Regulation of alternative splicing by human transformer-2-beta

Den Naturwissenschaftlichen Fakultäten

der Friedrich-Alexander-Universität Erlangen-Nürnberg

zur

Erlangung des Doktorgrades

vorgelegt von
Petar Stoilov
aus Plovdiv, Bulgarien

2002

Als Dissertation genehmigt von den Naturwisenschaftlichen Fakultäten

Der Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 2 Juli 2002

Vorsitzender Der Promotionskommision:	Prof. Dr. A. Margel
Erstberichterstatter:	Prof. Dr. W. Hillen
Zveitberichterstatter:	Prof. Dr. M. Wegner

To my parents.

Publications

Parts of this work are included in the following publications:

Stoilov P, Daoud R, Nayler O and Stamm S, Regulation and function of tra2-beta isoforms, manuscript in preparation

Annakaisa Haapasalo, Ilkka Sipola, Kim Larsson, Karl E.O. Åkerman, **Stoilov P**, Stefan Stamm, Garry Wong and Eero Castrén Regulation of TrkB Surface Expression by Brain-Derived Neurotrophic Factor and Truncated TrkB Isoforms, manuscript in preparation

Stoilov P, Eran Meshorer, Marieta Gencheva, David Glick, Hermona Soreq and Stefan Stamm, Defects in pre-mRNA processing as causes of and predisposition to diseases, submitted for publication

Stoilov P, Castren E, Stamm S. Analysis of the human TrkB gene genomic organization reveals novel TrkB isoforms, unusual gene length, and splicing mechanism. Biochem Biophys Res Commun. 2002;290:1054-65.

Stamm S., Zhu J., Nakai K., **Stoilov P**., Stoss, O. and Zhang M.Q. (2000) An alternative exon database (AEDB) and its statistical analysis, DNA and Cell Biol.,19:739-756

Daoud, R., **Stoilov P**, Stoss O., Hartmann AM, Olbrich M, and Stamm S. (2000) Control of pre-mRNA processing by extracellular signals: from molecular mechanisms to clinical applications. Gene Ther. Mol. Biol. 5: 141-150.

Stoss, O., **Stoilov, P.**, Daoud, R., Hartmann, A.M., Olbrich, M. and Stamm, S. (2000) Misregulation of pre-mRNA splicing that causes human diseases. Gene Ther. Mol. Biol. 5: 9-28

Stoss O, **Stoilov P**, Hartmann AM, Nayler O, Stamm S. The in vivo minigene approach to analyze tissue-specific splicing. Brain Res Brain Res Protoc. 1999 Dec;4(3):383-94.

Content

1. Abbreviations

ADAR	adenosine deaminases that act on RNA
ApoBec	apolypoprotein B editing complex
APP	amyloid precursor protein
ASF	alternative splicing factor
ATP	adenosine triphosphate
BDNF	brain-derived neurotrophic factor
cDNA	complementary DNA
CELF	CUG-BP and ETR3 like factors
CLB	clathrin light chain B
Clk	cdc2 like kinase
ConA	concanavalin A
CTD	carboxiterminal domain (of RNA polymerase II)
cTNT	cardiac troponin T
CUG-BP	CUG repeat binding protein
DMEM	dulbeco's modified Eagle's medium
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleotidtriphosphate
Dsx	doublesex
DTT	dithiothreitol
ECL	enhanced chemiluminiscence
EDTA	ethylenediamine tetra-acetic acid
EGFP	enhanced green flourescent protein
EJC	exon junction complex
ELAV	embrionic lethal abnormal vision
ESE	exonic splicing enhancer
ESI	exonic splicing inhibitor
EST	expressed sequence tag
ETR	elav-type RNA-binding protein
FCS	fetal calf serum
FTDP-17	frontotemporal dementia with parkinsonism linked to chromosome 17
GAPDH	glyceraldehyide-3-phosphateehdrogenase
HEK	human embryonic kidney
hnRNP	heterogenuos nuclear ribonucleoprotein
IG	immunoglobulin
ISE	intronic splicing enhancer
ISI	intronic splicing inhibitor
KH-domain	hnRNP K homology domain
mRNA	messenger RNA
NGF	nerve growth factor

NMD	nonsense mediated decay
NTRK	neuroptrophin receptor kinase
ORF	open reading frame
PAC	P1 artificial chromosome
PBS	phosphate bufferd saline
PCR	polymerase chain reaction
PEG	polyethylenglicol
perl	portable regular expression language
PLC	Phospholipase C
polII	RNA polymerase II
PTB	pyrimidine tract binding protein
RACE	rapid amplification od cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
RRM	RNA recognition motif
RT	reverse transcription
RT-PCR	reverse transcription followed by polymerase chain reaction
RTK	receptor tyrosine kinase
SAF	scaffold attachment factor (A or B)
SAM68	src associated in mithosis, 68kDa
SDS	sodium dodecyl phosphate
SF	splicing factor (1 or 2)
SLM	SAM68 like molecule
SMA	spinal muscle atrophy
SMN	Survival of motor neurons
snRNP	small nuclear ribonucleic particle
SRPK	SR-protein kinase
SR-protein	serin-Arginin-rich protein
SSC	sodium chloride, sodiumcitrate buffer
STAR	signal transduction and activation of RNA
SWAP	suppressor of whitener of apricot
TBE	Tris-Borate-EDTA buffer
TE	tris-EDTA
TEMED	N,N,N’,N’-Tetramethylethylenediamine
Tra2	transformer 2
Tris	Tris(hydroxymethyl)aminomethane
Trk	Tyrosine receptor kinase (A, B or C) also known as NTRK
U1-70	U1 snRNP 70kDa protein
U2AF	U2 snRNP auxiliary factor (35 or 65 kDa)
UTR	untranslated region

2. Zusammenfassung

Die Sequenzierung des menschlichen Genoms hat gezeigt, dass Menschen eine unerwartet geringe Anzahl von Genen besitzen. Alternatives Spleißen kristallisiert sich als einer der Hauptmechanismen heraus, mit dem Proteinvielfalt aus der begrenzten Anzahl von Genen hergestellt wird. In dieser Arbeit wurde ein Beitrag zum Verständnis des Mechanismus von alternativem Spleißen erbracht. Zunächst wurde die Regulation des alternativen Spleißens der prä-mRNA des htra2-beta Genes studiert. Es konnte gezeigt werden, dass htra2-beta1 das Spleißen seiner eigenen prä-mRNA in einer autoregulatorischen Schleife reguliert. Es wurden vier Bindungsstellen des TRA2-beta1 Proteins in Exon 2 der tra2-beta prä-mRNA identifiziert. Durch Mutagenese konnte eine in vivo Konsensus-Bindungsstelle bestimmt werden. Eine Durchforstung des menschlichen Genoms zeigte, das 12% aller beschriebenen Exons statistisch signifikante Übereinstimmungen mit diesem Konsensus besitzen. Die Bindung von TRA2-beta1 an prä-mRNA und die daraus folgende Prozessierung wird durch Protein Phosphorylierung reguliert.

Im weiteren wurde als erster Schritt zum Verständnis der TrkB Expression die Genstruktur und die alternativen Spleißmuster des menschlichen TrkB Genes bestimmt. TrkB ist ein Mitglied der Neurotrophin Rezeptor Tyrosin Kinase Familie und bindet bevorzugt an BDNF (brain derived neurotrophic factor). Es konnte gezeigt werden, dass TrkB durch alternatives Spleißen mehrere mRNAs bildet. Aus diesen mRNAs werden Proteine gebildet, die eine unterschiedliche Affinität zu BDNF aufweisen. Außderdem unterscheiden sich diese Proteine in der Fähigkeit, das BDNF Signal weiterzuleiten. In dieser Arbeit konnte gezeigt werden, dass das menschliche trkB Gen ungewöhnlich groß ist. Aus dem Gen können mehr als 100 verschiedene mRNAs durch alternatives Spleißen, sowie durch alternative Initiations- und Terminationsstellen entstehen. Durch die molekulare Analyse konnten wir eine neue stark exprimierte hirnspezifische mRNA (TrkB-T-Shc) identifizieren. Der hieraus entstehenden Protein Isoform fehlt die Tyrosin Kinase Domäne, dafür hat diese Form aber noch eine Shc Bindungsstelle.

3. Abstract

The rough draft of the human genome showed that humans have an unexpectedly small number of genes. Alternative splicing emerges as one of the major mechanisms involved in generating protein diversity from the relatively limited number of genes that appear to be contained in the genomes of the higher eukaryotes. The focus of this work is to understand the mechanisms of the alternative splicing regulation. First the regulation of the alternative splicing of the human tra2-beta pre-mRNA was studied. It shows that TRA2-beta1 protein regulates the splicing of its own pre-mRNA in a negative feedback autoregulatory loop. The study identified four binding sites for TRA2-beta1 in the tra2-beta exon 2. Through mutagenesis an *in vivo* TRA2-beta1 consensus binding site was derived. A search of the human genome shows that 12% of the annotated exons contain statistically significant matches to the consensus. The binding of TRA2-beta1 to RNA and respectively the splicing of its target genes is regulated through protein phosphorylation.

Second, as a first step in understanding the regulation of TrkB expression, this work characterized the structure and the alternative splicing of the human TrkB gene. TrkB is a member of the neurotrophin receptor tyrosine kinases family and a primary receptor for the brain derived neurotrophic factor (BDNF). It has been shown to generate through alternative splicing several mRNAs. These mRNAs are translated into proteins which differ in their affinities for BDNF and their ability to transmit the BDNF signal. This study showed that the human TrkB gene is unusually large. It can generate 100 forms through use of the alternative transcription initiation and termination sites and alternative splicing. The analysis of the splicing of the TrkB pre-mRNA led to the identification of a new major brain specific mRNA isoform (TrkB-T-Shc).This isoform lacks the tyrosine kinase domain, but retains a Shc binding site.

4. Introduction: Pre-mRNA Processing

Eukaryotic messenger RNA is subjected to a series of processing events. These events include capping of the 5' end, polyadenylation of the 3'end (Shatkin and Manley, 2000), splicing (Blencowe et al., 1994) and editing (Benne, 1996). They are crucial for the eukaryotic gene expression as they remove the intervening noncoding sequences from the pre-mRNA and stabilize it. In addition successful capping, polyadenylation and splicing facilitate mRNA export to the cytoplasm. Intact cap structure and poly(A) tail are also required for targeting of the mRNA to the ribosome and the initiation of the translation. The cap and the poly(A) tail are also subject to regulatory mechanisms that control the mRNA stability and as consequence the protein expression. Alternative polyadenylation and splicing as well as RNA editing are widely used to increase the protein diversity and to fine tune the properties of the proteins.

4.1. Capping and Polyadenylation

In all eukaryotic cells both 5' and 3' ends of the pre-mRNAs are modified to form the so called cap structure and poly(A) tails respectively (reviewed in Shatkin and Manley, 2000). The 5' triphosphate of the transcript is first shortened to diphosphate and then GMP is transferred generating an unusual triphosphate linkage: G(5')ppp(5')N. The guanosine base is subsequently methylated at N7 position. The cap can be further modified by methylating the 2' OH group of the ribose of the first and the second nucleotide. 3' end formation starts with cleavage of the RNA guided by an AAUAAA sequence 10 to 30 nucleotides upstream of the cleavage site and a less conserved GU or U rich stretches downstream of the cleavage site (Minvielle-Sebastia and Keller, 1999). Some pre-mRNAs require also U-rich motifs located upstream of the cleavage site. After the cleavage a tail of 200-250 adenines is added to the 3' end by the poly(A) polymerase. Both cap structure and poly(A) tail are essential for the mRNA transport to the cytoplasm, the initiation of the translation, and the protection of the mRNA from exonucleases. They therefore serve to identify the RNA as a messenger. In addition, cap formation is required for pre-mRNA splicing. The link between 3' end formation and splicing is less clear, but evidence suggests the two processes are linked and the presence of introns stimulates cleavage and polyadenylation.

4.2. Editing

Pre-mRNA can be chemically modified in a process designated as RNA editing. During editing either adenosine or cytosine residues are deaminated, which changes them into inosine and uracil residues, respectively. The editing of adenosines is catalyzed by adenosine deaminases (ADARs) (Reenan, 2001). There are three ADAR genes in humans and knock-out experiments show that

ADAR1 is essential in mice (Rueter et al., 1999). The frequency of editing of a particular nucleotide varies; despite the ubiquitous expression of ADAR1 and ADAR2, editing seems to be most prevalent in the brain, where one out of 17.000 nt is edited (Paul and Bass, 1998). Assuming an average length of 1 kb for an mRNA, this implies that ca. 1 in 20 brain transcripts is edited. The major determinant for ADAR editing seems to be perfectly matched RNA:RNA duplexes present on the pre-mRNA. Such duplexes are created either by regulatory antisense transcripts or by transcriptional read-through of adjacent genes in an antisense orientation. The editing of cytosines is catalysed by ApoBec-1 (Chester et al., 2000). In the case of ApoBec-1 editing the enzyme is directed to the site using auxiliary factor that recognizes an 11 nt sequence located 4-5 nt downstream of the edited cytosine (Chester et al., 2000). RNA editing can change the information content carried by a particular mRNA, both by changing the coding sequence or by creating new splicing sites. In addition adenosine deamination of viral RNAs is an essential part of the antiviral action of the interferons.

4.3. Splicing

4.3.1. Basic splicing reaction and spliceosome assembly

One of the most prominent features of eukaryotic genes is their discontinuity. With very few exceptions, notably the histone genes, the messages of almost all of the eukaryotic genes are disrupted by noncoding sequences termed introns. The analysis of the working draft of the human genome showed that on average introns account for 95% of the pre-mRNA (Lander et al., 2001).

There are three sequence elements (Figure 1) that mark the introns and are essential for their removal (Reed and Palandjian, 1997). The 5' splice site conforms to the consensus YRG/GURRGU (the slash denoting the exon – intron border). The two underlined nucleotides were believed until recently to be invariant. However both this study and research by Thanaraj and Clark (Thanaraj and Clark, 2001) showed that in 1% of all introns the first dinucleotide is GC instead of GU. The 3' splice site consensus is Y_{12}NYAG and is preceded by a polypyrimidine stretch. The third element is the branch point (YNCURAY) which is located 18 to 200 nucleotides upstream of the 3' splice site. An exception is the budding yeast *Saccharomyces cerevisiae*. In this organism a massive wave of retrotransposition events, due to the loss of the posttranscriptional silencing system, resulted in the elimination of most of the introns. This led to the subsequent disappearance of large number of spliceosome components that became dispensable (Aravind et al., 2000). As a result the splice sites of the few remaining introns strongly adhere to the consensus sequences, MAG/GURAGU and Y_{12}CAG/ for the 5' and 3' splice sites respectively (Staley and Guthrie, 1998). In addition the branch point is virtually invariant (UACUAAC) and is located 16 to 19 nucleotides upstream of the

3’ splice site. There is a second minor class of introns found in higher eucaryotes. They are named ATAC after the first and the last dinucleotides of the intron. The ATAC introns carry different splice site consensus sequences (5’ splice site: /AUAUCUU; 3’ splice site: CAC/; branch point: UCCUUAAC) and are processed by a separate splicing system (Levine and Durbin, 2001; Tarn and Steitz, 1996).

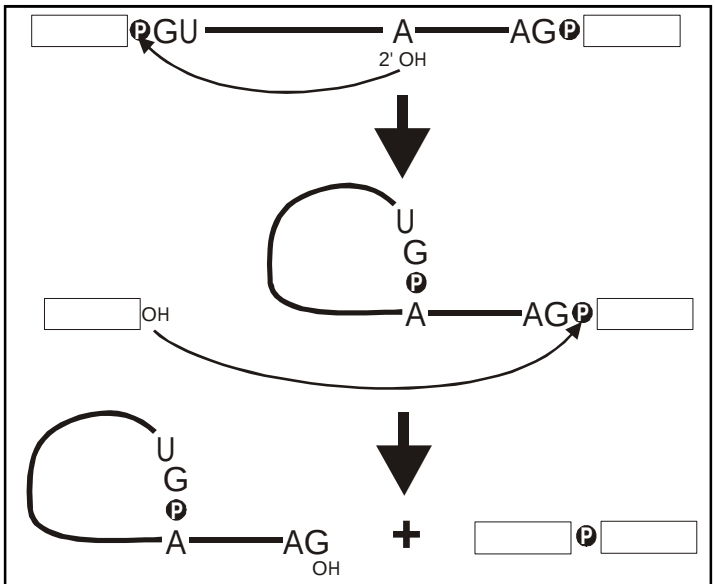


Figure 1 Basics of the splicing reaction

The intron excision proceeds through two transesterification reactions (Figure 1). Initially a nucleophilic attack of the 2’ hydroxyl group of the branch point adenosine at the 3’, 5’ phosphodiester bond of the 5’ splice site leads to its cleavage. As a result a 2’, 5’ phosphodiester bond forms between the branch point adenosine and the terminal guanosine. In the second step the newly formed 3’ hydroxyl group of the 5’ exon attacks the phosphodiester bond of the 3’ splice site, resulting in a 3’, 5’ phosphodiester bond between two exons and the release of the lariat intron.

The splicing reaction is catalyzed by a large ribonucleoprotein complex that assembles on the pre-mRNA, designated the spliceosome (reviewed in Reed and Palandjian, 1997). The fully assembled spliceosome sediments at 50-60S, suggesting a complexity comparable to that of the ribosome. In addition, the spliceosome is a highly dynamic structure. At different stages of its assembly it includes five small nuclear ribonucleoprotein particles (snRNPs), namely U1, U2, U4, U5 and U6 and multiple additional proteins (Will and Luhrmann, 2001). snRNPs assemble in an orderly fashion on the pre-mRNA and this assembly culminates in the formation of the catalytic spliceosome and the intron excision (Figure 2). The spliceosome assembly and the splicing reaction proceed through the formation of four distinct complexes. Each complex transition is ATP dependent and is accompanied by significant structural changes of the individual components. Initially the U1 snRNP binds to the 5’ splice site forming a RNA/RNA duplex with it. The polypirimidine tract and the AG dinucleotide at the 3’ splice site are recognized by the auxiliary proteins U2AF65 and U2AF35 respectively. The binding of the two U2AFs to the 3’ splicing site is facilitated by the U1 snRNP located at the 5’ splice site of the exon (Cunningham et al., 1995; Hoffman and Grabowski, 1992;

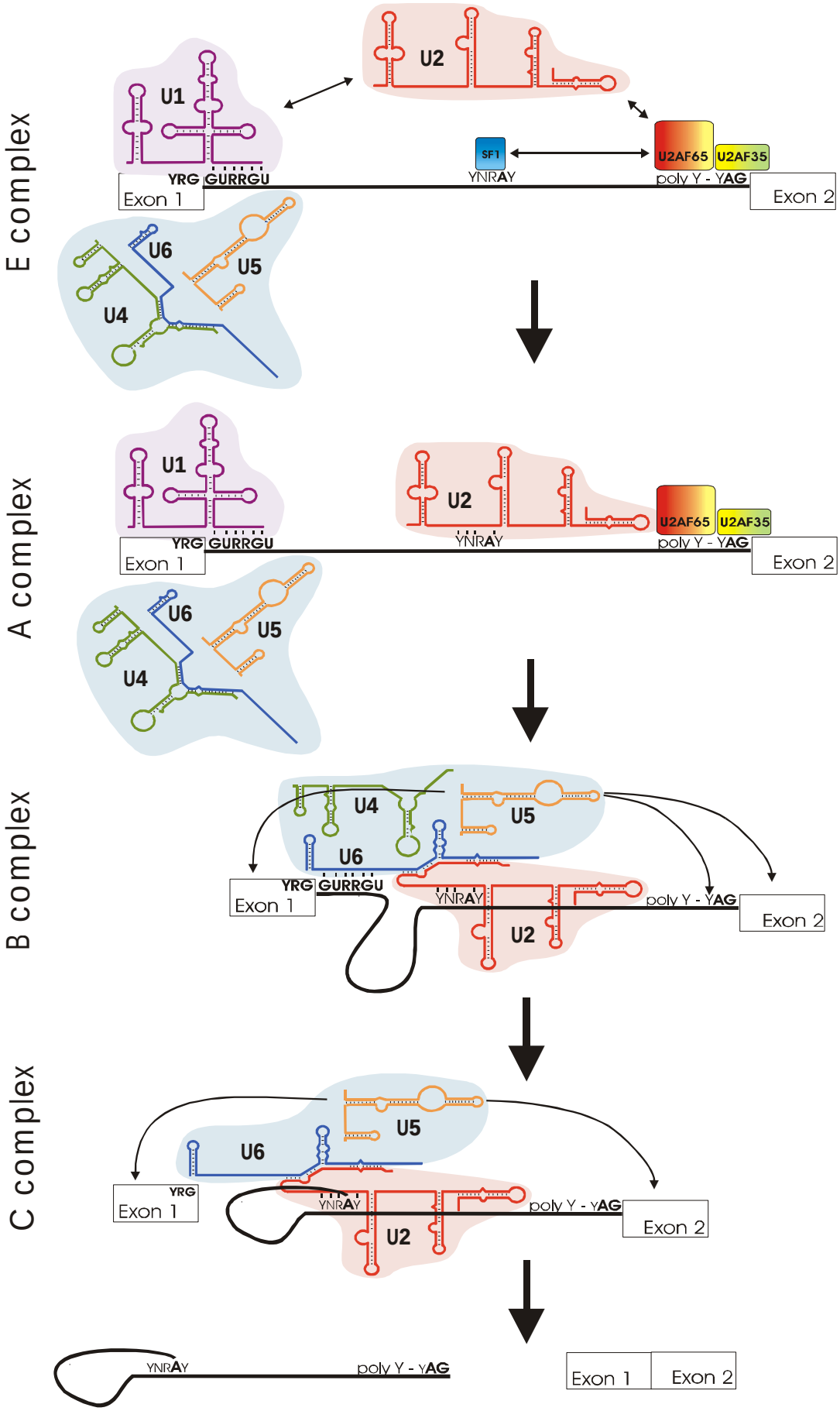


Figure 2. Spliceosome formation and rearrangement during the splicing reaction

Wu and Maniatis, 1993). There is also evidence that in the case of the 3' terminal exons the factors involved in the cleavage and polyadenylation of pre-mRNA can play a role in directing the splicing machinery to the 5' splice site (Niwa and Berget, 1991a; Niwa and Berget, 1991b; Niwa et al., 1990). The branch point is recognized by splicing factor1 (SF1). SF1 interacts with U2AF and this interaction is essential for its binding to the branch point. These interactions lead to the formation of the E-complex. This is the first complex specific to the intron containing RNAs and initially was described as the only ATP independent step in the spliceosome assembly. Later studies have shown that the U2 snRNP is also included in the E-complex through interactions with U1 snRNP, U2AF65 and possibly SF1 (Das et al., 2000). In addition U4/U6.U5 tri-snRNP can associate with the first exon near the 5' splice site in the E-complex. This association is ATP dependent (Maroney et al., 2000). The formation of the next complex (A-complex) involves RNA duplex formation between the U2 snRNP and the branch site. The next step of the spliceosome assembly involves major rearrangements of the snRNP components associated with the pre-mRNA. The U6/U4 duplex is disrupted and a new duplex between U6 and the 5' splice site is formed, displacing the U1 snRNP. In addition 5' splice site is brought close to the branch point and the 3' splice site through U6/U2 snRNA base pairing and interaction of U5 snRNP with both exons near the splice sites (Chiara et al., 1997; Collins and Guthrie, 2001). At this point, the U4snRNP leaves the complex and the first catalytic step of the splicing occurs, creating the intron lariat. In the new complex, termed the C complex, the two exons are kept in close proximity by the U5 snRNP. The second final step of the splicing reaction takes place in the C complex. After this step, the snRNPs leave the complex and are recycled for new splicing reactions. The exact mechanism of the catalysis is still poorly understood. However there are several lines of evidence that suggest that the reaction is catalyzed by RNA and is assisted by Prp8 (Collins and Guthrie, 1999), a protein component of the U6 snRNP (reviewed in Collins and Guthrie, 2000). The conservation of the RNA secondary structure of the snRNAs and more remarkably, the absolute conservation of the nucleotides in the proposed catalytic center offers strong evidence for this hypothesis. In addition a Mg^{2+} ion that is believed to be involved in the catalysis, is coordinated by highly conserved nucleotides in the U6 snRNA. Finally, there is a striking similarity between the splicing of the pre-mRNA and the excision of the group II self splicing introns (Sharp, 1991). Both reactions have the same stereochemistry of the phosphoryltransfer steps, and the nucleophile for the first step is the 2' hydroxyl group of the conserved branch point adenosine. The regions involved in catalysis have similar secondary structures and conserved sequences in bothe systems. This similarity implicates not only common splicing mechanism, but also suggests a common evolutionary origin.

Until recently recognition of the 3'splice site was poorly understood. Although U2AF65 and U2AF35 bind to the 3' splice site in the E-complex, they are released from the substrate relatively early in the spliceosome formation. Furthermore, U2AF35 has been shown to be dispensable for the splicing of many introns with 3' splice sites closely adhering to the consensus (Zhu and Krainer, 2000). A model involving scanning mechanism was initially proposed and there is some evidence for its existence (Chen et al., 2000). However, the results of these experiments seem to be compromised by the model system that was used. It used substrates where the 3' splice sites contained the AG dinucleotide but not the polypirimidine tract. It has been shown before that distal 3' splice sites can compete with the proximal site provided they have a polypyrimidine tract. Therefore, the resulting proximal splice site selection could be due to the lack of the functional distal 3' splice site. A more plausible model explaining the selection of proximal 3' splice site suggests that there is a molecular ruler selecting for splicing sites positioned at an optimal distance from the branch point (Chiara et al., 1997; Chua and Reed, 2001). This molecular ruler appears to be the U5 snRNP itself. Chiara and coworkers have shown that it interacts directly with the polypirimidine tract of the 3' splice site.

4.3.2. Exon recognition and intron bridging

The precise removal of the intron sequences is crucial for the gene expression. The initial finding that introns are marked by short loosely conserved sequences located near their borders was a surprise. The sequences of the splice sites and the branch point are clearly insufficient for the intron recognition and removal. Nevertheless the splicing machinery is able to accurately and efficiently locate and splice relatively small (100 nt or less) exons separated by introns that can reach lengths of 1Mbp or more. Therefore other factors than the core spliceosomal components should be involved in the early steps of splicing.

The first such factor to be isolated was SF2 (Krainer et al., 1991; Krainer et al., 1990; Krainer and Maniatis, 1985). It was biochemically purified based on its ability to complement cytoplasmic S100 extracts in an *in vitro* splicing reaction. S100 extracts are inactive in splicing although they contain all of the core spliceosomal components. SF2 has been shown to be a member of the large family of the SR proteins that share both structural and functional similarity (Valcárcel and Green, 1996). The SR protein family members are defined by their biochemical properties and the ability to be recognized by the mAB104 antibody (Mayeda et al., 1992; Roth et al., 1991; Zahler, 1999). The name of the family comes after the epitope that the mAB104 antibody recognizes, a phosphorylated stretch of alternating Serine (S) and Arginine (R) residues. This SR-repeat is currently recognized as a general feature characteristic for many of the proteins involved in the pre-mRNA

processing, including hnRNP and SR-like protein families and some of the snRNP proteins. The motif has been shown to be involved in protein-protein interaction. The second common structural feature among the SR proteins is the presence of one or two N-terminal RNA recognition motifs (RRMs).

All of the SR protein family members and some of the SR-like protein family members have been shown to activate splicing *in vitro*. They facilitate the formation of the E, A and B complexes by a set of protein/protein and protein/RNA interactions (reviewed in Kramer, 1996). Several distinct types of such interactions have been described. SR-proteins can bind in a sequence specific manner to cis-acting sequences located in the exons (Lui et al., 1999; Lui et al., 1998). They participate in the recruitment of U2AF65 and U2AF35 to the 3' splice site (Wang et al., 1995; Wu and Maniatis, 1993), as well as the recruitment of U1 snRNP to the 5' splice site (Eperon et al., 1993; Jamison et al., 1995; Kohtz et al., 1994; Staknis and Reed, 1994; Wu and Maniatis, 1993; Zahler and Roth, 1995) and of U2 snRNP to the branch point (Tarn and Steitz, 1995) by interactions mediated by the SR repeat. In addition SR proteins can bridge exons by interacting with both the U1-70K component of the U1 snRNP and the U2AF65-U2AF35 heterodimer (Wu and Maniatis, 1993). Exon bridging by SR proteins is also assisted by the SRm160/300 splicing coactivator complex. This complex interacts directly with multiple SR proteins and the U2 snRNP and indirectly with the U1 snRNP (Blencowe, 2000; Blencowe et al., 2000; Blencowe et al., 1998; Eldridge et al., 1999; Longman et al., 2001; McCracken et al., 2002).

SR and SR-like proteins can also bridge introns by interacting with themselves and the core spliceosome components (Fu and Maniatis, 1992; Stark et al., 1998). This activity is independent of their RNA binding ability (Hertel and Maniatis, 1999). Finally, they can displace from the pre-mRNA factors that inhibit splicing for example hnRNP family members by competing for binding sites on the target RNA (Eperon et al., 2000; Hanamura et al., 1998; Zhu et al., 2001). This activity can be demonstrated *in vitro* on certain substrates that have relatively strong 3' splice sites and it does not depend on the SR repeat (Zhu and Krainer, 2000).

Another group of proteins that plays an important role in splice site determination are the hnRNP proteins (reviewed in McAfee et al., 1997 and Krecic and Swanson, 1999). They were first described as packaging factors. Members of the hnRNP A, B and C families associate with RNA to form a regular array of 20-25 nm particles. The binding of the hnRNP proteins to RNA shows some sequence preference which could serve to 'phase' the packaging of the pre-mRNA. Thus it can influence the splice site selection, by blocking the access to certain splice sites. hnRNP proteins on

general antagonize SR and SR-like proteins function to activate splicing. Exon recognition and splice site selection is a combinatorial process involving both positive and negative signals provided by SR and SR like proteins and hnRNP proteins respectively.

Studies on the cis elements involved in splice site selection or activation of weak splice sites confirmed that splicing occurs on a general background of repression (Fairbrother and Chasin, 2000; Sun and Chasin, 2000) and can be stimulated by enhancer sequences. Several studies using the SELEX approach showed that the exonic splicing enhancer (ESE) sequences recognized by the SR and SR-like proteins are short (5 to 8nt) and highly degenerate (Amarasinghe et al., 2001; Cavaloc et al., 1999; Cooper, 1999; Coulter et al., 1997; Lui et al., 1999; Lui et al., 1998; Tacke and Manley, 1995; Wang et al., 1997). This degeneracy prevents them from interfering with the coding capacity of the exons. As in the case of the signals recognized by the core spliceosome components, however, they carry very low information content. To overcome this limited information content, ESEs act cooperatively (Hertel and Maniatis, 1998; Li et al., 2000; Lynch and Maniatis, 1995; Lynch and Maniatis, 1996; Nagel et al., 1998; Smith and Valcarcel, 2000). As SR and SR-like proteins attach to multiple nearby sites, their binding is stabilized by protein/protein interactions, forming a complex that marks the exon.

Due to degeneracy of their binding sites, different SR and SR-like proteins often act through the same ESEs and activate the splicing of the same exons. This raises the question of their functional redundancy. Double stranded RNA interference studies in *Caenorabditis elegans* have shown that some of the SR proteins are indeed redundant for survival (Kawano et al., 2000; Longman et al., 2000). However their function is still important as disruption of any two of the SR protein genes leads to the appearance of aberrant phenotypes. These results further support the notion that SR and SR-like proteins act cooperatively to activate exon splicing. When only one of the SR proteins is disrupted, the other proteins can substitute for it leading to mild or absent phenotype aberrations. Studies in other organisms have not been as extensive as those in *C. elegans*. Knockout experiments in mice, however, show that SR proteins might be much more important for survival in mammals that in *C. elegans*, as SRp20 knockout leads to early embryonic lethality (Jumaa et al., 1999) and conditional knockout of SC35 in T cells blocks their development (Wang et al., 2001).

The importance of the splicing enhancer and silencer sequences and the factors that bind to them in exon recognition is further emphasized by the increasing number of pathologies linked to them. Up to now, fifteen diseases have been linked to mutations that disrupt sequences involved in exon recognition and alternative splicing regulation (Table 1). Detailed studies have been performed

Table 1. Mutations in exon enhancers that cause diseases

Gene / Disorder	**Mutation**	**Effect**	**Reference**
SMN2 Spinal Muscle Atrophy (SMA)	Silent C>T conversion in exon 7	Disrupts ESE, skipping of exon 7	(Coovert et al., 1997; Jablonka et al., 2000; Lefebvre et al., 1995; Lefebvre et al., 1997; Lorson and Androphy, 2000; Lorson et al., 1999; Monani et al., 2000; Vitali et al., 1999)
SMN1 Spinal Muscle Atrophy (SMA)	425del5 W102X	Skipping of exon 3	(Sossi et al., 2001)
beta-hexaminidase beta-subunit Sandhof disease	exon 11 P417L C>T conversion at nucleotide 8	Disrupts an ESE. Causes use of a cryptic splice	(Fujimaru et al., 1998)
	intron 10 A>G conversion at position –17.	Causes usage of a critic splice site at position –37. Disrupts ISE. The conversion also disrupts a putative branch point	
Tau Frontotemporal Dementia with Parkinsonism linked to chromosome 17	Intron 10: +13 A>G, +14C>T, +16C>T, IVS10+3 G>A	Disrupt ISE IVS10+3 G>A improves slightly the splice site	(D'Souza et al., 1999; D'Souza and Schellenberg, 2000; Hasegawa et al., 1999; Hutton et al., 1998; Iijima et al., 1999; Spillantini et al., 1998; Stanford et al., 2000)
	Exon10: L284L T>C, S305S T>C, S305N G>A	Disrupt ESE S305N G>A improves the splice site	
	Exon10: N297K T>G, del280K (AAG deletion)	Disrupts ESE	
Porphobilinogen deaminase Acute intermittent porphyria	R28R, C>G	Skipping of exon 3	(Llewellyn et al., 1996)
Integrin GPIIIa Glanzmann thrombasthenia	C>A at position +16 and silent G>A at +134 of exon 9	Skipping of exon 9	(Jin et al., 1996)
Fumarylacetoacetate hydrolase Hereditary tyrosinemia type 1	N232N C>T	Skipping of exon 8	(Ploos van Amstel et al., 1996)
Pyruvate dehydrogenase E1 alpha Leigh's encephalomyelopathy	silent A>G	Aberrant splicing of exon 6	(De Meirleir et al., 1994)
MNK Menkes disease	Gly>Arg G>A	Skipping of exon 8	(Das et al., 1994)
Adenosine deaminase severe combined immunodeficiency disease	R142X G>A and C>T in the same codon	Skipping of exon 5	(Santisteban et al., 1995)
Arylsulfatase A metachromatic leukodystrophy	Thr409Ile C>T	Activates a cryptic splice site	(Hasegawa et al., 1994)
Fibrillin-1 Marfan syndrome	silent C>T in exon 51	Skipping of exon 51	(Liu et al., 1997)
CYP 27 cerebrotendinous xanthomatosis	silent G112G G>T in exon 2	Creates a cryptic splice site. In addition it causes skipping of the entire exon 2	(Chen et al., 1998)

Table 1. Mutations in exon enhancers that cause diseases (continued)

Gene / Disorder	**Mutation**	**Effect**	**Reference**
CD45 Individuals do not suffer from obvious immunodeficiency, but it is notable that no homozygotes have been described. Associated with Multiple sclerosis in the American population.	silent C>G at position 77 of the alternative exon 4	Disrupts an ESE. Constitutive inclusion of exon 4	(Jacobsen et al., 2000; Lynch and Weiss, 2001; Zilch et al., 1998)
beta-globin beta-Thalassemia	T->G transversion in position 705 of intron 2	Activates cryptic 3' splice site in intron 2	(Dobkin and Bank, 1983; Dobkin et al., 1983)
BRCA1 Breast and ovarian cancer	E1694X G>T	Disrupts ESE and causes skipping of exon 18	(Liu et al., 2001)
NF-1 Neurofibromatosis type 1	Y2264X C>A, C>G	Skipping of exon 37	(Ars et al., 2000a; Ars et al., 2000b; Hoffmeyer et al., 1998; Messiaen et al., 1997)
	R304X C>T	Skipping of exon 7	
	Q756X C>T	Skipping of exon 14	
DMPK Myotonic dystrophy	Expanded $(CUG)_{>40}$ in the 3´ UTR of DMPK	Two possibilities: CUG repeats sequestrate CUG-BP and therefore prevent it from its normal job in splicing certain genes or expanded CUG repeats alter the alternative splicing of the DMPK-mRNA by developing a new 3´splice site.	(Philips et al., 1998; Tiscornia and Mahadevan, 2000a; Tiscornia and Mahadevan, 2000b)

on the molecular mechanisms that cause the pathologies of spinal muscle atrophy (SMA) and frontotemporal dementia, with parkinsonism linked to chromosome 17 (FTDP-17). In both cases, the efficiency of exon recognition is affected by either disrupting or creating enhancer elements. In SMA an exonic splicing enhancer in SMN2 gene exon 7 is disrupted causing skipping of the exon (Table 1). In tau microfilament associated protein gene exon 10 and intron 10, multiple mutation were shown to improve splicing enhancers or disrupt splicing silencers. As a result the regulation of the alternative splicing of exon 10 is affected, leading to abnormally high levels of transcripts containing the exon causing the FTPD-17 (Table 1). Interestingly two of the missense mutations in tau exon 10 (N297K and S305S) the aminoacid residue replacement itself does not alter the protein properties. Instead the pathology is caused by increased exon 10 inclusion due to creation of strong exonic splicing enhancers (Hasegawa et al., 1999).

4.3.3. Fidelity of the splicing reaction

Although there is no direct measure for the fidelity of the splicing, it appears highly accurate. This accuracy is probably achieved by the presence of multiple energy-dependent steps in the spliceosome assembly. It should be explicitly noted that the two transesterification reactions do not require energy input and can occur spontaneously, as is the case of the group II self splicing introns.

Therefore the energy dependent steps in the spliceosome assembly are probably used to direct the reaction and to ensure the accurate splice site recognition. In addition to the energy dependence, individual spliceosome assembly steps depend also on the successful completion of the preceding steps.

A second checkpoint that controls the fidelity of the splicing reaction takes place during export of the RNA to the cytoplasm (reviewed in Nakielny and Dreyfuss, 1999). Factors that block mRNA export such as nonshuttling hnRNP proteins and components of the spliceosome have to be removed from the mRNA. This removal in most cases is coupled to the successful intron excision. In addition during splicing a 335 kDa protein complex (EJC) is formed (reviewed in Kim and Dreyfus, 2001). It is centered 20-24 nucleotides upstream of the exon-exon junctions. EJC plays dual role in pre-mRNA splicing quality control. It facilitates mRNA export by providing strong binding sites for export factors (Le Hir et al., 2001; Zhou et al., 2000). EJC also prevents translation of mRNAs containing premature stop codons that can result from inaccurate splicing. This is achieved by triggering the nonsense mediated decay (NMD) system through interaction with eRF1 and eRF3 bound to the stop codon, whenever it is located less than 55 nucleotides upstream of the exon-exon junction (Ishigaki et al., 2001; Kim et al., 2001; Lykke-Andersen et al., 2001).

4.3.4. Coupling of pre-mRNA processing to transcription

Several lines of evidence show that pre-mRNA processing and transcription are both physically and functionally linked. RNA polymerase II CTD is required for both 3' and 5' end formation (Hirose and Manley, 1998; McCracken et al., 1997a). It interacts directly with capping (Bentley, 1999; Steinmetz, 1997) and cleavage factors (McCracken et al., 1997b).

RNA polymerase II is also required for pre-mRNA splicing as intron containing transcripts generated by RNA polymerases I and III are not spliced *in vivo* (Sisodia et al., 1987; Smale and Tjian, 1985). *In vitro* splicing was accelerated and the spliced products associated with the template when transcription was carried out by RNA polymerase II compared to T7 RNA polymerase (Ghosh and Garcia-Blanco, 2000). Furthermore the carboxy terminal domain CTD of RNA polymerase II stimulates splicing both *in vitro* (Zeng and Berget, 2000) and *in vivo* (Fong and Bentley, 2001) and is required for splicing factor localization to the sites of transcription (Misteli and Spector, 1999). CTD has been shown to interact directly with SR and SR like proteins implicated in splice site selection (Kim et al., 1997; Yuryev et al., 1996). In yeast it interacts also with Prp40p (Morris and Greenleaf, 2000), a U1 snRNP-associated protein (Kao and Siliciano, 1996) acting in intron bridging (Abovich and Rosbash, 1997). Based on this evidence it has been suggested that the 5' splice site of

the newly transcribed exon is tethered to the CTD of the elongating RNA polymerase II (Morris and Greenleaf, 2000). When new exons are synthesized, they will emerge and be held in close proximity to the preceding exon, thus facilitating exon recognition and bridging over large introns.

Another link between transcription and splicing is provided by the discovery that transcriptional activator and coactivator proteins functionally interact with the splicing machinery and can stimulate splicing *in vitro*. The first such factor to be described is the transcriptional coactivator p52 (Ge et al., 1998). It interacts with SF2/ASF and colocalizes with it in the nuclear speckles. In *in vitro* splicing reactions using S100 extracts complemented with SF2/ASF, p52 enhances the SF2/ASF ability to activate splicing of the substrate pre-mRNA. The recent discovery that members of the SOX protein family (Pevny and Lovell-Badge, 1997; Wegner, 1999) are required for splicing is another illustration of the interplay between transcription and splicing. Ohe and coworkers show that SOX6, SOX9 and SRY proteins are necessary for the splicing of several substrates and their depletion prevents the formation of the C and B complexes (Ohe et al., 2002).

Current models of gene expression assume that chromatin is organized in topologically separated loops corresponding to separate functional units (Bode et al., 2000; Nakayama and Takami, 2001). In these units, transcription and RNA processing factors are organized in large complexes termed RNA factories (Jackson and Cook, 1995; Jackson et al., 1998). Their DNA is attached to the nuclear skeleton through scaffold or matrix attachment regions (S/MAR regions) which contain AT-rich sequences of high unwinding propensity (Benham et al., 1997). The studies of scaffold attachment factors A (SAF-A) and B (SAF-B) provide insight o how transcriptional factories are organized. SAF A and B bind to the S/MARs through a conservative domain (Kipp et al., 2000) and connect them to the nuclear scaffold. They have also been shown to interact with RNA through their RNA recognition motifs (RRM) and to be part of the hnRNP complexes (Kiledjian and Dreyfuss, 1992; Weighardt et al., 1999). SAF-B binds to RNA polymerase II CTD and factors involved in pre-mRNA processing (Arao et al., 2000; Nayler et al., 1998b; Nikolakaki et al., 2001; Weighardt et al., 1999). Data from these studies suggest that through interactions with pre-mRNA, DNA and multiple proteins, scaffold attachment factors link physically transcription, pre-mRNA processing and chromatin.

4.3.5. Alternative splicing

The recently released draft of the human genome sequence revealed a surprisingly low number of genes, about 35,000, with more recent unpublished estimates of under 25 000 genes (ENSEMBL release 4.28.1). To generate the estimated proteome of at least 250 000 proteins, diverse

posttranscriptional mechanisms are used. One major mechanism is alternative pre-mRNA splicing (Lander et al., 2001; Rubin, 2001). For example, it has been estimated that 59% of all human genes on chromosome 22 are alternatively spliced (Hide et al., 2001). Alternative splicing can generate an astonishing diversity. For example the *Drosophila* Dscam gene can generate more than 38,000 different mRNAs by alternatively splicing exons at four positions (Celotto and Graveley, 2001). There are no limits to the alternative splicing patterns. Virtually all possible splice sites combinations are used in the alternative splicing of pre-mRNA. Figure 3 summarizes the major modes of alternative splice site usage.

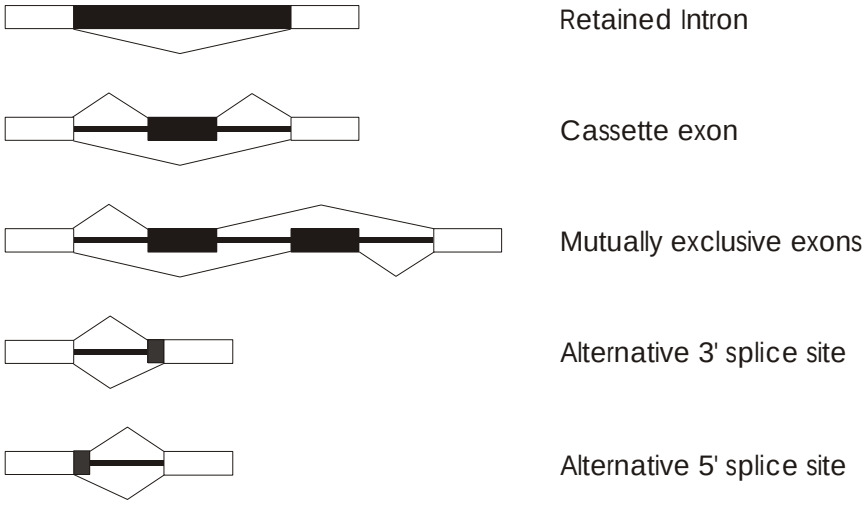


Figure 3. Types of alternative exons

Many aspects of the protein function can be regulated by alternative splicing. These include ligand affinity and signaling capabilities of receptors, intracellular localization of proteins, ion channel properties and DNA binding affinity and activity of the transcription factors, to name just a few. A substantial part of the alternatively spliced exons show tissue or cell type specific patterns of expression (Stamm et al., 1994; Stamm et al., 2000) and/or are regulated during development or in response to external stimuli. For example, insulin administration influences the incorporation of the alternative exon 11 of the insulin receptor (Sell et al., 1994) and activates exon ßII inclusion in the PKC gene (Chalfant et al., 1995; Chalfant et al., 1998; Patel et al., 2001). Serum deprivation alters usage of the serine/arginine-rich protein 20 (SRp20) exon 4 (Jumaa et al., 1997). Neuronal activity changes the alternative splicing pattern of clathrin light chain B, the NMDAR1 receptor, and c-fos (Daoud et al., 1999). In the brain, stress changes splicing patterns of potassium channels (Xie and Black, 2001; Xie and McCobb, 1998) and of acetylcholin esterase (Kaufer et al., 1998; Meshorer et al., 2002). ConA has been shown to change splicing patterns of the class 1b major histocompatibility complex molecule Qa-2 (Tabaczewski et al., 1994) and the splicing patterns of tumor necrosis factor ß are regulated by src kinases (Gondran and Dautry, 1999; Neel et al., 1995). Finally, programmed cell death is concomitant with a change in the alternative splicing patterns of several cell death regulatory proteins (reviewed in Jiang and Wu, 1999).

4.3.6. Regulation of the alternative splicing.

The tissue and cell type specific alternative splicing is regulated in two major ways: through tissue specific splicing factors or through regulation of the relative levels of the general splicing factors.

Few tissue specific splicing factors have been described (Table 2) and most of them are expressed in more than one tissue. Not much data is available on the mechanism of their action. However, a common theme appears from the studies on nPTB, SLM1 and SLM2. The three proteins nucleate the formation of tissue specific complexes of general splicing factors (Markovtsov et al., 2000; Stoss et al., 2001), which in turn activate or block exon inclusion. All of the tissue specific splicing factors are members of larger protein families. These families can also include splicing factors with ubiquitous expression (PTB, hnRNP G and the member of STAR family SAM68). Although PTB, hnRNP G and SAM68 are very similar to their tissue specific counterparts both in regard to the protein structure and RNA binding properties, they interact with different sets of proteins and therefore form different complexes on the same pre-mRNA sequences. As a result, different sets of exons will be spliced. This also appears to be true for the members of the CELF family. Although their protein interaction have not been studied in detail, they have the same RNA binding properties but differ in their ability to activate splicing of certain exons (Ladd et al., 2001; Lu et al., 1999).

The ubiquitously expressed general splicing factors also participate in regulation of the tissue specific splicing. Although they are expressed in virtually every tissue, a study by Hanamura and coworkers (Hanamura et al., 1998) has shown that the expression levels of several SR-proteins and hnRNP A are not the same for the different tissues. As discussed above exon recognition by the splicing machinery is based on cooperative assembly of SR and SR-like protein complexes on pre-mRNA that is antagonized by the hnRNP proteins. Deviations in the concentration of these proteins will reflect on the efficiency of the exon recognition. Alternative exons rely on SR and SR-like proteins for their splicing even more than constitutive exons as their splice sites deviate more from the consensus (Stamm et al., 1994; Stamm et al., 2000). Consequently the differences in the concentrations of the splicing factors will strongly influence their splicing.

Apart from the expression levels of the splicing factors, their posttranslational modifications also play a significant role in the alternative splicing regulation.

Protein phosphorylation has been shown to modulate the alternative splicing of a number of exons both *in vivo* (Du et al., 1998; Gondran and Dautry, 1999; Hartmann et al., 2001; Sanford and Bruzik, 1999) and *in vitro* (Prasad et al., 1999). It affects the ability of several splicing factors to

Table 2 Tissue specific splicing factors

Protein Family	Factor	Expression pattern	Notes	Reference
NOVA	Nova-1	CNS neurons		(Buckanovich et al., 1993)
	Nova-2	CNS neurons		
T-STAR	SLM1	Brain		(Di Fruscio et al., 1999; Stoss et al., 2001; Venables et al., 1999)
	SLM2	Brain, Heart, Testis	Detectable also in several other tissues. Testes expression is confined to the sertoly cells.	
PTB	nPTB	Brain, Testis	Detectable also in liver, lung, heart and thymus	(Polydorides et al., 2000)
hnRNP G	RBM1	Testis	Expression starts 4 days after birth	(Elliott et al., 1996)
CELF	ETR-3	Muscle, Brain		(Ladd et al., 2001)
	CUG-BP	Hearth, Diaphragm, Uterus, Spleen, Lung, Mammary Gland		
	CELF-3	Brain		
	CELF-4	Muscle, Kidney, Mammary Gland		
	CELF-5	Brain		

interact with RNA (Chen et al., 2001; Derry et al., 2000) and with other proteins (Hartmann et al., 1999; Wang et al., 1998a). Protein phosphorylation controls the release of the SR-proteins from the storage compartments in the nucleus (Hartmann et al., 1999; Wang et al., 1998a) and causes relocalization of hnRNP proteins to the cytoplasm (van der Houven van Oordt et al., 2000). In all cases the active concentration of the splicing factors will be altered and as result the alternative splicing pattern will change. Several protein kinases (SRPK1 and 2, Clk1, 2, 3 and 4) have been shown to specifically phosphorylate the SR repeats of splicing factors (Colwill et al., 1996a; Colwill et al., 1996b; Duncan et al., 1998; Gui et al., 1994; Nayler et al., 1997; Nikolakaki et al., 2002; Prasad et al., 1999). They are ubiquitously expressed, like the general splicing factors. The expression levels of the SRPK1, SRPK2 and CLK/STY, however, vary significantly between the different tissues (Wang et al., 1998b). It is unclear whether the tissue specific expression of these kinases can account for the tissue specific splicing events: they have partially overlapping substrate preferences and no direct evidence supports such hypothesis. Nevertheless a number of recent studies implicate well known signaling pathways in both direct or indirect regulation of alternative splicing (Daoud et al., 1999; Lynch and Weiss, 2000; van der Houven van Oordt et al., 2000; Weg-Remers et al., 2001).

The link between the signaling pathways and the regulation of alternative splicing is poorly understood. To date, only two cis acting sequences that confer response to signaling pathways have been identified (Konig et al., 1998; Xie and Black, 2001) and hnRNPA1, the STAR protein family members and YT521B have been identified as targets for the signal transduction pathways (Chen et al., 2001; Hartmann et al., 1999; Matter et al., 2000). STAR proteins and YT521B interact with and are tyrosine phosphorylated by Src tyrosine kinase family members Fyn, Sik and Rlk as well as Abl (Chen et al., 2001; Hartmann et al., 1999). The significance of splicing regulation by signal transduction and the mechanisms by which it is achieved still remain to be investigated.

4.4. Research overview

In this work two aspects of alternative pre-mRNA splicing regulation have been investigated. The first is the regulation of the alternative splicing of the human tra2-beta pre-mRNA. This study shows that the TRA2-beta protein regulates the splicing of its own pre-mRNA by binding to exon 2 and stimulating its inclusion in the mature pre-mRNA. In addition, the binding sites of TRA2-beta in exon 2 are identified and a consensus binding site is derived using sited directed mutagenesis. The work shows that the alternative splicing of the *tra2-beta* gene is regulated by protein phosphorylation. It can therefore be controlled by signal transduction pathways. Finaly the proteins that are synthesized from the tra2-beta locus are identified and their activity as alternative splicing regulators is examined.

The second is the regulation of the alternative splicing of the human TrkB pre-mRNA. This study is initial characterization of the gene structure of the TrkB gene and the alternative splicing of its transcripts. The aim of this study was to provide information for the construction of model minigenes to be used to investigate the alternative splicing regulation of the TrkB pre-mRNAs. The work showes that the human TrkB gene is extraordinarily large. It also identifies a new neuron specific TrkB isoform generated by alternative splicing. Two new alternative exons in the coding part of the mRNA and several new alternative exons in the 5' UTR are also discovered.

5. Materials and methods.

5.1. Contig assembly, general sequence manipulation, sequence alignment

Large scale contig assembly was done using either GCG package version 10 (Womble, 2000) running on HP/UX at the MPI for Biochemistry in Munich or by the Staden Package (http://www.mrc-lmb.cam.ac.uk/pubseq/) running on Windows NT. General sequence manipulation, editing and alignment, as well as Blast searches to small customized databases and small scale contig assembly, was performed using the BioEdit (Hall, 1999) program with bundled Clustalw (Thompson et al., 1994), CAP (Huang, 1992) and Blast3 (Altschul et al., 1997), running on Windows NT.

5.2. Pattern matching for splicing factor consensus binding sites.

The working draft of the human genome was downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/. Scripts written in Perl were used to extract the exons and introns from the genomic sequences according to the annotation data and to perform pattern searches of the datasets. The scripts were run on Slackware linux distribution (http://www.slackware.com). The probability of the matches in a given exon or intron to occur by chance was determined using the formula: $P ? e^{?\frac{Lm}{l}}$, where *l* is the length of the exon or intron, *m* the number of matches and *L* the length of the sequence with the same base composition as the inspected exon or intron, where one hit is expected to occur by chance.

5.3. Perl scripts:

5.3.1. gbk_parse.pl

This script strips the GeneBank files of the sequence data and most of the headers, keeping only the conting name and the information about the transcripts that is recorded in the “transcript” field (Figure 4). For each transcript a separate record is generated, which is idetified by ID, LOCUS, (derived from the transcript name) and CONTIG (name of the contig). This is followed by the information in the mRNA about the position of the exons.

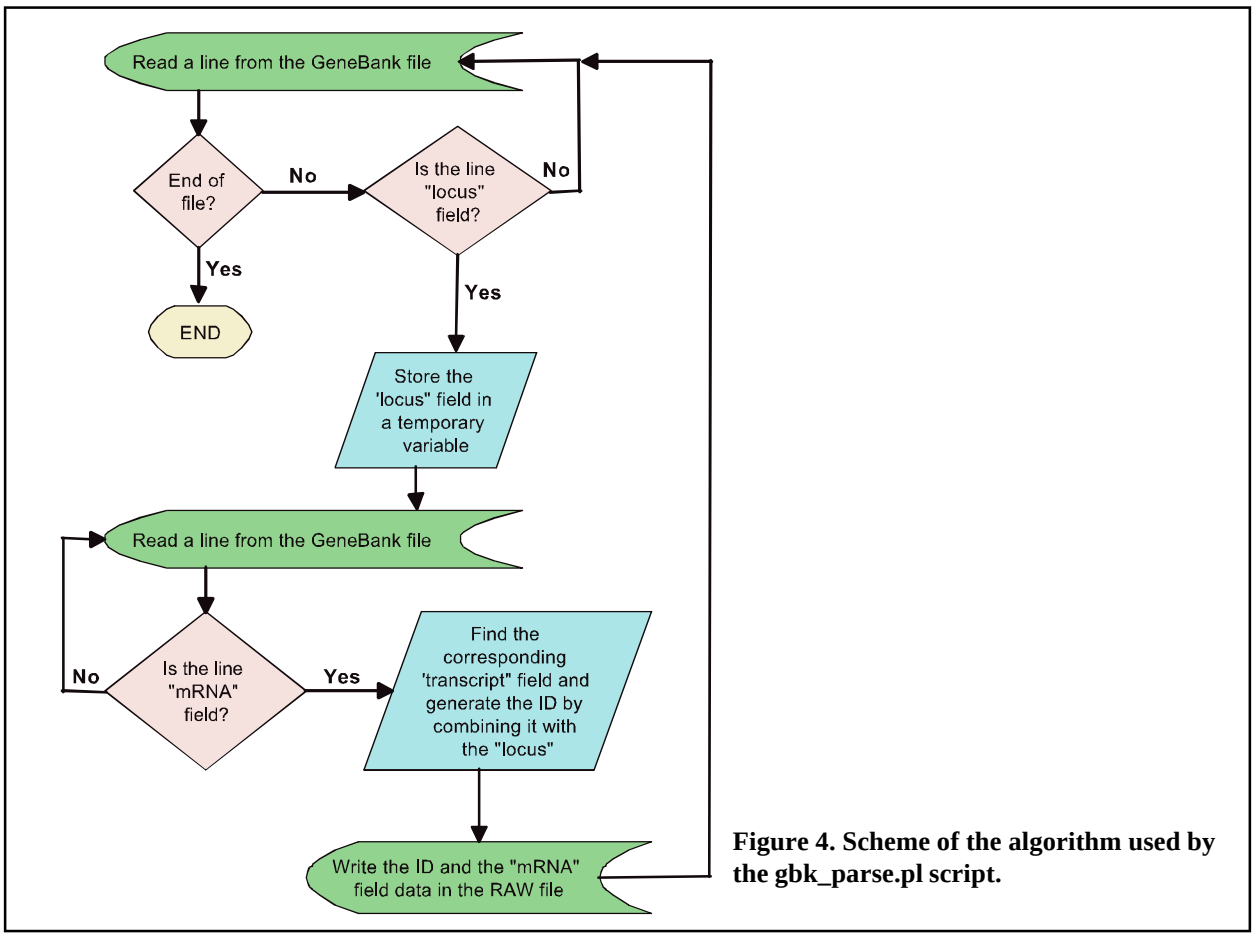


Figure 4. Scheme of the algorithm used by the gbk_parse.pl script.

```
#!/usr/bin/perl
# gbk_parse.pl
# Peter Stoilov
# This script strips the GeneBank files of the sequence data
# and most of the headers, keeping only the conting name and
# the information about the transcripts that is recorded in
# the “transcript” field. For each transcript a separate record
# is generated, which is idetified by ID, LOCUS, (derived from the
# transcript name) and CONTIG (name of the contig). This is followed
# by the information in the mRNA about the position of the exons


($#ARGV != -1) || die “Usage: first_parse.pl prefix\n”;
$prefix = $ARGV[0];

open(OUT, “>${prefix}.RAW”) || die “Could not create ${prefix}.RAW\n”;
open(GBK, “${prefix}.gbk”) || die “Could not open ${prefix}.gbk\n”;
$counter = 0;
print OUT “\n\n”;
while ($line = <GBK>){
                if ($line =~ /^LOCUS/) {
                  ($contig) = ($line =~ /^LOCUS\s+(\S+)/);
                }
LABEL2:         if ($line =~ /^\s+mRNA/ && $line =~ /join/){
                 ++$counter;
                 $entry=$line;
                 @locus=””;
                 $line=””;
                 until ($line =~ /^\s+mRNA/ || $line=~/^     CDS/){
                    $line = <GBK>;
LABEL1:      .      &dump_the_garbage;
        if ($line =~ /STS/){
                goto LABEL1;
        }
        elsif ($line =~ /variation/){
                goto LABEL1;
        }
                    if ($line =~ /^\s+\/transcript/){
                         @id = split /\”/, $line;      #”
                         push @locus, $id[1];
                    }
                    $entry .= $line;
                 }
                 $entry = “\$ID “.$counter.”\n”.”LOCUS       “.$locus[1].”\n”.”CONTIG      “.$contig.”\n”.$entry.”\n”;
                 print OUT $entry;
                 if ($line =~ /^\s+mRNA/){
                       goto LABEL2;
                 }
        }
}
```

```
print $counter."   Transcripts"."\n";

sub dump_the_garbage {
        if ($line =~ /STS/){
                &fast_forward;
        }
        elsif ($line =~ /variation/){
                &fast_forward;
        }
}
sub fast_forward {
          $line = <GBK>;
        until ($line !~ /^                    \//){
                $line = <GBK>;
        }
}
```

5.3.2. gbk_split.pl

This script reads the GenBank chromosome files and splits them in separate files that contain one contig each (Figure 5).

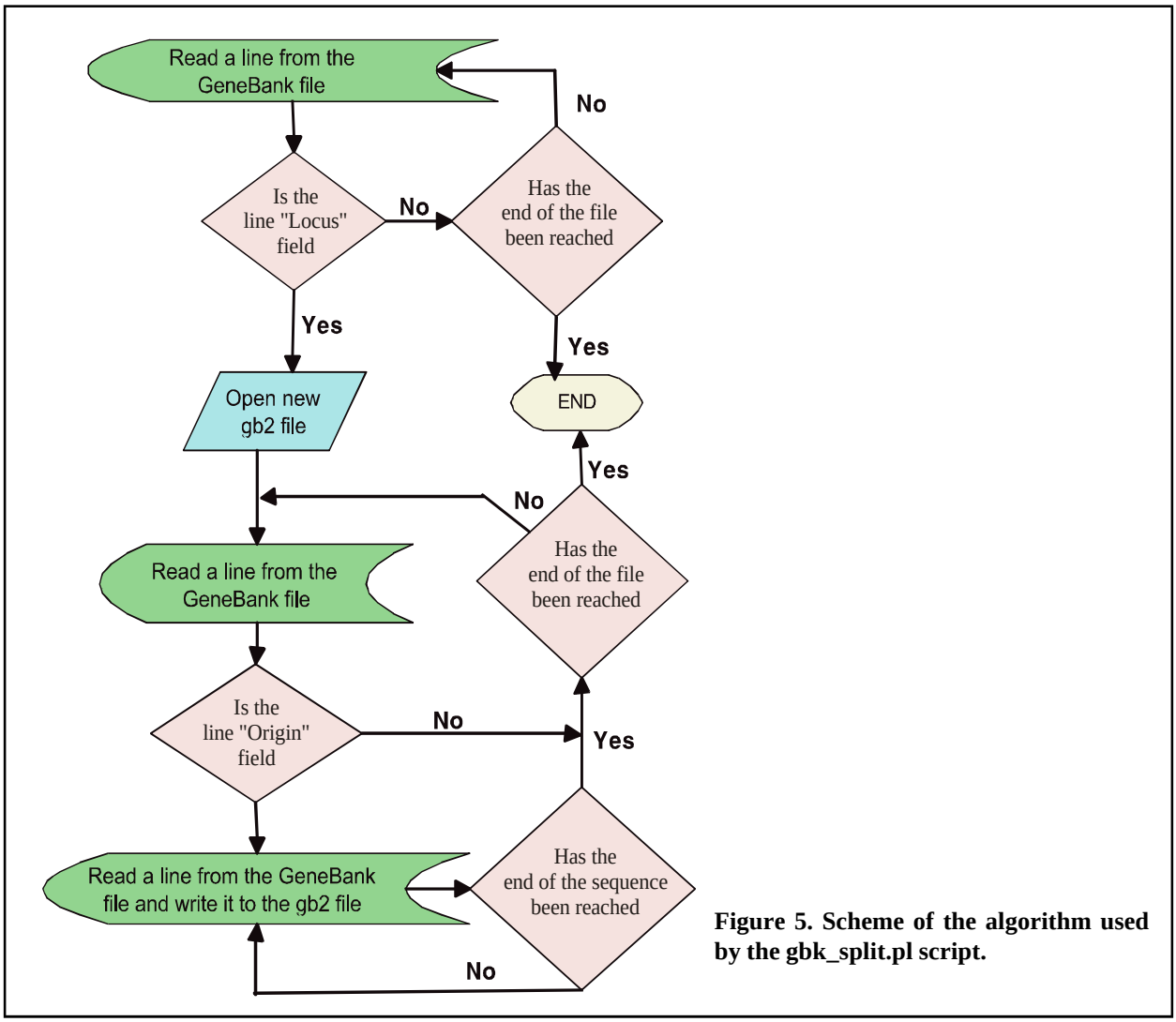


Figure 5. Scheme of the algorithm used by the gbk_split.pl script.

```
#!/usr/bin/perl
# gbk_split.pl
# Peter Stoilov
# This script reads the genbank chromosome files
# splits them in separate files that contain one contig each.

($#ARGV != -1) || die “Usage: first_parse.pl prefix\n”;
$prefix = $ARGV[0];

open(GBK, “${prefix}.gbk”) || die “Could not open ${prefix}.gbk\n”;
$counter = 0;
$line = <GBK>;
if ($line =~ /^LOCUS/){
          seek (GBK,0,0);
}
while ($line = <GBK>){
               if ($line =~ /^LOCUS/) {
                  ($locus) = ($line =~ /^LOCUS\s+(\S+)/);
                  open (OUT, “>${prefix}.${locus}.gb2”) || die;
               }
               print OUT $line;
               until ($line=~ /^ORIGIN/){
          $line=<GBK>;
               }
               until ($line=~ /^\/\//){
           print OUT $line;
           $line=<GBK>;
               }
               print OUT $line;
          }
```

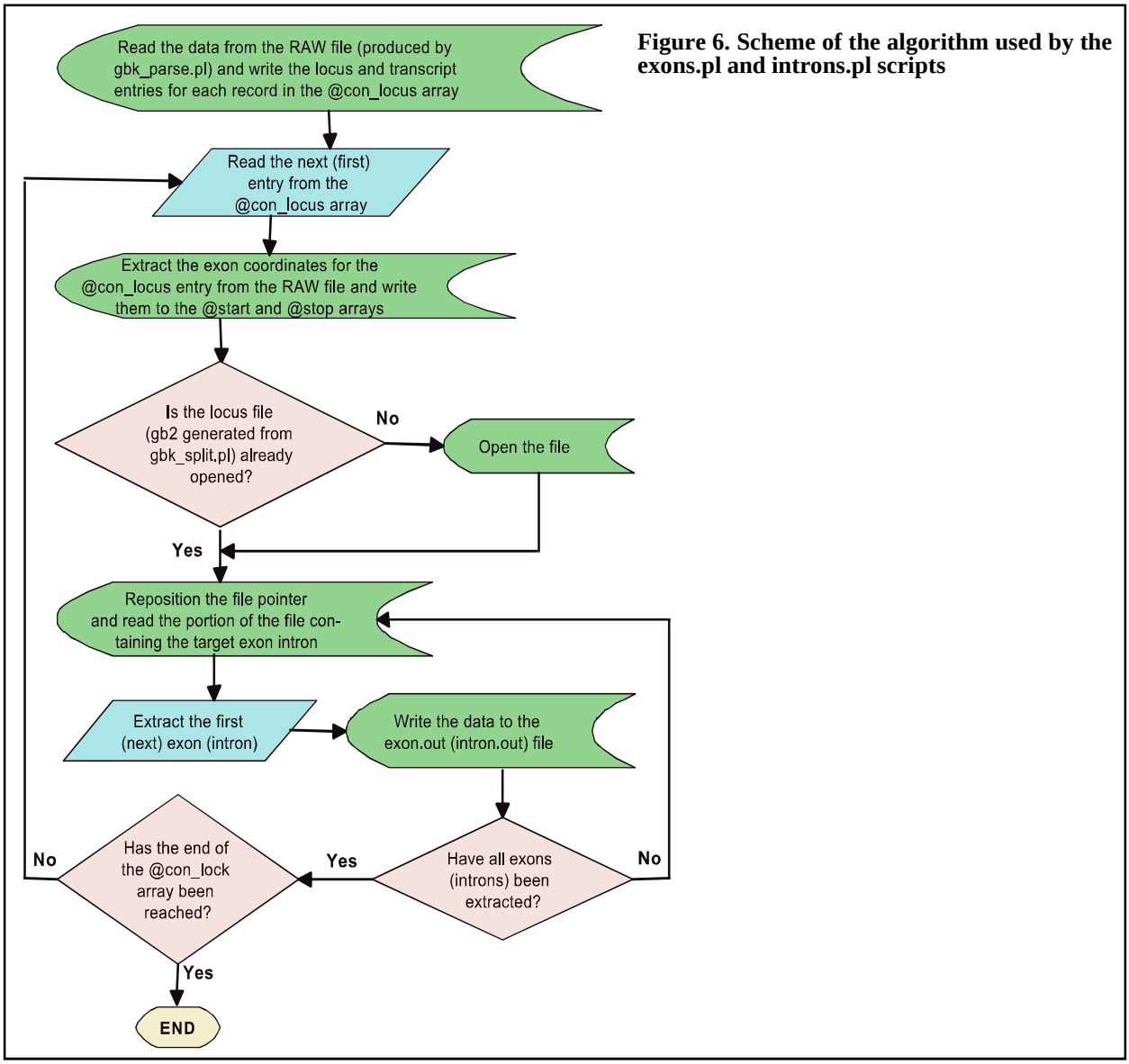


Figure 6. Scheme of the algorithm used by the exons.pl and introns.pl scripts

5.3.3. introns.pl

This script extracts, anotates and saves the intron sequences in FASTA format (Figure 6).

```
#!/usr/bin/perl    -w
# introns.pl
# Peter Stoilov
# Intron extraction script

($#ARGV != -1) || die “Usage: introns.pl prefix\n”;
$prefix = $ARGV[0];

# The original genebank file is GBK. the filter file with the
# extracted CDS joins and the CDS accecions is FILTER
# OUT is the output file

# First we read the FILTER file, that is generated by the gbk_parse.pl
# script to make a list of the transcripts wich is recorded in the
# @con_locus array

$cur_contig=””;
$base_number = 0;
@con_locus = &make_lists;
open (FILTER, “${prefix}.RAW”) || die;
open (OUTIN, “>${prefix}.in.out”) || die;

# The @con_locus array is read and each entry i split into contig and
# locus records that are kept in the @entry array. If the contig record
# is different that the name of the conting that was procesed last
# ($cur_conting) the $base_number counter and the $cur_contig are reset.
# Than the subroutine to extract the introns is launched. This is repeated
# until all of the records in the @con_locus array are processed.

$s_line=””;
$base_number=0;
$k=0;
$j=0;
$cur_contig = “”;
until ($k>$#con_locus){
        @entry=””;
        @entry = split (“&”, $con_locus[$k]);
        if ($cur_contig ne $entry[0]){
                $cur_contig = $entry[0];
                $base_number=0;
                    $pos=0;
        }
        print $con_locus[$k].”- transcript - “.$k.” – “;
        &get_them;
        ++$k;
}


sub make_lists {

          # This is the subroutine that makes the lists of the transcripts and
          # records them in the @con_locus array

        $j = 0;
        $temp=””;
        print “making lists\n”;
        # First we make a list of the locus entries:
        open (FILTER, “${prefix}.RAW”) || die;
        while ($line = <FILTER>){
                if ($line =~ /^LOCUS/) {
                  ($locus) = ($line =~ /^LOCUS\s+(\S+)/);
                  $line=<FILTER>;
                  ($contig)= ($line =~ /^CONTIG\s+(\S+)/);
                        $temp = $contig.”&”.$locus;
                        push (@con_lock, $temp);
                }
        }
        close FILTER;

         print “done making lists - “.($#con_lock+1).” genes\n”;
        $j=0;
        return @con_lock;
}

sub get_them {

          #Here we open the contig file (GBK) that was created by the gbk_split.pl
          # script. Than we search for the first/next transcript for the corresponding
          # contig in the FILTER file.

       open (GBK, “${prefix}.${entry[0]}.gb2”) || die;
LABEL1: while ($line = <FILTER> ){
          if($line =~ /^\$ID/){
                $id=$line;
                $line = <FILTER>;
                ($locus) = ($line =~ /^LOCUS\s+(\S+)/);

                    # There are some discrepancies in the GeneBank files wich initialy
                    # lead to generation of corrupted files from the gbk_parse.pl and
```

```
                # gbk_split.pl scripts. The two scripts now are supposed to deal
                # with this but I still keep the following two checks, just in case;)

                if(!defined($locus)){
                        die "Couldn't resolve the LOCUS line in $cur_entry\n";
                }
              if ($locus ne $entry[1]){
               print $id."—".$locus."—".$entry[1]."\n";
               die;
                      next LABEL1;
              }
              $cur_entry .= $locus;
        }

          # Now the exon locations will be read from the mRNA field and the start
          # and the end of each exon will be recorded in the @start and @stop arrays.
          # The $complement variable will be set to true if the transcript is located
          # on the complementary strand.

        elsif ($line =~ /^\s+mRNA/){
              @start="";
              @stop="";
              if ($line =~ /^\s+mRNA\s+complement\(join/) {
                      $complement = 1;
              }
              else {
                      $complement = 0;
              }
              $cds_line="";
              while ($line !~ /^\s+\//){
                      $line =~ s/ //g;
                      $cds_line .= $line;
                      $line  = <FILTER>;
              }
              while ($line !~ /^\s+\/gene/){
                      $line = <FILTER>;
              }
              @g = split /\"/, $line;   #"
              $gene = $g[1];
              $cds_line =~ s/[\(\)\n]//g;
              @raw_borders = split (/\.\./, $cds_line);
              $i=1;
              until ($line=~ /^\n/){
          $line=<FILTER>;
              }
              until ( $i == $#raw_borders ){
                      @b = split (/\,/, $raw_borders[$i]);
                      push (@start, $b[0]);
                      push (@stop, $b[1]);
                      ++$i;
              }
              last LABEL1;
        }
}

                      print $i." exons ".$stop[1]."\n";
                      $p=1;
                                  # Once we know where the exons are, we can start the extraction
                      &extract_introns;
}


sub extract_introns {
if ($pos==0){
          &find_origin;
}
else {
                    $bbb=$start[$p]-1;
                    if ($bbb < 0){
                              $bbb=0;
                    }
          if ($bbb < $base_number){
                    until ($bbb >= $base_number){
                              if ($base_number > 0){
                                        seek(GBK,$pos,0);            # The $pos variable keeps track of where we were
                                        $line=<GBK>;                 # in the GeneBank file. This greatly improves the
                                        $line=<GBK>;                 # speed, because we dont have to start at the
                                        $line=~s/[a-z]| |\n//g;      # begining of the file each time. Actualy it is
                                        $base_number=$line; # set  10 000 bytes before our last position
          $base_number=$base_number+59;
                                        &set_pointer;
                              }
                    }
                    if ($base_number==0){
                              &find_origin;
                    }
          }
          else{
                    seek(GBK,$pos,0);
                    $line=<GBK>;
                    $line=<GBK>;
                    $line=~s/[a-z]| |\n//g;
                    $base_number=$line;
                    $base_number=$base_number+59;
          }
}

LABEL2:       until ($p > $#start){    #$p keeps track of how many introns we have found
LABEL3:          until ($stop[$p] <= $base_number){

# In GeneBank the coordinates of the exon are marked as start and stop position, which
```

```
# has to be converted into start and lenght. We also do not want to keep the entire
# sequence of the locus in the memory, thats why we keep only a portion of it (starting
# maximum 59nt before the start position) and recalculate the start into the $offset variable

                        if ($base_number < $start[$p]){
                          $offset = abs ($base_number-$start[$p]);
                          $raw_intr="";     # here we record the part of the genomic
                                                                  # sequence where the intron is located
                                        if ($p==1){
                                              &set_pointer;
                                        }
                        }
                        else {
                           $raw_intr .= $s_line;
                        }
                        $s_line = <GBK>;
                        if ($s_line=~ /^LOCUS/){
                            die;
                            last LABEL2;
                        }
                        if ($s_line=~ /\// || $s_line=~ /^\n/){
                            last LABEL2;
                        }
                        $s_line =~ s/[0-9]| |\n//g; # strip the line of spaces and digits and EOL
                        $base_number += length $s_line;
                }
                    $raw_intr .= $s_line;
                    $len = $stop[$p] - $start[$p]-1;                # here we extract the intron
                    $intr = substr $raw_intr, $offset, $len;#
                    $intr_number=$p;                                #
                    if ( $complement == 1){                         #
          $intr = reverse $intr;                          #
          $intr =~ tr/gatcry/ctagyr/; #/           reverse complement if on the
          $intr_number = $#start+1-$p;           # complementary strand and
                    }                                                                   # set the counter

          # what is left now is to generate some sort of ID and save the intron in the output file

                    $idn=$locus."\#".$intr_number;
                  print OUTIN ">".$idn." : ".$cur_contig." : ".$locus." : ".$intr_number." : "." Gene :".$gene." : intron number
: ".$intr_number." between nucleotide ".$start[$p]." and ".$stop[$p]."\n".$intr."\n";
                    ++$p;
                    $raw_intr=$s_line;
        }
}


sub set_pointer{ # finds where are we in the GBK file and sets the new position
          $pos=tell(GBK);
          $pos=$pos-200;
          if ($pos<=0){
                    $pos=0;
          $base_number=0;
          }
}


sub find_origin{ # finds where the sequence starts
          $loc="";
          seek(GBK,0,0);
          until ($entry[0] eq $loc ){
          $s_line = <GBK>;
                    if ($s_line=~ /^LOCUS/){                                  # this is suposed to find the
                              ($loc) = ($s_line =~ /^LOCUS\s+(\S+)/);   # begining of the locus
          }
          }
      until  ($s_line =~ /^ORIGIN/){ #this is where the sequence starts
                $s_line = <GBK>;
      }
}
```

5.3.4. exons.pl

This script extracts, anotates and saves the exon sequences in FASTA format (Figure 6).

```
#!/usr/bin/perl     -w
# exons.pl
# Peter Stoilov
# Exon extraction script
# This script is almost identical to the intron extraction script.
# The major difference is taht it also records the splice site.

($#ARGV != -1) || die "Usage: introns.pl prefix\n";
$prefix = $ARGV[0];

$cur_contig="";
$base_number = 0;
@con_locus = &make_lists;

open (FILTER, "${prefix}.RAW") || die;
open (OUTEX, ">${prefix}.ex.out") || die;
open (OUTSS, ">${prefix}.ss.out") || die;

$s_line="";
$base_number=0;
$k=0;
```

```
$j=0;
$cur_contig = “”;
until ($k>$#con_locus){
        @entry=””;
        @entry = split (“&”, $con_locus[$k]);
             if ($cur_contig ne $entry[0]){
                             $base_number=0;
                             $pos=0;
                             $cur_contig = $entry[0];
              }
        print $con_locus[$k].”- transcript - “.$k.” – “;
        &get_them;
        ++$k;
}

sub make_lists {
        $j = 0;
        $temp=””;
        print “making lists\n”;
        open (FILTER, “${prefix}.RAW”) || die;
        while ($line = <FILTER>){
                if ($line =~ /^LOCUS/) {
                  ($locus) = ($line =~ /^LOCUS\s+(\S+)/);

                  $line=<FILTER>;
                  ($contig)= ($line =~ /^CONTIG\s+(\S+)/);
                        $temp = $contig.”&”.$locus;
                        push (@con_lock, $temp);
               }
        }
        close FILTER;
        print “done making lists - “.($#con_lock+1).” genes\n”;
        $j=0;
        return @con_lock;
}


sub get_them {
        open (GBK, “${prefix}.${entry[0]}.gb2”) || die;
LABEL1: while ($line = <FILTER> ){
          if($line =~ /^\$ID/){
                $id=$line;
                $line = <FILTER>;
                ($locus) = ($line =~ /^LOCUS\s+(\S+)/);
                    if(!defined($locus)){
                             die “Couldn’t resolve the LOCUS line in $cur_entry\n”;
                    }
                if ($locus ne $entry[1]){
                 print $id.”–”.$locus.”–”.$entry[1].”\n”;
                 die;
                         next LABEL1;
                }
                $cur_entry .= $locus;
        }
        elsif ($line =~ /^\s+mRNA/){
                @start=””;
                @stop=””;
                if ($line =~ /^\s+mRNA\s+complement\(join/) {
                        $complement = 1;
                }
                else {
                        $complement = 0;
                }
                $cds_line=””;
                while ($line !~ /^\s+\//){
                        $line =~ s/ //g;
                        $cds_line .= $line;
                        $line  = <FILTER>;
                }
                while ($line !~ /^\s+\/gene/){
                        $line = <FILTER>;
                }
                 @g = split /\”/, $line;   #”
                $gene = $g[1];
                until ($line=~ /^\n/){
                   $line=<FILTER>;
                }
                $cds_line =~ s/[a-z]||[A-Z]||[\(\)\n]//g;
                $i=1;
                 @raw_borders = split (/,/,$cds_line);
                for ($i=0; $i<=$#raw_borders; ++$i){
                        @b = split (/\./, $raw_borders[$i]);
                        push (@start, $b[0]);
                        push (@stop, $b[2]);
                }
                last LABEL1;
        }
}
                        print $i.” exons “.$stop[1].”\n”;
                        $p=1;
                        &extract_introns;
}



sub extract_introns {
          if ($pos==0){
                   &find_origin;
          }
else {
                    $bbb=$start[$p]-13;
```

```
                    if ($bbb < 0){
                              $bbb=0;
                    }
          if ($bbb < $base_number){
                    until ($bbb >= $base_number){
                              if ($base_number > 0){
                                        seek(GBK,$pos,1);
                                        $line=<GBK>;
                                        $line=<GBK>;
                                        $line=~s/[a-z]| |\n//g;
                                        $base_number=$line;
                                        $base_number=$base_number+59;

                              &set_pointer;
                              }
                    }
                    if ($base_number==0){
                              &find_origin;
                    }
          }
          else{
                    seek(GBK,$pos,1);
                    $line=<GBK>;
                    $line=<GBK>;
                    $line=~s/[a-z]| |\n//g;
                    $base_number=$line;
                    $base_number=$base_number+59;
          }
}
print $base_number."\n";
LABEL2:       until ($p > $#start){
LABEL3:          until (($stop[$p]+13) <= $base_number){
                              if ($base_number < ($start[$p]-12)){
                                $offset = abs ($base_number-$start[$p]);
                                $raw_intr="";
                                                  if ($p==1){
                                                            &set_pointer;
                                                  }
                              }
                              else {
                                 $raw_intr .= $s_line;
                              }
                              $s_line = <GBK>;
                                 if ($s_line=~ /^LOCUS/){
                                           die;
                                           last LABEL2;
                                  }
                                  if ($s_line=~ /\// || $s_line=~ /^\n/){
                                                last LABEL2;
                                  }
                              $s_line =~ s/[0-9]| |\n//g;
                              $base_number += length $s_line;
               }

                    $raw_intr .= $s_line;
                    $len = $stop[$p] - $start[$p];
                    $intr = substr $raw_intr, $offset-1, $len+1;
                    $SS3_i = substr $raw_intr, ($offset-13), 12;
                    $SS3_e = substr $intr, 0, 1;
                    $SS3 = lc($SS3_i)."\/".uc($SS3_e);
                    $SS5_i = substr $raw_intr, ($offset+$len), 6;
                    $SS5_e = substr $intr, ($len-2), 3;
                    $SS5 = uc($SS5_e)."\/".lc($SS5_i);
                    $intr_number=$p;

                    if ( $complement == 1){
          $SS5_i = substr $raw_intr, ($offset-7), 6;
          $SS5_e = substr $intr, 0, 3;
          $SS5 = lc($SS5_i)."\/".uc($SS5_e);
          $SS3_i = substr $raw_intr, ($offset+$len), 12;
          $SS3_e = substr $intr, $len, 1;
          $SS3 = uc($SS3_e)."\/".lc($SS3_i);
          $SS3 = reverse $SS3;
          $SS3 =~ tr/gatcryn\/GATCRYN/ctagyrn\/CTAGYRN/;
          $SS5 = reverse $SS5;
          $SS5 =~ tr/gatcryn\/GATCRYN/ctagyrn\/CTAGYRN/;
          $intr = reverse $intr;
          $intr =~ tr/gatcryn/ctagyrn/;#/
          $intr_number = $#start+1-$p;
                    }
                    if ($intr_number == 1){
          $SS3="";
                    }
                    if ($intr_number == $#start){
          $SS5="";
                    }
                    print OUTEX ">".$locus."#".$intr_number ." : ".$SS3." : ".$SS5." : ".$cur_contig." : "." Gene : ".$gene." :
Exon number : ".$intr_number." between nucleotide ".$start[$p]." and ".$stop[$p]."\n".$intr."\n";
                              print  OUTSS  "&ID            &".$locus."#".$intr_number."\n"."&3SS          &".$SS3."\n"."&5SS
&".$SS5."\n"."ENDRECORD\n";
                    ++$p;
                    $raw_intr=$s_line;
        }
}#/


sub set_pointer{
          $pos=tell(GBK);
          $pos=$pos-200;
          if ($pos<=0){
                    $pos=0;
```

```
         $base_number=0;
}
}

sub find_origin{
$loc="";
seek(GBK,0,0);
until ($entry[0] eq $loc ){
        $s_line = <GBK>;

        if ($s_line=~ /^LOCUS/){
                ($loc) = ($s_line =~ /^LOCUS\s+(\S+)/);
        }
}
        until  ($s_line =~ /^ORIGIN/){
                $s_line = <GBK>;
        }
}
```

5.3.5. makeeid.pl

This script copies the GeneBank files, decompresses them and consecutively launches the gbk_parse.pl, gbk_split.pl, exons.pl and itrons.pl scripts.

```
#!/usr/bin/perl
# makeeid.pl
# Peter Stoilov
# Automation script
$i=22;
until ($i ==0){
$prefix="hs_chr".$i;
print "processing ".$prefix."....\n";
$pre=$prefix.".gbk.gz";
$cp="/mnt/spock/genome/".$prefix.".gbk.gz ".$prefix.".gbk.gz";
print "Copying files....\n";
system("cp $cp");
print "unziping....\n";
system("gunzip $pre");
print "cleaning archives....\n";
system("rm -f *gz");
print "parsing headers....\n";
system("gbk_parse.pl $prefix");
print "spliting contigs....\n";
system("gbk_split.pl $prefix");
print "retreaving introns....\n";
system("introns.pl $prefix");
print "retreaving exons....\n";
system("exons.pl $prefix");
print "cleaning....\n";
system("rm -f $prefix.RAW");
system("rm -f $prefix.gbk");
system("rm -f *gb2");
print "moving files....\n";
system("mv  $prefix* /mnt/spock/genome/done");
$i=$i-1;
}
$prefix="hs_chrY";
print "processing ".$prefix."....\n";
$pre=$prefix.".gbk.gz";
$cp="/mnt/spock/genome/".$prefix.".gbk.gz ".$prefix.".gbk.gz";
print "Copying files....\n";
system("cp $cp");
print "unziping....\n";
system("gunzip $pre");
print "cleaning archives....\n";
system("rm -f *gz");
print "parsing headers....\n";
system("gbk_parse.pl $prefix");
print "spliting contigs....\n";
system("gbk_split.pl $prefix");
print "retreaving introns....\n";
system("introns.pl $prefix");
print "retreaving exons....\n";
system("exons.pl $prefix");
print "cleaning....\n";
system("rm -f $prefix.RAW");
system("rm -f $prefix.gbk");
system("rm -f *gb2");
print "moving files....\n";
system("mv  $prefix* /mnt/spock/genome/done");
$prefix="hs_chrX";
print "processing ".$prefix."....\n";
$pre=$prefix.".gbk.gz";
$cp="/mnt/spock/genome/".$prefix.".gbk.gz ".$prefix.".gbk.gz";
print "Copying files....\n";
system("cp $cp");
print "unziping....\n";
system("gunzip $pre");
print "cleaning archives....\n";
system("rm -f *gz");
print "parsing headers....\n";
system("gbk_parse.pl $prefix");
print "spliting contigs....\n";
system("gbk_split.pl $prefix");
```

```
print “retreaving introns....\n”;
system(“introns9.pl $prefix”);
print “retreaving exons....\n”;
system(“exons4.pl $prefix”);
print “cleaning....\n”;
system(“rm -f $prefix.RAW”);
system(“rm -f $prefix.gbk”);
system(“rm -f *gb2”);
print “moving files....\n”;
system(“mv  $prefix* /mnt/spock/genome/done”);
$i=$i-1;
print “DONE!\n”;
```

5.3.6. pattern.pl

This script searches the exons and the introns for matches to the patterns of the consensus binding sites for several splicing factors and estimates their statistical significance (Figure 7).

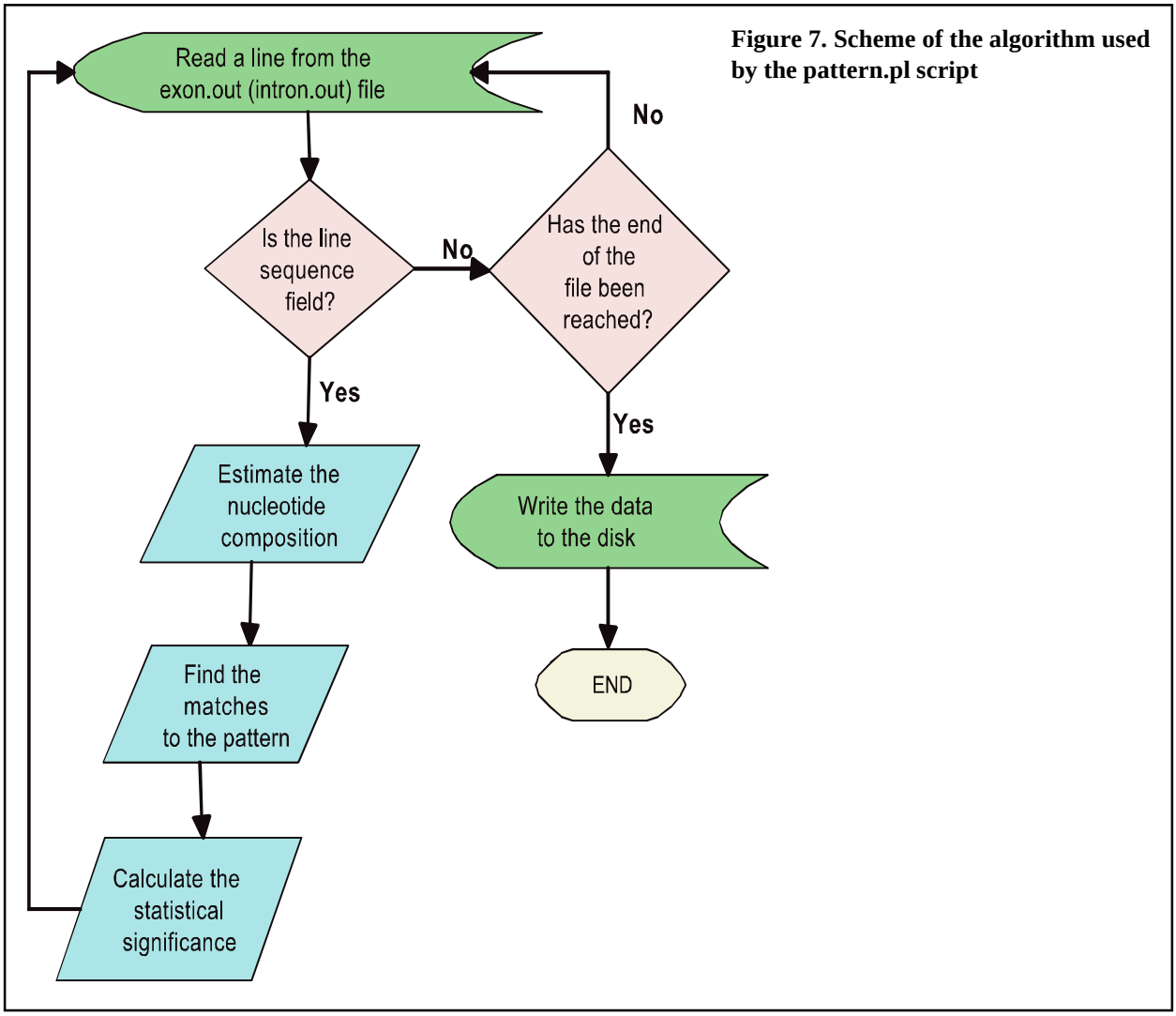


Figure 7. Scheme of the algorithm used by the pattern.pl script

```
#!/usr/bin/perl     -w
# pattern.pl
# Peter Stoilov
# This script searches the exons and the introns for matches
# to the patterns of the consensus binding sites for several
# splicing factors and estimates their statistical significance

$q=23; #Gues what is this;)
for ($d=1; $d<=20; $d++){ #reset some counters
          $ih[$d]=0;
          $eh[$d]=0;
          $ihh[$d]=0;
          $ehh[$d]=0;
}
# Define the splicing factors that we are interested in and their motifs
# set also the scale of the probabilities
```

```
%motifs=('tra','g[cat][acg][acg]ga[cgat][ag]','sf2','[gc][ag][gc]a[gc]ga','srp40','ac[atg]g[gc]','srp55','t[gc]cg[tg][ac]',
         'sc35','g[ag][ct][ct][gatc][cg][ct][ag]','hnh','ggga','ptb','[ct][ct][ct][ct][ct][ct][ct]');
@factors=('tra','sf2','srp40','srp55','sc35','hnh','ptb');
@scale=('10','1','0.1','0.01','0.001','0.0001','0.00001','0.000001','0.0000001','0.00000001','0.000000001','0.0000000001');
until ($q<1){
          $ade_e=0; #reset more counters
          $gua_e=0;
          $cyt_e=0;
          $thi_e=0;
          $ade_i=0;
          $gua_i=0;
          $cyt_i=0;
          $thi_i=0;
          $totalex =0;
          $lenex = 0;
          $totalin=0;
          $lenin = 0;
          $tte=0;
          $tti=0;
          $et=0;
          $ef=0;
          $it=0;
          $if=0;
          # specify the path to the database and open the files
          $prefix="/mnt/sda1/test/genome/done/hs_chr".$q;
          $prefix1="./done/hs_chr".$q;
          open (INEX, "${prefix}.ex.out") || die;
          open (OUTEX, ">${prefix1}.ep3xx.out") || die;
          open (OUTIN, ">${prefix1}.ip3xx.out") || die;
          open (STAT,  ">${prefix1}.st3xx.out") || die;

LABEL1:  while ($line = <INEX>){# finds the exon definition and sequence lines
            if ($line =~ /^>/){
                    ++$et;
                    $line=~s/\n//;
                    $id=$line;
             $line = <INEX>;
                    $line_len = length ($line);
                    $lenex += $line_len;
                    @a= ( $line =~ m/a/xg ); # check the base composition
                    @g= ( $line =~ m/g/xg ); # we need this for the statistics
                    @c= ( $line =~ m/c/xg );
                    @t= ( $line =~ m/t/xg );
                    &pattern_look; # find the patterns
             $tte+=$line_count;
                    #scale the hits
                    for ($d=1; $d<=10; $d++){
                              if ($test>$scale[$d] && $test<=$scale[$d-1]){
                                       $eh[$d]+=$hits;
                                       ++$ehh[$d];
                              }
                    }
if ($test>0.001){   # no need to save the exon specific data
          next LABEL1;        # if the hits are not significant
}
                    $ade_e+=($#a+1);
                    $gua_e+=($#g+1);
                    $cyt_e+=($#c+1);
                    $thi_e+=($#t+1);
                    ++$ef;
             $totalex += $line_count;
                    print OUTEX $id." tra sites: ".$line_count." size "."\n".$line;
             }
}

for ($d=1; $d<=10; $d++){
          print " ".$scale[$d]."—".$eh[$d]."—".$ehh[$d]."\n";
}
# print and save the data
print $prefix." Exon matches    : ".$tte." filtered ".$totalex." percent ".($totalex/$tte)." exon count ".$et." filtered exons
".$ef."  in ".$lenex." nucleotides of sequence\n";#/

print STAT $prefix." Exon matches   : ".$tte." filtered ".$totalex." percent ".($totalex/$tte)." exon count ".$et." filtered exons
".$ef."  in ".$lenex." nucleotides of sequence\n";#/

# repeat the same procedure for the introns
open (ININ, "${prefix}.in.out") || die;
LABEL2: while ($line = <ININ>){
        if ($line =~ /^>/){
                    ++$it;
                    $line=~s/\n//;
                    $id=$line;
                $line = <ININ> ;

                     $line_len = length ($line);
                  $lenin += $line_len;
                  @a= ( $line =~ m/a/xg );
                    @g= ( $line =~ m/g/xg );
                    @c= ( $line =~ m/c/xg );
                    @t= ( $line =~ m/t/xg );
                    &pattern_look;
             $line_count = $#line_matches+1;
                    $tti+=$line_count;
          for ($d=1; $d<=10; $d++){
                    if ($test>$scale[$d] && $test<=$scale[$d-1]){
                              $ih[$d]+=$hits;
                              ++$ihh[$d];
                    }
          }

if ($test>0.001){
```

```
         next LABEL2;
}
                    $ade_i+=($#a+1);
                    $gua_i+=($#g+1);
                    $cyt_i+=($#c+1);
                    $thi_i+=($#t+1);
                    ++$if;
           $totalin += $line_count;
                    print OUTIN $id."tra sites: ".$line_count."\n";
       }
}

for ($d=1; $d<=10; $d++){
print " ".$scale[$d]."–".$ih[$d]."–".$ihh[$d]."\n";
}

if($totalin==0){$totalin=1;}
if($ade_i==0){$ade_i=1;}
if($gua_i==0){$gua_i=1;}
if($cyt_i==0){$cyt_i=1;}
if($thi_i==0){$thi_i=1;}
$rat = ($totalex/$lenex)*($lenin/$totalin);
$R_to_Y_in =($ade_i+$gua_i)/($cyt_i+$thi_i);#/
$R_to_Y_ex =($cyt_e+$gua_e)/($cyt_e+$thi_e);#/
$GC_in=($ade_i+$gua_i)/($cyt_i+$thi_i+$ade_i+$gua_i);#/
$GC_ex=($ade_e+$gua_e)/($cyt_e+$thi_e+$ade_e+$gua_e);#/

print $prefix." Intron matches : ".$tti." filtered ".$totalin." percent ".($totalin/$tti)." intron count ".$it." filtered introns
".$if."  in ".$lenin." nucleotides of sequence\n";#/
print "ratio: ".$rat." R to Y intron ".$R_to_Y_in." R to Y exon ".$R_to_Y_ex." GC intron ".$GC_in." GC exon ".$GC_ex."\n";

print STAT $prefix." Intron matches : ".$tti." filtered ".$totalin." percent ".($totalin/$tti)." intron count ".$it." filtered
introns ".$if."  in ".$lenin." nucleotides of sequence\n";#/
print STAT "ratio: ".$rat." R to Y intron ".$R_to_Y_in." R to Y exon ".$R_to_Y_ex." GC intron ".$GC_in." GC exon ".$GC_ex."\n\n";
$q=$q-1;
}

open (HIST,  ">${prefix1}.st3xx") || die;
for ($d=1; $d<=10; $d++){
print HIST " ".$scale[$d]." exon ".$eh[$d]." – ".$ehh[$d]." intron ".$ih[$d]." –".$ihh[$d]."\n";
}

sub pattern_look{
my $lv;
$ex=0;
$hits=0;
foreach $lv(@factors){
                    # Searches the hits and puts them in @line_matches
           @line_matches = ( $line =~ m/$motifs{${lv}}/xg );
                    $hits+=($#line_matches+1); # adds up the number of the hits
           if (!defined $line_matches[0]){$line_count=0;}
                    else{
                              $line_count = $#line_matches+1;
                              $subf="e_".$lv;
                              &${subf};
                              $ex+=($line_count/($line_len*$ta));#/
                    }
}
$test=exp(-$ex);

}

# these subroutines calculate the statistical significance of the hits

sub e_tra{
                    $t1=($#g+1)/$line_len;
                    $t2=($#a+$#c+$#t+3)/$line_len;
                    $t3=($#a+$#c+$#g+3)/$line_len;
                    $t4=($#a+1)/$line_len;
                    $t5=($#a+$#g+2)/$line_len;#/
$ta=($t1*$t1*$t2*$t3*$t3*$t4*$t5);
}
sub e_sf2{
                    $t1=($#g+$#c+2)/$line_len;
                    $t2=($#a+$#g+2)/$line_len;
                    $t3=($#a+1)/$line_len;
                    $t4=($#g+1)/$line_len;
$ta=($t1*$t1*$t2*$t3*$t3*$t4);
}
sub e_srp40{
                    $t1=($#a+$#t+$#g+3)/$line_len;
                    $t2=($#g+$#c+2)/$line_len;
                    $t3=($#a+1)/$line_len;
                    $t4=($#c+1)/$line_len;
                    $t5=($#g+1)/$line_len;#/
$ta=($t1*$t2*$t3*$t4*$t5);
}
sub e_srp55{
                    $t1=($#a+$#c+2)/$line_len;
                    $t2=($#g+$#t+2)/$line_len;
                    $t3=($#t+1)/$line_len;
                    $t4=($#c+1)/$line_len;
                    $t5=($#g+1)/$line_len;#/
$ta=($t1*$t2*$t3*$t4*$t5);
}
sub e_sc35{
                    $t1=($#a+$#g+2)/$line_len;
                    $t2=($#g+$#c+2)/$line_len;
#                   $t3=($#c+1)/$line_len;
                    $t4=($#g+1)/$line_len;
                    $t5=($#c+$#t+2)/$line_len;#/
```

```
$ta=($t1*$t1*$t2*$t4*$t5*$t5*$t5);
}
sub e_ptb{
                    $t5=($#c+$#t+2)/$line_len;#/
$ta=(-$t5*$t5*$t5*$t5*$t5*$t5*$t5*$t5);
}
sub e_hnh{
                    $t1=($#a+1)/$line_len;
                    $t2=($#g+1)/$line_len;
$ta=(-$t1*$t2*$t2*$t2);
}
sub e_9g8{
                    $t1=($#g+$#t+2)/$line_len;
                    $t2=($#a+1)/$line_len;
                    $t3=($#c+1)/$line_len;
                    $t4=($#g+1)/$line_len;#/
$ta=($t1*$t2*$t2*$t3*$t3*$t4);
}
```

Table 3. Primers used for RT-PCR and 5' RACE experiments

Primer	Sequence	Target
hcTNT RT F	ATAGAAGAGGTGGTGGAAGAGTAC	Human cTNT minigene
hcTNT RT R	GTCTCAGCCTCTGCTTCAGCATCC	
pR(+)	ATTCACCACATTGGTGTGCA	APP minigene
p9(-)	ATTCTCTCTCGGTGCTTGGC	
pCIforward	GGTGTCCACTCCCAGTTCAA	SMN2 minigene
SMNex8rev	GCCTCACCACCGTGCTGG	
X16-T7	TAATACGACTCACTATAGGG	SRp20 minigene
X16F	CCCCAGCTGCAGACCATGCATCGTGATTCC	
X16R	CCTGGTCGACACTCTAGATTTCCTTTCATTT GACC	
5' miniG	AGAATTCCACGTGGAACAGAG	SWAP minigene
3MG2PCR	TAAGATCTGGATCCTAATAG	
pSG5 F	TAATACGACTCACTATAGGGC	Insulin receptor minigene
pSG5 R	GCTGCAATAAACAAGTTCTGC	
INS1	CAGCTACAGTCGGAAACCATCAGCAAGCAG	Tau minigene
INS3	CACCTCCAGTGCCAAGGTCTGAAGGTCACC	
N5Ins	GAGGGATCCGCTTCCTGCCCC	CD44v5 minigene, pET-LRR minigene
N3Ins	CTCCCGGGCCACCTCCAGTGCC	
cTNT F	CATTCACCACATTGGTGTGC	Chicken cTNT minigene
cTNT R	AGGTGCTGCCGCCGGGCGGTGGCTG	
TrkB E1 F	GGAGCAGGAGCCTCGCTGG	Human TrkB cDNA
TrkB E2 R	ATCTCGGAAATGCCACGATGCC	
TrkB E2 F	TCGGATCTGGTGCAGCGACC	
TrkB E6 R	TCTGCAGGTTTGCCAGGGGA	
TrkB E6 F	TGGTGGGCAATCCATTTACATGC	
TrkB E9 R	GGCCAGCCCATGAAGTGAGC	
TrkB E9 F	CACGGAGTACCACGGCTGCC	
TrkB E11 R	CGAGAGATGTTCCCGACCGGTT	
TrkB E11 F	CGAATGACATCGGGGACACCAC	
TrkB E13TK+ R	TTGAGCTGACTGTTGGTGATGCCA	
TrkB E13TK- R	CTTGAGGTGCTGCTTAGCTGCCTG	
TrkB E13TK+ F	TGATGATGACTCTGCCAGCCCA	
TrkB E18 R	CGGGGTCGCTGCAGGACTC	
TrkB E21 R2	GCAGGACTCGGCCCTGAGTG	
trkB E16 F	CCAGAGGTTCCCCCAAGAC	
trkB E16 R1	GACGTCAAGATGTTGTCTGGC	
trkB E12 F	GTCTATGCTGTGGTGGTGAT	
trkB E12 R	CTTTCATGCCAAACTTGGAG	
trkB E20 F	GTCGGTGGCCACACAATGC	
trkB E2 F2	GGACAGGCACTCGGGC	
trkB E21 R3	ATGTCCAGGTAGACCGGAGA	
mtrkb_ex_shc-r	CTCTTCTCCTCCATCAGCCT	Mouse TrkB cDNA
mtrkb_ex_shc_r2	ATTATTATCAGGCGGTCTTGG	
mtrkb_ex_14_f	GGCGAGACATTCCAAGTTTG	

Table 4. Clones from the Stamm's lab collection or outside sources

Clone name	Description	Source	Reference
ASF/SF2-EGFP	SF2/ASF cloned in pEGFP C2	Nayler O.	(Nayler et al., 1998b; Nayler et al., 1997)
Clk2-flag	Flag tagged mouse Clk2 cloned in pCDNA		
Clk2-EGFP	Mouse Clk2 cloned in pEGFP C2		
Clk2KR-EGFP	Mouse Clk2KR (kinase dead) cloned in pEGFP C2		
Clk1-flag	Flag tagged mouse Clk1 cloned in pCDNA		
Clk1-EGFP	Mouse Clk1 cloned in pEGFP C2		
Clk1KR-EGFP	Mouse Clk1KR (kinase dead) cloned in pEGFP C2		
CUG-BP	CUG binding protein cloned in pCDNA	Cooper T.A.	(Philips et al., 1998)
ESAF-EGFP	Rat SLM-2 cloned in pEGFP C2		(Stoss et al., 2001)
SLM2 del4 EGFP	Rat SLM-2 with deleted KH domain cloned in pEGFP C2		
hnRNP-G fl EGFP	Rat hnRNP G in pEGFP C2		
hnRNP-G TH EGFP	Rat hnRNP G with deleted RRM cloned in pEGFP C2		
Pht6 fl EGFP	Rat YT521B cloned in pEGFP C2		(Hartmann et al., 1999)
hTra2-beta1 EGFP	Human tra2-beta1 in pEGFP C2		(Beil et al., 1997; Nayler et al., 1998a)
hTra2-beta3 EGFP	Human tra2-beta3 in EGFP C2		
hTra2-beta del SR2 EGFP	Human tra2-beta1 with deleted second RS domain cloned in pEGFP C2		
hTra2-beta1 flag	Flag tagged human tra2-beta1 in pCDNA		
hTra2-beta3 flag	Flag tagged human tra2-beta3 in pCDNA		
SF2/ASF FF/DD	SF2 with Aspartic acid to Fenylalanine substitutions in the RRMs.	Caceres J	(Caceres et al., 1998)

This table lists only vectors that were not created in this work. It also does not contain the minigenes that were used. For minigenes used in this work see table 5 and 6

Table 4. Clones from the Stamm's lab collection or outside sources (continued)

Clone name	Description	Source	Reference
hTra2-beta del SR2 flag	Flag tagged human tra2-beta1 with deleted second RS domain cloned in pCDNA		(Nayler et al., 1998c)
hTra2-beta del Sr1 + SR2 flag	Flag tagged human tra2-beta RRM cloned in pCDNA		
rSAF-B EGFP	Rat SAF-B partial clone lacking the RRM cloned in pEGFP C2		
pEGFP-C1-NLS-HAP full	Full length human SAF-B cloned in pEGFP C1	Denegri M	(Chiodi et al., 2000)
SF1 EGFP	Human SF1/BBP clone in pEGFP C2	Stoss O	
SRp20 pCDM8	Human SRp20 cloned in pCDM8	Screaton G.R.	(Screaton et al., 1995)
SRp30c pCDM8	Human SRp30c cloned in pCDM8		
SRp55 pCDM8	Human SRp55 cloned in pCDM8		
SRp75 pCDM8			
hnRNP A1 pCDM8	Human hnRNP A1 cloned in pCDM8		
SRp30C EGFP	Human SRp30c cloned in pEGFP C2	Stamm S	none
SC35 EGFP C2	Human SC35 cloned in pEGFP C2	Nayler O	none
hnRNP L EGFP	Rat hnRNP L cloned in pEGFP C2	Stamm S	none
pSWAP12exp	Mouse SWAP cloned in pSG5	Lafyatis	(Denhez and Lafyatis, 1994)
Znf265 EGFP	Human Znf-265 clone in pEGFP C2	Adams D.J.	(Adams et al., 2001)
Znf265 EGFP del 4	Human Znf-265 with deleted RRM cloned in pEGFP C2		
TrkB.TK(+)flag pEF/BOS	Rat full length TrkB cloned in pEF/BOS	Castren E.	(Haapasalo et al., 2001)

This table lists only vectors that were not created in this work. It also does not contain the minigenes that were used. For minigenes used in this work see table 5 and 6

Table 5. Newly made clones (continued)

Construct	Notes	Primers
pCR-XL MGTra	Human tra2-beta exons 1 to 4 genomic sequence was amplified by long range PCR and cloned in pCR XL TOPO.	MGTra-Bam: GGGGATCCGACCGGCGCGTCGTGCGGGGCT MGTra-R-Xho: GGGCTCGAGTACCCGATTCCCAACATGACG
MGTra	The insert of pCR-XL MGTra was recloned using BamHI and XhoI sites in pCR 3.1	
MGTra-M1-RY MGTra M1-RR	Minigenes with binding site one substituted either by pyrimidine or by purine rich sequence. Derived from MGTra by Kunkel mutagenesis.	TradBS1-R: CGACTTCCGCATTTTGGAAGAAGAACATTAACCTTAATAG BS1-R->R-rev2: CGACTTCCGCATTTTTTCCTCCTAACATTAACCTTAATAG
MGTra-M2-RY MGTra-M2-RR	Minigenes with binding site two substituted either by pyrimidine or by purine rich sequence. Derived from MGTra by Kunkel mutagenesis.	BS4-R: GATAATTAGCATGCAAAGAAAGTTTTATTTTTTCAAATTTC BS4-R->R-rev: GATAATTAGCATGCACCTCCCTTTTTATTTTTTCAAATTTC
MGTra-M3-RY MGTra-M3-RR	Minigenes with binding site three substituted either by pyrimidine or by purine rich sequence. Derived from MGTra by Kunkel mutagenesis.	BS3-R: CGTTTCAAATTCTGAGAAGAAAGATCTTCCCCACTTCACAC BS3-R->R-rev: CGTTTCAAATTCTGATCCTCCCTATCTTCCCCACTTCACAC
MGTra-M4-RY MGTra-M4-RR	Minigenes with binding site four substituted either by pyrimidine or by purine rich sequence. Derived from MGTra by Kunkel mutagenesis.	TraBS2+3-R: CTTCATTTTTATTTCTTAAGAAGAAAGAAGCGTTTCAAATTCTGACT BS2-R->R-rev: CTTCATTTTTATTTCTTCCTCCTACCTCCTCGTTTCAAATTCTGACT
MGTra-sno	Control minigene with 15nt substitution near binding site four. Derived from MGTra by Kunkel mutagenesis.	MGTra-sno ATTACTTCTTATGCTCAATAGGATTACGTTTCTTTTCTTC
pEGFP N1 M/I pEGFP N3 M/I	ATG codon of EGFP is changed to ATC by Kunkel mutagenesis of pEGFP N1 and pEGFP N3 vectors.	egfp-M/I: TCGCCACCATCGTGAGCAAGGGCGAG
MGTra-N1-M/I MGTra-N3-M/I	The insert of pCR-XL MGTra was recloned using BamHI and XhoI sites in BglII/SalI sites of pEGFP N1 M/I or pEGFP N3 M/I.	
pCR4-beta1 pCR4-beta3	cDNA clones of the alternatively spliced region of the tra2-beta minigene. Contain exons 1, 3 and 4 and 1 and 4 respectively. Amplified by RT-PCR and cloned using TOPO cloning in pCR4.	MGTra-Bam, MGTra-R-Xho

Table 5. Newly made clones (continued)

Construct	Notes	Primers
MGTra-3C1, MGTra-3C2, MGTra-3C3, MGTra-3C4, MGTra-3C5, MGTra-3C6, MGTra-3C7, MGTra-3C8, MGTra-3T1, MGTra-3T2, MGTra-3T3, MGTra-3T4, MGTra-3T5, MGTra-3T6, MGTra-3T7, MGTra-3T8, MGTra-3A1, MGTra-3G2, MGTra-3G3, MGTra-3G4, MGTra-3A5, MGTra-3G6, MGTra-3G7, MGTra-3A8	Minigenes containing point mutations of binding site 3. Derived from MGTra by Kunkel mutagenesis.	BS3-1Y: CGTTTCAAATTCTGACTTCTTTRAT CTTCCCCACTTCACAC BS3-2Y: CGTTTCAAATTCTGACTTCTTRCAT CTTCCCCACTTCACAC BS3-3Y: CGTTTCAAATTCTGACTTCTRTCAT CTTCCCCACTTCACAC BS3-4Y: CGTTTCAAATTCTGACTTCRTTCAT CTTCCCCACTTCACAC BS3-5Y: CGTTTCAAATTCTGACTTRTTTCAT CTTCCCCACTTCACAC BS3-6Y: CGTTTCAAATTCTGACTRCTTTCAT CTTCCCCACTTCACAC BS3-7Y: CGTTTCAAATTCTGACRTCTTTCAT CTTCCCCACTTCACAC BS3-8Y: CGTTTCAAATTCTGARTTCTTTCAT CTTCCCCACTTCACAC BS3-1A: CGTTTCAAATTCTGACTTCTTTTAT CTTCCCCACTTCACAC BS3-2G : CGTTTCAAATTCTGACTTCTTCCAT CTTCCCCACTTCACAC BS3-3G: CGTTTCAAATTCTGACTTCTCTCAT CTTCCCCACTTCACAC BS3-4G: CGTTTCAAATTCTGACTTCCTTCAT CTTCCCCACTTCACAC BS3-5A: CGTTTCAAATTCTGACTTTTTTCAT CTTCCCCACTTCACAC BS3-6G: CGTTTCAAATTCTGACTCCTTTCAT CTTCCCCACTTCACAC BS3-7G: CGTTTCAAATTCTGACCTCTTTCAT CTTCCCCACTTCACAC BS3-8A: CGTTTCAAATTCTGATTTCTTTCAT CTTCCCCACTTCACAC
pCR4 SRp30C delRRM	cDNA clone of SRp-30c lacking the N-terminal RRM domain. Amplified using RT-PCR and cloned using TOPO cloning in pCR4.	SRp30c-delRRM-r: TAATGGATCCTCAGTAGGGCCTGA AAGGAG SRp30c-delRRM-f: GATGGAATTCGACACCAAATTCCG CTCTCA

Table 5. Newly made clones (continued)

Construct	Notes	Primers
beta1-N1 M/I beta3-N1 M/I beta1-N3 M/I beta3-N3 M/I	The inserts of pCR4 beta1 and pCR4 beta3 were recloned using BamHI and XhoI sites in BglII/SalI sites of pEGFP N1 M/I or pEGFP N3 M/I	
pCR4 Tra2-alpha	cDNA clone of Tra2-alpha. Amplified using RT-PCR and cloned using TOPO cloning in pCR4	Tra2-alpha-f: ATGGATCCACATGAGTGATGTGGAGGAAAAC Tra2-alpha-r: TAGGATCCATAGCGTCTTGGGCTGTAGG
Tra2-alpha-C2	Tra2-alpha clone from pCR4 Tra2-alpha was cloned in pEGFP C2 using BamHI site	
pCR4 TrkB-T-shc	cDNA clone of TrkB-T-Shc. Amplified using RT-PCR and cloned using TOPO cloning in pCR4	trkb E16 SalI: GGTGTCGACTGGGCGGTCTTGGGGGAAC trkB E1 SacI: CTGGAGCTCATGTCGTCCTGGATAAGGTGG
EGFP-N1-TrkB-T-shc	TrkB-T-Shc clone from pCR4 TrkB-T-shc was cloned in pEGFP N1 using SacI and SalI sites	
TrkB-T-shc-flag	Derived from EGFP-N1 TrkB-T-shc by inserting syntetic double stranded oligonucleotide in the SalI site	Flag-trkb sense: TCGAgCATGGACTACAAAGACGATGACGACAAGTAA Flag-trkb antisense: tcgattaCTTGTCGTCATCGTCTTTGTAGTCCATgc
hTrkB-EGFP-N1	Human full length TrkB cDNA was amplified by RT-PCR, digested with SalI and cloned in EcoRV and SalI sites of pEGFP N1	trkb E16 SalI: GGTGTCGACTGGGCGGTCTTGGGGGAAC trkB fl sal R: GATTGTCGACTGGCCTAGAATGTCCAGGTAGACCG
pET-LRR	Exons 6, 7 and 8 from the human TrkB gene were amplified by PCR and cloned PstI/NotI in pET1 (Mobitec)	trkB-lrrF2 AATCCTGCAGGCGAGGAGAAATGAAACCCAGTGC trkB-lrrR2 AGTGCGGCCGCTACAGCCGTTTTCACACAAAGCC
SRp30C delRRM-C2	SRp30c delRRM clone from pCR4 SRp30C delRRM was cloned in pEGFP C2 using EcoRI and BamHI sites	

5.4. Isolation of plasmid DNA

Large amounts of plasmid were prepared by the Qiagen Plasmid Maxi kit and according to the Qiagen protocol. Smaller amounts of plasmid were isolated using the following procedure, based on the protocol of Birnboim and Doly. Buffers content was according to the Qiagen protocols. Bacterial cells carrying the corresponding plasmid were cultured for 16 hours, at 37°C, in 5 ml LB medium. The cells were pelleted by centrifugation for 10 minutes at 4000 rpm and the pellet was resuspended in 250 µl buffer P1. Equal volume of lysis buffer P2 was then added and the solution was allowed to stay for 5 minutes at RT, followed by addition of 250 µl of neutralisation buffer P2. After centrifugation for 10 minutes at 12000 rpm, the resulting supernatant was precipitated by adding 0.6 volumes of isopropanol. DNA was pelleted by centrifugation at 12 000 rpm, washed with 70 % ethanol, air-dried and dissolved in 100 µl of TE buffer (10 mM Tris-HCl, pH 8.0/1 mM EDTA).

LB medium	10g Trypton
(1l)	10g NaCl
	5g yeast extract
Buffer P1	50 mM Tris-HCl, pH 8.0
	10 mM EDTA
	100 µg/ml RNase A
Buffer P2	200 mM NaOH
	1 % SDS
Buffer P3	3 M Potassium acetate, pH 5.5

5.5. Determination of DNA concentration

The concentration of DNA in solution was estimated by a spectrophotometer, measuring the absorbtion of the solution at 260 nm and using the following formulas:

$1\ A_{260}$ = 50 µg double stranded DNA

1 A_{260} = 33 µg single stranded DNA

1 A_{260} = 40 µg RNA

5.6. Electrophoresis of DNA

DNA was resolved on 0.7 - 2 % agarose gels prepared in 1 x TBE buffer (90 mM Tris-borate/ 20 mM EDTA). The electrophoresis was run for 1 hour at 100 V. The gels were stained for 30 minutes in 0.5 mg/ml Ethidium bromide and visualised under UV light, ? = 260 nm.

Gel loading buffer, 6x	0.25 % Bromophenol blue
	0.25 % Xylene cyanol
	15 % Ficoll 400

5.7. Extraction of DNA from agarose gels

DNA was purified from agarose gels in which Crystal violet was added to a concentration of 2 µg/ml, to detect DNA under visible light. Individual bands were excised and DNA was extracted using the Qiagen gel extraction kit and according to the Qiagen protocol.

Crystal violet loading buffer	15 % Ficoll 400 in dH_2O
	Crystal violet 0.025 %

5.8. PCR amplification of DNA

A standard PCR reaction to amplify DNA from plasmid template contained about 1-10 ng plasmid DNA, forward and reverse primers (0.5 µM each), dNTPs (200 µM), 1 x Taq-polymerase buffer, 1.5 mM $MgCl_2$ and 1 U Taq polymerase (AmpliTaq DNA polymerase, Perkin Elmer), in 25 µl. When the amplification was made for cloning purposes, a high-fidelity polymerase was used instead of Taq polymerase, i.e. Platinum Pfx polymerase (Invitrogen), 1 U per 25 µl reaction. The amplification was carried out in a GeneAmp PCR System 9700 thermocycler (Perkin Elmer). The amplification conditions were as follows: (1) initial denaturation for 4 minutes at 94°C; (2) 25-30 cycles of 30-40 seconds at 94°C, annealing at the Tm of the primers pair, extension of 1 min/kbp at 72°C (68°C for Pfx polymerase); (3) incubation for 5 minuntes at the extension temperature to allow for the complete amplification of all products.

5.9. DNA ligation

When the vector ends were blunt or compatible with each other, the vector was dephosphorylated prior to ligation to prevent self-ligation. To remove 5' phosphates from the vector, 2 U of Calf intestinal phosphatase (CIP, Boehringer) were added to 5 µg of linearised vector in 1 x CIP buffer in 20 µl. The reaction was incubated for 1 hour at 37°C. CIP was subsequently inactivated by heating the reaction to 68°C for 20 min. A typical ligation reaction contained vector and insert at a ratio of about 1:3 (500-1000 ng total DNA), 1 x ligase buffer, 1 mM ATP and 200-400 U T4 DNA Ligase (New England Biolabs) in 15 µl. The incubation was carried out at 16°C for 12-16 h. After that, one third to one half of the ligation mixture were then transformed in *E coli* cells.

5.10. Preparation of competent *E.coli* cells

5 ml of LB medium were inoculated with a single bacterial colony and grown overnight at 37°C. 4 ml of this culture were transfered to 250 ml LB and grown to early logarythmic phase (OD_{600} = 0.3-0.6). The culture was centrifuged at 2500 rpm for 10 minutes at 4°C and the pelleted bacteria were resuspended in 1/10 volume of cold TSB buffer. After 10 minutes on ice, the cells were aliquoted in cold Eppendorf tubes and freezed in liquid nitrogen. Competent bacterial cells were stored at -80°C for several months.

TSB buffer 10 % PEG 3500

5 % DMSO

10 mM $MgCl_2$

10 mM $MgSO_4$

in LB medium, pH 6.1

5.11. Transformation of *E.coli* cells

1-10 ng of plasmid DNA or an aliquot from a ligation reaction were added to 100 µl of 1 x KCM buffer. Equal amount of competent cells was then added to this mixture. After incubation on ice for 20 min, the cells were allowed to stay at RT for 10 min. The bacteria were then incubated for 1 hour at 37°C in 1 ml of LB medium. They were then plated on antibiotic-containing plates. The plates were incubated for at 37°C until colonies appeared, after about 16 h.

5 x KCM buffer	500 mM KCl
	150 mM $CaCl_2$
	250 mM $MgCl_2$

5.12. Radioactive labelling of DNA

5.12.1. Labelling by PCR

This protocol was used for probes shorter than 200 bp. A standard PCR reaction was made, The PCR product was then excised from agarose gel and purified. 10 to 100 pg of this fragment were used as template and were further amplified in the presence of 10 pmol of each primer, 10 µM dATP, dGTP, dTTP, 1 x polymerase buffer, 1 U Taq polymerase and 50 µCi ? -P^{32}-dCTP (3000 Ci/ mmol) in 45 µl. The reaction was denatured at 94°C for 3 minutes and then 15 cycles of 30 seconds 94°C/30 seconds Tm/2 minutes 72°C were performed. After this step cold dNTPs were added to a final concentration of 150 µM and 5 additional cycles were carried out.

Single stranded DNA probes were generated by cycle extension in a mixture containing 1 ng of gel purified PCR product as template, 10 pmol primer, 10 µM dATP, dGTP, dTTP, 1 x polymerase buffer, 1 U Taq polymerase and 50 µCi ? -P^{32}-dCTP (3000 Ci/mmol), for 35 cycles.

PCR labelled DNA was purified from unincorporated nucleotides by the dNTP removal Kit (Qiagen) according to the manufacturer's protocol.

5.12.2. Labelling by random priming

When generating DNA probes longer than 200 bp, random priming was used as a method of choice. 25 ng of DNA were used and the reaction was carried out with the Megaprime DNA Labelling System (Amersham) which uses random nonamers for priming the DNA, according to Amersham's protocol. Probes were labelled to specific activity of 1×10^9 dpm/µg. DNA was purified by the dNTP removal kit (Qiagen).

5.12.3. Labelling of 5' DNA ends

About 500 ng of purified DNA fragment (or 5-10 pmol oligonucleotide) were mixed with 1 x T4 polynucleotide kinase buffer, 50 ? Ci of ?-P^{32}-ATP (3000 Ci/mmol) and 10 U T4 polynucleotide kinase (New England Biolabs) in 10 ?l. The reaction was incubated for 1 hour at 37°C after which the kinase was inactivated at 68°C for 20 min.

5.13. Southern Blotting and hybridisation of DNA

DNA was separated on an agarose gel. After electophoresis the gel was placed in denaturation

buffer (1.5 M NaCl/0.5 N NaOH) for 30 min. The DNA was transferred on a nylon membrane in a denaturation buffer for 16 h. The DNA was crosslinked to the membrane on a UV transilluminator. Prehybridisation was carried out for 30 minutes at 68°C in a hybridisation buffer. Hybridisation was done in 10-20 ml of hybridisation buffer for 16 hours at 68°C. The membrane was washed two times in 2 x SSC/0.1% SDS for 10 minutes at 68°C. It was than washed (two times) in 0.5 x SSC/0.1% SDS for 10 minutes at 68°C. Finally, the membrane was washed in 0.1 x SSC/0.1% SDS for 10 minutes at 68°C. The membrane was then exposed to X-ray film or the signal was revealed on a phosphorimager.

Hybridisation buffer	0.5M phosphate buffer, pH 7.2
	7% SDS
20 x SSC	3 M NaCl
	0.3 M Na citrate

5.14. Library screening and DNA sequencing

For TrkB, the human BAC library screen was performed by the German Human Gene Center, Heidelberg. ESTs matching the human TrkB cDNA sequence were used. Additional library screens were made using as a probe PCR products generated by the TrkB-E18-F and TrkB-E19-R primer pair.

Double stranded sequencing of the BAC clones was performed on an ABI Prism 370 sequencer, with 2 - 5 µg of DNA, 10 pmols of sequencing primer and 5 µl of sequencing mix (BigDye sequencing kit, Perkin Elmer) in a final volume of 15 µl. Cycling conditions were 4 minutes denaturation at 96°C and then 100 cycles of 10 seconds at 96°C, 10 seconds at 50°C and 4 minutes at 60°C. Additional sequences were obtained by transposone insertion sequencing using the GPS-1 Kit (New England Biolabs) and the corresponding protocol.

Sequencing of cloned inserts for verification was performed according to the following protocol. 200 to 1000 ng of plasmid DNA were mixed with 5 pmols of a suitable primer and 4 µl of sequencing mix (BigDye sequencing kit, Perkin Elmer) in a final volume of 15 µl. DNA was denatured at 95°C for 30 sec, annealing was at 50°C for 15 seconds and extension was at 60°C for 3 min; a total of 25 cycles was performed. The amplified DNA was precipitated by the addition of 2 µl of 1.5 M Na acetate/250 mM EDTA and 60 µl of ethanol for 15 minutes on ice. The tube was centrifuged for

15 minutes at 14 000 rpm and the pellet was washed with 250 µl of 70 % ethanol before being air dried. The sequencing gels were run by the sequencing unit of the Max-Planck Institute of Neurobiology, Munich and by the sequencing unit of the Institute of Biochemistry, University of Erlangen.

5.15. Site-directed mutagenesis of DNA

Site-directed mutagenesis was performed according to the method of Kunkel (Kunkel, 1985). The DNA of interest was cloned in a vector carrying the f1 phage origin of replication and thus capable of existing in a single stranded form. The recombinant plasmid was transformed in *E.coli* strain CJ236. This strain is deficient in dUTPase (*dut*) and uracil N-glycosilase (*ung*). These mutations result in high rate of incorporation of deoxy-uracil residues in its DNA. After transformation, colonies were grown on chloramphenicol-containing plates in addition to the plasmid-specific antibiotic, to ensure the presence of the F' episome necessary for production of helper phage. To isolate single stranded DNA from the plasmid of interest, colonies were grown in 5 ml of 2 x YT medium for 1 hour 30 minutes and then 5 x 10^8 pfu helper phage M13KO7 was added. The culture was grown for 12-14 hours at 37°C, after which single stranded DNA was isolated with the Qiagen M13 kit. This uracil-containing DNA was used as a template in the *in vitro* mutagenesis reaction. Mutant phosphorylated oligonucleotides were annealed to the template at a molar ratio of 20:1 in 10 µl of 1 x T7-polymerase buffer (New England Biolabs). The DNA was denatured for 5 minutes at 94°C, after which the temperature was gradually decreased from 70°C to 37°C at a rate of 1°C/min. The extension of the annealed primer was carried out in 20 µl, by adding to the same tube 1 µl of 10 x T7-polymerase buffer, 0.8 µl of 10 mM dNTPs, 1.5 µl 10 mM ATP, 3 U T7-polymerase and 2 U Fast-Link T4 DNA Ligase (Biozyme). The reaction was incubated at 37°C for 45 min. The ligase was inactivated by heating to 65°C for 20 min. The mutagenesis reaction was transformed in competent XL1Blue *E.coli* cells. Replication of the plasmid in this strain leads to repair of the template strand and consequently to production of plasmid carrying the desired mutation. All mutant plasmids were verified by sequencing.

2 x YT 16 g Bactotryptone

5 g Yeast extract

10 g NaCl

dH2O to 1 l

5.16. Isolation of total RNA

Total RNA was isolated from eukaryotic cells grown in 6-well plates. Cells were washed in 1 x PBS, after which the RNeasy Mini kit (Qiagen) was used according to the manufacturer's protocol. One column was used for each 3 cm well. RNA was eluted in 50 µl RNAse free dH_2O.

5.17. Northern blotting

Human multiple tissue northern blots were obtained from Clontech and were hybridized with radioactively labeled single stranded DNA probes using the ExpressHyb solution (Clontech), according to the manufacturer's protocol.

5.18. Reverse transcription

400 ng total RNA (200 ng/µl) was mixed with 5 pmol reverse primer in 5 µl of RT buffer and 40 U MMLV reverse transcriptase (Superscript II, Gibco). To reverse transcribe the RNA the reaction was incubated at 42°C for 1 h.

RT buffer	5 x First strand synthesis buffer (Gibco)	300 µl
	0.1 M DTT (Gibco)	150 µl
	10 mM dNTPs	75 µl
	dH2O	475 µl

5.19. RT-PCR

1/8 of a typical reverse transcription reaction (made with 400 ng of total RNA) was used to amplify cDNA. The reaction was made in 25 µl and contained 10 pmol of specific forward and reverse primers, 200 mM dNTPs, 1 x Taq polymerase buffer and 1 U Taq polymerase (Perkin Elmer). The conditions of the PCR cycles were dependent on the template to be amplified (see 3.6. for conditions of amplifying minigene products from *in vivo* splicing assays).

5.20. 5' RACE

5' RACE was performed on total RNA isolated from SHSY-5Y cells using the GeneRacer kit (Invitrogen) according to the manufacturer's protocol. RNA was dephosphorylated and decapped. Next, RNA oligonucleotide was ligated to the 5' end to provide forward priming sites. The RNA was reverse transcribed and two nested PCR amplifications were performed (Figure 8). The following

gene specific primers were used for the procedure: TrkB E9 R for the reverse transcription; TrkB E6 R for the first PCR amplification; TrkB E2 R for the second 'nested' PCR amplification. The products were cloned directly from the PCR reactions using TOPO cloning and were sequenced.

5.21. Coupled *in vitro* transcription and translation

Coupled *in vitro* transcription and translation was performed using the T7 Quick TnT kit (Promega) according to the manufacturers protocol. Untagged or flag-tagged pCDNA clones were used as templates.

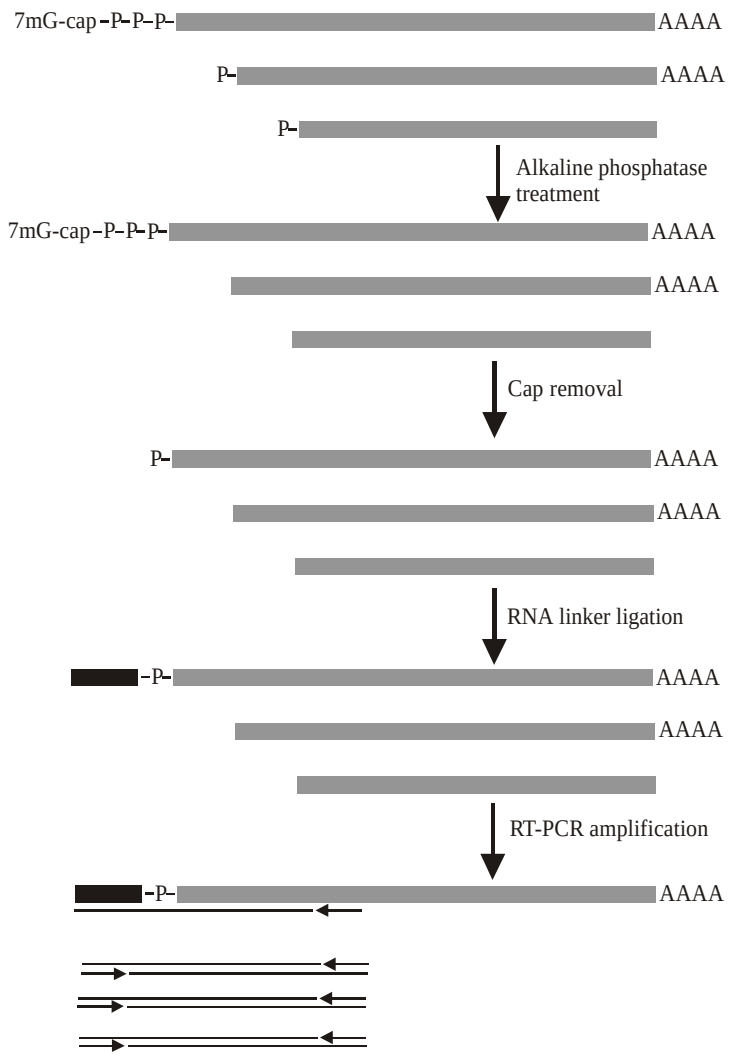


Figure 8. 5'-RACE. Initially the 5'-phosphate groups of the uncapped RNA molecules (ribosomal or other non mRNAs and degraded mRNAs) are removed by alkaline phosphatase treatment. Next the caps were removed using pirophosphatase and RNA linker (black box) is ligated to the 5'ends. The linker is not attached to the non mRNA or to the degraded mRNA molecules as they lack 5' phosphate group. The 5' ends of the RNA molecules that contain the ligated linker are amplified by PCR

5.22. In vivo splicing assay

To determine the influence of various proteins on the splicing of selected minigenes 1 to 2 µg of the minigene plasmid were transfected in eukaryotic cells together with an expression construct for the protein. A concentration dependent effect was usually assessed, in which case the protein was transfected in increasing amounts, in the range of 0.5 to 4 µg. To correct for the different quantity of plasmid in each transfection and possible promoter "quenching", empty vector was added in decreasing amount. Cells were plated in 6-well plates or 3 cm dishes and transfection was done after 24 h. After incubation for 24 hours at 3 % CO_2 total RNA was isolated from the cells (see 3.1.). 200 ng of RNA were used in a reverse transcription reaction (see 3.3.) with a reverse primer. This primer was specific for the vector in which the minigene was

cloned, to suppress reverse transcription of the endogenous RNA. A control reaction for specificity was included, with dH_2O instead of RNA. 1/8 of the reverse transcription reactions were used for PCR with minigene-specific primers. The primers were selected to amplify alternatively spliced minigene products. A control reaction with no template (RNA instead of cDNA) was included in the PCR. Half of the PCR reactions was resolved on 2 % agarose gel and the image was analysed by ImageQuant Analysis Software (Amersham/Pharmacia).

5.23. Pull down assays

The protocol was adapted from (Grabowski, 1990). 20 pmol of biotinylated RNA oligonucleotide immobilized on streptavidin coated metacrilate beads (Boehringer) were incubated overnight or for one hour with nuclear extract (10?g of protein) or 2?l of ^{35}S labeled *in vitro* translation mix and competitor RNA oligonucleotides in a total of 20?l RNA binding buffer (10mM HEPES pH7.5, 100mM KCl, 5?g/?l Heparin, 1.5?g/?l tRNA, 0.67?g/?l BSA, 0.1% Triton-X100, 20U RNase inhibitor). The beads were washed three times with RNA binding buffer and resuspended in SDS gel loading buffer. The bound protein was detected either using Western blot with TRA2-beta specific polyclonal sera or by autoradiography. The RNA oligonucleotide sequences were:

motif1 AAUGUUGAAGAAGGAAAAUGCGGAAGUCGdT
motif3 GUGGGGAAGAUGAAAGAAGUCAGAAUUUGdA
motif3-c2: GUGGGGAAGAUGCAAGAAGUCAGAAUUUGdA
motif3-t5: GUGGGGAAGAUGAAAUAAGUCAGAAUUUGdA
motif3-cons: GUGGGGAAGAUGCACGAAGUCAGAAUUUGdA
RNA1 ACAAAGCGUUCUACCAGCAGCCAGAUGCUG
RNA2 ACAAAGCGAAGAUGGUGCAGCCAGAUGCUG.

5.24. Freezing and thawing of cells

Monolayer cells were grown to mid logarythmic phase (about 75 % confluency) in 10 cm Petri dishes. They were collected by trypsinisation with 1 x Trypsin/EDTA (Gibco), resuspended in 90 % growth medium/10 % DMSO, 1 ml per 10 cm plate, aliquoted in tubes and frozen at -80^{o}C by decreasing the temperature by 1^{o}C/min. Cells were stored in liquid nitrogen.

Thawing of cells was performed at 37^{o}C. The entire content of the tube was transfered to a 10 cm dish. 15 ml of growth medium were added and the cells were maintained in an incubator at 37^{o}C, 5 % CO_2.

5.25. Plating cells

Cells were replated when reaching confluency. The monolayer was detached by addition of trypsin/EDTA and incubation at 37^{o}C until single cell suspension was formed. 1/5 to to 1/10 of this

suspension was transfered to a new dish and mixed with the corresponding growth medium. Cells were maintained in an incubator at 37°C, 5 % CO_2.

HEK 293 cells	10 % Fetal Calf Serum (Gibco)/90 % DMEM
SH-SY5Y cells	
NIH 3T3 cells	

5.26. Transfection of eukaryotic cells

5.26.1. Calcium phosphate precipitation

The procedure is based on the protocol of Chen and Okayama (Chen and Okayama, 1987) and was used with HEK-A293 cells. Exponentially growing cells were replated at a density of about 3 x 10^5 cells/8 cm^2. Growth medium was added and the cells were incubated at 37°C, 5 % CO_2 for about 24 h, to reach 60 - 70 % confluency. For most applictions cells were grown in 6-well plates, with 2 ml of growth medium per well. The transfection reaction for one well was made the following way. 1 to 5 µg of expression construct were mixed with 10 µl of 2.5 M $CaCl_2$ in 100 µl final volume. Equal volume of 2 x HBS buffer was added dropwise, with constant mixing. In order to form a precipitate, the solution was allowed to stay at RT for 20 min. After that, it was added to the growth medium. To express the transfected plasmid, cells were grown for additional 24 hours at 37°C, 3 % CO_2.

2 x HBS	280 mM NaCl
	10 mM KCL
	1.5 mM Na_2HPO_4 x 2 H_20
	12 mM Dextrose
	50 mM Hepes
	_____
	pH 6.95

5.26.2. Transfection by electroporation

Electroporation was performed on a Gel Pulser II Electroporation System (BioRad). 20 µg of expression construct were mixed with HeLa cells trypsinised from a 60-70 % confluent 10 cm

dish. Electroporation was done in 0.4 cm cuvette, at 250 V and 500 µF. After that, cells were plated on glass coverslips and grown at 37°C, 5 % CO_2.

5.27. Fixing cells on coverslips

Cells grown on coverslips were fixed in 4 % paraformaldehyde in 1 x PBS, pH 7.4 for 30 minutes at 4°C. The cells were washed twice with 1 x PBS and were permeabilised in 0.2 % Triton X-100 in PBS for 5 min. Cells were washed again in PBS prior to mounting on microscope slides with Gel/Mount (Biomeda). Slides were stored at 4°C.

5.28. Determination of protein concentration

Protein concentration was estimated according to the method of Bradford, using the BioRad Protein Assay Kit. Protein in the range of 5 to 20 µg in 800 µl of distilled H_2O was mixed with 200 µl of Bradford solution and incubated at RT for 5 min. Absorption of the solution was measured in a spectrophotometer at ? = 595 nm. The readings were compared to a control sample of 5, 10, 15 and 20 µg of BSA.

5.29. Isolation of nuclear protein extract from cultured eukaryotic cells

The protocol for nuclear extract preparation is from (Dignam et al., 1983). Cells from ten 15cm plates were harvested by scraping and washed twice with cold PBS. The pellet was washed with 5 volumes buffer A and after centrifugation resuspended in 3 volumes of buffer A. The cells were homogenized with Dounce homogenizer on ice. The nuclei were collected by centrifugation at 250g. The pellet was resuspended in buffer C and homogenized with tight homogenizer on ice. The homogenate was incubated for 30 minutes with stirring at 4°C and then centrifuged at 20,000g for 30 minutes. The supernatant was either dialyzed against buffer D or when the volume was small was diluted with 5 volumes of buffer D and concentrated by membrane filtration (Centricon).

Buffer A 10mHEPES/KOH pH7.5, 1.5mM $MgCl_2$, 10mM KCl, 0.5 mM DTT, 0.5 mM PMSF

Buffer C 20mM HEPES/KOH pH7.5, 420mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25 % glycerol, 0.5 mM DTT, 0.5 mM PMSF

Buffer D 20mM HEPES/KOH pH7.5, 100mM KCl, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT, 0.5 mM PMSF

5.30. Electrophoresis of proteins

Proteins were resolved on denaturing SDS polyacrylamide gels. Typically, the BioRad gel electrophoresis system was used, to make 10 cm/7.5 cm/0.5 cm gels. The separating gel was 7.5 -

15%, depending on the molecular weight of the proteins, and the stacking gel was 4%. The proteins were mixed with sample loading buffer, denatured at 96°C for 5 minutes and loaded on the gel. Electrophoresis was carried out at 100 V for 2 hours in SDS gel running buffer.

Separating gel(10 ml)	7.5%	12%	15%
dH_2O	4.85 ml	3.5 ml	2.5 ml
1.5 mM Tris-HCl, pH 8.8	2.5 ml	2.5 ml	2.5 ml
10% SDS	100 µl	100 µl	100 µl
30% Acrylamide/Bisacrylamide	2.5 ml	4 ml	5 ml
10% Ammonium persulfate	100 µl	100 µl	100 µl
TEMED	10 µl	10 µl	10 µl

Stacking gel(10 ml)

dH_2O 6.1 ml; 0.5 mM Tris-HCl, pH 6.8 2.5 ml; 10% SDS 100 µl;

30% Acrylamide/Bisacrylamide 1.3 ml;

10% Ammonium persulfate 100 µl; TEMED 10 µl

Sample loading buffer, 5x	225 mM Tris-HCl, pH 6.8
	50% Glycerin
	10% SDS
	2.5% ß-mercaptoethanol
	0.25% Bromophenolblue

SDS gel running buffer	14.4 g Glycine
	3 g Tris
	1 g SDS
	dH_2O to 1 l

5.31. Staining of protein gels

Coomassie staining was used to detect proteins in polyacrylamide gels. After electrophoresis, the gel was placed in staining solution (2. % Coomassie Brilliant Blue R250, 45% (v/v) Methanol, 10% (v/v) Acetic acid) for >2-3 hours at RT. After that the gel was washed 2-3 times for 30 minutes in 50% Methanol/10% Acetic acid and 2-3 more times in 20% Methanol/10% Acetic acid.

5.32. Western blotting

Proteins resolved on SDS polyacrylamide gels were transferred to nitrocellulose membrane (Schleicher & Schuell) in transfer buffer, for 45 minutes at 120 V. The membrane was blocked for 1 hour in 1 x NET-gelatine buffer (at 4°C, as well as all the subsequent prosedures in this protocol), after which a primary antibody was added and the incubation was allowed to proceed overnight. Then the membrane was washed three times for 20 minutes in 1 x NET-gelatine and incubated with a secondary antibody coupled to horseradish peroxidase for 1 h. The membrane was subsequently washed three times for 20 minutes in 1 x NET-gelatine and the bound antibodies were detected by the ECL system. Equal amounts of Solution 1 and Solution 2 were mixed and added to the membrane for 2 min. The membrane was then exposed to an X-ray film (Fuji SuperRX) and developed in an Amersham developing mashine.

Transfer buffer	192 mM Glycine
	25 mM Tris
	20 % Methanol
NET-gelatine	150 mM NaCl
	5 mM EDTA
	50 mM Tris-HCl, pH 7.5
	0.05 % Triton X-100
	0.25 % gelatine
Solution 1	4.5 mM Luminol
	4.3 mM p-Iodophenol
	100 mM Tris, pH 9.5

Solution 2	0.003 % H_2O_2
	100 mM Tris, pH 9.5

5.33. Immunoprecipitation of proteins

For HEK 293 cells, the plasmid/plasmids of interest were transfected using the calcium phosphate precipitation procedure. 24 hours after transfection cells were washed with PBS and lysed in RIPA buffer, 200 µl/8 cm^2. Incubation in RIPA buffer was done for 20 minutes at 4°C. The lysates were collected in Eppendorf tubes and cleared by centrifugation for 1 minutes at 12 000 rpm. The supernatant was diluted with 3 volumes of RIPA-rescue and an antibody recognizing the expressed protein was added to the lysate. It was incubated at 4°C with top-over-bottom rotation to allow for the protein binding. After 1 hour, 50 µl Protein A-Sepharose/Sepharose CL4B (1:1) were added to the solution and the incubation continued under the same conditions for > 6 h. The Sepharose beads were pelleted by centrifugation for 1 minutes at 1000 rpm in a microcentrifuge, after which they were washed three times with 500 µl HNTG buffer. Finally, 30 µl of sample loading buffer were added to the pellet and boiled for 5 minutes at 96°C to denature the proteins. The beads were then discarded by centrifugation and the supernatant loaded on SDS polyacrylamide gel. The resolved proteins on the gel were transfered to nitrocellulose membrane and analysed by Western blot.

Protein A-Sepharose/Sepharose CL4B was made as follows: Protein A-Sepharose beads were washed in 15 ml of distilled H_20 and pelleted at 500 rpm for 2 min. After a second wash with dH_20 equal volume of Sepharose CL4B was added and the beads were washed two more times in RIPA-rescue buffer and kept at 4°C.

RIPA buffer	1 % Nonidet P40
	1 % Na deoxycholate
	0.1 % SDS
	150 mM NaCl
	10 mM Na phosphate, pH 7.2
	2 mM EDTA
	5 mM ß-glycerolphosphate

freshly added	1 mM DTT
	1 mM PMSF
	2 µg/ml aprotinin
	100 U/ml benzonase

When protein phosphorylation was assayed, 50 mM NaF and 4 mM Na_3VO_4 were added to the buffer before use, to block dephosphorylation of the analysed proteins.

RIPA-rescue buffer	20 mM NaCl
	10 mM Na phosphate, pH 7.2
	5 mM ß-glycerolphosphate
freshly added	1 mM DTT
	1 mM PMSF
	2 µg/ml aprotinin
when needed	2 mM Na_3VO_4
	1 mM NaF

HNTG buffer	50 mM Hepes, pH 7.5
	150 mM NaCl
	10 % glycerol
	1 mM EDTA
	0.1 % Triton X-100
freshly added	1 mM PMSF
	2 µg/ml aprotinin
when needed	2 mM Na_3VO_4

100 mM NaF

The following antibodies were used most often for immunprecipitation:

anti-EGFP antibody (Boehringer)

anti-Flag Bio M2 antibody (Sigma)

5.34. Heterokaryon assay

Heterokaryon assay was performed with wild type and mutant Tra2 beta, to determine whether the proteins shuttle between the nucleus and the cytoplasm. The experiment determines whether a protein expressed in one type of cells can migrate to another. The protocol was according to (Pinol-Roma and Dreyfuss, 1991; Weighardt et al., 1995). A human cell line (HeLa(JW36) and a mouse cell line (NIH3T3) were used for the experiment. The nuclei of human and mouse cells are easily distinguishable by DAPI staining as mouse cells exibit a characteristic punctate pattern. NIH 3T3 cells were plated on coverslips at a density of 3 - 4 x 10^4 cells/cm^2 24 hours before fusion. 3 - 4 hours before fusion HeLa cells were trypsinised and were transfected by electroporation. They were subsequently plated onto the NIH3T3 cells at a ratio of about 1:1. After HeLa cells attached to the bottom of the dish, the coverslips were rinsed two times with pre-warmed PBS and placed upside-down on a drop of 50 % PEG 4000 (Gibco) in DPBS (without Ca++ and Mg++). The cells were incubated for 2 minutes, after which the coverslips were washed two times in PBS and placed back in growth medium for incubation to allow for the migration of the expressed protein. Cells were fixed in 4 % paraformaldehyde and the coverslips were mounted on microscope slides.

6. Tra2-beta

6.1. Introduction

Transformer-2 (TRA2) is a SR-like protein that was first discovered in *Drosophila* as part of the sex determination cascade that is regulated by alternative splicing (Baker, 1989). The predominant isoform of TRA-2, TRA-2^{264} binds together with TRA to the female specific exon of *dsx* and causes its inclusion into the mRNA. As a result, female flies lacking TRA-2^{264} develop into phenotypical males (Nagoshi et al., 1988; Watanabe, 1975). Two mammalian homologues of *Drosophila tra2* have been identified, *tra2-alpha* and *tra2-beta* (Beil et al., 1997; Dauwalder et al., 1996; Matsuo et al., 1995; Segade et al., 1996). All three proteins share a similar structure consisting of two RS-domains that flank a central RRM and regulate splice site selection in a concentration dependent manner. Mammalian *tra2-alpha* and *tra2-beta* can complement *tra2* loss in *Drosophila*. In both mammalian and *Drosophila* systems, additional variants of TRA-2 are generated by alternative splicing (Amrein et al., 1994; Mattox et al., 1990; Nayler et al., 1998a). Splicing events that regulate the usage of the first RS-domain are common to all systems. These events result in the proteins dTRA-2^{264}, dTRA-2^{226} and dTRA-2^{179} in *Drosophila*, and TRA2-beta1 and TRA2-beta3 in mammals, respectively. The concentration of TRA2 is tightly regulated. This is achieved by putative destabilizing elements in the 3' UTR (Beil et al., 1997; Segade et al., 1996) and regulation of tra2 alternative splicing. Posttranscriptional regulation is most likely responsible for the absence of TRA2-beta1 in lung, despite abundant mRNA expression (Nayler et al., 1998a).

The human *tra2-beta* gene consists of 10 exons, two of which are alternatively spliced (Figure 9). Through alternative splicing and usage of alternative promoters and polyadenylation sites, the human *tra2-beta* gene generates five different mRNAs. Two of these mRNAs are translated to give rise to the tra2-beta1 and tra2-beta3 proteins. Tra2-beta2 and tra2-beta4 mRNAs cannot be translated due to the presence of exon 2 which terminates the open reading frame. The conceptual translation of the tra2-beta5 mRNA yields the TRA2-beta3 protein. It remains to be determined whether this mRNA is translated into protein.

Regulation of tra2-beta splicing variants has been observed during epileptic episodes in the brain, where a change in concentration between the TRA2-beta1, TRA2-beta2 and TRA2-beta3 isoforms is observed (Daoud et al., 1999). The importance of regulating TRA2 concentration is highlighted by changes that mediate or indicate pathological states in models of ischemia (Matsuo et al., 1995), brain injury (Kiryou-Seo et al., 1998), atherosclerotic lesions (Tsukamoto et al., 2001) and inflammation (Segade et al., 1995; Segade et al., 1996). In *Drosophila*, proper concentration of

dTRA-2^{226}is required for normal spermatocyte development. dTRA-2^{226} concentration is controled by an autoregulatory loop where dTRA-2^{226} promotes inclusion of an alternative retained intron M1, that leads to formation of dTRA-2^{179}, a nonfunctional isoform lacking the first RS-domain (McGuffin et al., 1998). Similar to dTRA-2^{179}, mammalian TRA2-beta3 lacks the first RS-domain (Nayler et al., 1998a).

This work demonstrates that TRA2-beta3 protein is present in mammals and that the generation of its mRNA is regulated by phosphorylation of regulatory proteins. In contrast to TRA2-beta1, this isoform did not regulate alternative splicing patterns of a number of genes investigated. For example, it did not influence the splicing patterns of tra2-beta, survival motor neuron (SMN2) and amyloid precursor protein (APP) pre-mRNAs. TRA2-beta1 autoregulates its concentration by

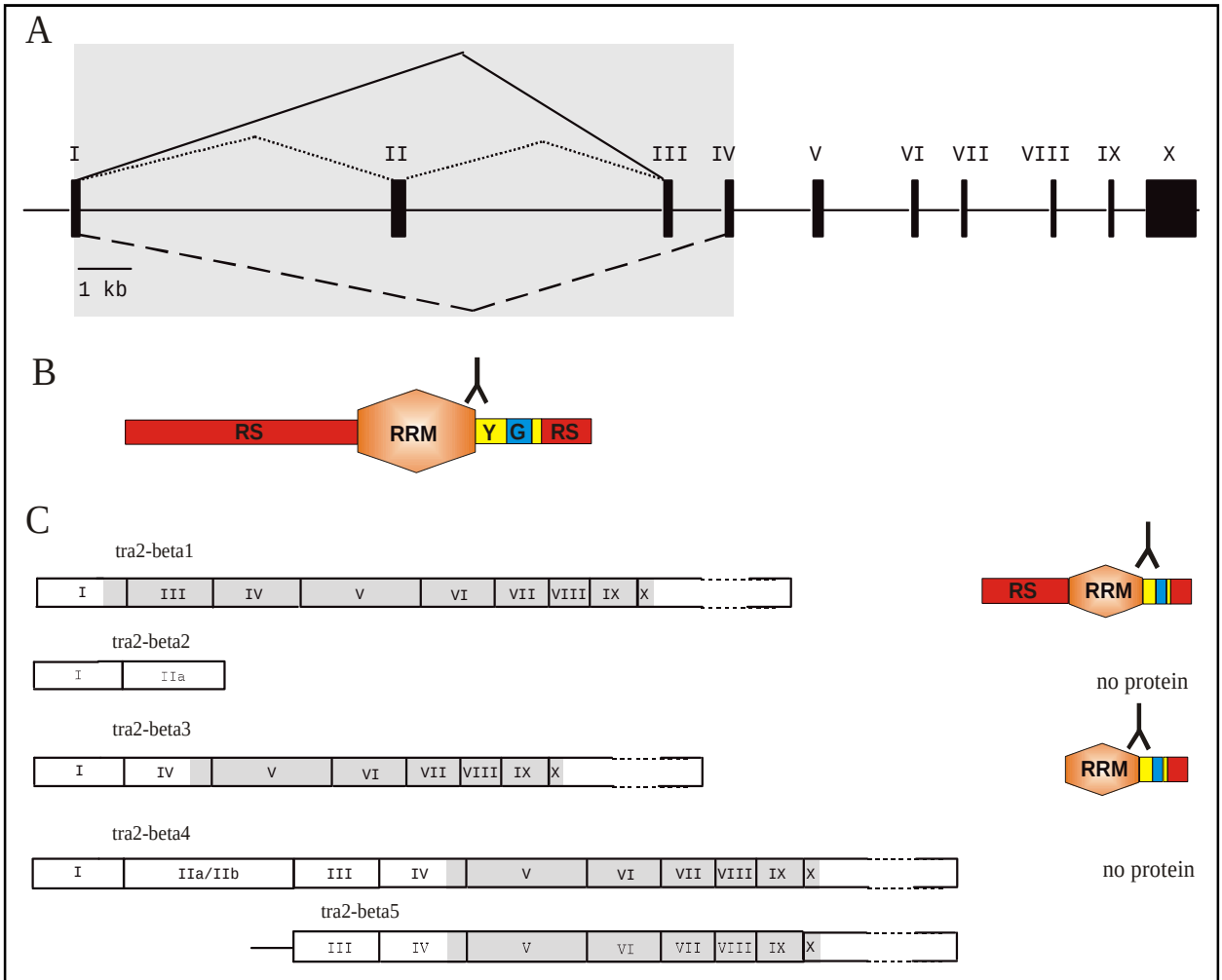


Figure 9. Tra2-beta gene structure. (A) Exon intron structure drawn to scale. Exons are shown as black boxes. Lines indicate the alternative splicing. The shaded region was cloned to create the tra2-beta minigene. (B) Structure of the protein. Tra2-beta1 protein consists of a RNA recognition motif (RRM), flanked by two SR repeats (red). The protein also has a tyrosine rich region (yellow) and glycine stretch (blue) located between the C-terminal SR repeat and the RRM. Position of the epitope in tra2-beta1 protein recognized by the pan-tra2 antiserum is shown on top. (C) Transcripts derived from the tra2-beta gene. Boxes indicate the individual exons. Shading shows the open reading frame. On the right are the proteins that are encoded by each of the transcripts. The position of the epitope in tra2-beta1 protein recognized by the pan-tra2 antiserum is shown on top.

binding to four exonic enhancers present in exon II of the *tra2-beta* pre-mRNA. This binding results in formation of a mRNA isoform that contains exon II, which is not translated into protein. Together, these data suggest that the concentration of TRA2-beta1 is maintained by an autoregulatory loop in which the protein causes the formation of a nonfunctional pre-mRNA isoform.

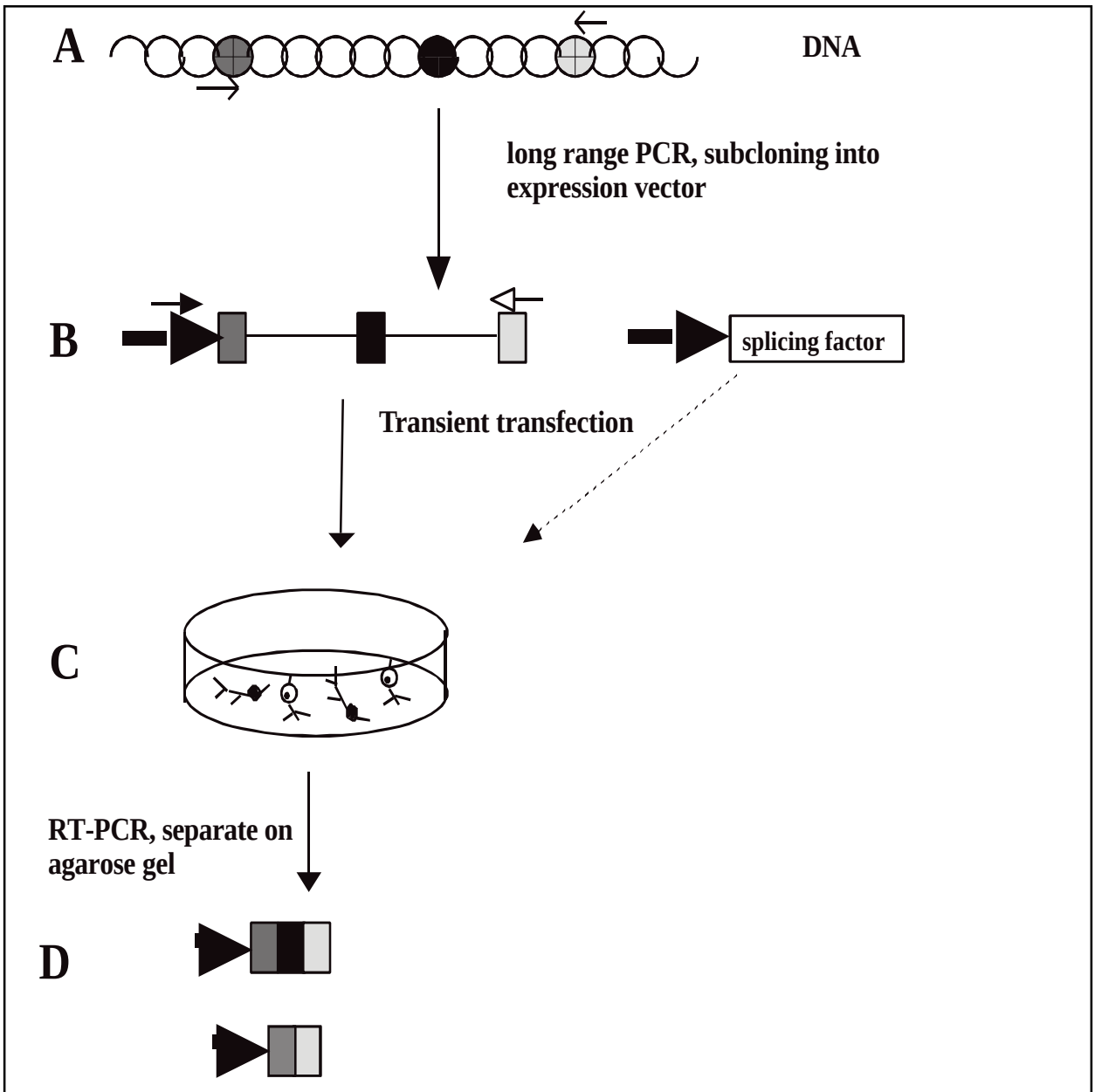


Figure 10. Overview of *in vivo* splicing analysis with the minigene approach. (A) Using long range PCR the alternatively spliced exon (black circle) and its flanking constitutive exons (shaded circles), as well as intergenic regions (open circles), are amplified from a genomic DNA clone. The restriction sites introduced by the PCR primers are indicated with a star and a box. (B) After subcloning into a suitable TOPO vector, the minigene is recloned into an eukaryotic expression vector using the unique restriction sites introduced by PCR. The eukaryotic promoter is indicated by a thick arrow. Exons are shown as boxes, introns as lines. After transfection, the splicing is analyzed by RT-PCR using an antisense primer against the downstream flanking exon (open arrow) and a sense primer against a vector derived sequence (closed arrow). (C). The minigene can be co-transfected with putative splicing factors. Alternatively it can be transfected into different cell types to analyze them for their splicing ability. (D) The PCR products can be separated by size or hybridization pattern, due to the presence or absence of the alternatively spliced exon.

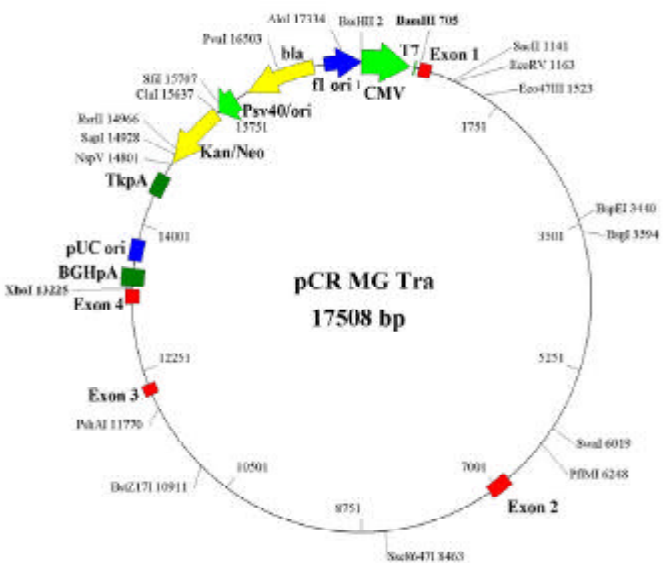
6.2. Results

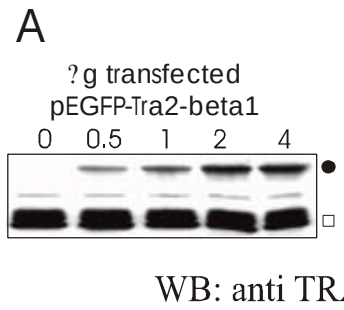
6.2.1. Minigene Construction and *in vivo* splicing assays

The tra2-beta pre-mRNA splicing was studied using the *in vivo* splicing assay approach (Stoss et al., 1999) (Figure 10). This approach involves transient cotransfection of a reporter minigene and one or more splicing regulators. The splicing pattern of the minigene is controlled using RT-PCR. The in vivo splicing assays are easy to perform and yield reproducible results. In contrast to the *in vitro* splicing assays there are no limitations in regard to the size or the complexity of the reporter minigene. They also do not require expression and purification of recombinant proteins, which in the case of proteins with long RS repeats is often difficult to achieve. The major drawback is that it is often difficult to distinguish whether the effect of the cotransfected splicing factor on the reporter minigene is direct or indirect.

A minigene containing the first four exons of the human *tra2-beta* gene was created (Figure 11), to study the factors and the cis-acting sequences that regulate the alternative splicing of the human tra2-beta pre-mRNA. The first four exons of the human *tra2-beta* gene together with the intervening introns were amplified by long range PCR, to construct the minigene. Consequently the PCR product was cloned in the pCR3.1 vector using BamHI and XhoI sites in the polylinker. The minigene splicing was studied by *in vivo* splicing assays. In a typical *in vivo* splicing assay the minigene is cotransfected in HEK A293 cells with increasing amounts of an expression vector carrying splicing factor cDNA. The splicing pattern was analyzed by RT-PCR 16 hours after the transfection. As seen from the control lane on Figure 13A three PCR products sized 318, 452 and 647 bp can be detected on the gel. They correspond to the three mRNAs derived from the alternative splicing of the exons 2 and 3. The major one migrating at 452 bp contains exons 1, 3 and 4. The 318 bp

Figure 11 Map of the Tra2-beta minigene. The tra2-beta genomic DNA is cloned in the BamHI and XhoI sites (bold). Exons are show as red boxes.





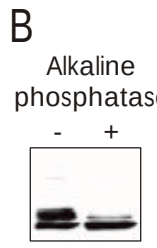


Figure 12. Tra2-beta1 protein. (A) Lysates of HEK A293 cells transfected with increasing amounts of tra2-beta1-EGFP fusion expression construct analyzed by western blot. The closed circle indicates the tra2-beta1-EGFP fusion and the open box shows the position of the endogenous protein. (B) Western blot of HeLa nuclear extract treated with alkaline phosphatase compared to mock treated extract. The phosphatase treatment eliminates the more slowly migrating bands

fragment contains exons 1 and 4. The 647 bp fragment contains all four exons. These isoforms correspond to the previously described mRNAs derived from the alternative splicing of exons 2 and 3 of the human and the rat *tra2-beta* genes, namely the tra2-beta1, tra2-beta3 and tra2-beta4 mRNAs. No product that contains exons 1, 2 and 4 but not exon 3 has been detected, which is also in agreement with the published data. Furthermore, the ratio between the different splicing isoforms is similar to the naturally occurring one. In conclusion, the splicing pattern of the minigene follows the one of the human *tra2-beta* gene. Therefore the tra2-beta minigene carries all the information that is necessary for its correct splicing.

6.2.2. Tra2-beta1 controls the splicing of its own pre-mRNA

As several splicing factors including the *Drosophila* tra2 (Chabot et al., 1997; Jumaa and Nielsen, 1997; Mattox and Baker, 1991) have been shown to regulate the splicing of their own pre-mRNAs, we first tested the effect of tra2-beta1 cDNA on the splicing of the tra2-beta minigene. In these experiments, constant amounts of the tra2-beta minigene was transfected with increasing amounts of vector expressing TRA2-beta1-EGFP fusion. The amount of the expressed protein was analyzed by western blot and was found to be lower than or comparable to the amount of the endogenous protein (Figure 12A). The multiple bands observed for the endogenous TRA2-beta1 are due to hyperphosphorylation (Figure 12B). Even the lowest amounts of tra2-beta1 cDNA tested induced dramatic increase in exon 2 inclusion, almost completely shutting off the synthesis of tra2-beta1 mRNA (Figure 13A). The exon 2 contains multiple stop codons that interrupt the open reading frame. It is expected that its inclusion will prevent either the export to the cytoplasm or the translation of the tra2-beta4 mRNA, creating an autoregulatory loop.

6.2.3. Tra2-beta1 regulates the splicing of the SMN2, APP and CLB minigenes

The next step in this work was to identify the sequences in the tra2-beta minigene that are recognized by the TRA2-beta1 protein. However, the large size of the minigene makes the deletion

Table 6 Minigenes that were tested for TRA2-beta1 dependent alternative splicing

Minigene	Description	Tra2-beta dependent change in splicing	Reference
pRSV-6789	Amyloid precursor protein (APP) exons 6,7,8 and 9	Skipping of exon 8	(Yamada et al., 1993)
CD44v5	CD44 exon v5	No effect	(Konig et al., 1998)
CLB	Clathrin light chain B (CLB) exons 3, 4 and 5	Skipping of exon 4	(Stamm et al., 1992)
RTB33.51	Chicken cardiac troponin T (cTNT) 30nt alternative exon	No effect	(Cooper, 1998; Ryan and Cooper, 1996)
RTB300	Human cardiac troponin T (cTNT) 30nt alternative exon	No effect	(Philips et al., 1998)
Insulin receptor	Human insulin receptor exons 10, 11 and 12	No effect	(Kosaki et al., 1998; Kosaki and Webster, 1993)
pSMN2	SMN2 exons 6, 7 and 8	Inclusion of exon 7	(Hofmann et al., 2000; Lorson and Androphy, 2000)
pXB	SRp20 (X16) exons 3, 4, 5 and 6)	No effect	(Jumaa and Nielsen, 1997)
tau-SV9/10L/11	Tau microfilament binding protein exons 9, 10 and 11	No effect	(Gao et al., 2000)
pSWAP-miniG2a	Human homologue of Drosophila suppressor of white apricot (SWAP) gene exons 2,2a and 3	No effect	(Denhez and Lafyatis, 1994)

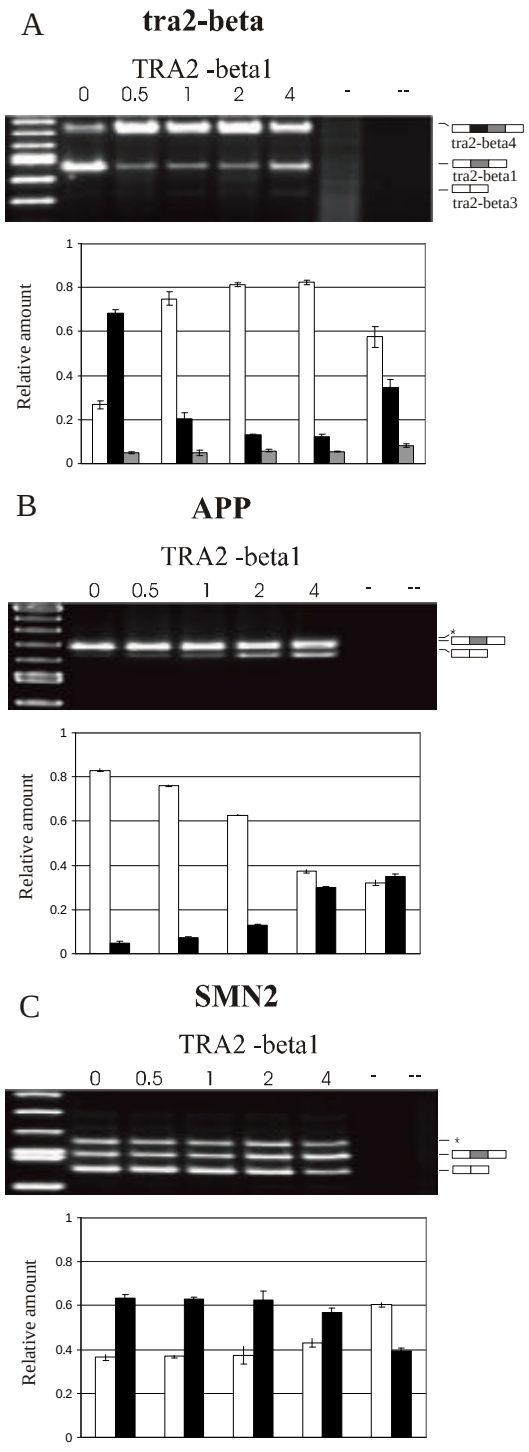


Figure 13. Alternative splicing regulated by the tra2-beta1 protein. Cotransfection of tra2-beta (A), APP (B) and SMN (C) minigenes with increasing amounts of tra-beta1 expression construct changes the utilization of the alternative exons. Exon composition of each of the bands is shown to the right. Asterisks indicate the positions of the heterodimers. Bar graphs show the quantified data from three independent experiments. The bars show the relative amount of each isoform as follows: A. Tra2-beta1 – black, tra2-beta3 – gray, tra2-beta4 white; B APP containing exon 9 – white, lacking exon 9 – black; C. SMN2 containing exon 7 – white, SMN2 lacking exon 7 – black.

analysis or the systematic mutagenesis difficult. The approach that was used here was to look for more pre-mRNAs that are targets for TRA2-beta1 and identify the common sequence motifs. Several minigenes were tested in *in vivo* splicing assays (Table 6) and on the splicing of two of them TRA2-beta1 had dose-dependent effect (Figure 13B and 13C). The CLB minigene has previously been reported to be a target for tra2-beta regulation (Stamm et al., 1992). TRA2-beta1 induced skipping of exon 8 of the APP minigene (Figure 13B) and inclusion of exon 7 of the SMN2 minigene (Figure 13C). It has been suggested that TRA2-beta1 binds purine rich sequences (Tacke et al., 1998). Using mutagenesis, a short GAR repeat has been identified as TRA2-beta1 dependent splicing enhancer for SMN2 exon 7 (Hofmann et al., 2000). In agreement with this the common sequence motifs that were identified in the tra2-beta, APP and CLB minigenes were 8nt GARR repeats (Figure 14A). Four such motifs were found in the tra2-beta exon 2 (Figure 14B). The affected alternative exons of the APP and CLB minigenes lacked the GARR repeats. Instead the motifs were located in the neighboring exons: one in exon 6, one in exon 7, two in exon 9 of the APP minigene and three in exon 3 and two in exon 5 of the CLB minigene (Figure 14A).

6.2.4. Four purine rich sequences in exon 2 are required for the Tra2-beta action

Site directed mutagenesis on the tra2-beta minigene was performed to verify the

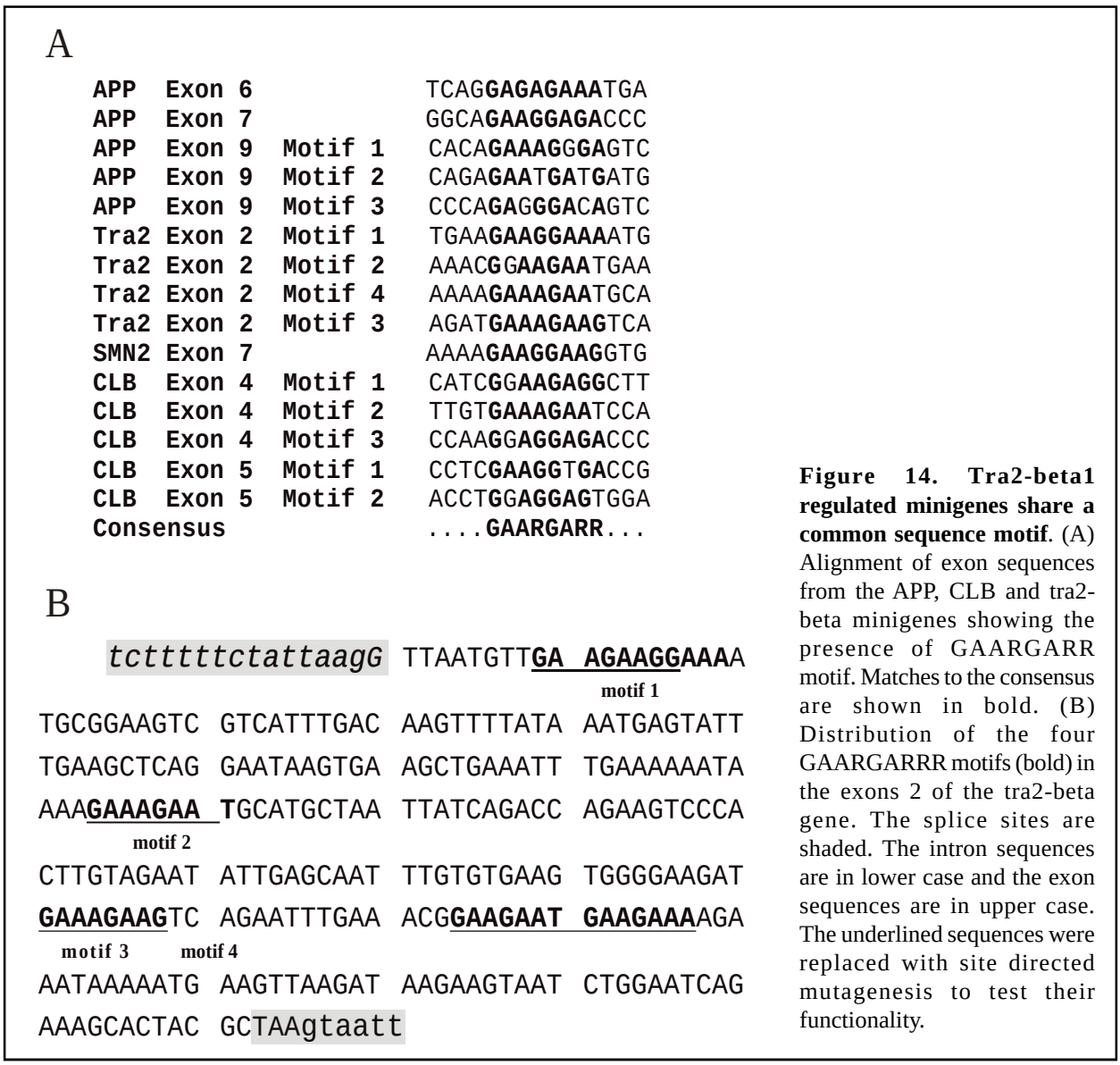


Figure 14. Tra2-beta1 regulated minigenes share a common sequence motif. (A) Alignment of exon sequences from the APP, CLB and tra2-beta minigenes showing the presence of GAARGARR motif. Matches to the consensus are shown in bold. (B) Distribution of the four GAARGARRR motifs (bold) in the exons 2 of the tra2-beta gene. The splice sites are shaded. The intron sequences are in lower case and the exon sequences are in upper case. The underlined sequences were replaced with site directed mutagenesis to test their functionality.

functionality of the four putative tra2-beta1 binding sites in the tra2-beta exon 2. Each of them was substituted with both purine and pyrimidine rich sequence (Figure 15A). Mutant minigenes were tested in cotransfection experiments with increasing concentration of TRA2-beta1. As shown in Figure 15B, the mutations decreased the inclusion of exon II with different efficiency ranging from two-fold decrease (motifs 1 and 2) to almost complete block (motifs 3 and 4). The decrease was independent of the sequence with which the site was substituted and a control mutation (Figure 15C) had no effect. Therefore the observed block of exon 2 inclusion is a result of disruption of an enhancer sequence rather than of an introduction of a splicing silencer. The substitution of motif 4 with purine rich sequence creates a new 5' splicing site, which is used concurrently to the original splicing site of exon 2. As a result, a fourth splicing product approximately 50nt shorter than the tra2-beta4 can be detected on the gel (Figure 15B).

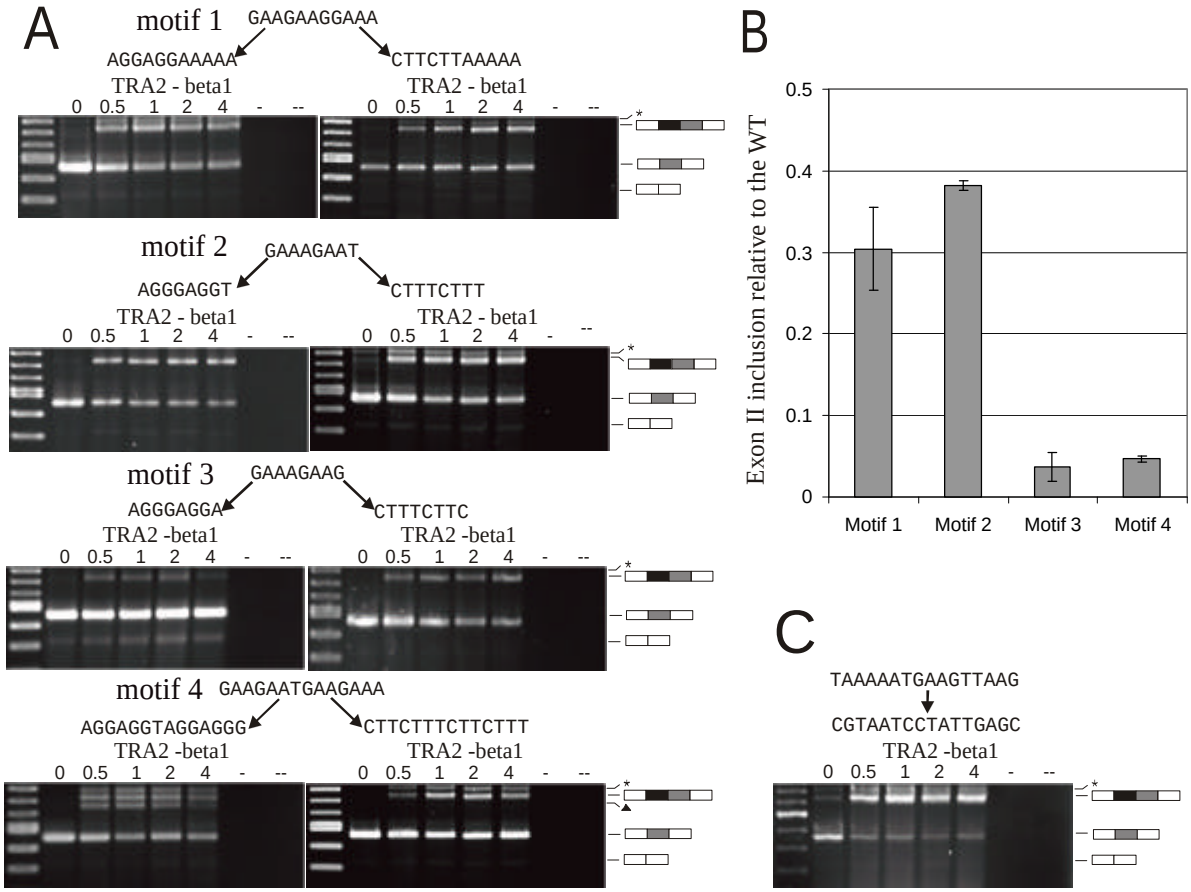


Figure 15. Site directed mutagenesis of the purine rich motifs results in decreased utilization of exon 2. (A) In vivo splicing assays using minigenes with substituted GAARGARR motifs in exon 2. Two substitutions of each motif were prepared (shown on top of each gel). Boxes on the right indicate the exon composition of each band. The asterisk shows the position of the heterodimers. The closed triangle shows the alternative splicing product that is produced using the splice cite introduced by motif 4 substitution for purine rich sequence. (B) Quantified data from three independent experiments showing the relative exon 2 inclusion for the purine rich mutants of each motif compared to the wild type. (C) Control substitution in exon 2 does not affect tra2-beta dependent inclusion.

6.2.5. Tra2-beta1 binds to the purine rich sequences in a sequence dependent way

To determine whether tra2-beta 1 interacts with the identified motifs directly a set of *in vitro* binding experiments using biotinylated RNA oligonucleotides were performed. The experimental layout is shown on Figure 16. In the first experiment, biotinylated oligonucleotide corresponding to the third binding site were used to pull down TRA2-beta1 from nuclear extracts. The binding was competed with non-biotinylated oligonucleotides for binding sites one and three and two unrelated RNA oligonucleotides of the same length. The bound protein was analyzed by western blot. The interaction is easily competed by non-biotinylated oligonucleotides containing binding sites one and three while the unrelated oligonucleotides RNA1 and RNA2 have little or no effect (Figure 17). Interestingly only the hypophosphorylated isoform of TRA2-beta1 was found to bind efficiently to

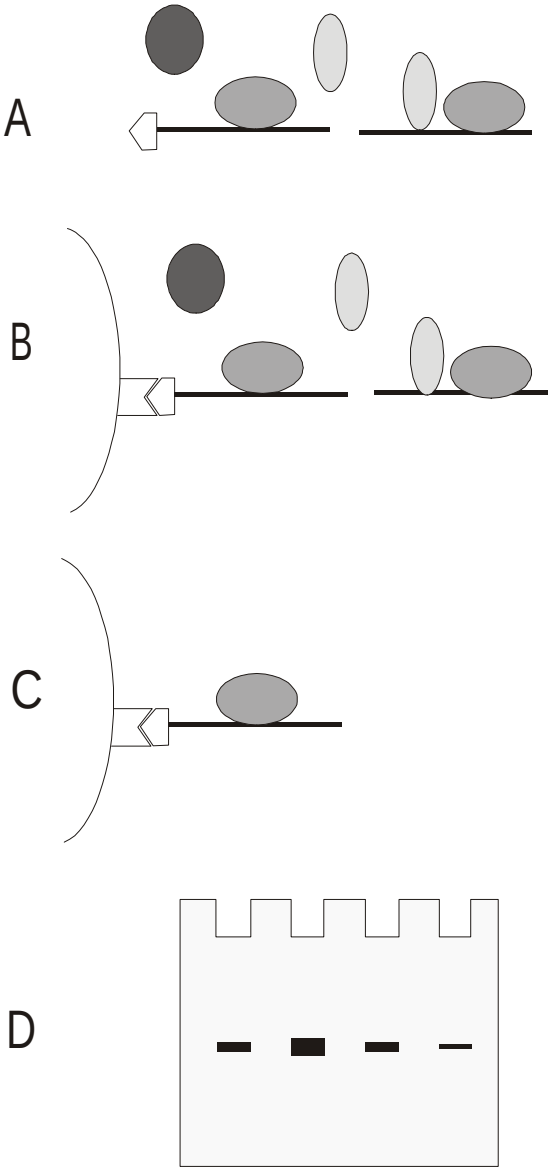


Figure 16. Pull down assay. (A) Biotinylated RNA oligonucleotide (left) is incubated with protein extract or *in vitro* translated protein (shaded circles) in the presence of non-biotinylated competitor (right). (B) The biotinylated RNA oligonucleotide is bound to strepatvidine coated resin. (C) The resin is washed to remove the nonbiotinylated competitor and the unbound protein. (D) The bound protein is analyzed on gel.

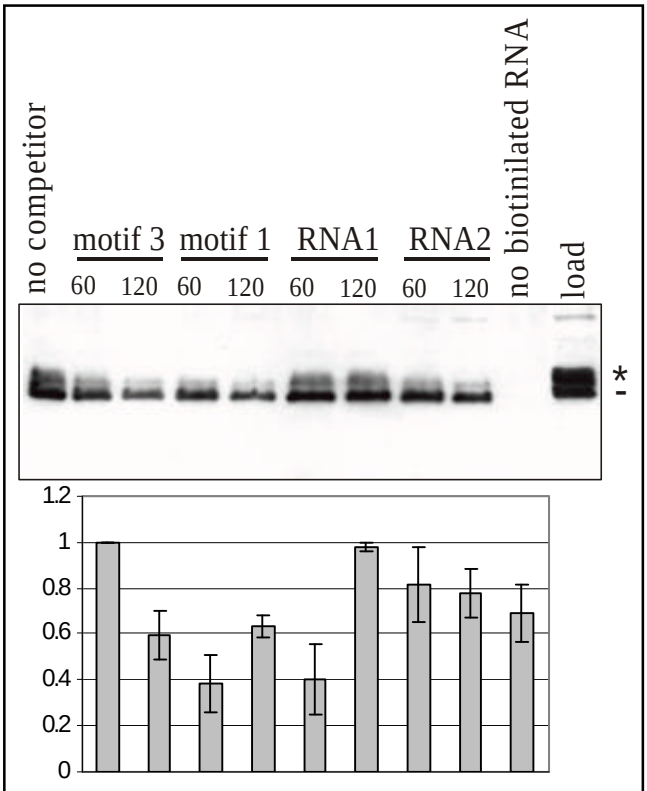


Figure 17. Pull down of tra2-beta1 from HeLa nuclear extract with biotinilated RNA oligonucleotide corresponding to motif 1. The binding was competed with non-biotinilated RNA oligonucleotides corresponding to binding sites 1 and 3 and two unrelated RNA oligonucleotides of the same size. The (-) indicates negative control in which the biotinilated RNA oligonucleotide was not included in the reaction. The asterisk indicates the hyperphosphorylated TRA2-beta1 protein. The bound TRA2-beta1 protein was detected using western blot.

the RNA oligonucleotide. To further prove that the interaction between TRA2-beta1 and the proposed binding sites is direct, pull down experiments were performed using *in vitro* translated, ^{35}S labeled protein. As in the experiments with nuclear extracts, the *in vitro* translated TRA2-beta1 was shown to bind to RNA in a sequence specific manner (Figure 18). Pull down was also performed using an *in vitro* translated deletion variant of tra2-beta that contains only the RNA recognition motif (RRM). Again, the binding was efficient and sequence specific (Figure 19). Therefore the RRM alone is sufficient for the RNA binding and the positively charged SR repeats are not required for the TRA2-beta1 interaction with RNA.

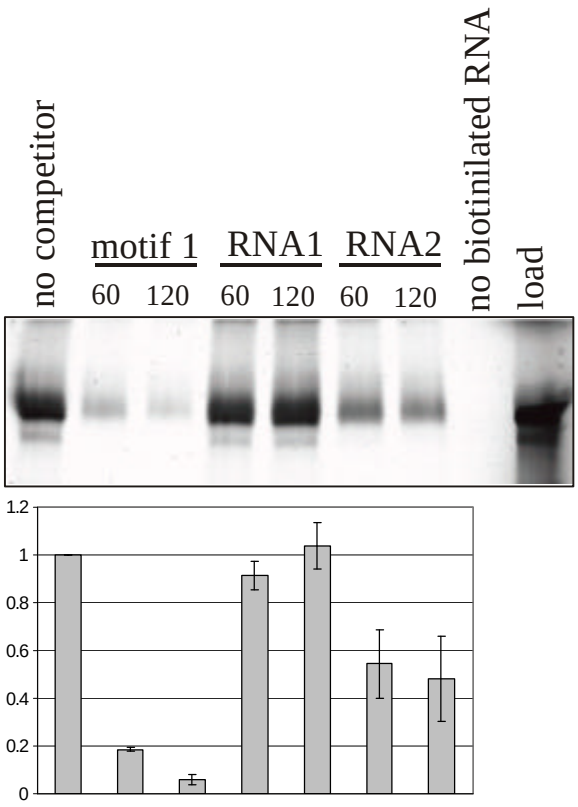


Figure 18. Pull down of *in vitro* translated ^{35}S labeled TRA2-beta1 protein with biotinylated RNA oligonucleotide corresponding to motif 1. The binding was competed with non-biotinylated RNA oligonucleotides corresponding to binding sites 1 and 3 and two unrelated RNA oligonucleotides of the same size. The (-) indicates negative control in which the biotinylated RNA oligonucleotide was not included in the reaction. The bound TRA2-beta1 protein was detected using autoradiography.

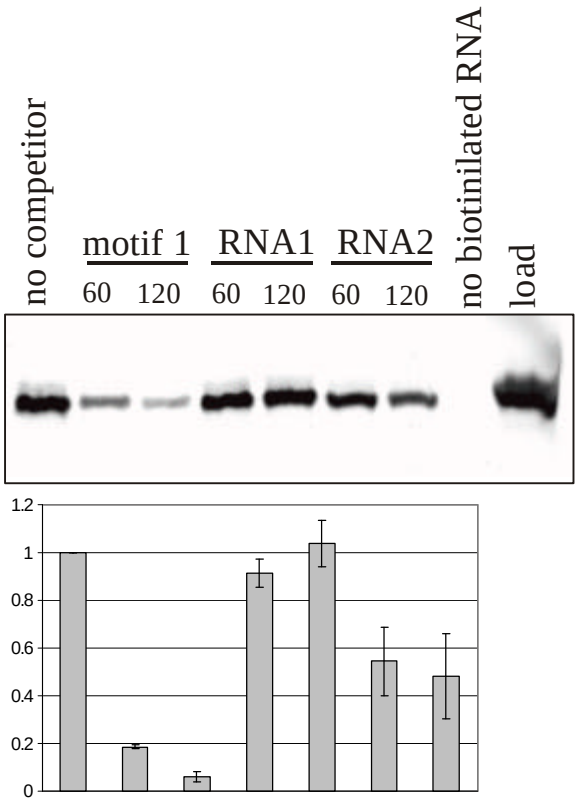
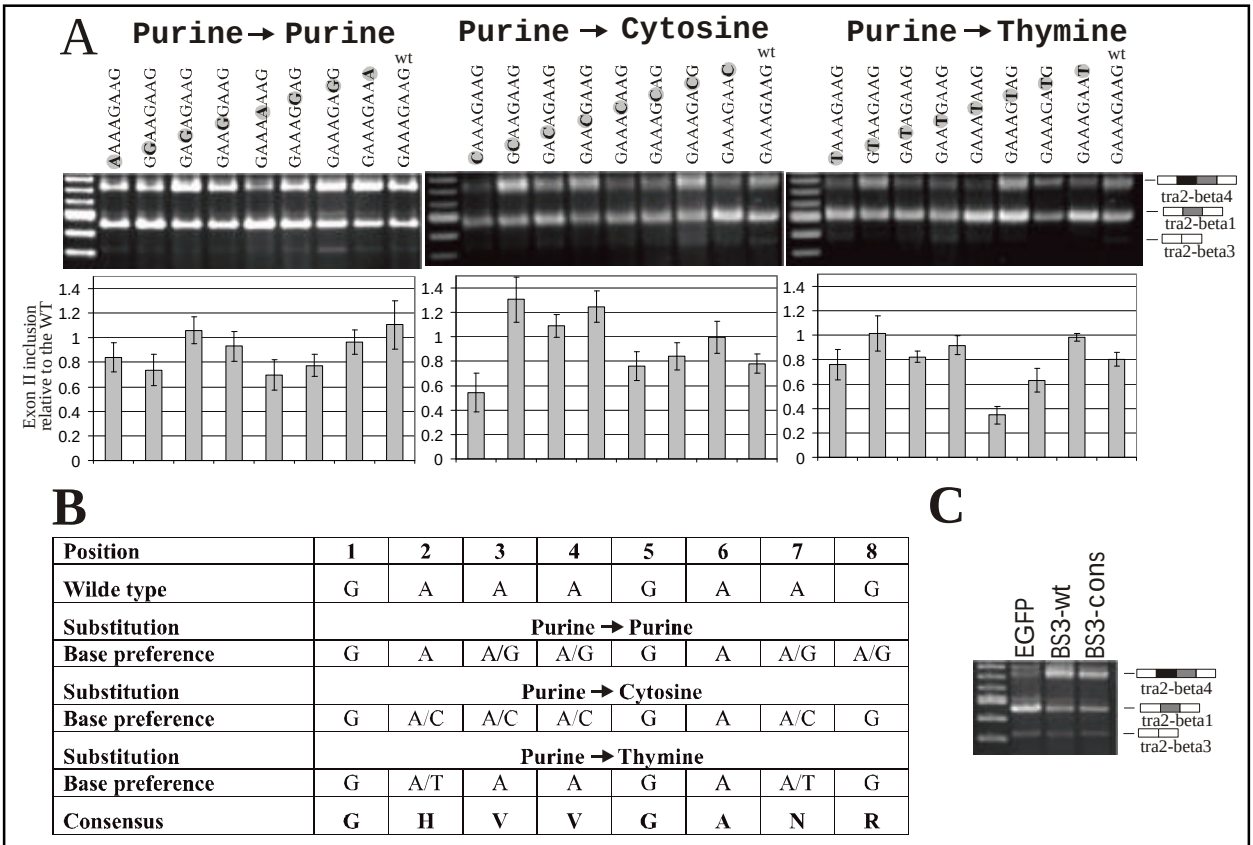


Figure 19. Pull down of *in vitro* translated ^{35}S labeled TRA2-beta RRM with biotinylated RNA oligonucleotide corresponding to motif 1. The binding was competed with non-biotinylated RNA oligonucleotides corresponding to binding sites 1 and 3 and two unrelated RNA oligonucleotides of the same size. The (-) indicates negative control in which the biotinylated RNA oligonucleotide was not included in the reaction. The bound TRA2-beta RRM was detected using autoradiography.

6.2.6. Single nucleotide mutagenesis of motif 3 reveals a highly degenerate TRA2-beta dependent enhancer

To derive a functional consensus binding site for TRA2-beta1 protein, single nucleotide substitutions were introduced in motif 3 in exon2 of the tra2-beta minigene. At each position the purine was replaced with the other purine, cytosine or thymine. Each mutant minigene was then analyzed in the presence of the same amount of TRA2-beta1 expressing cDNA and the ratio of exon II inclusion vs. exclusion for every mutant was compared to the wild type. Most of the mutations had either no effect or lowered the inclusion of exon 2 (Figure 20A). The two guanines at positions one and five and the adenine at position six were the most conservative. All substitutions of these nucleotides resulted in a significant decrease of exon inclusion. Surprisingly, two of the mutations gave stronger inclusion of exon 2 than the wild type. These mutations were cytosine for adenine replacements at position 2 and 4. However substituting both these positions with cytosine does not

have a cumulative effect and the inclusion of exon 2 is similar to the wild type (Figure 20C). Figure 20B summarizes the data from the single nucleotide mutagenesis of binding site 3. Nucleotides were assumed to be preferred for a certain position if they gave similar or higher inclusion of exon 2 compared to the wild type. Based on this assumption the following consensus has been derived: GMVVGANR, with only three conserved residues at positions one, five and six and preference for purine at position eight.



Position	1	2	3	4	5	6	7	8
Wilde type	G	A	A	A	G	A	A	G
Substitution	Purine → Purine							
Base preference	G	A	A/G	A/G	G	A	A/G	A/G
Substitution	Purine → Cytosine							
Base preference	G	A/C	A/C	A/C	G	A	A/C	G
Substitution	Purine → Thymine							
Base preference	G	A/T	A	A	G	A	A/T	G
Consensus	**G**	**H**	**V**	**V**	**G**	**A**	**N**	**R**

Figure 19. Single nucleotide mutagenesis of tre2-beta binding site 3 of tra2-beta exon 2. (A) In vivo splicing assays with minigenes carrying single nucleotide substitutions in binding site 3 cotransfected with 1??g tra2-beta1 expression construct. The quantified data from three independent experiments is shown under the gels. (B) Table showing the base preference for each position of the binding site. (C) The wildtype minigene was contransfected with 1??g pEGFPC2 (lane 1) or tra2-beta1 expression construct (lane 2). Lane 3: The minigene carrying the consensus binding site instead of binding site 3 was cotransfected with 1??g tra2-beta1 expression construct. The exon 2 inclusion levels are indistinguishable from the wild type.

6.2.7. Tra2-beta binds *in vitro* to the enhancer derived by mutagenesis

Pull down experiments were performed to verify that the results from the single nucleotide mutagenesis experiment reflect the binding affinity of TRA2-beta1 to the third binding site in exon 2. In these experiments the binding of *in vitro* translated TRA2-beta1 and TRA2-beta RRM to the wild type binding site 3 was competed with RNA oligonucleotides carrying mutations in the same binding site. Three oligonucleotides were used: one with cytosine to adenine substitution at position 2, one with thymine for guanine substitution at position 5 and one with two cytosine to adenine

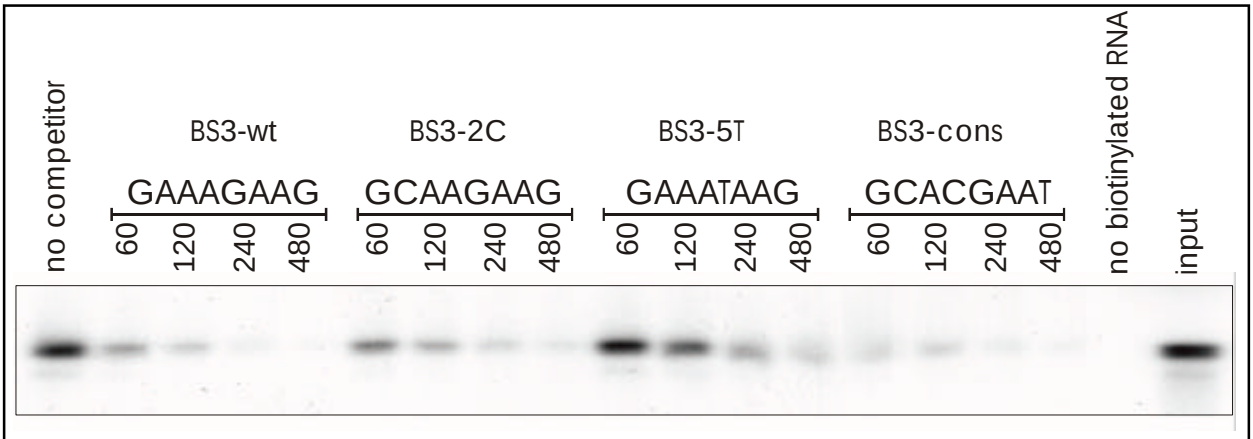


Figure 21. Pull down assays showing the specificity of TRA2-beta1 binding to binding site 3. 20pmol biotinylated RNA oligonucleotide corresponding to binding site 3 was used to pull down ^{35}S labeled in vitro translated TRA2-beta1 protein. The binding was competed with non-biotinylated RNA oligonucleotides corresponding to the wild type binding site 3 or to binding site 3 carrying single or double nucleotide substitutions. The amount of competitor in pmol is indicated on top of the gel.

substitutions at positions 2 and 4. These three oligonucleotides BS3-2C, BS3-5T and BS3-cons correspond to a mutation that gives better than the wild type inclusion (A to C substitution at position 2), worse than the wild type inclusion (G to T substitution at position 5) and the derived consensus (A to C substitutions at positions 2 and 4), respectively. The oligonucleotides carrying the A to C substitutions (BS3-2C and BS3-cons) competed the binding of both full length TRA2-beta1 (Figure 21) and the TRA2-beta1 RRM alone (Figure 22) to the binding site 3 with efficiency comparable to the wild type. The oligonucleotide with G to T substitution at position 5 was inefficient in competing the binding to the wild type binding site (Figures 21 and 22). Therefore the efficiency of the different variants of the TRA2-beta1 binding site as tra2-beta dependent enhancers directly correlates with the affinity of TRA2-beta1 binding.

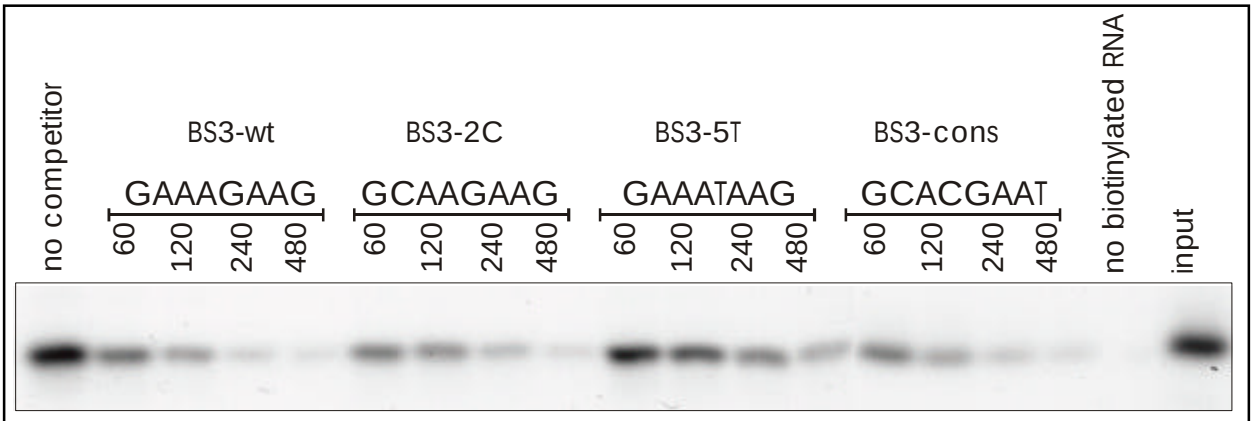
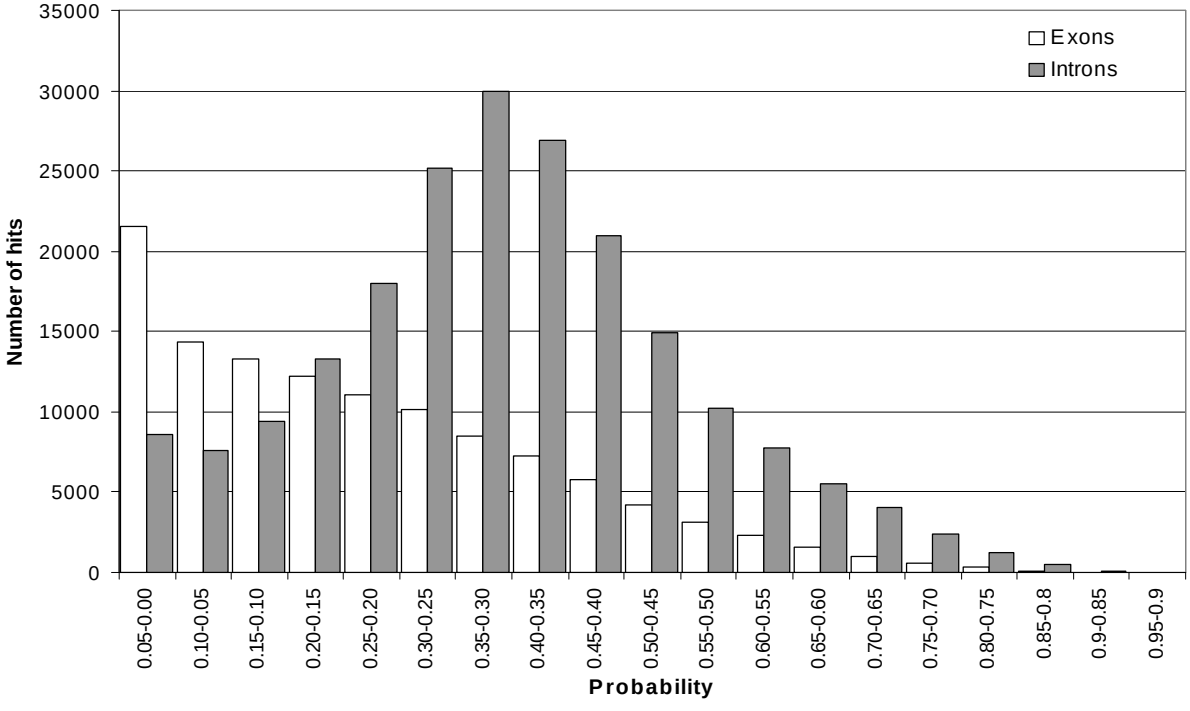


Figure 22. Pull down assays showing the specificity of TRA2-beta RRM binding to binding site 3. 20pmol biotinylated RNA oligonucleotide corresponding to binding site 3 was used to pull down ^{35}S labeled in vitro translated TRA2-beta RRM. The binding was competed with non-biotinylated RNA oligonucleotides corresponding to the wild type binding site 3 or to binding site 3 carrying single or double nucleotide substitutions. The amount of competitor in pmol is indicated on top of the gel.

6.2.8. The Tra2-beta consensus binding site is an exonic hallmark

As can be seen from Figure 20B the consensus (GMVVGANR) that has been derived from the mutagenesis experiment is strongly degenerated. Such degeneracy is typical for the binding sites of most of the splicing factors. Another feature of the splicing factors binding sites is their uneven distribution between the exons and the introns (Lui et al., 1998). To test whether TRA2-beta1 binding sites are preferentially located within the exons, the working draft of the human genome was screened with the consensus sequence. The human genome was downloaded from the NCBI site (ftp:// ftp.ncbi.nlm.nih.gov) and two datasets were extracted from it. One contained the introns and consisted of 252,288 sequences with total length of approximately 1.18 Gbp. The second, exonic dataset contained 280,064 sequences totaling approximately 42 Mbp. The two datasets were then screened with the pattern. Whenever a sequence contained a pattern match the probability of chance occurrence of the matches within it was determined. Figure 23 represents the probability distribution of the sequences containing the matches. The intron sequences (gray bars) peak at probability for random match 0.35 to 0.30. This roughly corresponds to the number of matches actually found equaling the number of matches expected to occur by chance. In contrast, the distribution of the exon sequences (white bars) is biased towards low probabilities of chance occurrence of the hits. 12.8% of all exons contain highly significant hits (P<0.1) that are unlikely to have occurred by chance, compared to 6.4% of all introns (Table 7). These exons concentrate approximately 27% of all hits in contrast to

Figure 23. Probability distribution of the hits to the TRA2-beta1 binding site in the exons (white bars) and the introns (gray bars) of the human genome.



only around 2% of the intronic hits (Table 7). Therefore, the TRA2-beta1 consensus binding site is preferentially found within exon sequences and is probably clustered there.

Table 7. Distribution of tra2-beta1 binding sites between the exons and the introns in the human genome.

	Exons	Introns	Exon Hits	Intron Hits
Total number	280064	252288	219212	3023488
Number with p<0.1	35904	16112	58824	68448
Percent	12.82%	6.39%	26.83%	2.26%

6.2.9. Clk kinases suppress exon 2 and exon 3 inclusion

The pull down experiments with biotinylated oligonucleotides have shown that the hyperphosphorylated form of TRA2-beta1 does not bind efficiently to RNA (Figure 17). This led to the prediction that if TRA2-beta1 is phosphorylated *in vivo*, it will prevent tra2-beta exon 2 inclusion. The clk kinases have been shown to bind and phosphorylate SR repeat containing proteins (Nayler et al., 1997). In addition, clk2 co-immunoprecipitates with TRA2-beta1. Therefore, clk kinases and particularly clk2, could block tra2-beta exon 2 inclusion when cotransfected with the tra2-beta minigene. As expected clk1 and clk2 kinases prevented tra2-beta1 exon 2 incusion in *in vivo* splicing assays (Figure 24). When higher amounts of the expression vector were transfected clk2 blocked both exon 2 and exon 3 inclusion in the mRNA (Figure 24). This leads to the synthesis of the brain

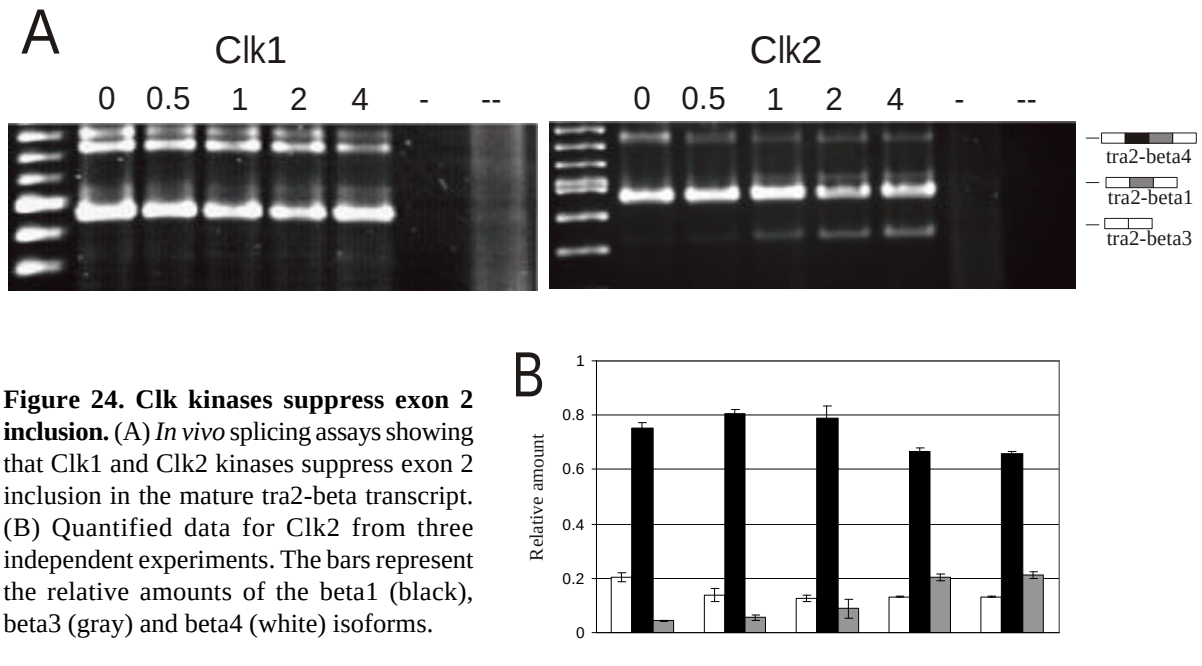


Figure 24. Clk kinases suppress exon 2 inclusion. (A) *In vivo* splicing assays showing that Clk1 and Clk2 kinases suppress exon 2 inclusion in the mature tra2-beta transcript. (B) Quantified data for Clk2 from three independent experiments. The bars represent the relative amounts of the beta1 (black), beta3 (gray) and beta4 (white) isoforms.

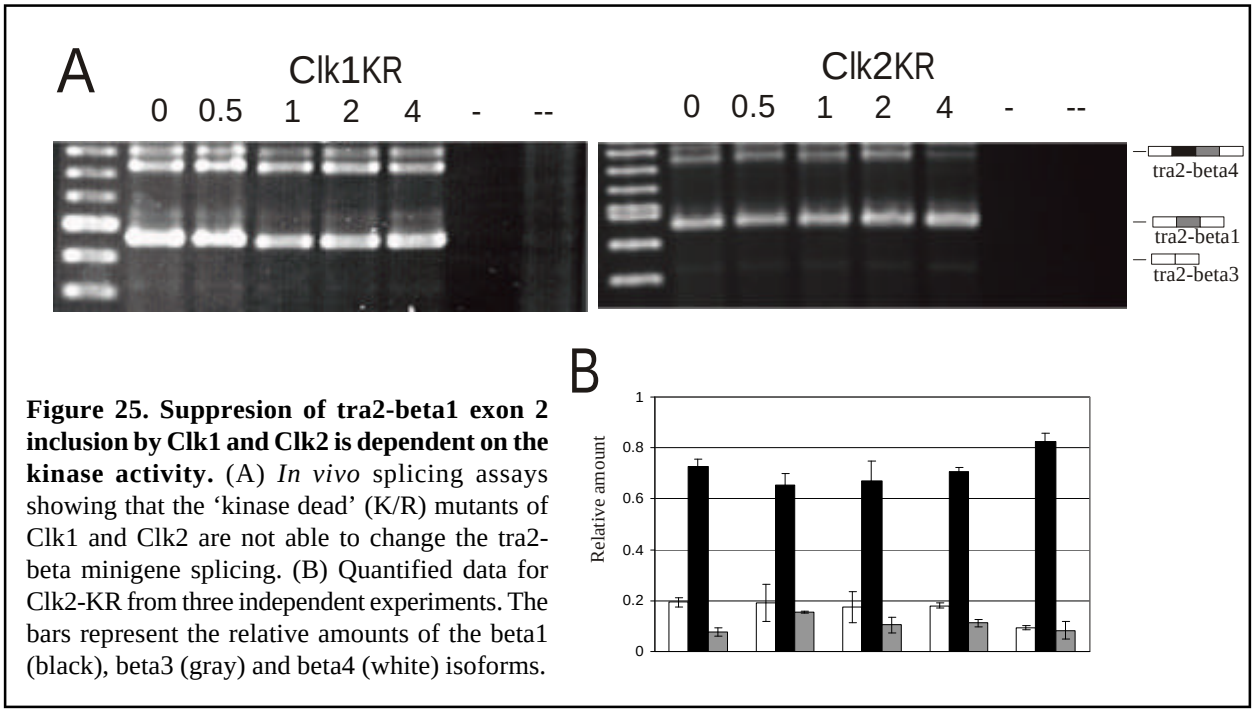


Figure 25. Suppresion of tra2-beta1 exon 2 inclusion by Clk1 and Clk2 is dependent on the kinase activity. (A) *In vivo* splicing assays showing that the 'kinase dead' (K/R) mutants of Clk1 and Clk2 are not able to change the tra2-beta minigene splicing. (B) Quantified data for Clk2-KR from three independent experiments. The bars represent the relative amounts of the beta1 (black), beta3 (gray) and beta4 (white) isoforms.

specific tra2-beta3 mRNA. Tra2-beta3 codes for a truncated protein that lacks the first SR domain. To corroborate this finding, *in vivo* splicing assays were performed with the inactive variants of the kinases (Figure 25). The splicing pattern of the tra2-beta minigene remained unchanged in these experiments, indicating that the effect of the kinases is due to their catalytic activity. Furthermore, when *in vivo* splicing assay using tra2-beta1 cDNA was performed in clk2 background (Figure 26), strong inhibition of the TRA2-beta1 dependent exon 2 inclusion was observed. These findings suggest that modulation of the TRA2-beta1 binding to mRNA through protein phosphorylation can regulate alternative splicing.

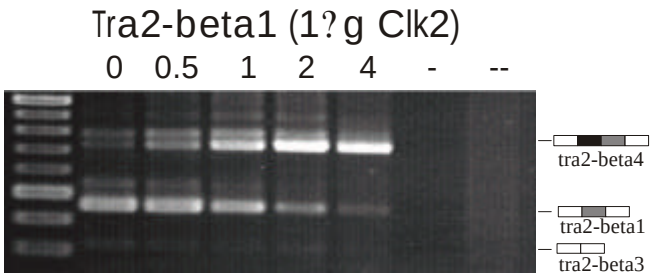
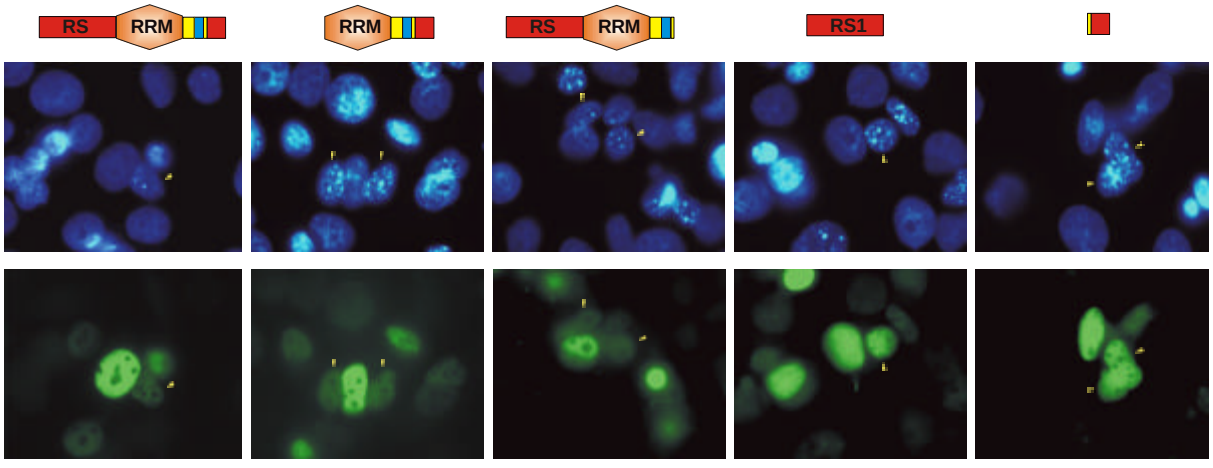


Figure 26. Tra2-beta1 effect on exon 2 inclusion is significantly suppressed in Clk2 background.

6.2.10. Tra2-beta1 shuttles between the nucleus and the cytoplasm, but Clk2 does not relocate it to the cytoplasm

Protein phosphorylation has been shown to induce relocation of hnRNP A (van der Houven van Oordt et al., 2000) and SF2/ASF from the nucleus to the cytoplasm (Caceres et al., 1998). In the case of hnRNP A, this relocation results in the change of the splicing pattern of the E1A reporter minigene. To test whether TRA2-beta1 shuttles between the nucleus and the cytoplasm, heterokaryon assays (Pinol-Roma and Dreyfuss, 1991; Weighardt et al., 1995) were performed. In these assays,



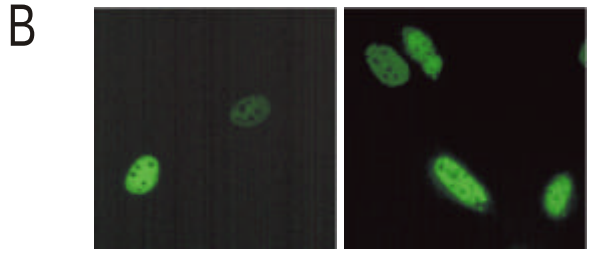


Figure 27. Nucleo-cytoplasmic shuttling of tra2-beta. (A) Heterokaryon assays. Top: DAPI staining; Bottom EGFP fluorescence. The full length tra2-beta1 protein as well as several deletion variant (domain structure is shown on top) were tested in heterokaryon assays and were found to be able to relocate to the mouse nucleus (indicated by yellow arrow). (B) EGFP-taged TRA2-beta1 protein does not relocate to the cytoplasm in cells transfected with Clk2.

HeLa cells (human origin) transiently transfected with expression clones for EGFP tagged proteins were fused with NIH 3T3 cells (mouse origin), in the presence of cycloheximide, to block the protein synthesis. After several hours, the fusions were examined for the presence of the fluorescent tag in the mouse nucleus. The human and mouse nuclei were distinguished based on their morphology after DAPI staining. As seen from Figure 27A TRA2-beta1 relocates from the human to the mouse nucleus. Several deletion variants of tra2-beta1 were also tested to map the region that is responsible for the shuttling. As seen from Figure 27A, all of them including the individual SR repeats shuttle in this assay. This suggests that TRA2-beta1 shuttling does not require RNA binding as has been shown for SF2 (Caceres et al., 1998). Next the TRA2-beta1 celular localization was examined after cotransfection with Clk2 kinase. No accumulation of the EGFP tagged TRA2-beta1 was observed regardless of the presence or absence of Clk2 (Figure 27B). Therefore the effect of Clk2 on the splicing of the tra2-beta minigene is due solely to the poor RNA binding ability of the phosphorylated TRA2-beta1 protein.

6.2.11. Tra2-beta interactors suppress exon 2 inclusion

In order to identify the factors that regulate tra2-beta1 splicing a number of splicing factors were tested in *in vivo* splicing assays with the tra2-beta minigene. Table 8 lists the tested factors. Of all the factors that have been tested, only TRA2-alpha promotes exon 2 inclusion (Figure 28A).

Table 8. Summary of the effect of the tested slicing factors on tra2-beta minigene splicing

Splicing factor	Effect on the tra2-beta minigene	Effect is dependent on RNA binding	Interacts with tra2-beta1 protein
SF1	No effect	NA	
SF2	Exon 2 skipping	No	Yes
SRp20	No effect	NA	No
SC35	No effect	NA	Yes
SRp30c	Exon 2 skipping	No	Yes
SRp55	No effect	NA	No
SRp75	No effect	NA	No
hnRNP G	Exon 2 skipping	No	Yes
hnRNP A	No effect	NA	No
hnRNP L	No effect	NA	No
SWAP	No effect	NA	No
YT 521B	Exon 2 skipping	NA	Yes
SAF-B	Exon 2 skipping	No	Yes
rSLM2	Exon 2 skipping	Yes	No
Tra2-alpha	Exon 2 inclusion	ND*	ND*
Tra2-beta1	Exon 2 inclusion	Yes	Yes
Znf265	Exon 2 skipping	Yes	ND

*Although the interaction between tra2-alpha and tra2-beta and the RRM dependency of the tra2-alpha effect have not been tested directly, due to the virtual identity of the two proteins it is likely that tra2-alpha behaves like tra2-beta in these assays

TRA2-alpha is a close homologue to TRA2-beta, with more than 90% identity. Therefore hey may act through the same mechanism on the tra2-beta pre-mRNA splicing. This hypothesis was tested by using tra minigene in which the third TRA2-beta1 binding site was mutated. As expected the tra2-alpha exon 2 inclusion was significantly lower compared to the wild type (Figure 28B).

SF2, SRp30c, hnRNP G, SAF-B, Znf265 and YT 521B suppressed exon 2 inclusion in the mature RNA (Figure 29). hnRNP and SR protein families have been reported to have opposite effects on splicing of the alternative exons. Usually, SR protein family members promote exon inclusion while hnRNPs prevent exons from being spliced in. Therefore it is surprising that SF2 and SRp30c, two SR-protein members and hnRNP G have the same effect. In addition, SAF-B, which is not a splicing factor but rather an organizing element that couples transcription and pre-mRNA

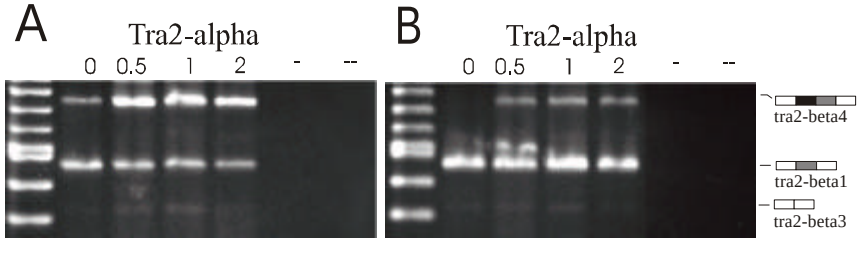


Figure 28. Tra2-alpha. (A) TRA2-alpha induces tra2-beta exon 2 inclusion with the same overall efficiency as TRA2-beta1. (B) TRA2-alpha is unable to induce exon 2 inclusion in a minigene with destroyed tra2-beta binding site 3.

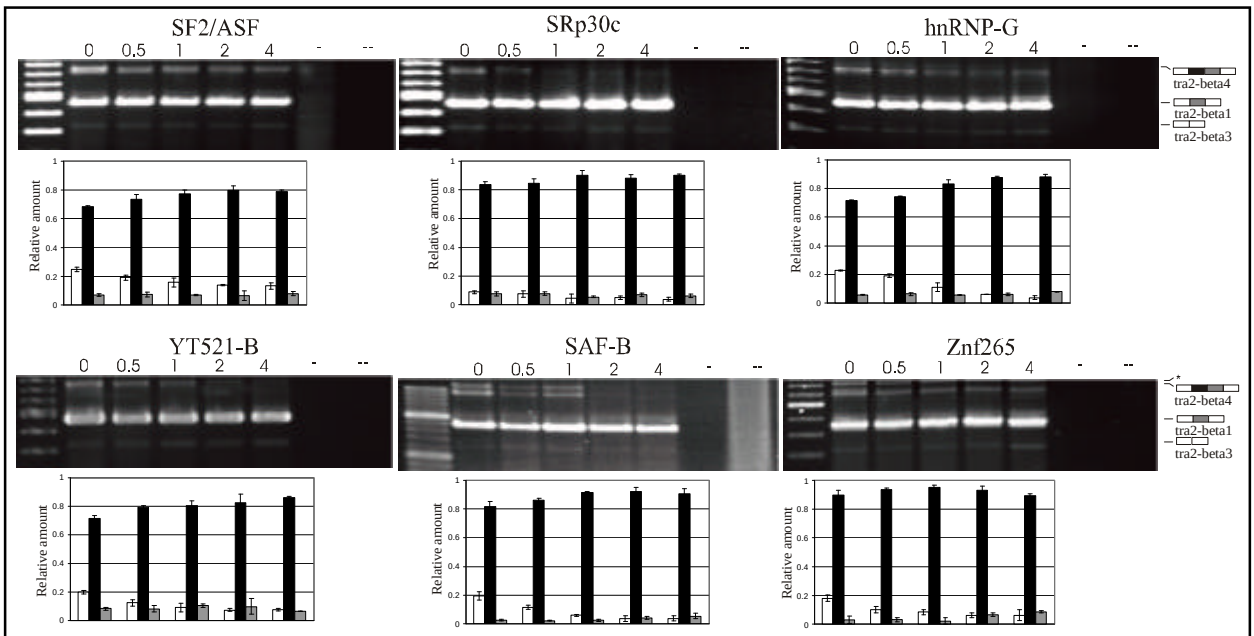


Figure 29. Several splicing factors suppress exon 2 inclusion. SG2/ASF, SRp30c, hnRNP-G, YT521-B, SAF-B and Zn265 prevent inclusion of exon 2 of the tra2-beta minigene in the mature transcript. Asterisk indicate the positions of the heterodimer. Below the gels is shown the quantified data from three independent experiments. The bars represent the relative amounts of the beta1 (black), beta3 (gray) and beta4 (white) isoforms.

processing to the nuclear matrix, also changes the splicing pattern of the tra2-beta minigene. All these factors, including YT 521B, interact with TRA2-beta1 both in two hybrid assays and co-immunoprecipitation. Therefore it is possible that they indirectly act on the tra2-beta minigene but by sequestering TRA2-beta1 through protein-protein interactions. To prove that this is the case, the *in vivo* splicing assays were repeated with cDNA clones that coded for proteins unable to bind RNA, but still retained the ability to interact with TRA2-beta1. In the case of SRp30c, hnRNP G, Zn265 and SAF-B these clones had the RNA recognition motifs deleted. In the case of SF2, two conserved phenylalanine residues in the RRM were substituted with aspartic acid. As expected, these truncated proteins were still able to suppress exon 2 inclusion (Figure 30). Moreover they were indistinguishable in their activity from the wild type clones. It was not possible to check whether YT 521B's effect on

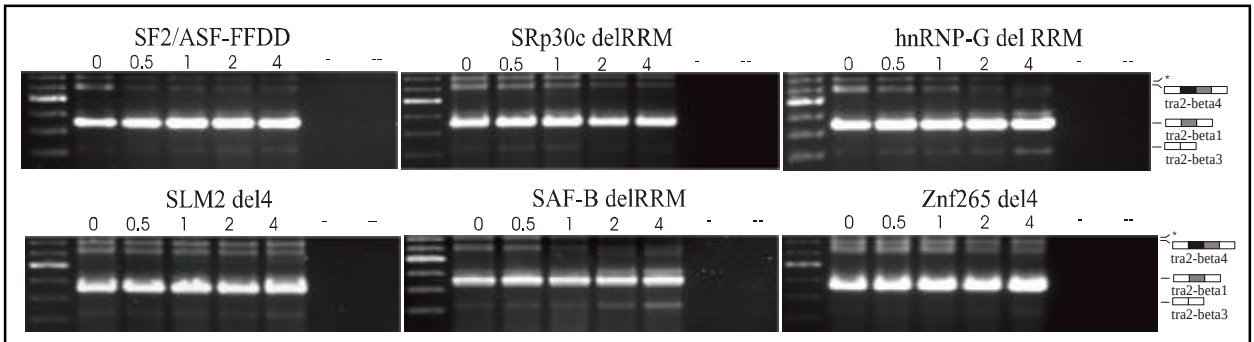


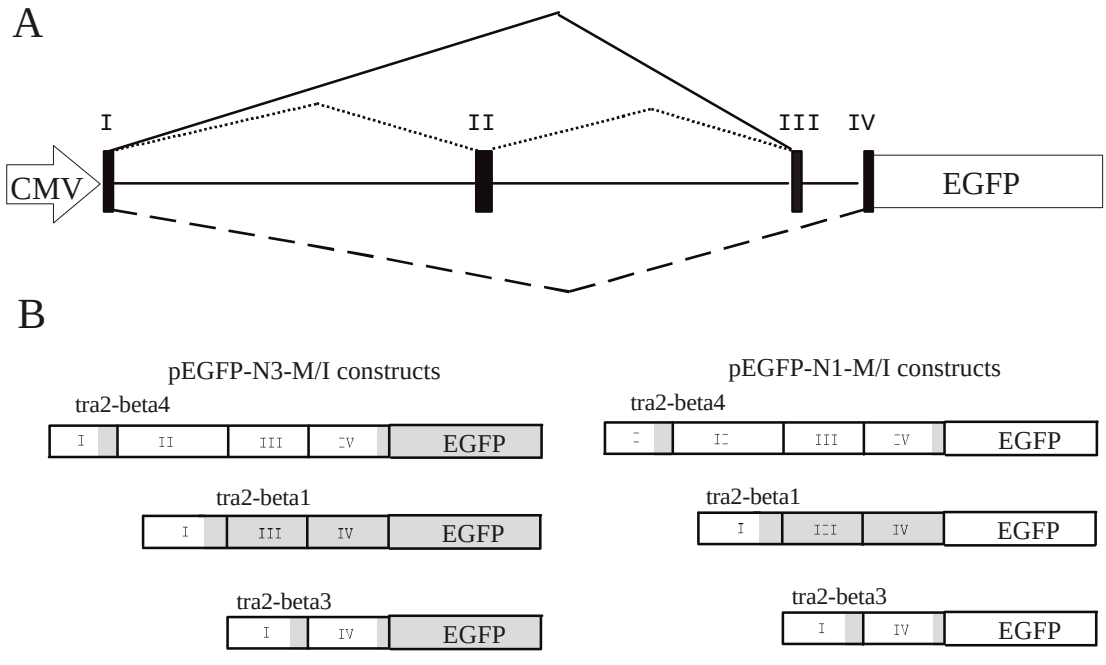
Figure 30. The effect of several splicing factors on tra2-beta minigene is not dependent on direct RNA binding. The SF2/ASF FF-DD double mutant and deletion variants of SRp30c, hnRNP-G, SAF-B, Zn265 that lack the RRM are still able to suppress exon 2 inclusion. Only SLM2 requires the presence of the RNA binding KH domain in order to modulate the splicing of the tra2-beta minigene

tra2-beta gene splicing is dependent on its RNA binding abilities, due to the absence of any identifiable RNA binding domain in its sequence that can be deleted or mutated. SLM2 has been shown before to suppress exon 2 splicing. It was postulated that it acts through a purine rich binding site that overlaps with tra2-beta1 binding site one (Stoss et al., 2001). However, it has also been shown that it forms a complex of several splicing factors including TRA2-beta1. To rule out the possibility that it can act through a sequestration mechanism, splicing assays were preformed using a deletion variant in which the RNA binding KH domain is destroyed. As can be seen from Figure 30, this variant is not able to influence tra2-beta exon 2 splicing.

6.2.12. Proteins expressed from the different tra2-beta mRNAs

The three mRNAs that are generated from the alternative splicing of exons 2 and 3 can be translated from two alternative codons. Tra2-beta1 RNA is most probably translated from the ATG codon in exon 1 to produce a protein consisting of two SR repeats flanking a RRM. In the tra2-beta4, mRNA this open reading frame is interrupted by multiple stop codons in exon 2, while in tra2-beta3 mRNA the ATG codon in exon 1 is out of frame with the rest of the message. These two mRNAs can be translated from an alternative codon located at the end of exon 4 and the resulting protein will lack the first SR domain. To assess the influence of the alternative splicing on the protein synthesis from the tra2-beta locus, two new tra2-beta minigenes were created (Figure 31).

Figure 31. MGTra-EGFP minigenes. (A) Structure of the tra2-beta-EGFP minigenes. The part of the tra2-beta gene containing exons 1,2,3 and 4 and the intervening introns was cloned in pEGFP-N1-M/I and pEGFP-N3-M/I in front of the EGFP ORF. The exons are shown as black boxes. The lines above and below the minigene indicate the splicing pattern. (B) Structure the mRNAs expected to be formed from the tra2-beta-EGFP-N3-M/I (top) and tra2-beta-EGFP-N1-M/I (bottom) minigenes. The shading shows the open reading frames.



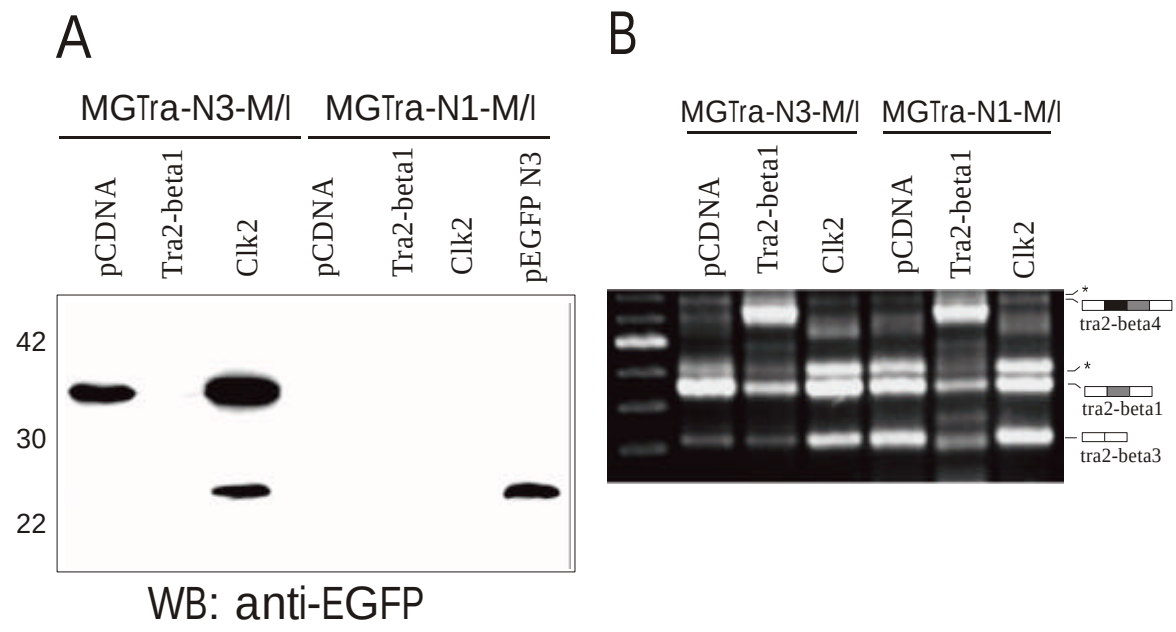


Figure 32. Proteins expressed from the different tra2-beta mRNA isoforms. (A) Western blot analysis of the proteins expressed from the MGtra-N3 M/I (lanes 1, 2 and 3) and MgTra-N1 M/I (lanes 4, 5 and 6) minigenes cotransfected with empty pCDNA vector and tra2-beta1 or Clk2 expression constructs. The splicing pattern of each cotransfection is shown on panel (B) The switch from tra2-beta1 to tra2-beta4 mRNAs induced by tra2-beta1 protein completely shuts off the protein synthesis (lane 2 compared to lanes 1 and 3). Induction of the tra2-beta3 isoform by Clk2 leads to the appearance of a faster migrating protein band that corresponds in size to the protein that should be produced by the tra2-beta3 ORF, e.g. approximately the same size as the EGFP protein (lane 7). No protein is detected when the tra2-beta minigene is cloned out of frame with EGFP (lanes 4, 5 and 6).

The *tra2-beta* gene part that spans the first four exons was cloned in front of EGFP, in pEGFP N1 M/I and pEGFP N3 M/I vectors. These vectors are derived from pEGFP N1 and N3 respectively by mutating the ATG start codon of EGFP to ATC (Isoleucine). This was done to force EGFP to be translated from start codons introduced by the cloned fragments. In the pEGFP N3 construct the tra2-beta1 and tra2-beta3 messages are in frame with the EGFP. The pEGFP N1 clone is a negative control in which the two messages are out of frame. The two constructs were cotransfected with flag-tagged TRA2-beta1 expression construct to induce exon 2 inclusion, with clk2 to block exon 2 and 3 inclusion or with empty pCDNA vector as a control. The splicing pattern was subsequently analyzed by RT-PCR (Figure 32B) and the protein expression was monitored by western blot (Figure 32A). As can be seen from Figure 32A, inclusion of exon 2 completely blocks protein synthesis from the tra2-beta1 locus. The tra2-beta4 mRNA is therefore not translated from the alternative ATG codon in exon 4. Skipping of exons 2 and 3 leads to increase of the amount of the protein that can be derived from the usage of the alternative start codon in exon 4. The band corresponding to the tra2-beta3 protein could still have been a result of proteolytic cleavage or alternative translation initiation. Two additional constructs were created to rule out these possibilities. In these constructs, the partial cDNAs comprising either exons 1, 3 and 4 (Tra2-beta1) or exon 1 and 4 (Tra2-beta3)

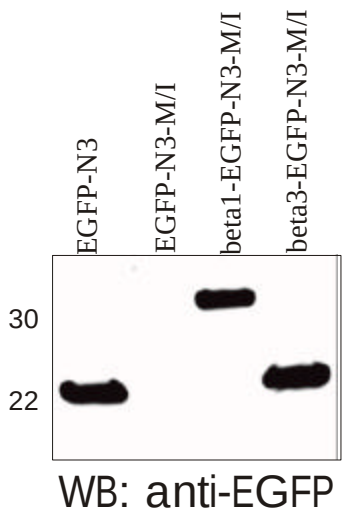


Figure 33. Protein expression from the tra2-beta1 and tra2-beta3 cDNAs. Tra2-beta1 (exons 1,3 and 4) and tra2-beta3 (exons 1 and 4) cDNAs cloned in front of EGFP produce on western blot single bands, showing that the tra2-beta3 band observed on figure 32A lane 3 is not product of protein degradation.

were cloned in the pEGFP-N3-M/I vector. After transfection in HEK A293 cells both clones expressed proteins of the expected size (Figure 33), indicating that both Tra2-beta1 and tra2-beta3 open reading frames are translated. In addition, no degradation products were observed for tra2-beta1 protein.

Previous studies (Nayler et al., 1998a) have shown that tra2-beta3 mRNA isoform is expressed predominantly in brain, testes and liver and is upregulated late in development. An anti-tra2 antibody (AK568) that detects both TRA2-beta3 and TRA2-beta1 proteins was used to probe rat brain and testes protein extracts for the presence of TRA2-beta3 protein. The western blot analysis shows that tra2-beta3 is expressed in all of the tested brain regions (Figure 34). The expression is also relatively high in testes. In conclusion, tra2-beta locus expresses two proteins: a full length protein consisting of two SR repeats flanking a RRM and a truncated form lacking the first SR repeat.

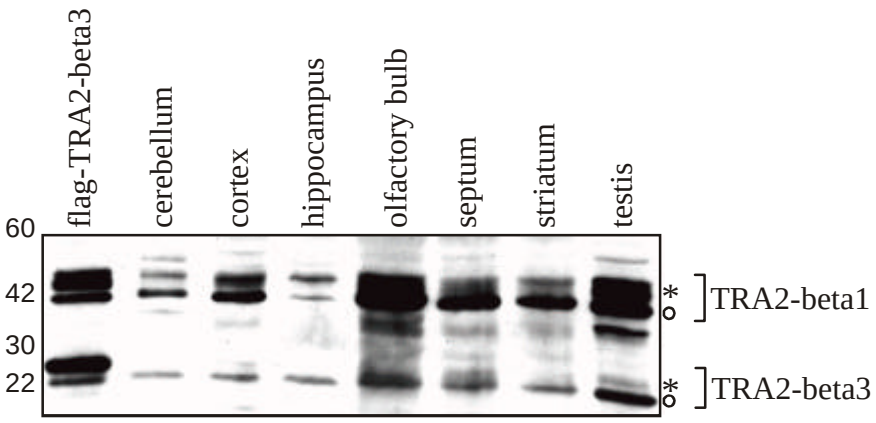


Figure 34. Expression of tra2-beta proteins in rat brain and testis as detected by the pan-Tra2 antibody. The first lane is protein from HEK A283 cells transiently transfected with flag-tagged tra2-beta3 expression construct. The open circle and the asterisk indicate the hyperphosphorylated and the hypophosphorylated forms of the TRA proteins.

6.2.13. Tra2-beta3 is inactive in splicing

Further *in vivo* splicing assays were carried out to investigate the role of TRA2-beta3 protein in splicing. In these assays tra2-beta3 expression construct was cotransfected with the tra2-beta, SMN2 and APP minigenes. The results are shown on Figure 35. In contrast to TRA2-beta1, TRA2-beta3 had no effect on the splicing of the three minigenes. The same difference between the two

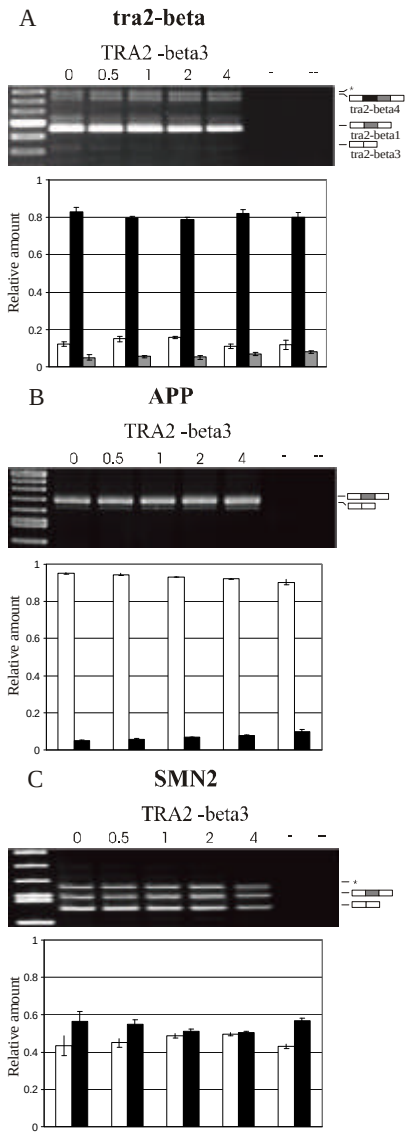


Figure 35. Tra2-beta3 is inactive in splicing. Cotransfection of tra2-beta (A), APP (B) and SMN (C) minigenes with increasing amounts of tra-beta3 expression construct does not change the utilization of the alternative exons. Exon composition of each of the bands is shown to the right. Asterisks indicate the positions of the heterodimers. Bar graphs show the quantified data from three independent experiments. The bars show the relative amount of each isoform as follows: (A) Tra2-beta1 – black, tra2-beta3 – gray, tra2-beta4 white; (B) APP containing exon 9 – white, lacking exon 9 – black; (C) SMN2 containing exon 7 – white, SMN2 lacking exon 7 – black.

isoforms has been reported for the CLB minigene. TRA2-beta3 is therefore inactive in regulating alternative splicing of the tested minigenes.

6.3. Discussion

6.3.1. Tra2-beta1 as splicing regulator

This work investigates both the regulation of the alternative splicing of tra2-beta pre-mRNA and the role of the proteins produced by the tra2-beta locus as splicing regulators. The TRA2-beta1 protein acts as a regulator of the alternative splicing of at least four genes, namely amyloid precursor protein, clathrin light chain B, survival of motor neuron 2 and the *tra2-beta* gene itself. The effect of TRA2-beta1 on the splicing of these genes is quite diverse both in respect to direction and strength. While TRA2-beta1 protein induces inclusion of the alternative exon in SMN2 and tra2-beta pre-mRNAs, it prevents the splicing of the alternative exon in APP and CLB. TRA2-beta1 recognizes a short and degenerate sequence on the target pre-mRNA. This consensus sequence was derived by single nucleotide mutagenesis and verified by pull down assays. Therefore TRA2-beta1 not only binds to this sequence but is also able to stimulate the inclusion of the exons that contain it in the mature RNA. A study by Tacke and Manley (Tacke et al., 1998) using *in vitro* SELEX has shown that TRA2-beta1 protein binds to purine rich sequences, consisting either of GAA repeats or poly(A) tracts. This work however shows that the tra2-beta1 binding site is much more degenerate and can carry a substantial number of pyrimidines. For example positions 2, 3 and 4 can be occupied by cytidine and position 7 can be occupied

by both cytidine and thymine.

To date, only four genes were identified to be targets for TRA2-beta1 regulation. I order to identify more targets the working draft of the human genome was screened with the TRA2-beta1 consensus sequence. The search identified 35,904 exons with statistically significant hits to the pattern compared to 16,112 introns. Therefore the splicing of 12.8% of all exons could be dependent on TRA2-beta1. This is a rough estimate which may deviate from the real number due to a number of issues:

1. The genome screen relied completely on the current annotation of the human genome to separate the exonic from the intronic sequences. This annotation is far from being complete and therefore some sequences are missed by the screen.

2. Whenever there are alternative transcription initiation or termination events, alternative splicing or simply multiple incomplete cDNAs, each one of them is annotated as an individual transcript. This results in portions of the genome that are annotated as both introns and exons, or the same exon or intron being annotated multiple times. The perl scripts used to perform the screen were not able to correct this and therefore some of the intron hits might actually be exon hits. This is often the case for alternative exons which can be included in the intron sequence when alternative transcripts in which they are skipped are used for the anotation. In addition, hits to the same exon (intron) are reported as hits to different exons (introns) when the same sequence is annotated several times.

3. The screen was preformed with a pattern rather than with a scoring matrix. This will identify hits that strongly deviate from the consensus as positive and therefore might also yield sites that are not functional.

4. The statistical analysis had taken only the TRA2-beta1 binding sites into account. As a result, it identified exons that depend exclusively on tra2-beta1 acting alone. However the splicing factors usually act cooperatively, each one of them binding to its own site in an exon. Therefore when the screen is performed with several consensus sequences simultaneously it could identify as statistically significant clusters of binding sites for different proteins that will otherwise be missed by screens performed using only a binding site for a single protein.

TRA2-beta1 has been shown to interact with several splicing factors as well as with core spliceosomal components, such as U1-70K. It can therefore facilitate the recognition of the tra2-

beta exon 2 and SMN2 exon 7, by recruiting such factors and organizing the assembly of the protein complex necessary for the exon recognition. No TRA2-beta1 binding sites are found in the alternative exons of APP and CLB, but several of them are located in the neighboring constitutive exons. In these genes, TRA2-beta1 molecules bound to the flanking exons can interact with each other bringing these exons together. This will loop out the alternative exon and prevent its recognition by the spliceosome complex.

6.3.2. Regulation of exon2 splicing – stechiometric ratio between TRA2-beta1 and its interactors

Even a slight increase of the TRA2-beta1 protein concentration leads to a dramatic shift in the tra2-beta pre-mRNA splicing pattern from tra2-beta1 to tra2-beta4 mRNA isoform. The shift is due to TRA2-beta1 dependent inclusion of exon 2 in the mature transcript. Exon 2 carries multiple stop codons that interrupt the tra2-beta1 open reading frame. A second start codon in exon 4 is also not utilized in this mRNA isoform, probably because it is located far from the 5' end of the message. As a result, whenever the TRA2-beta1 protein concentration rises above a certain level its synthesis is completely abolished. This threshold level is determined by the relative ratio between the concentration of the TRA2-beta1 protein and several other proteins: hnRNP-G, SAF-B, SF2, SRp30c, SLM2 and YT521B. On the molecular level at least four of these factors, namely hnRNP-G, SF2, SRp30c and SAF-B appear to control indirectly exon 2 inclusion in the tra2-beta pre-mRNA. Instead they interact with TRA2-beta1 protein and prevent it from binding to exon 2. The mode of action of YT521B is still unknown, as its RNA binding domain has not yet been identified. What is known is that this protein interacts directly with TRA2-beta1 and can therefore also act through a sequestration mechanism. SLM2 effect is dependent on the presence of the intact KH domain through which it binds to RNA. It has also been shown that SLM2 can act through purine rich sequences that in the case of tra2-beta exon 2, overlap with the tra2-beta1 binding sites (Stoss et al., 2001). SLM2 therefore, acts directly on the tra2-beta pre-mRNA, most probably by competing the TRA2-beta1 binding to exon 2.

Another set of factors that regulate the splicing of tra2-beta exon 2 are the clk protein kinases. They interact with and phosphorylate several SR repeat containing splicing factors, including TRA2-beta1 (Nayler et al., 1997; Nikolakaki et al., 2002). The hyperphosphorylation of tra2-beta1 prevents it from binding to RNA. In the case of tra2-beta minigene, TRA2-beta1 phosphorylation by clk kinases leads to a block of exon 2 inclusion and an increase in TRA2-beta1 protein synthesis.

6.3.3. Regulation of the tra2-beta3 isoform. Protein phosphorylation and signaling pathways

Protein phosphorylation by the Clk2 kinase induces skipping of exon 3 in addition to exon 2 of the tra2-beta minigene. The expression of the resulting tra2-beta3 mRNA isoform is restricted to brain, liver and testes and is developmentally regulated. The TRA2-beta3 expression is regulated by neuronal activity in rat hippocampus and cortex. Therefore the Clk2 kinase could be a member of the signaling pathways which control alternative splicing in the brain. No upstream or downstream members of this signaling pathway were identified.

Tra2-beta3 mRNA is translated from a start codon located at the end of the fourth exon. As a result, the TRA2-beta3 protein lacks the first SR repeat. Yeast two-hybrid screens and co-immunoprecipitation experiments show that TRA2-beta3 is very similar to TRA2-beta1 protein in respect to the proteins it interacts with. All of the protein interactors of TRA2-beta1 interact also with TRA2-beta3. This work shows that the two proteins are indistinguishable also in respect to their ability to recognize specific sequences in the pre-mRNA and bind to them. Furthermore, the RNA binding ability of TRA2-beta proteins depends solely on their RNA recognition motifs. It is therefore surprising that the TRA2-beta1 and TRA2-beta3 proteins show drastically different behaviour in *in vivo* splicing assay. While tra2-beta1 is acting as a splicing regulator affecting the alternative splicing patterns of at least four genes and possibly much more, TRA2-beta3 appears inactive. Some clues as of the function of the TRA2-beta3 come from studies on its *Drosophila* homologue dTRA2^{179}. In vitro splicing assays show that dTRA2^{179} is indistinguishable from the full length protein dTRA2^{226}. Nevertheless genetic studies convincingly show that it cannot substitute for the full length dTRA2^{226}. It is therefore possible that both TRA2-beta3 and dTRA2^{179} can support constitutive splicing, but can not influence the splicing of the alternative exons. One reason for this could be that the lack of the first SR repeat prevents the proteins from forming multiple protein-protein interactions simultaneously.

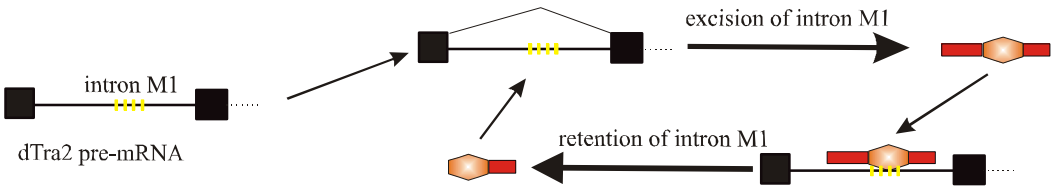


Figure 36. Regulation of the alternative splicing of the *D. melanogaster* tra2 gene. dTRA2^{226} protein regulates its level trough a negative feedback loop. It binds to multiple sites in intron M1 and prevents its excision. Intron M1 interrupts the ORF for dTRA2^{226}, forcing the usage of a downstream ATG codon resulting in the synthesis of the truncated dTRA2^{179} protein.

6.3.4. Autoregulation of the protein levels and signaling pathways controlling the ratio of the active versus inactive protein.

The splicing of the *Drosophila* tra2 pre-mRNA is controlled by an autoregulatory loop (Figure 36). Its most likely function is to keep the concentration of the active TRA2 protein constant. dTRA2 acts on its pre-mRNA by blocking the splicing of the M1 intron. The M1 intron containing mRNA is translated from a downstream codon to yield the $dTRA2^{179}$ protein. $dTRA2^{179}$ lacks the first SR repeat and can not influence the splicing of the alternative exons controlled by the full length $dTRA2^{226}$ protein. In mammals, there are two independent regulatory systems that control the tra2-beta pre-mRNA splicing (Figure 37). The first system is a negative feedback loop that is similar, but not identical to, the one in *Drosophila*. In this system TRA2-beta1 protein induces inclusion of exon 2, which prevents the mRNA from being translated. The role of this autoregulatory loop is to keep a steady state ratio between TRA2-beta1 and several other splicing factors. The second regulatory system comprises signaling pathways that control the splicing of exons 2 and 3 through protein phosphorylation. This allows for a rapid shift between the active TRA2-beta1 and the inactive TRA2-beta3 isoforms. Such a shift is observed in rat hippocampus and cortex, where changes in neuronal activity lead to rapid changes in the ratio of the tra2-beta1 and tra2-beta3 isoforms. Tra2-beta pre-mRNA is not an endpoint for the signaling pathways. Because TRA2-beta1 protein is a splicing factor, any changes in the splicing of the tra2-beta pre-mRNA will result in alterations in the splicing patterns of multiple genes. A search in the working draft of the human genome has identified 35,904 exons whose splicing could depend on tra2-beta1. In conclusion, this work shows that: (i) the splicing

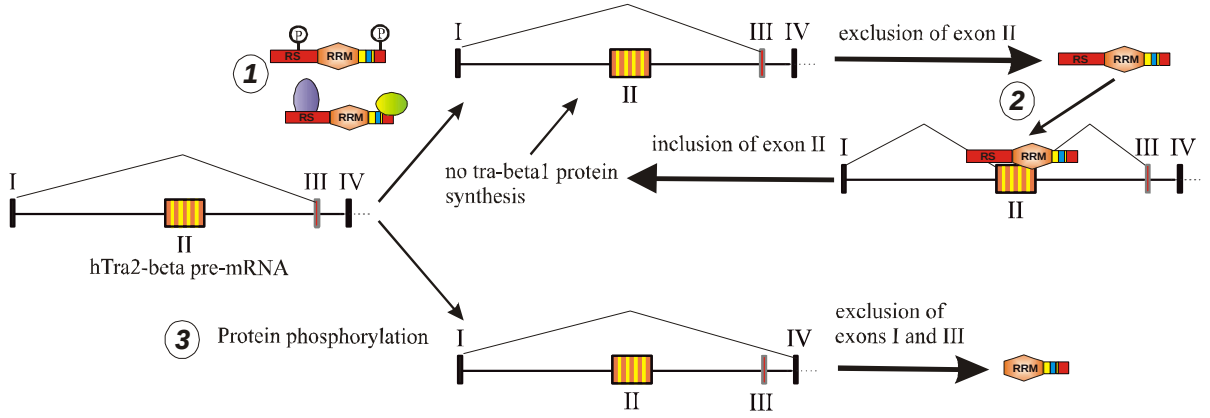


Figure 37. Regulation of the alternative splicing of the *H sapiense* tra2-beta gene. The human TRA2-beta1 protein also uses an autoregulatory loop to control its levels. It binds to four sites in exon 2 and induces its inclusion (2). Analogous to the *Drosophila* gene, this exon interrupts the open reading frame forcing the synthesis of a non functional truncated protein (tra2-beta3). Proteins that interact with TRA2-beta1 will sequester it, preventing exon 2 inclusion (1). Protein phosphorylation will prevent exon 2 inclusion by blocking tra2-beta1 binding to RNA (1). Through yet unknown mechanism, protein phosphorylation prevents splicing of both exons 2 and 3, leading to the synthesis of the truncated TRA2-beta3 protein (3). This allows signal transduction pathways to regulate TRA2-beta activity in the cells.

of the human tra2-beta1 pre-mRNA is controlled both by a negative feedback loop and signaling pathways; (ii) TRA2-beta1 protein binds in a sequence specific manner to exon 2 of its own pre-mRNA and activates its splicing; (iii) The consensus binding site has been identified and has been shown to be an exonic hallmark; (iv) RNA binding properties of TRA2-beta1 are modulated by protein phosphorylation and therefore can be regulated by signal transduction pathways; (v) TRA2-beta3 protein exists *in vivo* and has different splicing properties than the TRA2-beta1.

7. TrkB

7.1. Introduction

TrkB is a member of the protein tyrosine kinase family (Blume-Jensen and Hunter, 2001; Robinson et al., 2000). Together with trkA and trkC, TrkB forms the trk subfamily of RTKs. TrkB is the receptor for the brain derived neurotrophic factor (BDNF) and neurotrophin 4/5, whereas other neurotrophin family members preferentially bind to other trk receptors, i.e. nerve growth factor to trkA and neurotrophin-3 to trkC (Bibel and Barde, 2000; Lewin and Barde, 1996; Patapoutian and Reichardt, 2001; Thoenen, 1995).

The analysis of TrkB cDNA has revealed that it contains several protein domains common to other receptor tyrosine kinases (Allen et al., 1994; Middlemas et al., 1991; Schneider and Schweiger, 1991; Shelton et al., 1995). The extracellular domain of the receptor contains two cysteine rich regions separated by a leucine rich domain. Two IG like regions are located wetween those domains and the transmembrane domain. The intracellular part of the receptor contains the juxtamembrane domain that includes a Shc binding site, a tyrosine kinase domain and a tail region containing a PLC-gamma binding site (Figure 38A).

Northern blot analysis of TrkB messages has indicated that a number of mRNAs are generated from the TrkB locus (Middlemas et al., 1991) which is located on human chromosome 9 (Nakagawara et al., 1995). Molecular cloning revealed the existence of several isoforms generated by alternative

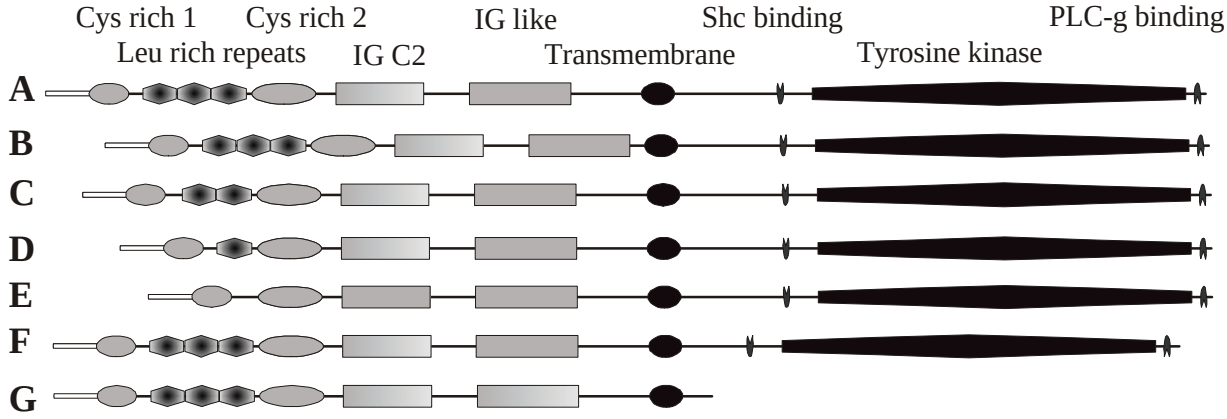


Figure 38. TrkB protein isoforms. (A) Full length receptor, (B) Receptor lacking an exon in the extracellular juxtamembrane region, (C, D and E) Receptors lacking one ore more of the leucine rich repeats, (F) Receptor lacking exon in the intracellular juxtamembrane region, (G) Truncated receptor lacking the kinase domain and Shc and PLC-gamma binding sites.

splicing. In the intracellular region, two isoforms are generated by this mechanism: a full length TrkB receptor containing the tyrosine kinase domain (Figure 38A, B, C, D, E and F) and a truncated form lacking it, named TrkB-T1(Figure 38G) (Klein et al., 1990; Middlemas et al., 1991; Shelton et al., 1995). In rodents, another truncated form (TrkB-T2) has been described, with an apparently identical splice site as that of TrkB-T1, but with a unique intracellular tail region (Middlemas et al., 1991). The TrkB-T1 form can bind to neurotrophins and is internalized upon ligand binding (Biffo et al., 1995) but due to the lack of a kinase domain cannot signal to the cytoplasm. Its exact function remains controversial. However, accumulating evidence suggests that the truncated TrkB isoforms act in a dominant negative way (Eide et al., 1996; Haapasalo et al., 2001), possibly suppressing neurotrophin action.

In addition to alternative splicing of the intracellular regions, alternatively spliced forms lacking one or all of the leucine rich motifs in the extracellular region (Figure 38C, D and E) have been described (Ninkina et al., 1997). The isoforms lacking the leucine rich motifs can no longer bind neurotrophins (Ninkina et al., 1997). In chicken, additional splice variants were described, one in the extracellular juxtamembrane region (Figure 38 B) which leads to a reduced affinity to NT 4/5 (Garner et al., 1996; Strohmaier et al., 1996) and others with unknown function in the cytoplasmic juxtamembrane region (Figure 38 F) (Garner et al., 1996). The splicing variant with reduced affinity to NT4/5 found in chicken was also found to be expressed in cultured human retinal pigmented epithelium cells (Hackett et al., 1998). Together, these data show that alternative splicing is used to generate molecules with different function from the TrkB locus. This is typical for higher eukaryotic gene regulation where about 60% of all genes are regulated by this mechanism (Lander et al., 2001; Stamm et al., 2000; Tupler et al., 2001). Since alternative splicing pathways can be regulated by extracellular signals (Stoss et al., 2001), e.g. neuronal activity (Daoud et al., 1999) the formation of splice variants is most likely important for neurotrophin signaling through TrkB.

The generation of multiple splice forms wich appear to be regulated both in developmental as well as tissue and cell type specific manner makes the TrkB gene an interesting model for alternative splicing research. As a first step to investigate the molecular mechanisms generating TrkB isoform diversity, the gene structure of the human TrkB gene was determined. The gene is unusually large, contains a non canonical intron and an extended repeat. TrkB gene can produce 100 mRNA isoforms and 10 different proteins. However three forms, the full length receptor and two truncated receptors, one of which has a binding site for the Shc adaptor molecule, are the major ones.

7.2. Results

7.2.1. Isolation of the human TrkB gene

To isolate clones corresponding to the TrkB gene, ESTs from the 5', 3' and middle parts of the TrkB cDNA were used to probe a human genomic PAC library. The screen was performed by German resource center in Berlin and identified 28 independent clones. These clones were verified by hybridization with TrkB specific oligonucleotides (Figure 39). Two of them were found to be false positives and one contained an internal recombination. Four clones (RPCIP704J22583, RPCIP704P04728, RPCIP704N1042 and RPCIP704F1647) covering the entire known mRNA sequence were characterized further. The clones were sequenced by primer walking starting from the known exons. The list of primers can be obtained from http://www.stamms-lab.net/pub_data/trkB_sequencing_primers.txt. Additional sequence information was obtained by sequencing random transposone insertions.

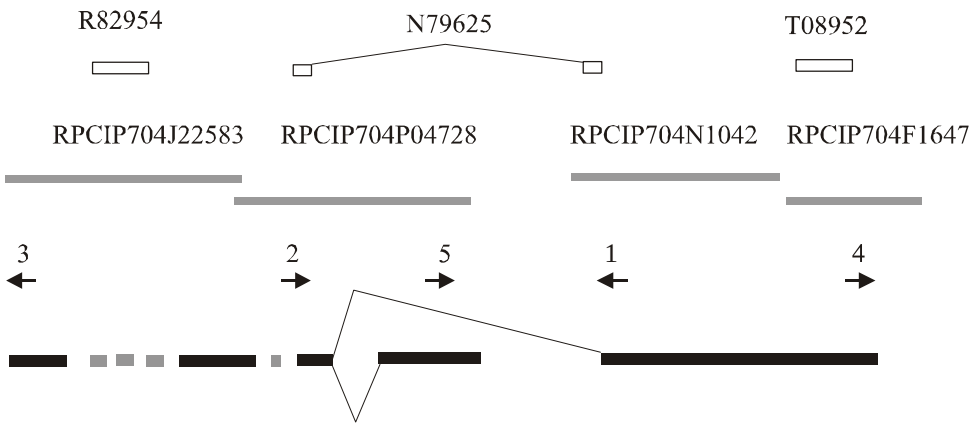


Figure 39. Screening strategy. The three ESTs used to screen the PAC library are shown as open boxes on the top. Numbers above the ESTs indicate their GeneBank accession numbers. EST N79625 hybridizes to two nonoverlapping genomic clones, because it contains two exons that are separated by a large intervening sequence in the genomic DNA. Positive clones are indicated by gray lines and were verified by rescreening with primers TrkBseq1-5 (arrows). Numbers on top of the clones indicate their names from the German resource center, Berlin. A schematic gene structure derived from the available cDNA data is shown below. The black blocks correspond to the parts of the mRNA with unknown exon structure. The grey blocks represent the previously described alternative exons. The black lines show the alternative splicing that generates the major isoforms encoding the full length TrkB and the TrkB-T1 receptors.

7.2.2. Structure of the gene

Sequencing of the gene shows that it contains 24 exons and 23 introns (Figure 40). One remarkable feature of the gene is its large size. The gene spans at least 590 kbp. The TrkB exons are located in five clusters. The first cluster contains the putative promoter region and exons 1-5. The

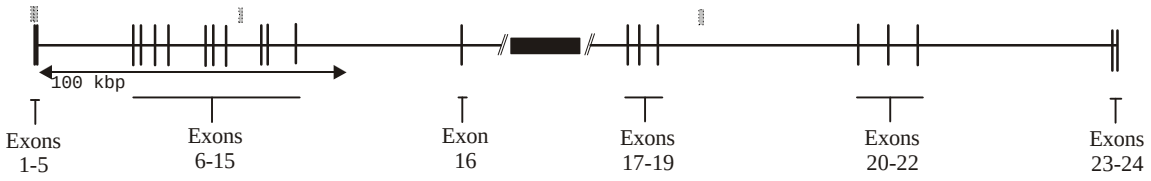


Figure 41. Structure of the TrkB gene. The structure of the TrkB gene is drawn to scale. Exons are indicated as vertical lines, introns are shown as horizontal lines. The ChAB4 repeat is indicated by a black box. The gap in the sequence surrounding the ChAB4 repeat is indicated by a double tilted line. CpG islands are indicated as gray boxes on top.

start codon is present in exon 5. Approximately 30 kbp downstream of this exon is the second exon cluster comprising exons 6-15. These exons encode the extracellular part of the protein, the transmembrane region and part of the juxtamembrane domain. Exon 16, which is used in the truncated receptor (TrkB-T1), is located 50 kbp downstream. A large intron of at least 300 kbp separates this exon from the third cluster. The third cluster is composed of exons 17-19. Exons 17 and 18 encode the intracellular juxtamembrane part of the receptor. Exon 19 is an alternative terminating exon which is used in a new truncated receptor form (TrkB-T-Shc). The fourth cluster containing the exons coding for part of the kinase domain is located approximately 60 kbp downstream. A 50 kbp intron separates it from the last two exons that compose the carboxy terminal part of the kinase domain and the phospholipase C gamma binding site. The exons vary in size from 36 to more than 6000 nt, with a median size of around 100 nt. The introns range from 115 nt to more than 300 kbp. Another unusual feature of the TrkB gene is the presence of a ChAB4 repeat between exons 16 and 17. This 250 kbp large repeat is found about 50 times in the human genome (Wohr et al., 1996). The translational phase of the exons changes within the gene: three exons are in first, eight exons in second phase and seven in the third phase (Table 9). The Grail program (Xu and Uberbacher, 1997) detects four CpG islands. Two of them are located in the promoter region. Two more CpG islands are found in introns 12 and 19 (Figure 40 and 41).

7.2.3. Structure of the 5'UTR

In order to analyze the 5' untranslated region of the gene, 5' RACE was performed on capped RNA using the GeneRacer kit (Invitrogen). This technique uses subsequent dephosphorylation, decapping and RNA oligonucleotide ligation (Figure 8) to amplify only capped pre-mRNA molecules, which results in the detection of putative transcription start sites (Maruyama and Sugano, 1994; Schaefer, 1995; Volloch et al., 1994). The 5' RACE was performed on RNA derived from the SH-SY5Y human neuroblastoma cell line. Sequencing of the clones revealed that the TrkB gene is

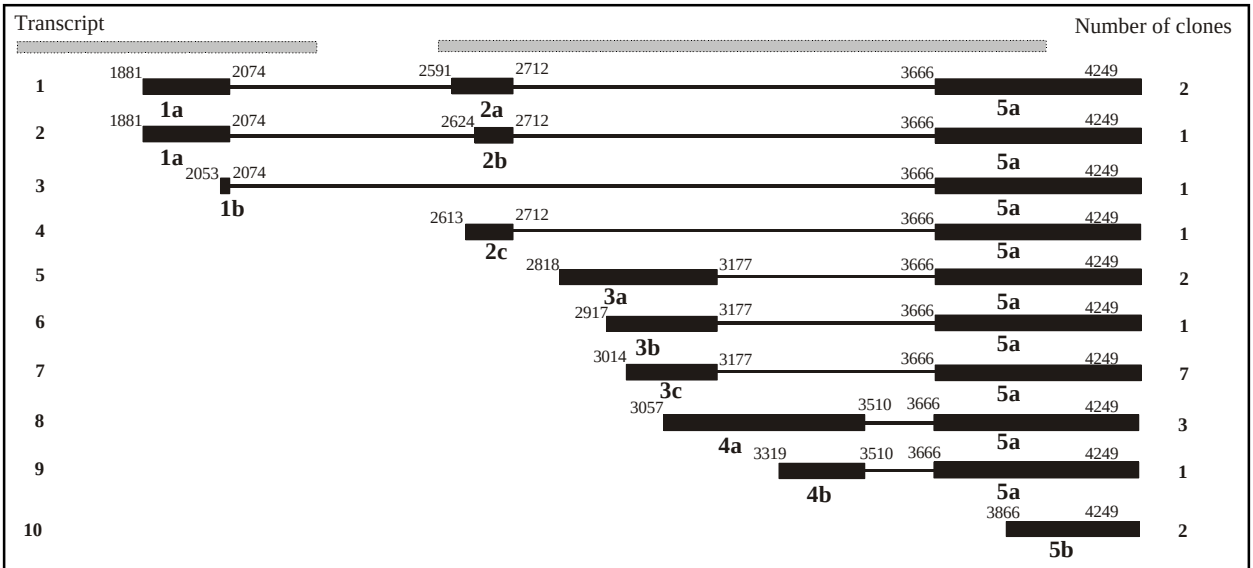


Figure 41. Structure of the TrkB gene 5' UTR. The first five exons are indicated as black boxes. Numbers indicate the exon positions in the genomic sequence (accession number AF410902). The CpG islands are indicated with gray boxes on the top.

transcribed from at least nine different start sites. Based on the number of clones, the transcriptional start at position 3014 (Figure 41) is the most common one. The putative promoter region shows no apparent initiator elements (i.e. TATA box) and is highly G/C rich. Computer prediction indicated the presence of two CpG islands in the promoter region (Figure 41). In conclusion, the TrkB promoter shows hallmarks of housekeeping genes and uses several transcriptional start sites.

7.2.4. The 5' UTR contains several alternative exons, one with unusual splice sites

In addition to alternative transcription start site usage, diversity at the 5' UTR is generated by alternative splicing. Exon 2 can be alternatively spliced, using two different 3' splice sites. Furthermore, exons 3 and 4 do not contain 3' splice sites and therefore are not joined to exons 1 or 2. A remarkable feature of the 5' UTR is the presence of a noncanonical exon. Exon 4 uses a GC rather then a GT dinucleotide in the 5' splice site of its downstream intron. The annotated genes in the human genome database at NCBI were searched for more introns containing such 5' splice site. The search resulted in 832 examples, which is 1% of all annotated introns. Similar results regarding the abundance of the GC-AG type of introns were obtained by Thanaraj and Clark (Thanaraj and Clark, 2001).

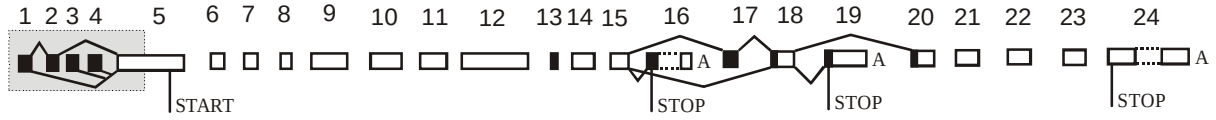


Figure 42. Splicing pattern of the TrkB gene. Alternative splicing patterns of the TrkB gene. Exons are shown as boxes and numbered. Constitutive exons are shown as white boxes, alternative exons are shown as black boxes. Alternative splicing patterns are indicated by lines. The area marked in gray is enlarged in figure 40. Dashed boxes indicate incomplete exon sequence.

7.2.5. Alternative splicing patterns

Systematic searches of the EST database with cDNA and genomic sequences form the TrkB locus, RT-PCR and RACE were used to identify novel exons and resulted in the discovery of 6 new exons: 1, 2, 3, 4, 17 and 19 (Figures 41 and 42). Two of them are cassette exons (exons 2 and 17), one has an alternative 3' splice site (exons 2a and 2b), one terminates with an alternative polyadenylation site (exon 19) and three exons are generated from alternative transcription start sites (exons 1, 3 and 4). When the nine different start sites are taken into account, the TrkB gene could generate 100, different mRNA isoforms that code for 10 different proteins. The protein domain composition is often reflected in the gene structure. In the human TrkB gene, several protein domains are encoded by a single exon. Exon 5 encodes the cysteine rich domain and each of exons 6 to 8 encodes a leucine rich motif (Figure 43). The IG C2 and IG like domains are formed by exons 10 and 11, as well as exon 12, respectively (Figure 43). The transmembrane domain corresponds to exon 15 and the tyrosine kinase domain is composed of exons 20 to 24 (Figure 43). The isoforms with different polyadenylation sites are formed by usage of the alternative exons 16 and 19 (Figures 42 and 43). This shows that the overall protein structure is reflected in the gene structure.

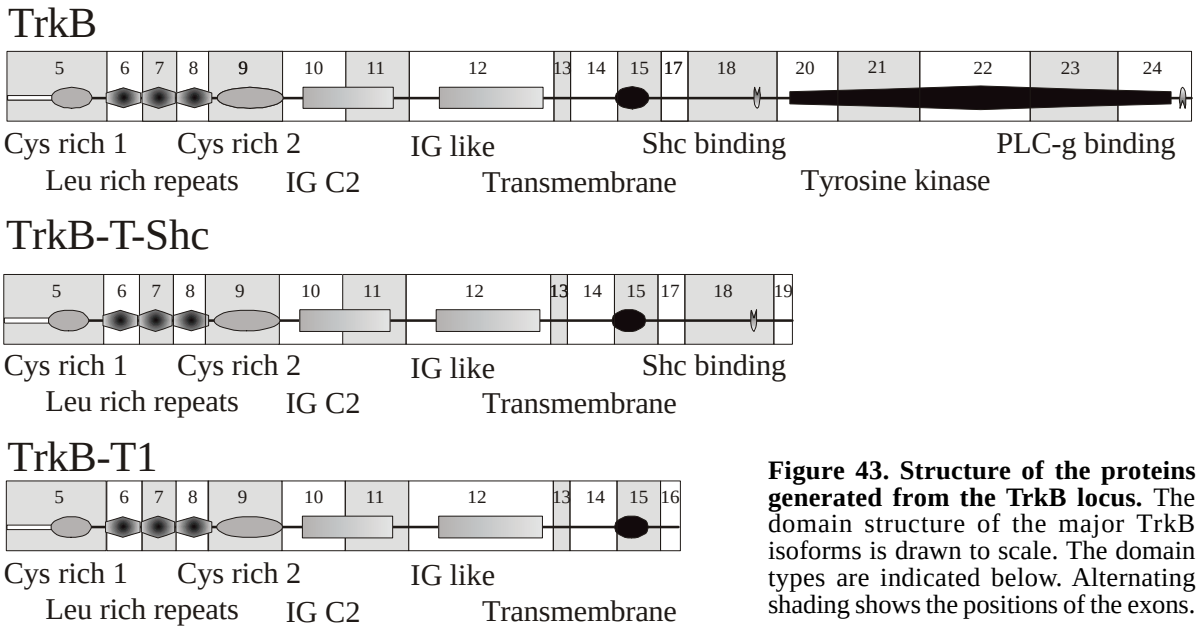


Figure 43. Structure of the proteins generated from the TrkB locus. The domain structure of the major TrkB isoforms is drawn to scale. The domain types are indicated below. Alternating shading shows the positions of the exons.

Table 9. Compilation of exon-intron boundaries. Protein coding exon sequences are indicated by boldface upper-case letters, untranslated exon sequences by upper-case letters and intronic sequences by lower case letters. Alternative exons are underlined. The amino acid usage is indicated by spaces between the triplets. Exon and intron sizes are indicated in nt. 5' and 3' scores are given for each splice site. A perfect match to U1 snRNA would give a theoretical maximum 5' score of 12.6 and the average score of constitutive exons is 8.0. Likewise, the theoretical maximum 3' score is 14.2 and the average score for constitutive exons is 9.4. Transcriptional start sites are in italic letters. Nucleotides found in the cDNA clones are indicated in capital italic letters.

Exon	**Size**	**3' splice site/*Transcriptional start site***		**5' splice site**		**Intron size**
		Site	**Score**	**Site**	**Score**	
1a	194	*ccccgcagccATCATTTAACT….*		….ACGCGCAAG/gtacgg	8.4	517 to Exon 2a
						550 to Exon 2b
1b	22	*cgcacaccctAGCACACATG*				1592 to Exon 5a
2a	122	cgcatttttcag/AGCTGAACCA….	6.5	….CATCTTAGAG/gtacct	5.2	1536
2b	89	catttcaaaaag/GGAGACAGCC….	1.1	….CATCTTAGAG/gtacct		
2c	100	*cacggtttccATTTCAAAAA*		….CATCTTAGAG/gtacct		
3a	360	*gtgaactcccACATGCTGCT….*		….CGGAGCCCGG/gtgagc	7.9	584
3b	261	*tttgtctggaGGGTGTTATG….*		….CGGAGCCCGG/gtgagc		
3c	165	*cagccctcacGTCACTTCGC….*		….CGGAGCCCGG/gtgagc		
4a	454	*gcggctcccgGAATTGGGTT….*		….CCATCGCGGG/gcaagt	1.3	251
4b	192	*attctgcagcATCATTCGGG….*		….CCATCGCGGG/gcaagt		
5a	584	gttttgtcctcag/CCTCGAGGTG….	7.0	….**TC ACC GAA AT**/gtgagt	8.9	31 198
5b	351	*ctgccggaacACTCTTCGCT….*		….**TC ACC GAA AT**/gtgagt		
6	75	gctttgttacag/**T TTC ATC GCA**….	5.9	….**TG AGA AAT CT**/gtgagt	5.4	115
7	72	gtgttttcacag/**G ACA ATT GTG**….	8.5	….**TG CAG CAC AT**/gtaagt	9.0	5 431
8	69	gattcctttcag/**C AAT TTT ACC**….	6.3	….**TG TCT GAA CT**/gtaagt	7.0	2 728
9	155	tgttccctgtag/**G ATC CTG GTG**….	10.0	….**CCC AAT TGT G**/gtaatt	6.0	12 782
10	137	cttgttccatag/**GT TTG CCA TC**….	9.3	….**CC AAA CAA TG**/gtaagg	9.1	513
11	133	attcatttgtag/**A AAT GAA ACA**….	4.4	….**ACT GTG CAT T**/gtacgt	4.4	3 296
12	306	ttttcaatttag/**TT GCA CCA AC**….	4.2	….**ATT GAC GAT G**/gtgagt	10.4	13 923
13	36	tttgccttttag/**GT GCA AAC CC….**	9.4	….**AAT TTA TAA G**/gtagct	5.9	3 046
14	101	tgcccacttaag/**AT TAT GGA AC**….	2.4	….**A CAT CTC TCG**/gtgagt	8.9	6 913
15	100	tcctttctctag/**GTC TAT GCT G**….	11.2	…**GGC ATG AAA G**/gtaaga	11.3	56 400 to Exon 16
						>110 584 to Exon 17
16	~6000	tgtttcttttag/**GT TTT GTT TT….**	9.4	Putative terminating exon		
17	48	tgtggttttcag/**AT TTC TCA TG….**	10.2	….**CAA GGT GTT G**/gtaagt	9.2	6 156
18	189	ttccatctccag/**GC CCA GCC TC….**	6.1	….**CCA GAC ACA T**/gtaagt	9.0	4 358 to Exon 19
						66 731 to Exon 20
19	6 109	actttatgcag/**GG CCC AGA GG….**	7.5	Terminating exon		
20	131	tttgttttgcag/**TT GTT CAG CA**….	8.6	….**TG GCA GTA AG**/gtaaga	11.3	14 170
21	174	cacccatccccag/**A ACC CTG AA**….	5.7	….**AG TTC CTC AG**/gtacag	4.8	6 649
22	234	gtccttccccag/**G GCA CAC GG**….	10.0	….**C TAC TAC AGG**/gtgagt	10.2	64 689
23	159	ctttctccccag/**GTC GGT GGC C**….	13.3	….**A AAC AAT GAG**/gtgtgc	6.7	888
24	~6000	atctccatccag/**GTG ATA GAG T**….	7.9	Putative terminating exon		

Alternative exon usage is often a result of suboptimal splice sites (Stamm et al., 2000). In order to assess the quality of the splice sites of the alternative exons, their score (Stamm et al., 1994) was calculated. The score expresses how well the splice site follows the consensus sequence and binds to the components of the spliceosome. With the exception of exon 4a, no strong deviations from the consensus were found in alternative exons (Table 9). Exon 4a is described by a weak score, because it uses a noncanonical GC rather than a GT. The good match of the alternative splice sites with the mammalian consensus sequence indicates that their regulation might involve repressing, rather than activating, splicing regulatory factors.

7.2.6. Expression of the splicing variants

To address the question of how abundantly the individual splice variants are expressed, regions of the mRNA containing alternative exons were amplified by RT-PCR (Figure 44). Total RNA derived from human cortex and from the SH-SY5Y human neuroblastoma cell line (Biedler et

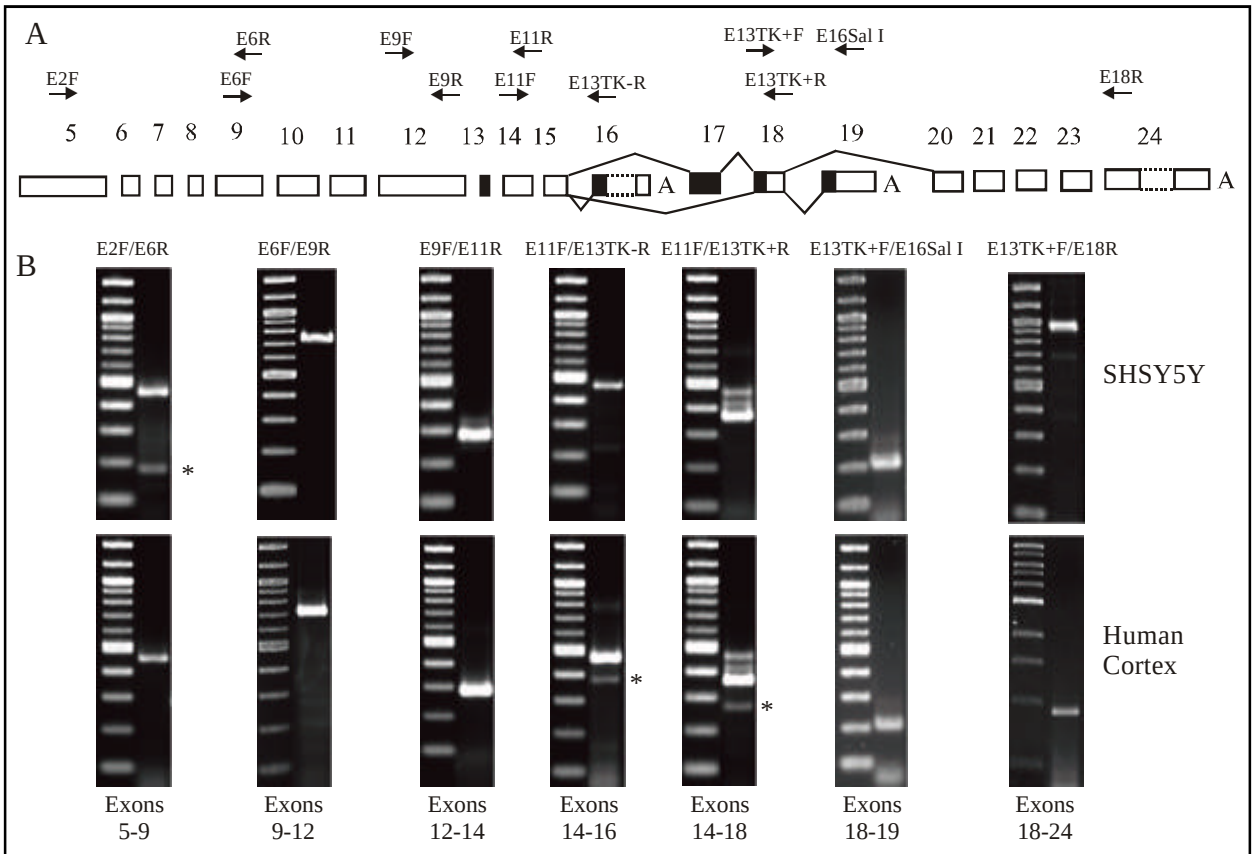


Figure 44. Expression of the alternative splice variants in SHSY5Y cells and human cortex. (A) Partial gene structure of TrkB with primer pairs used for RT-PR indicated on top. (B) RT-PCR with cDNA derived from human cortex and SH-SY5Y cells. All PCR products were subcloned and verified by sequencing, which showed that the minor bands indicated with stars were artificial. The marker represents a 100 bp ladder in all experiments, only in human cortex using primers E13TK+F/E18R a 1 kb ladder was used. The amplified exons are shown at the bottom.

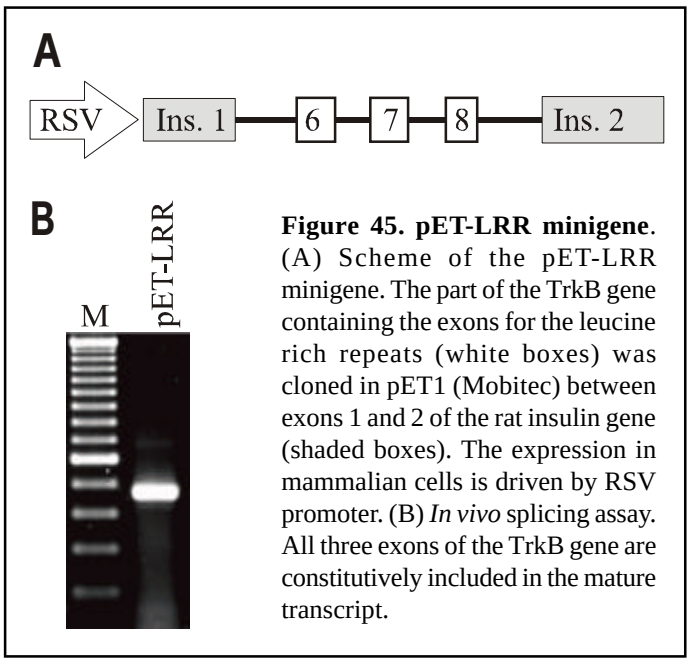


Figure 45. pET-LRR minigene. (A) Scheme of the pET-LRR minigene. The part of the TrkB gene containing the exons for the leucine rich repeats (white boxes) was cloned in pET1 (Mobitec) between exons 1 and 2 of the rat insulin gene (shaded boxes). The expression in mammalian cells is driven by RSV promoter. (B) *In vivo* splicing assay. All three exons of the TrkB gene are constitutively included in the mature transcript.

al., 1978) was used as template. Surprisingly, the exons encoding leucine rich repeats (exons 6, 7 and 8) were not alternatively spliced. To test whether these exons would be spliced alternatively in a different system a minigene was created (Figure 45A). In this minigene, the three exons were cloned between the rat insulin exons of pET1 exon-trap vector (Mobitec). The minigene was then transfected in HEK A293 cells and its splicing pattern was analyzed using RT-PCR. Again, no alternative splicing of exon 6, 7 and 8 was detected (Figure 45B). As shown in Figure 44, the newly identified exon 17 is used in both human cortex and SH-SY5Y cells. In contrast, all other cassette exons represent minor forms. In conclusion, the part of the TrkB juxtamembrane region that is encoded by exon 17 is regulated by alternative splicing.

7.2.7. Isoform diversity is generated by usage of different polyadenylation sites

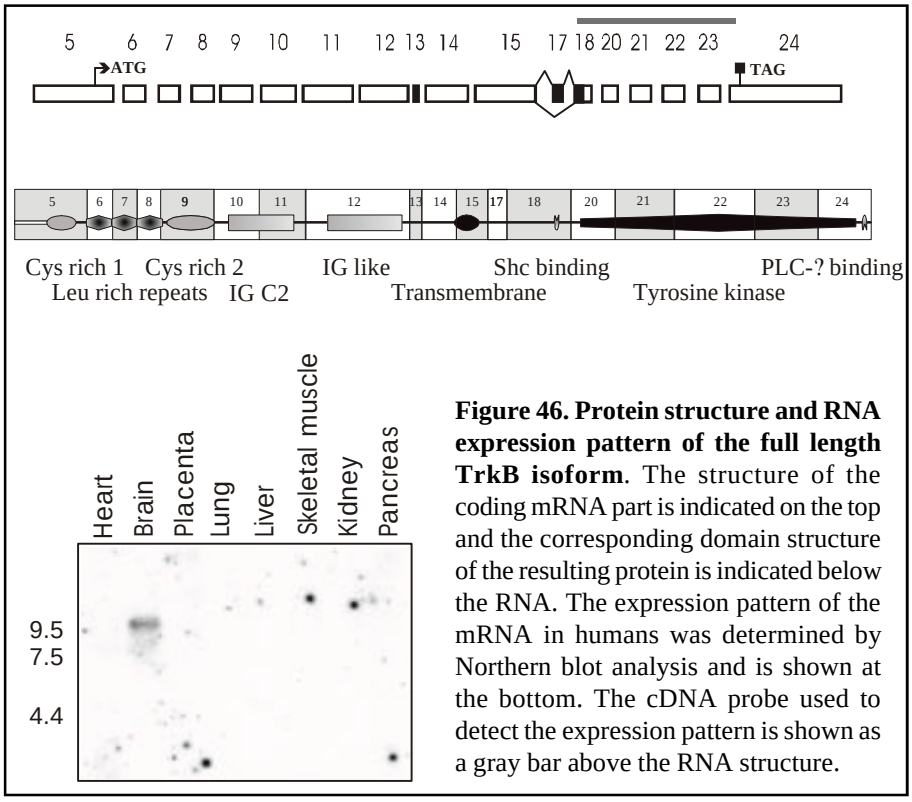
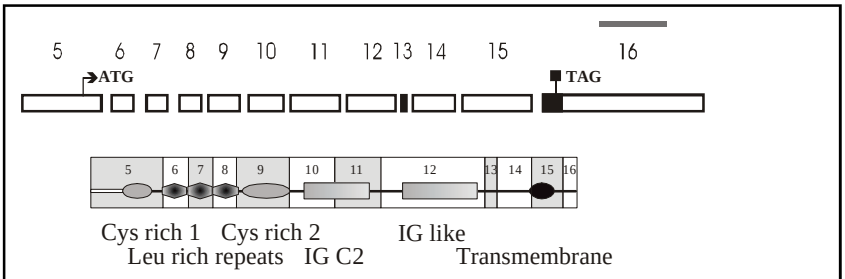


Figure 46. Protein structure and RNA expression pattern of the full length TrkB isoform. The structure of the coding mRNA part is indicated on the top and the corresponding domain structure of the resulting protein is indicated below the RNA. The expression pattern of the mRNA in humans was determined by Northern blot analysis and is shown at the bottom. The cDNA probe used to detect the expression pattern is shown as a gray bar above the RNA structure.

Next, the expression pattern of the human TrkB gene was analyzed by Northern blot. Human multiple tissue RNA blot (Clontech) was probed with single stranded antisense DNA probes targeted against exons 18-24, 16 and 19. Results show that the TrkB gene generates three major isoforms: the full length



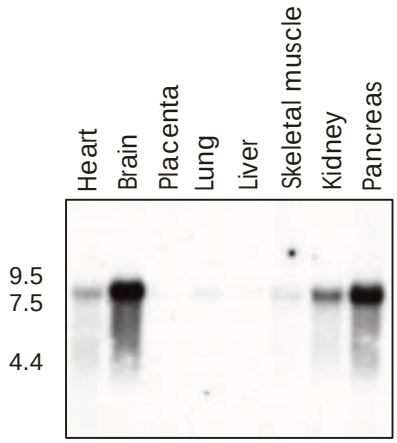
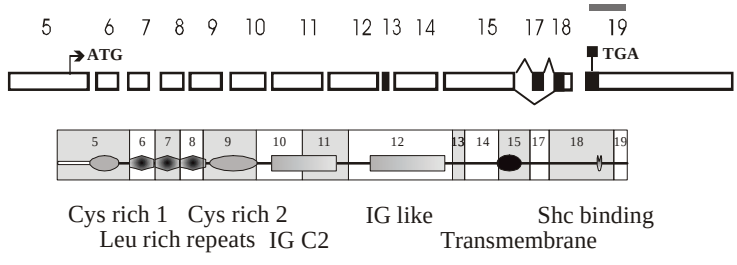


Figure 47. Protein structure and RNA expression pattern of the TrkB-T1 isoform. The structure of the coding mRNA part is indicated on the top and the corresponding domain structure of the resulting protein is indicated below the RNA. The expression pattern of the mRNA in humans was determined by Northern blot analysis and is shown at the bottom. The cDNA probe used to detect the expression pattern is shown as a gray bar above the RNA structure.

tyrosine receptor kinase (TrkB), a form lacking the catalytic domain (TrkB-T1) and a form lacking the catalytic domain but possessing a Shc site (TrkB-T-Shc). The full length TrkB form uses the stop codon and the polyadenylation site present in exon 24. It is predominantly expressed in the brain (Figure 46). The previously described TrkB-T1 form is expressed in multiple tissues. Its highest expression is in the brain, the pancreas, the kidney and the heart (Figure 47).

7.2.8. TrkB-T-Shc, a novel TrkB isoform predominantly expressed in brain

The EST database search with individual exon sequence detected an EST (accession number AL134306.1) containing a novel alternative spliced exon (exon 19). The EST clone was obtained from the German Resource center in Berlin and sequenced. The sequence showed that it contains part of exon 18 and a large single exon (exon 19) ending with a poly A tail (accession number AF410898). Since exon 19 contains a stop codon and a polyadenylation site, its usage leads to a truncated protein (accession number AF410901 and accession number



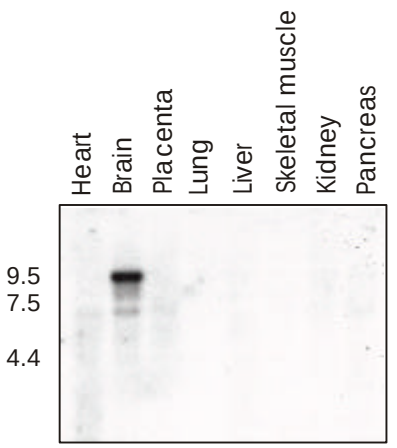


Figure 48. Protein structure and RNA expression pattern of the TrkB-T-Shc isoform. The structure of the coding mRNA part is indicated on the top and the corresponding domain structure of the resulting protein is indicated below the RNA. The expression pattern of the mRNA in humans was determined by Northern blot analysis and is shown at the bottom. The cDNA probe used to detect the expression pattern is shown as a gray bar above the RNA structure.

AF410900), similar to the TrkB-T1 isoform. In contrast to the TrkB-T1 isoform, this isoform results in a protein that retains the Shc binding site (Figure 48). Northern blot analysis reveals that this new isoform is predominantly expressed in the brain (Figure 48). Therefore, there are two different classes of tyrosine kinase deficient TrkB receptors in human brain. The TrkB-T1 form, which is also found in other tissues, cannot signal to Shc and the TrkB-T-Shc, which is brain specific and could bind to Shc.

```
mouse:  tttcatttaccaagtccagagatggtttgcaagattcttagagcaatgtgaacagagctg
        |||||||||| |||||||||||||||||| |||| | |||||||| ||||||| ||||
human:  tttcatttacaaagtccagagatggtttgcaggattgt-agagcaatctgaacagggctg

mouse: aatcgctcctgagcactttccaatagaaagtctgtctactttattgcagGGCCCAGAGGT
       |||  |||| ||||||||||||||| ||||||||||||||||||||||||||||||||
human: aattcctcccgagcactttccaatagcaagtctgtctactttattgcagGGCCCAGAGGT

mouse: TCCCCCAAGACCGCCTGATAATAATTTGGTATTTGGAGGCTCCTGCGTCACTGCAGGAAC
       |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
human: TCCCCCAAGACCGCCTGATAATAATTTGGTATTTGGAGGCTCCTGTGTCACTGCAGGAAC

mouse: CAAAGGAGGCTAAATCCAAGGCTGATGGAGGAGAAGAGTTCTATGGTTATCTGCAAATTC
       |||||||||||||||| | ||||||||||||||||||||||||||||||||||||||||
human: TAAAGGAGGCTAAATCCATGCCTGATGGAGGAGAAGAGTTCTATGGTTATCTGCAAATTC

mouse: TGGCCAGACAGCATCTGGAAGTCACTCCTTAGCTTCCATAA-CTAGCCAAGCAAGAAGTT
       |||||||||| ||||| || |||||||||||||||||||| ||||||||||||||||||
human: TGGCCAGACAACATCTTGACGTCACTCCTTAGCTTCCATAACCTAGCCAAGCAAGAAGTT

mouse: GCCTTCCCAGGACAAAACAGAGTGCTCTAATGACTAAGCCCTCAAAGTGCTGTGTCATTT
       ||||| ||| |||||| ||| |||||||||||||| |||||||||| || || || ||
human: GCCTTTCCAAGACAAAGCAGTGTGCTCTAATGACTAACCCCTCAAAGTACTATGCCACTT

mouse: AAACTATAGA-CCATCTCCTCGATCAATCAGGATGGCAAAATGGAGC
       ||||||||| |||||||||||||||||||||||||| |||||||
human: TAACTATAGACCCATCTCCTCGATCAATCAGGATGGCAAGATGGAGC
```

Figure 49. Alignment between mouse and human sequences in the beginning of exon 19. The intron sequence is in lower-case, the exon sequence is in upper-case and the coding part of the exon is shown in bold.

7.2.9. TrkB-T-Shc is present in mouse

To determine whether the newly found exons 17 and 19 are present in rodents, the ENSEMBL mouse trace repository was screened with the human sequences (ENSEMBL). Several traces were detected with exon 19 sequence and were assembled using the Staden Package (http://www.mrc-lmb.cam.ac.uk/pubseq/). The alignment with the human TrkB sequence shows high similarity. This simmilarity is not only confined to the coding sequence but extends into the 3' UTR and the upstream intron (Figure 49). RT-PCR analysis showed brain specific expression of exon 19 (Figure 50). Interestingly it is not detectable in primary astrocytes but is present in cultured neurons from embryonic

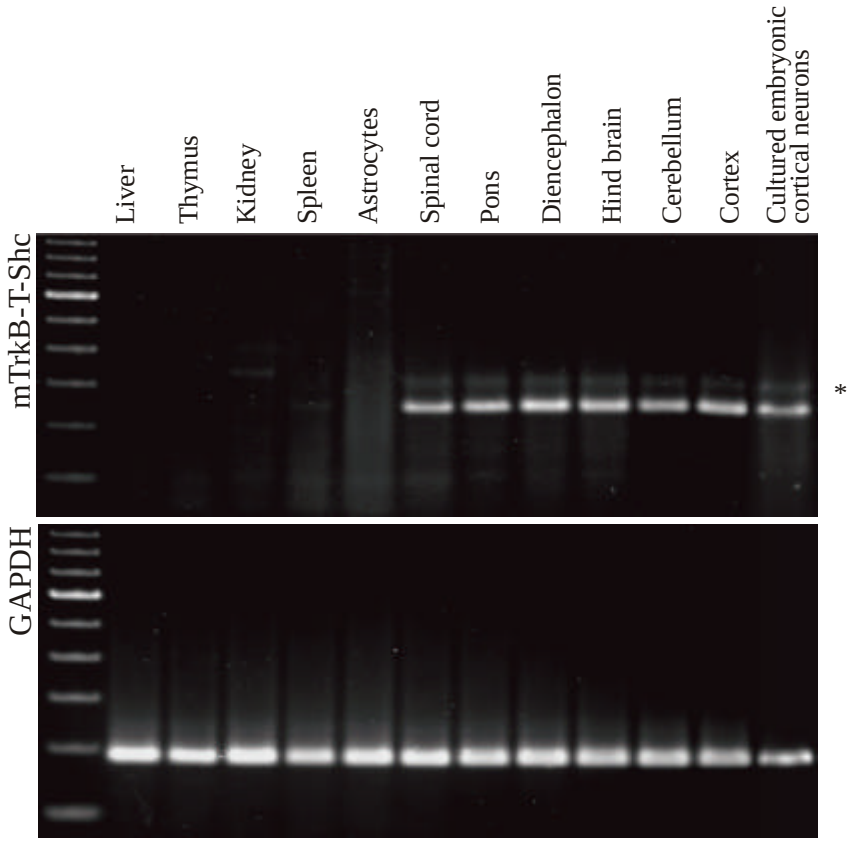


Figure 50. RT-PCR analysis of TrkB-T-Shc expression in mouse (top). The isoform is expressed in primary neurons from embryonic cortex and in cortex, cerebellum, hind brain, diencephalon, pons and spinal cord. It is absent from liver, thymus, kidney, spleen and cultured primary astrocytes. GAPDH was amplified from the same RNA samples to show that approximately equal amounts of RNA were used for the RT-PCR (bottom). The band marked with an asterisk is a PCR artifact

cortex (Figure 50). The RT-PCR and the trace repository screens show no evidence of a mouse exon that corresponds to the human exon 17.

7.2.10. TrkB-T-Shc is not phosphorylated by full length TrkB

In order to verify the role of the truncated TrkB-T-Shc isoform in the established TrkB signaling pathway, its ability to interact with the full length receptor was studied. Using fluorescence microscopy EGFP-tagged TrkB-T-Shc was shown to localize in the cell membrane (Figure 51). A significant amount of the protein is also localized in cytoplasmic compartments. This is in agreement with the data that Trk receptors are distributed between the cell membrane and cytoplasmic vesicles

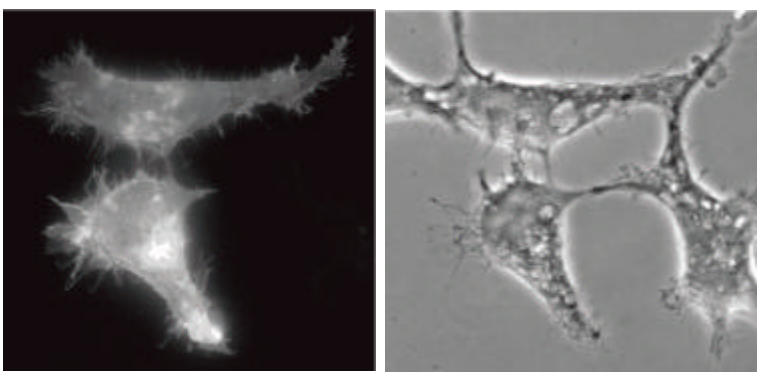


Figure 51. The TrkB-T-Shc isoform expressed on the cell membrane. Expression construct for TrkB-T-Shc/EGFP fusion was transfected into HEK A293 cells. The localization of the expressed protein was observed using fluorescent microscopy (left). Phase contrast view of the same cells is shown on the right.

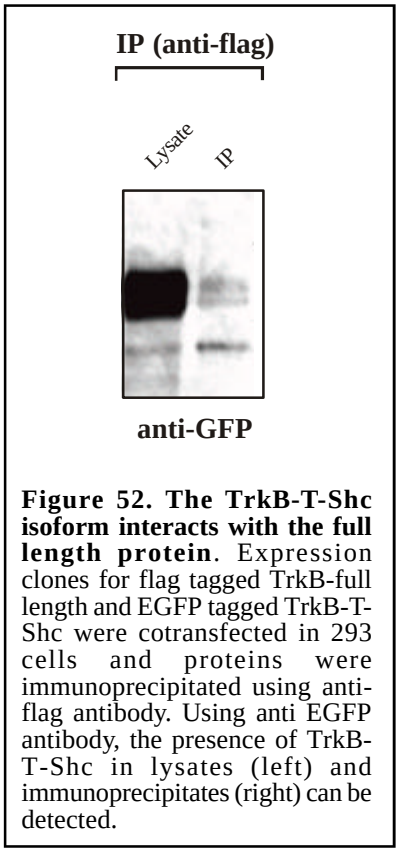


Figure 52. The TrkB-T-Shc isoform interacts with the full length protein. Expression clones for flag tagged TrkB-full length and EGFP tagged TrkB-T-Shc were cotransfected in 293 cells and proteins were immunoprecipitated using anti-flag antibody. Using anti EGFP antibody, the presence of TrkB-T-Shc in lysates (left) and immunoprecipitates (right) can be detected.

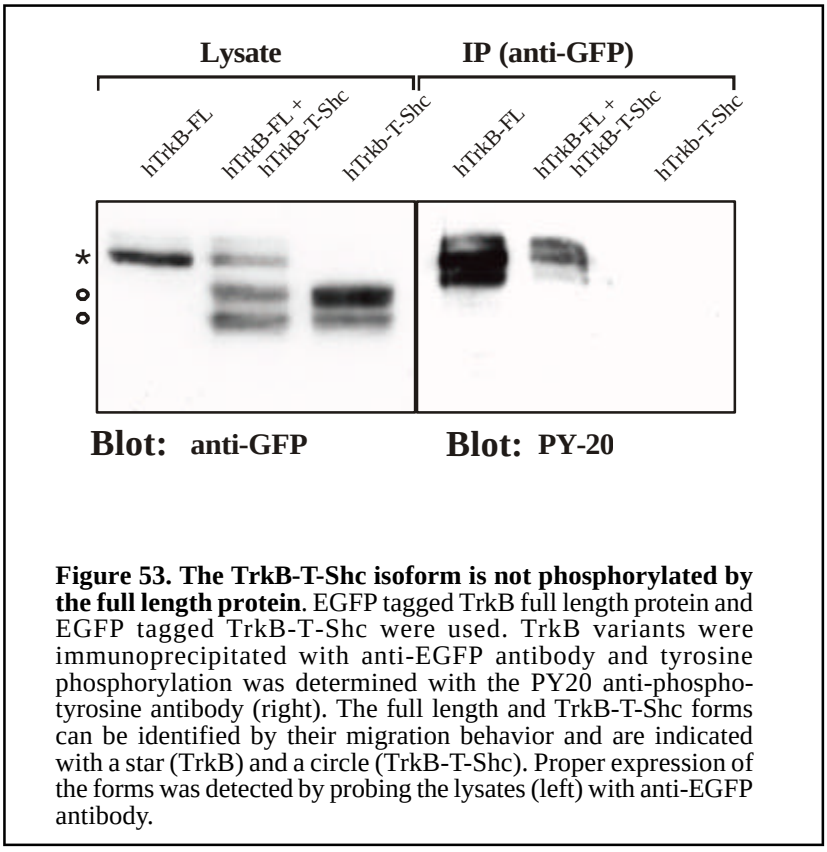


Figure 53. The TrkB-T-Shc isoform is not phosphorylated by the full length protein. EGFP tagged TrkB full length protein and EGFP tagged TrkB-T-Shc were used. TrkB variants were immunoprecipitated with anti-EGFP antibody and tyrosine phosphorylation was determined with the PY20 anti-phosphotyrosine antibody (right). The full length and TrkB-T-Shc forms can be identified by their migration behavior and are indicated with a star (TrkB) and a circle (TrkB-T-Shc). Proper expression of the forms was detected by probing the lysates (left) with anti-EGFP antibody.

and are rapidly internalized upon ligand binding. To determine whether TrkB-T-Shc binds the full length TrkB, flag-tagged full length TrkB was cotransfected with EGFP tagged TrkB-T-Shc. Immunoprecipitation with an anti-Flag-antibody followed by Western blot analysis with anti-EGFP antibody revealed that both forms coimmunoprecipitate (Figure 52). As the Shc site of the full length receptor is phosphorylated upon its activation, it was possible that the Shc site of TrkB-T-Shc can be phosphorylated by the catalytic variant. To test this human full length TrkB receptor, cDNA was cotransfected with TrkB-T-Shc expression constructs. The expressed protein was immunoprecipitated with anti-EGFP antibody and the immunoprecipitates were analyzed for tyrosine phosphorylation. Whereas the full length receptor was tyrosine phosphorylated, no phosphorylation of the TrkB-T-Shc variant is detectable (Figure 53).

7.3. Discussion

7.3.1. General structure of the TrkB gene

Approximately 100 kbp of the TrkB gene were sequenced and the latest assembly of the human genome estimates the gene size to be about 350 kbp. However, this sequence contains a

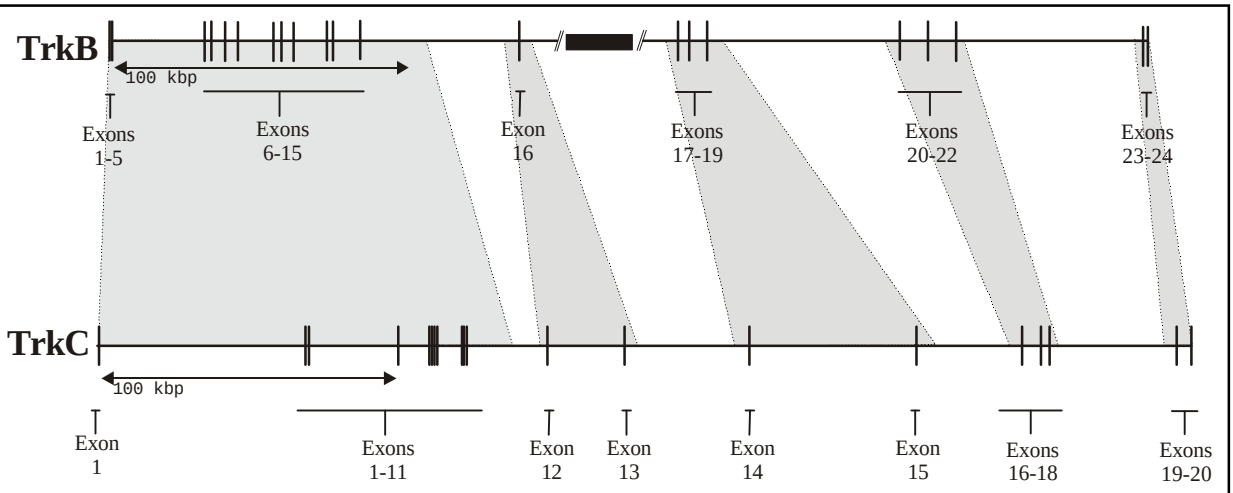


Figure 54. Comparison of the TrkB and TrkC gene structures. The structures of the TrkB and TrkC genes are drawn to scale. Exons are indicated as vertical lines, introns are shown as horizontal lines. The ChAB4 repeat is indicated by a black box. The gap in the sequence surrounding the ChAB4 repeat is indicated by a double tilted line. The shading shows the corresponding regions of the two genes.

clone gap. This works shows the presence of a ChAB4 repeat at the gap location (Figure 40). ChAB4 repeats are characteristic for humans and chimpanzee. Only a few copies of these repeats are found in gorilla and orangutan. ChAB4 repeats have been estimated to be about 240 kbp in size. This suggest a gene size of at least 590 kbp. One large intron, located between exons 16 and 17, accounts for at least half its size. This is well above the average for the human genes that have a median of 14 kbp and a mean of 27 kbp in their size distribution. The presence of the ChAB4 repeat is another unusual feature, as there are only about 50 such repeats present on ten different human chromosomes (Kehrer-Sawatzki et al., 1998). The TrkB gene consists of 24 exons which can be grouped into six clusters. The analysis of the 5' UTR showed that the gene uses at least nine different start sites. The promoter is located in a CpG island. No apparent initiator elements could be detected. The combination of alternative exon and promoter usage generates TrkB mRNAs with considerable variation in their 5' UTRs. The alternative promoters do not create protein isoforms. The exact function of this divergence remains to be determined, but alterations in the 5' UTRs are generally associated with differences in translation efficiency (Gray and Wickens, 1998). The overall gene structure of the human TrkB gene is similar to the gene structure of TrkC (Figure 54) (Ichaso et al., 1998). Furthermore, similarly to the mouse TrkB gene (Barettino et al., 1999) and the human TrkC gene (Ichaso et al., 1998), it uses TATA-less promoters located in CpG islands.

7.3.2. Alternative splicing variants

A combination of alternative splicing, alternative polyadenylation and alternative promoter usage leads to 100 isoforms that can encode 10 different proteins. The RT-PCR analysis reveals that

most of these alternative exons are not abundantly used, suggesting that they might fulfill specialized roles only in small cell populations or during certain developmental stages. Most alternative spliced exons of the TrkB gene are flanked by splice sites following the human consensus, indicating that their alternative use is due to repression mechanisms that remain to be elucidated. The only exon with suboptimal splice sites is exon 4 that uses a GC instead of the canonical GT at its 5' splice site. A recent survey found 162 GC-AG introns derived from 145 genes (Thanaraj and Clark, 2001). The search in the working draft of the human genome found 832 GC-AG introns, which corresponds to about 1% of all annotated introns. The exact splicing mechanism of those introns remains to be determined. It could involve either optimized base pairing at the GC 5' splice site to U1 snRNA, or different mechanism, such as changing GC to GU by an editing event. The large size of the TrkB introns makes it likely that introns will be removed stepwise, as found for the ultrabithorax genes (Hatton et al., 1998). It could also be used to regulate the alternative splicing pattern of the gene, as three of the largest introns (15, 16 and 19) separate the alternative exons that generate the three major isoforms.

7.3.3. Novel isoform

Similarly to TrkC (Lamballe et al., 1991; Menn et al., 1998; Tsoulfas et al., 1993), the TrkB gene generates two types of noncatalytic isoforms, which are conserved in human and rodents. The usage of exon 19 results in a new truncated TrkB isoform, TrkB-T-Shc. In contrast to the previously described truncated isoforms TrkB-T1 and TrkB-T2, this variant has a Shc binding site. Another difference is its expression pattern. The formerly described TrkB-T1 variant is expressed in most tissues, but the novel TrkB-T-Shc isoform is expressed only in brain and is neuron specific. Sequence analysis shows that this isoform is highly conserved between humans and mouse. This indicates that it could play an important role in TrkB signal transduction. Surprisingly, the full length receptor could bind to the TrkB-T-Shc isoform, but was unable to phosphorylate it. Overexpressed TrkB-T-Shc is present in the plasma membrane and similarly to the full length TrkB receptor accumulates inside the cell. Since the TrkB-T-Shc isoform is present only in neurons, it might regulate the function of the catalytic form of the TrkB receptor. It could act as a negative regulator in regards to the full length receptor, the same way the TrkB-T1 isoform does. Alternatively, it could have a specific function in the brain allowing for regulation of the TrkB receptor by other kinases (Figure 55). This hypothesis is supported by the finding that TrkC paralogues of TrkB-T-Shc can promote neuronal differentiation in cooperation with $p75^{NTR}$ (Hapner et al., 1998). Trk-T-Shc has been shown recently

to relocalize the full length TrkB receptor from the cell interior to the cell surface (Haapasalo et al 2002 submitted for publication). This suggest that it can play a crucial role in BDNF signaling by bringing the full length TrkB receptor to the cell membrane where it can interact with its ligand. Furthermore, the existence of a brain specific truncated isoform raises the question about the function of the TrkB-T1 isoform that is present in most tissues.

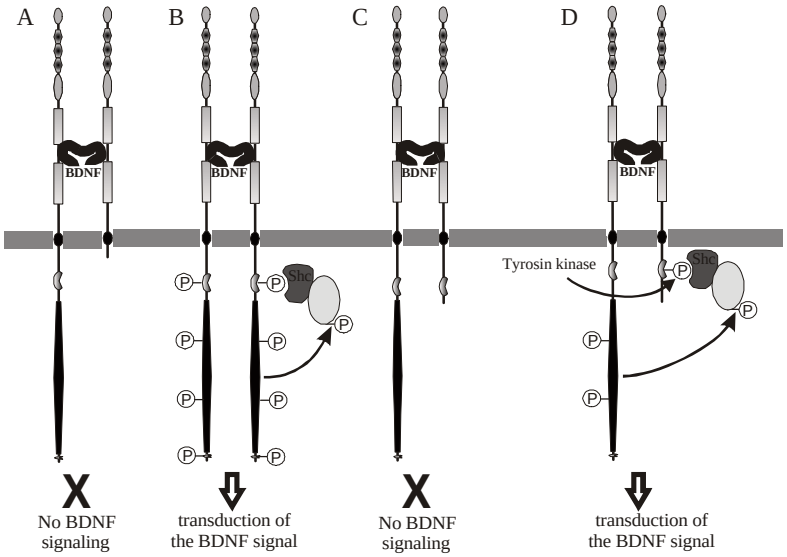


Figure 55. Possible roles for the TrkB-T-Shc isoform in BDNF signaling. (A) The TrkB-T1 truncated receptor blocks BDNF signaling from the full length receptor. (B) BDNF induces dimerization and rapid autophosphorylation of the TrkB receptor. It can then phosphorylate targets bound brought to it by the Shc adaptor molecule. (C) By analogy with the TrkB-T1, TrkB-T-Shc can block the signaling from the full length receptor. Alternatively signaling trough it can be enabled by other tyrosine kinases that phosphorylate the Shc binding site, allowing for crosstalk between different signaling pathways (D).

8. Conclusion

Alternative pre-mRNA splicing plays a major role in generating protein diversity in higher eukaryotes. The second part of this work, focusing on the gene structure and alternative splicing of the human TrkB gene, illustrates how a single gene can generate proteins with different tissue distribution and properties. It also provides insight into how a complex regulatory and development patterns can be generated from relatively small number of genes. The vast size of the human TrkB gene also exposes the limitations of the current techniques for studying alternative splicing regulation. It makes virtually impossible the creation of model minigene systems as well as the deletion and mutation analysis. In addition TrkB's tightly controlled tissue specific expression further impedes the research on the molecular mechanisms controlling its splicing. In this respect, the study of the human tra2-beta both as a model system for alternative splicing regulation and as splicing regulator provides some clues how this difficulties can be overcome. This work is unique in providing a new approach in studying the factors that regulate alternative splicing and their targets. It shows that it is possible to derive the binding sequence of a splicing factor by identifying the target transcripts and performing comparative analysis of their sequences. When taken to a large scale, using DNA chips to monitor the alternative splicing of multiple genes in combination with Gibbs sequence analysis to derive the common sequence patterns, it can provide a powerful approach in studying alternative splicing.

9. References

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell *89*, 403-412.

Adams, D. J., Weyden, L. v. d., Mayeda, A., Stamm, S., Morris, B. J., and Rasko, J. E. J. (2001). ZNF265-A novel spliceosomal protein able to induce alternative splicing. J Cell Biol, in press.

Allen, S. J., Dawbarn, D., Eckford, S. D., Wilcock, G. K., Ashcroft, M., Colebrook, S. M., Feeney, R., and MacGowan, S. H. (1994). Cloning of a non-catalytic form of human trkB and distribution of messenger RNA for trkB in human brain. Neuroscience *60*, 825-834.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Amarasinghe, A. K., MacDiarmid, R., Adams, M. D., and Rio, D. C. (2001). An in vitro-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element. Rna *7*, 1239-1253.

Amrein, H., Hedley, M. L., and Maniatis, T. (1994). The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. Cell *76*, 735-746.

Arao, Y., Kuriyama, R., Kayama, F., and Kato, S. (2000). A nuclear matrix associated factor, SAF-B, interacts with specific isoforms of AUF/hnRNP D. Arch Biochem Biophys *380*, 228-236.

Aravind, L., Watanabe, H., Lipman, D. J., and Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proc Natl Acad Sci U S A *97*, 11319-11324.

Ars, E., Serra, E., de la Luna, S., Estivill, X., and Lazaro, C. (2000a). Cold shock induces the insertion of a cryptic exon in the neurofibromatosis type 1 (NF1) mRNA. Nucl Acids Res *28*, 1307-1312.

Ars, E., Serra, E., Garcia, J., Kruyer, H., Gaona, A., Lazaro, C., and Estivill, X. (2000b). Mutations affecting mRNA splicing are the most common molecular defects in patients with

neurofibromatosis type 1. Hum Mol Genet *9*, 237-247.

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

Barettino, D., Pombo, P. M. G., Espliguero, G., and Rodriguez-Pena, A. (1999). The mouse neurotrophin receptor trkB gene is transcribed from two different promotors. Biochim Biophys Acta *1446*, 24-34.

Beil, B., Screaton, G., and Stamm, S. (1997). Molecular cloning of htra2-beta-1 and htra2-beta-2, two human homologs of tra-2 generated by alternative splicing. DNA Cell Biol *16*, 679-690.

Benham, C., Kohwi-Shigematsu, T., and Bode, J. (1997). Stress-induced Duplex DNA Destabilization in Scaffold/Matrix Attachment Regions. JMB *274*, 181-196.

Benne, R. (1996). RNA editing: how a message is changed. Curr Opin Genet Dev *6*, 221-231.

Bentley, D. (1999). Coupling RNA polymerase II transcription with pre-mRNA processing. Curr Opin Cell Biol *11*, 347-351.

Bibel, M., and Barde, Y. A. (2000). Neurotrophins: key regulators of cell fate and cell shape in the vertebrate nervous system. Genes Dev *14*, 2919-2937.

Biedler, J. L., Roffler-Tarlov, S., Schachner, M., and Freedman, L. S. (1978). Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. Cancer Res *38*, 3751-3757.

Biffo, S., Offenhauser, N., Carter, B. D., and Barde, Y. A. (1995). Selective binding and internalisation by truncated receptors restrict the availability of BDNF during development. Development *121*, 2461-2470.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci *25*, 106-110.

Blencowe, B. J., Baurén, G., Eldridge, A. G., Issner, R., Nickerson, J. A., Rosonina, e., and Sharp, P. A. (2000). The SRm160/300 splicing coactivator subunits. RNA *6*, 111-120.

Blencowe, B. J., Issner, R., Nickerson, J. A., and Sharp, P. A. (1998). A coactivator of pre-mRNA splicing. Genes Dev *12*, 996-1009.

Blencowe, B. J., Nickerson, J. A., Issner, R., Penman, S., and Sharp, P. A. (1994). Association of nuclear matrix antigens with exon-containing splicing complexes. J Cell Biol *127*,

593-607.

Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. Nature *411*, 355-365.

Bode, J., Benham, C., A., K., and Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements). Crit Rev Eukar Gene Exp *10*, 73-90.

Buckanovich, R. J., Posner, J. B., and Darnell, R. B. (1993). Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. Neuron *11*, 657-672.

Caceres, J. F., Screaton, G. R., and Krainer, A. R. (1998). A specific subset of SR proteins shuttles continously between the nucleus and the caytoplasm. Genes Dev *12*, 55-66.

Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999). The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. Rna *5*, 468-483.

Celotto, A. M., and Graveley, B. R. (2001). Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated. Genetics *159*, 599-608.

Chabot, B., Blanchett, M., Lapierre, I., and La Branche, H. (1997). An intron element modulating the splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNp A1. Mol Cell Biol *17*, 1776-1786.

Chalfant, C. E., Mischak, H., Watson, J. E., Winkler, B. C., Goodnight, J., Farese, R. V., and Cooper, D. R. (1995). Regulation of alternative splicing of protein kinase C beta by insulin. J Biol Chem *270*, 13326-13332.

Chalfant, C. E., Watson, J. E., Bisnauth, L. D., Kang, J. B., Patel, N., Obeid, L. M., Eichler, D. C., and Cooper, D. R. (1998). Insulin regulates protein kinase CbetaII expression through enhanced exon inclusion in L6 skeletal muscle cells. A novel mechanism of insulin- and insulin-like growth factor-i-induced 5' splice site selection. J Biol Chem *273*, 910-916.

Chen, C., and Okayama, H. (1987). High efficiency transformation of plasmid DNA. Mol Cell Biol *7*, 2745-2752.

Chen, S., Anderson, K., and Moore, M. J. (2000). Evidence for a linear search in bimolecular 3' splice site AG selection. Proc Natl Acad Sci U S A *97*, 593-598.

Chen, T., Cote, J., Carvajal, H. V., and Richard, S. (2001). Identification of Sam68

arginine glycine-rich sequences capable of conferring nonspecific RNA binding to the GSG domain. J Biol Chem *276*, 30803-30811.

Chen, W., Kubota, S., Teramoto, T., Nishimura, Y., Yonemoto, K., and Seyama, Y. (1998). Silent nucleotide substitution in the sterol 27-hydroxylase gene (CYP 27) leads to alternative pre-mRNA splicing by activating a cryptic 5' splice site at the mutant codon in cerebrotendinous xanthomatosis patients. Biochemistry *37*, 4420-4428.

Chester, A., Scott, J., Anant, S., and Navaratnam, N. (2000). RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. Biochim Biophys Acta *1494*, 1-13.

Chiara, M. D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. Embo J *16*, 4746-4759.

Chiodi, I., Biggiogera, M., Denegri, M., Corioni, M., Weighardt, F., Cobianchi, F., Riva, S., and Biamonti, G. (2000). Structure and dynamics of hnRNP-labelled nuclear bodies induced by stress treatments. J Cell Sci *113 (Pt 22)*, 4043-4053.

Chua, K., and Reed, R. (2001). An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. Mol Cell Biol *21*, 1509-1514.

Collins, C. A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. Genes Dev *13*, 1970-1982.

Collins, C. A., and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? Nat Struct Biol *7*, 850-854.

Collins, C. A., and Guthrie, C. (2001). Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. Rna *7*, 1845-1854.

Colwill, K., Feng, L. L., Gish, G. D., Cáceres, J. F., Pawson, T., and Fu, X. D. (1996a). SRKP1 and Clk/Sty protein kinases show distinct substrate specificities for serine/arginine-rich splicing factors. J Biol Chem *271*, 24569-24575.

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

Cooper, T. A. (1999). In vivo SELEX in vertebrate cells. Methods Mol Biol *118*, 405-417.

Coovert, D. D., Le, T. T., McAndrew, P. E., Strasswimmer, J., Crawford, T. O., Mendell, J. R., Coulson, S. E., Androphy, E. J., Prior, T. W., and Burghes, A. H. (1997). The survival motor neuron protein in spinal muscular atrophy. Hum Mol Genet *6*, 1205-1214.

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

Cunningham, T. P., Hagan, J. P., and Grabowski, P. J. (1995). Reconstitution of exon-bridging activity with purified U2AF and U1 snRNP components. Nucleic Acids Symp Ser, 218-219.

Daoud, R., Da Penha Berzaghi, M., Siedler, F., Hubener, M., and Stamm, S. (1999). Activity-dependent regulation of alternative splicing patterns in the rat brain. Eur J Neurosci *11*, 788-802.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

Das, S., Levinson, B., Whitney, S., Vulpe, C., Packman, S., and Gitschier, J. (1994). Diverse mutations in patients with Menkes disease often lead to exon skipping. Am J Hum Genet *55*, 883-889.

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

De Meirleir, L., Lissens, W., Benelli, C., Ponsot, G., Desguerre, I., Marsac, C., Rodriguez, D., Saudubray, J. M., Poggi, F., and Liebaers, I. (1994). Aberrant splicing of exon 6 in the pyruvate dehydrogenase-E1 alpha mRNA linked to a silent mutation in a large family with Leigh's encephalomyelopathy. Pediatr Res *36*, 707-712.

Denhez, F., and Lafyatis, R. (1994). Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the Drosophila splicing regulator, suppressor-of-white-apricot. J Biol Chem *269*, 16170-16179.

Derry, J. J., Richard, S., Valderrama Carvajal, H., Ye, X., Vasioukhin, V., Cochrane, A. W., Chen, T., and Tyner, A. L. (2000). Sik (BRK) phosphorylates Sam68 in the nucleus and negatively regulates its RNA binding ability. Mol Cell Biol *20*, 6114-6126.

Di Fruscio, M., Chen, T., and Richard, S. (1999). Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. Proc Natl Acad

Sci U S A *96*, 2710-2715.

Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucl Acids Res *11*, 1475-1489.

Dobkin, C., and Bank, A. (1983). A nucleotide change in IVS 2 of a beta-thalassemia gene leads to a cryptic splice not at the site of the mutation. Prog Clin Biol Res *134*, 127-128.

Dobkin, C., Pergolizzi, R. G., Bahre, P., and Bank, A. (1983). Abnormal splice in a mutant human beta-globin gene not at the site of a mutation. Proc Natl Acad Sci U S A *80*, 1184-1188.

D'Souza, I., Poorkai, P., Hong, M., Nochlin, D., Lee, V. M.-Y., Bird, T. D., and Schellenberg, G. D. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc Natl Acad Sci USA *96*, 5598-5603.

D'Souza, I., and Schellenberg, G. D. (2000). Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. J Biol Chem *275*, 17700-17709.

Du, C., McGuffin, E., Dauwalder, B., Rabinow, L., and Mattox, W. (1998). Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in drosophila. Mol Cell *2*, 741-750.

Duncan, P. I., Stojdl, D. F., Marius, R. M., Scheit, K. H., and Bel, J. C. (1998). The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing. Exp Cell Res *241*, 300-308.

Eide, F. F., Vining, E. R., Eide, B. L., Zang, K., Wang, X. Y., and Reichardt, L. F. (1996). Naturally occurring truncated trkB receptors have dominant inhibitory effects on brain-derived neurotrophic factor signaling. J Neurosci *16*, 3123-3129.

Eldridge, A. G., Li, Y., Sharp, P. A., and Blencowe, B. J. (1999). The SRm160/300 splicing coactivator is required for exon-enhancer function. Proc Natl Acad Sci U S A *96*, 6125-6130.

Elliott, D. J., Ma, K., Kerr, S. M., Thakrar, R., Speed, R., Chandley, A. C., and Cooke, H. (1996). An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. Hum Mol Genet *5*, 869-874.

ENSEMBL. ENSEMBL traces repository (http://trace.ensembl.org) holds sequence traces provided by the Mouse sequencing consorcium (http://www.sanger.ac.uk/Info/Press/001006.shtml).

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

Eperon, I. C., Makarova, O. V., Mayeda, A., Munroe, S. H., Caceres, J. F., Hayward, D. G., and Krainer, A. R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. Mol Cell Biol *20*, 8303-8318.

Fairbrother, W. G., and Chasin, L. A. (2000). Human genomic sequences that inhibit splicing. Mol Cell Biol *20*, 6816-6825.

Fong, N., and Bentley, D. L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev *15*, 1783-1795.

Fu, X. D., and Maniatis, T. (1992). The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3' splice site. Proc Natl Acad Sci U S A *89*, 1725-1729.

Fujimaru, M., Tanaka, A., Choeh, K., Wakamatsu, N., Sakuraba, H., and Isshiki, G. (1998). Two mutations remote from an exon/intron junction in the beta- hexosaminidase beta-subunit gene affect 3'-splice site selection and cause Sandhoff disease. Hum Genet *103*, 462-469.

Garner, A. S., Menegay, H. J., Boeshore, K. L., Xie, X.-Y., Voci, J. M., Johnson, J. E., and Large, T. H. (1996). Expression of TrkB receptor isoforms in the developing avian visaul system. J Neurosci *16*, 1740-1752.

Ge, H., Si, Y., and Wolffe, A. P. (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. Mol Cell *2*, 751-759.

Ghosh, S., and Garcia-Blanco, M. A. (2000). Coupled in vitro synthesis and splicing of RNA polymerase II transcripts. Rna *6*, 1325-1334.

Gondran, P., and Dautry, F. (1999). Regulation of mRNA splicing and transport by the tyrosine kinase activity of src. Oncogene *18*, 2547-2555.

Grabowski, P. J. (1990). Isolation and analysis of splicing complexes. Methods Enzymol *184*, 319-327.

Gray, N. K., and Wickens, M. (1998). Control of translation initiation in animals. Annu Rev Cell Dev Biol *14*, 399-458.

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

Haapasalo, A., Koponen, E., Hoppe, E., Wong, G., and Castren, E. (2001). Truncated trkB.T1 is dominant negative inhibitor of trkB.TK+-mediated cell survival. Biochem Biophys Res Commun *280*, 1352-1358.

Hackett, S. F., Friedman, Z., Freund, J., Schoenfeld, C., Curtis, R., DiStefano, P. S., and Campochiaro, P. A. (1998). A splice variant of trkB and brain-derived neurotrophic factor are co-expressed in retinal pigmented epithelial cells and promote differentiated characteristics. Brain Res *789*, 201-212.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser *41*, 95-98.

Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. RNA *4*, 430-444.

Hapner, S. J., Boeshore, K. L., Large, T. H., and Lefcort, F. (1998). Neural differentiation promoted by truncated trkC receptors in collaboration with p75(NTR). Dev Biol *201*, 90-100.

Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A., and Stamm, S. (1999). The interaction and colocalization of Sam68 with the splicing- associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). Mol Biol Cell *10*, 3909-3926.

Hartmann, A. M., Rujescu, D., Giannakouros, T., Nikolakaki, E., Goedert, M., Mandelkow, E. M., Gao, Q. S., Andreadis, A., and Stamm, S. (2001). Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors. Mol Cell Neurosci *18*, 80-90.

Hasegawa, M., Smith, M. J., Iijima, M., Tabira, T., and Goedert, M. (1999). FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. FEBS lett *443*, 93-96.

Hasegawa, Y., Kawame, H., Ida, H., Ohashi, T., and Eto, Y. (1994). Single exon mutation in arylsulfatase A gene has two effects: loss of enzyme activity and aberrant splicing. Hum Genet *93*, 415-420.

Hatton, A. R., Subramaniam, V., and Lopez, A. J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. Mol Cell *2*, 787-796.

Hertel, K. J., and Maniatis, T. (1998). The function of multisite splicing enhancers. Mol Cell *1*, 449-455.

Hertel, K. J., and Maniatis, T. (1999). Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. Proc Natl Acad Sci U S A *96*, 2651-2655.

Hide, W. A., Babenko, V. N., van Heusden, P. A., Seoighe, C., and Kelso, J. F. (2001). The contribution of exon-skipping events on chromosome 22 to protein coding diversity. Genome Res *11*, 1848-1853.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

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

Hoffmeyer, S., Nurnberg, P., Ritter, H., Fahsold, R., Leistner, W., Kaufmann, D., and Krone, W. (1998). Nearby stop codons in exons of the neurofibromatosis type 1 gene are disparate splice effectors. Am J Hum Genet *62*, 269-277.

Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J., and Wirth, B. (2000). Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). Proc Natl Acad Sci U S A *97*, 9618-9623.

Huang, X. (1992). A contig assembly program based on sensitive detection of fragment overlaps. Genomics *14*, 18-25.

Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., *et al.* (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature *393*, 702-705.

Ichaso, N., Rodriguez, R. E., Martin-Zanca, D., and Gonzalez-Sarmiento, R. (1998). Genomic characterization of the human trkC gene. Oncogene *17*, 1871-1875.

Iijima, M., Tabira, T., Poorkaj, P., Schellenberg, G. D., Trojanowski, J. Q., Lee, V. M., Schmidt, M. L., Takahashi, K., Nabika, T., Matsumoto, T., *et al.* (1999). A distinct familial

presenile dementia with a novel missense mutation in the tau gene. Neuroreport *10*, 497-501.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L. E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell *106*, 607-617.

Jablonka, S., Schrank, B., Kralewski, M., Rossoll, W., and Sendtner, M. (2000). Reduced survival motor neuron (Smn) gene dose in mice leads to motor neuron degeneration: an animal model for spinal muscular atrophy type III. Hum Mol Genet *9*, 341-346.

Jackson, D. A., and Cook, P. R. (1995). The structural basis of nuclear function. Int Rev Cytol *162A*, 125-149.

Jackson, D. A., Iborra, F. J., Manders, E. M., and Cook, P. R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Mol Biol Cell *9*, 1523-1536.

Jacobsen, M., Schweer, D., Ziegler, A., Gaber, R., Schock, S., Schwinzer, R., Wonigeit, K., Lindert, R. B., Kantarci, O., Schaefer-Klein, J., *et al.* (2000). A point mutation in PTPRC is associated with the development of multiple sclerosis. Nat Genet *26*, 495-499.

Jamison, S. F., Pasman, Z., Wang, J., Will, C., Luhrmann, R., Manley, J. L., and Garcia-Blanco, M. A. (1995). U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. Nucleic Acids Res *23*, 3260-3267.

Jiang, Z.-H., and Wu, J. Y. (1999). Alternative splicing and programmed cell death. Proc Soc Exp Biol Med *220*, 64-72.

Jin, Y., Dietz, H. C., Montgomery, R. A., Bell, W. R., McIntosh, I., Coller, B., and Bray, P. F. (1996). Glanzmann thrombasthenia. Cooperation between sequence variants in cis during splice site selection. J Clin Invest *98*, 1745-1754.

Jumaa, H., Guenet, J.-L., and Nielsen, P. J. (1997). Regulated expression and RNA processing of transcripts from the Srp20 splicing factor gene during the cell cycle. Mol Cell Biol *17*, 3116-3124.

Jumaa, H., and Nielsen, P. J. (1997). The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. EMBO J *16*, 5077-5085.

Jumaa, H., Wei, G., and Nielsen, P. J. (1999). Blastocyst formation is blocked in mouse embryos lacking the splicing factor SRp20. Curr Biol *9*, 899-902.

Kao, H. Y., and Siliciano, P. G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. Mol Cell Biol *16*, 960-967.

Kaufer, D., Friedman, A., Seidman, S., and Soreq, H. (1998). Acute stress facilitates long-lasting changes in cholinergic gene expression. Nature *393*, 373-377.

Kawano, T., Fujita, M., and Sakamoto, H. (2000). Unique and redundant functions of SR proteins, a conserved family of splicing factors, in Caenorhabditis elegans development. Mech Dev *95*, 67-76.

Kehrer-Sawatzki, H., Wohr, G., Schempp, W., Eisenbarth, I., Barbi, G., and Assum, G. (1998). Mapping of members of the low-copy-number repetitive DNA sequence family chAB4 within the p arms of human acrocentric chromosomes: characterization of Robertsonian translocations. Chromosome Res *6*, 429-435.

Kiledjian, M., and Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. Embo J *11*, 2655-2664.

Kim, E., Du, L., Bregman, D. B., and Warren, S. L. (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. J Cell Biol *136*, 19-28.

Kim, V. N., and Dreyfus, G. (2001). Nuclear mRNA binding proteins couple pre-mRNA splicing and post-splicing events. Mol Cells *12*, 1-10.

Kim, V. N., Yong, J., Kataoka, N., Abel, L., Diem, M. D., and Dreyfuss, G. (2001). The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. Embo J *20*, 2062-2068.

Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C. M., van Driel, R., Przybylski, M., and Fackelmayer, F. O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. Mol Cell Biol *20*, 7480-7489.

Kiryou-Seo, S., Matsuo, N., Wanaka, A., Ogawa, S., Tohyama, M., and Kiyama, H. (1998). A sequence-specific splicing activator, Tra2beta, is up-regulated in response to nerve injury. Mol Brain Res *62*, 220-223.

Klein, R., Conway, D., Parada, L. F., and Barbacid, M. (1990). The trkB tyrosine protein kinase gene codes for a second neurogenic receptor that lacks the catalytic kinase domain. Cell *61*, 647-656.

Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia, B. M., and Manley, J. L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-124.

Konig, H., Ponta, H., and Herrlich, P. (1998). Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. Embo J *17*, 2904-2913.

Krainer, A., Mayeda, A., Kozak, D., and Binns, G. (1991). Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U170K, and drosophila splicing regulators. Cell *66*, 383-394.

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

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.

Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu Rev Biochem *65*, 367-409.

Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol *11*, 363-371.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A *82*, 488-492.

Ladd, A. N., Charlet, N., and Cooper, T. A. (2001). The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol *21*, 1285-1296.

Lamballe, F., Klein, R., and Barbacid, M. (1991). trkC, a new member of the trk family of tyrosine protein kinases, is a receptor for neurotrophin-3. Cell *66*, 967-979.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M. J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. Embo J *20*, 4987-4997.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B.,

Cruaud, C., Millasseau, P., Zeviani, M., and et al. (1995). Identification and characterization of a spinal muscular atrophy- determining gene. Cell *80*, 155-165.

Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997). Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet *16*, 265-269.

Levine, A., and Durbin, R. (2001). A computational scan for U12-dependent introns in the human genome sequence. Nucleic Acids Res *29*, 4006-4013.

Lewin, G. R., and Barde, Y. A. (1996). Physiology of the neurotrophins. Annu Rev Neurosci *19*, 289-317.

Li, X., Shambaugh, M. E., Rottman, F. M., and Bokar, J. A. (2000). SR proteins Asf/SF2 and 9G8 interact to activate enhancer-dependent intron D splicing of bovine growth hormone pre-mRNA in vitro. Rna *6*, 1847-1858.

Liu, H. X., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet *27*, 55-58.

Liu, W., Qian, C., and Francke, U. (1997). Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nat Genet *16*, 328-329.

Llewellyn, D. H., Scobie, G. A., Urquhart, A. J., Whatley, S. D., Roberts, A. G., Harrison, P. R., and Elder, G. H. (1996). Acute intermittent porphyria caused by defective splicing of porphobilinogen deaminase RNA: a synonymous codon mutation at -22 bp from the 5' splice site causes skipping of exon 3. J Med Genet *33*, 437-438.

Longman, D., Johnstone, I. L., and Caceres, J. F. (2000). Functional characterization of SR and SR-related genes in Caenorhabditis elegans. Embo J *19*, 1625-1637.

Longman, D., McGarvey, T., McCracken, S., Johnstone, I. L., Blencowe, B. J., and Caceres, J. F. (2001). Multiple interactions between SRm160 and SR family proteins in enhancer-dependent splicing and development of C. elegans. Curr Biol *11*, 1923-1933.

Lorson, C. L., and Androphy, E. J. (2000). An exonic enhancer is required for inclusion of an essential exon in the SMA-determining gene SMN. Hum Mol Genet *9*, 259-265.

Lorson, C. L., Hahnen, E., Androphy, E. J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad

Sci USA *96*, 6307-6311.

Lu, X., Timchenko, N. A., and Timchenko, L. T. (1999). Cardiac elav-type RNA-binding protein (ETR-3) binds to RNA CUG repeats expanded in myotonic dystrophy. Hum Mol Genet *8*, 53-60.

Lui, H.-X., Chew, S. L., Cartegni, L., Zhang, M. Q., and Krainer, A. R. (1999). Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. Mol Cell Biol *20*, 1063-1071.

Lui, H.-X., Zhang, M., and Krainer, A. R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes and Dev *12*, 1998-2012.

Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. Science *293*, 1836-1839.

Lynch, K. W., and Maniatis, T. (1995). Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev *9*, 284-293.

Lynch, K. W., and Maniatis, T. (1996). Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. Genes Dev *10*, 2089-2101.

Lynch, K. W., and Weiss, A. (2000). A Model system for activation-induced alternative splicing of CD45 pre-mRNA in T cells implicates protein kinase C and Ras. Mol Cell Biol *20*, 70-80.

Lynch, K. W., and Weiss, A. (2001). A CD45 Polymorphism Associated with Multiple Sclerosis Disrupts an Exonic Splicing Silencer. J Biol Chem *276*, 24341-24347.

Markovtsov, V., Nikolic, J. M., Goldman, J. A., Turck, C. W., Chou, M. Y., and Black, D. L. (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol *20*, 7463-7479.

Maroney, P. A., Romfo, C. M., and Nilsen, T. W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol Cell *6*, 317-328.

Maruyama, K., and Sugano, S. (1994). Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. Gene *138*, 171-174.

Matsuo, N., Ogawa, S., Imai, Y., Takagi, T., Tohyama, M., Stern, D., and Wanaka, A. (1995). Cloning of a novel RNA binding polypeptide (RA301) induced by hypoxia/ reoxygenation. J Biol Chem *270*, 28216-28222.

Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P., and Konig, H. (2000). Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. J Biol Chem *275*, 35353-35360.

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.

Mattox, W., Palmer, M. J., and Baker, B. S. (1990). Alternative splicing of the sex determination gene transformer-2 is sex-specific in the germ line but not in the soma. Genes and Dev *4*, 789-805.

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

McAfee, J. G., Huang, M., Soltaninassab, S., Rech, J. E., Iyengar, S., and Lestourgeon, W. M. (1997). The packaging of pre-mRNA. In Eucaryotic mRNA processing, A. Krainer, ed. (New York, Oxford University Press Inc., New York), pp. 68-102.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D. L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306-3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

McCracken, S., Lambermon, M., and Blencowe, B. J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. Mol Cell Biol *22*, 148-160.

McGuffin, M. E., Chandler, D., Somaiya, D., Dauwalder, B., and Mattox, W. (1998). Autoregulation of transformer-2 alternative splicing is necessary for normal male fertility in Drosophila. Genetics *149*, 1477-1486.

Menn, B., Timsit, S., Calothy, G., and Lamballe, F. (1998). Differential expression of

TrkC catalytic and noncatalytic isoforms suggests that they act independently or in association. J Comp Neurol *401*, 47-64.

Meshorer, E., Erb, C., Gazit, R., Pavlovsky, L., Kaufer, D., Friedman, A., Glick, D., Ben-Arie, N., and Soreq, H. (2002). Alternative splicing and neuritic mRNA translocation under long-term neuronal hypersensitivity. Science *295*, 508-512.

Messiaen, L., Callens, T., De Paepe, A., Craen, M., and Mortier, G. (1997). Characterisation of two different nonsense mutations, C6792A and C6792G, causing skipping of exon 37 in the NF1 gene. Hum Genet *101*, 75-80.

Middlemas, D. S., Lindberg, R. A., and Hunter, T. (1991). trkB, a neural receptor protein-tyrosine kinase: evidence for a full-length and two truncated receptors. Mol Cell Biol *11*, 143-153.

Minvielle-Sebastia, L., and Keller, W. (1999). mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. Curr Opin Cell Biol *11*, 352-357.

Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. Mol Cell *3*, 697-705.

Monani, U. R., Sendtner, M., Coovert, D. D., Parsons, D. W., Andreassi, C., Le, T. T., Jablonka, S., Schrank, B., Rossol, W., Prior, T. W., *et al.* (2000). The human centromeric survival motor neuron gene (SMN2) rescues embryonic lethality in Smn(-/-) mice and results in a mouse with spinal muscular atrophy. Hum Mol Genet *9*, 333-339.

Morris, D. P., and Greenleaf, A. L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem *275*, 39935-39943.

Nagel, R. J., Lancaster, A. M., and Zahler, A. M. (1998). Specific binding of an exonic splicing enhancer by the pre-mRNA splicing factor SRp55. Rna *4*, 11-23.

Nagoshi, R. N., McKeown, M., Burtis, K. C., Belote, J. M., and Baker, S. (1988). The control of alternative splicing at genes regulating sexual differentiation in D. melanogaster. Cell *53*, 229-236.

Nakagawara, A., Liu, X. G., Ikegaki, N., White, P. S., Yamashiro, D. J., Nycum, L. M., Biegel, J. A., and Brodeur, G. M. (1995). Cloning and chromosomal localization of the human TRK-B tyrosine kinase receptor gene (NTRK2). Genomics *25*, 538-546.

Nakayama, T., and Takami, Y. (2001). Participation of Histones and Histone-Modifying

Enzymes in Cell Functions through Alterations in Chromatin Structure. J Biochem (Tokyo) *129*, 491-499.

Nakielny, S., and Dreyfuss, G. (1999). Transport of proteins and RNAs in and out of the nucleus. Cell *99*, 677-690.

Nayler, O., Cap, C., and Stamm, S. (1998a). Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization, and generation of a tissue-specific isoform. Genomics *53*, 191-202.

Nayler, O., Stamm, S., and Ullrich, A. (1997). Characterisation and comparison of four SR protein kinases. Biochem J *326*, 693-700.

Nayler, O., Strätling, W., Bourquin, J.-P., Stagljar, I., Lindemann, L., Jasper, H., Hartmann, A. M., Fackelmayer, F. O., Ullrich, A., and Stamm, S. (1998b). SAF-B couples transcription and pre-mRNA splicing to SAR/MAR elements. Nucleic Acids Res *26*, 3542-3549.

Neel, H., Gondran, P., Weil, D., and Dautry, F. (1995). Regulation of pre-mRNA processing by src. Curr Biol *5*, 413-422.

Nikolakaki, E., Du, C., Lai, J., Giannakouros, T., Cantley, L., and Rabinow, L. (2002). Phosphorylation by LAMMER Protein Kinases: Determination of a Consensus Site, Identification of In Vitro Substrates, and Implications for Substrate Preferences. Biochemistry *41*, 2055-2066.

Nikolakaki, E., Kohen, R., Hartmann, A. M., Stamm, S., Georgatsou, E., and Giannakouros, T. (2001). Cloning and characterization of an alternatively spliced form of SR protein kinase 1 that interacts specifically with scaffold attachment factor-B. J Biol Chem *276*, 40175-40182.

Ninkina, N., Grashchuck, M., Buchman, V. L., and Davies, A. M. (1997). TrkB variants with deletions in the leucine-rich motifs of the extracellular domain. J Biol Chem *272*, 13019-13025.

Niwa, M., and Berget, S. M. (1991a). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev *5*, 2086-2095.

Niwa, M., and Berget, S. M. (1991b). Polyadenylation precedes splicing in vitro. Gene Expr *1*, 5-14.

Niwa, M., Rose, S. D., and Berget, S. M. (1990). In vitro polyadenylation is stimulated by

the presence of an upstream intron. Genes Dev *4*, 1552-1559.

Ohe, K., Lalli, E., and Sassone-Corsi, P. (2002). A direct role of SRY and SOX proteins in pre-mRNA splicing. Proc Natl Acad Sci U S A *29*, 29.

Patapoutian, A., and Reichardt, L. F. (2001). Trk receptors: mediators of neurotrophin action. Curr Opin Neurobiol *11*, 272-280.

Patel, N. A., Chalfant, C. E., Watson, J. E., Wyatt, J. R., Dean, N. M., Eichler, D. C., and Cooper, D. R. (2001). Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells. J Biol Chem *276*, 22648-22654.

Paul, M. S., and Bass, B. L. (1998). Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. Embo J *17*, 1120-1127.

Pevny, L. H., and Lovell-Badge, R. (1997). Sox genes find their feet. Curr Opin Genet Dev *7*, 338-344.

Philips, A. V., Timchenko, L. T., and Cooper, T. A. (1998). Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science *280*, 737-741.

Pinol-Roma, S., and Dreyfuss, G. (1991). Transcription-dependent and transcription-independent nuclear transport of hnRNP proteins. Science *253*, 312-314.

Ploos van Amstel, J. K., Bergman, A. J., van Beurden, E. A., Roijers, J. F., Peelen, T., van den Berg, I. E., Poll-The, B. T., Kvittingen, E. A., and Berger, R. (1996). Hereditary tyrosinemia type 1: novel missense, nonsense and splice consensus mutations in the human fumarylacetoacetate hydrolase gene; variability of the genotype-phenotype relationship. Hum Genet *97*, 51-59.

Polydorides, A. D., Okano, H. J., Yang, Y. Y. L., Stefani, G., and Darnell, R. B. (2000). A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. Proc Natl Acad Sci USA *97*, 6350-6355.

Prasad, J., Colwill, K., Pawson, T., and Manley, J. (1999). The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. Mol Cell Biol *19*, 6991-7000.

Reed, R., and Palandjian, L. (1997). Spliceosome assembly. In Eucaryotic mRNA processing, A. Krainer, ed. (New York, Oxford University Press Inc., New York), pp. 103-129.

Reenan, R. A. (2001). The RNA word meets behavior: A to I pre-mRNA editing in animals. Trends Genet *17*, 53-56.

Robinson, D. R., Wu, Y.-M., and Lin, S.-F. (2000). The protein tyrosine kinase family of the human genome. Oncogene *19*, 5548-5557.

Roth, M. B., Zahler, A. M., and Stolk, J. A. (1991). A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. J Cell Biol *115*, 587-596.

Rubin, G. M. (2001). Comparing species. Nature *409*, 820-821.

Rueter, S. M., Dawson, T. R., and Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75-80.

Sanford, J. R., and Bruzik, J. P. (1999). Developmental regulation of SR protein phosphorylation and activity. Genes Dev *13*, 1513-1518.

Santisteban, I., Arredondo-Vega, F. X., Kelly, S., Loubser, M., Meydan, N., Roifman, C., Howell, P. L., Bowen, T., Weinberg, K. I., Schroeder, M. L., and et al. (1995). Three new adenosine deaminase mutations that define a splicing enhancer and cause severe and partial phenotypes: implications for evolution of a CpG hotspot and expression of a transduced ADA cDNA. Hum Mol Genet *4*, 2081-2087.

Schaefer, B. C. (1995). Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem *227*, 255-273.

Schneider, R., and Schweiger, M. (1991). A novel modular mosaic of cell adhesion motifs in the extracellular domains of the neurogenic trk and trkB tyrosine kinase receptors. Oncogene *6*, 1807-1811.

Screaton, G. R., Caceres, J. F., Mayeda, A., Bell, M. V., Plebanski, M., Jackson, D. G., Bell, J. I., and Krainer, A. R. (1995). Identification and characterization of three members of the human SR nfamily of pre-mRNA splicing factors. EMBO J *14*, 4336-4349.

Segade, F., Claudio, E., Wrobel, K., Ramos, S., and Lazo, P. S. (1995). Isolation of Nine Gene Sequences Induced by Silica in Murine Macrophages. J Immunol *154*, 2384-2392.

Segade, F., Hurle, B., Claudio, E., Ramos, S., and Lazo, P. S. (1996). Molecular cloning of a mouse homologue for the Drosophila splicing regulator Tra2. FEBS lett *387*, 152-156.

Sell, S. M., Reese, D., and Ossowski, V. M. (1994). Insulin-inducible changes in insulin receptor mRNA splice variants. J Biol Chem *269*, 30769-30772.

Sharp, P. A. (1991). “Five easy pieces”. Science *254*, 663.

Shatkin, A. J., and Manley, J. L. (2000). The ends of the affair: capping and polyadenylation. Nat Struct Biol *7*, 838-842.

Shelton, D. L., Sutherland, J., Gripp, J., Camerato, T., Armanini, M. P., Phillips, H. S., Carroll, K., Spencer, S. D., and Levinson, A. D. (1995). Human trks: molecular cloning, tissue distribution, and expression of extracellular domain immunoadhesins. J Neurosci *15*, 477-491.

Sisodia, S. S., Sollner-Webb, B., and Cleveland, D. W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. Mol Cell Biol *7*, 3602-3612.

Smale, S. T., and Tjian, R. (1985). Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. Mol Cell Biol *5*, 352-362.

Smith, C. W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.

Sossi, V., Giuli, A., Vitali, T., Tiziano, F., Mirabella, M., Antonelli, A., Neri, G., and Brahe, C. (2001). Premature termination mutations in exon 3 of the SMN1 gene are associated with exon skipping and a relatively mild SMA phenotype. Eur J Hum Genet *9*, 113-120.

Spillantini, M. G., Murrell, J. R., Goedert, M., Farlow, M. R., Klug, A., and Ghetti, B. (1998). Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. Proc Natl Acad Sci U S A *95*, 7737-7741.

Staknis, D., and Reed, R. (1994). SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol *14*, 7670-7682.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

Stamm, S., Casper, D., Dinsmore, J., Kaufmann, C. A., Brosius, J., and Helfman, D. (1992). Clathrin light chain B: gene structure and neuron-specific splicing. Nucl Acids Res *20*, 5097-5103.

Stamm, S., Zhang, M. Q., Marr, T. G., and Helfman, D. M. (1994). A sequence compilation and comparison of exons that are alternatively spliced in neurons. Nucleic Acids Res *22*, 1515-1526.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., and Zhang, M. Q. (2000). An

alternative-exon database and its statistical analysis. DNA Cell Biol *19*, 739-756.

Stanford, P. M., Halliday, G. M., Brooks, W. S., Kwok, J. B., Storey, C. E., Creasey, H., Morris, J. G., Fulham, M. J., and Schofield, P. R. (2000). Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: expansion of the disease phenotype caused by tau gene mutations. Brain *123*, 880-893.

Stark, J. M., Bazett-Jones, D. P., Herfort, M., and Roth, M. B. (1998). SR proteins are sufficient for exon bridging across an intron. Proc Natl Acad Sci U S A *95*, 2163-2168.

Steinmetz, E. J. (1997). Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? Cell *89*, 491-494.

Stoss, O., Olbrich, M., Hartmann, A. M., Konig, H., Memmott, J., Andreadis, A., and Stamm, S. (2001). The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites. J Biol Chem *276*, 8665-8673.

Stoss, O., Stoilov, P., Hartmann, A. M., Nayler, O., and Stamm, S. (1999). The in vivo minigene approach to analyze tissue-specific splicing. Brain Research Protocols *4*, 383-394.

Strohmaier, C., Carter, B. D., Urfer, R., Barde, Y.-A., and Dechant, G. (1996). A splice variant of the neurotrophin receptor trkB with increased specificity for brain-derived neurotrophic factor. EMBO J *15*, 3332-3337.

Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. Mol Cell Biol *20*, 6414-6425.

Tabaczewski, P., Shirwan, H., Lewis, K., and Stroynowski, I. (1994). Alternative splicing of class Ib major histocompatibility complex transcripts in vivo leads to the expression of soluble Qa-2 molecules in murine blood. Proc Natl Acad Sci U S A *91*, 1883-1887.

Tacke, R., and Manley, J. L. (1995). The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. Embo J *14*, 3540-3551.

Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998). Human Tra2 proteins are sequence-specific activators of

pre-mRNA splicing. Cell *93*, 139-148.

Tarn, W. Y., and Steitz, J. A. (1995). Modulation of 5' splice site choice in pre-messenger RNA by two distinct steps. Proc Natl Acad Sci U S A *92*, 2504-2508.

Tarn, W. Y., and Steitz, J. A. (1996). A novel spliceosome containing U11, U12, and U5

snRNPs excises a minor class (AT-AC) intron in vitro. Cell *84*, 801-811.

Thanaraj, T. A., and Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res *29*, 2581-2593.

Thoenen, H. (1995). Neurotrophins and neuronal plasticity. Science *270*, 593-598.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Tiscornia, G., and Mahadevan, M. S. (2000a). Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios. Mol Cell *5*, 959-967.

Tiscornia, G., and Mahadevan, M. S. (2000b). Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios. Mol Cell *5*, 959-967.

Tsoulfas, P., Soppet, D., Escandon, E., Tessarollo, L., Mendoza-Ramirez, J. L., Rosenthal, A., Nikolics, K., and Parada, L. F. (1993). The rat trkC locus encodes multiple neurogenic receptors that exhibit differential response to neurotrophin-3 in PC12 cells. Neuron *10*, 975-990.

Tsukamoto, Y., Matsuo, N., Ozawa, K., Hori, O., Higashi, T., Nishizaki, J., Tohnai, N., Nagata, I., Kawano, K., Yutani, C., *et al.* (2001). Expression of a novel RNA-splicing factor, RA301/Tra2beta, in vascular lesions and its role in smooth muscle cell proliferation. Am J Pathol *158*, 1685-1694.

Tupler, R., Perini, G., and Green, M. R. (2001). Expressing the human genome. Nature *409*, 832-833.

Valcárcel, J., and Green, M. R. (1996). The SR protein family: pleiotropic functions in pre-mRNA splicing. Trends Biochem Sci *21*, 296-301.

van der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J., and Caceres, J. F. (2000). The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. J Cell Biol *149*, 307-316.

Venables, J. P., Vernet, C., Chew, S. L., Elliott, D. J., Cowmeadow, R. B., Wu, J., Cooke, H. J., Artzt, K., and Eperon, I. C. (1999). T-STAR/ÉTOILE: a novel relative of SAM68 that

interacts with an RNA-binding protein implicated in spermatogenesis. Human Molecular Genetics *8*, 959-969.

Vitali, T., Sossi, V., Tiziano, F., Zappata, S., Giuli, A., Paravatou-Petsotas, M., Neri, G., and Brahe, C. (1999). Detection of the survival motor neuron (SMN) genes by FISH: further evidence for a role for SMN2 in the modulation of disease severity in SMA patients. Hum Mol Genet *8*, 2525-2532.

Volloch, V., Schweitzer, B., and Rits, S. (1994). Ligation-mediated amplification of RNA from murine erythroid cells reveals a novel class of beta globin mRNA with an extended 5'-untranslated region. Nucleic Acids Res *22*, 2507-2511.

Wang, H.-Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L., and Fu, X.-D. (1998a). SRPK2: A differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalina cells. J Cell Biol *140*, 737-750.

Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998b). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J Cell Biol *140*, 737-750.

Wang, H. Y., Xu, X., Ding, J. H., Bermingham, J. R., Jr., and Fu, X. D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. Mol Cell *7*, 331-342.

Wang, J., Dong, Z., and Bell, L. R. (1997). Sex-lethal interactions with protein and RNA. Roles of glycine-rich and RNA binding domains. J Biol Chem *272*, 22227-22235.

Wang, Z., Hoffmann, H. M., and Grabowski, P. J. (1995). Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. RNA *1*, 21-35.

Watanabe, T. K. (1975). A new sex-transforming gene on the second chromosome of Drosophila melanogaster. Japan J Genet *50*, 269-271.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res *27*, 1409-1420.

Weg-Remers, S., Ponta, H., Herrlich, P., and Konig, H. (2001). Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. Embo J *20*, 4194-4203.

Weighardt, F., Biamonti, G., and Riva, S. (1995). Nucleo-cytoplasmic distribution of

human hnRNP proteins: a search for the targeting domains in hnRNP A1. J Cell Sci *108 (Pt 2)*, 545-555.

Weighardt, F., Cobianchi, F., Cartegni, L., Chiodi, I., Villa, A., Riva, S., and Biamonti, G. (1999). A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. J Cell Sci *112*, 1465-1476.

Will, C. L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol *13*, 290-301.

Wohr, G., Fink, T., and Assum, G. (1996). A palindromic structure in the pericentromeric region of various human chromosomes. Genome Res *6*, 267-279.

Womble, D. D. (2000). GCG: The Wisconsin Package of sequence analysis programs. Methods Mol Biol *132*, 3-22.

Wu, J. Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Xie, J., and Black, D. L. (2001). A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. Nature *410*, 936-939.

Xie, J., and McCobb, D. P. (1998). Control of alternative splicing of potassium channels by stress hormones. Science *280*, 443-446.

Xu, Y., and Uberbacher, E. C. (1997). Automated gene identification in large-scale genomic sequences. J Comput Biol *4*, 325-338.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R. V., Gentile, C., Gebara, M., and Corden, J. L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc Natl Acad Sci U S A *93*, 6975-6980.

Zahler, A. M. (1999). Purification of SR protein splicing factors. Methods Mol Biol *118*, 419-432.

Zahler, A. M., and Roth, M. B. (1995). Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. Proc Natl Acad Sci U S A *92*, 2642-2646.

Zeng, C., and Berget, S. M. (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol Cell Biol *20*, 8290-8301.

Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature *407*, 401-405.

Zhu, J., and Krainer, A. R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev *14*, 3166-3178.

Zhu, J., Mayeda, A., and Krainer, A. R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell *8*, 1351-1361.

Zilch, C. F., Walker, A. M., Timon, M., Goff, L. K., Wallace, D. L., and Beverley, P. C. (1998). A point mutation within CD45 exon A is the cause of variant CD45RA splicing in humans. Eur J Immunol *28*, 22-29.

Curriculum Vitae

Petar Stoilov

AG Prof. Stefan Stamm

Institute of Biochemistry, Friedrich-Alexander-University

17 Fahrstrasse, 91054 Erlangen, Germany

e-mail: pesho@stamms-lab.net

phone: +49 9131 85 4617

fax: +49 9131 85 24605

Scientific interests

Regulation of RNA splicing.

RNA splicing mechanism.

Education and Employment

Since 1999

Prof. Stefan Stamm group

Max-Plank-Institute of Biochemistry, Martinsried, Germany

Friedrich-Alexander-University, Erlangen, Germany

Graduate Student

1995- 1999

Bulgarian Academy of Sciences, Institute of Molecular Biology

Sofia, Bulgaria

Research Assistant

1990 - 1995

Sofia University, Faculty of Biology

Sofia, Bulgaria

M.Sc. in Biochemistry and Microbiology, Diploma work: A study of the DNA repair activity over human beta-globin gene region (see publications)

1986 - 1990

National High School of Natural Sciences and Mathematics

Biotechnology class

Sofia, Bulgaria

1979 - 1986

Aleko Konstantinov elementary school

Plovdiv, Bulgaria

Courses

1997

Institute of contagious and parsitic deseases: Flow Cytometry and Cell sorting.

Languages

English, Russian, Bulgarian

Personal Information

Date of Birth: 22nd December 1972

Marital status: Married

Nationality: Bulgarian

Publications

Peter Stoilov, Eran Meshorer, Marieta Gencheva, David Glick, Hermona Soreq and Stefan Stamm, Defects in pre-mRNA processing as causes of and predisposition to diseases, DNA and Cell Biology, 2002, accepted for publication

Stoilov P, Castren E, Stamm S. Analysis of the human TrkB gene genomic organization reveals novel TrkB isoforms, unusual gene length, and splicing mechanism. Biochem Biophys Res Commun. 2002 Jan 25;290(3):1054-65.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O. and Zhang, M.Q. (2000) An alternative exon database (AEDB) and its statistical analysis, DNA and Cell Biol.,19:739-756

Daoud, R., P. Stoilov, O. Stoss, A.M. Hartmann, M. Olbrich, and S. Stamm. (2000) Control of pre-mRNA processing by extracellular signals: from molecular mechanisms to clinical applications. Gene Ther. Mol. Biol. 5: 141-150.

Stoss, O., Stoilov, P., Daoud, R., Hartmann, A.M., Olbrich, M. and Stamm, S. (2000) Misregulation of pre-mRNA splicing that causes human diseases. Gene Ther. Mol. Biol. 5: 9-28

Stoss O, Stoilov P, Hartmann AM, Nayler O, Stamm S. The in vivo minigene approach to analyze tissue-specific splicing. Brain Res Brain Res Protoc. 1999 Dec;4(3):383-94.

Chakarov St, Stoilov P, Alexandrov A, Russev G (1997) Repair pattern in thebeta-globin gene cluster of human fibroblasts after ultraviolet irradiation, Eur. J. Biochem 248, p 669

Manuscripts in preparation

Stoilov P, Daoud R, Nayler O and Stamm S, Regulation and function of tra2-beta isoforms, manuscript in preparation

Annakaisa Haapasalo, Ilkka Sipola, Kim Larsson, Karl E.O. Åkerman, Peter Stoilov, Stefan Stamm, Garry Wong and Eero Castrén Regulation of TrkB Surface Expression by Brain-Derived Neurotrophic Factor and Truncated TrkB Isoforms, manuscript in preparation

Stoilov P, Rafalska I, Stamm S, YTH: a new domain in nuclearproteins involved in pre-mRNA metabolism, manuscript in preparation

Patents

EP 1097941-A 20 09-MAY-2001

Grants

Deutsche Forschungsgemeinschaft (DFG) RNA conference 2001 travel sponsorship: KON 952/2001