived from wild-type and mutations NM5'cons and NMbpup following transfection into myoblasts. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. Numbers in brackets indicate that the percentage given is the average of duplicate or triplicate transfections.

Figure 4.4B. shows the results obtained when the exon NM uprated 5'-splice-site (NM5'cons) and the background (NMSK) construct were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video image of the autoradiograph using the Image™ system and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.2b.

Table 4.4.2b

Mutation	Percentage exon use in myotubes			
	4-NM-SK-6	4-NM-6	4-SK-6	4-6
wild-type	0%	20% ₍₂₎	32% ₍₂₎	48% ₍₂₎
NM5'cons	0%	59% ₍₂₎	30% ₍₂₎	12% ₍₂₎

Table 4.4.2b. The percentage isoform use in transcripts derived from the wild-type and NM5'cons mutation following transfection into myoblasts. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

These results indicate that the 5'splice site of exon NM can be improved by mutation to the consensus, and that this will shift the splicing patterns produced. Thus, the presence of a sub-optimal splice-site may be exploited as a mechanism by which the NM exon is skipped in myotubes. However, this requires that either a specific activator is present to induce incorporation of the NM exon in myoblasts, or that general splicing factors are present in different amounts in myoblasts and myotubes; this results in a shift from inclusion to skipping during differentiation. Changing the branch-point to the consensus did not improve the splicing of the NM exon. This construct was not tested in myotubes.

4.4.3 Is the *cis*-acting sequence within the introns?

To test whether the intron sequences that surround an exon are the determinants of splicing the exons were swapped such that SK was placed in the upstream site, and NM placed in the downstream site (SKNM*). For technical reasons this mutation also included a deletion of part of a potential stem loop encompassing the 5' splice site of the downstream (SK) site. Figure 4.4A. shows the results obtained when this construct, and its background construct containing only a deletion of the stem loop (SKNM*) were transfected into myoblasts. If intron sequence alone determined splicing, the use of the upstream and downstream sites would be identical to the background mutation. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and its associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.3a

Table 4.4.3a

Mutation	Percentage exon use in myoblasts		
	Upstream	Downstream	Jump
NMSK*	84%,63%	13%,6%	3%31%
SKNM*	0%	100%(2)	0%(2)

Table 4.4.3a. The percentage exon use in transcripts derived from the background mutation NMSK* and mutation SKNM* following transfection into myoblasts. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. Numbers in brackets indicate that the percentage given is the average of duplicate or triplicate transfections.

As already indicated, the delstem mutation led to an inconsistent drop in the level of jump-splicing (analyses in detail in section 4.5.3) However, the proportions of exon use are obviously significantly different.

Figure 4.4B. shows the results obtained when the exon swap (SKNM*) and the background (NMSK) construct were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video image of the autoradiograph using the Image™ system and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.3b.

Table 4.4.3b

Mutation	Percentage exon use in myotubes		
	Upstream	Downstream	Jump
NMSK*	28% ₍₂₎	48% ₍₂₎	22% ₍₂₎
SKNM*	0% ₍₂₎	98% ₍₂₎	2% ₍₂₎

Table 4.4.3b. The percentage isoform use in transcripts derived from the background and SKNM* mutations following transfection into myoblasts. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

These data indicate that in myoblasts, when the exons were swapped, the exon use was also swapped indicating that exon sequence and not its context (i.e. bounding intron sequence) was the point of control. In myotubes, when the exons were swapped, upstream site use was markedly reduced, and downstream site use increased. These data are consistent with two models. First, the

upstream NM exon site is effectively quiescent unless activated by NM exon sequences bound by a trans-acting factor. Second, a combination of the presence of residual *trans*-acting repressor which binds to sequences within the SK exon in myoblasts and poor splice-signals surrounding the NM site causes skipping of the upstream site. Thus in myotubes, the intron sequences that surround the NM exon may play a role in regulation of alternative splicing, but are not the sole requirements.

4.4.4 Identification of elements within SK exon

A series of mutations were made which changed the sequences of regions of the SK exon into the corresponding NM exon sequence. These mutations were made in a background which included an uprating of the branch-point of SK (SKbpup⁺). Three of the constructs (16-30⁺⁺, 31-45⁺⁺, and 46-60⁺⁺) also contained a deletion of the stemloop*. Figure 4.4.C shows the results obtained when these constructs and the background construct (SKbpup) were transfected into myoblasts. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and its associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.4a.

4.4C Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 6 hours and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results from cells transfected with no vector (mock) and with a background mutation construct (SKbpup) are compared to the results obtained from cells transfected with a series of constructs in which consecutive 15bp regions of the SK exon were replaced by the corresponding region of the NM exon.

mock SKbpup 1-15 16-30 31-45 46-60 61-75

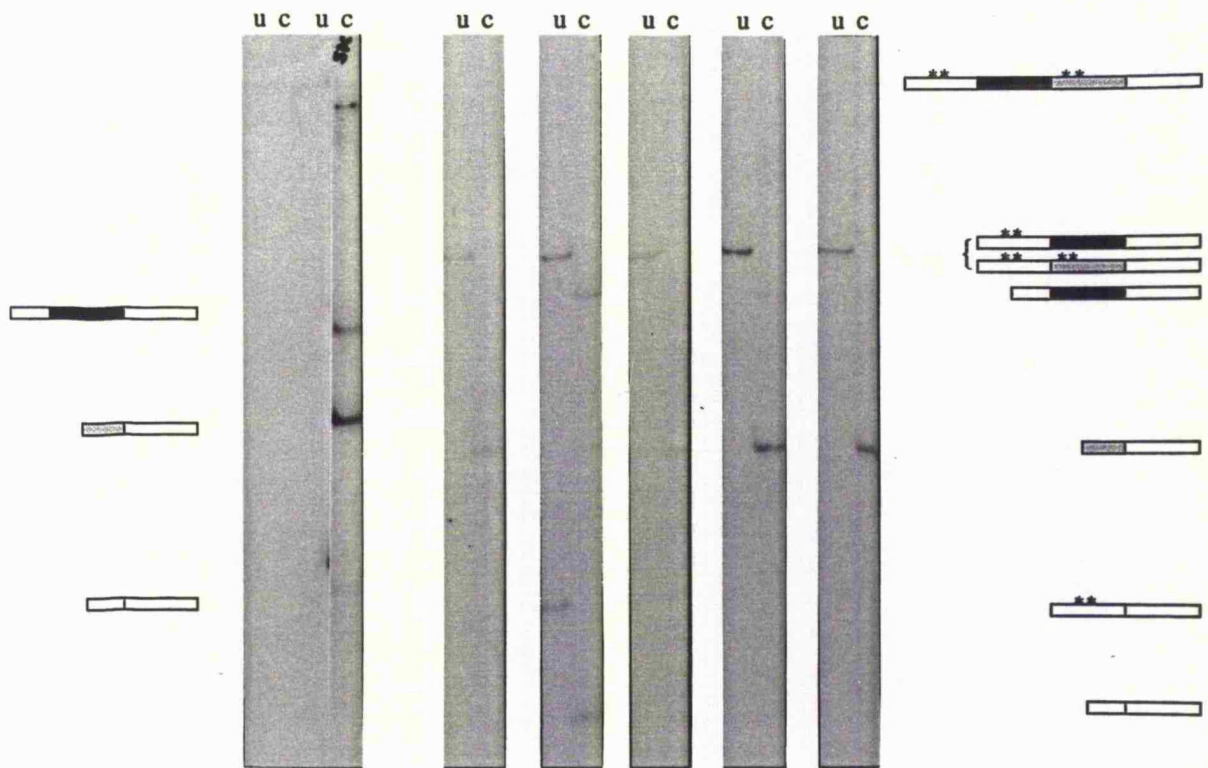


Table 4.4.4a

Mutation	Percentage exon use in myoblasts			
	4-NM-SK-6	4-NM-6	4-SK**-6	4-6
SKbpup ⁺	0%	30%	58%	12%
1-15 ⁺	0%	2%	97%	1%
16-30 ⁺⁺	0%	27%	12%	61%
31-45 ⁺⁺	0%	28%	59%	13%
46-60 ⁺⁺	0%	1%	98%	1%
61-75 [*]	0%	1%	97%	2%

Table 4.4.2a. The percentage exon use in transcripts derived from a series of mutations in which the SK exon was replaced by corresponding regions of the NM exon, compared to the background mutation SKbpup. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. **Indicates that this SK site was mutated to contain NM sequences.

Figure 4.4D. shows the results obtained when these constructs, and the background (SKbpup) construct were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video image of the autoradiograph using the Image™ system and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.4.4b.

4.4D. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 92 hours (when the cells had formed myotubes) and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut product (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results from cells transfected with no vector (mock) and with a background mutation construct (SKbpup) are compared to the results obtained from cells transfected with a series of constructs in which consecutive 15bp regions of the SK exon were replaced by the corresponding region of the NM exon.

mock	SKbpup				
c	c	c 1-15	c 16-30	c 31-45	c 46-60
		c 61-75			

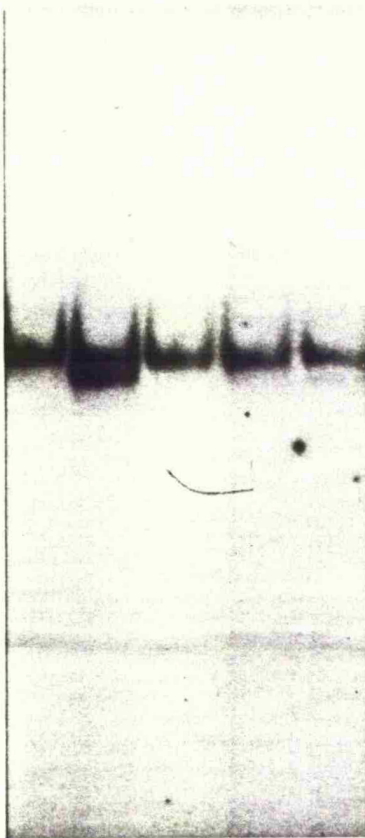


Table 4.4.4b

	Percentage exon use in myotubes			
Mutation	4-NM-SK-6	4-NM-6	4-SK**-6	4-6
SKbpup ⁺	0%	0%	100%	0%
1-15 ⁺	0%	0%	100%	0%
16-30 ⁺⁺	0%	0%	100%	0%
31-45 ⁺⁺	0%	0%	100%	0%
46-60 ⁺⁺	0%	0%	100%	0%
61-75 ⁺	0%	0%	100%	0%

Table 4.4.2b. The percentage exon use in transcripts derived from a series of mutations in which the SK exon was replaced by corresponding regions of the NM exon, compared to the background mutation SKbpup. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. **Indicates that this SK site was mutated to contain NM sequence.

From these data it is apparent that sequences within the first 15bp of the SK exon, bases 46-60 and 61-75 contain elements that are recognised either by sequence-specific repressors of SK, or by sequence-specific activators of NM. This repression/activation is operating in myoblasts, but not in myotubes.

4.5 A role for competition in alternative splicing in C-2/C-12 cells

4.5.1 *Are the splice-sites sub-optimal?*

If alteration of the 5' splice site of NM can perturb the splicing patterns, it is possible that alteration of the splice sites of exons 4 and SK can also perturb splicing. Three constructs were made which altered the native splice-sites to the consensus: the first changed the 5'splice site of exon SK to the consensus (SK5'cons*), the second changed the branch-point of SK to the consensus (SKbpup) and the third changed the 5'splice-site of exon 4 to the consensus (4 5'cons). Again for technical reasons the first (SK5'cons*) was made in a background containing a deletion of the stem-loop of exon SK (NMSK*).

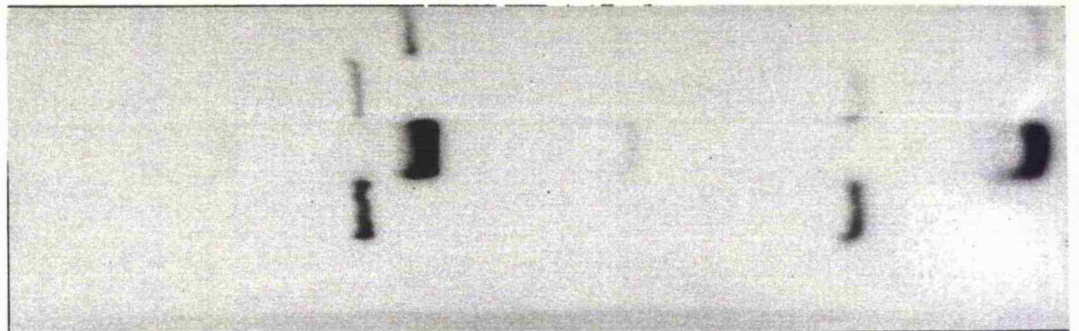
Figure 4.5A. shows the results obtained when these constructs were transfected into myoblasts. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the ImageTM system and associated software, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and it's associated quantification software. The percentage use derived from this data (given in appendix 2) is shown in Table 4.5.1a

4.5A. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 6 hours and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), with wild-type construct (wt) and with a background mutation construct (delstem) are compared to the results obtained from cells transfected with constructs containing: an uprated 5' splice-site of SK (SK5'cons), an uprated branch-point of SK (SKbpup) and an uprated 5' splice-site of exon 4 (4 5'cons).

mock wt NMSK*

u c u c u c



SK5'cons SKbpup 4 5'cons

u c u c u c

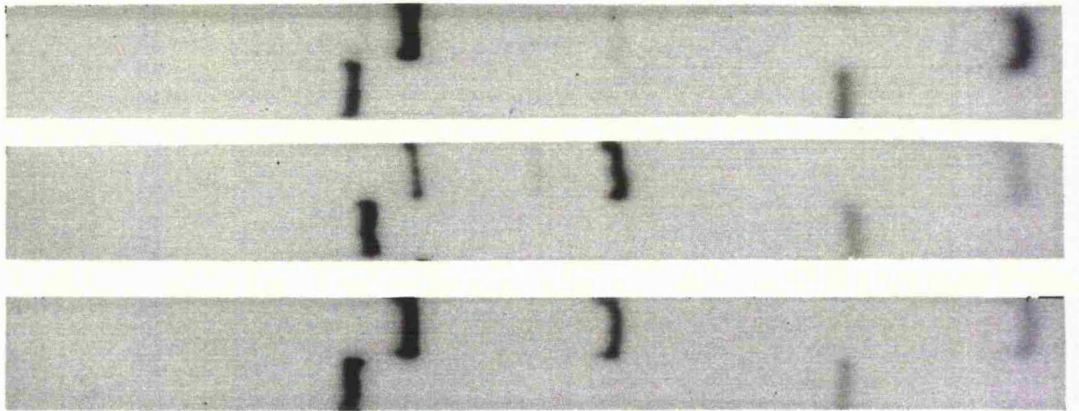


Table 4.5.1a

	Percentage exon use in myoblasts			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
NMSK*	0% ₍₂₎	84%,63%	13%,6%	3%,31%
SK5'cons*	0% ₍₂₎	62%,60%	38%,26%	0%,15%
wild-type	0% ₍₂₎	55% ₍₂₎	6% ₍₂₎	39% ₍₂₎
SKbpup	0% ₍₃₎	30% ₍₃₎	58% ₍₃₎	12% ₍₃₎
2 x bpup	0% ₍₂₎	7% ₍₂₎	74% ₍₂₎	19% ₍₂₎
4 5'cons	0% ₍₂₎	56% ₍₂₎	5% ₍₂₎	39% ₍₂₎

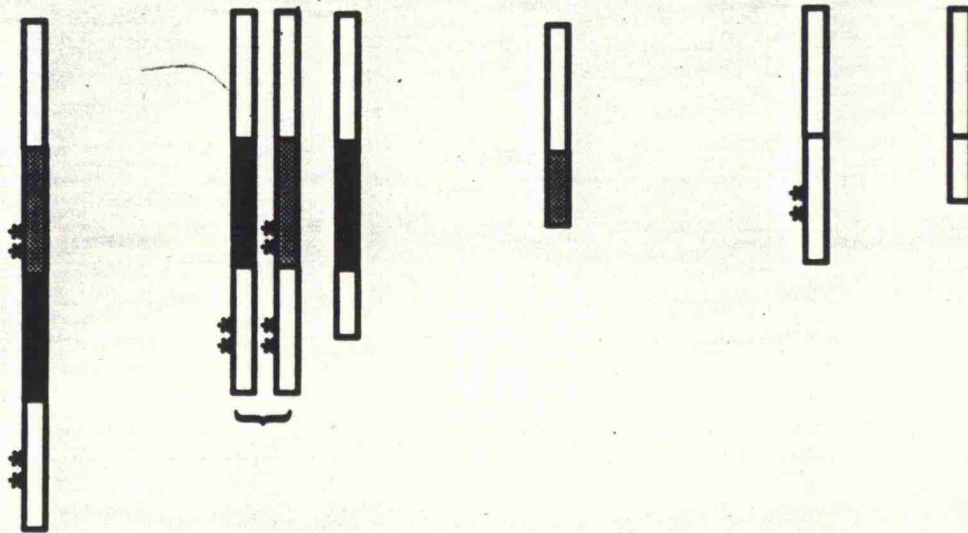
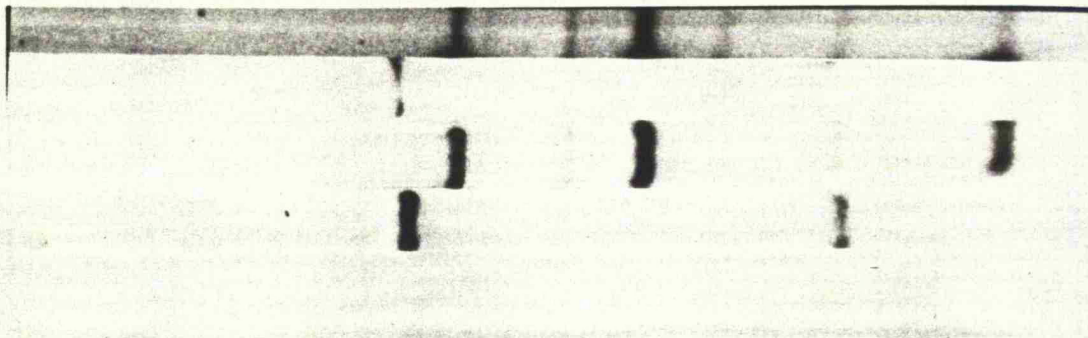
Table 4.5.1a. The percentage isoform use in transcripts derived from wild-type and mutations NM5'cons and NMbpup following transfection into myoblasts. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. Numbers in brackets indicate that the percentage given is the average of duplicate or triplicate transfections.

Figure 4.5B. shows the results obtained when these mutations and the background constructs (NMSK* and wild-type) were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.5.1b.

4.5B. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 92 hours (when the cells had formed myotubes) and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the length of uncut PCR product (u) and AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), with wild-type construct (wt) and with a background mutation construct (delstem) are compared to the results obtained from cells transfected with constructs containing: an uprated 5' splice-site of SK (SK5'cons), and an uprated branch-point of SK (SKbpup).

mock wt NMSK*
u c u c u c



SK5'cons SKbpup
u c u c

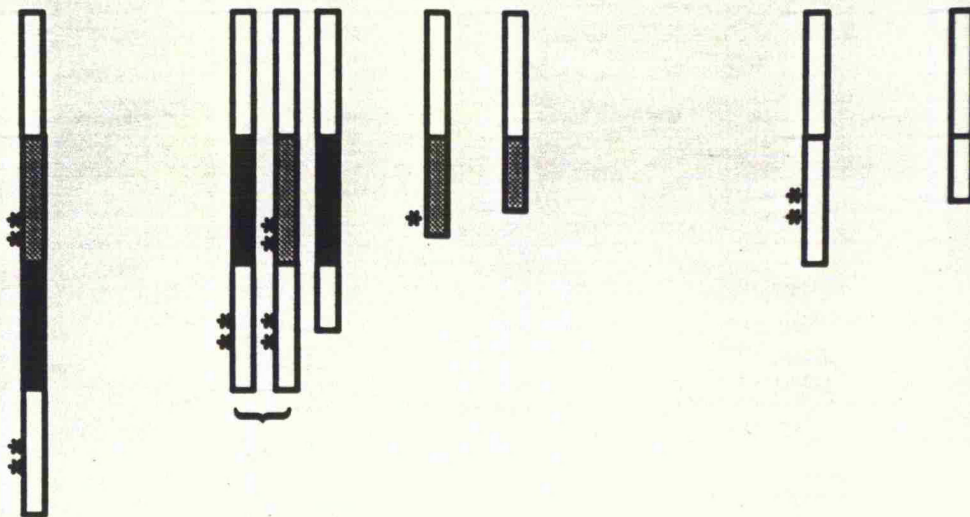
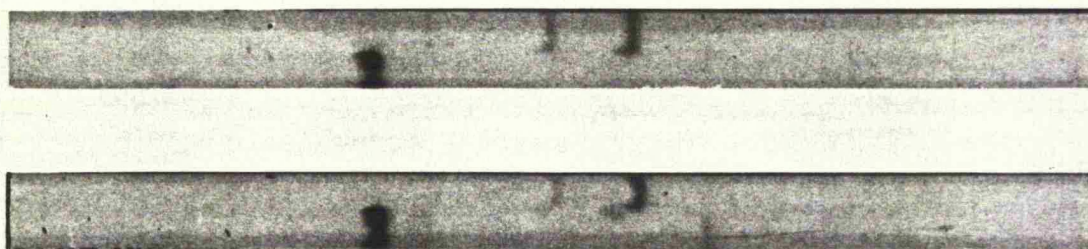


Table 4.5.1a

	Percentage exon use in myotubes			
Mutation	4-NM-SK-6	4-NM-6	4-SK-6	4-6
NMSK*	0% ₍₂₎	29% ₍₂₎	49% ₍₂₎	22% ₍₂₎
SK5'cons*	0% ₍₂₎	15% ₍₂₎	84% ₍₂₎	3% ₍₂₎
wild-type	0%	22%	41%	37%
SKbpup	0% ₍₂₎	0% ₍₂₎	100% ₍₂₎	0% ₍₂₎

Table 4.4.2b. The percentage isoform use in transcripts derived from the SK5'cons and SKbpup mutations compared to their background constructs NMSK* and wild-type respectively, following transfection into myoblasts. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

These data show that in myoblasts and myotubes, improving the 5'splice site and the branch point of exon SK resulted in an increased incorporation of the SK exon. However, in COS cells, although improving the branch-point led to increased use of the SK exon, improving the 5' splice site did not. This may indicate an important difference between the regulation of alternative splicing in true non-muscle cells and muscle cell precursors (myoblasts). Thus, improving the splicing signals does affect the level of incorporation of the SK exon. This indicates that the splice signals of the SK exon are sub-optimal and that this may be a requirement for appropriate incorporation of the SK exon.

4.5.2 Do exons compete for a limited splicing component?

Could the switch from NM to SK use in myotubes, in the absence of SK repressor, be explained by competition between these exons for a *limited* general splicing factor. If this were the case then the deletion of the SK exon, or its splicing signals, in a construct tested in myotubes should result in the use of the NM exon. In myoblasts deletion of the NM exon should result in an increase of jump-spliced products because of the repression of SK. Four constructs were made to test a competition model. Two of the constructs contained deletions of the 5'splice-sites (NM5'del and SK5'del). The third contained a deletion of the branch-point of exon SK (SKbpdel). The fourth contained a deletion of the entire NM exon including 625 nucleotides of the upstream intron and 30 nucleotides of the downstream intron (NMdel). Figure 4.5C. shows the results obtained when these constructs were transfected into myoblasts. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the Image™ system and associated soft-ware, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and it's associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.5.2a

4.5C. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 6 hours and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), and with wild-type construct (wt) are compared to the results obtained from cells transfected with constructs containing: a deletion of the 5' splice-site of NM (delNM 5'), a deletion of the 5' splice-site of SK (delSK5') a deletion of the branch-point of SK (delSKbp) and a deletion of the entire NM exon and some flanking intron sequence (delNM).

NM5'del SK5'del SKbpdel delNM

mock wt

c

c

c

c

c

c



Table 4.5.2a

Mutation	Percentage exon use in myoblasts			
	4-NM-SK-6	4-NM-6	4-SK-6	4-6
wild-type	0%	55% ₍₂₎	6% ₍₂₎	39% ₍₂₎
NM5'del	0%	0% ₍₂₎	8% ₍₂₎	92% ₍₂₎
NMdel	0%	0% ₍₂₎	10% ₍₂₎	90% ₍₂₎
NMSK*	0%	84%,63%	13%,6%	3%,31%
SK5'del*	0%	62%,49%	4%,6%	34%,45%
SKbpdel*	0%	50%	8%	42%

Table 4.5.2a. The percentage isoform use in transcripts derived from NM5'del, NMdel, Sk5'del and SKbpdel (compared to their background constructs: wild-type and NMSK*) following transfection into myoblasts. RNA was isolated 6 hours post-transfection, while the cells were still myoblasts. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

Figure 4.5D. shows the results obtained when these mutations and the background constructs (NMSK* and wild-type) were transfected into myoblasts which were then induced into the myogenic pathway. RNA was harvested 5 days post-transfection when cells had differentiated into myotubes. Quantification of isoform use was achieved by analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.5.2b.

4.5D. Analysis of α_s -tropomyosin mRNA derived from transient transfection of C-2 cells with mutant α_s -tropomyosin mini-genes. RNA was isolated after 92 hours (when the cells had formed myotubes) and analysed by a combined RT-PCR. The effect of each mutation is shown by the relative proportions of the alternative isoforms, identified by the presence of AluI sites in exons 4 and SK, but not in NM. The relative proportion of each isoform is deduced by comparing the AluI cut products (c). The expected positions of each isoform are also illustrated, with open boxes representing the flanking exons 4 and 6, dark hatching the alternatively spliced NM exon, and pale hatching the alternatively spliced SK exon.

The results of cells transfected with no vector (mock), and with wild-type construct (wt) are compared to the results obtained from cells transfected with constructs containing: a deletion of the 5' splice-site of SK (delSK5') a deletion of the branch-point of SK (delSKbp) and a deletion of the entire NM exon and some flanking intron sequence (delNM).

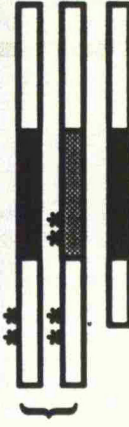
mock wt

u c

24

SK5'del

u u c



SK5'del

u c

NMdel

u c



Table 4.5.2b

Mutation	Percentage exon use in myotubes			
	4-NM-SK-6	4-NM-6	4-SK-6	4-6
wild-type	0%	20%	32%	48%
NMdel	0%	0% ₍₂₎	27% ₍₂₎	73% ₍₂₎
NMSK*	0%	29% ₍₂₎	49% ₍₂₎	22% ₍₂₎
SK5'del*	0%	18%,37%	4%,5%	78%,58%
SKbpdel*	0%	0% ₍₂₎	0% ₍₂₎	100% ₍₂₎

Table 4.4.2b. The percentage isoform use in transcripts derived from the NMdel,SK5'del and SKbpdel mutations (compared to their background constructs wild-type and NMSK*) following transfection into myoblasts. RNA was isolated 5 days post-transfection, when cells had differentiated into myotubes. Numbers in brackets indicate that the percentage given is the average of duplicate transfections.

In myoblasts the results are as predicted for a mechanism involving the repression of SK, by a sequence-specific factor, rather than competition. In myotubes the deletion of the NM exon did not lead to an increase in the level of incorporation of the SK exon; the proportion of transcripts containing SK remained unchanged, but the proportion which contained jump-spliced transcripts increased. When the SK exon was disabled a different outcome was observed depending on the mutation used to inactivate the exon. When the splice site was removed, splicing switched from inclusion of the SK exon to an increase in the proportion with jump-splices; the proportion that included the NM exon remained constant. When the branch-point was deleted all transcripts contained a jump-splice with a dramatic loss of transcripts containing the NM exon.

4.5.3 Does the potential stem-loop play a role?

A potential stem loop had been identified which would sequester the 5'splice-site of exon SK (Graham 1992b). Deletion of the downstream portion of the stem had no effect in COS cells (Graham 1992b). However, it was important to investigate whether this mutation (NMSK*) had an effect in myoblasts and myotubes. Figure 4.4A. shows the results obtained when this constructs and the wild-type construct were transfected into myoblasts. Quantification of isoform use was achieved by the analysis of a video image of the autoradiograph using the ImageTM system and associated soft-ware, and by the phosphoimage of the original polyacrylamide gel using a molecular dynamics phospho-imager, and it's associated quantification soft-ware. The percentage use derived from this data (given in appendix 2) is shown in Table 4.5.3a. The results obtained from duplicate transfections of the stem deletion (NMSK*) in myoblasts cannot be pooled because the data is heterogenous.

Table 4.5.3a

	Absorbance in myoblasts			
Mutation	d	n m	sk	j
NMSK*	0	83	13	3
NMSK*	0	276	27	135

Table 4.5.3a. Results for duplicate transfections of the mutation NMSK* The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm), SK (sk) jump-spliced (j) isoforms from samples cut with AluI.

However, further analysis of the data indicates the reason for the hererogeneity: there are significant differences in the level of jump-spliced product. When the proportions of NM to SK use are compared directly, (as shown in Table 4.5.3b) without the inclusion of jump-spliced product, there is no difference between the sets of data ($\chi_{(1df)}^2=1.3$).

Table 4.5.3b

	Absorbance in myoblasts		
Mutation	d	n m	sk
NMSK*	0	83	13
NMSK*	0	276	27

Table 4.5.3a. Results for duplicate transfections of the mutation NMSK*. The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm) and SK (sk) isoforms from samples cut with AluI.

This indicates that the level of jump-splicing is irreproducible, but that the relative proportions of NM to SK are consistent

If the delstem transfections into myoblasts are now compared to those produced by transfection with a wild-type construct (shown in Table 4.5.3c), there is no significant difference ($\chi_{(2df)}^2= 2.1$) in use of the NM and SK exons.

Table 4.5.3c

	Absorbance in myoblasts		
Mutation	d	nm	sk
wild-type	0	170	22
NMSK*	0	83	13
NMSK*	0	276	27

Table 4.5.3a. Results for duplicate transfections of wild-type construct, and the mutation NMSK*. The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm) and SK (sk) isoforms from samples cut with AluI.

Therefore, the delstem mutation did not alter the proportion of NM and exon SK use relative to each other in myoblasts, but did cause a drop in the level of jump-splicing.

The effect of the delstem mutation was also tested in myotubes, the results for duplicate transfections are given in Table 4.5.3d.

Table 4.5.3a

	Absorbance in myotubes			
Mutation	d	nm	sk	j
NMSK*	0	70	121	54
NMSK*	0	13	30	10

Table 4.5.3d. Results for duplicate transfections of the mutation NMSK*. The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm), SK (sk) jump-spliced (j) isoforms from samples cut with AluI.

However, unlike the data from the myoblasts, the delstem transcripts from transfected myotubes had a more consistent level of jump-slicing (there is no significant difference between these data $\chi^2=0.8$) such that the data are homogenous and could be pooled. This pooled data can be compared to that obtained from transfection of the wild-type construct. The results are given in Table 4.5.3e

Table 4.5.3a

	Absorbance in myotubes			
Mutation	d	n m	sk	j
wild-type	0	66	104	156
NMSK*	0	83	151	64

Table 4.5.3d. The combined results for the duplicate transfections of the mutation NMSK*. The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm), SK (sk) jump-spliced (j) isoforms from samples cut with AluI.

A comparison between the mutant and the wild-type constructs shows that the level of jump-splicing is lower with the delstem mutation. However, the proportion of NM to SK containing transcripts is the same (shown in Table?)

Table 4.5.3a

	Absorbance in myotubes		
Mutation	d	n m	sk
wild-type	0	66	104
NMSK*	0	83	151

Table 4.5.3d. The combined results for the duplicate transfections of the mutation NMSK*. The absorbance readings for the intensity of bands corresponding to double incorporation (d) was deduced from uncut samples; incorporation of NM (nm) and SK (sk) isoforms from samples cut with AluI.

Statistical analysis of these data indicates that there is no significant difference ($\chi^2 = 1.1$) between transcripts encoded by the wild-type and the NMSK* mutation in exon NM and SK use.

Therefore, in myotubes, as in myoblasts, a deletion of the down-stream portion of the potential stem-loop did not alter the relative proportion of NM to SK but did cause a drop in the level of jump-spliced transcripts.

Chapter 5

Discussion

5.1 Development of methods

5.1.1 Development of RT-PCR methods.

The development of a combined reverse-transcription/polymerase chain reaction method (Eperon and Hamshire 1993) was central to the analysis of alternative splicing of N-CAM and α_5 -tropomyosin pre-mRNA. Two basic techniques were employed in the analysis of the heterogeneous DNA, formed as a consequence of RT-PCR amplification. Either the products were cloned into m13 derived vectors and sequenced, or the products were re-amplified using internal PCR primers and separated on polyacrylamide gels by size. The later method was developed to address certain reservations concerning the cloning and sequencing protocol. First, although the method was suitable for a limited number of investigations (e.g. the incorporation of the MSD of N-CAM in myoblasts and myotubes (Hamshire et al 1991), I thought it would be unwieldy for the analysis of large numbers of RNA samples. These would be produced by either a developmental analysis of alternative splicing in different tissues (as was undertaken for N-CAM (Hamshire et al 1991) or for a thorough investigation of the effect of mutations of mini-gene constructs in transient transfections in COS cells, myoblasts and myotubes (Graham et al 1992; this thesis). Second, and more important, the quantification of alternative isoform use was an important part of this investigation. The cloning step provided a potential source of error which would have been difficult to assess: it was possible that different products (although with identical terminal sequences) would be cloned with different efficiencies. This could be the consequence either because of the sequence of the product (early attempts to clone RT-PCR products from N-CAM were unsuccessful in Rec A⁺ strains of *E.coli*, but were successful in Rec A⁻ strains) possibly by the introduction of sequences which could be lethal or recombine with the vector, or because of the different length of the products; shorter products perhaps being more favourably incorporated into vector than longer

products. However, the direct re-amplification protocol could be tested for its ability to maintain input isoform ratios. This was achieved by the co-amplification of different cDNAs (12-13 and 12-a-b-c-13; derived from the original cloning and sequencing protocol) in a series of amplifications in which the input concentrations of each isoform was varied systematically. The addition of α -³²PdATP to the reaction mixture allowed for the identification and quantification of the different products on a polyacrylamide gel. A constant level of a different cDNA (12-a-13) was also incorporated into each reaction to act as an internal standard. An internal standard was necessary in order that differences in the amount of sample loaded onto the gel could be corrected. This allowed me to establish not only that input ratios were maintained over a wide range of isoform ratios (from 61:1 to 1:160 for 12-13:12-a-b-c-13), but also that the amplification efficiency was constant at different input concentrations.

A comparison of results obtained by the two protocols for the identification of MSD exon use in N-CAM transcripts showed that no difference was noted between the results obtained from myoblasts, but differences were apparent in the results obtained from myotubes. This appeared to be due to a relative lack of the isoform 12-a-b-c-AAG-13 in the cloning protocol and an excess of the isoform 12-13 using the direct re-amplification method. This finding indicated either that one of the methods was not maintaining original input RNA ratios in myotubes, or that the RNA used for the two sets of data had a slightly different isoform profile, although the RNA was isolated at a similar stage of development. I thought the latter was the more likely explanation, but, as I could not exclude the possibility that a cloning and sequencing protocol might affect the final ratios of isoforms, this protocol was not employed in future analyses.

The use of poly (A)⁺ selected RNA enabled the amplification of scarce endogenous N-CAM RNA. Several methods were assessed for their capacity to isolate mRNA from total RNA, all of which gave comparable results although at widely different cost. The Hybond mAP paper from Amersham was found to be the most practical method. The more abundant α_s -tropomyosin mRNA could be amplified directly from total RNA, without the requirement for poly (A)⁺ selection.

The use of internal "nested" primers for the direct re-amplification protocol significantly reduced the level of non-specific background amplification. In the re-amplification of N-CAM products a single nested primer was used, whereas for analysis of the α_s -tropomyosin both primers for the re-amplification were internal to the original RT-PCR primers.

The direct analysis of RT-PCR products from N-CAM transcripts was achieved by separation on polyacrylamide gels, by virtue that all combinations of exon use would produce a different length PCR product. However, consistent results were only obtained using native gels rather than separation on denaturing gels either with or without the addition of 20% formamide. When denaturing gels were used the products had a tendency to migrate in a position which was unrelated to their actual length (as deduced from co-separation on the gel of size-markers), although they would migrate at the expected position if the gel was run at a very high voltage such that it was physically hot. It was thought that the reason for this aberrant migration may have been due to the small size of some of these products (less than 100bp). It was possible that following the pre-gel denaturation of products (incubation at 80 °C in F-dyes for 5 minutes) the products could reanneal either to themselves or to other products by virtue of their common sequences, even though samples were loaded immediately onto the gel.

Unless the polyacrylamide gel was run at a high voltage, duplex or partial duplex molecules could form, or persist during the electrophoretic separation of the products. This mechanism forms the basis for identification of single point mutations by single-strand conformational polymorphisms (SSCP). This is not the same as the identification of heteroduplex molecules produced as a consequence of aberrant re-annealing during the final rounds of a PCR reaction (Zorn and Kreig 1991) in which the aberrant products of the PCR migrate in an unexpected position on native gels, but are undetectable under strong denaturing conditions. Thus the generalised errant migration on denaturing gels, but correct migration on native gels was presumed to be due to duplex formation when denaturing gel electrophoresis was used, and not due to heteroduplex product formation during the PCR. However, two errant bands were obtained during native gel electrophoresis of N-CAM RT-PCR products. These were shown to contain the isoforms 12-13 and 12-a-b-c-13 by a PCR-based sequencing strategy, and they were included in the quantification of the final isoform ratios as such. Recently, heteroduplex formation during the latter stages of a PCR has been demonstrated (Zorn and Kreig 1991). It is possible that these bands were formed as a consequence of this, and that the PCR based sequencing strategy would have identified only one of the partners of a heteroduplex. Therefore, the proportion of isoforms 12-13 and 12-a-b-c-13 may have been slightly overestimated. However, re-analysis of the data indicates that the overestimate is small and has no significant effect on the overall shift in patterns of isoform use which is observed during the development of skeletal muscle.

In developing the method for direct analysis of RT-PCR products from α_s -tropomyosin, an alternative method had to be devised in order to assess the level of incorporation of the NM versus the SK exon in α_s -tropomyosin as both isoforms would yield PCR products of identical length. Initially I

investigated the possibility that a restriction endonuclease would recognise a sequence within one of the exons, but not the other. Analysis of the restriction enzymes which fulfilled the criteria for cutting in one exon but not the other revealed that only a few enzymes could be used. Also, a control would have been required for each reaction to ensure that the enzyme had cut to completion. However, AluI sites were identified within exon 4 and exon SK, but not within exon NM. The position of these sites indicated not only a mechanism to differentiate between the alternative isoforms, but gave a perfect internal control for complete cleavage by the enzyme: all products would be truncated (by cleavage in exon 4) if the enzyme cut to completion.

This modified RT-PCR method has provided a means for the quantitative analysis of RNA molecules, which would not have been feasible by more traditional methods such as S1 mapping, screening of cDNA libraries or northern blotting.

5.1.2 Development of a system for analysis of sequences determining alternative splicing in muscle cells.

Several problems have been noted by other workers in their attempts to produce transient transfection assays in COS cells (Guo et al 1991) and in myoblasts and myotubes (Libri et al 1989), which had led to an emphasis on the investigation into the regulation of the non-muscle pattern of alternative splicing in non-muscle cells, rather than the regulation which produces the muscle-type splice in muscle cells. Therefore, the development of a simple transient transfection method, which could be used to study the regulation of alternative splicing in differentiated muscle cells, was of great importance. The development of such a method was central to my investigation of the regulation of alternative splicing of α_s -tropomyosin in myoblasts (which normally incorporate the NM exon) and myotubes (which incorporate the SK

exon). This would not only provide important information about the regulatory mechanisms of α_s -tropomyosin, but address the wider question as to whether heterologous genes shared common regulatory mechanisms.

In order to produce a method for successful transfection of myoblasts and myotubes (Eperon and Hamshire 1993), several modifications to standard transfection methods were employed. First, a modified calcium phosphate/DNA co-precipitation method, already a standard method within this laboratory for the transfection of COS and HeLa cells, was further adapted by the inclusion of a glycerol shock rather than a DMSO shock to encourage uptake of vector DNA. Second, post-transfection cells were maintained in medium which contained sodium butyrate. Sodium butyrate enhances transcription from exogenous DNA (Sambrooke et al 1989). Post-transfection myotubes maintained in a medium which had not been supplemented with sodium butyrate lost the ability to produce α_s -tropomyosin transcripts. Third, the type of precipitate that was formed was also observed to be important. Heavy precipitates appeared to yield the best transfections, light precipitates (produced when solutions were not fresh) gave poor transfections, as determined by RT-PCR of RNA derived from the transfected vector. Poor transfection of myoblasts resulted in a loss of vector encoded RNA in differentiated myotubes.

It has been reported (Guo et al 1991) that over-expression of the vector in a non-muscle cell can itself cause a switch in splicing preference, indicating a possible role for a repressor factor which can be titrated out. In order to test this possibility, a series of concentrations of vector were transfected into myoblasts. No such switch in splicing preference was observed even when 40 μ g of vector was used. However, there does appear to be a trend towards exclusion of both exons (jump-splicing) at higher concentrations. It is

interesting to speculate that this could be due to the titration of an NM activator factor. However, all concentrations produced a fairly high level of jump-spliced product. Subsequent transfections used a concentration of 5 μ g; the lowest concentration for which reproducible results could be obtained.

Jump-spliced products were seen when the wild-type mini-gene construct was expressed in both myoblasts and myotubes, but were not usually seen following transfection of COS cells. However, the expected isoform switch from inclusion of NM to inclusion of the SK exon upon myotube formation, at the expense of the NM exon, could still be superimposed on a constant level of background jump-splicing. It was apparent that the mechanism which controlled this switch was operating in this system, and could be investigated by the transfection of mini-gene mutations. The raised level of jump-spliced product has, however, indicated two potential points for control which were not apparent in the transfection of COS cells. First, it may indicate that the transfection of a human α_s -tropomyosin gene into a mouse cell line contained some partial incompatibility. This could be due to an incompatibility of either a general splicing factor, such that there was a reduced efficiency of all splicing, or a factor specific to the alternatively spliced region. The presence of a *cis*-acting repressor had been implicated by the results obtained in COS cells (Graham et al 1992). Thus in myoblasts, incompatibility of repressor should result in an increase in the proportion of transcripts which contained the SK exon. In myoblasts the SK exon was incorporated in about 20% of transcripts. However, the incompatibility of a repressor should also reduce, not increase, the level of jump-spliced product. Incompatibility of a *trans*-acting activator would, however, cause an increase in jump-splicing. Evidence from the transfection of increasing concentrations of wild-type mini-gene construct, and from transfections of mutant constructs (discussed below) indicated that this may have been the case. Also, the

deletion of a region of intron downstream from exon SK (NMSK*) caused the level of jump-splicing to fall dramatically, the significance of this is discussed later. This phenomenon, and the significance for the "stem" region in the control of alternative splicing may have been revealed only because the muscle cell system was operating to the limit of its capacity.

An alternative hypothesis was that the α_s -tropomyosin homologue in the mouse also produced jump-spliced product. This was tested by RT-PCR of endogenous mRNA (from mock-transfected cells) using human α_s -tropomyosin directed oligonucleotides. RNA from myoblasts, myotubes and HeLa cells all gave a similar length product which contained AluI restriction sites at the same position. Myotubes also produced a second isoform which contained AluI sites at a position identical to that predicted from the human SK exon sequence. No jump-spliced product was seen in any of these samples. The jump-spliced product was therefore assumed to be due to a partial incompatibility of splicing of human α_s -tropomyosin and the splicing factors (general or specific) in a mouse cell line.

5.2 Analysis of splicing in the MSD region of mouse N-CAM.

Part of the reason for the investigation of alternative splicing patterns of the MSD in the mouse was a discrepancy between the data derived from the screening of a human muscle cDNA library (Dickson et al 1987) and the sequencing of a genomic clone from the mouse (Santoni et al 1988). Sequencing of the mouse genomic fragment identified the MSD1a homologue in a position about 1kb downstream of the constitutively spliced exon 12, but indicated a putative site for the triplet AAG exon with no evidence of the homologues to exons MSD1b and MSD1c in between. Sequence analysis of a cDNA library from the chicken also indicated a

discrepancy between the human and the chicken MSD (Prediger et al 1988). Re-alignment of the sequence between exons 12 and 13 in the chicken indicated that the MSD homologue to exon MSD1b was 15 base-pairs shorter. Therefore, it was important to discover whether the mouse gene did not contain the MSD1b and MSD1c homologues (as would be predicted from the genomic clone), contained truncated versions of any exon (as found in the chicken), or did contain homologues to the human exons (indicating that the putative site for the triplet AAG exon in the mouse genomic clone was not the site used by the gene). RT-PCR and cloning and sequencing of the products answered these queries. The mouse N-CAM gene did contain exons with homology to the MSD1b and MSD1c that were the same length as the human MSD exons.

5.2.1 Discrete and coordinated regulation of the MSD: in muscle cells

In C-2 myoblast cells the predominant isoforms of N-CAM excluded the MSD exons, although there were substantial levels of the triplet exon AAG. During fusion of the cells the incorporation of AAG declined transiently and MSD1a appeared at a low level; later, in myotubes the MSD1b and MSD 1c were incorporated almost exclusively together and in conjunction with MSD1a. Thus, during development of muscle, the MSD exons were present as a coordinated unit. In neural cells, most transcripts did not contain exons of the MSD, although low levels of the single MSD1a exon were seen. From these data, I concluded that the incorporation of exon MSD1a is regulated independently of the incorporation of exons MSD1b and MSD1c, but that inclusion of the latter two requires MSD1a. Incorporation of the triplet AAG exon does not appear to require or be coordinated with the MSD exons.

5.2.2 Effect of denervating muscle

In RNA isolated from skeletal muscle seven days after surgical denervation the predominant isoform was 12-a-b-c-13; little inclusion of the triplet exon AAG was detected, and the isoform 12-13 had almost disappeared. Surgical denervation of adult skeletal muscle is known to cause a marked increase in N-CAM expression (Moore and Walsh 1986). Therefore, it is very likely that the apparent switch in splicing results in part from the nerve-dependent transcriptional activation of multinucleated myofibres, such that these cells increase their contribution to the N-CAM mRNA isoforms isolated from these tissues. Likewise, the contributions of Schwann and satellite cells might be expected to alter in response to nerve degeneration and muscle atrophy. It is also possible, that the destruction of a particular cell type (i.e. nerve) within the complex tissue causes a loss of a *trans*-cellular factor which induces or maintains the alternative splicing pattern of a neighbouring (i.e. muscle) cell (see below).

5.2.3 Discrete and coordinated regulation of the MSD: in neural cells

During development of the mouse brain there appears to be an increase in the proportion of transcripts which contain exon MSD1a, although there was no evidence for the inclusion of exons MSD1b and MSD1c. More interesting was the differences noted for exon usage in cell lines derived from specific rat neural cell lines. All of these cells appeared to include the triplet exon AAG to a greater extent than mouse brain cells. This could represent a species difference, a consequence of immortalisation of the cell lines, or a consequence of a lack of interaction between different cell types which might influence their overall expression pattern. If the lack of an influence of surrounding (but different) cells is invoked as an explanation for the different profile of isoform expression exhibited by these cell lines, this could indicate that a *trans*-acting factor produced by one cell type influences the splicing

pattern of the MSD region of N-CAM in another cell. This would be analogous to the situation induced by Wieczorek in which muscle cells which were supplemented with neural cell extracts gave an altered splicing profile (Wieczorek 1988). Therefore, inter-cellular signalling by factors which then influence the splicing patterns of heterologous cells could be an important mechanism for regulation of alternative splicing, particularly in complex differentiated tissues.

5.2.4 Identification of putative *cis*-acting elements

This study, and others (Santoni et al 1989) indicated that exon MSD1a could be incorporated into a mature transcript in the absence of exons MSD1b and MSD1c (and was found in both muscle and neural cells) but that MSD1b and MSD1c were generally incorporated with the MSD1a exon as a unit (a single exception was identified by the sequencing of one clone which contained the isoform pattern 12-c-AAG-13). This raised the possibility that a single factor was used by exons MSD1b and MSD1c for their coordinated incorporation, either as a repressor to prevent incorporation of the exons in non-muscle cells, or as an activator to enhance their incorporation in skeletal muscle. If a single factor was involved, then a common binding site sequence would be predicted. This binding site could be either shared with other alternatively spliced exons such that a common mechanism was invoked, or unique to the MSD: the alternative splicing of the MSD being regulated by a transcript specific mechanism.

Few sequences of *cis*-acting elements, which are involved in the regulation of alternative splicing in mammals, have been formally identified (Strueli and Saito 1989; Tsai et al 1989; Cooper and Ordahl 1989; Guo et al 1991; Libri et al 1992). Only one of these sequences (an imperfect repeat with the sequence YRYYYYRYRY, where Y=pyrimidine and R=purine), identified in the rat β -

tropomyosin gene just upstream of exon 7 between the exon and the branch-point (Guo et al 1991), shares any similarity with exon sequences of the MSD. A single copy of this sequence (TGTTCCGTTGT) is found in MSD1b. However, the significance of this sequence is doubtful especially as this sequence would arise by chance once in every 2048 bases in random DNA, and the identified sequence was isolated as multiple copies in intron, rather than as a single copy in an exon sequence.

Another possibility was that an MSD-specific sequence was present in exons MSD1b and MSD1c which regulated their co-ordinated inclusion. A search within the exons of the MSD revealed 7 sequences which could have been imperfect copies of a 12 base pair repeat, one copy of which was located in exon MSD1a, three copies were found in MSD1b and three in MSD1c. These sequences were then analysed in two ways. First, it was possible that any similarity between the exons of the MSD was due to common ancestry, the exons being the consequence of exon duplication events. This was analysed by two cluster analyses which identified the degree of similarity between the 7 sequences and produced maximum parsimony trees which would indicate the possible evolutionary tree of these sequences. Two methods, simple cluster analysis and the more complex unweighted pair group method with arithmetic mean, gave broadly similar results. The evolutionary history predicted by these methods was difficult to attribute to exon duplication events unless gene conversion within the exons and recombination events between sequences within the exons were invoked. Also, an analysis of the number of back mutations, and the type of mutations gave further evidence that these sequences were unlikely to have arisen because of shared common descent.

Second, having reasoned that the sequences were not common because of evolution from a common ancestor, the possibility that they had arisen by chance needed to be addressed. This was achieved by analysing the chance that seven random sequences would contain this number of identical bases in the same position (i.e. the strength of match to a consensus sequence). If all the sequences were identical a consensus score of 84 would be produced (7 x 12). A score of 51 was obtained for these sequences. The chance that a score of 51 would be obtained from seven random sequences is 1 in 17,074. At first sight, this result indicated that these sequences were not the consequence of random distribution. However, the seven twelve base-pair sequences chosen were only one of 40,080 possible combinations of twelve base-pair sequences which could be produced from the exons of the MSD. Unless mutational studies indicated that these sequences play a role (in splicing), and are thereby fixed at a single location the possibility that they have arisen by chance must remain.

5.2.5 Possible role for MSD exons

All combinations of splicing patterns of the MSD exons would maintain the reading frame, but it is impossible to predict whether the minor isoforms produced by alternative splicing would exencode functional N-CAM proteins. The physiological role of the protein domain encoded by this region has not been studied in detail, although incorporation of MSD1a may introduce a 'hinge' into the molecule, and the domain as a whole is known to undergo significant post-translational modification, notably O-linked glycosylation. Inclusion of the MSD as a unit has also been shown to induce enhanced myogenesis in myoblasts which have been transfected with full length cDNA encoding the MSD (Dickson et al 1990). Inclusion of the triplet AAG exon results in an inclusion of a single amino acid into the primary structure. In place of arginine, which is found in all the isoforms which do not include

the AAG exon, glutamine and glycine are encoded by the isoforms 12-AAG-13 and 12-a-AAG-13, and lysine and glycine by 12-a-b-c-AAG-13. If the inclusion of the AAG exon is indeed regulated, these subtle changes in the amino acid structure may represent an important point for modulation of cell-cell interaction which is mediated by the regulation of alternative splicing of an unusually short exon in N-CAM. It is interesting to speculate how the regulated inclusion/ommission of such a short exon could be achieved. With only three nucleotides of exon sequence, any *cis*-acting element must be located within flanking intron sequences, possibly within the splice-sites themselves.

5.2.6 Genomic structure of the MSD region of N-CAM

Initially the analysis of regulation of alternative splicing of the MSD exons of N-CAM was to form part of a collaboration into the wider implications for the alternative splicing of the MSD. This was to include an investigation into the regulation of alternative splicing, a continued identification of the genomic structure of the gene, and the physiological role for the MSD in cell adhesion. Unfortunately, the emphasis of the research by the other partners in the collaboration switched towards isolation and characterisation of the promoter region, rather than the 12-13 MSD containing region of N-CAM. Therefore, the isolation of genomic clones which included the entire 12-13 region was not achieved, as had originally been hoped.

In order to assess the mechanism which regulated the alternative splicing of this region, a mini-gene system which incorporated both intronic and exonic sequences from the MSD was required. I was concerned that isolation of MSD containing genomic clones would be problematic, as each exon is very short. Indeed several attempts to map the positions of the MSD exons by probing a Southern blot of genomic DNA fragments cut with specific endonucleases,

with end labelled oligonucleotides were unsuccessful. I thought that isolating positive clones from a library via colony hybridisation techniques was also likely to be unsuccessful. This led to an attempt to isolate and amplify the genomic region between exons 12 and 13 by PCR amplification. Several PCRs were attempted using a variety of conditions. The only successful amplifications were between exons 12 and MSD1a and between exons MSD1b and MSD1c. Unfortunately, all attempts to sequence these products either directly (by a PCR based sequencing strategy) or after cloning failed. The cloning and sequencing of the MSD1b to MSD1c fragment provided sequences which corresponded not to genomic DNA but were presumed to have been derived by the reverse transcription action of Taq polymerase on contaminating RNA. These sequences contained no intron and were identical to the sequences derived from the earlier cloning and sequencing results for the identification of the patterns of expression of the MSD in myotubes. The contamination of the ligation reaction with products as short as those produced by RT-PCR of mRNA which contained exons MSD1b and MSD1c was unexpected, as the product of the PCR reaction had been separated on a ChromaSpin 400 column in order to remove both unincorporated primers and any such contaminating product. At this point, a decision was made to switch the emphasis to the regulation of alternative splicing of a gene for which the genomic fragment was available and of a manageable size. At the same time, there was tentative evidence that distal sites could affect the regulation of splicing. Therefore, a mini-gene construct for N-CAM, which it appeared would have had to contain large deletions in at least two of its introns, may have led to the loss of relevant *cis*-acting elements. The decision was taken that the regulation of alternative splicing in muscle cells could be studied better with the central exons of human α_s -tropomyosin.

5.3 Alternative splicing in human α_s -tropomyosin

The SK and NM exons of human α_s -tropomyosin are alternatively spliced in non-muscle and muscle cells, with inclusion of the NM exon predominating in non-muscle and the SK exon in skeletal muscle. The mutually exclusive use of one of two exons in different tissues can be achieved by one of three general mechanisms. First, by interfering with the mechanism of constitutive splicing. This might be mediated by the formation of a stable secondary structure or by the binding of a *trans*-acting repressor or by the incompatibility of signals due to mutual interference. Second, by specific activation of splicing of one of the exons if the exons were otherwise unusable. The activator could interact with the exon or with intron sequences to promote splicing (possibly by activating "dormant" splice-site signals). Third, the exons could be in competition for a limited resource. This resource could be a splicing factor, the splice-site of one of the flanking constitutive exons, or time. If a splicing factor is limited, then an exon which has a low disassociation rate is more likely to be associated with that factor than an exon with a higher disassociation rates and would therefore be spliced to a neighbouring exon, to the exclusion of the exon with a lower affinity for the factor. However, the factor would have to be present at equimolar concentrations, with respect to the transcript, or less, to ensure that the incorporation of both exons, in the same transcript, was unlikely. If the competition is between the splicing of alternative exons to the flanking constitutive exons then it is also important to remember that the flanking constitutive exons are also involved in the competition. The competition is not four-way between, say, the NM (upstream) and SK exons splicing to exons 4 or 6, but five-way between the splicing of the two alternative exons to the constitutive exons and jump-splicing (4 to 6) of the constitutive exons. The concept that time is a limited resource is important if relative strengths of splice sites are invoked as the control point for the competition. If time were unlimited (i.e. splicing factors

assembled rapidly at exon/intron borders but splicing commitment was relatively slow) then all exons with competent splicing signals would be incorporated into the mature transcript. If time is limited then only exons with splice-sites which form sufficiently stable associations with splicing components would be available for splicing. Exons with poor splice sites would be lost by the splicing together of flanking exons. If competition is the controlling mechanism in one tissue, a different mechanism must account for the use of the "weaker" exon in the alternative tissue.

These mechanisms can be summarized as follows:

1. Interfering with constitutive splicing
 - a) via secondary structure
 - b) via exon repressor sequences
 - c) via mutual interference
2. Activation of unusable exons
 - a) via exon sequences
 - b) via intron sequences
3. Competition for a limited resource
 - a) which could be a splicing factor
 - b) which could be another splice-site
 - c) which could be time

These mechanisms were investigated for their role in the regulation of alternative splicing of the NM and SK exons of α_s -tropomyosin by transient expression of α_s -tropomyosin mini-gene mutants in non-muscle COS cells, pre-muscle myoblasts and myotubes.

5.3.1 Identification of signals which regulate expression in non-muscle (COS cells and myoblasts)

These were identified by transfection of mutant constructs into mouse C-2 myoblasts and subsequent analysis by RT-PCR, and by the re-analysis of RNA samples which had been produced by the transient expression in monkey COS cells (Graham 1992).

A mechanism based on the formation of stable secondary structures (1a) was tested by the transfection of constructs which contained exon rearrangements (SKSK*, (--)NM*, SKNM*). If formation of a stable secondary structure around the SK exon was a key point in suppression of SK incorporation, then placing the SK exon in the natural NM site should allow for its incorporation. This was not observed, the SK exon was never incorporated at the same level as the NM exon when placed in the natural NM site in either COS cells or myoblasts. These results also invalidate a mechanism in which the natural SK site requires activation via *cis*-acting sequences within the intron (2b). Under such a mechanism, swapping the exons (SKNM*) would cause use of the (upstream) SK but not the (downstream) NM exon. This was not found in either COS cells or myoblasts. Mutual interference of these exons, as a mechanism which prevents their splicing to each other (mechanism 1c), had been tested by transfection of COS cells with the construct delSK-6 and analysis by S1 mapping (Graham et al 1992). However, an ambiguity had arisen concerning the level of splicing of SK to NM. I analysed the RNA by RT-PCR which established that there was no intrinsic barrier to the splicing of NM with SK. *In-vitro* splicing experiments (I.Eperon personal communication) and analysis of the genomic sequence upstream of SK indicated that unlike the mutually exclusive exons 2 and 3 in rat α -tropomyosin (Smith and Nadal-Ginard 1989) the branch-point of exon SK is almost certainly accessible for formation of a lariat with the 5' splice-site of NM. The possibility that the exons were in competition (mechanisms 3a, 3b

5.3.1 Identification of signals which regulate expression in non-muscle (COS cells and myoblasts)

These were identified by transfection of mutant constructs into mouse C-2 myoblasts and subsequent analysis by RT-PCR, and by the re-analysis of RNA samples which had been produced by the transient expression in monkey COS cells (Graham 1992).

A mechanism based on the formation of stable secondary structures (1a) was tested by the transfection of constructs which contained exon rearrangements (SKSK*, (--)NM*, SKNM*). If formation of a stable secondary structure around the SK exon was a key point in suppression of SK incorporation, then placing the SK exon in the natural NM site should allow for its incorporation. This was not observed, the SK exon was never incorporated at the same level as the NM exon when placed in the natural NM site in either COS cells or myoblasts. These results also invalidate a mechanism in which the natural SK site requires activation via *cis*-acting sequences within the intron (2b). Under such a mechanism, swapping the exons (SKNM*) would cause use of the (upstream) SK but not the (downstream) NM exon. This was not found in either COS cells or myoblasts. Mutual interference of these exons, as a mechanism which prevents their splicing to each other (mechanism 1c), had been tested by transfection of COS cells with the construct delSK-6 and analysis by S1 mapping (Graham et al 1992). However, an ambiguity had arisen concerning the level of splicing of SK to NM. I analysed the RNA by RT-PCR which established that there was no intrinsic barrier to the splicing of NM with SK. *In-vitro* splicing experiments (I.Eperon personal communication) and analysis of the genomic sequence upstream of SK indicated that unlike the mutually exclusive exons 2 and 3 in rat α -tropomyosin (Smith et al 1989) the branch-point of exon SK is almost certainly accessible for formation of a lariat with the 5' splice-site of NM. The possibility that the exons were in competition (mechanisms 3a, 3b and 3c) was

discounted by testing a construct containing a deletion of the NM exon. Inclusion of the SK exon was not the default pattern. However, perturbation of exon use could be achieved by uprating the splice-sites surrounding the exons (NM5'cons, SK5'cons and SKbpup). This may have indicated a role for a competitive event in non-muscle cells. However, I think that these results were obtained because of the underlying sub-optimal nature of the splice-sites; these are exploited as a control mechanism in myotubes (see below), but play no direct role in the regulation of alternative splicing in non-muscle cells. Thus, mechanisms based on secondary structure (1a), mutual interference (1c), activation of a quiescent exon via intronic *cis*-acting elements (2b) and competition (3a, 3b and 3c) were not controlling expression of the non-muscle isoform 4-NM-6 in COS cells and myoblasts. Therefore, a mechanism involving SK repression via sequences within its exon (1b) and/or activation of NM via *cis*-acting sequences within its exon (2a) must be invoked.

The position of a putative *cis*-acting repressor was identified by a series of mutations in the SK exon: two regions were identified from the COS cell transfections (1-15 and 61-75), and three regions from transfections in myoblasts (1-15, 46-60 and 61-75). The much stronger effect of mutations at the 3' end of the SK exon in myoblast transfections may indicate either that the repression of the SK exon in kidney and myoblast is not the same (although both exhibit non-muscle tropomyosin phenotypes), or that the repressor factor(s) have different sequence specificity in monkey COS cells and mouse C-2 cells.

In COS cells the mutation of repressor sequences (1-15 and 61-75) resulted in the double incorporation of NM and mutated SK sequences in a significant proportion of transcripts (40% and 31% respectively). This result is predicted

by a model in which the downstream site would be available to the splicing machinery, but is repressed because of an exon-specific repressor element. Relief of that repression led to the double incorporation of exons (NM+SK). In myoblasts, no such double incorporation was noted for the mutations of the repressor sequences (1-15, 46-60 and 61-75). The difference between these data could be attributed partly to the high background level of jump-spliced product which was produced by the wild-type mini-gene in myoblasts, although the background mutation for these constructs only produced 8% jump-spliced transcripts. Therefore, some incompatibilities may have been inherent in the system; a reduced splicing efficiency of these two exons was the consequence. In such a situation, enabling the use of the downstream site by mutation of the repressing exon sequences resulted in the loss of jump-spliced products rather than formation of double incorporated (NM+SK) transcripts. However, this does not fully explain the lack of double incorporation in myoblasts, as would be predicted from the results of COS cell transfections, and may indicate that a mechanism is operating in myoblasts which prevents the splicing together of NM and SK.

An unusual and unexpected result from the mutational analysis of repression of the SK exon in both COS cells and myoblasts was the apparent re-repression of SK exon use in the construct where 15 bases from NM were swapped with positions 16-30 of SK. In COS cells the 31-45 and the 46-60 mutations resulted in levels of incorporation of SK (21% and 32% respectively) similar to the background mutation (33%). However, the 16-30 mutation led to a dramatic loss of SK use (0%) to levels well below that of the background mutation. In myoblasts the same phenomenon was noted; mutation of region 31-45 gave a background level of incorporation of SK (59%), but mutation 16-31 resulted in a drop in SK incorporation (12%). These data indicate that the region 16-30 of SK appears to be required for inclusion of

the SK exon: alternatively, the 16-30 region of NM, which was used to form these mutations, contained an NM repressor sequence. However, it seems less likely that the sequence was a repressor of NM because it was inactive in constructs in which the entire exon was swapped with the SK exon (SKNM), and must have been inactive in its native position within the NM exon. Further investigation of these sequences will be required in order to ascertain their role in the regulation of alternative splicing.

A difference between the level of jump-spliced products formed by transfection of mutants into COS cells, and transfection of the same mutants into myoblasts was also noted. In the background construct (SKbpup) no jump-spliced product was formed in COS cells and 12% was formed in myoblasts. However, the mutations which did not appear to alter the level of SK splicing (31-45 and 46-60 in COS; 31-45 in myoblasts) produced very different levels of jump-spliced product. In COS cells this was 45% and 23% respectively, in myoblasts this was 13%. The myoblast level is much closer to the background level and therefore appeared to be a null mutation. So, although the 31-45 and 46-60 region mutations were null mutations with respect to the level of SK (downstream site) incorporation, they appeared to prevent use of the distal NM exon (upstream site) in COS cells. One possible role for these sequences is as an NM exon activator (mechanism 2a). Insertion of these sequences into the SK exon did not over-ride the repression of the SK exon, but could cause the titration of a limited NM activator such that the upstream exon could no longer be fully activated for incorporation into mature transcript. The results obtained cannot exclude the possibility that there is an activator sequence within bases 31-60 of NM which acts in cis to promote use of a quiescent upstream (natural NM) site. If the activator sequence was intronic (mechanism 2b), then a construct which contained a duplication of the NM exon should produce double incorporation of NM: the

upstream NM exon being activated as in a wild-type construct, the downstream NM (replacing the natural SK exon) would be freed from repression by SK specific repressors and therefore also available to the splicing machinery. Unfortunately, this construct has recently been shown to contain a deletion of the upstream NM exon (A. Smith personal communication). Therefore, the lack of double incorporation in myoblasts and COS cells from this construct (now called (--)NM*) cannot now be cited as evidence against an intronic sequence activator of NM and this mechanism cannot now be excluded.

The results obtained in myoblasts differ from those from COS as a limiting exon specific activator did not appear to act to promote inclusion of the NM exon in myoblasts. However, improving the 5'splice-site of the NM exon did cause an increase in exon NM use, at the expense of jump-spliced product. Thus, improving the 5' splice-site may have promoted the upstream (natural NM) site from a quiescent to a constitutively used exon. This negated the requirement for an activator of NM, and the site was now used in 91% of transcripts from myoblasts. The high level of jump-spliced products formed by transfection with wild-type mini-gene may themselves indicate a partial incompatibility between the *trans*-acting NM activator in mouse, and human NM sequence.

Thus, in COS cells there was evidence in favour of activation of a quiescent NM exon via NM exon sequences. However, a mechanism which also involved intron sequence activators could not be excluded. In myoblasts, there was no evidence for an exon-sequence activator; the possibility that intron sequence was involved could not be excluded.

Cis-acting activator sequences have been identified in the cardiac troponin T gene (Cooper and Ordahl 1989; Cooper 1992) and in the *dsx* gene of *Drosophila* (Hedley and Maniatis 1991) as mechanisms for the regulation of alternative splicing. Further investigation of these sequences will be required in order to ascertain their role in the regulation of alternative splicing of human α_s -tropomyosin.

The results obtained from the transfection of mutant α_s -tropomyosin mini-gene constructs into COS cells and myoblasts were consistent with the following mechanism:

The SK exon was omitted from incorporation into mature transcripts because of *cis*-acting repressor sequences contained within the first 15 and the last 30 bases of the SK exon. The binding of a SK sequence-specific *trans*-acting factor, present in non-muscle cells but absent from muscle cells, caused the repression. Incorporation of the NM exon into mature transcripts may have been aided by the binding of a *trans*-acting activator (in COS cells) to a sequence contained within the central region of the NM exon. A mechanism in which for intron sequence acted as an activator of splicing, of the NM exon, could not be excluded. In the absence of *trans*-acting factor(s) which could bind the activator sequence(s) the NM exon was recognised only poorly by the splicing machinery. In myoblasts, recognition of the human activator sequence by a mouse factor may have been poor. This could account for the high levels of jump-spliced products from the wild-type construct.

5.3.2. Identification of signals which regulate expression in muscle (myotubes)

The reduced incorporation of the NM exon in myotubes could be produced by the same mechanisms as listed earlier:

1. Interfering with constitutive splicing
 - a) via secondary structure
 - b) via exon repressor sequences
 - c) via mutual interference
2. Activation of unusable exons
 - a) via exon sequences
 - b) via intron sequences
3. Competition for a limited resource
 - a) which could be a splicing factor
 - b) which could be another splice-site
 - c) which could be time

Each possibility is not necessarily mutually exclusive.

If the reduced incorporation of the NM exon in myotubes was due to the formation of a large secondary structure (mechanism 1a) which sequestered the NM exon from the splicing machinery, then replacing the NM exon with SK sequence would be predicted to disrupt this structure and de-repress the exon. Under this mechanism, constructs which replaced NM with SK (SKNM*, SKSK*) should result in the use of both the upstream and the downstream sites. This was not observed. In neither case was the upstream site activated, even though the entire exon sequence was changed.

An alternative mechanism, which could produce the the reduced incorporation of the NM exon in myotubes, is repression via exon sequences (mechanism 1b). Under this form of control, then the construct which swapped the exons (SKNM*) should have incorporated the SK exon and a

construct which placed the NM exon in the SK site, without a competing upstream exon (--)NM* should have produced more jump-spliced product. This was clearly not the case: all transcripts from these constructs included the NM exon. Therefore, exon sequence repressor elements did not control the exclusion of the NM exon in myotubes.

If the lack of a *trans*-acting activator of NM is the mechanism which causes the lack of incorporation of NM in muscle cells (2a and/or 2b), the NM exon site must be dormant as a default pattern, but can be activated in non-muscle cells. Although repression of the SK exon in COS cells and myoblasts was clearly due to the presence of *cis*-acting repressor sequences within the SK exon, an additional role for NM sequences (31-45, 46-60) as *cis*-acting activator sequences was indicated in COS cells, although required that the factor which bound to the sequence was limited and was only required for activation of the upstream site (NM sequences could be incorporated if flanked by the SK splice-sites and intron). Also, the possibility that an intron sequence activator of NM was active in COS cells and myoblasts had not been excluded. Therefore, in muscle cells, a mechanism based on a lack of activation of the NM exon by a *trans*-acting activator binding to sequence(s) associated with the NM exon, could not be discounted. This requires that the activating factor is produced in non-muscle cells, but is absent from muscle cells. Also, in muscle cells, the absence of activating factor predicts that the upstream site is quiescent even in the absence of downstream exon, and that double incorporation would never be seen. The results obtained in testing for a competition mechanism, discussed below, are consistent with a NM activator mechanism.

Two models can be proposed which would produce a reduction in activation of a quiescent exon. First, the loss of NM-specific activator protein (which targeted either exon or intron sequence). Second, a reduction in concentration of a general splicing factor. In the second mechanism if the disassociation constant for binding of the factor to the NM exon is high, and the concentration of the factor is low, then the chances of the factor being associated with that exon must be very low. The NM exon never reaches a threshold level of binding of the limiting factor and is effectively hidden from the splicing machinery. In the presence of the SK exon, splicing of the flanking constitutive exons to the SK exon will occur. However, even in the absence of the SK exon the NM exon would not be incorporated because a jump-splice between the flanking constitutive exons would form before components could associate around the poor NM exon. Therefore, a competition does exist in myotubes, but it is a competition between splicing of the SK exon to its flanking constitutive exons and splicing between the constitutive exons (to the exclusion of the SK exon); the NM exon plays no part in this competition. The slightly higher levels of jump-splicing seen in myotubes could be explained by the reduction in general splicing factor before the loss of SK repressor (also discussed later).

If the low exon NM incorporation, in myotubes, was due to competition between the splice-sites (mechanism 3b), then altering the strength of the splice-sites should perturb the splicing preferences. The 5' splice-sites of SK (SK5'cons*) and NM (NM5'cons) were improved to match the consensus sequence, and the branch-point of SK was mutated to the consensus UACUAAC sequence (SKbpup). All these constructs caused a change in the pattern of splicing, indicating that competition could play some part in the control. However, if competition for inclusion into mature transcript (mechanisms 3a or 3c), or competition for a limited splicing factor

(mechanism 3a) was the sole control mechanism, then inactivation of the SK exon should result in a proportional increase in use of the NM exon. In this case, the level of NM incorporation should be the same as that found in myoblasts (where the SK exon is repressed). This was not found to be the case: two different constructs were tested (SK5'del*, SKbpdel*) neither of which was capable of producing NM containing transcripts at a level which would be predicted from a simple competition model. Therefore, competition was not the sole mechanism for determining the outcome of splicing in myotubes. As indicated earlier, these results were consistent with a mechanism in which the NM exon required activation: myotubes appeared to lack the necessary activator.

The results obtained from the transfection of mutant α_s -tropomyosin mini-gene constructs into myotubes were consistent with the following mechanism:

In myotubes, the weak 5'splice-site of NM, a reduction in the concentration of a general splicing factor and the absence of an exon specific activator cause the upstream site (NM) to be ineffective (and not recognised by the splicing machinery). Repression of the SK exon is relieved by a reduction in concentration of repressor protein(s). The reduction in concentration of these proteins is achieved either because transcription of these genes is switched off, or by the alteration of the protein sequence via regulated alternative splicing (in a manner analagous to the cascade which regulates sex determination in *Drosophila*) or post-translational processing. The reduction in the level of active repressor protein may lag behind a reduction in the level of splicing factor and/or active activator to prevent errant double incorporation. These factors produce a situation in which incorporation of the SK exon, and skipping of the NM exon is greatly favoured. This leads to the incorporation of SK exon in mature transcripts, and unfavourable conditions for the

incorporation of both exons in a single transcript.

5.3.3 Proposed mechanism which regulates the switch in α_s -tropomyosin isoform use.

In myoblasts the SK exon is sequestered by binding of a *trans*-acting repressor to sequences within the exon. The NM exon is activated by the binding of a *trans*-acting activator which enhances the weak 5'splice-site. General splicing factors are abundant. The NM exon is recognised and spliced as if it were a constitutive exon. This results in formation of the 4-NM-6 isoform. In myotubes the repressor and activator proteins are absent or inactive and there is a reduction in concentration of a general splicing factor. This leads to the skipping of the now quiescent NM exon and the incorporation of the SK exon to produce transcripts containing the 4-SK-6 isoform.

5.4 Relevance of results to mechanisms involved in constitutive splicing

The transfection of two constructs into myotubes produced an interesting result which may have more relevance toward an understanding of constitutive splicing mechanisms, than to alternative splicing. Two constructs were made (SK5'del, SKbpdel) such that the SK exon was inactivated. One construct (sk5'del) produced some transcripts which incorporated the NM exon (18% and 37%), the other construct (skbpdel) produced transcripts which contained only jump-spliced product. The difference between these results was unexpected. A similar disparity of results had also been noted in mutational analysis of *cis*-acting repressor sequences of the alternatively spliced exons 6 and 7 of rat β -tropomyosin (Helfman et al 1990). In HeLa cells, rat β -tropomyosin usually incorporates the upstream exon 6. However, when the branch-point and polypyrimidine tract of exon 7 were deleted, but exon 7 and its downstream 5' splice-site were left intact, jump-spliced (5-8) transcripts rather than wild-type pattern (5-6-8) were

produced. These two independent findings may indicate that deletion of the branch-point but retention of a 5' splice-site can affect the splicing of an upstream exon.

Two models can be envisaged which explain these data. First, the branch-point allows for definition of the upstream 5' splice-site and downstream 3' splice-site. Without a branch-point in the downstream intron an exon cannot be spliced. Second, the retention of an exon with a 5' splice-site but no branch-point causes the formation of an abortive complex. This complex cannot dissociate into its constituent splicing components. If one of these components is limited, then the splicing of other exons may be affected. This phenomenon requires further investigation.

5.5 Evolutionary considerations for the analysis of alternative splicing

5.5.1 Evolution of the tropomyosin family

In humans, four tropomyosin gene loci have been identified, whereas in the rat and chicken there appear to be only three, and in *Drosophila* there are only two. It is interesting to speculate as to which gene arrangement most closely reflects the ancestral organisation. I surmised that the extant human α_s -tropomyosin NM and SK exons have been derived by the recombination of two ancestral genes, one encoding the NM exon the other the SK exon. The two ancestral genes were themselves the consequence of a gene duplication event. By this mechanism, a single more complex gene fulfils the functions of two distinct genes. Is it possible that a similar event has occurred in rats and chickens, since their divergence from man, such that the ancestral homologues to the α_F - and α_s -tropomyosins have recombined to form a single, more complex α -tropomyosin. Similarly, could the complex *drosophila* gene (TmII) have arisen by a recombination of its α_F -, α_s -, β - and δ -tropomyosin homologues.

It is an interesting thought that the tropomyosins may be at a later stage of evolutionary development in *Drosophila* than in man. This model re-draws the evolutionary history of the tropomyosins. If true, this also indicates that the organism of choice for the investigation of regulation of alternative splicing may not be man, rat or chicken but *Drosophila*, because the mechanisms which regulate its alternative splicing are descendants of the mechanisms extant in man, rat and chicken. *Drosophila* also has the advantage of a wealth of information concerning genomic structure and the availability of myriad mutants that allow for extensive genetic analysis.

5.5.2 Evolution of the NM and SK exons

As indicated earlier, the central exons NM and SK of the α_s -tropomyosin gene could have arisen as a consequence of one of two general mechanisms: duplication of an ancestral exon, or duplication of the gene followed by non-homologous recombination within the region bounded by exons 4 and 6. The mechanism which caused the evolution of the extant tropomyosin gene was cited as an important consideration in the analysis of alternative splicing mechanisms.

However, a conundrum exists in the evolution of the NM and SK exons of α_s -tropomyosin. They are un-symmetrical exons (as they have 76 nucleotides) and would induce a frame shift if incorporated, or skipped in tandem. If the exons arose as a consequence of exon duplication, the extant exon must have been incorporated as a constitutively spliced exon. Unless the duplication event itself caused the partial loss of function of one of the duplicated sets of splice-sites, the new and the extant exon must both be recognised as constitutive exons. This, however, would lead to double incorporation, a frame shift, and thereby formation of a truncated protein. The alternative

evolutionary pathway is via gene duplication followed by recombination which brought the NM and SK exons together. In this situation, the genes can diverge such that the region of the gene of one copy encodes a protein which confers an advantage in non-muscle cells (corresponding to the extant NM exon) and the other which confers an advantage in muscle cells (corresponding to the extant SK exon). Any mechanisms which repressed the use of the alternative isoform, in a particular tissue, would be advantageous. The adaptation or evolution of a *trans*-acting repressor of the NM exon in muscle cells and the SK exon in non-muscle cells is the obvious solution. Once the exon specific repressors were active, a recombination between the genes would be neutral (and so could spread by random genetic drift) or may have been advantageous and thereby selected for. However, repression of the NM exon was not operating as the control mechanism in myotubes. The only other mechanisms which can explain the presence of the NM exon in non-muscle cells and its exclusion in muscle cells are because of an activation of a "dormant" NM exon in non-muscle cells and/or a competition between the exons for inclusion into the splicing. In an activation model, a *trans*-acting factor must be produced in non-muscle, but not in muscle. The NM exon must be "dormant" or incorporated at a low level in the absence of this factor. In the ancestral NM gene, the trend toward "weakening" the exon could initially have been the result of random genetic drift. A slightly weakened NM exon may have exploited a *trans*-acting factor present in non-muscle. A mutation which then caused the dormancy of the NM exon in skeletal muscle would have been advantageous. The combination of the two led to a mechanism whereby the NM exon was incorporated in non-muscle and skipped in muscle. In a competition model, the pertinent question must be "how did the exons get into this situation in the first place?" Mutations which "strengthened" the NM exon would be deleterious as double incorporation would be the consequence in muscle; mutations that

“weakened” the NM exon could lead to its skipping in all cell types, and only jump-spliced transcripts would be produced non-muscle cells. This model implies that the (presumably) fine balance between losing a competition in one cell but not being skipped in the other cell can be maintained. It also predicts that if the extant gene containing NM and SK exons arose as consequence of a recombination event, that the splice-sites of the NM exon were already sub-optimal, and were selected as such because in muscle cells the ancestral NM gene would lose in a competition for a limited splicing factor, and produce only jump-spliced product. This may appear to be altruistic by the NM exon; the maintenance of the sub-optimal splice-sites providing an increased fitness for the organism, but not for the exon itself. However, if this is combined with the model in which the NM exon is enhanced in non-muscle cells the following model can be proposed, based on evolutionary considerations in combination with observed data: The ancestral NM and SK exons and their control mechanisms arose on separate genes. Recombination of the ancestral genes produced the extant hTMnm gene with its central alternatively spliced NM and SK exons. The incorporation of the NM exon (to the exclusion of the SK exon) in non-muscle involves NM enhancement and SK repression. The incorporation of the SK exon in muscle involves a lack of NM enhancement and a lack of SK repression which leads to a competition in which the SK exon is favoured.

This model, proposed for the regulation of alternative splicing of human α_s -tropomyosin, indicates that specific protein factors play a significant role in non-muscle cells, but play no role in muscle cells. It is interesting to speculate that because of the increasing specialisation of differentiated tissues, the potential for exploitation of a *trans*-acting factor is reduced purely on the basis of choice. When more data becomes available concerning the regulation of other alternatively spliced genes, it will be interesting to note whether this

predominance of specific factors in non-muscle is widespread and in itself indicates a common feature in the regulation of alternative splicing.

5.5.3 Linkage and co-evolution of *cis*- and *trans*-acting elements

In either model, during the evolution of alternative splicing it would be advantageous for the genes which regulate the switch (i.e. those encoding *trans*-acting factors) to be closely linked to the genes on which they act. At present it is only possible to assess this in *Drosophila*, for which detailed linkage maps are available. Interestingly, analysis of some of the genes involved in the cascade of sex determination shows that they are closely linked in a pair-wise manner. However, I am uncertain of the statistical significance of this finding.

5.5.4 Evidence for common regulatory mechanisms.

The results presented in this thesis indicates an important role for splice-site signals in regulation of alternative splicing, and postulated that exploitation of a reduction in a general factor in differentiated muscle cells could induce a muscle-specific switch in exon choice. This combination of features has also been suggested as the mechanism for the regulation of incorporation of exon 18 of N-CAM in neural cells (Tacke and Goridis 1991).

Several general splicing factors have now been shown to play an important role in alternative splice-site selection in *in-vitro* splicing experiments (Fu et al 1992; Mayeda and Krainer 1992; Mayeda et al 1992). ASF/SF2 has been shown to favour the use of proximal 5' splice-sites, whereas hnRNP protein A1 appears to favour the use of distal sites in exon duplication experiments. Recently, this observation has been tested, in *in-vitro* splicing experiments, using naturally occurring alternatively spliced transcripts (Mayeda et al 1993). These experiments indicated a role for ASF/SF2 to prevent exon skipping,

and a role for hnRNP A1 to promote exon skipping. ASF/SF2 is a member of the conserved SR protein family which include a conserved RNA binding domain and a region of arginine-serine repeats. The roles for other members of the family is uncertain although it is clear that they share the capacity to complement the inactive S100 splicing extract to allow correct splicing of transcripts *in-vitro* (Mayeda et al 1992; Zahler et al 1993); and SC35 and SF2 have been shown to have equivalent splice-site selection activities *in-vitro* (Fu et al 1992). Recently, it has been shown that at least four of the SR proteins are present in different absolute, and relative concentrations in different tissues (Zahler et al 1993; see Mayeda et al 1993); that the hnRNP A1 protein has different strand-annealing properties in different phosphorylation states (Cobianchi et al 1993); and that ASF/SF2 is likely to be alternatively spliced, two forms of which would not contain the SR domain (Tacke et al 1992). Therefore, different concentrations of active SR protein(s) could be produced not only by altering the absolute concentration of the factor, but by post-transcriptional and post-translational modifications. Thus, regulated exon selection or skipping could be controlled by the relative proportions and concentrations of just a few related proteins.

Interestingly, a mechanism based on exploitation of splicing signals by differing levels of general splicing-related factors in undifferentiated and differentiated cells might also enable regulated inclusion of even the smallest exons (i.e. the triplet exon AAG of the MSD region in N-CAM) which in themselves cannot contain enough nucleotides to harbour a regulatory *cis*-acting element. The importance of sub-optimal splice sites has also been implicated in regulation of alternative splicing in α - and β -tropomyosins (Mullen et al 1989; Libri et al 1992). Therefore, the data presented here provides further evidence that a common mechanism may be exploited by different genes in their regulation of alternative splicing, and that the

inclusion of tissue-specific exons in different tissues (neural and muscle) may depend on an exploitation of the same regulatory mechanism.

5.6 Conclusions

In order to address questions about the regulation of alternative splicing, I have developed two important techniques which are now routine "in house" methods. The quantitative RT-PCR method has widespread application for the analysis of alternative splicing of endogenously produced RNA, of RNA expressed from a vector introduced into cells by transfection, and of RNA spliced during an *in-vitro* splicing reaction. The transient transfection of muscle cells in culture has allowed for an analysis of the mechanisms which govern alternative splicing in differentiated cells, without the requirement for the formation of stable cell lines.

Using these methods I have established the tissue- and stage-specific alternative splicing patterns exhibited by the MSD region of N-CAM in RNA isolated from the mouse. Two exons were previously unidentified (MSD1b and MSD1c) in the mouse N-CAM. These exons have been sequenced and the sequence deposited in the EMBL data base. The analysis of the alternative splicing of this region indicated that the exons were probably incorporated as a unit in mature muscle. A search for a common sequence to which a common *trans*-acting factor might bind indicated the presence of several 12 nucleotide repeat units within these exons. However, statistical analysis could not disprove the null hypothesis that this common sequence was the consequence of random distribution. Thus, if a common sequence is bound by a *trans*-acting factor, this sequence must be within the intron sequences. The possibility that the splice-sites play a crucial role still remains.

Transfection of mutant α_5 -tropomyosin mini-gene constructs into COS cells

myoblasts and myotubes provided evidence for the mechanism by which mutually exclusive alternative pre-mRNA splicing of the NM and SK exons could be achieved. The results were consistent with the following model: The SK exon was omitted from incorporation into mature transcripts because of *cis*-acting repressor sequences contained within the first 15 and the last 30 bases of the SK exon. The binding of a SK sequence-specific *trans*-acting factor, present in non-muscle cells but absent from muscle cells, caused the repression. Incorporation of the NM exon into mature transcripts may have been aided by the binding of a *trans*-acting activator (in COS cells) to a sequence contained within the central region of the NM exon. In myotubes, the weak 5'splice-site of NM, a reduction in the concentration of a general splicing factor and the absence of an exon specific activator cause the upstream site (NM) to be ineffective (and not recognised by the splicing machinery). Repression of the SK exon is relieved by a reduction in concentration of repressor protein(s).

In the analysis of the mechanisms which regulate alternative pre-mRNA splicing I thought it prudent to address the question of the evolution of this strategy for producing isoform diversity, and also to address the question as to how this gene arrangement may have been produced. In particular, the conundrum of how mutually exclusive exons, which if incorporated together would cause a frame-shift, could have evolved by simple exon duplication if the frame-work for regulation of alternative splicing was not already in place. Under this evolutionary pathway, mechanisms exploited by different transcripts would be shared, rather than each gene being regulated by its own mechanism. Therefore, evolution considerations argue more strongly that common mechanisms underlie the regulation of alternative splicing.

Thus, on the basis of evolutionary considerations, and from analysis of the data presented here, I conclude that regulation of alternative splicing may be founded upon a common mechanism which is largely dependent upon the presence of sub-optimal splice-signals and the potential for variation in the relative concentrations of certain splicing factors. Identification and characterisation of these factors will be the next goal in the elucidation of mechanisms involved in alternative and constitutive pre-mRNA splicing.

APPENDIX 1: Sequences of mini-gene mutants

Wild-type (wt)

AGGAACAAAA GGTGGAGCTA TGAGCACAGA TAAAGACTCA AGTCTGGGGA
 CCTCCTGGTC ACTCAGGCAG CACCCCTTCT TTCTTGCCCC AGTCTCCAGT
 TCTCCAGTGT TCACAGGTGA GCCTACCAAC AGCCACTGCT CATGATGGAG
 GCCATCAAGA AAAAGATGCA GATGCTGAAG TTAGACAAGG AGAATGCTCT
 GGATCGGGCT GAGCAAGCTG AAGCTGAGCA GAAGCAGGCA GAAGAAAGAA
 GTAAACAGCT GGAGGATGAG CTGGCAGCCA TGCAGAAGAA GCTGAAAGGG
 ACAGAGGATG AGCTGGACAA GTATTCTGAA GCTTTGAAGG ATGCCCAGGA
 GAAGCTGGAA CTGGCAGAGA AGAAGGCTGC TGATGCTGAG GCTGAGGTGG
 CCTCCTTGAA CCGTAGGATC CAGCTGGTTG AAGAAGAGCT GGACCGTGCT
 CAGGAGCGCC TGGCCACTGC CCTGCAAAAAG CTGGAAGAAG CTGAAAAAGC
 TGCTGATGAG AGTGAGAGAG GTATGAAGGT TATTGAAAAC CGGGCCTTAA
 AAGATGAAGA AAAGATGGAA CTCCAGGAAA TCCAACCTAA AGAAGCTAAG
 CACATTGCAG AAGAGGCAGA TAGGAAGTAT GAAGAGGTGG CTCGTAAGTT
 GGTGATCATT GGGTCACCTT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
 CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
 atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
 tcaacctgtg taaacagggc cagggaaatgg gagtcatcta tagtgctttt
 ttaggcccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
 gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
 agccaaataa ggactaatac atggtgaaca gttaatgga agatcctact
 gggtccttca cctggggttg aagacatcaa agtggctatc ctgatttaag
 atctagtgtc tggctgggtg tgggtggtca aacctgtagt cccagctact
 tagtatgcca aggcaggaa gttgcttgaa cccaggagtt ggaggccagc
 ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
 aaatctggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
 ggcaagggtg gatattaact gctctcccgt cggtagcttc tgggccttga
 ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
 tcccgtctgc cagacataac atatcatgac catggtgggtg ttctcttccc
 attttttgaa tttttttttt tcgttttgct gcttggtttc ttgctgtgtt
 cctttttgac ctttctccct cccacaacc tggctgtgac catcccgtg
 actatcacta acagCCCTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
 GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
 ttaagccac agcaagggtt tatatgtggt gtggttttgt tttgttttgt
 tttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
 catctcttca tggaaatgct cattcttttc cacttacttg gcatctttca
 tctcttcttt cctagtgtct tcaactgcctc tccacagggt ctccttccct
 tgggtgatgt gggggactgt ttgggttggg catttgttcc caccctccacc
 ccaggtagca gttatgactt gttcctggta gtggattttc ctcaagtgtc
 ttattagtat tcaaacattt ttgagccag gaaggtctag ctcctgacac
 gttctatggt agagggagga ggggttgatgc ttgctcagggt tacttgggaa
 catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
 tcaagtgat gtgtctctct cccttttttt attctcttcc ttttcccttt
 tcctctctg tggattttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
 GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
 GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
 cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
 cagttaggcy gttcctcttc ctacctctt accatgtatt cagggtcttg
 gtttaagtct attttctgtg cacctttgat tgctttaatt aaaataacct
 tcttggtctc aggttactgg agtcatgtgc tccttaggct gatacctgct
 gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
 tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
 gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg

ggggtgctact actcttgtct atcaagcagg ggaaaggaaa aacacctcct
 tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
 tggagaatth taacctgagg atctcatggc tctaaagaca tagatttctg
 aatttcaagg acctattgtc cagagccagg ccttgctctt actcttggac
 cattcagcac atttagtgc agaggatatt caatgtaagt taaatgtagg
 gaggggggta ttgcctgtgt gactaatttc tactctgaaa tttttaatct
 acttatctta cagTACTCTC **AAAAAGAAGA TAAATATGAG GAAGAACTCG**
AGATTCTTAC TGATAAACTC AAGGAGGCAG AGACCCGTGC TGAGTTTGCT
GAGAGATCGG TAGCCAAGCT GGAAAAGCGG CGGCGATGAC CTGGAAGATG
AGCTCTATGC CCAGAACTG AAGTACAAGG CCATTAGCGA GGAGCTGGAC
CACGCCCTCA ATGACATGAC CTCTATATAA TTATCACCGT TTCTGCTCTG
TTCTGGATCT GCCCCCTTA CTCTCGGGG AACCCAAGGC CCCACTCTCG
CTCTGGATTC CATTGGGTC AGCCTGGCTG GTCCCCAAGG CATTAGGATG
GGGGAGCAAA AAGCAACTTA TGTATTTTCT TCCACCCCA CCCCAAATTA
AAATGTTAAG CTGCTGGAA

The full length SK cDNA sequence is indicated by capital letters, intron sequences
 between exons 4 and 6 by lowercase letters, and the alternatively spliced exons NM
 and SK are shown in bold capital letters.

NMSK* (*)

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcacta tagtgctttt
ttaggcccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atgggtgaaca gtttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt ccagctact
tagtatgcc aaggcaggaag gttgcttgaa ccaggaggt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccctg cggtagcttc tgggccttga
ttaacaaagt cagtgtctgaa tgcataatga tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catgggtgtg ttctcttccc
attttttgaa tttttttttt tctgttttgc gcttggtttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagcccac agcaagggtta tatatgtggt gtgggtttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgctc cattcttttc cacttacttg gcatctttca
tctcttcttt cctagtgtct tcaactgcctc tccacagggt ctccttcctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggtctag ctcctgacac
gttctatggt agagggagga gggttgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctcctg tggtattttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctgggt tgtgaaagt acaggctttg gggcc-----
----- gatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt caggttcttg
gtttaatgct attttctgtg cacttttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaatagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgagc
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SKSK*

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggccttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaaatgg gagtcactta tagtgctttt
ttagggccaa gaatgggttt aggccttcctt attctcctga ctagtgacaa
gtgacaaaag aggcctttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atgggtgaaca gtttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctggggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtaacctc tgggccttga
ttaacaaagt cagtgcgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catgggtggtg ttctcttccc
attttttgaa tttttttttt tctgttttgc gcttggtttc ttgctgtgtt
cctttttgac cttttccctt ccccaaaccc tggctgtgtc catccccgtg
actatcacta acag**TAAGTG** **TTCTGAGCTG** **GAGGAGGAGC** **TGAAGAATGT**
CACCAACAAC **CTCAAGTCTC** **TTGAGGCTCA** **GGCGGAGAAG** gtactaacca
ttaagcccac agcaagggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tgggaattgt cactcttttc cacttacttg gcatctttca
tctcttcttt cctagtgtct tcaactgcctc tccacagggt ctccttcctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggtctag ctcctgacac
gttctatggt agaggggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cctttttttt attctcttcc ttttcccttt
tctctccttg tggattttaa aatattcaca g**TAAGTGTTT** **TGAGCTGGAG**
GAGGAGCTGA **AGAATGTCAC** **CAACAACCTC** **AAGTCTCTTG** **AGGCTCAGGC**
GGAGAAGgta ggggctgggt tgtgaaagt acaggctttg gggcc-----
----- gatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt cagggttctgg
gtttaatgct attttctgtg cacttttgat tgccttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaaatc agctctgaga cctaatagac aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgga
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cag**TACTCTC** **AAAAAGAAGA** **TAAATATGAG** **GAAGAACTCG**
AGATTCTTAC **TGATAAACTC** **AAGGAG**

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SKNM*

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggcctg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gtagcatcta tagtgctttt
ttaggcccaa gaatggtttt aggcttcctt attctcctga ctagtacaa
gtgacaaaag aggctttcct attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccgt cggtaacctc tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgctctg cagacataac atatcatgac catgggtggg ttctcttccc
atthtttgaa tttttttttt tctgttttgc gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catccccgtg
actatcacta acagTAAGTG TTCTGAGCTG GAGGAGGAGC TGAAGAATGT
CACCAACAAC CTCAAGTCTCT TGAGGCTCA GCGGAGAAAG gtactaacca
ttaagccac agcaaggta tatatgtgt gtggttttgt tttgttttgt
ttttgtttg cagctgcctc cctccactt gatthttgtt agtgctgtct
catctcttca tggaaatgctc cattcttttc cacttacttg gcatcttca
tctcttctt ctagttgtct tcaactgcctc tccacagggc ctccttctct
tgggtgattt gggggactgt ttgggttggg catttgttcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagccag gaaggtctag ctccgtacac
gttctatggt agagggagga ggggtgatgc ttgtcagggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgat gtgtctctct cctttttttt attctcttcc ttttcccttt
tcctctctg tggattttaa aatattcaca **GCGTTCGCC AGAGATGGAT**
GAGCAGATTA GACTGATGGA CCAGAACCTG AAGTGTCTGA GTGCTGCTGA
AGAAAAGgta ggggctggtt tgtgaaagt acaggctttg gggcc-----
-----gatgaaa gactggagac acgcattcca tatataagtc
cagttaggcg tgtcctcttc ctacctctt accatgtatt caggttctgg
gtttaatgct atthttctgt cactttgat tgccttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tcttaggct gatactgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagtagc aatcaagaca
gatgacttca ctttccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaagggaaa aacacctct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgga
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatctc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCCTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

(--)NM*

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaatth
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaacctgtg taaacagggc cagggaatgg gagtcacta tagtgctttt
ttaggcccac gaatgggttt aggcctcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gtttaattgga agatcctact
gggtccttca cctgggtggt aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt ccagctact
tagtatgcca aggcaggag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtagcttc tgggccttga
ttaacaaagt cagtgtgaa tgcataatg tttgtgctc cctcctacat
tccgctgtgc cagacataac atatcatgac catggtggtg ttctcttccc
atthtttgaa tttttttttt tctgtttgct gcttggtttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catcc-----
----- (ctgtggtatttaaacatt) -----
-----cac agcaaggtta tatatgtgt gtggtttgt tttgtttgt
ttttgtttt cagctgctc cctccactt gattttgtt agtgctgtct
catctcttca tggaaatgtc cattctttt cacttacttg gcatcttca
tctcttctt ctagttgct tcaactgctc tccacagggt ctccttctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgt
ttattagat tcaaacattt ttgagcccag gaaggtctag ctcctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cctttttttt attctcttcc ttttccctt
tcctctctg ttgtatttaa aatattcaca **gCCGTTGCCG** **AGAGATGGAT**
GAGCAGATTA **GACTGATGGA** **CCAGAACCTG** **AAGTGTCTGA** **GTGCTGCTGA**
AGAAAAAGgta ggggctggt tgtgaaagt acaggctttg gggcc-----
-----gatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctctc ctacctcct accatgtatt caggttctgg
gtttaatgct attttctgt cactttgat tgccttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaaac agctctgaga cctaagtgc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgacctt
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggt tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgga
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC **AAAAAGAAGA** **TAAATATGAG** **GAAGAACTCG**
AGATTCCTAC **TGATAAACTC** **AAGGAG**

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters. The nineteen bases shown in brackets were introduced, as a cloning artefact, into the site of the deleted NM exon during construction of this mini-gene.

delSK-6

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcaccta tagtgctttt
ttaggcccaa gaatggtttt aggcttcctt attctcctga ctagtgacaa
gtgacaaaag aggctttcct attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt ccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtagcctc tgggccttga
ttaacaaagt cagtgtgaa tgcataatga tttgtgcttc cctcttacat
tcccgctctg cagacataac atatcatgac catgggtggtt ttctcttccc
attttttgaa tttttttttt tctgtttgct gcttggtttc ttgctgtgtt
cctttttgac ctttctccct cccacaaacc tggcttgtgc catccccgtg
actatcacta acag**CCGTTG** **CCGAGAGATG** **GATGAGCAGA** **TTAGACTGAT**
GGACCAGAAC **CTGAAGTGTC** **TGAGTGCTGC** **TGAAGAAAA**gtactaacca
ttaagcccac agcaagggtt tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgctc cattcttttc cacttacttg gcatctttca
tctcttcttt cctagttgct tcaactgcctc tccacagggt ctccttcctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cacctccacc
ccaggtagca gttatgactt gttcctggtg gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggtctag ctctgacac
gttctatggt agagggagga gggttgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctcctg tggatattaa aatattcaca g**TAAGTGTT** **TGAGCTGGAG**
GAGGAGCTGA **AGAATGTAC** **CAACAACCTC** **AAGTCTCTTG** **AGGCTCAGGC**
GGAGAAGTACTCTC **AAAAAGAAGA** **TAAATATGAG** **GAAGAACTCG**
AGATTCTTAC **TGATAAACTC** **AAGGAG**

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SKbpup (+)

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtoca ttaccaattt
atcccttccc ctttggttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcatcta tagtgctttt
ttaggcccaa gaatggtttt aggcttcctt attctcctga ctagtacaa
gtgacaaaag aggcttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gtttaattgga agatcctact
gggtccttca cctggggtgg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggtca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtaacctc tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catggtggtg ttctcttccc
attttttgaa tttttttttt tctgtttgct gcttggtttc ttgctgtgtt
cctttttgac ctttctccct ccccacaacc tggctgtgac catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagcccac agcaagggtt tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgcgtctc
catctcttca tggaaatgct cattcttttc cacttacttg gcactcttca
tctcttcttt cctagtgtgt tcaactgcctc tccacagggc ctcccttccc
tgggtgattt gggggactgt ttgggtggg catttggttc cacctccacc
ccaggtagca gttatgactt gttcctggtg gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggtctag ctccgtgacac
gttctatggt agagggagga gggttgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctcctg ttgtatttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattoca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt caggttctgg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaatgagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaagggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgagc
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggtg ttgctgtgt gactaattc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAAGCTC
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

1-15+

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttgggtcca ttaccaatth
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaaatgg gagtcatcta tagtgctttt
ttaggcccaa gaatgggttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atgggtgaaca gttaatggga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctggggtcga tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtaacctc tgggccttga
ttaacaaaag cagtgcctgaa tgcataatgta tttgtgcttc cctcttacat
tcccgctctgc cagacataac atatcatgac catggtgggtg ttctcttccc
attttttgaa tttttttttt tctgtttgct gcttgttttc ttgctgtgtg
cctttttgac ctttctccct cccacacaac tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagcccac agcaagggtt tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tgggaatgct cattcttttc cacttacttg gcatctttca
tctcttcttt cctagtgtgt tcaactgcctc tccacagggc ctccctcctt
tgggtgattt gggggactgt ttgggttggg cattgttcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggtctag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cctttttttt attctcttcc ttttcccttt
tcctctcctg tggattttaa aatattcaca gCCGTTGCCG AGAGATGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtgcggcg tgtcctcttc ctacctctt accatgtatt caggttcttg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttgggtctc aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacacaatc agctctgaga cctaatgagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgacctg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgcctct actctggac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggtg ttgctgtgt gactaatctt tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAAGCTC
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

16-30*+

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcaccta tagtgctttt
ttaggcccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
gtgacaaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctggggtgg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcagggaag gttgcttgaa cccaggagt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccg cggtaacctc tgggccttga
ttaacaaagt cagtgcctgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catggtgggtg ttctcttccc
attttttgaa tttttttttt tctgttttgc gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacaaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccac agcaaggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgctc cattcttttc cacttacttg gcatctttca
tctcttcttt cctagtgtgt tcaactgcctc tccacagggt ctccttctct
tgggtgattt gggggactgt ttgggttggg catttgttcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagccag gaaggtctag ctccctgacac
gttctatggt agaggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cctttttttt attctcttcc ttttcccttt
tctctctctg tggatattaa aatattcaca gTAAGTGTTG TGAGCTGGAT
GAGCAGATTA GGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGGta ggggctggtt tgtgaaagt acaggctttg gggcc-----
----- gatgaaa gactggagac acgcattcca tatataagtc
cagttaggagc tgtcctcttc ctacctcctt accatgtatt caggttcttg
gtttaatgct attttctgtg cacctttgat tgccttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatactgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaatgagc aatcaagaca
gatgacttca ctttgcctac atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgcct atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgcctct actcttggac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatctt tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

31-45*+

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG

CAGAGTCgtg agtatctagc tccccaaact tcttgggtcca ttaccaatth
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaaatgg gagtcactta tagtgctttt
ttaggcccaa gaatgggttt aggccttcctt attctcctga ctagtgcaca
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gtttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccg cggtaacctc tgggcttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgctctg cagacataac atatcatgac catgggtggg tctcttccc
atthtttgaa tttttttttt tctgtttgct gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacacaac tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT

GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccac agcaagggtt tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tgggaatgct cactcttttc cacttacttg gcatcttca
tctcttcttt cctagtgtgt tcaactgcctc tccacagggc ctccctcctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cacctocacc
ccaggtagca gttatgactt gttctcggtg gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggcttag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cctttttttt attctcttcc ttttcccttt
tcctctcctg tggatthtaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA **AAGTATGGA** **CCAGAACCTC** AAGTCTCTTG AGGCTCAGGC

GGAGAAgta ggggctggtt tgtgaaagt acaggctttg gggcc-----
-----gatgaaa gactggagac acgcattcca tatataagtc
cagtgaaggc tgtcctcttc ctacctctt accatgtatt caggttcttg
gtttaatgct atthttctgt cacctttgat tgctthaat aaaataacct
tcttgggtctc aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgacctg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaatth taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgtctct actctggac
cattcagcac atthagtga agaggatatt caatgtaagt taaatgtagg
gaggggggtg ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC **AAAAAGAAGA** TAAATATGAG GAAGAAGCTC
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

46-60**

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttgggtcca ttaccaatth
atcccttccc ctttgggttg ggcacataat atgaataaat tagcattccct
tcaaccttgt taaacagggc caggggaatgg gagtcaccta tagtgctttt
ttagggccaa gaatggtttt aggccttcctt attctcctga ctagtgcaca
gtgacaaaag aggcctttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atgggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtgg aagacatcaa agtgggtatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcc aaggcaggaag gttgcttgaa cccaggagtt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgtttttg ccactttgcc aatctgctag ggaaccttag
ggcaagggt gatattaact gctctccgt cggtacctt tgggcttga
ttaacaaagt cagtgtgaa tgcataatga tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catggtgggt ttctcttccc
attttttgaa tttttttttt tctgtttgtc gcttggtttc ttgctgtgtt
cctttttgac ctttctccct ccccaaac tggcttgtgc catcccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccac agcaagggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tggaaatgct cattcttttc cacttacttg gcatctttca
tctcttctt cctagtgtgt tcaactgcct tccacagggt ctccttctt
tgggtgattt gggggactgt ttgggttggg catttgttcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtagt tcaaacatth ttgagcccag gaaggcttag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cctttttttt attctcttcc ttttccctt
tctctctctg tggattttaa aatattcaca gTAAGTGTTG TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACTG ATGGACCAGA AGGCTCAGGC
GGAGAAGgta ggggctgggt tgtgaaagt acaggctttg gggcc-----
----- -gatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt cagggttctgg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttgggtct aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagtgc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaggaaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaatth taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATCTTAC TGATAAATC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

61-75+

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcacta tagtgctttt
ttaaggcccaa gaatgggttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctggggtgg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggtca aacctgtagt ccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccgt cggtacctc tgggacctga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgctc cctcttacat
tcccgctgc cagacataac atatcatgac catggtggtg ttctcttccc
attttttgaa tttttttttt tcgttttgct gcttggtttc ttgctgtgtt
cctttttgac ctttctccct ccccaaac tggcttgtgc catcccgtg
actatcacta acagccggtg CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccac agcaagggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tggaaatgct cattcttttc cacttacttg gcatcttca
tctcttctt cctagtgtct tcaactgctc tccacagggt ctccttctt
tgggtgattt ggggactgtt ttgggttggg catttggtcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttatttagtat tcaaacatatt ttgagcccag gaaggctag ctccgtgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cctttttttt attctcttcc ttttccctt
tcctctctg tggattttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAAATGTCAC CAACAACCTC AAGTCTCTTG ATGCTGCTGA
AGAAAAGgta ggggctgggt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtggggcg tgtcctcttc ctactcctt accatgtatt cagggtcttg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttgggtct aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgcct atcaagcagg ggaagggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

NMbpup

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcaccta tagtgctttt
ttaggcccaa gaatggtttt aggccttctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gtaattgga agatcctact
gggtccttca cctggggtgg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggtca aacctgtagt ccagctact
tagtatgcca aggcagggaag gttgcttgaa cccaggagt ggaggccagc
ctgggctga tagggagacc ctgtccctaa acaataaaa tttaaaaaa
aaatctggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccgt cggtaacctc tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgctc cctcttacct
tcccgtctgc cagacataac ata**TACTAAC** catggtggtg ttctcttccc
atthttttaa tttttttttt tctgtttgct gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTG TGAGTGTGTC TGAAGAAAAG gtactaacca
ttaagccac agcaagggtt tatatgtgtt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgtc cattcttttc cacttacttg gcatctttca
tctcttctt ctagttgtt tcaactgcctc tccacagggt ctccttccct
tgggtgattt gggggactgt ttgggttggg catttgttcc cacctccacc
ccaggtagca gttatgactt gttcctggtg gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagccag gaaggtctag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcagg tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctctct ggtattttaa aatattcaca gTAAGTGTTC TGAGCTGGAG
GAGGAGCTGA AGAATGTCTAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagttaggcg tgtcctcttc ctacctctt accatgtatt caggttcttg
gtttaatgct attttctgtg cacttttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg accattgtc cagagccagg ccttgctctt actctggac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

2xbpup

GGTGATCATT GGGTCACCCT TGGAACGCAC GTGG CTCGTAAGTT
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaatct
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaaatgg gagtcacta tagtgctttt
ttaggcccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtaggggtt acaaatgag
agccaaataa ggactaatac atgggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctc aacctgtagt cccagctact
tagtatgcc aaggcaggaag gttgcttgaa cccaggagtt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccctg cggtagcttc tgggcttga
ttaacaaagt cagtgtgaa tgcataatga tttgtgcttc cctcttacat
tcccgtctgc cagacataac ata**TACTAAC** catggtggtg ttctcttccc
attttttgaa tttttttttt tctgttttgc gcttggtttc ttgctgtgtt
cctttttgac ctttctccct ccccaaac tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagcccac agcaagggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gatgtttgtt agtgctgtct
catctcttca tggaaatgct cattcttttc cacttacttg gcatctttca
tctcttctt cctagtgtgt tcaactgcctc tccacagggt ctccttctc
tgggtgattt ggggactgt tggggttggg catttggtcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggcttag ctccgtgacac
gttctatggt agagggagga gggttgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttca**TAC**
TAACgtgtat gtgtctctct cccttttttt attctcttcc ttttccctt
tctctctctg tggattttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctgggt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgcctcttc ctacctcctt accatgtatt caggttctgg
gtttaatgct attttctgtg cacctttgat tgccttaatt aaaataacct
tcttgggtct aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaagggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actctggac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cag**TACTCTC** AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SK5 'cons'

GTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCT TGGAAACGCAC AGAGGAACGA GCTGAGCTGG

CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt

atcccttccc ctttggttg ggcacataat atgaataaat tagcattcct

tcaaccttgt taaacagggc cagggaatgg gagtcacta tagtgctttt

ttaggcccac gaatgggttt aggccttcctt attctcctga ctagtacaa

gtgacaaaag aggccttctt attttttgtt ggtaggggtt acaaattgag

agccaaataa ggactaatac atgggtgaaca gtttaattgga agatcctact

gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag

atctagtgtc tggctgggtg tgggtggctca aacctgtagt ccagctact

tagtatgcc aaggcaggaag gttgcttgaa cccaggagtt ggaggccagc

ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa

aaatctgggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag

ggcaagggtg gatattaact gctctcccgt cggtaacctt tgggccttga

ttaacaaagt cagtgttgaa tgcataatgta tttgtgcttc cctcttacat

tcccgtctgc cagacataac atatcatgac catgggtgggt ttctcttccc

attttttgaa tttttttttt tcgttttgct gcttggtttc ttgctgtgtt

cctttttgac ctttctccct ccccaacaac tggcttgctc catccccgtg

actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT

GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca

ttaagcccac agcaagggtta tatatgtggt gtgggtttgt tttgttttgt

ttttgttttg cagctgcctc ccctccactt gatgtttgtt agtgctgtct

catctcttca tggaaatgct cattcttttc cacttacttg gcatctttca

tctcttctt cctagtgtgt tcaactgcctc tccacagggt ctccttccct

tgggtgattt gggggactgt ttgggttggg catttggttc cactccacc

ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc

ttatttagtat tcaaacattt ttgagcccag gaagggtctag ctctgacac

gttctatggt agagggagga ggggtgatgc ttgtcagggt tacttgggaa

catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt

tcacgtgtat gtgtctctct cctttttttt attctcttcc ttttcccttt

tcctctctg tggattttaa aatattcaca gTAAGTGTTC TGAGCTGGAG

GAGGAGCTGA AGAATGTAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC

GGAGAAGgta AgTgctgggt tgtgaaagt acaggctttg gggcc-----

----- --gatgaa gactggagac acgcattcca tatataagtc

cagtgaggcg tgtcctcttc ctacctcctt accatgtatt cagggttctgg

gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct

tcttgggtct aggttactgg agtcatgtgc tccttaggct gataacctgt

gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga

tgggggcaac agacagaatc agctctgaga cctaatagac aatcaagaca

gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg

gggtgctact actcttgctc atcaagcagg ggaagggaaa aacacctcct

tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca

tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg

aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgga

cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg

gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct

acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG

AGATTCTTAC TGATAAATC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

4 5'cons

GGTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAAAGgtA agtatctagc tccccaaact tcttggtcca ttaccaatth
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattccct
tcaaccttgt taaacagggc caggggaatgg gagtcaccta tagtgctttt
ttaggcccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggcctttctt attttttgtt ggtaggggtt acaaattgag
agccaaataa ggactaatac atgggtgaaca gtttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcc aaggcaggaag gttgcttgaa cccaggagtt ggaggccagc
ctgggctgca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggtg tctgtttttg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctccctg cggtaccttc tgggcttga
ttaacaaagt cagtgtctgaa tgcataatgta tttgtgcttc cctcttacat
tcccgctctgc cagacataac atatcatgac catgggtgggt ttctcttccc
atthttttgaa tttttttttt tctgttttgc gcttggtttc ttgctgtgtt
cctttttgac ctttctccct cccacaaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccccc agcaagggtta tatatgtggt gtgggtttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tgggaatgctc cattcttttc cacttacttg gcatctttca
tctcttctt cctagtgtgt tcactgcctc tccacagggt ctccttctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgt
ttattagtat tcaaacatth ttgagcccag gaaggcttag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cctttttttt attctcttcc ttttccctt
tcctctcctg tggtatthaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAATGTCTC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgtg ggggctgggt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt cagggtcttg
gtttaatgct attttctgtg cacttttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaattgagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgacctg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaatth taacctgagg atctcatggc tctaaagaca tagatttctg
aatthcaagg acctattgtc cagagccagg ccttgctctt actcttgac
cattcagcac atthagtgc agaggatatt caatgtaagt taaatgtagg
gaggggggtg ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

NM5' cons

GGTGG CTCGTAAGTT

GGTGATCATT GGGTCAACCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttgggtcca ttaccaattt
atcccttccc ctttggcttg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc caggaatgg gagtcacta tagtgctttt
ttaggccccaa gaatggtttt aggccttcctt attctcctga ctagtgacaa
gtgacaaaag aggccttcctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggtca aacctgtagt cccagctact
tagtatgcc aaggaggaag gttgcttgaa cccaggaggt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtaacctt tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catggtggtg ttctcttccc
attttttgaa tttttttttt tctgttttgc gcttgttttc ttgctgtgtt
cctttttgac ctttctccct ccccaacaac tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gta**AGT**acca
ttaagccccc agcaagggtt tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgct cattcttttc cacttaactg gcactcttca
tctcttcttc cctagtgtct tcaactgcctc tccacagggt ctctcttccc
tgggtgattt gggggactgt ttgggttggg catttgttcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaagggtctag ctctgacac
gttctatggt agaggaggga ggggtgatgc ttgctcagg tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctctctg tggattttaa aatattcaca gTAAGTCTTC TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagcctgccc cttgatgaaa gactggagac acgcattcca tatataagtc
cagttagggc tgtcctcttc ctacctcctt accatgtatt caggttctgg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttgggtctc aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttggac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SK5'del*

```

GGTGG CTCGTAAGTT
GGTGATCATT GGGTCACCTT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaattt
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaaccttgt taaacagggc cagggaatgg gagtcacta tagtgctttt
ttaggcccga gaatggtttt aggccttcctt attctcctga ctagtgacaa
gtgacaaaag aggcctttctt attttttgtt ggtagggttt acaaattgag
agccaaataa ggactaatac atgggtgaaca gtttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcc aaggcaggaag gttgcttgaa cccaggagtt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggtg tctgtttttg ccactttgcc aatctgctag ggaaccttag
ggcaagggta gatattaact gctctcccgt cggtagcttc tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catgggtggtg ttctcttccc
attttttgaa tttttttttt tctgtttgtc gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccccc agcaagggtta tatatgtggt gtgggtttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaaatgctc cattcttttc cacttacttg gcatctttca
tctcttctct cctagtgtgt tcaactgcctc tccacagggt ctccttccct
tgggtgattt gggggactgt ttgggttggg catttgttcc caccctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagccag gaaggtctag ctccgtgacac
gttctatggt agagggagga ggggttgatg ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tctctctctg tggtatattaa aatattcaca gTAAGTGTTC TGAGCTGGAG
GAGGAGCTGA AGAATGTTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAG--a ggggctgggt tgtgaaagtg acaggctttg gggcc-----
-----gatgaaa gactggagac acgcattcca tatataagtc
cagttagggcg tgctcctctc ctacctcctt accatgtatt cagggttctgg
gtttaatgct attttctgtg cacttttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccctaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

```

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

NM5'del

GGTGG CTCGTAAGTT

GGTGATCATT GGGTCACCCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaatth
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaacctgtg taacacagggc cagggaatgg gagtcatcta tagtgctttt
ttaggcccga gaatgggtttt aggccttcctt attctcctga ctagtgacaa
gtgacaaaag aggcctttcctt attttttgtt ggtagggttt acaaattgag
agccaaaataa ggactaatac atgggtgaaca gttaattgga agatcctact
gggtccttca cctgggtgag aagacatcaa agtggctatc ctgatttaag
atctagtgtc tgggtgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaaataaa tttaaaaaaa
aaatctgggtg tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccgt cggtagcttc tgggccttga
ttaacaaagt cagtgtctgaa tgcataatgta tttgtgcttc cctcttacat
tcccgctctgc cagacataac atatcatgac catgggtggtg ttctcttccc
atthttttgaa tttttttttt tctgttttgc gcttgttttc ttgctgtgtt
cctttttgac ctttctccct cccacacacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG --actaacca
ttaagcccac agcaagggtta tatatgtggt gtgggttttgt tttgttttgt
ttttgttttg cagctgcctc cctccactt gattttgttg agtgctgtct
catctcttca tgggaatgctc cattcttttc cacttacttg gcatcttca
tctcttcttt cctagtgtct tcaactgcctc tccacagggt cctctcctt
tgggtgattt gggggactgt ttgggttggg cattgttcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggcttag cctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcaggt tacttgggaa
catctcttcc ccagtatgcc ttccaactct ctctacatat aggttcaagt
tcacgtgtat tgtgtctctc cctttttttt attctcttcc ttttcccttt
tctctcctg ttgtatttaa aatattcaca gTAAGTGTTT TGAGCTGGAG
GAGGAGCTGA AGAATGTAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtggagcg tgtctcttct ctacctcctt accatgtatt caggttcttg
gtttaatgct attttctgtg cacctttgat tgccttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tcttaggtc gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaattgagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgagc
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggtc ttgcctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATTCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase.

delNM (spe)

```

GGTGG CTCGTAAGTT
GGTGATCATT GGGTCACCTT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaatth
atcccttccc ctttggtctg ggcacataat atgaataaat tagcattcct
tcaacctgtg taaacagggc cagggaatgg gagtcatcta tagtgctttt
ttaggcccaa gaatgggttt aggccttcctt attctcctga -----
----- (731) -----
----- tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gattttgttg agtgctgtct
catctcttca tggaatgctc cattcttttc cacttacttg gcatctttca
tctcttcttt cctagtgtgt tcaactgcctc tccacagggt ctcttcctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cacctccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgtc
ttattagtat tcaaacattt ttgagcccag gaaggcttag ctctgacac
gttctatggt agagggagga ggggtgatgc ttgctcagg tacttgggaa
catctcttcc ccagtagtgc ttccaactct ctctacatat aggttcaagt
tcacgtgtat gtgtctctct cccttttttt attctcttcc ttttcccttt
tcctctcctg tggattttaa aatattcaca gTAAGTGTTC TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcctggcc
cagccctgcc cttgatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt caggttctgg
gtttaatgct attttctgtg cacctttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tccttaggct gatactgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaatgagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaatth taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttggac
cattcagcac atttagtga agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgcctgtgt gactaatthc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAAGTCG
AGATTCTTAC TGATAAACTC AAGGAG

```

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

SKbpdel*

GTGG CTGTAAGTT

GGTGATCATT GGGTCACCCT TGGAACGCAC AGAGGAACGA GCTGAGCTGG
CAGAGTCgtg agtatctagc tccccaaact tcttggtcca ttaccaatth
atcccttccc ctttggttg ggcacataat atgaataaat tagcattccct
tcaaccttgt taaacagggc cagggaaatgg gagtcacta tagtgctttt
ttagggccaa gaatggtttt aggccttcctt attctcctga ctagtacaa
gtgacaaaag aggccttctt attttttgtt ggtaggggtt acaaattgag
agccaaataa ggactaatac atggtgaaca gttaattgga agatcctact
gggtccttca cctgggggtg aagacatcaa agtggctatc ctgatttaag
atctagtgtc tggctgggtg tgggtggctca aacctgtagt cccagctact
tagtatgcca aggcaggaag gttgcttgaa cccaggagt ggaggccagc
ctgggcgtca tagggagacc ctgtccctaa aacaaataaa tttaaaaaaa
aaatctgggt tctgttttgg ccactttgcc aatctgctag ggaaccttag
ggcaagggtg gatattaact gctctcccg cggtacctt tgggccttga
ttaacaaagt cagtgtgaa tgcataatgta tttgtgcttc cctcttacat
tcccgtctgc cagacataac atatcatgac catggtggtg ttctcttccc
attttttgaa tttttttttt tcgttttgct gcttggtttc ttgctgtgtt
cctttttgac ctttctccct cccacaacc tggcttgtgc catccccgtg
actatcacta acagCCGTTG CCGAGAGATG GATGAGCAGA TTAGACTGAT
GGACCAGAAC CTGAAGTGTC TGAGTGCTGC TGAAGAAAAG gtactaacca
ttaagccac agcaagggtta tatatgtggt gtggttttgt tttgttttgt
ttttgttttg cagctgcctc ccctccactt gatthttgtt agtgctgtct
catctcttca tggaaatgct cattcttttc cacttacttg gcatctttca
tctcttctt cctagtgtgt tcaactgcctc tccacagggt ctccttctt
tgggtgattt gggggactgt ttgggttggg catttggtcc cactccacc
ccaggtagca gttatgactt gttcctggta gtggatttct ctcaagtgt
ttattagtat tcaaacattt ttgagcccag gaagggtctag ctctgacac
gttctatggt agaggagga ggttgatgc ttgctcagg tacttgggaa
catctcttcc ccagtgacc ttccaactct ctctacatat aggttca**GTT**
AGTAgtgtat gtgtctctct cccttttttt attctcttcc ttttccctt
tctctctctg tggattttaa aatattcaca gTAAGTGTTC TGAGCTGGAG
GAGGAGCTGA AGAATGTCAC CAACAACCTC AAGTCTCTTG AGGCTCAGGC
GGAGAAGgta ggggctggtt tgtgaaagt acaggctttg gggcc-----
----- gatgaaa gactggagac acgcattcca tatataagtc
cagtgaggcg tgtcctcttc ctacctcctt accatgtatt cagggttctg
gtttaatgct attttctgtg cacttttgat tgctttaatt aaaataacct
tcttggtctc aggttactgg agtcatgtgc tcttaggct gatacctgct
gtgctgaaaa gcaataatca agggccctaa caggctttct agcagacaga
tgggggcaac agacagaatc agctctgaga cctaagagc aatcaagaca
gatgacttca ctttgccatc atgtgggtgt gtttcagtca tgtgaccttg
gggtgctact actcttgctc atcaagcagg ggaaggaaa aacacctcct
tgggttgatc tgaataatat caaagcttag actccatctt aaaaagatca
tggagaattt taacctgagg atctcatggc tctaaagaca tagatttctg
aatttcaagg acctattgtc cagagccagg ccttgctctt actcttgac
cattcagcac atttagtgca agaggatatt caatgtaagt taaatgtagg
gaggggggta ttgctgtgt gactaatttc tactctgaaa tttttaatct
acttatctta cagTACTCTC AAAAAGAAGA TAAATATGAG GAAGAACTCG
AGATCTTAC TGATAAACTC AAGGAG

Mini-gene sequence from exons 4 to NM to SK to 6 (inclusive). Exon sequences shown in underlined capitals, intron sequences in lowercase, deletions indicated by dotted lines, and point mutations by bold letters.

Appendix 2: Myoblasts

Absorbance/emmission data derived from analysis of RT-PCR products. Reactions were labelled with ^{32}P dATP, the resulting product cleaved with AluI, and separated by polyacrylamide gel electrophoresis. Quantification of isoform use was assessed by absorbance readings of autoradiographs (using the Image system by UVP), or by direct analysis of the gel by phosphorimager. Each result indicates a separate transfection into myoblasts and subsequent analysis.

APPENDIX 2: Myoblasts

Construct	Absorbance/Emmision (arbitrary units)		
	4-NM-6	4-SK-6	4-6
wt	2333	153	1621
wt	170	22	124
NMSK*	83	13	3
NMSK*	276	27	135
SKNM*	504	24	0
SKSK*	0	251	754
(-)-NM*	286	0	0
1-15*	2	94	1
16-30**	18	8	40
31-45**	43	91	20
46-60**	1	91	1
61-75*	1	85	2
NM5'cons	151	5	6
NM5'cons	192	8	15
NM5'cons	1519	14	55
NMbpup	394	37	401
NMbpup	69	20	53
SK5'cons*	90	55	0
SK5'cons*	925	413	230
SKbpup	32	48	9
Skbpup	62	130	12
SKbpup	369	824	311
4 5'cons	120	14	86
4 5'cons	1320	59	889
NM5'del	0	20	141
NM5'del	0	42	736
NMdel	3	9	65
NMdel	0	48	582
SK5'del	126	15	116
SK5'del	383	25	206
SKbpdel	103	17	88
2 x bpup	13	91	12
2 x bpup	30	352	99

Appendix 2: Myotubes

Absorbance/emmission data derived from analysis of RT-PCR products. Reactions were labelled with ^{32}P dATP, the resulting product cleaved with AluI, and separated by polyacrylamide gel electrophoresis. Quantification of isoform use was assessed by absorbance readings of autoradiographs (using the Image system by UVP), or by direct analysis of the gel by phosphor-imager. Each result indicates a separate transfection into myoblasts. RNA was isolated after cells had differentiated into myotubes.

APPENDIX 2: Myotubes

Construct	Absorbance/Emmision (arbitrary units)		
	4-NM-6	4-SK-6	4-6
wt	66	104	156
wt	424	797	707
NMSK*	70	121	54
NMSK*	13	30	10
SKNM*	235	0	5
SKNM*	81	0	0
SKSK*	0	52	50
SKSK*	0	15	15
(-)-NM*	161	0	2
(-)-NM*	46	1	0
1-15*	100%		
16-30+*	100%		
31-45+*	100%		
46-60+*	100%		
61-75*	100%		
NM5'cons	187	82	29
NM5'cons	32	18	8
SK5'cons*	17	81	4
SK5'cons*	14	92	1
SKbpup	100%		
SKbpup	100%		
NMdel	0	33	80
NMdel	0	8	30
SK5'del	45	6	72
SKbpdel	0	0	19
SKbpdel	0	0	21

- Adema, G.J., Bovenberg, R.A.L., Jansz, H.S. and Baas, P.D. (1988). Unusual branchpoint selection involved in splicing of the alternatively processed calcitonin/CGRP-1 pre-mRNA. *Nucleic Acids Res.*, **16**, 9513-9526.
- Adema, G.J., van Hulst, K.L. and Baas, P.D. (1990). Uridine branch acceptor is a *cis*-acting element involved in regulation of the alternative processing of calcitonin/CGRP-1 pre-mRNA. *Nucleic Acids Res.*, **18**, 5365-5373.
- Aebi, M., Hornig, H., Padgett, R.A., Reiser, J. and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell*, **47**, 555-565.
- Aebi, M., Hornig, H. and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved GU. *Cell*, **50**, 237-246.
- Aebi, M. and Weissmann, C. (1987). Precision and orderliness in splicing. *Trends in Genetics*, **3**, 102.
- Akins, R.A. and Lambowitz, A.M. (1987). A protein required for splicing group-I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-transfer RNA-synthetase or a derivative thereof. *Cell*, **50**, 331-345.
- Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982). Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature*, **298**, 240-244.
- Amrein, H., Gorman, M. and Nöthiger, R. (1988). The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. *Cell*, **55**, 1025-1035.
- Andreadis, A., Gallego, M. and Nadal-Ginard, B. (1987). Generation of protein isoform diversity by alternative splicing: Mechanistic and biological implications. *Ann. Rev. Cell Biol.*, **3**, 207-242.
- Arrigo, S. and Beeman, K. (1988). Regulation of rows sarcoma virus RNA splicing and stability. *Mol. Cell. Biol.*, **8**, 4858-4867.
- Auffray, C. and Rougeon, F. (1980). Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumour RNA. *Eur. J. Biochem.*, **107**, 303-314.

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

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

Barbas, J., Chaix, J.-C., Steinmetz, A. and Goridis, C. (1988). Differential splicing and alternative polyadenylation generates distinct NCAM transcripts and proteins in the mouse. *EMBO J.*, **7**, 625-632.

Barthels, D., Santoni, M.J., Willie, W., Ruppert, C., Chaix, J.C., Hirsch, M.R., Fontecillacamps, J.C. and Goridis, C. (1987). Isolation and nucleotide sequence of mouse NCAM cDNA that encodes for a Mr 79,000 polypeptide without a membrane spanning region. *EMBO J.*, **6**, 907-914.

Barton, C., Dickson, G., Gower, H., Rowett, L., Putt, W., Elsom, V., Moore, S., Goridis, C. and Walsh, F. (1988). Complete sequence and in vitro expression of a tissue-specific phosphatidylinositol-linked N-CAM isoform from skeletal muscle. *Development*, **104**, 165-173.

Basi, G.S. and Storti, R.V. (1986). Structure and DNA sequence of the tropomyosin I gene from *Drosophila melanogaster*. *J. Biol. Chem.*, **261**, 817-827.

Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988). *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell*, **55**, 1037-1046.

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

Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol. Cell. Biol.*, **12**, 3165-3175.

Berget, S.M. and Roberson, B.L. (1986). U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell*, **46**, 691-696.

- Bernués, J., Beltrán, R., Casasnovas, J.M. and Azorín, F. (1989). Structural polymorphism of homopurine-homopyrimidine sequences: the secondary DNA structure adopted by a d(GA.CT)₂₂ sequence in the presence of zinc ions. *EMBO J.*, **8**, 2087-2094.
- Bindereif, A. and Green, M.R. (1986). Ribonucleoprotein complex formation during pre-mRNA splicing in vitro. *Mol. Cell. Biol.*, **6**, 2582-2592.
- Bindereif, A. and Green, M.R. (1987). An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.*, **6**, 2415-2424.
- Bingham, P.M., Chou, T.-Z., Mims, I. and Zachar, Z. (1988). On/off regulation of gene expression at the level of splicing. *Trends in Genetics*, **4**, 134-138.
- Black, D.L., Chabot, B. and Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell*, **42**, 737-750.
- Black, D.L. and Steitz, J.A. (1986). Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell*, **46**, 697-704.
- Black, D.L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron specific exon in nonneuronal cells. *Genes and Development*, **5**, 389-402.
- Blake, C.C.F. (1978). Do genes-in-pieces imply proteins-in-pieces. *Nature*, **273**, 267-268.
- Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. and McKeown, N. (1987). Regulation of sexual differentiation in *Drosophila melanogaster* via alternative splicing of RNA from the transformer gene. *Cell*, **50**, 739-747.
- Bradac, J.A., Gruber, C.E., Forry-Schaudies, S. and Hughes, S.H. (1989). Isolation and characterization of related cDNA clones encoding skeletal muscle β -tropomyosin and a low-molecular-weight nonmuscle tropomyosin isoform. *Mol. Cell. Biol.*, **9**, 185-192.
- Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.*, **50**, 349-383.

Breitbart, R.E. and Nadal-Ginard, B. (1986). Complete nucleotide sequence of the fast skeletal troponin T gene: Alternatively spliced exons exhibit unusual interspecies divergence. *J. Mol. Biol.*, **188**, 313-324.

Breitbart, R.E., Andreadis, A. and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Ann. Rev. Biochem.*, **56**, 467-495.

Breitbart, R.E. and Nadal-Ginard, B. (1987). Developmentally induced, muscle-specific *trans* factors control the differential splicing of alternative and constitutive troponin T exons. *Cell*, **49**, 793-803.

Bruzik, J.P., Van Doren, K., Hirsh, D. and Steitz, J.A. (1988). *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature*, **335**, 559.

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

Casasnovas, J.M., Ellison, M.J., Rodriguezcampos, A., Martinezbalbas, A. and Azorin, F. (1989). In vivo assessment of the Z-DNA-forming potential of d(CA.GT)_n and d(CG.GC)_n sequences cloned into SV40 minichromosomes. *J. Mol. Biol.*, **208**, 537-549.

Cavalier-Smith, T. (1985). Selfish DNA and the origin of introns. *Nature*, **315**, 283-284.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. *Trends in Genetics*, **7**, 145-148.

Cech, T.R. (1990). Self-splicing of group I introns. *Ann. Rev. Biochem.*, **59**, 543-568.
Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science*, **230**, 1344-1349.

Chabot, B. and Steitz, J.A. (1987). Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. *Mol. Cell. Biol*, **7**, 281-293.

- Chabot, B. and Steitz, J.A. (1987). Recognition of mutant and cryptic 5' splice sites by the U1 small nuclear ribonucleoprotein in vitro. *Mol. Cell. Biol.*, **7**, 698-707.
- Chakrabartty, A., Schellman, J.A. and Baldwin, R.L. (1991). Large differences in the helix propensities of alanine and glycine. *Nature*, **351**, 586-588.
- Choi, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. (1986). Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science*, **231**, 1534-1539.
- Chou, T.-B., Zachar, Z. and Bingham, P.M. (1987). Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J.*, **6**, 4095-4104.
- Clackson, T. and Winter, G. (1989). Sticky feet-directed mutagenesis and its application to swapping antibody domains. *Nucleic Acids Res.*, **17**, 10163-10170.
- Clayton, L., Reinach, F.C., Chumbley, G.M. and Macleod, A.R. (1988). Organization of the hTM_{nm} gene: Implications for the evolution of muscle and non-muscle tropomyosins. *J. Mol. Biol.*, **201**, 507-515.
- Colleaux, L., D'Auriol, L., Gilibert, F. and Dujon, B. (1988). Recognition and cleavage site of the intron-encoded *omega* transposase. *Proc. Natl. Acad. Sci. USA*, **85**, 6022-6026.
- Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M., and Riva, S. (1993). Phosphorylation of human hnRNP protein A1 abrogates in vitro strand annealing activity. *Nucleic Acids Res.*, **21**, 949-955.
- Cooper, T.A., Cardone, M.H. and Ordahl, C.P. (1988). *Cis* requirements for alternative splicing of the cardiac troponin T pre-mRNA. *Nucleic Acids Res.*, **16**, 8443-8465.
- Cooper, T.A. and Ordahl, C.P. (1989). Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. *Nucleic Acids Res.*, **17**, 7905-7921.
- Cooper, T.A. (1992). In vitro splicing of cardiac troponin T precursors (exon mutations disrupt splicing of the upstream intron). *J. Biol. Chem.*, **267**, 5330-5338.
- Cote, G.J., Nguyen, I.N., Lips, C.J.M., Berget, S.M. and Gagel, R.F. (1991). Validation of an *in vitro* RNA processing system for CT/CGRP precursor mRNA. *Nucleic Acids Res.*, **19**, 3601-3606.

Cote, G.J., Stolow, D.T., Peleg, S., Berget, S.M. and Gagel, R.F. (1992). Identification of exon sequences and an exon binding protein involved in alternative RNA splicing of calcitonin CGRP. *Nucleic Acids Res.*, **20**, 2361-2366.

Cunningham, B., Hemperley, J., Murray, B., Prediger, E., Brackenbury, R. and Edelman, G. (1987). Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. *Science*, **236**, 799-806.

Cunningham, S.A., Else, A.J., Potter, B.V.L. and Eperon, I.C. (1991). Influences of separation and adjacent sequences on the use of alternative 5' splice sites. *J. Mol. Biol.*, **217**, 265-281.

D'Orval, B.C., D'Aubenton-Carafa, Y., Sirand-Pugnet, P., Gallego, M., Brody, E. and Marie, J. (1991). RNA secondary structure repression of a muscle-specific exon in HeLa nuclear extracts. *Science*, **252**, 1823-1828.

Datta, B. and Weiner, A.M. (1991). Genetic evidence for base pairing between U2 and U6 SnRNA in mammalian mRNA splicing. *Nature*, **352**, 821-824.

Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989). Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell*, **56**, 431-441.

Dibb, N.J. and Newnan, A.J. (1989). Evidence that introns arose at proto-splice sites. *EMBO J.*, **8**, 2015-2021.

Dibb, N.J. (1991). Proto-splice site model of intron origin. *J. Theor. Biol.*, **151**, 405-416.

Dickson, G., Gower, H., Barton, C., Prentica, H., Elsom, V., Moore, S., Cox, R., Quinn, C., Putt, W. and Walsh, F. (1987). Human muscle neural cell adhesion molecule (N-CAM): Identification of a muscle-specific sequence in the extracellular domain. *Cell*, **50**, 1119-1130.

Dickson, G., Peck, D., Moore, S.E., Barton, C.H. and Walsh, F.S. (1990). Enhanced myogenesis in N-CAM-transfected mouse myoblasts. *Nature*, **344**, 348-351.

Doherty, P., Barton, C., Dickson, G., Seaton, P., Rowett, L., Moore, S., Gower, H. and Walsh, F.S. (1989). Neuronal process outgrowth of human sensory neurons

on monolayers of cells transfected with cDNAs for 5 human N-CAM isoforms. *J. Cell Biol.*, **109**, 789-798.

Doherty, P., Fruns, M., Seaton, P., Dickson, G., Barton, C., Sears, T. and Walsh, F. (1990). A threshold effect of the major isoforms of NCAM on neurite outgrowth. *Nature*, **343**, 464-466.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.*, **19**, 4008.

Doolittle, F.W. and Stoltzfus, A. (1993). Genes in pieces revisited. *Nature*, **361**, 403.

Dorit, R.L., Schoenbach, L. and Gilbert, W. (1990). How big is the universe of exons. *Science*, **250**, 1377-1382.

Dreyfuss, G., Swanson, M.S. and Piñol-Roma, S. (1988). Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.*, **13**, 86-91.

Dujon, B. (1980). Sequence of the intron and flanking exons of the mitochondrial 21S ribosomal-RNA gene of yeast strains having different alleles at the omega-loci and rib-1-loci. *Cell*, **20**, 185-197.

Edelman, G.M. and Crossin, K.L. (1991). Cell-Adhesion molecules: implications for a molecular histology. *Ann. Rev. Biochem.*, **60**, 155-190.

Emeson, R.B., Hedjran, F., Yeakley, J.M., Guise, J.W. and Rosenfeld, M.G. (1989). Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific splice acceptor. *Nature*, **341**, 76-80.

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

Eperon, I.C. (1986a). M13 vectors with T7 polymerase promoters: transcription limited by oligonucleotides. *Nucleic Acids Res.*, **14**, 2830.

Eperon, I.C. (1986b). Rapid preparation of bacteriophage DNA for sequence analysis in sets of 96 clones, using filtration. *Analyt. Biochem.*, **156**, 406-412.

Eperon, L.P., Graham, I.R., Griffiths, A.D. and Eperon, I.C. (1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? *Cell*, **54**, 393-401.

Eperon, I.C. and Hamshire, M. (1993). Applications of gene transfer to study RNA splicing in mammalian cell lines. *J. Tissue Culture Methods*, in press,

Erlich, H.A. (1989). *PCR Technology: Principles and Applications for DNA Technology*. Stockton Press. London.

Fedorov, A., Suboch, G., Buyakov, M. and Fedorova, L. (1992). Analysis of nonuniformity in intron phase distribution. *Nucleic Acids Res.*, **20**, 2553-2557.

Forry-Schaudies, S., Maihle, N.J. and Hughes, S.H. (1990). Generation of skeletal, smooth and low molecular weight isoforms from the chicken tropomyosin 1 gene. *J. Mol. Biol.*, **211**, 321-330.

Fu, X.D, Mayede, A., Maniatis, T. and Krainer, A.R. (1992). General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. *Proc. Natl. Acad. Sci. USA*, **89**, 11224-11228.

Galli, G., Guise, J., Tucker, P.W. and Nevins, J.R. (1988). Poly(A) site choice rather than splice site choice governs the regulated production of IgM heavy-chain mRNAs. *Proc. Natl. Acad. Sci. USA*, **85**, 2439-2443.

Gally, J.A. and Edelman, G.M. (1992). Evidence for gene conversion in genes for cell-adhesion molecules. *Proc. Natl. Acad. Sci. USA*, **89**, 3276-3279.

Garcia-Blanco, M.A., Jamison, S.F. and Sharp, P.A. (1989). Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.*, **3**, 1874-1886.

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

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

Gerke, V. and Steitz, J.A. (1986). A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. *Cell*, 47, 973-984.

Gil, A., Sharp, P.A., Jamison, S.F. and Garcia-Blanco, M.A. (1991). Characterization of cDNAs encoding the polypyrimidine tract-binding protein. *Genes Dev.*, 5, 1224-1236.

Gilbert, W. (1978). Why genes in pieces. *Nature*, 271, 501.

Gilbert, W. (1986). The RNA world. *Nature*, 319, 618.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. *Cell*, 23, 175-182.

Gò, M. (1982). Correlation of DNA exomic regions with protein structural units in haemoglobin. *Nature*, 291, 90-92.

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

Gorman, C. (1985). High efficiency gene transfer into mammalian cells. In *DNA Cloning* Vol 2, ed. D.M. Glover. IRL Press, pp. 143-190.

Goux-Pelletan, M., Libri, D., d'Aubenton-Carafa, Y., Fiszman, M.Y., Brody, E. and Marie, J. (1990). *In vitro* splicing of mutually exclusive exons from the chicken β -tropomyosin gene: role of the branch point location and very long pyrimidine stretch. *EMBO J.*, 9, 241-249.

Gower, H., Barton, C., Elsom, V., Thompson, J., Moore, S., Dickson, G. and Walsh, F. (1988). Alternative splicing generates a secreted form of N-CAM in muscle and brain. *Cell*, 55, 955-964.

Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984). Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell*, 37, 415-427.

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

- Grabowski, P.J. and Sharp, P.A. (1986). Affinity chromatography of splicing complexes: U2, U5, and U4+U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science*, **233**, 1294-1299.
- Graham, F.L. and van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, **52**, 456-467.
- Graham, I.R. (1992). Alternative splicing of tropomyosin pre-mRNA: control in non-muscle cells. *PhD Thesis*. University of Leicester.
- Graham, I.R., Hamshire, M. and Eperon, I.C. (1992). Alternative splicing of a human α -tropomyosin muscle-specific exon: identification of determining sequences. *Mol. Cell. Biol.*, **12**, 3872-3882.
- Grandino, B., Campuzano, S. and Sánchez, L. (1990). The *Drosophila melanogaster* $\delta I(2)d$ gene is needed for the female-specific splicing of sex-lethal RNA. *EMBO J.*, **9**, 2597-2602.
- Green, M.R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Ann. Rev. Cell. Biol.*, **7**, 559-599.
- Gunning, P., Gordon, M., Wade, R., Gahlmann, R., Lin, C.-S. and Hardeman, E. (1990). Differential control of tropomyosin mRNA levels during myogenesis suggests the existence of an isoform competition-autoregulatory compensation control mechanism. *Dev. Biol.*, **138**, 443-453.
- Guo, W., Mulligan, G.J., Wormsley, S. and Helfman, D.M. (1991). Alternative splicing of β -tropomyosin pre-mRNA: *cis*-acting sequences and cellular factors that block the use of a skeletal muscle exon in non-muscle cells. *Genes Dev.*, **5**, 2096-2107.
- Guthrie, C. and Patterson, B. (1988). Spliceosomal SnRNAs. *Ann. Rev. Genet.*, **22**, 387-419.
- Hall, K.B., Green, M.R. and Redfield, A.G. (1988). Structure of a pre-mRNA branch point/3' splice site region. *Proc. Natl. Acad. Sci. USA*, **85**, 704-708.
- Hall, D.H., Liu, Y. and Shub, D.A. (1989). Exon shuffling by recombination between self splicing introns of bacteriophage T4. *Nature*, **340**, 574-576.

Hamshire, M., Dickson, G. and Eperon, I.C. (1991). The muscle specific domain of mouse N-CAM: structure and alternative splicing patterns. *Nucleic Acids Res.*, **19**, 4709-4716.

Hanke, P.D. and Storti, R.V. (1988). The *Drosophila melanogaster* tropomyosin II gene produces multiple proteins by use of alternative tissue-specific promoters and alternative splicing. *Mol. Cell. Biol.*, **8**, 3591-3602.

Harper, J.E. and Manley, J.L. (1991). A novel protein factor is required for use of distal alternative 5' splice sites *in vitro*. *Mol. Cell. Biol.*, **11**, 5945-5953.

Hedley, M.L. and Maniatis, T. (1991). Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to the *tra-2* protein *in vitro*. *Cell*, **65**, 579-586.

Hedrick, P.W. (1985). *Genetics of Populations*. Jones and Bartlett Publications inc.,

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

Helfman, D.M., Ricci, W.M. and Finn, L.A. (1988). Alternative splicing of tropomyosin pre-mRNAs *in vitro* and *in vivo*. *Genes Dev.*, **2**, 1627-1638.

Helfman, D.M., Erster, S., Lees-miller, J., Cheley, S. and Finn, L. (1988). Molecular and cellular aspects of non-muscle and muscle tropomyosins. *J. Muscle Res. Cell Mobility*, **9**, 102.

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

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

Hernandez, N. and Keller, W. (1983). Splicing of *in vitro* synthesized messenger RNA precursors in HeLa cell extracts. *Cell*, **35**, 89-99.

- Hickey, D.A., Bally-Cuif, L., Abukashawa, S., Payant, V. and Benkel, B.F. (1991). Concerted evolution of duplicated protein-coding genes in *Drosophila*. *Proc. Natl. Acad. Sci. USA*, 88, 1611-1615.
- Hickley, D.A. (1989). A general model for the evolution of nuclear pre-mRNA introns. *J. Theor. Biol.*, 137, 41-53.
- Hoch, B., Maier, R.M., Appel, K., Igbi, G.L. and Kössel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, 353, 178-180.
- Hodgkin, J. (1989). *Drosophila* sex determination: a cascade of regulated splicing. *Cell*, 56, 905-906.
- Hornig, H., Aebi, M. and Weissmann, C. (1986). Effect of mutations at the lariat branch acceptor site on β -globin pre-mRNA splicing *in vitro*. *Nature*, 324, 589-591.
- Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. and Shimura, Y. (1991). Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science*, 252, 833-836.
- Htun, H. and Dahlberg, J.E. (1989). Topology and formation of triple-stranded M-DNA. *Science*, 243, 1571-1576.
- Hunkapiller, T. and Hood, L. (1986). The growing immunoglobulin gene superfamily. *Nature*, 323, 15-16.
- Inoue, K., Hoshijima, K., Sakamoto, H. and Shimura, Y. (1990). Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript. *Nature*, 344,
- Jeffreys, A., Wilson, V., Neumann, R. and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.*, 16, 10953-10971.
- Jiménez-García, L.F. and Spector, D.L. (1993). In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell*, 73, 47-59.
- Johansen, S., Johansen, T. and Haugli, F. (1992). Structure and evolution of myxomycete nuclear group I introns: a model for horizontal transfer by intron homing. *Current Genetics*, 22, 297-304.

- Jones, D.S.C. and Schofield, J.P. (1990). A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Res.*, **18**, 7463-7464.
- Kadman, G., Kowitz, A., Altevogt, P. and Schachner, M. (1990a). Functional co-operation between the neural adhesion molecules L1 and N-CAM is carbohydrate dependent. *J. Cell Biol.*, **110**, 209-218.
- Kadman, G., Kowitz, A., Altevogt, P. and Schachner, M. (1990b). The neural cell adhesion molecule N-CAM enhances L1-dependent cell-cell interactions. *J. Cell Biol.*, **110**, 193-208.
- Katzenberg, D.R., Tilley, S.A. and Birshstein, B.K. (1989). Nucleotide sequence of an unequal sister chromatid exchange site in a mouse myeloma cell line. *Mol. Cell Biol.*, **9**, 1324-1326.
- Keller, E.B. and Noon, W.A. (1984). Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci. USA*, **81**, 7417-7420.
- Kole, R. and Weissman, S.M. (1982). Accurate in vitro splicing of human β globin RNA. *Nucleic Acids Res.*, **10**, 5429-5445.
- Konarska, M.M., Padgett, R.A. and Sharp, P.A. (1985). *Trans* splicing of mRNA precursors in vitro. *Cell*, **42**, 165-171.
- Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985). Characterization of the branch site in lariat RNA produced by splicing of mRNA precursors. *Nature*, **313**, 552-557.
- Konarska, M.M. and Sharp, P.A. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell*, **49**, 763-774.
- Konarska, M.M. and Sharp, P.A. (1988). Association of U2, U4, U5 and U6 small nuclear ribonucleoproteins in a spliceosome-type complex in absence of precursor RNA. *Proc. Natl. Acad. Sci. USA*, **85**, 5459-5462.
- Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984). Normal and mutant human β -globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell*, **36**, 993-1005.

Krainer, A.R. and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell*, **42**, 725-736.

Krainer, A.R. and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell*, **42**, 725-736.

Krainer, A.R., Conway, G.C. and Kozak, D. (1990a). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.*, **4**, 1158-1171.

Krainer, A.R., Conway, G.C. and Kozak, D. (1990b). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*, **62**, 35-42.

Krainer, A.R., Mayeda, A., Kozak, D. and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and Drosophila splicing regulators. *Cell*, **66**, 383-394.

Kramer, A., Frick, M. and Keller, W. (1987). Separation of multiple components of HeLa cell nuclear extracts required for pre-messenger RNA splicing. *J. Biol. Chem.*, **262**, 17630-17640.

Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984). The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors.

Krämer, A. (1988). Pre-splicing complex formation requires 2 proteins and U2 snRNP. *Genes Dev.*, **2**, 1155-1167.

Krämer, A. and Utans, U. (1991). Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *EMBO J.*, **10**, 1503-1509.

Kusakabe, S., Baba, H., Koga, A., Bewley, G.C. and Mukai, T. (1990). Gene duplication and concerted evolution of the BPDH locus in natural populations of *Drosophila melanogaster*. *Proc. Roy. Soc. Lond. B*, **242**, 157-162.

Kühne, T., Wieringa, B., Reiser, J. and Weissmann, C. (1983). Evidence against a scanning model of RNA splicing. *EMBO J.*, **2**, 727-733.

Lambowitz, A.M. (1989). Infectious introns. *Cell*, **56**, 323-326.

Lamond, A.I., Konarska, M.M. and Sharp, P.A. (1987). A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.*, **1**, 532-543.

Lamond, A.I. (1991a). Nuclear RNA processing. *Current Opinion in Biology*, **3**, 493-501.

Lamond, A.I. (1991b). ASF/SF2 - a splice site selector. *Trends Bioch. Sci.* **16** 452-453.

Lang, K.M. and Spritz, R.A. (1983). RNA splice site selection: evidence for a 5'→3' scanning model. *Science*, **220**, 1351-1355.

Laski, F.A., Rio, D.C. and Rubin, G.M. (1986). Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell*, **44**, 7-19.

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 Dev.*, **3**, 720-728.

Lear, A.L., Eperon, L.P., Wheatley, I.M. and Eperon, I.C. (1990). Hierarchy for 5' splice site preference determined *in vivo*. *J. Mol. Biol.*, **211**, 103-115.

Lees-Miller, J.P., Goodwin, L.O. and Helfman, D.M. (1990a). Three novel brain tropomyosin isoforms are expressed from the rat α -tropomyosin gene through the use of alternative promoters and alternative RNA processing. *Mol. Cell. Biol.*, **10**, 1729-1742.

Lees-Miller, J.P., Yan, A. and Helfman, D.M. (1990b). Structure and nucleotide sequence of the gene encoding rat fibroblast tropomyosin 4. *J. Mol. Biol.*, **213**, 399-405.

Leff, S.E., Evans, R.M. and Rosenfeld, M.G. (1987). Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. *Cell*, **48**, 517-524.

- Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980). Are snRNPs involved in splicing? *Nature*, **283**, 220-224.
- Levitt, N., Briggs, D., Gil, A. and Proudfoot, N.J. (1989). Definition of an efficient poly(A) site. *Genes Dev.*, **3**, 1019-1025.
- Libri, D., Lemonnier, M., Meinnel, T. and Fiszman, M.Y. (1989a). A single gene codes for the β subunits of smooth and skeletal muscle tropomyosin in the chicken. *J. Biol. Chem.*, **264**, 2935-2944.
- Libri, D., Marie, J., Brody, E. and Fiszman, M.Y. (1989b). A subfragment of the β tropomyosin gene is alternatively spliced when transfected into differentiating muscle cells. *Nucleic Acids Res.*, **17**, 6449-6462.
- Libri, D., Mouly, V., Lemonnier, M. and Fiszman, M.Y. (1990a). A nonmuscle tropomyosin is encoded by the smooth/skeletal β -tropomyosin gene and its RNA is transcribed from an internal promoter. *J. Biol. Chem.*, **265**, 3471-3473.
- Libri, D., Goux-Pelletan, M., Brody, E. and Fiszman, M.Y. (1990b). Exon as well as intron sequences are *cis*-regulating elements for the mutually exclusive alternative splicing of the β tropomyosin gene. *Mol. Cell. Biol.*, **10**, 5036-5046.
- Libri, D., Piseri, A. and Fiszman, M.Y. (1991). Tissue-specific splicing in vivo of the β -tropomyosin gene: dependence on an RNA secondary structure. *Science*, **252**, 1842-1845.
- Libri, D., Balvay, L. and Fiszman, M.Y. (1992). In vivo splicing of β tropomyosin pre-mRNA: a role for branch point and donor site competition. *Mol. Cell. Biol.*, **12**, 3204-3215.
- Lindquester, G.J., Flach, J.E., Fleenor, D.E., Hickman, K.H. and Devlin, R.B. (1989). Avian tropomyosin gene expression. *Nucleic Acids Res.*, **17**, 2099-2118.
- Liskay, R.M., Lietsou, A. and Stachelek, J.L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics*, **115**, 161-167.
- Loomis, W.F. and Gilpin, M.E. (1986). Multigene families and vestigial sequences. *Proc. Natl. Acad. Sci. USA*, **83**, 2143-2147.

Macleod, A.R. and Talbot, K. (1983). A processed gene defining a gene family encoding a human non-muscle tropomyosin. *J. Mol. Biol.*, **167**, 523-537.

Macleod, A.R., Houlker, C., Reinach, F.C., Smillie, L.B., Talbot, K., Modi, G. and Walsh, F.S. (1985). A muscle-type tropomyosin in human fibroblasts: Evidence for expression by an alternative RNA splicing mechanism. *Proc. Natl. Acad. Sci. USA*, **82**, 7835-7839.

Macleod, A.R., Houlker, C., Reinach, F.C. and Talbot, K. (1986). The mRNA and RNA-copy pseudogenes encoding TM30_{nm}, a human cytoskeletal tropomyosin. *Nucleic Acids Res.*, **14**, 8413-8426.

Macleod, A.R., Talbot, K., Reinach, F.C., Houlker, C., Smillie, L.B., Giometti, C.S. and Anderson, N.L. (1986). Molecular characterization of human cytoskeletal tropomyosins. In *Molecular biology of muscle development* Vol ed. C. Emerson, D.A. Fishman, B. Nadal-Ginard and M.A.Q. Siddiqui. Alan R. Liss, Inc, New York. pp. 445-455.

Macleod, A.R., Talbot, K., Smillie, L.B. and Houlker, C. (1987). Characterization of a cDNA defining a gene family encoding TM30_{pl}, a human fibroblast tropomyosin. *J. Mol. Biol.*, **194**, 1-10.

Macleod, A.R. and Gooding, C. (1988). Human hTM α gene: Expression in muscle and nonmuscle tissue. *Mol. Cell. Biol.*, **8**, 433-440.

Madhani, H.D. and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 SnRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**, 803-817.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maniatis, T. and Reed, R. (1987). The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature*, **325**, 673-678.

Maniatis, T. (1991). Mechanisms of alternative pre-mRNA splicing. *Science*, **251**, 33-34.

Mattaj, I.W. (1989). A binding consensus: RNA-protein interactions in splicing, snRNPs and sex. *Cell*, **57**, 1-3.

Mattox, W. and Baker, B.S. (1991). Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila*. *Genes Dev.*, 5, 786-796.

Mayeda, A. and Ohshima, Y. (1988). Short donor site sequences inserted within the intron of β -globin pre-mRNA serve for splicing in vitro. *Mol. Cell. Biol.*, 8, 4484-4491.

Mayeda, A. and Krainer, A.R. (1992). Regulation of alternative pre-messenger-RNA splicing by hnRNP A1 and splicing factor SF-2. *Cell*, 68, 365-375.

Mayeda, A., Zahler, A.M., Krainer, A.R., and Roth, M.B. (1992). 2 members of a conserved family of nuclear phosphoproteins are involved in pre-messenger-RNA splicing. *Proc. Natl. Acad. Sci. USA*, 89, 1301-1304.

Mayeda, A., Helfman, D.M., and Krainer, A.R. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-messenger RNA splicing factor SF2/ASF. *Mol. Cell. Biol.*, 13, 2993-3001.

Mohr, G. and Lambowitz, A.M. (1991). Integration of a group I intron into a ribosomal RNA sequence promoted by a tyrosyl-tRNA synthetase. *Nature*, 354, 164-167.

Moore, S., Thompson, J., Kirkness, V., Dickson, J. and Walsh, F. (1986). Skeletal muscle neural cell adhesion molecule (N-CAM) changes in protein and mRNA species during myogenesis of muscle cell lines. *J. Cell Biol.*, 105, 1377-1386.

Mount, S.M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.*, 10, 459-472.

Mullen, M.P., Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1991). α -Tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev.*, 5, 642-655.

Muscarella, D.E. and Vogt, V.M. (1989). A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell*, 56, 443-454.

Nagoshi, R.N., McKeown, M., Burtis, K.C., Belote, J.M. and Baker, B.S. (1988). The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. *Cell*, 53, 229-236.

Nagoshi, R.N. and Baker, B.S. (1990). Regulation of sex-specific RNA splicing at the *Drosophila doublesex* gene: *cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes Dev.*, **4**, 89-97.

Nelson, K.K. and Green, M.R. (1988). Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev.*, **2**, 319-329.

Newman, A. and Norman, C. (1991). Mutations in yeast U5 snRNA alter the specificity of 5' splice site cleavage. *Cell*, **65**, 115-123.

Newman, A.J. and Norman, C. (1992). U5 SnRNA interacts with exon sequences at 5' and 3' splice sites. *Cell*, **68**, 743-754.

Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D.,

Kinzler, K.W. and Vogelstein, B. (1991). Scrambled exons. *Cell*, **64**, 607-613.

Niwa, M., Macdonald, C.C. and Berget, S.M. (1992). Are vertebrate exons scanned during splice-site selection. *Nature*, **360**, 277-280.

Owens, G., Edelman, G. and Cunningham, B. (1987). Organisation of the neural cell adhesion molecule (N-CAM) gene: alternative exon usage as the basis for different membrane-associated domains. *Proc. Natl. Acad. Sci. USA*, **84**, 294-298.

Padgett, R.A., Mount, S.M., Steitz, J.A. and Sharp, P.A. (1983). Splicing of messenger RNA precursors is inhibited by antisera to small nuclear ribonucleoprotein. *Cell*, **35**, 101-107.

Padgett, R.A., Hardy, S.F. and Sharp, P.A. (1983). Splicing of adenovirus RNA in a cell-free transcription system. *Proc. Natl. Acad. Sci. USA*, **80**, 5230-5234.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, **225**, 898-903.

Padgett, R.A., Konarska, M.M., Aebi, M., Hornig, H., Weissmann, C. and Sharp, P.A. (1985). Nonconsensus branch-site sequences in the *in vitro* splicing of transcripts of mutant rabbit β -globin genes. *Proc. Natl. Acad. Sci. USA*, **82**, 8349-8353.

Padgett, R.A., Grabowski, P.J., Konarska, M.M. and Sharp, P.A. (1985). Splicing messenger RNA precursors: branch sites and lariat RNAs. *Trends Biochem. Sci.*, **10**, 154-157.

Palton, J.G. and Nadal-Ginard, B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein. A component of a complex necessary for pre-messenger RNA splicing. *Genes and Development*, **5**, 1237-1251.

Paterson, M.L. and Perry, R.D. (1989). The regulated production of μ m and μ s mRNA is dependant on the efficiencies of μ s Poly(A) site usage and the C μ 4-to-M1 splice. *Mol. Cell. Biol.*, **9**, 726-738.

Patton, J.G., Mayer, S.A., Tempst, P. and Nadal-Ginard, B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev.*, **5**, 1237-1251.

Pearson-While, S.H. and Emerson, C.P. (1987). Anovel hybrid α -tropomysin in fibroblasts is produced by alternative splicing off transcripts from the skeletal muscle α -tropomyosin gene. *J. Biol. Chem.*, **262**, 15998-16010.

Pompliano, D.L. and Knowles, J.R. (1992). Exons as microgenes. *Science*, **257**, 1489-1490.

Prediger, E., Hoffman, S., Edelman, G. and Cunningham, B. (1988). Four exons encode a 93-base-pair insert in three neural cell adhesion molecule mRNAs specific for chicken heart and skeletal muscle. *Proc. Natl. Acad. Sci. USA*, **85**, 9616-9620.

Quirk, S.M., Bell-Pedersen, D. and Belfort, M. (1989). Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell*, **56**, 455-465.

Rautmann, G. and Breathnach, R. (1985). A role for branchpoints in splicing *in vivo*. *Nature*, **315**, 430-432.

Reed, R. and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell*, **41**, 95-105.

- Reed, R. and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell*, **46**, 681-690.
- Reed, R. and Maniatis, T. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev.*, **2**, 1268-1276.
- Reed, R., Griffith, J. and Maniatis, T. (1988). Purification and visualization of native spliceosomes. *Cell*, **53**, 949-961.
- Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113-2123.
- Reyes, A.A., Small, A.J. and Akeson, R. (1991). At least 27 alternatively spliced isoforms of the neural cell adhesion molecule messenger RNA are expressed during rat-heart development. *Mol. Cell. Biol.* **11**, 1654-1661.
- Robberson, B.L., Cote, G.J. and Berget, S.M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.*, **10**, 84-94.
- Rothstein, D.M., Saito, H., Streuli, M., Schlossman, S.F. and Morimoto, C. (1992). The alternative splicing of the CD45 tyrosine phosphate is controlled by negative trans-acting splicing factors. *J. Biol. Chem.*, **267**, 7139-7147.
- Ruby, S.W. and Abelson, J. (1988). An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science*, **242**, 1028-1035.
- Ruiz-Opazo, N., Weinberger, J. and Nadal-Ginard, B. (1985). Comparison of α -tropomyosin sequences from smooth and striated muscle. *Nature*, **315**, 67-70.
- Ruiz-Opazo, N. and Nadal-Ginard, B. (1987). α -tropomyosin gene organization. *J. Biol. Chem.*, **262**, 4755-4765.
- Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, **38**, 317-331.
- Ruskin, B., Greene, J.M. and Green, M.R. (1985). Cryptic branch point activation allows accurate in vitro splicing of human β -globin intron mutants. *Cell*, **41**, 833-844.

- Ruskin, B. and Green, M.R. (1985a). An RNA processing activity that debranches RNA lariats. *Science*, **229**, 135-140.
- Ruskin, B. and Green, M.R. (1985b). Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature*, **317**, 732-734.
- Ruskin, B., Zamore, P.D. and Green, M.R. (1988). A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*, **52**, 207-219.
- Rutishauser, U., Acheson, A., Hall, A., Mann, D. and Sunshine, J. (1988). The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. *Science*, **240**, 53-57.
- Ryner, L.C. and Baker, B.S. (1991). Regulation of doublesex pre-mRNA processing occurs by 3' splice site activation. *Genes and Development*, **5**, 2071-2085.
- Saiki, R.K., Scharf, S., Falsona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science*, **230**, 1350-1354.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanders, C. and Smillie, L.B. (1985). Amino acid sequence of chicken gizzard β -tropomyosin. *J. Biol. Chem.*, **260**, 7264-7275.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.*, **74**, 5463-5467.
- Santoni, M., Barthels, D., Vopper, G., Boned, A., Goridis, C. and Willie, W. (1988). Differential exon usage involving an unusual splicing mechanism generates at least eight types of NCAM cDNA in mouse brain. *EMBO J.*, **8**, 385-392.
- Seidel, H.M., Pompliano, D.L. and Knowles, J.R. (1992a). Exons as microgenes. *Science*, **257**, 1489-1490.
- Seidel, H.M., Pompliano, D.L. and Knowles, J.R. (1992b). Phosphate biosynthesis - molecular cloning of the gene for phosphoenolpyruvate mutase from *Tetrahymena pyriformis* and overexpression of the gene-product in *E. coli*. *Biochemistry*, **31**, 2598-2608.

Senapathy, P. (1986). Origin of eukaryotic introns: A hypothesis, based on codon distribution statistics of genes, and its implications. *Proc. Natl. Acad. Sci. USA*, **83**, 2133-2137.

Senapathy, P. (1988). Possible evolution of splice-junction signals in eukaryotic genes from stop codons. *Proc. Natl. Acad. Sci. USA*, **85**, 1129-1133.

Shapiro, M.B. and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*, **15**, 7155-7174.

Sharp, P.A. (1981). Speculations on RNA splicing. *Cell*, **23**, 643-646.

Sharp, P.A. (1985). On the origin of RNA splicing and introns. *Cell*, **42**, 397-400.

Sharp, P.A. (1987). Splicing of messenger RNA precursors. *Science*, **235**, 766-771.

Siebel, C.W. and Rio, D.C. (1990). Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. *Science*, **248**, 1200-1208.

Sisodia, S.S., Cleveland, D.W. and Sollner-Webb, B. (1987). A combination of RNase H and S1 nuclease circumvents an artefact inherent to conventional S1 analysis of RNA splicing. *Nucleic Acids Res.*, **15**, 1995-2011.

Small, S.J. and Akeson, R. (1990). Expression of the unique N-CAM VASE is independently regulated in distinct tissues during development. *J. Cell Biol.* **111**, 2089-2096.

Small, S.J., Shull, G.E., Santoni, M.J. and Akeson, R. (1987). Identification of a cDNA clone that contains the complete coding sequence for a 140 kD rat NCAM polypeptide. *J. Cell. Biol.*, **105**, 2335-2345.

Smith, C.W.J., Porro, E.B., Patton, J.G. and Nadal-Ginard, B. (1989). Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature*, **342**, 243-247.

Smith, C.W.J. and Nadal-Ginard, B. (1989). Mutually exclusive splicing of α -tropomyosin exons enforced by an unusual lariat branch point location: Implications for constitutive splicing. *Cell*, **56**, 749-758.

Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989). Alternative splicing in the control of gene expression. *Annu. Rev. Genet.*, **23**, 527-577.

Smith, C.W.J., Gallego, M.E., Andreadis, A., Gooding, C. and Nadal-Ginard, B. (1989). *Investigation of the mechanism of mutually exclusive splicing in the α -tropomyosin, troponin-T and myosin 1/3 genes*. Alan R. Liss inc.,

Solnick, D. (1985). Alternative splicing caused by RNA secondary structure. *Cell*, **43**, 667-676.

Solnick, D. (1985). *Trans* splicing of mRNA precursors. *Cell*, **42**, 157-164.

Solnick, D. and Lee, S.I. (1987). Amount of secondary structure required to induce an alternative splice. *Mol. Cell. Biol.*, **7**, 3194-3198.

Sorkin, B.C., Gallin, W.J., Edelman, G.M. and Cunningham, B.A. (1991). Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome. *Proc. Natl. Acad. Sci. USA*, **88**, 11545-11549.

Sosnowski, B.A., Belote, J.M. and McKeown, M. (1989). Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. *Cell*, **58**, 449-459.

Sri-Widada, J., Ferraz, C., Colote, S., Liautard, J.-P. and Marti, J. (1989). *Re-evaluation of the tropomyosin isoform diversity: gene classification and structure function relationship*. Alan R. Liss inc.,

Steitz, J.A. (1992). Splicing takes a Holliday. *Science*, **257**, 888-889.

Streuli, M. and Saito, H. (1989). Regulation of tissue-specific alternative splicing: exon-specific *cis*-elements govern the splicing of leukocyte common antigen pre-mRNA. *EMBO J.*, **8**, 787-796.

Swanson, M.S. and Dreyfus, G. (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J.*, **7**, 3519-3529.

Tacke, R. and Goridis, C. (1991). Alternative splicing in the neuronal cell-adhesion molecule pre-messenger RNA - regulation of exon 18 skipping depends on the 5' splice site. *Genes and Development*, **5**, 149.

Tacke, R., Boned, A., and Goridis, C. (1992). ASF alternative transcripts are highly conserved between mouse and man. *Nucl. Acids Res.* **20**, 5482

Takenaga, K., Nakamura, Y., Tokunagag, K., Kageyama, H. and Sakiyama, S. (1988). Isolation and characterization of a cDNA that encodes mouse fibroblast tropomyosin isoform 2. *Mol. Cell. Biol.*, **8**, 5561-5565.

Talbot, K. and Macleod, A.R. (1983). Novel form of non-muscle tropomyosin in human fibroblasts. *J. Mol. Biol.*, **164**, 159-174.

Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986). A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. *Cell*, **47**, 755-766.

Thompson, J., Dickson, G., Moore, S., Gower, H., Putt, W., Kenimer, J., Barton, C. and Walsh, F. (1989). Alternative splicing of the neural cell adhesion molecule gene generates variant extracellular domain structure in skeletal muscle and brain. *Genes and Development*, **3**, 348-357.

Tian, M. and Maniatis, T. (1992). Positive control of pre-mRNA splicing in vitro. *Science*, **256**, 237-240.

Tittiger, G., Whyard, S. and Walker, V.K. (1993). A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. *Nature*, **361**, 470-472.

Tsai, A.Y.M., Streuli, M. and Saito, H. (1989). Integrity of the exon 6 sequence is essential for tissue-specific alternative splicing of human leukocyte common antigen pre-mRNA. *Mol. Cell. Biol.*, **9**, 4550-4555.

Tseng, J.C., Zollman, S., Chain, A.C. and Laski, F.A. (1991). Splicing of the *Drosophila* P element ORF2-ORF3 intron is inhibited in a human cell extract. *Mechanisms of Development*, **35**, 65-72.

Tsurushita, N. and Korn, L.J. (1987). Effects of intron length on differential processing of mouse μ heavy-chain mRNA. *Mol. Cell. Biol.*, **7**, 2602-2605.

Walsh, F. and Dickson, G. (1989). Generation of multiple N-CAM polypeptides from a single gene. *Bioessays*, **11**, 83-88.

- Weiner, A. (1993). mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism. *Cell*, 72, 161-164.
- Wells, R.D. (1988). Unusual DNA Structures. *J. Biol. Chem.*, 263, 1095.
- Wenzlau, J.M., Saldanha, R.J., Butow, R.A. and Perlman, P.S. (1989). A latent intron-encoded maturase is also an endonuclease needed for intron mobility. *Cell*, 56, 421-430.
- Wieczorek, D.F. (1988). Regulation of alternatively spliced α -tropomyosin gene expression by nerve extract. *J. Biol. Chem.*, 263, 10456-10463.
- Wieczorek, D.F., Smith, C.W.J. and Nadal-Ginard, B. (1988). The rat α -tropomyosin gene generates a minimum of six different mRNAs coding for striated, smooth, and nonmuscle isoforms by alternative splicing. *Mol. Cell. Biol.*, 8, 679-694.
- Wieringa, B., Meyer, F., Reiser, J. and Weissmann, C. (1983). Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β -globin gene. *Nature*, 301, 38-43.
- Witkowski, J.A. (1988). The discovery of 'split' genes: a scientific revolution. *Trends Biochem. Sci.*, 13, 110-113.
- Wolfner, M.F. (1988). Sex-specific gene expression in somatic tissues of *Drosophila melanogaster*. *Trends in Genetics*, 4, 333-337.
- Woodson, S.A. and Cech, T.R. (1989). Reverse self-splicing of the tetrahymena group I intron: implications for the directionality of splicing and for intron transposition. *Cell*, 57, 335-345.
- Wu, J. and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base-pairing. *Genes Dev.*, 3, 1553-1561.
- Wysocki, L.J. and Gefter, M.L. (1989). Gene conversion and the generation of antibody diversity. *Ann. Rev. Biochem.*, 58, 509-531.
- Yamawaki-Kataoka, Y. and Helfman, D.M. (1987). Isolation and characterization of cDNA clones encoding a low molecular weight nonmuscle tropomyosin isoform. *J. Biol. Chem.*, 262, 10791-10800.

Zachar, Z., Chou, T.-B. and Bingham, P.M. (1987). Evidence that a regulatory gene autoregulates splicing of its transcript. *EMBO J.*, **6**, 4105-4111.

Zahler, A.M., Neugebauer, K.M., Lane, W.S. and Roth, M.B. (1993). Distinct functions of SR proteins in alternative pre-messenger RNA splicing. *Science*, **260**, 219-222.

Zamore, P.D. and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. USA*, **86**, 9243-9247.

Zamore, P.D. and Green, M.R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.*, **10**, 207-214.

Zamore, P.D., Patton, J.G. and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor UZAF. *Nature*, **355**, 609-614.

Zillmann, M., Rose, S.D. and Berget, S.M. (1987). U1 small nuclear ribonucleoproteins are required early during spliceosome assembly. *Mol. Cell. Biol.*, **7**, 2877-2883.

Zoller, M.J. and Smith, M. (1983). Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. *Meths. Enzymol.*, **100**, 468-500.

Zorn, A.M. and Kreig, P.A. (1991). PCR analysis of alternative splicing pathways. Identification of artefacts generated by heteroduplex formation. *Biotechniques*, **11**, 180.