

**Alternative splicing regulation in the human system:
Mechanistic studies of hnRNP L and CA-rich elements**

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie und Chemie
der Justus-Liebig-Universität Gießen

vorgelegt von

Monika Heiner
aus Marburg

Gießen, Dezember 2008

Die vorliegende Arbeit wurde am Institut für Biochemie des Fachbereichs 08 der Justus-Liebig-Universität Gießen in der Zeit von August 2004 bis Dezember 2008 unter der Leitung von Prof. Dr. Albrecht Bindereif angefertigt.

Dekan: Prof. Dr. Peter R. Schreiner
Institut für Organische Chemie
Justus-Liebig-Universität Gießen

1. Gutachter: Prof. Dr. Albrecht Bindereif
Institut für Biochemie
Justus-Liebig-Universität Gießen

2. Gutachter: Prof. Dr. Rainer Renkawitz
Institut für Genetik
Justus-Liebig-Universität Gießen

Contents

CONTENTS	I
ZUSAMMENFASSUNG	V
SUMMARY	VII
1. INTRODUCTION	1
1.1 Splicing of RNA	1
1.2 The splicing reaction	2
1.3 Spliceosome assembly	3
1.4 Alternative splicing	5
1.5 Splicing enhancer and silencer	7
1.6 <i>Trans</i> -acting factors	8
1.7 HnRNP L	11
1.8 Splicing and disease	13
1.9 Global analysis of alternative splicing	14
1.10 Aim of the work	15
2. MATERIALS AND METHODS	17
2.1 Materials	17
2.1.1 Chemicals and reagents	17
2.1.2 Nucleotides	18
2.1.3 Enzymes and enzyme inhibitors	18
2.1.4 Reaction buffers	18
2.1.5 Molecular weight markers	19
2.1.6 Kits	19
2.1.7 Materials for mammalian cell culture	19
2.1.8 Plasmids	19
2.1.9 <i>E.coli</i> strains and mammalian cell lines	20
2.1.10 Antibodies	20
2.1.11 DNA oligonucleotides	20
2.1.12 RNA oligonucleotides	21
2.1.13 Other materials	22
2.2 Methods	22
2.2.1 DNA cloning	22
2.2.1.1 Preparation of plasmid DNA	22

2.2.1.2 Agarose gel electrophoresis	22
2.2.1.3 Preparation of DNA fragments	23
2.2.1.4 Restriction endonuclease digestion	23
2.2.1.5 Dephosphorylation	23
2.2.1.6 Ligation	23
2.2.1.7 Transformation	23
2.2.2 Minigene constructs	24
2.2.2.1 pcDNA3-SLC2A2	24
2.2.2.2 pcDNA3-TJP1	24
2.2.2.3 pcDNA3-ITGA2	25
2.2.3 <i>In vivo</i> splicing analysis	26
2.2.3.1 Cell culture	26
2.2.3.2 Transient transfection	26
2.2.3.3 Isolation of total RNA from HeLa cells	27
2.2.3.4 RQ1 DNase treatment	27
2.2.3.5 Analysis of <i>in vivo</i> splicing by RT-PCR	27
2.2.4 <i>In vitro</i> transcription	28
2.2.4.1 Transcription of ³² P-labelled RNA	28
2.2.4.2 Removal of unincorporated nucleotides by gel filtration	28
2.2.4.3 Transcription without ³² P-label	29
2.2.4.4 Determination of RNA concentrations	29
2.2.5 <i>In vitro</i> splicing of pre-mRNAs	29
2.2.5.1 Splicing reaction	29
2.2.5.2 Proteinase K treatment	30
2.2.5.3 Analysis of <i>in vitro</i> splicing by RT-PCR	30
2.2.6 Depletion of hnRNP L from HeLa nuclear extract	30
2.2.7 SDS polyacrylamide gel electrophoresis (SDS-PAGE)	31
2.2.8 Coomassie staining	31
2.2.9 Western blot	31
2.2.10 Purification of recombinant proteins	32
2.2.11 Spliceosome assembly reaction	33
2.2.12 Psoralen crosslinking	33
2.2.13 Preparation of total RNA from HeLa nuclear extract	34
2.2.14 RNase H cleavage	34
2.2.15 Silver staining	34
2.2.16 Electromobility shift assay (band shift)	35
2.2.17 UV crosslinking and immunoprecipitation	35
2.2.17.1 UV crosslinking with purified proteins	35
2.2.17.2 UV crosslinking in HeLa nuclear extract	35
2.2.17.3 Immunoprecipitation	36
2.2.18 RNAi in HeLa cells	36

2.2.18.1 siRNA knockdown	36
2.2.18.2 RNA isolation	37
2.2.18.3 Real-time PCR analysis	37
2.2.19 Microarray analysis	38
2.2.19.1 RT-PCR validation of hnRNP L target genes	38
3. RESULTS	39
3.1 <i>SLC2A2</i>	39
3.1.1 Intronic CA-repeat sequence represents a splicing silencer	40
3.1.2 HnRNP L mediates skipping of <i>SLC2A2</i> exon 4 <i>in vitro</i>	42
3.1.3 Intronic splicing silencer affects spliceosome assembly	43
3.1.4 HnRNP L binding to intronic splicing silencer interferes with 5' splice site recognition by the U1 snRNP	45
3.2 <i>TJP1</i>	49
3.2.1 HnRNP L knockdown in HeLa cells by RNA interference	49
3.2.2 Genome-wide search for hnRNP L target genes by a combined microarray and RNAi analysis	50
3.2.3 <i>TJP1</i> , an hnRNP L target gene	52
3.2.4 Identification of an intronic splicing silencer in the <i>TJP1</i> gene	53
3.2.5 Alternative splicing of <i>TJP1</i> exon 20 is mediated by hnRNP L	56
3.2.6 Mutational analysis of the intronic splicing silencer in <i>TJP1</i> gene	57
3.2.7 HnRNP L binds to <i>TJP1</i> CA-rich intronic silencer sequence with high affinity	58
3.2.8 U2AF65 interaction with polypyrimidine tract of <i>TJP1</i> intron 19 is very weak	61
3.2.9 HnRNP L interferes with binding of U2AF65 to <i>TJP1</i> intron 19	62
3.3 <i>ITGA2</i>	65
3.3.1 Analysis of <i>ITGA2</i> minigene splicing leads to identification of a cryptic exon	66
3.3.2 Splicing analysis of additional <i>ITGA2</i> minigene constructs	67
3.4 <i>ASAH1</i>	71
3.4.1 HnRNP L mediates alternative poly(A) site selection in the <i>ASAH1</i> gene	71
4. DISCUSSION	74
4.1 Microarray analysis: Genome-wide search for hnRNP L target genes	74
4.2 Crossregulation of the hnRNP L proteins	75
4.3 Diverse roles of hnRNP L in splicing regulation	76
4.4 The mechanism of alternative splicing repression by hnRNP L	78
4.4.1 HnRNP L interferes with 5' splice site recognition	79
4.4.2 HnRNP L interferes with 3' splice site recognition	80
4.4.2.1 Short <i>TJP1</i> RNA transcripts form different secondary structures	83
4.5 HnRNP L regulates alternative poly(A) site selection	85
4.6 HnRNP L represses inclusion of cryptic exons	86
4.7 Perspectives	87

5. REFERENCES	89
6. APPENDICES	100
Abbreviations	100
Curriculum vitae	102
Acknowledgements	104

Zusammenfassung

Alternatives Spleißen von prä-mRNAs trägt hauptsächlich dazu bei, aus einer verhältnismäßig kleinen Anzahl von Genen ein komplexes Proteom zu erzeugen. Es sind verschiedene Formen des alternativen Spleißens bekannt, die streng reguliert werden müssen, um eine fehlerfreie Expression von Proteinen zu gewährleisten.

Repetitive CA-Sequenzen bilden eine neue Gruppe spleißregulatorischer Sequenzelemente. CA-Dinukleotidwiederholungen sowie CA-reiche Sequenzen können als Spleiß-*Enhancer* oder –*Silencer* auf alternative Spleißvorgänge wirken. HnRNP L gehört zur großen Gruppe der heterogenen nukleären Ribonukleoproteine (hnRNPs) und bindet CA repetitive Sequenzen mit hoher Affinität. Die Identifizierung von Genen, deren alternatives Spleißen durch hnRNP L reguliert wird, soll dazu beitragen, weitere Erkenntnisse über den Regulationsmechanismus alternativer Spleißvorgänge zu gewinnen.

Die Regulation von alternativen Spleißvorgängen wurde in dieser Arbeit anhand von drei ausgewählten Beispielen untersucht. Im Gen *SLC2A2* wurden CA-Dinukleotidwiederholungen als intronischer Spleiß-*Silencer* identifiziert, dessen Funktion von hnRNP L als *trans*-agierenden Faktor abhängt. Weitergehende Studien zum Mechanismus der Spleißregulation zeigten, dass hnRNP L mit der Erkennung der 5' Spleißstelle durch den U1 snRNP interferiert. Eine kurze intronische CA-reiche Sequenz im *TJP1* Gen konnte als Spleiß-*Silencer* charakterisiert werden, dessen Funktion ebenfalls von hnRNP L vermittelt wird. Im Gegensatz zu *SLC2A2*, konnte für *TJP1* gezeigt werden, dass hnRNP L mit U2AF65 um die Bindung zum Polypyrimidine Trakt konkurriert. Für einen CA-reichen Abschnitt im ersten Intron des *ITGA2* Gens wurde gezeigt, dass dieser, abhängig von der Sequenz des gesamten Introns, entweder die Benutzung der nahe liegenden 5' Spleißstelle aktiviert oder die Erkennung eines kryptischen Exons unterdrückt.

Darüber hinaus konnte mit Hilfe einer Untersuchung, kombiniert aus Microarray und RNAi, eine neue Funktion von hnRNP L identifiziert werden, nämlich die Beteiligung an alternativer Polyadenylierung. Für das Gen *ASAH1* wurde gezeigt, dass hnRNP L die Verwendung einer internen Polyadenylierungsstelle verhindert.

Zusammenfassend ist zu sagen, dass die Ergebnisse der vorliegenden Arbeit neue Einblicke in den Mechanismus der Regulation von alternativen Spleißen durch

hnRNP L erlauben. Zudem konnte hnRNP L eine Beteiligung an der Auswahl alternativer Poly(A)-Stellen nachgewiesen werden. Im humanen System bildet hnRNP L damit ein vielseitiges Regulationsprotein.

Summary

Alternative splicing of pre-mRNAs is the major contributor in the human system to generate complex proteomes from a comparatively low number of genes. Several modes of alternative splicing are known today which are tightly regulated to ensure accurate protein expression.

We have identified intronic CA repeat and CA-rich sequences as a new class of regulatory elements acting as enhancers or silencers on alternative splicing. Their function is mediated by the heterogenous ribonucleoprotein (hnRNP) L which has been characterised as the main CA binding protein. Considering that CA repetitive sequences are very common in the human genome, the identification and analysis of hnRNP L target genes should give further insights into the mechanism of alternative splicing regulation.

In this work, I studied alternative splicing regulation of three recently identified hnRNP L target genes. First, in the *SLC2A2* gene CA repeats were identified as an intronic splicing silencer element. Their function was shown to depend on hnRNP L as the *trans*-acting factor. Further studies on the mechanism of splicing regulation demonstrated that hnRNP L interfered with 5' splice site recognition by the U1 snRNP. Second, I characterised a short intronic CA-rich cluster in the *TJP1* gene as an hnRNP L-dependent splicing silencer. In contrast to *SLC2A2*, hnRNP L was shown to compete with U2AF65 for binding to the polypyrimidine tract thus impairing 3' splice site recognition. Third, an intronic CA repeat in the *ITGA2* gene either activated splicing of the corresponding exon or repressed recognition of a cryptic exon in a sequence-dependent manner.

Furthermore, a combined microarray and RNAi analysis revealed new modes of hnRNP L-mediated splicing regulation and, moreover, a novel role for hnRNP L in alternative polyadenylation. In the *ASAH1* gene, hnRNP L repressed usage of an internal poly(A) site.

Taken together, I have studied different mechanism of splicing regulation on the basis of three genes, *SLC2A2*, *TJP1*, and *ITGA2*. The results revealed new functions of CA repeat and CA-rich sequences and hnRNP L in alternative splicing regulation.

In sum, hnRNP L was shown to be a global and versatile regulator protein in the human system with roles in alternative splicing and polyadenylation.

1. Introduction

1.1 Splicing of RNA

Splicing of mRNAs is an essential step in the expression of genetic information since almost all eukaryotic protein-coding genes are interrupted by introns. Human introns have a mean length of about 3500 nucleotides and show, except for the splice sites, no sequence conservation (Deutsch & Long, 1999). Considering that on average, one human gene contains just four introns, the human dystrophin gene represents an extreme example (Pozzoli *et al.*, 2002). It is the largest gene known so far, consisting of 79 exons, and spanning more than two million basepairs.

Exons, on the other hand, are normally much shorter (100-200 nucleotides) than introns. Moreover, they contain the coding sequence and are therefore highly conserved. For gene expression, it is very important that the excision of introns and the joining of exons occur most accurately. In Eukarya there are four splicing mechanisms known (Abelson *et al.*, 1998). These are two types of self-splicing introns, group I and group II, tRNA splicing and splicing of nuclear pre-mRNAs. Mechanistically, splicing of tRNAs is very different from the other types of splicing since it is catalysed by proteins. Group I and group II introns, on the other hand, are classical ribozymes, which catalyse their own excision without the help of *trans*-acting proteins (Toor *et al.*, 2008). Only in some cases proteins are needed to form the catalytic centre (Cech, 1990). Splicing of the protein-coding genes of higher eukaryotes occurs by a mechanism similar to that of group II introns although it is not autocatalytic. It proceeds in two subsequent transesterification reactions and is catalysed by the spliceosome.

For recognition and subsequent catalysis by the spliceosome, three conserved intronic sequence elements are essential (Fig. 1.1). These are the exon-intron boundaries, namely the 5' and 3' splice sites, and the branch point sequence (Green, 1986; Sheth *et al.*, 2006). The polypyrimidine tract, a pyrimidine-rich region of variable length upstream of the 3' splice site, represents another conserved sequence element.

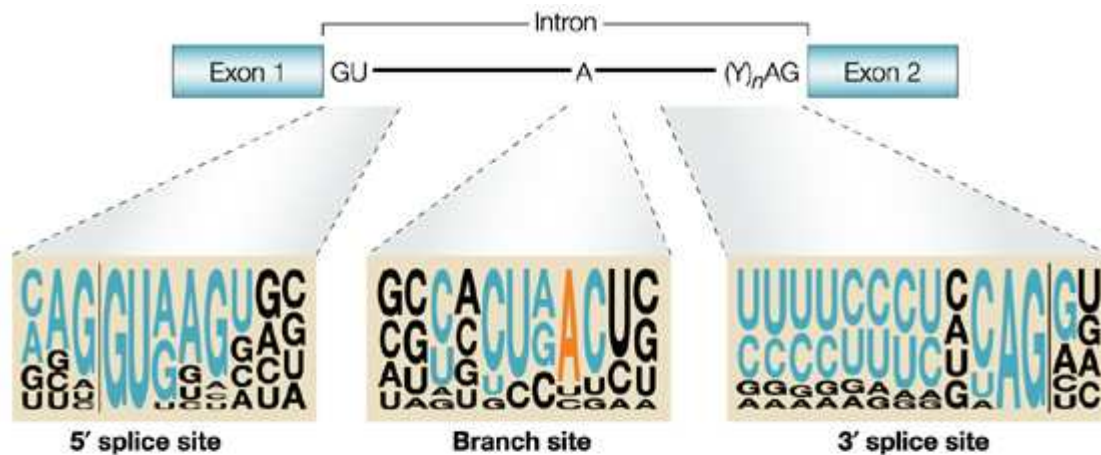


Figure 1.1

Conserved sequence elements of major-class introns in the pre-mRNA. A two-exon pre-mRNA is schematically shown with conserved sequence motifs of the 5' splice site, branch site, and 3' splice site with preceding polypyrimidine tract (Y)_n. The height of a letter at a given position represents the frequency of the corresponding nucleotide at that position as determined by alignment of conserved sequences from 1,683 human introns. Nucleotides that are part of the classical consensus motifs are shown in blue, except for the branch point A, which is shown in orange. The vertical lines indicate the exon–intron boundaries (adapted from Cartegni *et al.*, 2002).

A distinction is drawn between the major (U2-dependent) and minor (U12-dependent) spliceosomes according to their composition. The vast majority of introns are U2-dependent having the terminal dinucleotides GU-AG. Only few introns are spliced by the minor spliceosome frequently possessing AT-AC dinucleotides at their ends (Will & Lührmann, 2005).

1.2 The splicing reaction

The splicing reaction itself starts with the 2' hydroxyl group of the branch site adenosine attacking the phosphodiester bond at the 5' splice site (Krämer, 1996). Through this nucleophilic attack, the linkage between exon and intron is broken resulting in two splicing intermediates, released exon 1 and exon 2 with intron lariat (Fig. 1.2). The characteristic lariat structure results from an unusual 2'-5' phosphodiester bond of the branch site adenosine which is thereafter linked to three nucleotides. The phosphodiester bond of the 3' splice site is broken in the second step by the nucleophilic attack of the free 3' hydroxyl group of the first exon. This leads to release of the intron lariat and joining of the two exons (Valadkhan &

Manley, 2001). The spliceosome, which catalyses the two transesterification steps, is described in more detail in the next chapter.

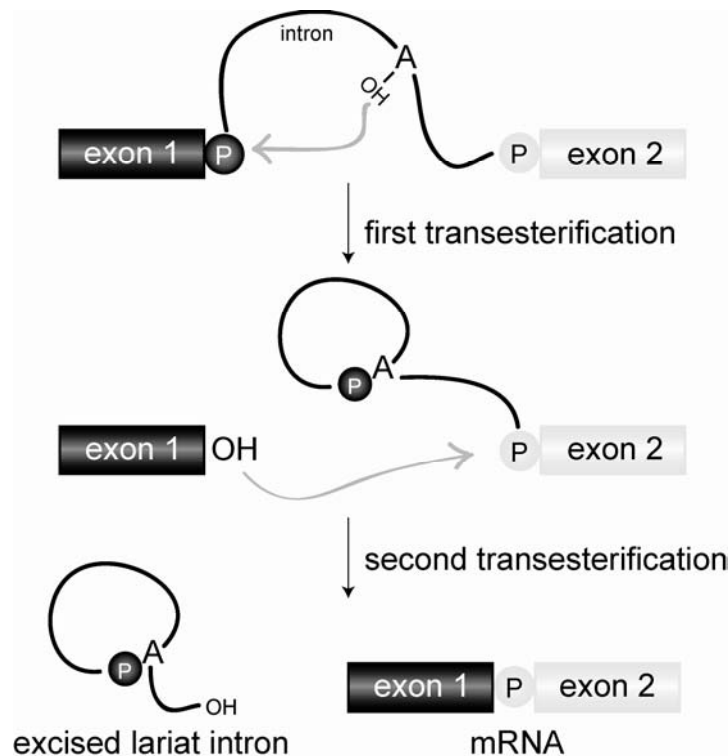


Figure 1.2

Two-step transesterification pathway of pre-mRNA splicing. Exons are depicted as boxes, the intron as a line. Phosphate groups are illustrated by rings. The branch point adenosine is highlighted. The removal of the intron from the pre-mRNA is carried out in two subsequent transesterification reactions leading to joining of the exons and release of the intron lariat.

1.3 Spliceosome assembly

The spliceosome, a large and dynamic ribonucleoprotein machine, is responsible for splicing of most eukaryotic mRNAs (Staley & Guthrie, 1998). Small nuclear ribonucleoproteins (snRNPs) represent the major components of the spliceosome, each composed of a small nuclear RNA (snRNA) and several proteins. Besides the snRNPs, more than 100 additional non-snRNP splicing factors, which are also required for the removal of introns, are associated with the spliceosome (Jurica & Moore, 2003). The five spliceosomal snRNAs are termed U1, U2, U4, U5, and U6, due to their uridine-rich sequence (Will & Lührmann, 2001). Each snRNA binds a set of seven Sm or Sm-like proteins and several specific proteins altogether forming the snRNP particle.

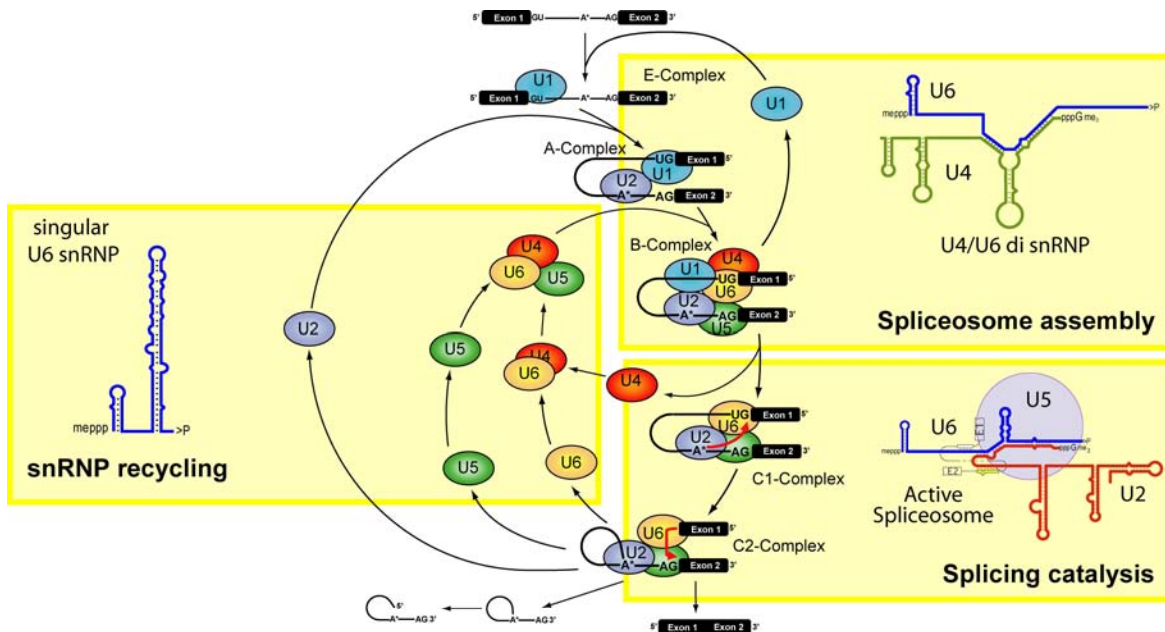


Figure 1.3

The spliceosome cycle. Exons are represented by black boxes, the intron by a line. The branch point adenosine (A*), 5' (GU), and 3' (AG) splice sites are highlighted. Each snRNP is depicted as a coloured circle. The nucleophilic attacks during the splicing reaction are shown by red arrows. The spliceosome cycle passes through three main phases, spliceosome assembly, splicing catalysis, and snRNP recycling which are marked by yellow boxes. The U4/U6 base-pairing in the di-snRNP is shown on the right as well as the base-pairing of U2 and U6 in the active spliceosome. The secondary structure of the singular U6 snRNP is depicted on the left (kindly provided by Dr. Jan Medenbach).

The catalysis itself is mediated by extensive structural rearrangements of the spliceosomal complex (Nilsen, 1994). The assembly occurs in a highly ordered and stepwise manner on every intron starting with the formation of the ATP-independent E (early) complex (Fig. 1.3) (Brow, 2002; Hastings & Krainer, 2001; Schellenberg *et al.*, 2008). The spliceosomal E complex is characterised by the association of the U1 snRNP with the 5' splice site through base-pairing of its snRNA component. Branch site, polypyrimidine tract, and 3' splice site are recognised by non-snRNP splicing factors, such as SF1 (splicing factor 1) and U2AF (U2 auxiliary factor). Recruitment of the U2 snRNP to the branch site leads to formation of the A complex, the first ATP-dependent step in spliceosome assembly. The subsequent B complex is generated by association of the U4/U6•U5 tri-snRNP with the pre-mRNA. In the tri-snRNP the U4 and U6 snRNA are extensively base-paired whereas the U5 snRNP is associated via protein-protein interactions. After recruitment of the tri-snRNP the spliceosome has to undergo several RNA-RNA and RNA-protein rearrangements to form the catalytic core. Unwinding of the U4/U6 RNA-duplex,

destabilisation and release of the U1 and U4 snRNP lead to formation of the catalytically active C complex which carries out the first step of splicing. To facilitate the second step further structural rearrangements take place. After release of the spliced product and intron lariat, the tri-snRNP has to be reconstituted from postspliceosomal single U4, U5 and U6 snRNPs in order to participate in further rounds of splicing. In a recycling process U4 and U6 snRNAs reanneal to form the U4/U6 di-snRNP followed by association of the U5 snRNP yielding the splicing-competent tri-snRNP (Raghuathan & Guthrie, 1998).

1.4 Alternative splicing

Based on expressed-sequence clustering, the human genome was initially estimated to contain 150,000 genes (Modrek & Lee, 2002). Therefore, the actual number of “only” 32,000 genes came as a surprise (Lander *et al.*, 2001; Venter *et al.*, 2001). Moreover, since the comparatively lower organism of the fruit fly *Drosophila melanogaster* contains already 14,000 genes. How can a relatively low number of human genes produce a much higher number of mRNAs? Alternative splicing was the answer to that question.

Alternative splicing is the major contributor in the human system to generate complex proteomes. The analysis of alternative splicing using bioinformatics revealed a much greater number of alternatively spliced genes than were initially expected (Ast, 2004). It is currently estimated that more than 60% of all human genes undergo alternative splicing. Some bioinformatic analyses, however, anticipate an even greater number (Lee & Wang, 2005). Most genes encode 2-3 different protein isoforms but there are also some extreme examples (Olson *et al.*, 2007). By taking into account all possible combination of exons the *D. melanogaster* *Dscam* (homolog of human Down syndrome cell adhesion molecule) gene can generate 38,016 isoforms.

A constitutive exon is always spliced or included into the final mRNA whereas some exons are regulated (Black, 2003). Fig. 1.4 gives an overview of the five basic modes of alternative splicing. A regulated exon can be either skipped or included (panel A). By altering the 5' or 3' splice site exons can also be lengthened or shortened (panels B, C). In some cases, exons are mutually exclusive, always including only one of several possible exon choices into the final mRNA (panel D).

Finally, introns can be either removed or retained thereby contributing to alternative splicing modes as well (panel E). One mode of alternative splicing, however, is not restricted to one pre-mRNA species. Frequently, several alternative splicing events lead to a family of related proteins expressed from a single gene.

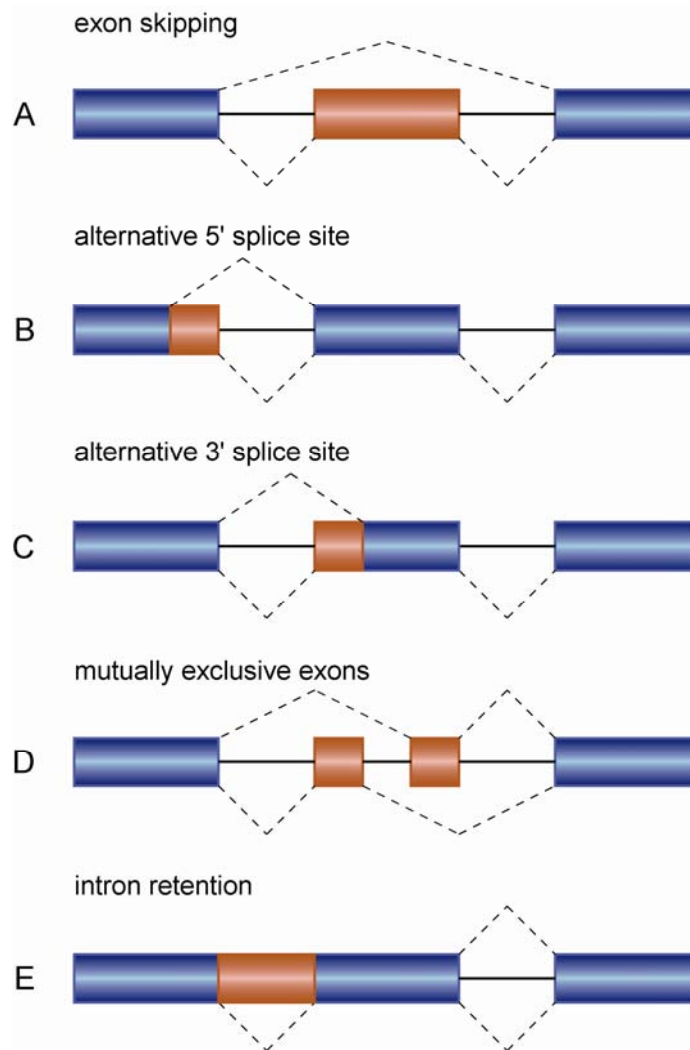


Figure 1.4

Alternative splicing patterns. The constitutive exons are represented by blue boxes, alternative exons by red boxes. Introns are illustrated by black lines. Alternative splicing pattern are indicated in each case (adapted from Ast, 2004).

The effects of alternative splicing on the encoded proteins are very diverse. For example, it leads to the formation of protein isoforms that differ in functional domains, subcellular localisation or binding specificity (Black, 2000). Alternative splicing can be regulated in a tissue-, sex-, or developmental-specific manner. The somatic sex determination pathway in *Drosophila melanogaster* is one of the best-studied examples of alternative splicing regulation (Black, 2003; Robida *et al.*,

2007). The sex determination genes are spliced differently in male and female flies controlled by a series of alternative splicing events.

Approximately one third of all alternative splicing events leads to the introduction of a premature termination codon (PTC) subjecting the mRNA to degradation by nonsense-mediated decay (NMD) (McGlinchy & Smith, 2008). NMD represents one RNA surveillance pathway to ensure the fidelity of gene expression.

1.5 Splicing enhancer and silencer

For accurate splicing of pre-mRNAs, the spliceosome has to recognise comparatively small exons in often large stretches of intronic RNA (Cartegni *et al.*, 2002; Maniatis & Tasic, 2002). The only poorly conserved splice site sequences are not sufficient for correct identification of exons since cryptic splice sites, which loosely match the consensus sequence, are very common in introns. Besides the canonical splice site signals, additional sequence elements are therefore required to define genuine exon-intron boundaries. Splicing enhancer and silencer are *cis*-acting regulatory elements containing the necessary information for either stimulating (enhancer) or repressing (silencer) splicing (Wang *et al.*, 2005). They can be classified according to their location in either exons or introns as exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs), and intronic splicing silencers (ISSs).

Exonic splicing enhancers are the best characterised splicing regulatory elements. Sequences with enhancer activity were identified using functional *in vitro* and *in vivo* SELEX (systematic evolution of ligands by exponential enrichment) (Coulter *et al.*, 1997; Schaal & Maniatis, 1999). The two major classes are purine-rich and adenosine/cytosine-rich ESE sequences. These motifs, however, are generally very short, only 6-8 nucleotides long, degenerate and partially overlapping. By means of a web-based program called ESEfinder putative ESEs can be predicted (Cartegni *et al.*, 2003; Smith *et al.*, 2006). This is especially important for the prediction of disease-associated point mutations or polymorphisms which often result in pre-mRNA splicing defects. ESEs mediate regulation through recruitment of *trans*-acting factors, mainly members of the serine-arginine-rich (SR) protein family (Wang & Burge, 2008).

Exonic splicing silencers are another major class of splicing regulatory sequences. ESSs interact with heterogeneous nuclear ribonucleoproteins (hnRNPs) to inhibit use of adjacent splice sites (Wang *et al.*, 2004). They have also been predicted based on mutational or computational approaches but most of these sequences show little similarity and are highly degenerate.

Besides their role in constitutive splicing, ESEs and ESSs play an important role in the regulation of alternative splicing (Black, 2003).

So far, only a few intronic splicing regulatory elements (ISREs), such as ISEs and ISSs, have been characterised. Yeo and co-workers found that up to 50% of ISREs were enriched near alternatively spliced exons suggesting their importance in regulation of alternative splicing (Yeo *et al.*, 2007). CA-rich and CA repeat sequences represent one class of ISREs (Hui *et al.*, 2005; Hung *et al.*, 2008). These sequence elements have been shown to function both, as intronic splicing enhancer and silencer. The YCAY motif, which is bound by the neuron-specific Nova protein family, represents another example of an intronic splicing regulatory sequence, regulating a large number of splicing events in the brain (Ule *et al.*, 2003; Ule *et al.*, 2006).

1.6 *Trans*-acting factors

Splicing enhancer and silencer mediate their function through binding of *trans*-acting factors, which can be divided into two major groups, SR proteins and hnRNPs.

Serine-arginine-rich (SR) proteins are a family of highly conserved non-snRNP splicing factors with diverse roles in constitutive and alternative splicing. Characteristic for all SR proteins is their variable-length arginine-serine-rich (RS) domain at the C-terminus which is required for protein-protein interactions with other RS domain-containing proteins (Graveley, 2000). The N-terminal part of SR proteins contains one or two RNA-recognition motifs (RRMs) which are sufficient for sequence-specific RNA binding. Other proteins, distinct from the SR proteins, which also contain an RS domain, are referred to as SR-related proteins. These proteins include the U2AF and U1 snRNP 70 kDa (U1 70K) protein (Table 1.1).

Table 1.1

Human SR proteins and SR-related proteins (adapted from Graveley, 2000)

SR proteins		SR-related proteins
SRp20	RRM RS	U2 auxiliary factor
SC35	RRM RS	U2AF65 U2AF35
SRp46	RRM RS	snRNP components
SRp54	RRM RS	U1 70K U5 100K U4/U6•U5 27K hLuc7p
SRp30c	RRM RRMH RS	Splicing regulators
ASF/SF2	RRM RRMH RS	hTra2 α hTra2 β
SRp40	RRM RRMH RS	Splicing coactivators
SRp55	RRM RRMH RS	SRm160 SRm300
SRp75	RRM RRMH RS	RNA helicases
9G8	RRM Z RS	hPrp16 HRH1
		Protein kinases
		Clk/Sty

SR proteins are involved in the regulation of alternative splicing mostly through recognition of ESE sequences (Fig 1.5). SR proteins display several modes of action, however, one of their best-characterised functions is splice site activation (Graveley, 2000; Matlin *et al.*, 2005). Through binding to an ESE close to a weak 5' splice site, SR proteins can stimulate splicing by recruitment of the U1 snRNP mediated by interaction with U1 70K (Lam *et al.*, 2003; Ryner *et al.*, 1996). Activation of the 3' splice site on the other hand is mediated by recruitment of U2AF65 to a weak pyrimidine tract (Zuo & Maniatis, 1996).

Heterogeneous nuclear ribonucleoproteins (hnRNP) are factors that bind RNA polymerase II transcribed primary transcripts of protein-coding genes in the nucleus (Dreyfuss *et al.*, 1993). These transcripts are called heterogeneous nuclear RNAs (hnRNAs) which is a historical term describing their size heterogeneity and cellular localisation. HnRNP proteins belong to the most abundant nuclear proteins in higher eukaryotes and participate in several RNA-related biological processes for example transcriptional regulation, splicing, 3' end processing, and mRNA export (Kim *et al.*, 2000). Over 20 major hnRNPs and several isoforms have been identified so far,

designated from hnRNP A1 (34 kDa) to U (120 kDa) (Dreyfuss *et al.*, 2002). A modular structure is a common feature of most hnRNP proteins, which usually contain one or more RNA-binding motifs and auxiliary domains for protein-protein interactions. HnRNP proteins, similar as SR proteins, participate in the regulation of alternative splicing. In contrast to SR proteins, most hnRNPs function as splicing repressors through binding to splicing silencer sequences.

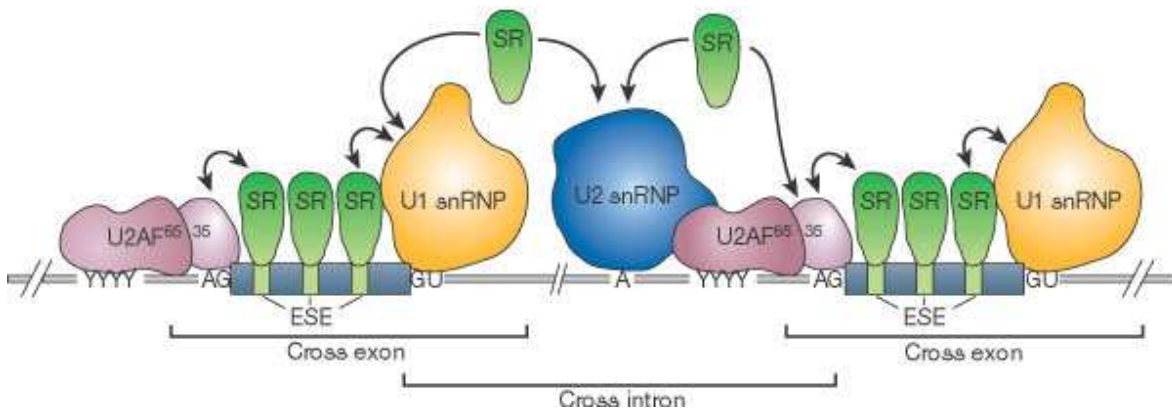


Figure 1.5

Splicing regulation by SR proteins binding to exonic splicing enhancers. The 5' splice site (GU), branchpoint (A), polypyrimidine tract, and 3' splice site (AG) are recognised by the splicing machinery. The exons contain exonic splicing enhancers (ESE) that are binding sites for SR proteins. SR proteins stimulate splicing by recruitment of the U1 snRNP to the downstream 5' splice site and/or U2AF (65 and 35 kDa subunits) to the upstream polypyrimidine tract and 3' splice site. Thereafter U2AF recruits the U2 snRNP to the branchpoint. SR proteins can also function across the intron (adapted from Maniatis and Tasic, 2002).

HnRNP A1 and I (PTB, polypyrimidine tract binding protein) are well-characterised examples of alternative splicing regulators. HnRNP A1 has been shown to bind to a G-rich intronic splicing silencer in the chicken β -tropomyosin gene antagonising the function of a splicing enhancer (Expert-Bezancon *et al.*, 2004). This displays, however, only one example of the genes regulated by hnRNP A1. Several target genes could be identified for PTB which generally acts as a splicing repressor as well (Spellman *et al.*, 2005). Wollerton and co-workers demonstrated that PTB even autoregulates its own expression by alternative splicing (Wollerton *et al.*, 2004). Like other hnRNP proteins, PTB shows an RNA binding specificity by preferentially binding to UCUU in a pyrimidine-rich context (Perez *et al.*, 1997). HnRNP L, another hnRNP protein shall be introduced in detail in the next chapter.

1.7 HnRNP L

The heterogeneous nuclear ribonucleoprotein (hnRNP) L is an abundant nuclear protein of 64 kDa (Pinol-Roma *et al.*, 1989). Several functions could be assigned to hnRNP L. First of all, it was identified in connection with the nuclear export of intronless mRNAs (Guang *et al.*, 2005; Liu & Mertz, 1995). It was shown that hnRNP L binds to a pre-mRNA processing enhancer derived from the intronless herpes simplex virus type 1 thymidin kinase (HSV-TK) gene, thereby enhancing cytoplasmic accumulation of mRNAs in an intron-independent manner. HnRNP L was also found to interact with the 3' border of the hepatitis C virus (HCV) internal ribosomal entry site (IRES) (Hahm *et al.*, 1998b; Hwang *et al.*, 2008). Binding of hnRNP L to the IRES correlated with increased translation efficiencies of the HCV mRNAs. Thirdly, hnRNP L plays a role in mRNA stability. Shih and Claffey identified hnRNP L as the protein binding to a CA-rich region in the 3'-untranslated region of the human vascular endothelial growth factor (VEGF) gene (Shih & Claffey, 1999). This interaction mediated regulation of VEGF mRNA stability under hypoxic conditions. The human endothelial nitric oxide synthase (eNOS) gene represents another case where hnRNP L affects RNA stability (Hui *et al.*, 2003a). The eNOS gene revealed, in addition, a novel function of hnRNP L in splicing regulation (Hui *et al.*, 2003b). Intron 13 of the eNOS gene carries a polymorphic CA repeat sequence which we identified as a splicing enhancer element. The CA repeat length correlated thereby with the splicing activation mediated by hnRNP L. Moreover, the number of CA repeats represents an independent risk factor for coronary artery disease since eNOS plays an important role in vascular homeostasis (Stangl *et al.*, 2000). With an *in vitro* SELEX approach, we determined the binding specificity of hnRNP L (Fig. 1.6). The sequences obtained showed an enrichment of CA dinucleotides with ACAC and CACA representing the minimal high-score binding motifs for hnRNP L which also recognises certain CA-rich sequences (Hui *et al.*, 2005).

The identification of further target genes besides eNOS revealed hnRNP L as a global regulator of alternative splicing. HnRNP L was also reported to interact with an exonic splicing silencer in exon 4 of the human *CD45* gene mediating repression of the corresponding exon (House & Lynch, 2006; Rothrock *et al.*, 2005). *CD45* encodes a haematopoietic-specific transmembrane protein tyrosine phosphatase

which is important for T-cell development and signalling. Exon 4 is one of three variable exons in the *CD45* gene which are repressed upon T-cell activation.

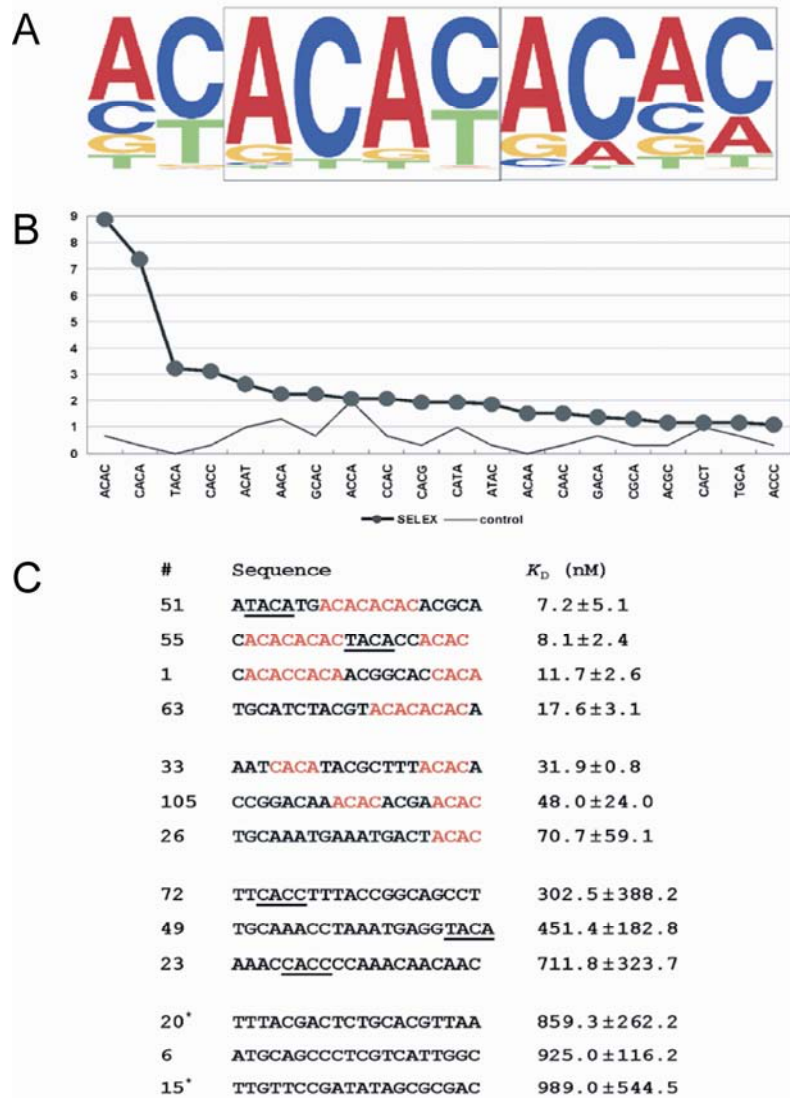


Figure 1.6
hnRNP L RNA binding specificity defined by *in vitro* SELEX. (A) Consensus sequence of hnRNP L binding. The height of a letter at a given position represents the frequency of the corresponding nucleotide. The boxes mark the two high-score hnRNP L binding motifs. (B) Tetranucleotide frequency in sequences selected by SELEX. The 20 most common tetranucleotide sequences are given in order of their frequencies in the 108 selected SELEX sequences (heavy line) and in 20 sequences taken from the initial pool (thin line, control). Both frequencies are diagrammed as percentage of the total. (C) Characteristics of 11 SELEX-derived (clone numbers on the left) and two control sequences (with asterisks; #20 and 15). Given are the individual sequences (with high-score motifs in red, low-score motifs underlined) and the K_D values (in nM; with standard deviations, $p < 0.05$) (adapted from Hui *et al.*, 2005).

Recently, hnRNP L-like (hnRNP LL), a closely related paralog of hnRNP L, was shown to regulate alternative splicing of CD45 exon 4 as well (Oberdoerffer *et al.*,

2008; Topp *et al.*, 2008). hnRNP LL expression was induced upon T-cell stimulation and promoted CD45 exon 4 skipping during T-cell activation.

hnRNP L and LL share 58% amino acid identity and are very similar in size and domain organisation (Fig. 1.7). Both proteins contain four RNA recognition motifs (RRMs) and a glycine-rich region at the N-terminus which is less pronounced in hnRNP LL. In HeLa cells hnRNP LL is about ten times less abundant than hnRNP L (Hung *et al.*, 2008). This observation and its participation in T-cell activation-induced alternative splicing suggest a tissue-specific role for hnRNP LL.



Figure 1.7

Domain structure of the hnRNP L proteins. Schematic representation of the domain structures of hnRNP L (P14866; 589 amino acids) and the closely related hnRNP L-like protein (Q53T80; 542 amino acids). Four canonical RNA recognition motifs (RRM) are represented by the red boxes. Glycine- and proline-rich regions are shown in blue and green, respectively.

1.8 Splicing and disease

Years ago it was estimated that 15% of point mutations leading to human genetic diseases disrupt splicing (Krawczak *et al.*, 1992). This estimation, however, is likely to be an underestimate since mutations in splicing regulatory elements, such as enhancer and silencer, had not been taken into account yet. By disrupting a splicing *cis*-element, either canonical splice site signals or additional regulatory elements, mutations can affect a single gene (Faustino & Cooper, 2003). The expression of multiple genes can be affected when a splicing regulatory factor is concerned. Some examples of splicing-associated diseases shall be illustrated in the following.

The first example, spinal muscular atrophy (SMA), is a recessive autosomal disorder characterised by degeneration of spinal cord motor neurons leading to muscle atrophy (Cartegni *et al.*, 2006; Frugier *et al.*, 2002). SMA represents one of the most common genetic causes of childhood mortality. The survival of motor neuron (SMN1) gene, which encodes an essential protein for assembly of ribonucleoprotein

complexes, is affected in the disease (Meister *et al.*, 2002). The *SMN1* gene is duplicated, resulting in the homologous copy called *SMN2*. The *SMN2* gene, however, is not able to completely compensate for the loss of *SMN1* protein since it carries a point mutation in exon 7 leading predominantly to skipping of the exon and resulting in a truncated protein that is non-functional. The nucleotide substitution is thought to disrupt a splicing regulatory element located in exon 7.

Myotonic dystrophy (DM) represents another example for a splicing-linked disease. DM is an autosomal dominant disorder caused by a CTG expansion in the 3' untranslated region of the DM protein kinase (DMPK) (Lukong *et al.*, 2008; Philips & Cooper, 2000). Unaffected individuals have less than 40 repeats whereas patients with the severe type of DM can have up to 1,500 CTG repeats. The disease severity correlates thereby with the repeat length. The repeat-containing transcripts accumulate in the nucleus and alter the function of RNA-binding proteins that are involved in alternative splicing (Kuyumcu-Martinez *et al.*, 2007; Lukong *et al.*, 2008). At least two proteins, muscleblind-like 1 (MBNL1) and CUG-binding protein (CUGBP1), are now so far to bind to the expanded CTG repeats. The observed effects of DM correspond to a loss of MBNL1 and a gain of CUGBP1 function supporting the idea that these proteins are misregulated. MBNL1 loss of function is due to sequestration of the protein on the expanded CTG repeats, whereas steady-state levels of CUGBP1 are increased in DM.

1.9 Global analysis of alternative splicing

Ever since alternative splicing was discovered the number of genes, which are thought to have alternative splice variants, increased constantly. Because of its importance high-throughput experimental approaches are used today for the genome-wide identification of alternative splicing events (Blencowe, 2006). The analysis of alternative splicing based on bioinformatics has become an important field over the past few years (Lee & Wang, 2005). The computational evaluation of expressed sequence tags (EST) was the first approach to identify alternative splicing events in larger scale. ESTs that come from the same gene are aligned to the genomic sequence which allows identification of differences that are consistent with alternative splicing (Modrek & Lee, 2002). This method, however, has its limitations due to experimental artefacts and biases. Some of the limitations have

been overcome recently by the development of custom microarrays and computational tools.

To identify alternative splicing targets of hnRNP L on a genome-wide level we used the Affymetrix GeneChip Human Exon 1.0 Array (Hung *et al.*, 2008). The array contains 1.4 million probe sets which interrogate more than 1 million exon clusters (<http://www.affymetrix.com>). One gene is covered by roughly 40 probes. Each exon is thereby represented by approximately four probes which allow detection of alterations in exon usage (Fig. 1.8).

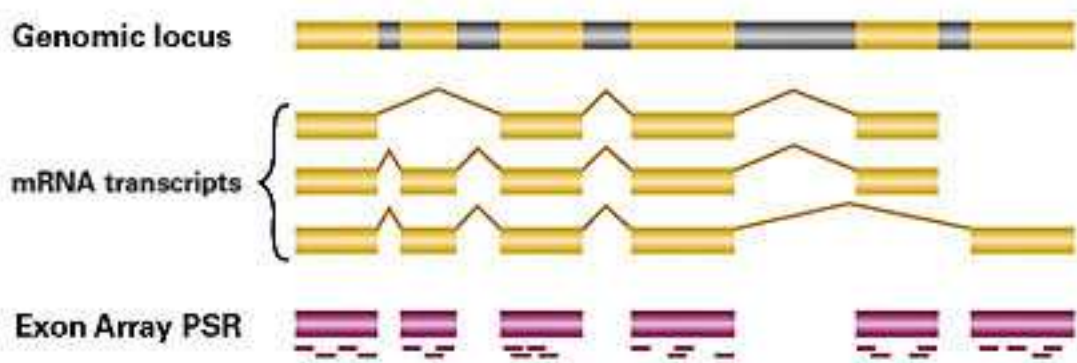


Figure 1.8

Schematic representation of the Affymetrix GeneChip Human Exon 1.0 Array design. The yellow and purple boxes display exons, the grey regions represent introns. Alternative splicing events are indicated. In the probe selection region (PSR) each exon is covered by several probes, which are shown as dashes (adapted from <http://www.affymetrix.com>).

1.10 Aim of the work

As described above, we identified the abundant nuclear protein hnRNP L as a global regulator of alternative splicing binding to CA-rich and CA repeat sequences. A genome-wide database search yielded few alternative splicing targets of hnRNP L (Hui *et al.*, 2005). Recently, we validated additional hnRNP L target genes with a combined microarray and RNAi approach (Hung *et al.*, 2008). On the basis of selected target genes obtained from database search as well as from the microarray approach I studied the mechanism of alternative splicing regulation by hnRNP L.

This work is divided into four parts. First, I characterised an intronic splicing silencer in the *SLC2A2* gene. This hnRNP L target gene contains an intronic CA-repeat sequence in close proximity to the 5' splice site of the regulated exon. I tested the

hypothesis that hnRNP L binding to the splicing silencer element interferes with recognition of the 5' splice site by the U1 snRNP.

TJP1, the second hnRNP L target gene I investigated, was identified by a combined microarray and RNAi approach. I demonstrated that a CA-rich cluster close to the 3' splice site of the regulated exon represents an intronic splicing silencer. Moreover, I studied the mechanism of exon repression mediated by hnRNP L binding to the silencer element.

Third, in the *ITGA2* gene I tested the influence of a CA-rich region in intron 1 on splicing of the corresponding exon. Recently, a polymorphic CA repeat in the mouse *ITGA2* gene was shown to enhance splicing efficiency through binding of hnRNP L (Cheli & Kunicki, 2006).

Finally, validation of the microarray data revealed a new role for hnRNP L in the regulation of alternative polyadenylation.

Taken together, I have obtained further insights into hnRNP L's mode of action as a global regulator of alternative splicing.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

2-mercaptoethanol	Roth
5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal)	Roche
Acetic acid	Roth
Acrylamide	Bio-Rad
Acrylamide/bisacrylamide 30, 37.5:1	Roth
Acrylamide/bisacrylamide 40, 19:1	Roth
Agarose ultra pure	Roth
Ammonium persulfate (APS)	Bio-Rad
Ampicillin	Roche
Bacto-agar	Roth
Bacto-tryptone	Roth
Bacto-yeast extract	Roth
Bisacrylamide	Bio-Rad
Boric acid	Roth
Bovine serum albumin, RNase free	Roche
Bromphenol blue	Merck
Calcium chloride	Merck
Chloroform	Roth
Coomassie brilliant blue R250	Merck
Creatine phosphate	Roche
Dimethyl pyrocarbonate (DMPC)	Sigma
Di-sodium hydrogenphosphate (Na_2HPO_4)	Merck
Dithioereitol (DTT)	Roche
Ethanol absolute	Roth
Ethidium bromide	Roth
Ethylendiaminetetraacetic acid (EDTA)	Roth
Formamide	Roth
Glucose	Sigma
Glycerol	Roth
Glycine	Roth
Glycogen	PeqLab
Guanidium thiocyanate	Roth
Heparin	Sigma
Imidazole	Roth
Isoamyl alcohol	Roth
Isopropanol	Roth
Isopropyl-1-thio- β -D-galactoside (IPTG)	Roche
Magnesium chloride	Merck
Methanol	Roth
N,N,N',N'-tetramethylenediamine (TEMED)	Bio-Rad

2. Materials and Methods

N-2-hydroxyethylpiperazine (HEPES)	Roth
Nonidet P-40 (NP-40)	Sigma
Phenylmethylsulfonyl fluoride (PMSF)	Roth
Polyoxyethyleneorbiten monolaurate (Tween 20)	Sigma
Polyvinylalcohol	Merck
Potassium chloride (KCl)	Roth
Psoralen	Sigma
Roti-phenol	Roth
Roti-phenol/chloroform	Roth
Sodium acetic acid (NaAc)	Merck
Sodium chloride (NaCl)	Roth
Sodium citrate	Roth
Sodium dihydrogen phosphate monohydrate ($\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$)	Merck
Sodium dodecyl sulfate (SDS)	Roth
Tris-hydroxymethylaminomethane (Tris)	Roth
Triton X-100	Merck
Trizol	Invitrogen
tRNA from yeast	Roche
Urea	Roth
Xylenxanol	Fluka

2.1.2 Nucleotides

$[\alpha\text{-}^{32}\text{P}]\text{ATP}$ (3,000 Ci/mmol)	Hartmann Analytic
Deoxynucleosidetriphosphate set (dNTP), 100 mM	Roth
m^7GpppG cap analog	Biozym
Ribonucleosidetriphosphate set (NTP), 100 mM	Roche

2.1.3 Enzymes and enzyme inhibitors

Expand reverse transcriptase, 50 U/ μl	Roche
Protease inhibitor cocktail tablets	Roche
Proteinase K, 10 $\mu\text{g}/\mu\text{l}$	Roth
Restriction endonucleases	New England Biolabs
RNase A, 100 mg/ml	Qiagen
RNase H, 10 U/ μl	Ambion
RNase inhibitor (RNase out),	Invitrogen
RQ1 RNase free DNase, 1 U/ μl	Promega
Shrimp alkaline phosphatase, 1 U/ μl	Roche
SP6 RNA polymerase, 20 U/ μl	New England Biolabs
T4 DNA ligase, 400 U/ μl	New England Biolabs
T7 RNA polymerase, 20 U/ μl	Fermentas
<i>Taq</i> DNA polymerase	Own purification

2.1.4 Reaction buffers

10x PCR buffer	Promega
----------------	---------

10x restriction enzyme buffer	New England Biolabs
10x RNase H buffer	PeqLab
10x RQ1 DNase buffer	Promega
10x SAP buffer	Roche
10x SP6 reaction buffer	New England Biolabs
5x Expand RT buffer	Roche
5x T4 DNA ligase buffer	New England Biolabs

2.1.5 Molecular weight markers

DNA Dig-labelled molecular weight marker VIII	Roche
GeneRuler™ DNA ladder mix	Fermentas
peqGold protein marker IV	PeqLab

2.1.6 Kits

Qiagen plasmid maxi kit	Qiagen
QIAprep spin miniprep kit	Qiagen
QIAquick gel extraction kit	Qiagen
RNeasy mini kit	Qiagen
Silver stain kit	Bio-Rad
SYBR Green Jumpstart Taq Readymix	Sigma
TOPO TA cloning kit	Invitrogen

2.1.7 Materials for mammalian cell culture

10x Posphate-buffered saline (PBS)	Invitrogen
1x trypsin-EDTA	Invitrogen
Dulbecco's modified Eagle's medium (DMEM)	Invitrogen
Fetal calf serum (FCS)	Invitrogen
GlutaMAX-1	Invitrogen
Opti-MEM	Invitrogen
Tissue culture dish	Greiner

2.1.8 Plasmids

pcDNA3.0	Invitrogen
pcDNA3-SLC2A2-WT, -sub	described in Hui <i>et al.</i> , 2005
pCR2.1 TOPO	Invitrogen
pFAST-BAC Htb-hnRNP L-GST	described in Hui <i>et al.</i> , 2005
pSP65-MINX	described in Zillman <i>et al.</i> , 1988

2.1.9 *E.coli* strains and mammalian cell lines

HeLa (human cervix carcinoma cells)	ATCC No. CCL-2
JM109 high-competent cells	Promega
TOP 10 high-competent cells	Invitrogen

2.1.10 Antibodies

Anti-goat Immunoglobulin-Peroxidase	Sigma
Anti-hnRNP L monoclonal antibody (4D11)	Sigma
Anti-hnRNP L peptide polyclonal antibody (D-17)	Santa Cruz Biotechnology
Anti-mouse Immunoglobulin-Peroxidase	Sigma
Anti-rabbit Immunoglobulin-Peroxidase	Sigma
Anti-U2AF65 monoclonal antibody (MC3)	Sigma
Anti- γ -tubulin monoclonal antibody (GTU88)	Sigma

2.1.11 DNA oligonucleotides

SLC2A2-C1	5' -TTAAAAGCTTGGGCTGAGGAAGAGACTGTG-3'
SLC2A2-C2	5' -TTAACTCGAGACTAATAAGAATGCCCGTGACG-3'
SLC2A2-C3	5' -CAGGGATATTGAGGGGCTTTCATTCAAGATA-3'
SLC2A2-C4	5' -TAAGAGCAATAGCTATTCCACAAGAAGAAAGA-3'
SLC2A2-C5	5' -TGGAATAGCTATTGCTCTTAGGTTAAAAAAAATC-3'
SLC2A2-C6	5' -AAAGCCCCTCAATATCCCTGAGTGCTACCA-3'
SLC2A2-C7	5' -CTCGGATCCACTAGTAACGGCCGCCCTACCTTTGTCTGAAAGTA-3'
SLC2A2-C8	5' -AGTGGATCCGAGCTCGGTACCAAGCTACTTACCACAATATAGTCCT-3'
M-SLC2A2 fwd	5' -CATTTCAATTCTGAAGCAGTCCAATGACTACCTACCTTTGTCTCGGAAAGTA-3'
M-SLC2A2 rev	5' -GACTGCTTCAGAATGAAATGCAATAATGCACTTACCACAATATAGTCCTG-3'
T7-SLC2A2s fwd	5' -TAATACGACTCACTATAGGGCATATCAGGACTATATTGTGG-3'
SLC2A2s rev	5' -ACATCCGCCTTTAGAGTTAC-3'
HU6-6	5' -TGTATCGTTCCAATTTTA-3'
U1 140-124	5' -CCCACTACCACAAATTA-3'
U1 14-1	5' -TGCCAGGTAAGTAT-3'
BGH rev	5' -TAGAAGGCACAGTCGAGG-3'
TJP1	5' -ATATCCTCCTTACTCACCACAAGC-3'
TJP2	5' -TTCAAACATGGTTCTGCCTC-3'
TJP-C1	5' -GATGAAGCTTCTGCTTTCTATAAAATATTTAAATATTTTAAATATAGTATTTCTGTTT ACTGCTAACT-3'
TJP-C2	5' -ACCGGAGTCTGCCATTACAC-3'
T7-TJP_exon19 fwd	5' -TAATACGACTCACTATAGGGATATCCTCCTTACTCACCAC-3'
TJP_exon20 rev	5' -GCAGAGGTTGATGATGCTG-3'

T7-TJP-1 fwd	5' -TAATACGACTCACTATAGGGTGGAAAGTTAGCAGTAAACAG-3'
TJP-22 rev	5' -AGGGACTGGAGATGAAGCT-3'
TJP-33 rev	5' -GCAACACCGCAGCACAGG-3'
M-TJP rev	5' -GATGAAGCTTCTGCTTTCTGCGAAGCGTTTAAAATATTTTAAATATAGTATTTCTGTTT ACTGCTAACTT-3'
M2-TJP rev	5' -GATGAAGCTTCTGCTTTCTGCGAAGCGTTTAAAATATTTTAAATATAGTG-3'
ITGA5	5' -TTAAAAGCTTGTCTCAGACCCAGGATGG-3'
ITGA6	5' -GCTTGACCTAAGTTGGGCTGCAGGACTC-3'
ITGA7	5' -CAGCCCAACTTAGGTCAAGCAAGTTTTCTTAA-3'
ITGA8	5' -TTAAGAATTCCAGTTGCCTTTTGGATTTATA-3'
ITGA9	5' -GTTTTAGGTAAGCATGGACAGTGTGGGG-3'
ITGA10	5' -TGTCCATGCTTACCTAAAACAAAGCACTCAC-3'
ITGA11	5' -CTCGGATCCACTAGTAACGGCCGCCAGTGTGCTAGCATGGACA-3'
ITGA12	5' -CCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTACCTAAAACA-3'
ITGA13	5' -AAGCTTGTCTCAGACCCAGGAT-3'
ITGA14	5' -ATTTTGTCTTCTGGGAGACC-3'
ITGA6-2	5' -TCTTACTAATCAGGGGAAGTTGGGCTG-3'
ITGA6-5	5' -TTGACCTAATCAAAGCCAGCAAGCACCG-3'
ITGA7-2	5' -ACTTCCCCTGATTAGTAAGATAATGAATTATGC-3'
ITGA7-5	5' -GCTGGCTTTGATTAGGTCAAGCAAGTTTTCTTA-3'
ASAH1	5' -GAGGAAATGAAGGGTATTGCC-3'
ASAH2	5' -ACTCCAAAATCCATGTTTCTCC-3'
ASAH3	5' -ATCACACCTCAATGGAACTTG-3'
ASAH-2	5' -GGTAAAGTTCACTTAGAAGCT-3'
hnRNP L fwd	5' -TTCTGCTTATATGGCAATGTGG-3'
hnRNP L rev	5' -GACTGACCAGGCATGATGG-3'
hnRNP LL fwd	5' -ACCATTCCTGGTACAGCACTG-3'
hnRNP LL rev	5' -TGGCCAGCACTTGTAAAGC-3'
β-actin 703	5' -TGGACTTCGAGCAAGAGATG-3'
β-actin 994	5' -GTGATCTCCTTCTGCATCCTG-3'

These DNA oligonucleotides were ordered from Sigma or MWG Biotech.

2.1.12 RNA oligonucleotides

5'-Biotin-(CA) ₃₂ -3'	Xeragon
human hnRNP L H1	5' -GAAUGGAGUUCAGGCGAUGTT-3'
human hnRNP LL	5' -AGUGCAACGUUUGUUUUAUATT-3'
luciferase GL2	5' -CGUACGCGGAUACUUCGATT-3'

The siRNA oligonucleotides were ordered from MWG Biotech.

2.1.13 Other materials

Eppendorf tube, 1.5 ml, 2 ml	Eppendorf
Falcon tube, 15 ml, 50 ml	Greiner
HeLa cell nuclear extract	4C Biotech
Hybond ECL nitrocellulose membrane	GE Healthcare
Nickel-nitrilotriacetic acid (Ni-NTA) agarose	Qiagen
Protein A-Sepharose	GE Healthcare
Roti-Block	Roth
Streptavidin-agarose	Sigma
Lipofectamin 2000	Invitrogen
X-ray film	Kodak

2.2 Methods

2.2.1 DNA cloning

2.2.1.1 Preparation of plasmid DNA

Plasmid DNA was isolated from bacterial culture using either QIAprep spin miniprep kit or QIAGEN plasmid maxi kit according to the manufacturer's instructions. The concentration of the plasmid DNA was determined by UV light absorption at 260 nm using a spectrophotometer (Eppendorf).

2.2.1.2 Agarose gel electrophoresis

Agarose was melted in 0.5x TBE buffer (100 mM boric acid, 100 mM Tris, 2 mM EDTA pH 8.8) using a microwave oven. After cooling, ethidium bromide was added in a 1:20,000 dilution and the gel was poured into a casting platform. DNA samples were mixed with 6x loading buffer (30% (v/v) glycerol, 0.025% (w/v) bromophenol blue) and loaded into the wells. 1-2% agarose gels were used to analyse DNA. Gels were run at 130 V for the appropriate time and visualised with a gel documentation system (SynGene).

2.2.1.3 Preparation of DNA fragments

DNA bands were excised from agarose gels and purified using QIAquick gel extraction kit according to the manufacturer's instructions.

2.2.1.4 Restriction endonuclease digestion

The DNA sample was mixed with 1x reaction buffer and restriction enzyme in a final volume of 50 µl. An amount of 1 to 5 units of enzyme was used to digest 1 µg of DNA. The reaction was incubated for 1 to 2 h at the recommended temperature. DNA was purified by gel extraction or phenolisation.

2.2.1.5 Dephosphorylation

After linearisation of plasmid vector DNA by endonuclease digestion, the vector was dephosphorylated at the 5'-terminus to prevent self-ligation. 10 µg of linearised vector DNA was incubated with 1x SAP buffer and 1 U/µl SAP in a total volume of 50 µl for 30 min at 37°C. The enzyme was heat inactivated afterwards by incubation at 65°C for 15 min.

2.2.1.6 Ligation

Appropriate amounts of linearised vector DNA and purified DNA fragment were incubated with 1x T4 ligation buffer and 1 µl T4 DNA ligase (400 U/µl) in a final volume of 10 µl. The incubation was carried out at either 16°C o/n or for 2 h at room temperature.

2.2.1.7 Transformation

10 µl of ligation reaction was mixed with 200 µl of competent JM109 *E.coli* cells and incubated on ice for 30 min. Cells were heat shocked at 42°C for 90 sec and

immediately chilled on ice. 500 µl of LB medium (1% (w/v) Bacto-tryptone, 0.5% (w/v) Bacto-yeast extract, 1% (w/v) NaCl) was added to each transformation culture. Cultures were grown at 37°C for 1 h, aliquots of each culture were plated on pre-warmed LB plates (1.5% (w/v) Bacto-agar in LB medium) containing 100 µg/ml ampicillin. The plates were incubated overnight at 37°C. Single-cell colonies were picked and grown in LB medium. Plasmids isolated from these clones were analysed by PCR and digestion with restriction endonucleases. Finally, plasmids were confirmed by sequence analysis (SeqLab, Göttingen).

TOPO cloning was carried out in a similar way. Aliquots of the ligation reaction were transformed in TOP10 high-competent *E.coli* cells. For selection of recombinants by blue/white screening 40 µl of X-Gal (40 mg/ml) and 40 µl of IPTG (100 mM) were spread on each LB plate.

2.2.2 Minigene constructs

2.2.2.1 pcDNA3-SLC2A2

Construction of *SLC2A2* wildtype minigene construct was described by Hui *et al.*, 2005. For construction of the substitution minigene, a series of PCRs was performed using oligonucleotides SLC2A2-C1 and -C2 in combination with M-SLC2A2 fwd and rev. The whole fragment, including restriction sites, was obtained by a second-step PCR using the previous PCR products as templates, cut with *HindIII* and *XhoI*, and inserted into the corresponding sites of the pcDNA3 vector. The CA repeat was replaced by a random sequence.

DNA templates for [α -³²P]ATP-labelled *in vitro* transcription were obtained by PCR using pcDNA3-SLC2A2 plasmid DNAs as template and oligonucleotides T7-SLC2A2s fwd and SLC2A2s rev.

2.2.2.2 pcDNA3-TJP1

TJP1 minigene constructs consist of three exons and two introns. The *TJP1* genomic sequence for the minigene constructs was amplified using genomic DNA

isolated from primary HUVEC cells as template (kindly provided by Dr. Karl Stangl, Charité, Berlin) and oligonucleotides TJP-C2/2. The minigene unit was first TOPO-cloned into pCR2.1 and then released by *EcoRI* digestion. The released fragment was re-cloned into pcDNA3 vector. To construct the *TJP1* mutant derivative, a PCR was carried out using oligonucleotides T7 and TJP-C1, which carried five point mutations. The PCR product was digested with *HindIII*, followed by substituting the *HindIII* fragment in the wildtype minigene with it. For the *TJP1* mutG and mutTG constructs the oligonucleotide combinations T7/M-TJP rev respectively T7/M2-TJP rev were used. These constructs were made by Dr. Jingyi Hui, Institute of Biochemistry and Cell Biology, Chinese Academy of Science, Shanghai.

As described above for *SLC2A2*, *TJP1* templates for *in vitro* transcription were obtained by PCR. The shorter PCR product of 84 nt was obtained using the oligonucleotide combination T7-TJP-1 fwd/TJP-22 rev, whereas for the longer product of 318 nt the oligonucleotide TJP-33 rev was used as reverse primer.

2.2.2.3 pcDNA3-ITGA2

All *ITGA2* minigene constructs contain the first two exons and a shortened intron 1 in between.

In the ITGA2a minigene constructs most of the intron sequence was deleted, leaving ~400 nt downstream of the 5' splice site and ~100 nt upstream of the 3' splice site. The minigene unit was constructed using oligonucleotide combinations ITGA 5/6, ITGA 7/8, and ITGA 5/8 for a second-step PCR, and HUVEC genomic DNA as template. The PCR product was digested with *HindIII* and *EcoRI* and cloned into the corresponding sites of the pcDNA3 vector. To obtain the substitution derivative, a similar two-step PCR was carried out using oligonucleotide combinations ITGA 5/12, ITGA 8/11, and ITGA 5/8. In all *ITGA2* substitution minigene constructs, the CA-rich sequence was replaced by a non-specific sequence amplified from the polylinker region of pcDNA3. In the CA0 minigene construct, the complete CA-rich sequence in intron 1 was deleted using oligonucleotide combinations ITGA 5/10, ITGA 8/9, and ITGA 5/8. These constructs were made by Marius Prohm.

In the ITGA2b minigene constructs more of the sequence of intron 1 was included, leaving ~500 nt downstream of the 5' splice site and again ~100 nt upstream of the

3' splice site. The oligonucleotide combinations for the two-step PCR were ITGA 5/6-5, ITGA 7-5/8, and ITGA 5/8. For the substitution derivative the two-step PCR was carried out with the same oligonucleotides already described for the ITGA2a substitution construct.

In the ITGA2c minigene constructs intron sequence from ~400 nt downstream of the 5' splice site to ~200 nt upstream of the 3' splice site was deleted. ITGA 5/6-2, ITGA 7-2/8, and ITGA 5/8 were the oligonucleotide combinations used for the two-step PCR. The substitution derivative was constructed as described above.

2.2.3 *In vivo* splicing analysis

2.2.3.1 Cell culture

HeLa cells were maintained in DMEM supplemented with 10% FCS at 37°C, and 5% CO₂. When the cells were 100% confluent (every 2-3 days), they were washed once with 1x PBS, detached from the plate by 1x trypsin-EDTA, split, and reseeded into new dishes.

2.2.3.2 Transient transfection

One day before transfection, 6×10^5 cells were seeded in 6 cm dishes. For transfection cells were not more than 80% confluent. 8 µg of plasmid DNA was mixed with 65 µl 1M calcium chloride solution in a polystyrene tube (Greiner). As a negative control (mock), plasmid DNA was left out. Sterile ddH₂O was added to each reaction to a final volume of 260 µl. 260 µl of 2x HBS (50 mM HEPES pH 6.96, 280 mM NaCl, 10 mM KCl, 1.5 mM Na₂HPO₄, 12 mM glucose) was added dropwise to the transfection mix with constant vortexing. After 20 min incubation at room temperature, the precipitate was applied dropwise to the cell medium.

2.2.3.3 Isolation of total RNA from HeLa cells

Two days after transfection cells were harvested and total RNA was isolated using guanidium thiocyanate. First, cells were washed once with 3 ml of ice-cold 1x PBS. Then, cells were lysed by applying 2 ml of solution A (50% phenol, 2 M guanidium thiocyanate, 12.5 mM sodium citrate pH 7.0, 100 mM sodium acetate pH 4.0, 100 mM β -mercaptoethanol). The lysate was transferred into a 15-ml falcon tube; 300 μ l of chloroform/isoamyl alcohol (24:1) were added and mixed by vortexing. The lysate was kept on ice for 30 min and then distributed into two 2 ml tubes for centrifugation. After centrifugation at 14,000 rpm for 20 min at 4°C the upper aqueous phase was transferred into a new tube. Nucleic acids were precipitated by addition of 0.7 volumes of isopropanol and incubation at -80°C for 15 min. Total RNAs were pelleted by centrifugation at 14,000 rpm for 20 min at 4°C, washed once with 500 μ l of 70% ethanol, dried, and resuspended in 10 μ l of DMPC-H₂O. The concentration of total RNA was determined by UV light absorption at 260 nm.

2.2.3.4 RQ1 DNase treatment

To remove any co-precipitated DNA, total RNA was treated with RQ1 DNase. 20 μ g of total RNA was mixed with 5 μ l 10x RQ1 DNase buffer, 5 μ l 100mM DTT, 1 μ l RNase inhibitor, and 5 μ l RQ1 DNase. DMPC-H₂O was added to a final volume of 50 μ l. After incubation at 37°C for 30 min, 150 μ l DMPC-H₂O was added to each reaction. Total RNAs were extracted with 200 μ l phenol/chloroform/isoamyl alcohol (25:24:1) and then precipitated with 20 μ l 3M NaAc pH 5.2 (1/10 vol), 600 μ l ethanol (3 vol), and 1 μ l glycogen. After pelleting, washing, and drying, total RNA was dissolved in 10 μ l DMPC-H₂O. The RNA concentration was determined again.

2.2.3.5 Analysis of *in vivo* splicing by RT-PCR

Reverse transcriptions were carried out by *Expand* RT. 2.5 μ l of total RNA was mixed with 4 μ M BGH rev primer in a total volume of 6.5 μ l and incubated at 65°C for 10 min. Reactions were immediately chilled on ice. The RT reaction containing 1x RT buffer, 8 mM DTT, 800 μ M dNTPs, 1 U/ μ l RNase inhibitor, 2.5 U/ μ l *Expand*

RT and DMPC-H₂O in a final volume of 20 µl was added to each reaction. Reverse transcriptions were carried out for 1 h at 42°C.

The PCR was done in 25-µl reactions with 5 µl RT reaction, 1x PCR buffer, 400 µl dNTPs, 1 mM MgCl₂, 600 µM forward primer, 600 µM reverse primer, and 0.2 U/µl *Taq* DNA polymerase. The following amplification profile was applied: 2 min of denaturation at 95°C, 30 cycles of amplification (30 sec at 95°C, 30 sec at 58°C, and 30 sec at 72°C), and a final elongation step for 7 min at 72°C. In the case of *TJP1*, the annealing temperature was increased to 62°C. 5 µl of PCR reaction were analysed on a 2% agarose gel by ethidium bromide staining. The GeneTools software (version 3.07; SynGene) was used for quantification of stained bands.

2.2.4 *In vitro* transcription

2.2.4.1 Transcription of ³²P-labelled RNA

RNAs were radioactively-labelled internally by T7 *in vitro* transcription. PCR products were used as templates for transcription. Only in the case of MINX, RNA was transcribed from a linearised plasmid. 5 µl of template DNA or 1 µl of plasmid DNA (1 µg/µl) were mixed with 5 µl 5x transcription buffer, 2.5 µl 100 mM DTT, 1.25 µl 2 mM ATP, 1.25 µl 10 mM CTP, 1.25 µl 10 mM UTP, 1.25 µl 2 mM GTP, 1.25 µl m⁷GpppG cap analog, 0.5 µl RNase inhibitor, 2 µl [α-³²P]ATP (3000 Ci/mmol), and 1 µl T7 RNA polymerase (20 U/µl). With addition of DMPC-H₂O, the final volume was adjusted to 25 µl. For MINX transcription, SP6 RNA polymerase was used instead of T7. Transcriptions were carried out at 37°C for 1 h. 2 µl RQ1 DNase was added to each reaction and incubation continued for 30 min at 37°C.

2.2.4.2 Removal of unincorporated nucleotides by gel filtration

Unincorporated nucleotides were removed from the transcription reactions using RNA spin columns following the manufacturer's instructions.

Transcribed RNAs were precipitated with 600 µl ethanol, 20 µl 3M NaAc pH 5.2, and 1 µl glycogen. After pelleting, washing and drying, the amount of RNA was

measured using a scintillation counter ($\text{RNA [ng]} = \text{ATP}_{\text{total cold}} [\mu\text{M}] \times \text{volume of reaction } [\mu\text{l}] \times \% \text{incorporation} \times 0.0132$). The transcripts were dissolved in an appropriate volume of DMPC-H₂O.

2.2.4.3 Transcription without ³²P-label

Transcription reactions were carried out as described above. Only, ATP was added in a final concentration of 500 μM instead of 100 μM because [α -³²P]ATP was omitted. After transcription and DNase treatment, transcribed RNAs were extracted with 200 μl phenol/chloroform/isoamyl alcohol (25:24:1) and then precipitated. After pelleting, washing, and drying, transcripts were dissolved in 5 μl of DMPC-H₂O.

2.2.4.4 Determination of RNA concentrations

1 μl of each transcription reaction was mixed with 4 μl DMPC-H₂O and 5 μl 2x formamide loading buffer (80% (v/v) formamide, 10 mM EDTA pH 8.0). Samples were heated for 2 min at 95°C and immediately chilled on ice. Transcripts were analysed on a 2% agarose gel and ethidium bromide staining. The concentration of RNA was estimated by comparison to tRNA standards. Transcripts were diluted with DMPC-H₂O to a final concentration of 10 ng/ μl .

2.2.5 *In vitro* splicing of pre-mRNAs

2.2.5.1 Splicing reaction

In vitro transcribed pre-mRNAs were spliced in HeLa cell nuclear extract. 10 ng of pre-mRNA was incubated in a 25- μl reaction containing 60% HeLa nuclear extract, 0.5 mM ATP, 3.2 mM MgCl₂, 20 mM creatine phosphate, 1.6 U/ μl RNase inhibitor, and 2.66% (v/v) PVA. The splicing reaction was incubated at 30°C. 20 μl aliquots were taken at different time points and stored at -20°C.

2.2.5.2 Proteinase K treatment

Aliquots of the splicing reaction were mixed with 100 μ l 2x PK buffer (200 mM Tris/HCl, 300 mM NaCl, 25 mM EDTA, 2% SDS), 4 μ l PK (10 mg/ml), and 76 μ l DMPC-H₂O. Reactions were incubated for 30 min at 37°C. The RNA was extracted with phenol, precipitated, washed, dried and dissolved in 5 μ l DMPC-H₂O.

2.2.5.3 Analysis of *in vitro* splicing by RT-PCR

Reverse transcriptions were carried out as described above (2.2.3.5). 1 μ l of RNA, purified from a splicing reaction, was mixed with 2 μ M of a gene-specific reverse primer. 5 μ l of the RT reaction was used as template in the PCR assay. 5 μ l of each PCR reaction was analysed by agarose gel electrophoresis and ethidium bromide staining.

2.2.6 Depletion of hnRNP L from HeLa nuclear extract

HnRNP L was depleted from nuclear extract with a 5'-biotinylated (CA)₃₂ RNA oligonucleotide that was bound to streptavidin agarose beads. First, 200 μ l of packed streptavidin agarose beads were blocked in 500 ml blocking solution (4 mM HEPES pH 8.0, 0.2 mM DTT, 2 mM MgCl₂, 20 mM KCl, 0.002% (v/v) NP-40, 0.2 mg/ml tRNA, 1 mg/ml BSA, 0.2 mg/ml glycogen) at 4°C o/n. The blocked beads were washed four times with 1 ml of WB 400 (20 mM HEPES pH 8.0, 1 mM DTT, 10 mM MgCl₂, 400 mM KCl, 0.01% (v/v) NP-40). For each depletion reaction, 20 μ l of packed beads were then incubated with 6 μ g of the 5'-biotinylated (CA)₃₂ RNA oligonucleotide in 200 μ l of WB 400 for 4 h at 4°C with rotation. A mock depletion was done in the absence of RNA oligonucleotide. Beads were washed four times with 1 ml of WB 400 and one time with buffer D (20 mM HEPES pH 8.0, 100 mM KCl, 0.5 M EDTA, 20% (v/v) glycerol, 1 mM DTT, 1 mM PMSF), followed by incubation with 200 μ l of HeLa nuclear extract for 30 min at 30°C with rotation. Then, the KCl concentration was increased to 600 mM and the incubation continued for 20 min at 4°C. After removal of the streptavidin beads, depleted nuclear extracts were dialysed against buffer D for 2 h.

2.2.7 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

4x stacking gel buffer containing 0.5 M Tris/HCl pH 6.8, 0.4% (w/v) SDS and a 4x separating gel buffer with 1.5 M Tris/HCl pH 8.8, 0.4% (w/v) SDS were used. The stacking gel (5% Acrylamide/bisacrylamide 37.5:1, 1x stacking gel buffer, 200 μ l APS, 20 μ l TEMED in 20 ml) covers the separating gel (10-12% Acrylamide/bisacrylamide 37.5:1, 1x separating gel buffer, 100 μ l APS, 10 μ l TEMED in 10 ml). Protein gels were run in SDS running buffer (25 mM Tris, 250 mM glycine pH 8.3, 0.1% (w/v) SDS) at 100 V until the samples reached the separating gel. Then, the voltage was increased to 150 V and run for an appropriate time. Gels were subjected to either Western blot analysis or Coomassie blue staining.

2.2.8 Coomassie staining

For Coomassie staining, the gel was incubated for 1 h in Coomassie blue staining solution (0.25% (w/v) Coomassie brilliant blue R250, 50% (v/v) methanol, 10% (v/v) acetic acid) with shaking and then destained with Coomassie blue destaining solution (50% (v/v) methanol, 10% (v/v) acetic acid) until the background became clear. The gel was dried using a vacuum gel dryer.

2.2.9 Western blot

Protein samples were first separated on a 10% SDS-PAGE and then transferred to Hybond ECL nitrocellulose membrane for 30 min at 300 mA using a semi-dry transfer cell (Bio-Rad). The protein transfer buffer contained 50 mM Tris, 380 mM glycine, 20% (v/v) methanol, 0.02% (w/v) SDS. The membrane was blocked overnight in blocking buffer (1x PBS, 1x Roti-Block, 0.2% (v/v) Tween). Before addition of the first antibody, fresh blocking buffer was applied to the membrane. Antibodies were diluted in blocking buffer as follows: anti-hnRNP L polyclonal peptide antibody (D-17), 1:500; anti- γ -tubulin monoclonal antibody (GTU-88), 1:10,000; anti-U2AF65 monoclonal antibody (MC3), 1:10,000. Incubation with the first antibody was carried out for 1 h. The membrane was washed three times for 10 min with washing buffer (1x PBS, 0.2% (v/v) Tween). The second, peroxidase

conjugated antibody was applied for 1 h in blocking buffer in the following dilutions: anti-goat IgG peroxidase-conjugate, 1:500; anti-mouse IgG peroxidase-conjugate 1:10,000. The membrane was washed again three times for 10 min with washing buffer. The chemoluminescents reaction was carried out by application of peroxidase substrate (1x ECL, 30% (v/v) H₂O₂ 1000:1) for 1 min and detection by autoradiography.

For Western blot analysis of hnRNP L and LL knockdown, cells were lysed after RNAi in WB100 (20 ml HEPES pH 7.5, 1 mM DTT, 10 mM MgCl₂, 100 mM KCl, 1% (v/v) NP-40) on ice for 10 min. For hnRNP LL, cell lysates were incubated with 20 µl of streptavidin beads with pre-bound 5'-biotinylated (CA)₃₂ RNA oligonucleotide (see 2.2.6). After washing once with WB400 and three times with WB100, SDS-PAGE loading buffer was added to the beads. Cell lysates (for hnRNP L) and (CA)₃₂ affinity selected material (for hnRNP LL) were separated on a 10% SDS-PAGE. The Western blot for hnRNP L was carried out as described above. For hnRNP LL detection, anti-hnRNP LL polyclonal antibody (1:100) was used as primary antibody and anti-rabbit IgG peroxidase-conjugate as the secondary antibody.

2.2.10 Purification of recombinant proteins

The generation of baculovirus-expressed His-tagged hnRNP L was described before (Hui *et al.*, 2005). Note that this hnRNP L expression construct (558 aa) is 31 amino acids shorter than the annotated hnRNP L protein (NCBI, P14866) because it was annotated before with a different transcriptional start. Ni-NTA beads were used to purify recombinant hnRNP L from Sf21 cell pellet. The cell pellet from one 150 cm² TC flask was resuspended in 10 ml of lysis buffer (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 10 mM imidazole, 1% (v/v) NP-40, protease inhibitor) and incubated on ice for 10 min. After centrifugation at 14,000 rpm for 10 min at 4°C, the supernatant was incubated with 100 µl of packed Ni-NTA beads for 2 h at 4°C with rotation. The beads were washed four times with 1 ml of wash buffer (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.05% (v/v) NP-40). Finally, bound proteins were eluted from the beads with 200 µl of elution buffer (50 mM NaH₂PO₄ pH 8.0, 300 mM NaCl, 250 mM imidazole, 0.05% (v/v) NP-40). The eluate was dialysed against buffer D for 2 h at 4°C. Recombinant GST-U2AF65 was kindly provided by Prof. Dr. Juan Valcárcel, CRG, Barcelona.

2.2.11 Spliceosome assembly reaction

The protocol described here is a modification of the method described by Das & Reed, 1999.

5 ng of ^{32}P -labelled substrate RNAs were incubated in 60% HeLa nuclear extract with 0.8 U/ μl RNase inhibitor under E complex conditions in a total volume of 25 μl . ATP was depleted from the nuclear extract before use by incubation at room temperature for 30 min. The spliceosome assembly reaction was carried out at 30°C. After 0, 10, and 30 min of incubation, a 3- μl aliquot was taken and mixed with 3 μl of 5x loading dye (1x TBE, 20% (v/v) glycerol, 0.25% (w/v) BPB, 0.25% (w/v) XC). Reactions were separated at 4°C on a 1.5% horizontal low-melting-point agarose gel for 2.5 h at 100 V. The running buffer contained 50 mM Tris and 50 mM glycine. The gel was fixed in 10% acetic acid, 10% methanol for 30 min, and afterwards dried under vacuum at 70°C. Spliceosomal complexes were visualized by autoradiography.

2.2.12 Psoralen crosslinking

The protocol for Psoralen crosslinking was modified after Wassarman, 1993.

^{32}P -labelled substrate RNAs (100,000 cpm) were incubated in 60% HeLa nuclear extract under splicing conditions (2.2.5.1). After 30 min of incubation at 30°C, psoralen was added to a final concentration of 20 $\mu\text{g}/\text{ml}$. For the time-course experiment, 25- μl aliquots were taken at different time points. The samples were transferred, as drops, to a sheet of parafilm that has been placed on a bed of ice. The ice-container was then covered with a glass plate. Samples were placed under an UV lamp (350 nm) with a distance of 5 cm to the light source and irradiated for 30 min. Then, samples were transferred into a 1.5-ml tube. Proteinase K treatment was carried out as described before (2.2.5.2). Total RNA was extracted with phenol and recovered by ethanol precipitation. The purified RNAs were then either separated directly on an 8% denaturing polyacrylamide gel or subjected to RNase H cleavage.

2.2.13 Preparation of total RNA from HeLa nuclear extract

15 µl of HeLa nuclear extract was incubated with 100 µl 2x PK buffer, 4 µl PK (10 mg/ml), and 81 µl DMPC-H₂O for 30 min at 37°C. Total RNA was extracted with phenol and precipitated with ethanol. After washing and drying, total RNA was dissolved in 15 µl of DMPC-H₂O.

2.2.14 RNase H cleavage

DNA oligonucleotides used: U1 14-1, U1 140-124, HU6-6.

5 µl of RNA, purified after psoralen crosslinking or from HeLa nuclear extract, was mixed with 1x RNase H buffer, 5 U/µl RNase H, and 50 ng/µl DNA oligonucleotide in a final volume of 20 µl. Reactions were incubated for 30 min at 37°C followed by phenol extraction and ethanol precipitation. Finally, RNAs were analysed on an 8% denaturing polyacrylamide gel and visualised by autoradiography. RNase H-treated total RNA, purified from HeLa nuclear extract, was analysed by silver staining.

For U1 inactivation in HeLa nuclear extract the U1 14-1 DNA oligonucleotide was used. 20 ng/µl of the DNA oligonucleotide was mixed with 60% HeLa nuclear extract, 5 U/µl RNase H under splicing conditions (2.2.5.1), and incubated for 30 min at 30°C. The U1-inactivated nuclear extract was used in either psoralen crosslinking assays or subjected to Proteinase K treatment (2.2.5.2). Total RNA was purified and analysed on an 8% denaturing polyacrylamide gel by silver staining.

2.2.15 Silver staining

RNA samples were separated on an 8% denaturing polyacrylamide gel. A Silver staining kit (BioRad) was used according to the manufacturer's protocol. First, the gel was fixed for 1 h in 40% methanol, 10% acetic acid, followed by rinsing twice with 10% ethanol, 5% acetic acid for 15 min. The gel was washed with distilled water before impregnation with silver reagent for 20 min. Again, the gel was washed with distilled water and developed until the bands reached the desired intensity. To reduce background staining, the developer was changed several times. The reaction was stopped by incubation in 5% acetic acid for 15 min.

2.2.16 Electromobility shift assay (band shift)

100 nM of ^{32}P -labelled substrate RNAs were incubated with different amounts of recombinant hnRNP L protein (200, 400, or 800 nM). The reactions contained in addition 300 ng/ μl tRNA and 2 U/ μl RNase inhibitor, adjusted to a final volume of 10 μl with buffer D (20 mM HEPES pH 8.0, 100 mM KCl, 100 mM MgCl_2 , 0.5 M EDTA, 20% (v/v) glycerol, 1 mM DTT). Reactions were incubated on ice for 30 min and afterwards separated on a 6% native gel (6% acrylamide/bisacrylamide 80:1, 1x TBE, 400 μl 10% (v/v) APS, 40 μl TEMED filled up with H_2O to 50 ml). The gel was pre-run for 30 min at 250 V. After loading of the samples, the gel was run for 30 min at 50 V and then 1 h 40 min at 250 V. Samples were visualised by autoradiography. For electromobility shift assays with U2AF65, 30 nM of ^{32}P -labelled substrate RNAs were used. They were incubated as described above with increasing amounts of recombinant U2AF65 (0.3, 1, 4, 5, 10, 20 μM)

2.2.17 UV crosslinking and immunoprecipitation

2.2.17.1 UV crosslinking with purified proteins

^{32}P -labelled substrate RNA (20,000 cpm) was incubated with 1.5 mM ATP, 4.1 mM MgCl_2 , 30 mM creatine phosphate, 300 ng/ μl tRNA, 300 ng recombinant hnRNP L adjusted to a final volume of 10 μl with buffer D. Reactions were incubated at either 4 or 30°C for 30 min. As negative control, RNA was incubated alone without addition of recombinant protein. UV crosslinking was carried out at 500 mJ for 3 min. Crosslinked samples were incubated with 1.5 $\mu\text{g}/\mu\text{l}$ RNase A at 37°C for 30 min. Crosslinked proteins were separated on a 12% SDS-PAGE and visualised by autoradiography.

2.2.17.2 UV crosslinking in HeLa nuclear extract

^{32}P -labelled substrate RNA (200,000 cpm) was incubated in 25- μl reactions containing 60% HeLa nuclear extract (hnRNP L- or mock-depleted), 32 nM HEPES

pH 8.0, 0.5 mM ATP, 1.6 mM MgCl₂, 20 mM creatine phosphate, 0.8 U/μl RNase inhibitor. For complementation analysis, 200 ng of recombinant hnRNP L was added to the hnRNP L-depleted nuclear extract. The splicing reactions were incubated at 30°C for 30 min and then irradiated with UV-light (254 nm) at 400 mJ for 72 sec. Then, samples were treated with 60 ng/μl RNase A for 30 min at 37°C.

2.2.17.3 Immunoprecipitation

For immunoprecipitation of crosslinked proteins, either anti-hnRNP L monoclonal antibody (4D11) or anti-U2AF65 monoclonal antibody (MC3) were used. After RNase A treatment crosslinked samples were incubated with 3 μl of anti-hnRNP L or anti-U2AF65 antibody, respectively, in 25 μl NP-40 buffer (50 mM Tris/HCl pH 8.0, 150 mM NaCl, 1% (v/v) NP-40) for 1h at 4°C with constant shaking. Then, 25 μl of packed Protein A-sepharose beads and 100 μl of NP-40 buffer were added to each reaction. Incubation was carried out for 1h at 4°C on a rotating wheel. Beads were washed three times with 500 μl of High Salt buffer (50 mM Tris/HCl pH 8.0, 500 mM NaCl, 1% (v/v) NP-40), followed by washing once with 1 ml of NP-40 buffer. After removal of residual buffer, beads were boiled in 10 μl 4x SDS loading buffer (250 mM Tris/HCl pH 6.8, 40% (v/v) glycerol, 4% (w/v) SDS, 0.025% (w/v) bromophenol blue) for 10 min. Proteins were separated on a 10% SDS-PAGE and visualised using a phosphor imager.

2.2.18 RNAi in HeLa cells

2.2.18.1 siRNA knockdown

One day before transfection, 4.3×10^5 HeLa cells were seeded in 10-cm culture dishes to be 30% confluent on the next day. The siRNA duplex (at a final concentration in culture medium of 30 nM) was transfected with Lipofectamine 2000 according to the manufacturer's protocol.

2.2.18.2 RNA isolation

Four days after transfection, total RNA was isolated using Trizol and RNeasy kit. Cells were washed once with 3 ml of ice-cold 1x PBS. Then, cells were lysed by resuspending them in 5 ml Trizol. The lysate was transferred into 15-ml falcon tubes and vortexed for 1 min. After 2-3 min incubation at room temperature, 1 ml of chloroform (0.2 vol.-%) was added and mixed by vortexing. Samples were incubated for 5 min at room temperature and then centrifuged at 12,000 g for 15 min at 4°C. The upper aqueous phase was transferred into a new 15-ml falcon tube and 1 vol.-% of 70% ethanol was added to each sample and after mixing applied to RNeasy columns. The columns were centrifuged at 10,000 rpm for 15 min at room temperature to allow binding of RNA. Columns were washed with 700 µl RW1 buffer and after centrifugation transferred into a new 1.5-ml tube. Columns were washed again with 500 µl of RPE buffer and centrifuged at 10,000 rpm for 2 min. To remove residual buffer, columns were centrifuged again for 1 min. The RNA was eluted from the column with 30 µl of DMPC-H₂O.

2.2.18.3 Real-time PCR analysis

The first step of reverse transcription was carried out using 5 µg of total RNA, 8 µM oligo d(T)₁₈ primer in a total volume of 6.5 µl and incubation at 65°C for 10 min. The reactions were immediately chilled on ice. The RT reaction containing 1x RT buffer, 8 mM DTT, 800 µM dNTPs, 1 U/µl RNase inhibitor, 2.5 U/µl *Expand* RT and DMPC-H₂O in a final volume of 20 µl was added to each reaction. Reverse transcriptions were carried out for 1 h at 42°C.

Real-time PCR assays for hnRNP L, LL and β-actin were performed in an iCycler (Bio-Rad) using SYBR Green Jumpstart Readymix kit and oligonucleotide combinations hnRNP L fwd/rev, hnRNP LL fwd/rev, and β-actin 703/994. 5 µl of the RT reaction was mixed with 12.5 µl 2x SYBR Green mix, 800 µM forward primer, and 800 µM reverse primer in a final volume of 20 µl. The following amplification profile was applied: 3 min of denaturation at 95°C, 40 cycles of amplification (30 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C). The relative expression levels of hnRNP L and LL normalised to β-actin were determined with the Gene Expression

Macro (version 1.1; Bio-Rad) and presented as fold change in gene expression relative to luciferase control. A melting curve analysis confirmed that no primer dimers were generated. All primer pairs resulted in almost 100% amplification efficiency.

2.2.19 Microarray analysis

For data analysis and detection of alternative splicing targets see Hung *et al.* (2008).

2.2.19.1 RT-PCR validation of hnRNP L target genes

Total RNA was isolated after knockdown of hnRNP L, LL, or L/LL in HeLa cells using Trizol and RNeasy kit (see 2.2.18.2). RT-PCR reactions were carried out as described above (2.2.3.5). For *TJP1*, the oligonucleotide combination TJP1/TJP2 was used. For *ASAH1*, a combination of three oligonucleotides ASAH1/ASAH2/ASAH3 was used for RT-PCR analysis. In a second assay, reverse transcription was carried out using an oligo d(T)₁₈ primer. For PCR the oligonucleotide combination ASAH-2/oligo d(T)₁₈ was used. Ethidium bromide-stained bands were quantitated by TINA software, version 2.07d.

3. Results

3.1 *SLC2A2*

The *SLC2A2* (solute carrier family 2) gene, also referred to as *GLUT2* (glucose transporter 2), encodes a glucose transporter-like protein expressed predominantly in liver, insulin-producing β -cells, small intestine and kidney (Fukumoto *et al.*, 1988; Thorens *et al.*, 1988). It is an integral glycoprotein of the plasma membrane, mediating facilitated bidirectional transfer of glucose. Due to its low affinity to glucose, the protein may be part of a glucose-sensing mechanism in β -cells. Defects in the *SLC2A2* gene are the cause of Fanconi-Bickel syndrome, a rare inherited disorder of carbohydrate metabolism (Santer *et al.*, 1997), and suspected as the possible basis of inherited susceptibility to diabetes. The *SLC2A2* gene (NM_000340), spanning approximately 30 kb, consists of 11 exons (Fig. 3.1), one of those, exon 4, is alternatively spliced. In intron 4 resides a (CA)₁₉ repeat sequence starting at position +7.

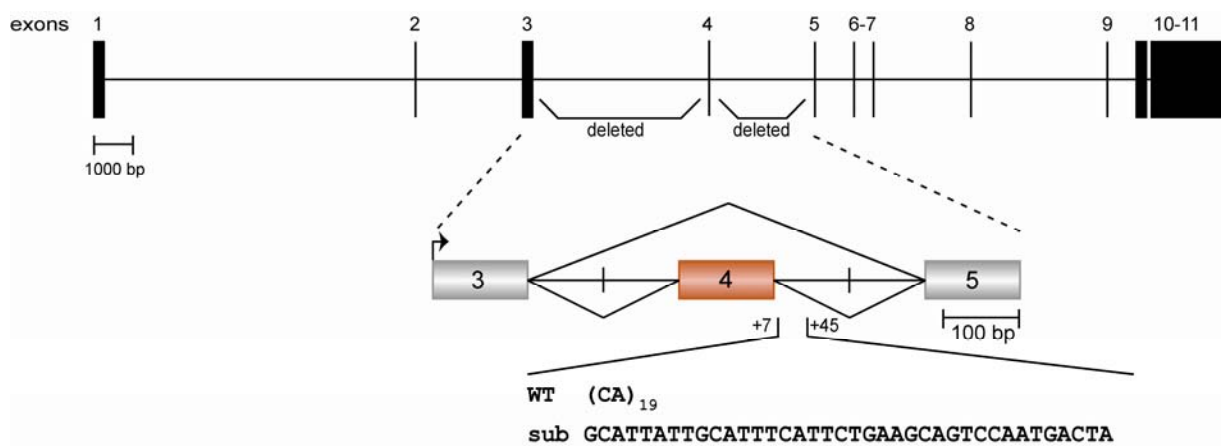


Figure 3.1

Schematic representation of the *SLC2A2* gene structure and minigene constructs. Boxes display the exons, lines the introns (sizes in scale). The *SLC2A2* minigene comprises exons 3 to 5 with introns shortened (marked by the vertical lines within the introns). The colour highlights the alternatively used exon 4. The lines below the minigene represent the normal splicing pattern; the lines above show the alternative splicing pattern. Wildtype (WT) and substituted (sub) sequence elements are given with boundaries relative to the 5' splice site. The minigene constructs were placed under the control of T7 and CMV promoter (depicted as an arrow) for *in vivo* and *in vitro* splicing analysis.

Interestingly, Matsutani and co-workers characterised a polymorphism concerning this CA dinucleotide repeat sequence (Matsutani *et al.*, 1992). Comparing nine alleles of the gene in three racial groups they identified a polymorphism of the CA repeat sequence ranging from 14 to 24 repeats. However, a linkage between the *SLC2A2* polymorphic repeat region and development of non-insulin-dependent diabetes mellitus could not be found (Janssen *et al.*, 1994).

3.1.1 Intronic CA-repeat sequence represents a splicing silencer

The *SLC2A2* gene was selected, based on a genome-wide database search for alternative splicing targets of hnRNP L (Hui *et al.*, 2005). As mentioned above, the gene contains a CA repeat sequence in intron 4 close to the alternatively spliced exon 4 (Fig. 3.1). To investigate the influence of this CA repeat on splicing of exon 4, a mutational analysis was performed using minigene constructs. As shown in Fig. 3.1, these minigene constructs consist of exons 3 to 5 with the alternatively spliced exon 4 in between. Due to their length, introns 3 and 4 had to be shortened. In a substitution derivative, the CA repeat element was replaced by a non-specific control sequence.

Splicing of minigene constructs was first tested *in vivo* (Fig. 3.2A). Two days after transient transfection of HeLa cells, total RNA was isolated and splicing was analysed by RT-PCR using gene-specific primers. Splicing of the wildtype *SLC2A2* construct resulted primarily in mRNA containing all three exons (lane 1). Only around 10% of exon skipping could be detected. In case of the substituted minigene construct, however, almost no exon skipping was determined (lane 2). As a control, a mock transfection was done in the absence of DNA (lane 3) showing that there was no detectable background from endogenous *SLC2A2* mRNA. Taken together, these data represented first evidence that the intronic (CA)₁₉ repeat sequence is important for alternative splicing of *SLC2A2* exon 4 acting as an intronic splicing silencer *in vivo*. To support this finding splicing of the same minigene constructs was analysed in an *in vitro* splicing assay (Fig. 3.2B). *In vitro* transcribed pre-mRNAs were incubated in HeLa cell nuclear extract under splicing conditions and the splicing activity of each construct was analysed by semi-quantitative RT-PCR. After 240 min, around 42% exon inclusion could be detected for the wildtype minigene (lane 3). For comparison, splicing of the

substitution derivative resulted in 89% exon 4 inclusion (lane 6). These results confirmed consistent with the data obtained from the *in vivo* splicing experiment that the CA repeat sequence acts on alternative splicing of *SLC2A2* exon 4 as a splicing silencer. As shown in Fig. 3.2B, an additional splicing product appeared for both constructs representing a case of intron retention. This alternative splicing event, however, was not investigated further.

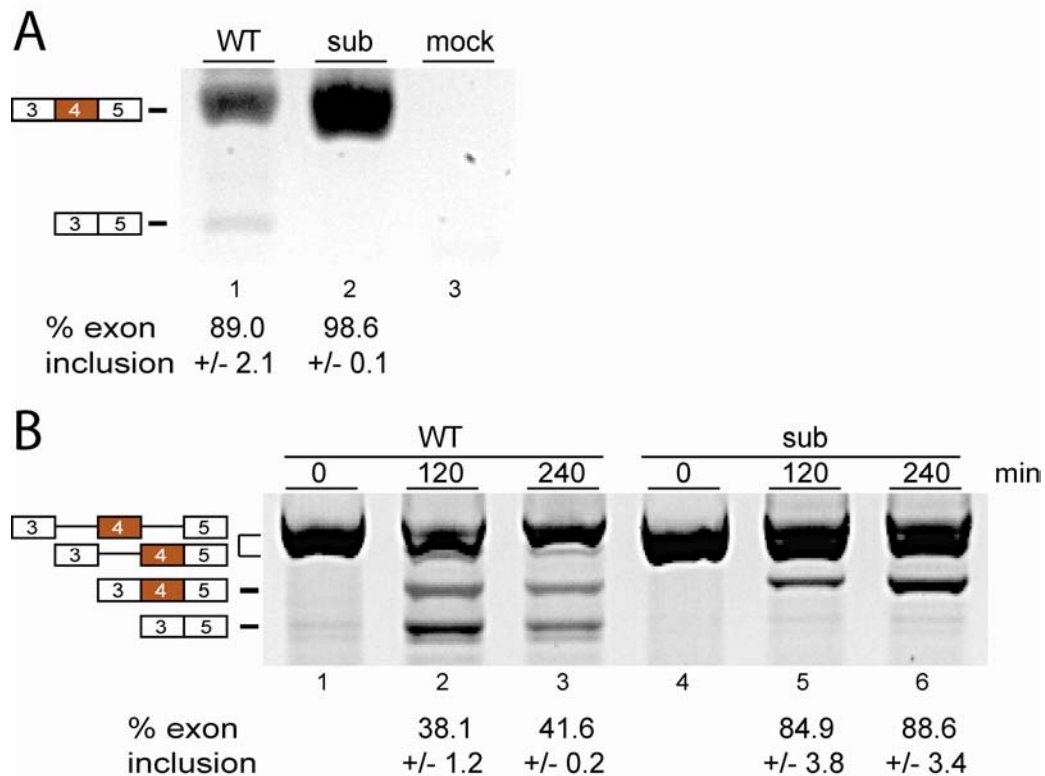


Figure 3.2

Characterisation of *SLC2A2* alternative splicing using minigene constructs.

(A) HeLa cells were transiently transfected and alternative splicing of the minigene constructs was tested by RT-PCR (lanes *WT*, *sub*); the control transfection (lane *mock*) was done in the absence of DNA. The percentages of exon inclusion with standard deviations ($n=3$) are given below the corresponding lanes. (B) *In vitro* transcribed pre-mRNAs were spliced in HeLa nuclear extract (lanes *WT*, *sub*). Aliquots were taken at the indicated times and alternative splicing was tested by RT-PCR. Positions of PCR products corresponding to pre-mRNA and splice variants are schematically shown on the left. The percentages of exon inclusion with standard deviations ($n=3$) are given below the corresponding lanes.

In summary, results from *in vivo* and *in vitro* splicing assays confirm that an intronic CA repeat sequence close to the 5' splice site of *SLC2A2* exon 4 negatively influences the inclusion of the exon into the mRNA thus acting as a splicing silencer.

3.1.2 HnRNP L mediates skipping of *SLC2A2* exon 4 *in vitro*

To investigate the importance of hnRNP L on alternative splicing of *SLC2A2*, I performed depletion/complementation experiments. HnRNP L was depleted from HeLa nuclear extract with a 5'-biotinylated (CA)₃₂ RNA oligonucleotide. It was shown before that hnRNP L binding to (CA)₃₂ is specific and occurs with high affinity (Hui *et al.*, 2003b). Via binding to streptavidin agarose, hnRNP L can be removed from nuclear extract. As control, a mock depletion was done in the absence of RNA oligonucleotide.

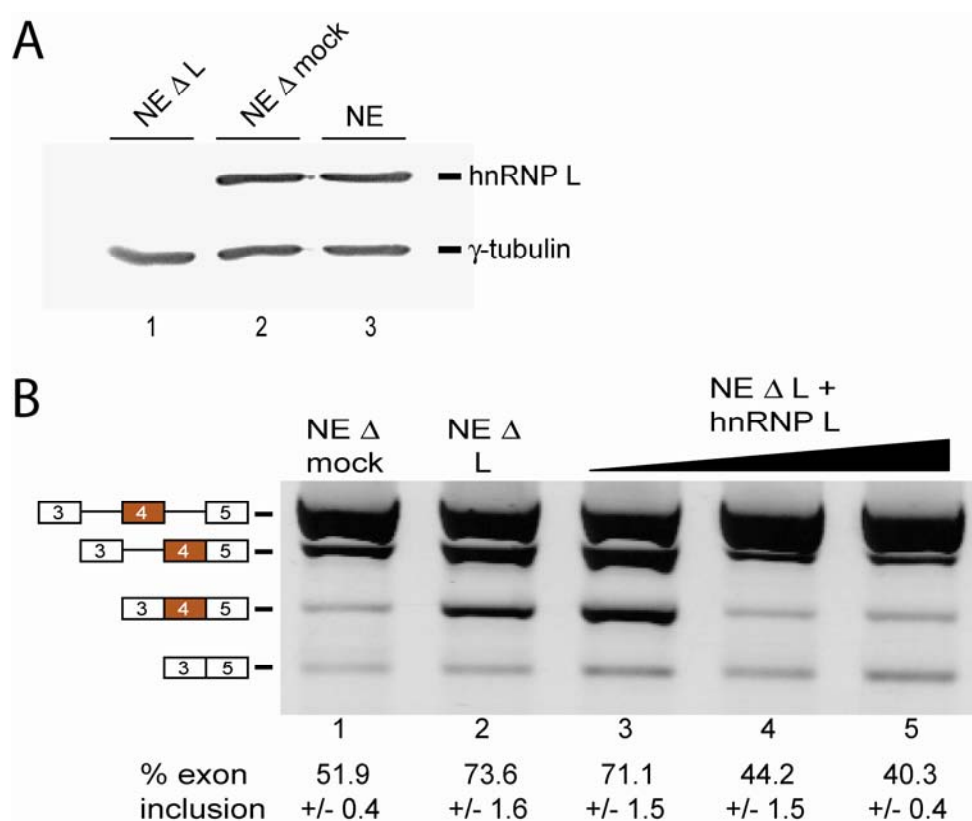


Figure 3.3

HnRNP L mediates alternative splicing of *SLC2A2* exon 4. (A) A biotinylated (CA)₃₂ RNA oligonucleotide was used to deplete hnRNP L from HeLa nuclear extract. As control, mock-depletion was done in the absence of RNA oligonucleotide. Untreated (lane NE), mock-depleted (lane NE Δ mock), and hnRNP L-depleted (lane NE Δ L) nuclear extract were analysed by Western blot using anti-hnRNP L peptide antibody D-17. Additionally, γ-tubulin detection was used as internal standard. (B) In vitro splicing of *SLC2A2* pre-mRNA in hnRNP L-depleted (lane NE Δ L) or mock-depleted (lane NE Δ mock) HeLa nuclear extract. The hnRNP L-depleted nuclear extract was complemented by the addition of increasing amounts of recombinant hnRNP L protein (100, 200, and 300 ng per 25-μl splicing reaction) to the splicing reaction (lanes NE Δ L + hnRNP L).

Western blot analysis using an hnRNP L-specific antibody confirmed efficient depletion from HeLa nuclear extract (Fig. 3.3A). With that antibody, no remaining

hnRNP L was detected in the depleted extract (lane 1). Moreover, the level of hnRNP L in the mock-depleted extract remained unchanged when compared to untreated nuclear extract (compare lanes 2, 3). Detection of γ -tubulin was used as an internal loading control.

To show that repression of *SLC2A2* exon 4 inclusion is mediated by hnRNP L, wildtype pre-mRNA was spliced in hnRNP L-depleted or mock-depleted HeLa cell nuclear extract and alternative splicing was analysed by RT-PCR using gene-specific primers (Fig. 3.3B). In mock-depleted nuclear extract, I detected 52% exon inclusion (lane 1), whereas over 70% exon inclusion was observed in hnRNP L-depleted extract (lane 2). Moreover, addition of recombinant hnRNP L to the depleted extract restored the inhibitory effect on exon 4 inclusion in a dose-dependent manner (lanes 3-5). Addition of 200 or 300 ng per 25- μ l splicing reaction of recombinant hnRNP L resulted in an even higher rate of exon skipping than observed in mock-depleted extract (compare lane 1 with 4, 5). However, with the addition of 100 ng per 25- μ l splicing reaction I detected only a slight decrease in exon inclusion. Note that per 25- μ l splicing reaction 200 ng of hnRNP L correspond to the normal level in HeLa nuclear extract (Hung *et al.*, 2008).

Taken together these data clearly show that hnRNP L works as a repressor on *SLC2A2* exon 4 inclusion. Depletion of hnRNP L led to a significant increase in exon inclusion levels whereas adding back recombinant protein reversed the effect. Moreover, exon repression depended on the presence of the CA repeat sequence since no change in exon inclusion levels was detected upon splicing of the *SLC2A2* substitution minigene construct in mock- or hnRNP L-depleted nuclear extract (data not shown).

3.1.3 Intronic splicing silencer affects spliceosome assembly

I obtained first evidence about the mechanism of splicing repression by analysing spliceosome assembly or, more precisely, formation of the spliceosomal E complex (Fig. 3.4). Wildtype and substitution *SLC2A2* substrate RNAs comprise part of exon 4 (20 nt) and the 5' splice site with the following (CA)₁₉ repeat or substituted sequence in intron 4, respectively (panel A). The RNA substrates were labelled internally with [α^{32} P]ATP by T7 *in vitro* transcription and incubated in HeLa nuclear extract. Since E complex formation is ATP-independent, ATP was

3. Results

depleted from the extract by incubation at room temperature for 30 min (Das & Reed, 1999). Electrophoresis on polyacrylamide gels represents the commonly used method for resolution of spliceosomal complexes (Das & Reed, 1999). Under these conditions however, the E complex cannot be resolved since it comigrates with the non-spliceosomal H complex. Therefore, I used native agarose mini-gel electrophoresis to resolve specifically spliceosomal E complex (Fig. 3.4).

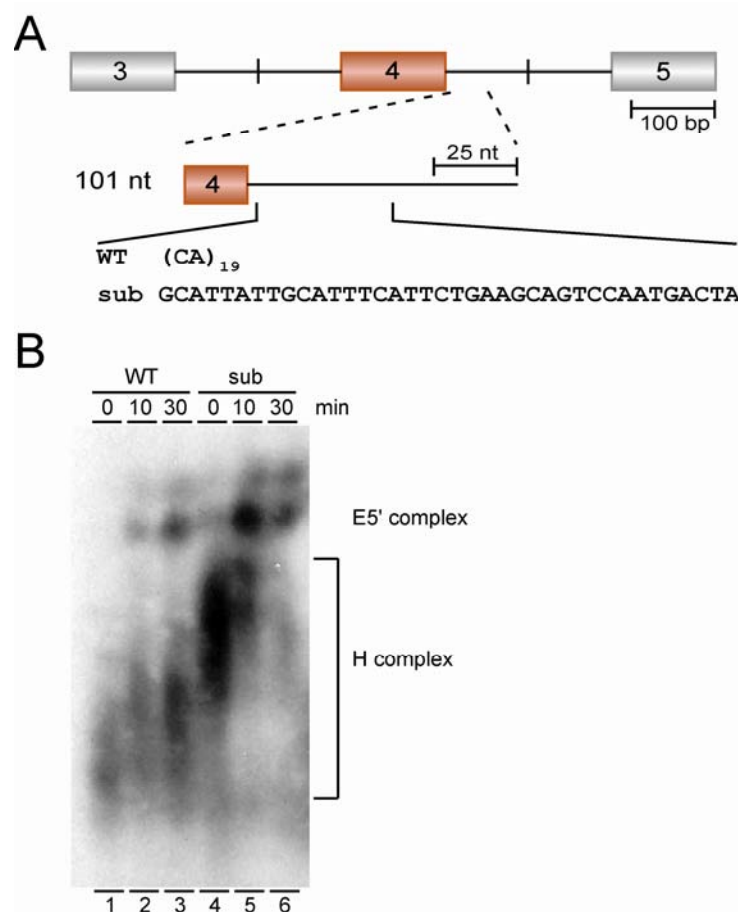


Figure 3.4

Analysis of spliceosome assembly. (A) Schematic representation of the *SLC2A2* minigene constructs and RNA transcripts (101 nt). Intron and exon parts are shown in scale. The colour highlights the alternatively used exon 4. Wildtype and substituted sequence elements are given below. (B) Resolution of the spliceosomal E5' complex. Short *SLC2A2* RNA probes were labelled internally with [α^{32} P]ATP by T7 *in vitro* transcription. Wildtype and substituted RNAs were incubated in HeLa cell nuclear extract under E complex conditions (without ATP) for the indicated times. Aliquots were separated on a 1.5% agarose gel. Complexes formed are indicated on the right.

The first complex to be assembled on the pre-mRNA is the ATP-independent H complex (Fig.3.4B). This complex lacks the spliceosomal snRNPs and assembles on every RNA independently of splicing (Bennett *et al.*, 1992). The E complex represents the first distinct spliceosomal complex. The E complex was termed E5'

in this experiment because the *SLC2A2* substrates did not contain a full-length intron but only its 5' part (Michaud & Reed, 1993). Therefore, it represents only a "half-substrate" lacking a branchpoint and 3' splice site. Note that the E5' complex cannot proceed further in spliceosome assembly due to the fact that the next step, formation of the A complex, requires the branchpoint sequence. A slow-mobility complex, which was not present at time point zero and accumulated with time, was identified as the spliceosomal E complex due to its sensitivity to the addition of heparin (data not shown). A second, fast-mobility complex corresponded to the H complex and was detected also at time point zero.

As shown in Fig. 3.4B, formation of the E5' complex could be observed for both transcripts, wildtype and substitution. But complex formation with the substitution derivative was more efficient than with the *SLC2A2* wildtype construct (compare lanes 2, 3 to 4, 5). I could detect a strong signal for the E5' complex with the substitution construct already after 10 min of incubation (lane 5).

In summary, substitution of the CA repeat sequence led to an enhancement of spliceosomal E complex formation which is likely to be mediated by hnRNP L.

3.1.4 HnRNP L binding to intronic splicing silencer interferes with 5' splice site recognition by the U1 snRNP

Next, I addressed the question for the mechanism of splicing repression by hnRNP L. The CA repeat sequence in *SLC2A2* intron 4 is located in the immediate vicinity of the 5' splice site of the regulated exon. Therefore, binding of hnRNP L may interfere with recognition of the splice site by the U1 snRNP.

To test this hypothesis, psoralen crosslinking experiments were performed. Psoralen is a photochemical reagent, which intercalates in nucleic acid helices. Upon irradiation with long wavelength UV-light (~360 nm) it forms covalent bonds to pyrimidine bases. Psoralen can therefore be used to crosslink double-stranded nucleic acids (Wassarman, 1993). In this case, I used psoralen crosslinking to investigate RNA-RNA interactions between U1 snRNA and pre-mRNA.

Again, short ³²P-labelled RNA transcripts were used (see Fig. 3.4A). *SLC2A2* wildtype and substitution RNA transcripts were incubated in HeLa nuclear extract under splicing conditions. After addition of psoralen, RNAs were crosslinked at 350 nm for 30 min, purified and separated on an 8% denaturing polyacrylamide gel.

3. Results

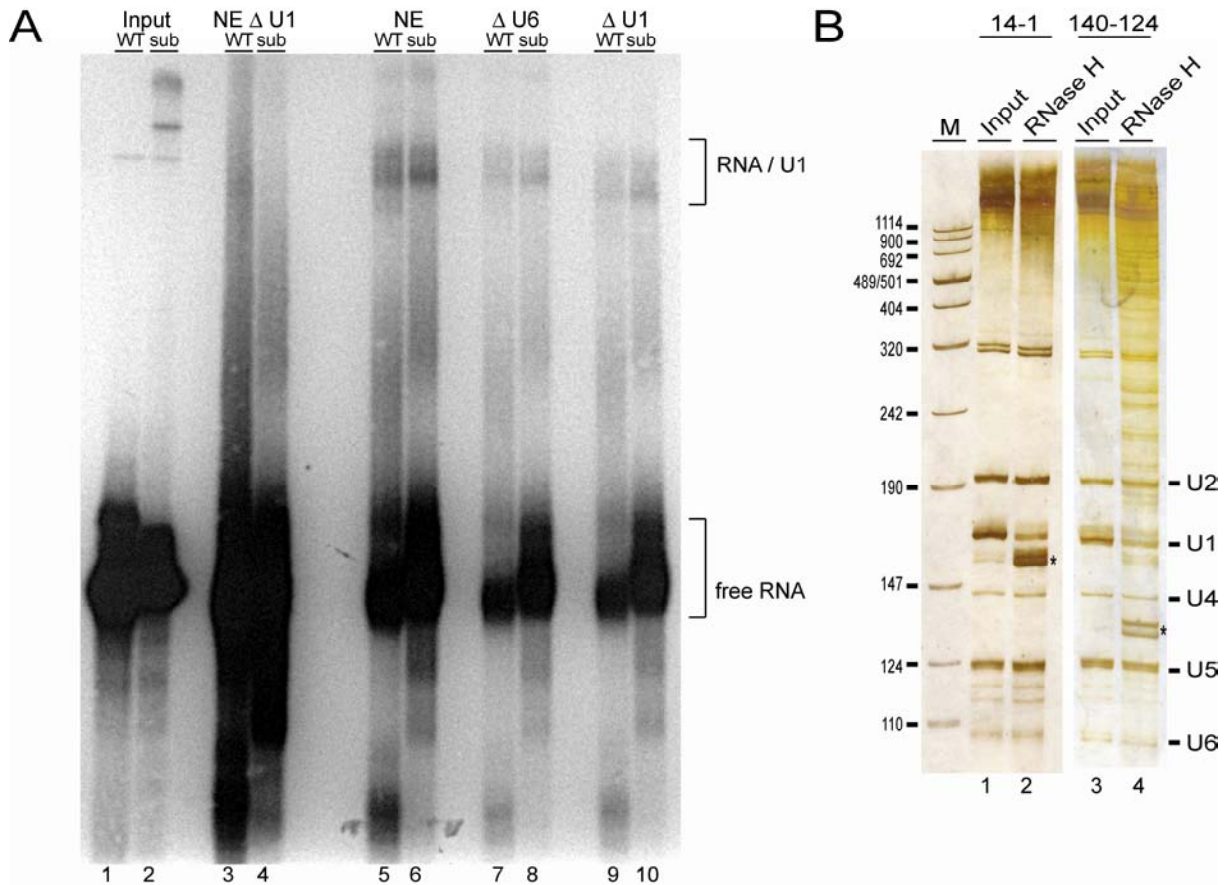


Figure 3.5

Identification of crosslinked bands by RNase H treatment. (A) Psoralen crosslinking. *SLC2A2* wildtype and substitution RNA substrates (lanes *Input*) were labelled with [α^{32} P]ATP by T7 *in vitro* transcription and incubated under splicing conditions in HeLa nuclear extract (lanes *NE*) or in U1-inactivated extract (lanes *NE Δ U1*). RNAs were purified and analysed on an 8% denaturing polyacrylamide gel. Crosslinked species and free RNA are indicated on the right. For identification of the crosslinked bands, purified RNAs were treated with RNase H and an oligonucleotide complementary to U1 snRNA (lanes $\Delta U1$). As control, a U6-specific oligonucleotide was used (lanes $\Delta U6$). (B) RNase H mediated knockout of U1 snRNA. HeLa nuclear extract (lanes 1, 2) or total RNA (lanes 3, 4) was treated with RNase H and an oligonucleotide complementary to U1 snRNA. The oligonucleotide was directed against the 5' end (14-1) or the Sm site (140-124) of the U1 snRNA, respectively. RNA samples were separated on an 8% polyacrylamide gel, each representing 10% of total, and visualised by silver staining. Positions of snRNAs are indicated on the right. Cleavage products are indicated by an asterisk. The DNA molecular weight marker VIII (M) is given in bp.

As shown in Fig. 3.5A, psoralen crosslinking of *SLC2A2* RNA transcripts in HeLa nuclear extract resulted in two distinct bands (lanes 5, 6), the lower prominent band representing free RNA. Two independent RNase H cleavage assays were used to identify the U1/RNA crosslink.

RNase H is an endonuclease that specifically cleaves the RNA component in a DNA/RNA duplex. First, U1 inactivation by RNase H cleavage was established in

HeLa nuclear extract (Fig. 3.5B, lanes 1, 2). No crosslinks were observed in subsequent psoralen crosslinking experiments when the U1-inactivated nuclear extract was used (Fig. 3.5A, lanes 3, 4). All crosslinked bands, detected with untreated nuclear extract, were therefore considered as U1-specific. Second, confirmation on the identity of crosslinks was also obtained by RNase H treatment of total RNA purified after psoralen crosslinking. U1-containing crosslinks were expected to shift in size due to cleavage of the U1 snRNA by RNase H. With a U1-specific oligonucleotide the expected shift was observed but not with a U6-specific oligonucleotide, which was used as a control for specificity (panel A, compare lanes 7, 8 to 9, 10).

Due to experimental requirements of the RNase H cleavage assays, it was necessary to use different U1-specific oligonucleotides. The oligonucleotide used for U1 inactivation in HeLa nuclear extract was directed against the 5' end of the U1snRNA (14-1) which is not bound by proteins and therefore freely accessible. But during spliceosome assembly the 5' part of the U1 snRNA base-pairs with the 5' splice site of the pre-mRNA rendering it inaccessible to the 5' end oligonucleotide (Krämer *et al.*, 1984). Therefore, another U1-specific oligonucleotide had to be used for cleavage of U1-containing crosslinks after psoralen crosslinking. This oligonucleotide was directed against the Sm site of the U1 snRNA. In nuclear extract, the U1 snRNA exists as part of the snRNP particle where the Sm site is bound by a set of seven Sm proteins, called B/B', D1, D2, D3, E, F, and G, which form a heptameric ring structure (Raker *et al.*, 1999). For this reason, the Sm site DNA oligonucleotide (140-124) could only be used in protein-free reactions on purified RNA. Due to the different target sites of the U1-specific oligonucleotides in the U1 snRNA, distinct U1 cleavage products can be observed for either oligonucleotide after RNase H treatment but both oligonucleotides led to a specific cleavage of the U1 snRNA whereas the other snRNAs were not affected (Fig. 3.5B).

Comparison between the *SLC2A2* wildtype and substitution derivative revealed different crosslinking efficiencies (Fig. 3.5A). The U1 crosslink with the wildtype RNA transcript was less efficient than with the substitution derivative (Fig. 3.5A, compare lanes 5, 6), supporting the idea that hnRNP L binding to the intronic silencer sequence interferes with 5' splice site recognition by the U1 snRNP.

To investigate the U1 snRNP interaction with *SLC2A2* RNA transcripts further, I performed a time course experiment. Again, ³²P-labelled *SLC2A2* wildtype and

3. Results

substitution RNA substrates were incubated in HeLa nuclear extract. After 5 and 30 min samples were subjected to psoralen crosslinking and the purified RNAs were separated on an 8% denaturing polyacrylamide gel (Fig. 3.6). Crosslinked bands were observed for both substrates already after 5 minutes. Again, crosslinking of the *SLC2A2* wildtype RNA transcript was less efficient (compare lanes 3, 5) and disappeared almost completely after 30 min of incubation (lane 4). A degradation band became visible instead. The U1 crosslink with the substituted transcript was also reduced but to a lesser extent (compare lanes 4, 6). A degradation band became visible instead. The U1 crosslink with the substituted transcript was also reduced but to a lesser extent (compare lanes 4, 6).

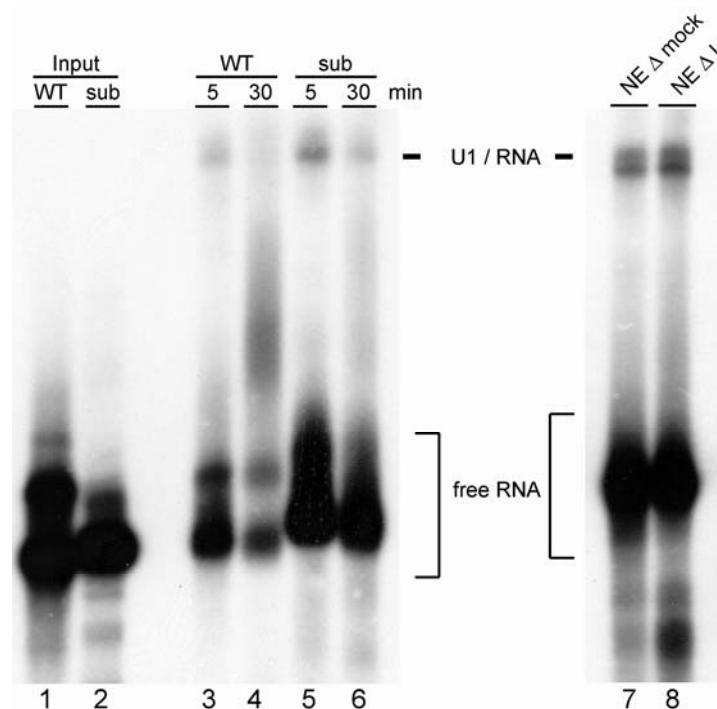


Figure 3.6
Depletion of hnRNP L promotes interaction of the U1 snRNP with *SLC2A2* RNA. Psoralen crosslinking. *SLC2A2* wildtype (WT) and substituted (sub) RNA substrates (lanes *Input*) were labelled with [α^{32} P]ATP by T7 *in vitro* transcription and incubated under splicing conditions in HeLa nuclear extract (lanes *WT*, *sub*) for the indicated times. Additionally, crosslinking of *SLC2A2* wildtype RNA transcript was performed in hnRNP L-depleted (lane *NE Δ L*) or mock-depleted (lane *NE Δ mock*) HeLa nuclear extract. Purified RNA samples were separated on an 8% denaturing polyacrylamide gel each representing 10% of total and visualised by autoradiography. Crosslinked species and free RNAs are indicated.

I performed psoralen crosslinking experiment with *SLC2A2* wildtype RNA substrate in hnRNP L- and mock-depleted HeLa nuclear extract to confirm that hnRNP L is directly involved in the observed effect (Fig. 3.6, right panel). Compared to mock-depleted nuclear extract the crosslinked band in hnRNP L-depleted extract was more prominent (compare lanes 7, 8) supporting the idea that the U1 interaction is impaired by hnRNP L. U1 crosslinking to *SLC2A2*

wildtype RNA was more efficient in the absence of hnRNP L. Taken together, the crosslinking experiments clearly show that the CA repeat sequence in *SLC2A2* intron 4 impairs the U1/RNA interaction and moreover, that this is mediated by hnRNP L. HnRNP L interfered with binding of the U1 snRNP to the 5' splice site of *SLC2A2* exon 4, resulting in skipping of the exon.

3.2 *TJP1*

3.2.1 HnRNP L knockdown in HeLa cells by RNA interference

In the last few years, RNA interference (RNAi) has become a powerful tool for sequence-specific silencing of genes by small interfering RNAs (siRNAs). The siRNAs are usually between 21 to 25 nucleotides in size, complementary to both strands of the silenced gene, and can be transfected into cells to knockdown the expression of any gene (McManus & Sharp, 2002).

I used siRNAs directed against hnRNP L and hnRNP LL, a closely related paralog, for knockdown protein expression in HeLa cells for the identification of target genes (Fig. 3.7). Knockdown of firefly luciferase served as a control. Each siRNA consisted of two 21-nucleotide single-stranded RNAs that formed a 19-base-pair duplex with two nucleotides overhang at the respective 3' ends. The siRNA duplexes were transfected into HeLa cells. Four days after transfection, total RNA was isolated and the knockdown efficiencies were analysed for mRNA and protein expression. First, I assayed the reduction of mRNA levels by real-time PCR using gene-specific primer pairs (Fig. 3.7A). For quantification, the relative mRNA expression levels of hnRNP L and LL were normalised to the housekeeping gene β -actin and diagrammed as fold change in gene expression relative to the luciferase control. As shown in Fig. 3.7A, the hnRNP L mRNA was down-regulated to 19% and hnRNP LL to 12%. Simultaneous silencing of both factors resulted in reduced mRNA levels of hnRNP L and LL to 31% and 23%, respectively. Interestingly, upon knockdown of hnRNP L, the mRNA level of hnRNP LL increased 1.5-fold, whereas in the hnRNP LL knockdown no significant increase of hnRNP L mRNA could be detected.

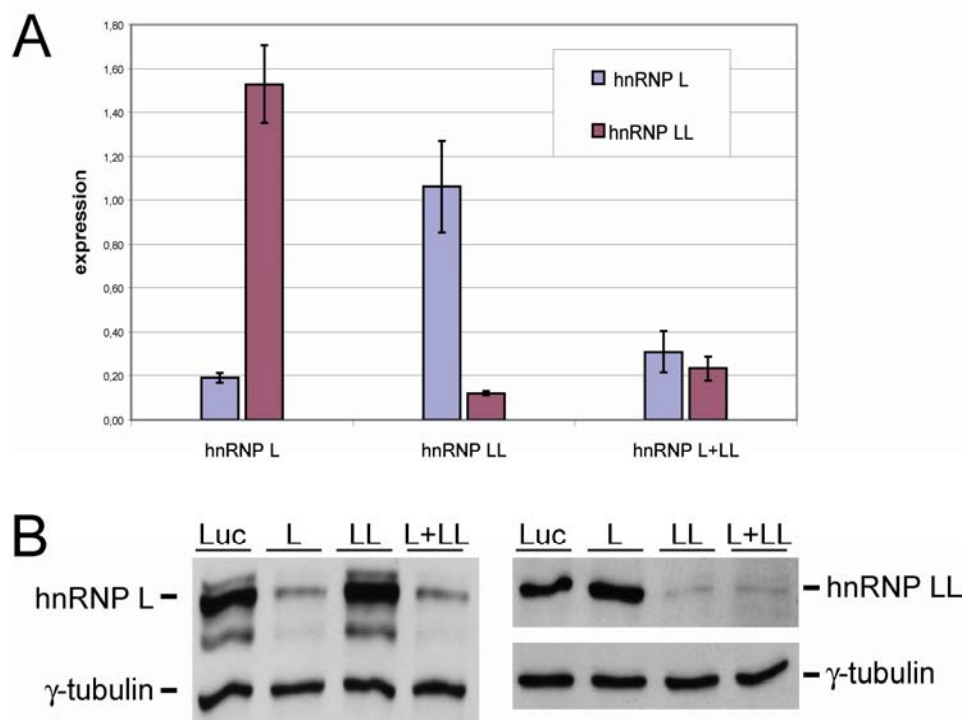


Figure 3.7
RNAi-mediated knockdown of hnRNP L and hnRNP L-like in HeLa cells. (A) Analysis of RNAi knockdown of hnRNP L and hnRNP L-like (LL) in HeLa cells by quantitative RT-PCR. HeLa cells were treated with siRNA oligonucleotides against hnRNP L, hnRNP LL, or both hnRNP L and LL. A siRNA oligonucleotide for luciferase serves as a control. Expression levels of hnRNP L and LL were normalised to β -actin and diagrammed as fold change in gene expression relative to the luciferase control. (B) Western blot analysis of hnRNP L (left panel) and hnRNP LL (right panel) knockdown using paralog-specific antibodies. Detection of γ -tubulin served as an internal standard (adapted from Hung *et al.*, 2008).

The RNAi-mediated knockdown of hnRNP L and LL was also confirmed by Western blot analysis. Fig. 3.7B shows the efficient and paralog-specific reduction of protein levels using γ -tubulin as an internal reference protein. HnRNP L was detected by monoclonal antibody 4D11, whereas for hnRNP LL detection a polyclonal antibody was used.

Taken together, these data clearly demonstrate that RNA interference is a suitable tool for silencing of hnRNP L as well as hnRNP LL expression in HeLa cells. I could show a significant and paralog-specific reduction of mRNA and protein levels by quantitative RT-PCR and Western blot analysis, respectively.

3.2.2 Genome-wide search for hnRNP L target genes by a combined microarray and RNAi analysis

After establishment of an efficient hnRNP L knockdown in HeLa cells, we used the

Affymetrix GeneChip Human Exon 1.0 Array to detected differences of exon expression signals in response to hnRNP L depletion.

So far, we could identify only a few alternative splicing target genes of hnRNP L by an initial database search, which was based on hnRNP L's binding specificity (Hui *et al.*, 2005). Considering the abundance of CA repetitive sequences in the human genome, it seemed most likely that there are more target genes to be found. The microarray technology enabled us to search the overall human genome for more hnRNP L target genes. Furthermore, we wanted to identify alternative splicing targets for the hnRNP L paralog hnRNP LL.

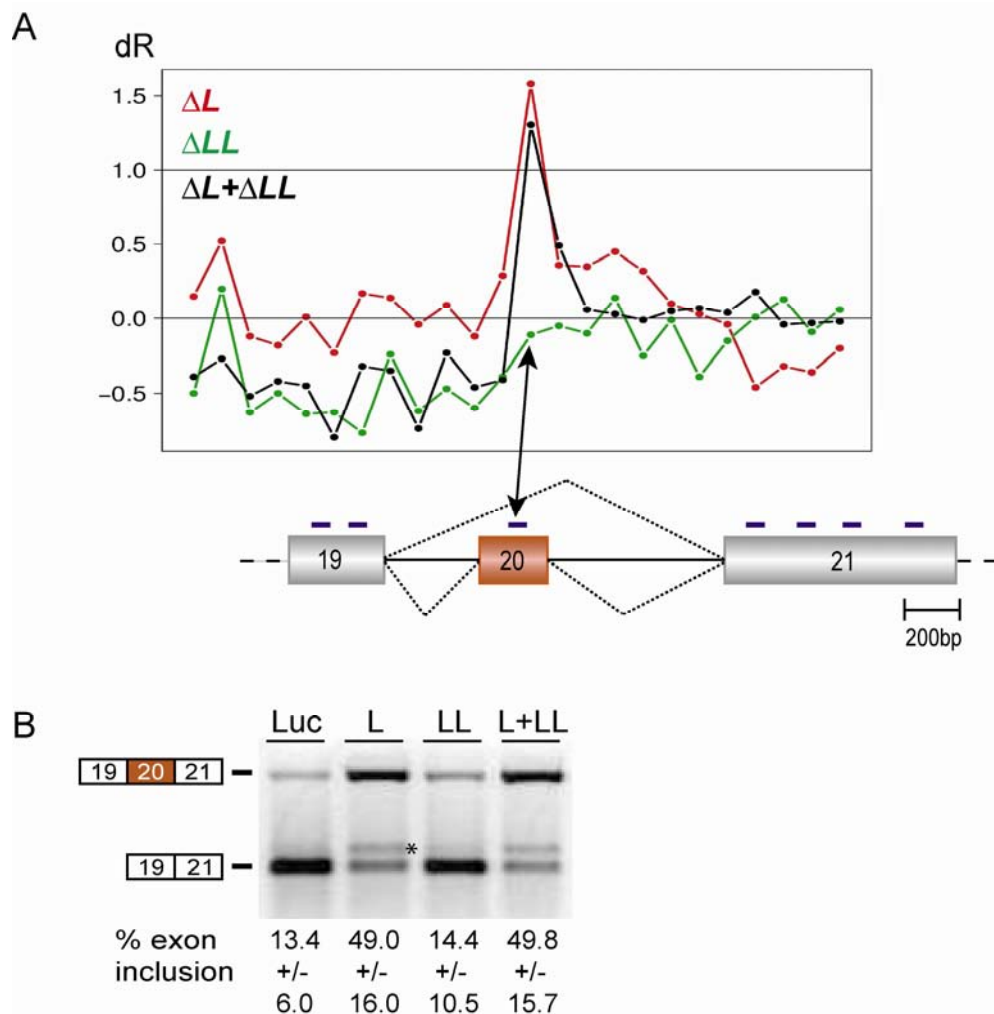


Figure 3.8

Analysis and validation of microarray data for the *TJP1* target gene. (A) Analysis of the microarray data; the diagram shows \log_2 ratios of probe-set signal intensities (dR) each relative to the luciferase control values across the *TJP1* gene. Three values are given (Y-axis: ΔL , knockdown of hnRNP L, in red; ΔLL , knockdown of hnRNP LL, in green; $\Delta L+\Delta LL$, knockdown of hnRNP L and LL, in black) for each probe set (X-axis). Probe-set positions in *TJP1* exons are shown below. (B) RT-PCR validation of *TJP1* alternative splicing. Total RNA was prepared after knockdown in HeLa cells and alternative splicing was analysed by RT-PCR. The asterisk marks an additional minor band representing an RT-PCR product due to mispriming within exon 20. The percentages of exon inclusion are given with standard deviations ($n=3$) below the corresponding lanes.

3. Results

The design of the Affymetrix human exon array allows detection of alternative exon usage (see Fig. 1.8). In combination with RNAi-mediated knockdown of hnRNP L, we identified several new alternative splicing target genes, 11 of these could be validated experimentally (Hung *et al.*, 2008). *TJP1*, one of the newly identified hnRNP L target genes, was investigated in further detail.

Our microarray data analysis was based on the assumption that the ratio of mRNA concentration between knockdown and luciferase control samples corresponds to the ratio of microarray expression signals. Fig. 3.8 shows the ratio of expression signals as \log_2 values (dR) diagrammed across the *TJP1* gene. A dR value around zero indicates that there is either no significant change in mRNA concentration between knockdown and control samples or that the probe yields no information. In the *TJP1* gene, a single probe in exon 20 showed significantly higher dR values after hnRNP L and double knockdown than the rest of the probe sets indicating an alternative splicing event (panel A). By RT-PCR assays, it was confirmed experimentally that exon 20 inclusion increased upon hnRNP L and L/LL double knockdown (panel B). The percentage of exon inclusion increased from 13% in the luciferase control to almost 50% upon hnRNP L and L/LL double knockdown (compare lane *Luc* with lanes *L*, *L+LL*). In contrast, no change in the alternative splicing pattern of *TJP1* exon 20 was observed upon knockdown of hnRNP LL (compare lanes *Luc*, *LL*).

In summary, analysis and validation of the microarray data yielded several new target genes whose alternative splicing is regulated by hnRNP L. In the case of *TJP1*, hnRNP L could be identified as a repressor on exon 20 inclusion.

3.2.3 *TJP1*, an hnRNP L target gene

The tight junction protein 1 (TJP1), also referred to as ZO-1 (zona occludens), is located on the surface of cytoplasmic membranes in tight junctions. TJP1, a large protein of 225 kDa, is most likely involved in signal transduction, which is required for tight junction assembly. Spanning approximately 122 kb, the *TJP1* gene consists of 28 exons (Fig. 3.9).

Willott and co-workers reported an alternative splicing event resulting in the expression of two distinct isoforms (Willott *et al.*, 1992). The two isoforms of *TJP1* differ by 80 amino acids, the so-called “motif- α ”, and are generated by alternative

splicing of exon 20. They are likely to contribute to tight junction diversity in epithelia. Kurihara and co-workers demonstrated diverse distribution among renal junctions (Kurihara *et al.*, 1992). The motif- α containing isoform is present in most epithelial junctions, whereas the shorter isoform of the protein occurs in endothelial cells.

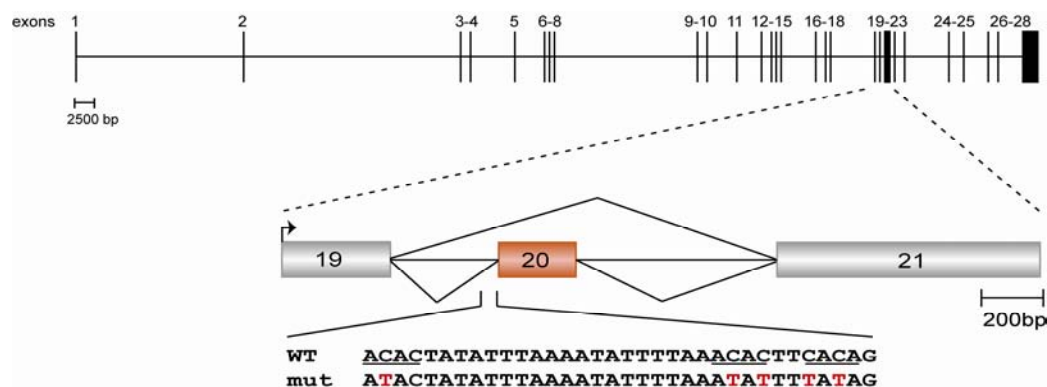


Figure 3.9

Schematic representation of the *TJP1* gene structure and minigene constructs. Boxes display the exons, lines the introns (sizes in scale). The *TJP1* minigene comprises exons 19 to 21. The colour highlights the alternatively used exon 20. The lines below the minigene represent the normal splicing pattern; the line above shows the alternative splicing pattern. Wildtype (WT) and mutated (mut; shown in red) sequence elements are given below. The underlined sequences represent hnRNP L high-score binding motifs as determined by SELEX. The minigene constructs were placed under the control of T7 and CMV promoter (depicted as an arrow) for *in vivo* and *in vitro* splicing analysis.

3.2.4 Identification of an intronic splicing silencer in the *TJP1* gene

As demonstrated by the microarray data, *TJP1* exon 20 inclusion is regulated by hnRNP L acting as a repressor (see Fig. 3.8). Examination of the sequence surrounding exon 20 revealed ten putative hnRNP L-binding motifs (Fig. 3.10A). We decided to further investigate a small cluster of high-score hnRNP L-binding motifs (motif no. 4) close to the 3' splice site, which provides an intronic splicing silencer element (Fig. 3.10B; for calculation of scores see Hung *et al.*, 2008). Mutational analysis of this putative silencer should give more insight into hnRNP L's mode of action. For this reason, a set of minigenes was made, each consisting of exons 19 to 21 with the full-length introns in between. Five point mutations (C→T) were intended to abolish hnRNP L binding in the mutant minigene (see Fig. 3.9).

3. Results

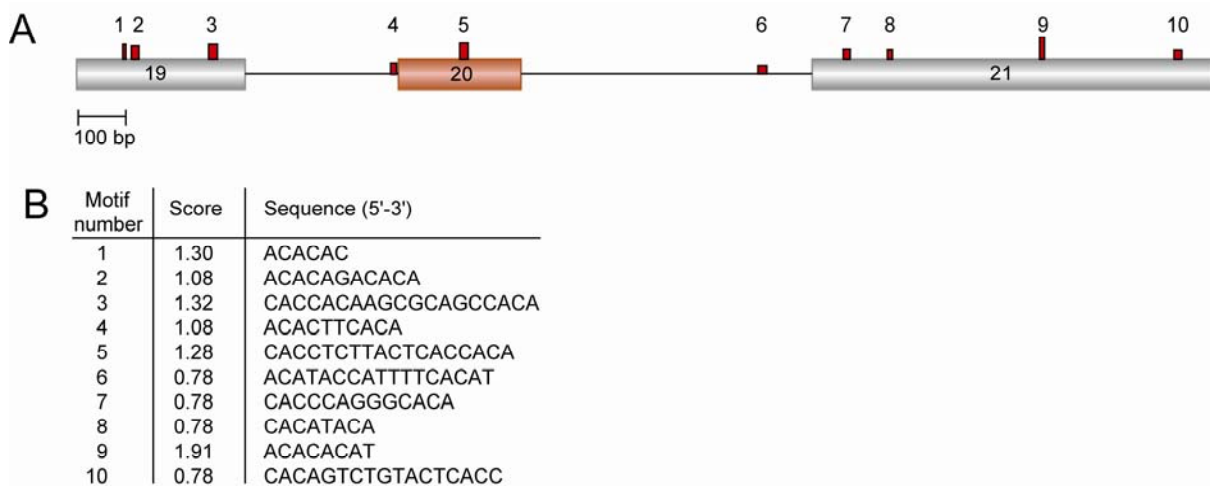


Figure 3.10

Sequence and location of potential hnRNP L-binding motifs in the *TJP1* target region. (A) Schematic representation of the *TJP1* target region. Exons are represented by boxes, introns by lines. The alternatively used exon 20 is highlighted. The red bars above show positions of hnRNP L-binding motifs. The width of the bar corresponds to the length of the motif, the height to the score (for calculation of scores see Hung *et al.*, 2008). All parts are shown in scale. (B) List of hnRNP L-binding motifs in the *TJP1* target region. Given are the score and the sequence in 5'-3' direction for each putative binding motif (adapted from Hung *et al.*, 2008).

I studied splicing of the minigene constructs *in vivo* by transient transfection of HeLa cells (Fig. 3.11A). Cells were harvested two days after transfection and total RNA was prepared using guanidium thiocyanate. Alternative splicing products were analysed by RT-PCR with primers specific for exon 19 and BGH rev (a vector-specific primer binding downstream of exon 21). Transfection of the wildtype *TJP1* minigene resulted predominantly in exon 19-21 mRNA with exon 20 skipped (lane 1), whereas the mutant derivative gave more than 50% exon 20 inclusion (lane 2). A mock control done in the absence of transfected DNA gave almost no products (lane 3), demonstrating that RT-PCR products were due to the transfected minigenes and that endogenous *TJP1* mRNA is not significant. Taken together, these data were first evidence that the CA-rich cluster in *TJP1* intron 19 represents a genuine intronic splicing silencer.

For further support of this idea, I also tested splicing of *TJP1* minigene constructs in *in vitro* assays. The three-exon minigenes, used for analysis of splicing *in vivo*, were too long to allow efficient *in vitro* transcription (data not shown). Therefore, I used shorter templates comprising only exons 19 and 20 with the full-length intron in between, which I obtained by PCR (Fig. 3.11B, right). *TJP1* pre-mRNAs were synthesised by T7 *in vitro* transcription and spliced in HeLa nuclear extract for one

and two hours (Fig. 3.11B, left). The splicing activity was determined by semi-quantitative RT-PCR with primers specific for the flanking exons 19 and 20. As shown in Fig. 3.11B, different splicing efficiencies of *TJP1* wildtype and mutant constructs could be detected. With over 30%, the splicing efficiency was significantly higher with the mutant derivative (lane 2) than with the wildtype (15.2%, lane 4). A control for DNA contamination (-RT control) gave no products (data not shown).

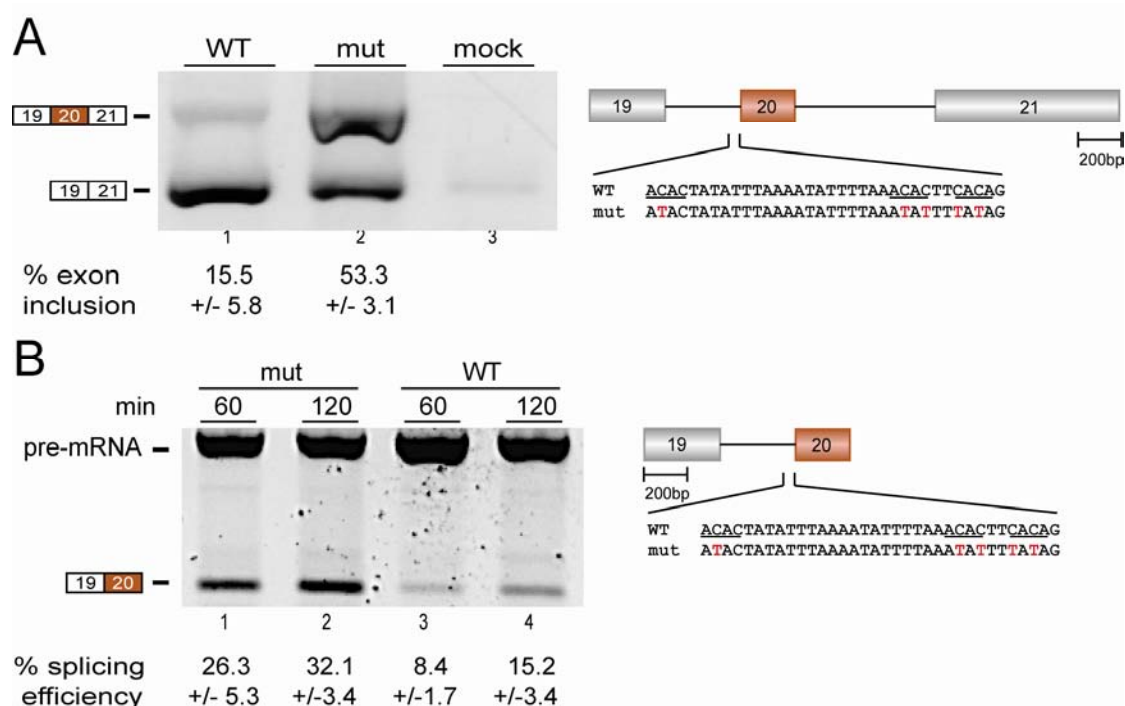


Figure 3.11

CA-rich sequence acts as an hnRNP L-dependent splicing silencer on *TJP1* exon 20 inclusion. (A) Characterisation of *TJP1* alternative splicing *in vivo* using minigenes illustrated on the right. Exons are shown by boxes, introns by lines. The alternatively spliced exon is highlighted. All exon and intron parts are shown in scale. Wildtype and mutated sequence elements are given. Minigene constructs were transfected into HeLa cells; the control transfection (lane *mock*) was carried out in the absence of DNA. Alternative splicing was analysed by RT-PCR (lanes *WT*, *mut*). Positions of the PCR products corresponding to the splice variants are indicated on the left. The percentages of exon inclusion with standard deviations ($n=4$) are given below the corresponding lanes. (B) *In vitro* transcribed pre-mRNAs were spliced in HeLa nuclear extract (lanes *WT*, *mut*). Aliquots were taken at the indicated time points and splicing tested by RT-PCR. Positions of PCR products corresponding to pre-mRNAs and spliced mRNAs are schematically shown on the left. Splicing efficiencies with standard deviations ($n=3$) are given below the corresponding lanes.

As determined with MaxEntScan (a web-based tool to predict the strength of human splice sites) the 3' splice site of exon 20 is very poor due to a weak

3. Results

polypyrimidine tract (Yeo & Burge, 2004). Several adenosines interrupt the polypyrimidine tract sequence leading to in a low score of this 3' splice site (WT 5.2, mut 5.8). Due to this fact, it was not surprising that the overall splicing efficiency of both *TJP1* minigene constructs was very low.

In summary, these data confirm the *in vivo* splicing result identifying the CA-rich cluster in *TJP1* intron 19 as a splicing regulatory sequence element. I could clearly show that this sequence element functions as an intronic splicing silencer on *TJP1* exon 20.

3.2.5 Alternative splicing of *TJP1* exon 20 is mediated by hnRNP L

To support the data, obtained by microarray analysis, that repression of *TJP1* exon 20 is mediated by hnRNP L, I performed depletion experiments. As illustrated before for *SLC2A2*, hnRNP L was depleted from HeLa nuclear extract using a biotinylated (CA)₃₂ RNA oligonucleotide (see Fig. 3.3A). Splicing of the *in vitro* transcribed *TJP1* two-exon wildtype pre-mRNA was assayed by RT-PCR (Fig. 3.12). Although splicing efficiency was very low, a significant difference between hnRNP L- and mock-depleted nuclear extract could be detected in splicing. The splicing efficiency in the hnRNP L-depleted extract was considerable higher than in mock-depleted nuclear extract (compare lanes 2, 4). Depletion of hnRNP L leads to a significant increase in *TJP1* exon 20 inclusion, supporting the idea that hnRNP L acts as a splicing repressor.

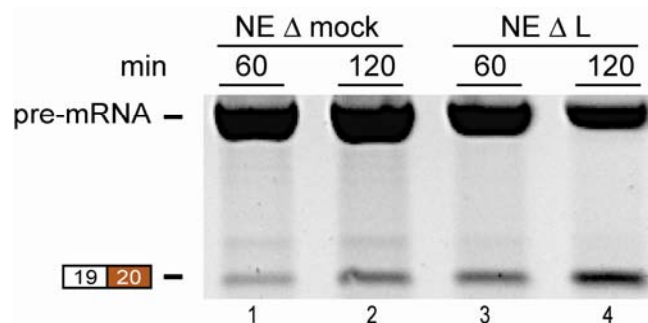


Figure 3.12

Repression of *TJP1* exon 4 inclusion is mediated by hnRNP L. *TJP1* wildtype pre-mRNAs were spliced *in vitro* in hnRNP L-depleted (lanes *NE Δ L*) or mock-depleted (lanes *NE Δ mock*) HeLa nuclear extract for the indicated times. Splicing efficiencies were determined by RT-PCR. Positions of PCR products corresponding to pre-mRNAs and spliced mRNAs are schematically shown on the left.

In summary, by mutational analysis, I identified a CA-rich cluster in *TJP1* intron 19 as splicing silencer sequence. Furthermore, I could show that splicing repression of *TJP1* exon 20 is mediated by hnRNP L, confirming the data obtained from our microarray analysis.

3.2.6 Mutational analysis of the intronic splicing silencer in *TJP1* gene

In order to obtain further insight into the sequence requirements of the newly-discovered intronic splicing silencer in *TJP1* gene, I made a new set of mutant minigene constructs (Fig. 3.13A). In the first new mutant derivative (mutTG), I introduced three point mutations. Taking the hnRNP L RNA binding specificity determined by *in vitro* SELEX (see Fig. 1.6) as a basis, one nucleotide was mutated in each of the three *TJP1* intron 19 high-score binding motifs. In the second mutant minigene (mutG), I changed only two nucleotides (A→G), one in each of the two motifs closest to the 3' splice site of exon 20. In contrast to mutantTG, the hnRNP L high-score binding motif ACAC, further upstream of the splice site, was kept intact.



Figure 3.13

Mutational analysis of the *TJP1* intronic splicing silencer. (A) Schematic representation of the *TJP1* minigene construct. Exons are shown as boxes, introns as lines. The alternatively used exon 20 is highlighted. Wildtype (WT) and mutated sequences are given below. Exon 20 sequence is marked by a box. HnRNP L high-score binding motifs are shown in red. C to T mutations are depicted in bold, A to G mutations in blue. (B) HeLa cells were transiently transfected with minigene constructs. A control transfection (lane *mock*) was done in the absence of DNA. Alternative splicing was assayed by RT-PCR. Positions of alternative spliced products are displayed on the left.

Splicing of the new minigene constructs was tested *in vivo*. As described before, minigene constructs were transiently transfected into HeLa cells and alternative splicing was assayed by RT-PCR (Fig. 3.13B). Splicing of both new mutant constructs resulted predominantly in the exon 19-21 mRNA (lanes 3, 4). Compared to the wildtype, no increase in exon 20 inclusion could be detected (compare lane 1 with 3, 4). The introduced mutations were apparently not sufficient to impair the silencing effect on *TJP1* exon 20 inclusion. Only with the initial mutant construct (mut), I could detect an enhanced level of exon 20 inclusion (lane 2), which was already shown before (see Fig. 3.11A).

A new inspection of the sequence of the new *TJP1* mutant minigenes revealed that both still contained potential hnRNP L binding motifs. ATAC, ACGC, and CGCA resulted from mutation of the three hnRNP L high-score binding motifs in the mutantTG derivative (see Fig. 3.13A). All of them were among the 20 most common tetranucleotide sequences selected by SELEX (see Fig. 1.6B) meaning that hnRNP L may bind to these sequences although with low affinity. The mutantG, however, still contains one of the three high-score binding motifs since only two motifs were mutated.

3.2.7 HnRNP L binds to *TJP1* CA-rich intronic silencer sequence with high affinity

Next, I wanted to investigate binding of hnRNP L to *TJP1* RNA by electrophoretic mobility shift assays (EMSA). The *TJP1* substrate RNAs used for this purpose consisted of the 5' end of exon 20 and part of intron 19 including branch point and CA-rich cluster (Fig. 3.14A).

The *TJP1* substrates were labelled internally with [α^{32} P]ATP via T7 *in vitro* transcription and incubated with increasing amounts of baculovirus-expressed recombinant hnRNP L. As a non-specific competitor for hnRNP L binding, tRNA was added to each reaction. Complexes were resolved on a native polyacrylamide gel and detected by autoradiography.

As shown in Fig. 3.14B, complex formation on wildtype substrate was very efficient (lanes 2-4). Addition of a two-fold molar excess of protein (200 nM) over RNA resulted already in formation of a complex (lane 2). With further addition of hnRNP L (400 nM), a second complex became visible (lane 3), most likely due to multiple

hnRNP L binding motifs within the wildtype substrate RNA. No complexes could be detected for the first mutant (mut) transcript (lanes 6-8). As expected from the *in vivo* splicing study, the other two mutant derivatives (mutTG and G) formed complexes with hnRNP L (lanes 10-12, 14-16). Apparently, both mutated silencer sequence in *TJP1* intron 19 were still able to bind hnRNP L, although with lower affinity than the wildtype (compare lanes *WT* with *mutTG*, *mutG*). However, mutTG and mutG displayed different affinities for hnRNP L binding (compare lanes *mutTG* with *mutG*). MutG still contained one of the three high-score binding motifs for hnRNP L (Fig. 3.14A) which correlated with the more efficient complex formation of this substrate compared to the mutTG derivative.

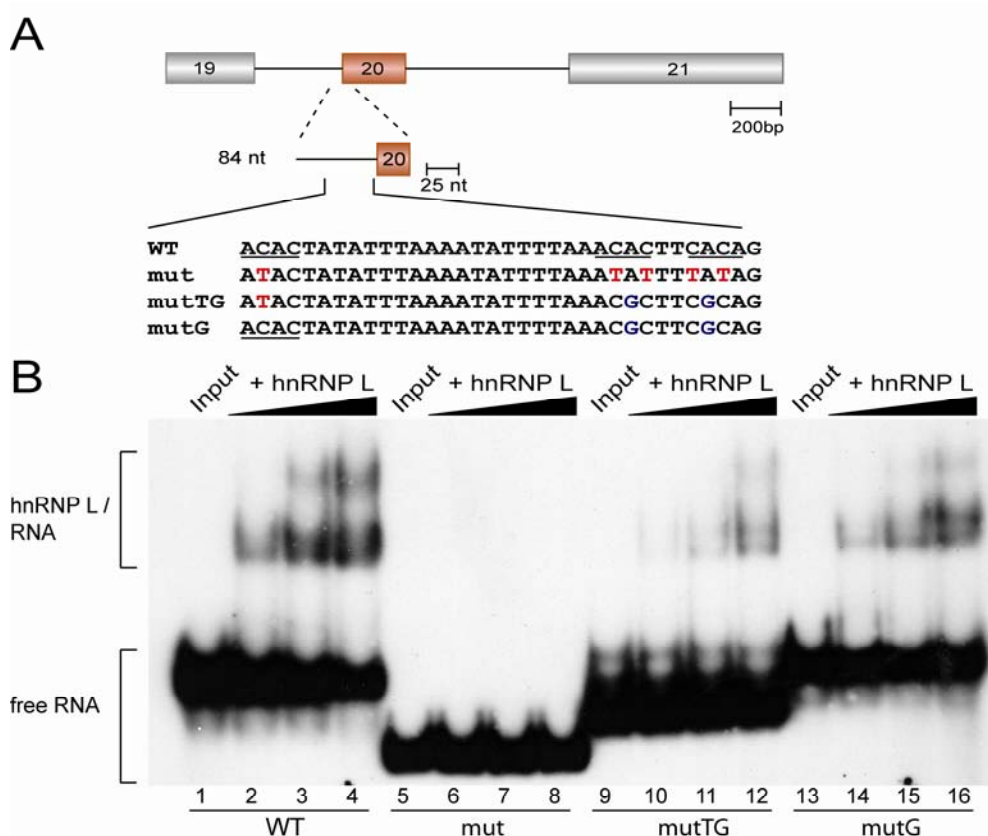


Figure 3.14

HnRNP L binding to *TJP1* RNA. (A) Schematic representation of the *TJP1* minigene constructs and RNA transcripts (84 nt). Intron and exon parts are shown in scale. The colour highlights the alternatively used exon 20. Wildtype and mutated (C→T shown in red, A→G in blue) sequence elements are given below. The underlined sequences represent hnRNP L high-score binding motifs. (B) Band shift assay, using RNA transcripts (shown in A) labelled with [α^{32} P]ATP. *TJP1* RNA substrates were incubated with increasing amounts of recombinant hnRNP L (200, 400, and 800 nM) and analysed on a native acrylamide gel. Positions of resulting RNA-protein complexes and free RNAs are indicated on the left.

Surprisingly, I observed differences in the migration behaviour of free wildtype and

3. Results

mutant transcripts although all transcripts had the same length. A different secondary structure of the transcripts is a likely explanation for this finding.

In summary, I could show that hnRNP L directly interacts with the *TJP1* wildtype, mutTG, and mutG RNA but not with the initial mutant (mut).

I further studied the interaction between hnRNP L and *TJP1* RNA by UV-induced crosslinking experiments. For that purpose, ³²P-labelled RNA transcripts were incubated with baculovirus-expressed recombinant hnRNP L at different temperatures (4 and 30°C) and afterwards subjected to UV crosslinking (Fig. 3.15).

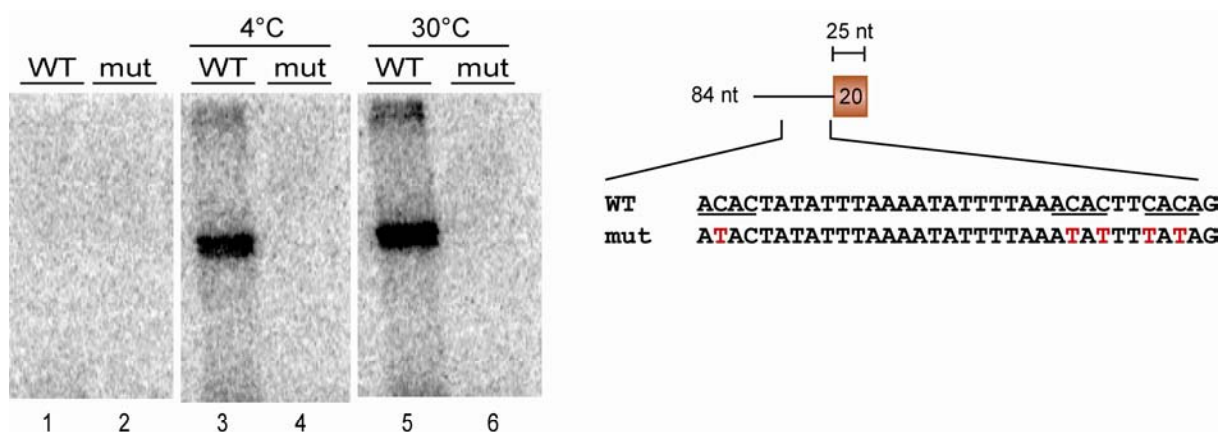


Figure 3.15

Analysis of the interaction between hnRNP L and *TJP1* RNA by UV crosslinking. Short ³²P-labelled *TJP1* wildtype (WT) and mutant (mut) RNA transcripts were incubated with 200 ng per 10 µl reaction of recombinant hnRNP L at either 4°C (lanes 4°C) or 30°C (lanes 30°C) and afterwards subjected to UV crosslinking. Crosslinked proteins were analysed on a 10% SDS-PAGE. As control, crosslinking was performed in the absence of protein (lanes 1 and 2). A schematic representation of the RNA transcripts is given at the right. The exon is represented by a box, the intron by a line. Wildtype and mutated sequence elements are given below. Exon and intron parts are shown in scale.

To degrade unprotected RNAs all reactions were digested with RNase A. Crosslinked proteins were separated on a 10% SDS-PAGE and visualised by autoradiography. Additionally, crosslinking was done in the absence of recombinant protein as control for specificity (lanes 1, 2). As shown in Fig. 3.15, hnRNP L could be crosslinked only to *TJP1* wildtype RNA (lanes 3, 5). No crosslink was observed with the mutant derivative (lanes 4, 6). Comparison of the two incubation temperatures revealed no significant differences in the crosslinking efficiencies.

Within this purified system, the binding specificity of hnRNP L is well defined. I could show, consistent with the band shift data, that hnRNP L specifically binds to CA-containing transcripts. The interaction of hnRNP L with the substrate RNA occurs also at 4°C indicating that it is splicing independent.

3.2.8 U2AF65 interaction with polypyrimidine tract of *TJP1* intron 19 is very weak

How is *TJP1* exon 20 skipping facilitated by hnRNP L? To address this question, I took a closer look at U2AF (U2 snRNP auxiliary factor). U2AF is a heterodimer (65 and 35 kDa subunits) which plays an important role in spliceosome assembly (Zamore & Green, 1989). The 65 kDa subunit (U2AF65) directly contacts the polypyrimidine tract whereas U2AF35 interacts with the 3' splice site of pre-mRNAs. Binding of U2AF facilitates the subsequent recruitment of the U2 snRNP to the branch site. The hnRNP L binding sites in *TJP1* intron 19 are located within the polypyrimidine tract. In the following, I therefore tested the hypothesis that hnRNP L binding impairs recognition of the polypyrimidine tract by U2AF65 thereby mediating exon skipping.

I first tested the interaction of U2AF65 with *TJP1* RNA by electromobility shift assays. As described before for hnRNP L, *TJP1* RNA transcripts were internally labelled with [$\alpha^{32}\text{P}$]ATP and incubated with increasing amounts of recombinant U2AF65 (Fig. 3.16). Purified GST-U2AF65 was analysed before on a protein gel (Fig. 3.16A).

As shown in Fig. 3.16B, complex formation on *TJP1* wildtype substrate was only detected at high U2AF65 concentrations (lanes 4, 6-8). Formation of a complex was first detected after addition of a 128-fold molar excess of protein (4 μM) over RNA (lane 4). Even with a 640-fold molar excess of U2AF65 (20 μM), free RNA was still visible (lane 8). MINX, a well characterised splicing substrate for *in vitro* applications (Zillmann *et al.*, 1988), was used as a positive control (Fig. 3.16C). In the case of MINX, formation of a complex was detected already with an 8-fold molar excess of U2AF65 (0.3 μM , lane 2). Addition of increasing amounts of U2AF65 shifted all free RNA into complexes (lanes 3, 4). As mentioned before, *TJP1* intron 19 contains a poor polypyrimidine tract whereas the one of MINX is very good. It is most likely that the strength of the polypyrimidine tract affects its

3. Results

ability to interact with U2AF65.

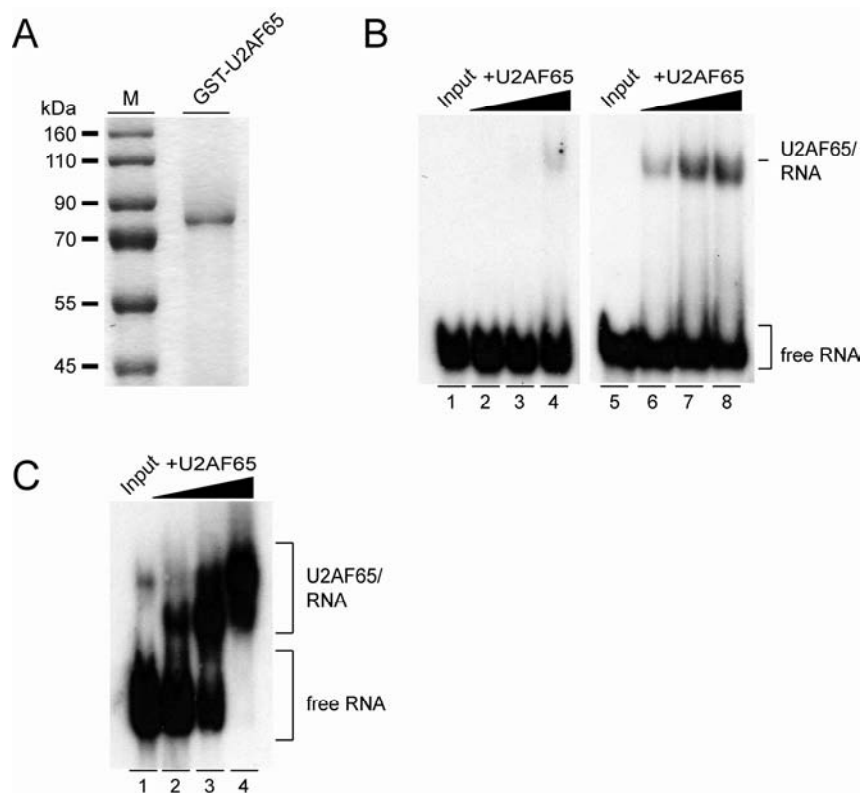


Figure 3.16

U2AF65 binds to *TJP1* RNA with low affinity. (A) Bacterially expressed GST-U2AF65 was purified, separated on a 10% SDS-PAGE and visualised by Coomassie staining. (B) Electromobility shift assays. ³²P-labelled *TJP1* wildtype RNA transcripts were incubated with increasing amounts of recombinant U2AF65 (left panel 0.3, 1, and 4 μM; right panel 5, 10, and 20 μM). Complexes were separated on native polyacrylamide gels and visualised by autoradiography. Positions of RNA-protein complexes and free RNAs are indicated on the right. (C) MINX pre-mRNA, consisting of two exons, one intron, was incubated with increasing amounts of recombinant U2AF65 (0.3, 1, and 4 μM). Complexes were separated on a native polyacrylamide gel and visualised by autoradiography. Positions of RNA-protein complexes and free RNAs are indicated on the right.

Taken together, these data showed that binding of U2AF65 to *TJP1* RNA occurred with very low efficiency. Nevertheless, the interaction between U2AF65 and *TJP1* RNA was investigated further in the following.

3.2.9 HnRNP L interferes with binding of U2AF65 to *TJP1* intron 19

I tested the hypothesis that hnRNP L interferes with binding of U2AF65 by UV crosslinking experiments using *TJP1* wildtype RNA transcripts which included the branch site, polypyrimidine tract, and 3' splice site of exon 20 as well as the full-length exon 20 with the 5' splice site (Fig. 3.17A).

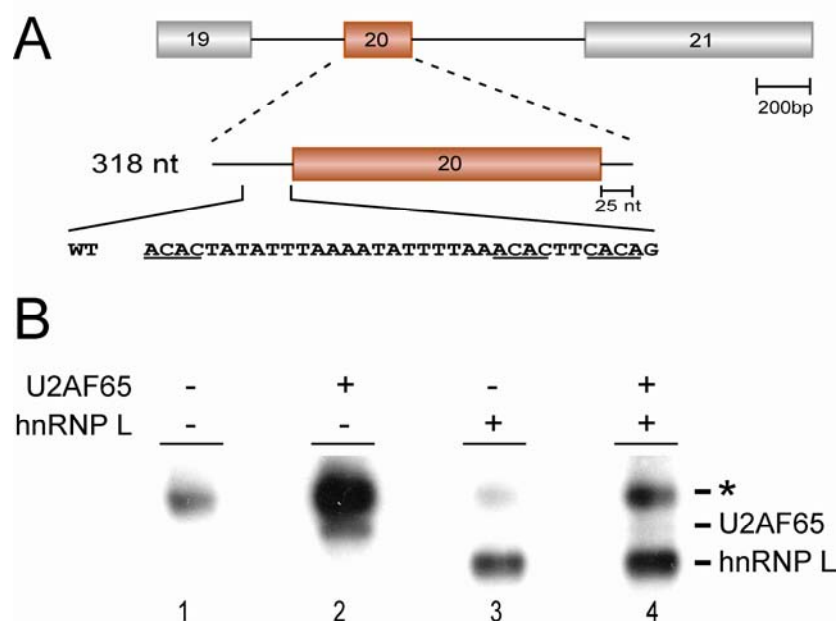


Figure 3.17

HnRNP L antagonises binding of U2AF65 to *TJP1* RNA. (A) Schematic representation of the *TJP1* wildtype minigene construct and RNA transcript (318 nt). Intron and exon parts are shown in scale. The colour highlights the alternatively used exon 20. The wildtype sequence element is given below. The underlined sequences represent hnRNP L high-score binding motifs. (B) UV crosslinking assay. 32 P-labelled *TJP1* wildtype (WT) RNA transcripts (shown in A) were incubated with either 600 ng of recombinant U2AF65 (lane 2), 200 ng of recombinant hnRNP L (lane 3) or both proteins at the same time (lane 4) in 13.5 μ l reactions. Reactions were subjected to UV crosslinking; unprotected RNAs were degraded by RNase A. The crosslinked proteins were analysed on a 10% SDS-PAGE. As control, crosslinking was performed in the absence of protein (lane 1). The positions of crosslinked proteins are indicated on the right. An unspecific band is marked by the asterisk.

TJP1 transcripts were internally labelled with [α^{32} P]ATP by T7 *in vitro* transcription and incubated with either GST-U2AF65 (600 ng per 13.5- μ l reaction), His-HnRNP L (200 ng per 13.5- μ l reaction), or both of them at 4°C for 30 min before UV crosslinking at 250 nm. Unprotected RNAs were degraded by subsequent RNase A treatment. Crosslinked proteins were separated on a 10% SDS-PAGE and visualised by autoradiography. As negative control, crosslinking was done in the absence of protein.

As shown in Fig. 3.17B, a band (asterisk) could be detected already without addition of recombinant protein (lane 1) most likely due to not properly degraded RNA. Besides the unspecific band, addition of recombinant U2AF65 resulted in a second band (lane 2) representing the crosslinked U2AF65 protein. Incubation with hnRNP L resulted in a specific crosslinked band as well with the expected size (lane 3). To test for competition, both proteins were incubated with the wildtype transcript simultaneously (lane 4). Only hnRNP L was observed to

3. Results

crosslink to the *TJP1* substrate RNA efficiently. The crosslink to U2AF65 on the other hand decreased significantly by comparison with incubation of U2AF65 alone (compare lanes 2, 4). These results show that in this minimal system, hnRNP L was capable to prevent binding of U2AF65 to the *TJP1* target RNA most likely by displacement from the polypyrimidine tract.

I obtained further support for this idea by UV-induced crosslinking and immunoprecipitation experiments. *TJP1* wildtype RNA transcripts were incubated under splicing conditions in hnRNP L- or mock-depleted HeLa nuclear extract. Fig. 3.18A shows the result of the subsequent UV crosslinking experiment. Several crosslinked proteins could be detected in nuclear extract by autoradiography. Only one protein around 65 kDa, however, was reduced in hnRNP L-depleted nuclear extract (compare lane 2 with lanes 1, 3). By immunoprecipitation using anti-hnRNP L monoclonal antibody the crosslinked protein could be identified as hnRNP L (Fig. 3.18B). In hnRNP L-depleted extract, immunoprecipitation of hnRNP L was strongly reduced compared to mock-depleted nuclear extract (compare lanes 1, 2). With addition of recombinant hnRNP L (200 ng per 25- μ l splicing reaction), protein crosslink and immunoprecipitation could be restored (lane 3). Levels of U2AF65 in the total crosslink were not visible (panel A). Since U2AF65 (65 kDa) and hnRNP L (64 kDa) are of similar size, they were expected to run approximately at the same position in a SDS-PAGE. Taking into account that hnRNP L bound *TJP1* RNA with high affinity (see Fig. 3.14) it was most likely that hnRNP L signals concealed signals of U2AF65. Therefore, I determined U2AF65 levels by UV crosslinking and immunoprecipitation with anti-U2AF65 monoclonal antibody (Fig. 3.18C). Detection of U2AF65 by immunoprecipitation revealed that hnRNP L reduction was associated with an increase in the level of crosslinked U2AF65 (compare lanes 1, 2). With addition of recombinant hnRNP L, this effect could be reversed (lane 3) showing that increased binding of U2AF65 is directly linked to hnRNP L levels.

In sum, these results clearly show that hnRNP L impairs binding of U2AF to *TJP1* RNA. Since the binding sites of both proteins reside in the same region of *TJP1* intron 19 hnRNP L competes with U2AF65 for binding, most likely displacing it from the RNA.

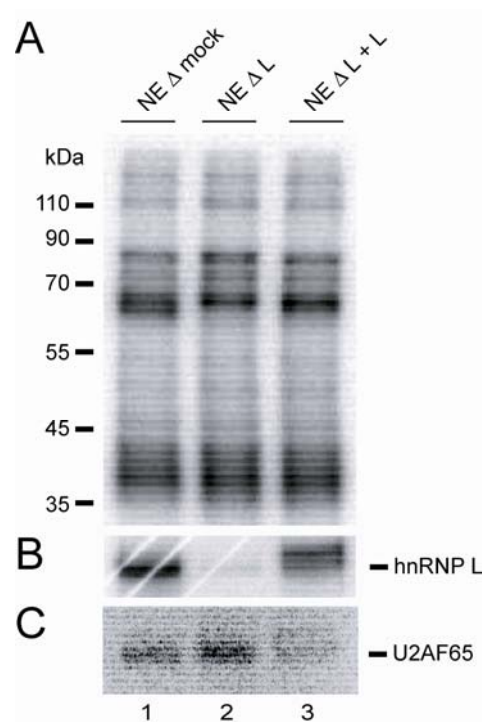


Figure 3.18

HnRNP L levels influence binding of U2AF65 to *TJP1* RNA. (A) UV crosslinking in HeLa nuclear extract. 32 P-labelled *TJP1* RNA transcripts were incubated in mock- (lane *NE Δ mock*) or hnRNP L-depleted (lane *NE Δ L*) nuclear extract under splicing conditions and subjected to UV crosslinking. HnRNP L-depleted extract was complemented by addition of purified recombinant hnRNP L (200 ng per 25- μ l splicing reaction; lane *NE Δ L + L*). Crosslinked proteins were separated on a 10% SDS-PAGE. Molecular size marker is given on the left. (B) UV crosslinking and immunoprecipitation using anti-hnRNP L monoclonal antibody. The position of immunoprecipitated hnRNP L is indicated on the right. (C) UV crosslinking and immunoprecipitation using anti-U2AF65 monoclonal antibody. The position of immunoprecipitated U2AF65 is indicated on the right.

3.3 *ITGA2*

Recently, a study on the mouse *ITGA2* (integrin alpha-2) gene identified a polymorphic CA-repeat region, (CA)₂₁ versus (CA)₆, in intron 1 which represents an intronic splicing enhancer (Cheli & Kunicki, 2006). Splicing activation was shown to be mediated by hnRNP L binding to the silencer sequence. The enhancement of splicing correlated thereby with the length of the CA repeat because hnRNP L bound to (CA)₂₁ with higher affinity than to (CA)₆.

The human *ITGA2* gene consists of 30 exons with a very long first intron of 37.2 kb which carries a CA-rich region similar to the mouse gene (Fig. 3.19). The *ITGA2* gene, also referred to as VLA-2 (very late activation protein 2 receptor), encodes the integrin alpha-2 subunit of a cell surface heterodimer (Takada & Hemler, 1989). Together with a common beta chain it forms a collagen receptor

3. Results

which plays an important role in adhesion of platelets by coupling platelets to collagen thus contributing to blood coagulation (Inoue *et al.*, 2003).

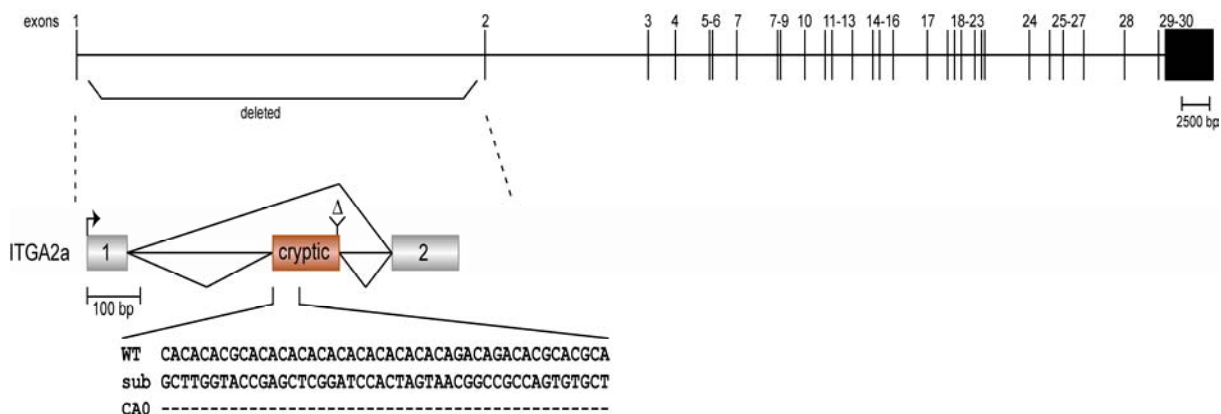


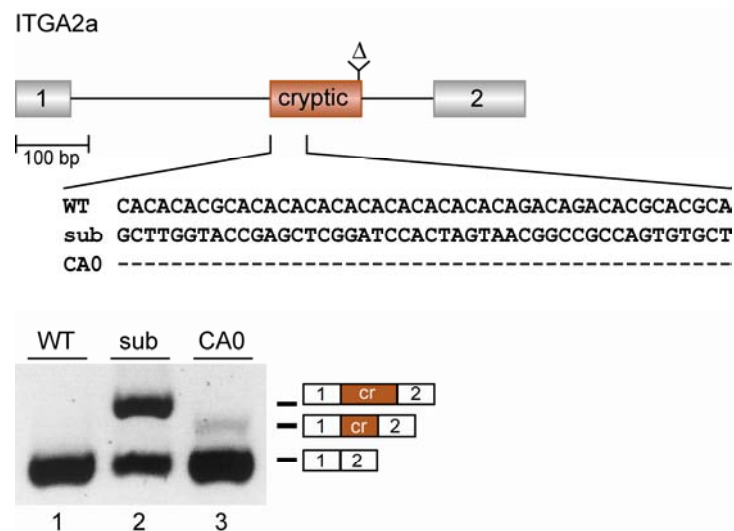
Figure 3.19

Schematic representation of the *ITGA2* gene structure and minigene constructs. Exons are represented by boxes, introns by lines (all sizes in scale). The *ITGA2* minigene consists of exons 1 and 2 with a shortened intron in between. The position of intron deletion is marked above the minigene by a Δ . The colour highlights a newly identified cryptic exon in intron 1. The lines above the minigene represent the normal splicing pattern; the lines below show an alternative splicing pattern. Wildtype (WT) and substituted (sub) sequence elements are given below. The minigene constructs were placed under the control of T7 and CMV promoter (shown as an arrow) for *in vitro* and *in vivo* splicing analysis.

3.3.1 Analysis of *ITGA2* minigene splicing leads to identification of a cryptic exon

In order to test the influence of the CA-rich region in *ITGA2* intron 1 on splicing a set of minigene constructs was made (Fig. 3.19). Each minigene consists of the first two exons with a shortened intron 1 in between (*ITGA2a*-WT). In addition, the CA-rich region was either substituted (*ITGA2a*-sub) or deleted (*ITGA2a*-CA0).

I first tested splicing of the minigene constructs *in vivo*. For that purpose, HeLa cells were transiently transfected, total RNA was isolated, and splicing was analysed by RT-PCR. As shown in Fig. 3.20, splicing of wildtype minigene resulted exclusively in two-exon mRNA (lane 1) whereas additional splicing products were observed when the CA repeats were either substituted or deleted (lanes 2, 3). Sequencing of the additional products revealed that they reflect inclusion of a cryptic exon into the mRNA (Fig. 3.20, top).

**Figure 3.20****CA-rich sequence in *ITGA2* intron 1 suppresses cryptic exon inclusion.**

Characterisation of *ITGA2* splicing using minigene constructs (ITGA2a). Exons are represented by boxes, introns by lines. All exon and intron parts are shown in scale. The cryptic exon in intron 1 is highlighted. HeLa cells were transiently transfected and splicing of minigene constructs was analysed by RT-PCR. Positions of PCR products corresponding to splice variants are given schematically on the right. The CMV promoter is depicted as an arrow.

The PCR product for the cryptic exon inclusion was shorter for the ITGA2a-CA0 construct than for ITGA2a-sub corresponding to the size of the deleted CA-rich region (compare lanes 2, 3). Moreover, less cryptic exon inclusion was detected for the ITGA2a-CA0 construct than for the substitution derivative.

Taken together, the *in vivo* splicing analysis indicated that the CA-rich region suppresses recognition of a cryptic exon in *ITGA2* intron 1.

3.3.2 Splicing analysis of additional *ITGA2* minigene constructs

Splicing analysis of *ITGA2* minigenes required shortening of intron 1 due to its enormous length of 37.2 kb (see Fig. 3.19). Therefore, the ITGA2a minigene constructs displayed an artificial context which may have led to the creation of the cryptic exon in intron 1. As shown in Fig. 3.19 and 3.20, the intron 1 deletion lies within the cryptic exon. In the natural context the splice sites of the cryptic exon would be separated by around 37 kb. For this reason, it is very unlikely that within the full-length intron the cryptic exon would be recognised in the same manner.

3. Results

A set of new minigene constructs was made to further investigate splicing of *ITGA2* (Fig. 3.21). Each of the new minigenes contained more intronic sequence than the initial ITGA2a constructs (Fig. 3.21, top) to test if cryptic exon inclusion could still take place.

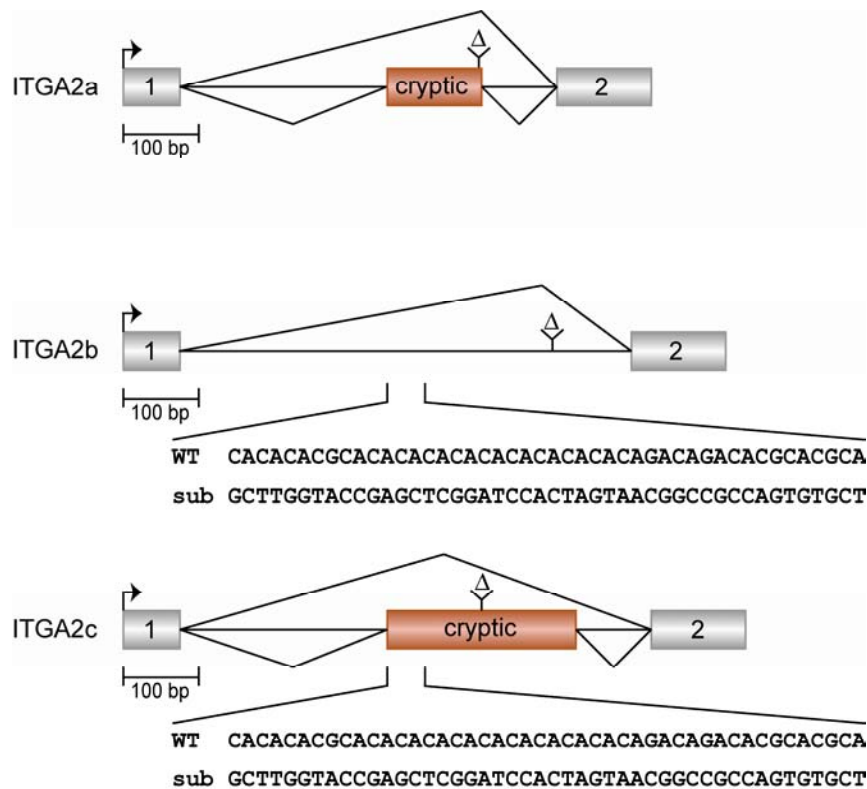


Figure 3.21

Comparative representation of *ITGA2* minigene constructs. Exons are represented by boxes, introns by lines (all sizes in scale). The position of intron deletion is indicated above each minigene by a Δ . The cryptic exon is highlighted. Lines above each minigene show the normal splicing pattern, lines below the alternative pattern. Wildtype (WT) and substituted (sub) sequence elements are given. The CMV promoter is depicted as an arrow.

In ITGA2b more of the 5' part of the intron was included into the minigene construct (middle), whereas more of the 3' part was retained in ITGA2c (bottom). A sequence alignment of the three *ITGA2* minigene constructs is shown in Fig. 3.22.

ITGA2a	1	GCTGCTGGTG	TTAGCGCTCA	GTCAAG ↓ GTAA	GCGGGGATTT	CGCTCTGCAT
ITGA2b	1	GCTGCTGGTG	TTAGCGCTCA	GTCAAG ↓ GTAA	GCGGGGATTT	CGCTCTGCAT
ITGA2c	1	GCTGCTGGTG	TTAGCGCTCA	GTCAAG ↓ GTAA	GCGGGGATTT	CGCTCTGCAT
ITGA2a	51	CGGCTGCAGG	AGGGACCCGC	GCGGCTTCCT	GGGGCTCGCT	GGAGGGATTG
ITGA2b	51	CGGCTGCAGG	AGGGACCCGC	GCGGCTTCCT	GGGGCTCGCT	GGAGGGATTG
ITGA2c	51	CGGCTGCAGG	AGGGACCCGC	GCGGCTTCCT	GGGGCTCGCT	GGAGGGATTG
ITGA2a	101	GGCGGAACTA	GGGAGCGAGC	TCCGTGTGTT	CCCTCGGATT	CCACGTGGAC
ITGA2b	101	GGCGGAACTA	GGGAGCGAGC	TCCGTGTGTT	CCCTCGGATT	CCACGTGGAC
ITGA2c	101	GGCGGAACTA	GGGAGCGAGC	TCCGTGTGTT	CCCTCGGATT	CCACGTGGAC
ITGA2a	151	TCGGGCTCCC	AGCCACCAGG	ACCTGCTTGG	AAAGAGATAC	TTCTTCGTTC
ITGA2b	151	TCGGGCTCCC	AGCCACCAGG	ACCTGCTTGG	AAAGAGATAC	TTCTTCGTTC
ITGA2c	151	TCGGGCTCCC	AGCCACCAGG	ACCTGCTTGG	AAAGAGATAC	TTCTTCGTTC
ITGA2a	201	CCTTCCTGCT	ATCCCAGGCC	TTGCGTCCCC	ATTAGGCGCG	TCTCGCCCTG
ITGA2b	201	CCTTCCTGCT	ATCCCAGGCC	TTGCGTCCCC	ATTAGGCGCG	TCTCGCCCTG
ITGA2c	201	CCTTCCTGCT	ATCCCAGGCC	TTGCGTCCCC	ATTAGGCGCG	TCTCGCCCTG
ITGA2a	251	TCTGTTAAGC	ACCCTCCAAA	GTCATCTCAG	TGAGTGCTTT	GTTTTAG GT A
ITGA2b	251	TCTGTTAAGC	ACCCTCCAAA	GTCATCTCAG	TGAGTGCTTT	GTTTTAG GT A
ITGA2c	251	TCTGTTAAGC	ACCCTCCAAA	GTCATCTCAG	TGAGTGCTTT	GTTTTAG GT A
ITGA2a	301	CACACACGCA	CACACACACA	CACACACACA	GACAGACACG	CACGCACTCA
ITGA2b	301	CACACACGCA	CACACACACA	CACACACACA	GACAGACACG	CACGCACTCA
ITGA2c	301	CACACACGCA	CACACACACA	CACACACACA	GACAGACACG	CACGCACTCA
ITGA2a	351	AGCATGGACA	GTGTGGGGAT	CCGCGGGTGT	CCTGGAGACC	GCGAGTCCTG
ITGA2b	351	AGCATGGACA	GTGTGGGGAT	CCGCGGGTGT	CCTGGAGACC	GCGAGTCCTG
ITGA2c	351	AGCATGGACA	GTGTGGGGAT	CCGCGGGTGT	CCTGGAGACC	GCGAGTCCTG
ITGA2a	401	CAGCCCAAC ---	-----	-----	-----	-----
ITGA2b	401	CAGCCCAACT	TCCCCTGGCT	CAGACAGGTG	CCGCGCCTGC	CAGAGACCAG
ITGA2c	401	CAGCCCAACT	TCCCCTG ---	-----	-----	-----
ITGA2a	451	-----	-----	-----	-----	-----
ITGA2b	451	GGAAATCCGG	AATCCATTGC	TGATGGCGGC	CTGCAGCTCC	CCAGCGCGGT
ITGA2c	451	-----	-----	-----	-----	-----
ITGA2a	501	-----	-----	-----	-----	-----
ITGA2b	501	GCTTGCTGGC	TTTG-----	-----	-----	-----
ITGA2c	501	-----	----- ATTAG	TAAGATAATG	AATTATGCAT	ATATAGTATT
ITGA2a	551	-----	-----	-----	-----	-----
ITGA2b	551	-----	-----	-----	-----	-----
ITGA2c	551	TAGCTTGGTG	CCGGGCCTCC	AAATAAGTAC	TTAAAAAGCA	GTAGTTATTT
ITGA2a	601	-----	-----	-----	↓ TTAG GTCAA	GCAAGTTTTTC
ITGA2b	601	-----	-----	-----	-TTAGGTCAA	GCAAGTTTTTC
ITGA2c	601	TTTCTGGGTA	AACGTTGATA	AGTAATAATT	ATTAG GTCAA	GCAAGTTTTTC
ITGA2a	651	TTAAAAATGG	CCTTTGAGAA	TTATACGACA	CACTTATATT	CATGAATTGT
ITGA2b	651	TTAAAAATGG	CCTTTGAGAA	TTATACGACA	CACTTATATT	CATGAATTGT
ITGA2c	651	TTAAAAATGG	CCTTTGAGAA	TTATACGACA	CACTTATATT	CATGAATTGT
ITGA2a	701	AAATATGTGC	TATTTTCATAT	TTTTCTTAAT	↓ AGGCATTTTA	AATTGTTGTT
ITGA2b	701	AAATATGTGC	TATTTTCATAT	TTTTCTTAAT	AGGCATTTTA	AATTGTTGTT
ITGA2c	701	AAATATGTGC	TATTTTCATAT	TTTTCTTAAT	AGGCATTTTA	AATTGTTGTT

Figure 3.22

Sequence alignment of *ITGA2* wildtype minigene constructs. Given are the complete intron 1 sequence and parts of the exon 1 and 2 sequence of the respective *ITGA2* minigenes (ITGA2a, ITGA2b, and ITGA2c). Exons are shown in bold, the cryptic exon in blue, the CA-rich sequence in red. All splice sites are marked by an arrow.

3. Results

In ITGA2b more of the 5' part of the intron was included into the minigene construct (middle), whereas more of the 3' part was retained in ITGA2c (bottom). A sequence alignment of the three *ITGA2* minigene constructs is shown in Fig. 3.22.

I tested splicing of the new minigene constructs *in vitro* in HeLa cell nuclear extract. Aliquots were taken after 0, 60, and 120 minutes and splicing was analysed by RT-PCR using gene-specific primers. Fig. 3.23 shows that the two sets of minigene constructs (ITGA2b, panel A; ITGA2c, panel B) displayed different splicing pattern.

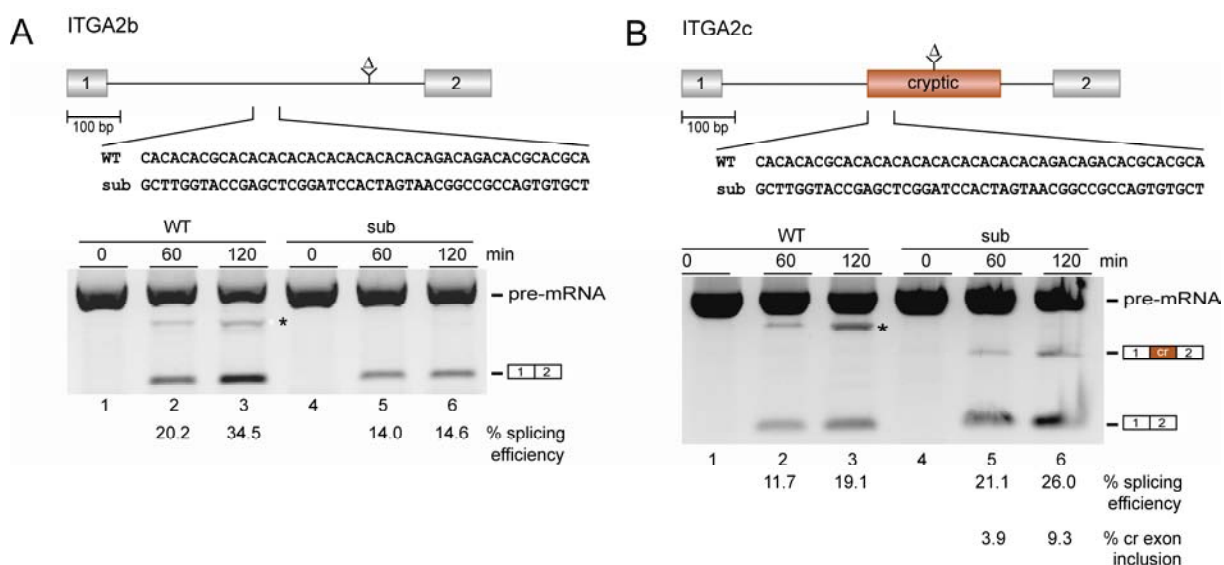


Figure 3.23

***In vitro* splicing analysis of different *ITGA2* minigene constructs.** (A) Characterisation of *ITGA2* splicing using ITGA2b wildtype (WT) and substitution (sub) minigene constructs. Exons are represented by boxes, introns by lines. All exon and intron parts are shown in scale. The position of intron deletion is indicated by a Δ . Wildtype and substituted sequence elements are given below the minigene. *In vitro* transcribed pre-mRNAs were incubated in HeLa nuclear extract for the indicated times and splicing was analysed by RT-PCR. Positions of PCR products corresponding to pre-mRNA and spliced product are given schematically on the right. The asterisk marks an unspecific PCR product. The percentages of splicing efficiency are given below the corresponding lanes. (B) Schematic representation of ITGA2c minigene constructs. The cryptic exon is highlighted. *In vitro* splicing of ITGA2c wildtype (WT) and substitution (sub) minigene constructs was analysed by RT-PCR. The positions of pre-mRNA and splicing variants are schematically shown on the right. The percentages of splicing efficiency and cryptic exon inclusion are given below the corresponding lanes.

For the ITGA2b constructs cryptic exon inclusion was not detected (Fig. 3.23A). Substitution of the CA-rich sequence resulted in a decrease of splicing efficiency from 34.5 to 14.6% after 120 minutes (Fig. 3.23A, compare lanes 3, 6). Note that the splice sites of the cryptic exon which was recognised in the ITGA2a substitution and deletion constructs were also present in ITGA2b and c. In contrast

to ITGA2b, splicing of the ITGA2c substitution derivative again resulted in inclusion of the cryptic exon (Fig. 3.23B, lanes 5, 6). Sequencing of the spliced product revealed that the same splice sites, as in ITGA2a, were used. Splicing of the wildtype pre-mRNA yielded exclusively the two-exon product (lanes 2, 3). Moreover, the overall splicing efficiency of the substitution construct was with 26% better than the wildtype (19%; compare lanes 3, 6).

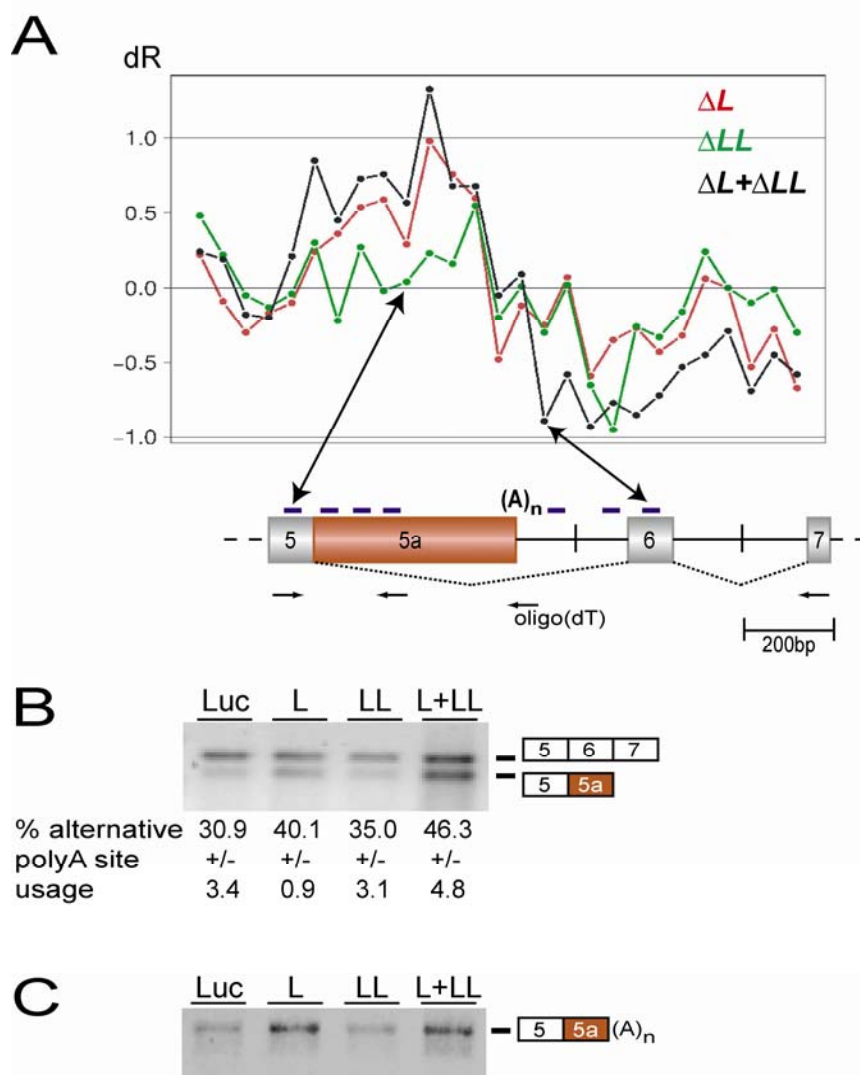
Taken together, these data indicated that the CA-rich sequence in the ITGA2b minigene context functions as an enhancer of splicing efficiency. The ITGA2c minigene constructs on the other hand showed the same splicing pattern than the ITGA2a constructs, where the CA region suppressed inclusion of a cryptic exon.

3.4 *ASAH1*

The human *ASAH1* gene (N-acylsphingosine amidohydrolase 1) encodes an acid ceramidase, a lysosomal enzyme which catalyses the degradation of ceramide into sphingosine and fatty acid (Li *et al.*, 1999). A deficiency in the acid ceramidase activity leads to a lysosomal storage disorder known as Farber disease. Three missense mutations in the human *ASAH1* gene have been associated with the disease. The gene itself spans approximately 30 kb and contains 14 exons. The primary gene product represents a 50 kDa precursor polypeptide that is posttranscriptionally cleaved to give rise to the mature heterodimer which consists of a nonglycosylated α subunit (14 kDa) and a glycosylated β subunit (40 kDa). Until now, alternative polyadenylation was not reported for the *ASAH1* gene.

3.4.1 HnRNP L mediates alternative poly(A) site selection in the *ASAH1* gene

The *ASAH1* gene was one of the 11 hnRNP L target genes identified by our combined RNAi and microarray analysis (as described before for *TJP1*). This gene was particularly interesting since three probe sets downstream of exon 5 showed higher dR values upon hnRNP L and L/LL double knockdown whereas most of the downstream probe-set signals were reduced (Fig. 3.24A).

**Figure 3.24**

Alternative poly(A) selection in the *ASAH1* target gene. (A) Analysis of the microarray data for the *ASAH1* gene. The diagram shows log₂ ratios of probe-set signal intensities (dR) each relative to the luciferase control values across the *ASAH1* gene. Three values are given (Y-axis: ΔL , knockdown of hnRNP L, in red; ΔLL , knockdown of hnRNP LL, in green; $\Delta L+\Delta LL$, knockdown of hnRNP L and LL, in black) for each probe set (X-axis). Probe-set and primer positions in *ASAH1* are shown below. (B) Total RNA was prepared after knockdown in HeLa cells (as indicated above the lanes) and splicing/alternative polyadenylation was assayed by RT-PCR. A combination of three gene-specific primers was used to detect spliced mRNA and internally polyadenylated mRNA as indicated on the right. The percentages of internal polyadenylation with standard deviations ($n=3$) are given below the corresponding lanes (because different reverse primers were used these values only allow comparison between knockdown and control samples and do not yield absolute numbers on the mRNA variants). (C) RT-PCR analysis using an oligo(dT) reverse primer and a gene-specific forward primer to detect use of the internal poly(A) site (as indicated on the right).

An implication from this observation was that this may be a case of alternative polyadenylation. The usage of a poly(A) site in *ASAH1* intron 5 may have produced an mRNA missing all downstream exons. To test this hypothesis, RT-

PCR validation was performed using a combination of three primers (indicated in panel A). The forward primer was placed in exon 5 whereas two different reverse primers were used; one in exon 7 and one in intron 5 to detect the spliced mRNA as well as the internally polyadenylated mRNA. As shown in Fig. 3.24B, two distinct products were detected by RT-PCR analysis and which were also confirmed by sequencing. The upper band represented the normal splicing product consisting of exons 5 to 7 whereas the lower band corresponded to an mRNA containing exons 5 and part of intron 5 which we termed 5a. This alternative PCR product increased from 31% in the control knockdown (lane *Luc*) to 40% and 46% in hnRNP L and L/LL double knockdown, respectively (lanes *L*, *L+LL*). No significant increase was detected for the hnRNP LL knockdown (35%; lane *LL*). Direct evidence for alternative polyadenylation was obtained by RT-PCR analysis using a different reverse primer (shown in panel A). An oligo(dT) reverse primer in combination with the exon 5 forward primer allowed detection of internally polyadenylated mRNA (panel C). The level of alternatively polyadenylated mRNA increased upon hnRNP L and L/LL double knockdown, showing that the internal poly(A) site was properly polyadenylated, and supporting the previous data. In addition, usage of the internal polyadenylation site was confirmed by sequencing of the PCR product, which also allowed identification of the definite cleavage site. In summary, analysis of the microarray data revealed a new mode of regulation for hnRNP L. In the *ASAH1* gene it was shown for the first time that hnRNP L can also function in alternative polyadenylation by repressing the use of an internal poly(A) site.

4. Discussion

4.1 Microarray analysis: Genome-wide search for hnRNP L target genes

The DNA microarray technology has developed over the last decade as a powerful tool for large-scale analysis of gene expression (Gershon, 2002). The underlying principle is based on the preference of single-stranded nucleic acids to bind complementary sequences. A microarray can contain up to millions of DNA molecules (probes) with known sequences, which are immobilised in an ordered manner on a solid support (e.g. glass or silicone). Samples, which are applied on the microarray, are fluorescently labelled and hybridise to the complementary probes on the array (Heller, 2002). After excitation by a laser beam, the emitted fluorescence, which is proportional to the degree of hybridisation, is detected and analysed by computational tools. There is a variety of applications for the microarray technology, for example gene expression analysis or screening of samples for single nucleotide polymorphisms (SNPs). In addition, microarrays are widely used for medical purposes e.g. in genetic disease and cancer diagnostics.

More recently, the DNA microarray technology has emerged as a tool for the genome-wide analysis of alternative splicing (Lee & Roy, 2004). In order to allow detection of alternative splice variants, the microarray probe design had to be modified. In a standard microarray each gene is typically targeted by a single probe set placed at the 3' end of the transcript (Okoniewski & Miller, 2008). For detection of alternative splicing on the other hand, it is necessary that probe sets cover each exon along the entire length of the gene (exon array). Exon junction arrays represent another possibility for the high-throughput analysis of alternative splicing. In this kind of microarrays, probes are designed to cover each exon-exon junction which might be spliced together in an alternative splicing event.

Relevant for this work was the Affymetrix GeneChip Human Exon 1.0 Array, a microarray platform that allows detection of alternative splicing on a genome-wide level. Total RNA, prepared from the hnRNP L and LL knockdown as well as luciferase control reactions, was processed through the Affymetrix Exon Array. The huge amount of data obtained from the microarray was analysed using

computational tools to identify differences in exon expression signals between knockdown and control reactions (Hung *et al.*, 2008). From an initial list of about 50 target genes, we could experimentally validate reproducible differences in response to hnRNP L depletion for alternative splicing of 11 genes.

Several recent studies showed that the Affymetrix exon array is a viable tool for the global analysis and identification of novel alternative splicing events. McKee and co-workers found stimulus-induced changes in alternative splicing that contribute to gene regulation (McKee *et al.*, 2007). Cheung and co-workers validated 14 alternative splicing changes associated with primary brain tumors, seven of which were novel events (Cheung *et al.*, 2008). A lot more examples of cancer-related alternative splicing events could be identified using exon array analysis (Gardina *et al.*, 2006; Jhavar *et al.*, 2008; Thorsen *et al.*, 2008). Another major application for exon arrays is the search for tissue-specific alternative splicing (Clark *et al.*, 2007; Das *et al.*, 2007). Oberdoerffer and co-workers identified hnRNP L-like (hnRNP LL), a paralog of hnRNP L, as an inducible regulator of alternative splicing (Oberdoerffer *et al.*, 2008). Exon array data suggested that hnRNP LL functions as a global regulator of alternative splicing in activated T-cells. This finding was especially important for our work since our microarray analysis yielded no alternative splicing targets in HeLa cells which points to a tissue-specific role of hnRNP LL (Hung *et al.*, 2008). Several other splicing factors including hnRNP L have been shown to participate in alternative splicing of *CD45* (House & Lynch, 2006; Rothrock *et al.*, 2005). None of them, however, was induced upon T-cell activation. HnRNP L and LL were demonstrated to have overlapping but distinct binding specificities in the *CD45* gene (Topp *et al.*, 2008).

4.2 Crossregulation of the hnRNP L proteins

RNAi knockdown of hnRNP L and LL in HeLa cells revealed a reciprocal regulation of the two proteins. Quantitative real-time PCR analysis confirmed efficient reduction of hnRNP L and LL mRNA levels after either single or double knockdown (see Fig. 3.7A). Interestingly, with the reduction of hnRNP L mRNA levels a simultaneous 1.5-fold increase of hnRNP LL expression was detected. In another experiment we also observed the reciprocal effect that hnRNP L levels

increased upon knockdown of hnRNP LL (Hung *et al.*, 2008). Recently, we found a crossregulatory mechanism explaining the reciprocal regulation of the two proteins (Rossbach *et al.*, 2009). It was demonstrated that hnRNP L autoregulates its own expression by alternative splicing. Inclusion of a “poison exon”, which introduces a premature termination codon into the mRNA leads to degradation of the respective mRNA by nonsense-mediated decay. HnRNP L was also shown to activate inclusion of a similar “poison exon” into the mRNA of hnRNP LL. But why did I observe no upregulation of hnRNP L levels upon hnRNP LL knockdown? Most likely, the effect on hnRNP L was more sensitive to biological variations than hnRNP LL, whose upregulation was very strong. Note that hnRNP L levels in HeLa cells are already very high. Quantitation of hnRNP L and LL protein levels by Western blot analysis showed that hnRNP L is approximately ten times more abundant in HeLa cells than hnRNP LL (Hung *et al.*, 2008). High levels of hnRNP L upregulation in HeLa cells are therefore unlikely. Considering that hnRNP LL but not hnRNP L expression was induced upon T-cell activation it would be interesting to test other types of immune cells for hnRNP LL expression levels to confirm a tissue-specific role of the protein. Moreover, to find cell types, which express a higher level of hnRNP LL than HeLa cells, is an important task in order to investigate crossregulation between the hnRNP L proteins.

4.3 Diverse roles of hnRNP L in splicing regulation

HnRNP L was initially identified as a splicing activator of the human *eNOS* gene (Hui *et al.*, 2003b). Determination of the binding specificity revealed that hnRNP L preferentially binds to CA repeats and certain CA-rich sequences (see Fig. 1.6 and Hui *et al.*, 2005). Considering that CA-repetitive sequences represent the most common simple sequence repeat in the human genome (Waterston *et al.*, 2002), it was very likely that hnRNP L also regulates splicing of other genes besides *eNOS*. The identification of new target genes revealed the involvement of hnRNP L in different types of splicing regulation mainly in regulation of alternative splicing. Fig. 4.1 gives an overview on what we currently know about hnRNP L's mode of action.

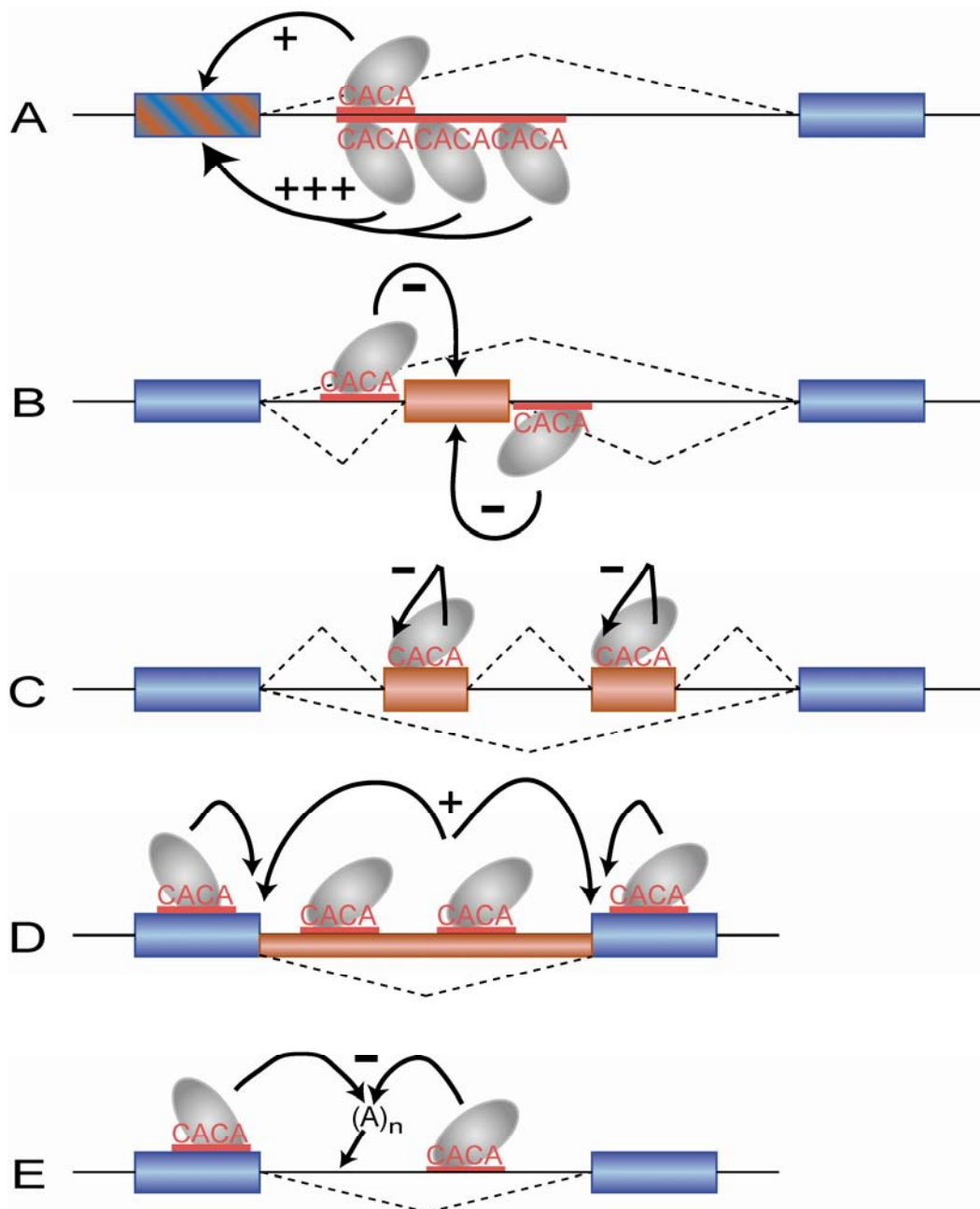


Figure 4.1

Splicing regulation by hnRNP L. Schematic representation of alternative splicing events regulated by hnRNP L (depicted in grey) which acts as an activator (+) or repressor (-). Constitutive exons are represented by blue boxes; alternative exons are shown in red (adapted from Hung *et al.*, 2008). **(A)** HnRNP L functions as an activator of exon inclusion or on splicing efficiency sometimes in a length-dependent manner (e.g. *eNOS*, *ITGA2*). **(B)** HnRNP L represses splicing of cassette-type exons by binding to intronic CA-rich silencer sequences either downstream or upstream of the regulated exon (e.g. *SLC2A2*, *TJP1*). **(C)** HnRNP L represses splicing of multiple alternative exons or pseudoexons by binding to exonic silencer elements (e.g. *ARGBP2*, *LIFR*, *ITGA2*, *CD45*). **(D)** HnRNP L acts as an activator of intron retention (e.g. *DAF*; note that this represents a hypothetical mechanism). **(E)** HnRNP L represses internal polyadenylation (e.g. *ASAHI*; note that this represents a hypothetical mechanism).

First, hnRNP L was shown to determine the splicing efficiency of an intron by binding to an intronic CA-rich enhancer (panel A). In case of the human *eNOS* and

the mouse *ITGA2* gene splicing activation correlated with the length of the CA repeat (Cheli & Kunicki, 2006; Hui *et al.*, 2003b). HnRNP L can furthermore activate inclusion of an alternative exon into the mRNA (Hui *et al.*, 2005). Second, it was shown for several cases that hnRNP L functions as a splicing repressor by binding to an intronic CA-rich silencer element which can be located either upstream or downstream of the regulated exon (panel B). *TJP1* and *SLC2A2* fall into this category.

Third, hnRNP L binds to an exonic CA-rich silencer and represses inclusion of the exon into the mRNA (panel C). This mode of regulation by hnRNP L was first identified in the *CD45* gene where hnRNP L represses inclusion of the variable exon 4 upon T-cell activation (Rothrock *et al.*, 2005). Suppression of multiple alternative exons in one intron as shown for *ARGBP2* and *LIFR* as well as suppression of a cryptic exon in the case of *ITGA2* also belong in this category (Hung *et al.*, 2008). The regulatory sequence resides in each case in an exon. Fourth, two genes, namely *DAF* and *STRA6* are known where hnRNP L acts as an activator of intron retention (panel D). So far, it is not known however if the regulatory sequence is located in the intron or the flanking exons. Finally, alternative poly(A) site selection represents a new mode of hnRNP L-mediated regulation (panel E). In the *ASAH1* gene, hnRNP L was shown to repress usage of an internal polyadenylation site. The mechanism of action is yet unknown.

Taken together, hnRNP L displayed a surprising diversity of regulatory functions in both splicing and polyadenylation. This versatility clearly distinguishes hnRNP L from other proteins of the hnRNP family whose function in splicing is often restricted to splicing repression e.g. PTB, also referred to as hnRNP I, is homologous to hnRNP L and represent a well characterised repressor of alternative splicing (Spellman *et al.*, 2005).

4.4 The mechanism of alternative splicing repression by hnRNP L

HnRNP L was revealed as a global regulator of alternative splicing. For the two genes *SLC2A2* and *TJP1* I have obtained further insights into the mechanism of splicing repression of cassette-type exons mediated by hnRNP L.

4.4.1 HnRNP L interferes with 5' splice site recognition

SLC2A2 was identified as an hnRNP L candidate gene by a genome-wide database search. By mutational analysis I could demonstrate that a CA repeat sequence in intron 4 of the gene represents a splicing silencer element (Fig. 3.2). Subsequent depletion and complementation experiments confirmed that alternative splicing of *SLC2A2* is regulated by hnRNP L (Fig. 3.3). In addition, we could show in UV crosslinking and immunoprecipitation experiments that hnRNP L directly binds to the intronic CA-repeat silencer element (Hui *et al.*, 2005). The position of the CA repeat, namely close to the 5' splice site of the regulated exon, suggested that hnRNP L may interfere with spliceosome assembly.

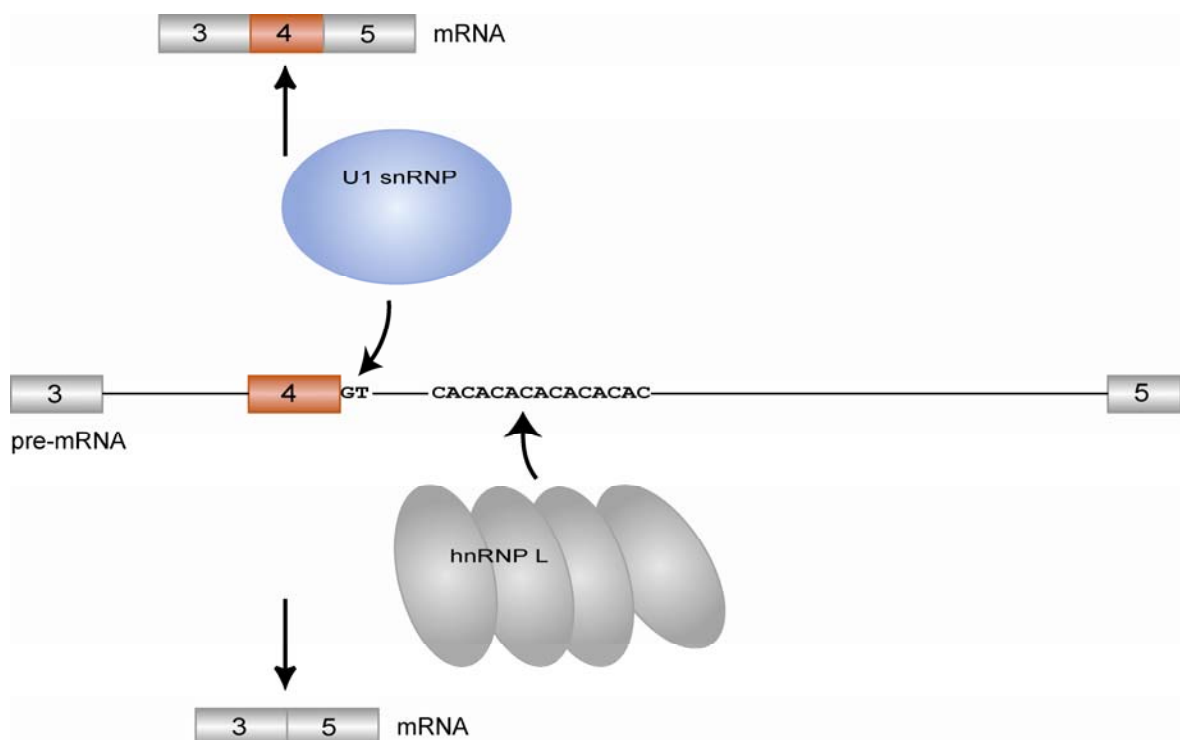


Figure 4.2

Model of splicing regulation by hnRNP L in the *SLC2A2* gene. Exons are represented by boxes, introns by lines. The alternatively spliced exon 4, the CA repeat sequence, and the 5' splice site (GT) are highlighted. The 5' splice site of *SLC2A2* pre-mRNA is recognised by the U1 snRNP in the absence of hnRNP L (top). This leads to inclusion of the regulated exon 4 in the mRNA. Bound hnRNP L proteins block 5' splice site recognition by the U1 snRNP which leads to skipping of the regulated exon (bottom).

Analysis of spliceosomal E complex formation which is characterised by the U1 snRNP recognising the 5' splice site of the pre-mRNA provided first evidence on the mechanism of *SLC2A2* splicing regulation (Fig. 3.4). The E5' complex formed

less efficiently on CA repeats-containing pre-mRNA than on the substitution derivative. Interaction of the U1 snRNP is mediated in part by base-pairing of the snRNA component with the first six nucleotides of the intron including the 5' splice site (Krämer *et al.*, 1984). Psoralen crosslinking experiments showed an increased U1/RNA interaction in the absence of CA-repeats or hnRNP L protein (Fig. 3.5 and 3.6). Together with the data obtained from analysis of E complex formation these findings proposed a mechanism for repression of *SLC2A2* exon 4 inclusion (Fig. 4.2). The 5' splice site of *SLC2A2* exon 4 is recognised by the U1 snRNP in the absence of hnRNP L which leads to inclusion of the exon into the mRNA. If hnRNP L binds to the CA repeat sequence recruitment of the U1 snRNP is impaired resulting in exon 4 skipping.

It has been proposed for PTB (hnRNP I) that in some cases, where multiple PTB binding sites flank the regulated exon, multimerisation of PTB proteins creating a zone of silencing (Wagner & Garcia-Blanco, 2001). It remains to be investigated if a similar mechanism where hnRNP L proteins bind cooperatively to long CA repeats can be shown for *SLC2A2* splicing regulation.

4.4.2 HnRNP L interferes with 3' splice site recognition

TJP1 represents one of 11 hnRNP L target genes identified by our microarray analysis (Fig. 3.8). *TJP1* contains a variable exon whose inclusion was significantly increased upon knockdown of hnRNP L. I could demonstrate by mutational analysis that a CA-rich cluster in intron 19 functions as an hnRNP L-dependent splicing silencer (Fig. 3.11). Band shift as well as UV crosslinking experiments confirmed that hnRNP L directly interacts with the intronic CA-rich silencer element (Fig. 3.14 and 3.15). The CA-rich splicing regulatory sequence resides in the polypyrimidine tract of *TJP1* intron 19 and consists of three hnRNP L high-score binding motifs, one of those next to the 3' splice site of the regulated exon (Fig. 3.9). These findings suggested that *TJP1* exon 20 repression by hnRNP L is mediated through impaired 3' splice site recognition.

Early recognition of 3' splice sites is mediated by the U2 snRNP auxiliary factor (U2AF) (Ruskin *et al.*, 1988; Zamore *et al.*, 1992). U2AF is composed of a large (U2AF65) and a small (U2AF35) subunit which interact with each other to form a stable heterodimer. Both subunits belong to the family of SR-related proteins (see

Table 1.1). UV crosslinking and immunoprecipitation experiments revealed that hnRNP L interferes with binding of U2AF65 to the polypyrimidine tract of *TJP1* intron 19 (Fig. 3.18). U2AF65 binds to the polypyrimidine tract of pre-mRNAs, interacts with U2AF35, SF1, and SF3b155, and finally promotes recruitment of the U2 snRNP to the branch site (Jenkins *et al.*, 2008). The U2AF65 protein contains several distinct domain structures to promote all of these interactions (Fig. 4.3). Two RRM s contact the polypyrimidine tract; a C-terminal U2AF homology motif (UHM) interacts with SF1 and SF3b155; a tryptophan-containing UHM ligand motif (ULM) positions the U2AF35 small subunit at the 3' splice site and an N-terminal RS domain promotes recruitment of the U2 snRNP. Therefore, U2AF65 binding to the polypyrimidine tract represents a crucial step in 3' splice site recognition.

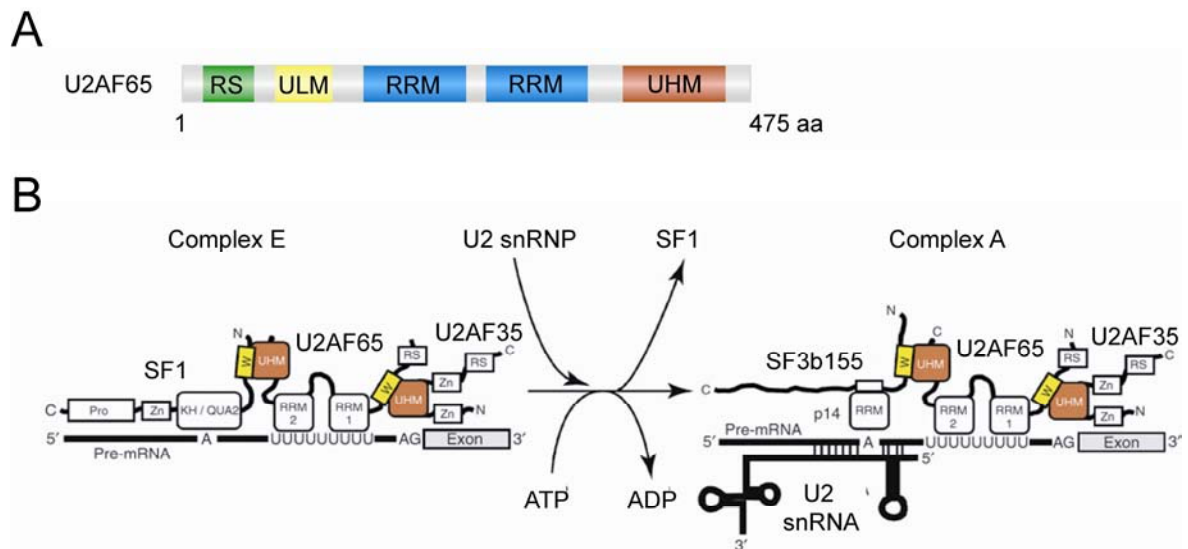


Figure 4.3
Schematic representation of U2AF65 domain structure and interactions in spliceosomal E and A complex. (A) U2AF65 (P26368; 475 amino acids) protein contains two canonical RNA recognition motifs (RRM, blue), U2AF homology motif (UHM, red), UHM ligand motif (ULM, yellow), and an arginine-serine-rich domain (RS, green). (B) Schematic representation of the 3' splice site in spliceosomal complexes E and A. UHM (red) interaction with tryptophan-containing ULM (W, yellow) is highlighted (adapted from Corsini *et al.*, 2007).

Regulation of alternative splicing is often mediated by promotion or inhibition of early events in spliceosome assembly such as 5' splice site or 3' splice site recognition. Several cases are known where splicing regulators interfere with binding of U2AF to the 3' splice site region leading to exon skipping. One well studied example belongs to the sex determination pathway in *Drosophila melanogaster* (Black, 2003). In female flies, sex lethal (sxl) protein binds to the 3'

4. Discussion

splice site region of exon 2 in the *transformer* (*tra*) pre-mRNA thereby blocking recruitment of U2AF. This leads to the subsequent skipping of the exon which deletes a stop codon from the *tra* mRNA and allows expression of *tra* protein. Further examples for inhibition of 3' splice site recognition are given by the polypyrimidine tract binding protein (PTB/hnRNP I) which is involved in the repression of several tissue-specific exons by preferentially binding to pyrimidine-rich intronic silencers upstream or downstream of the regulated exon (Wagner & Garcia-Blanco, 2001). PTB has been shown to compete with U2AF65 for binding to the polypyrimidine tract, thereby blocking splicing of the corresponding exon (Matlin *et al.*, 2007; Sauliere *et al.*, 2006). For alternative splicing of *TJP1* exon 20 I could demonstrate a similar mechanism where hnRNP L represses splicing by interfering with U2AF binding most likely leading to inhibition of 3' splice site recognition due to impaired recruitment of the U2 snRNP (Fig. 4.4)

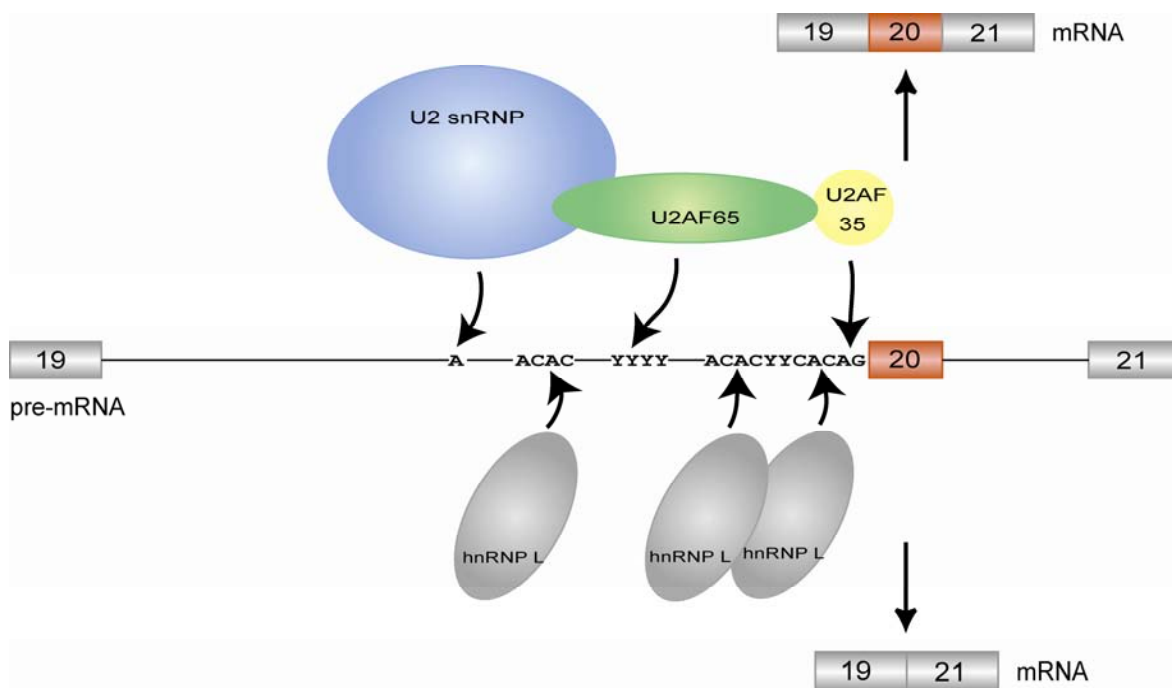


Figure 4.4

Model of splicing regulation by hnRNP L in the *TJP1* gene. Exons are represented by boxes, introns by lines. The alternatively spliced exon 20, the branch point adenosine (A), polypyrimidine tract (Y), hnRNP L binding motifs, and the 3' splice site (AG) are highlighted. The 3' splice site, polypyrimidine tract, and branch site of *TJP1* pre-mRNA are recognised by U2AF (65 and 35 kDa subunits) and the U2 snRNP in the absence of hnRNP L (top). This leads to inclusion of the regulated exon 20 in the mRNA. Binding of hnRNP L proteins blocks 3' splice site recognition by U2AF which leads to skipping of the regulated exon (bottom).

This hypothesis was supported by the finding that *TJP1* intron 19 contains a

comparatively weak polypyrimidine tract (MaxEntScan score 5.2) whereas the 3' splice site of exon 21 has an average strength (score 8.2). For comparison, the 3' splice site of the good splicing substrate MINX has very high score of 12.5. The affinity of U2AF65 for *TJP1* intron 19 was, in correlation with the low score, very low whereas hnRNP L showed a high affinity for the same RNA substrate (Fig. 3.14 and 3.16). It has been shown that although dispensable for splicing of introns with a strong polypyrimidine tract (so-called AG-independent) U2AF35 is essential for splicing of introns that contain a weak polypyrimidine tract (Pacheco *et al.*, 2006). The small subunit of U2AF recognises the 3' splice site consensus signal (Wu *et al.*, 1999). Considering that one of the hnRNP L binding motifs resides next to the 3' splice site in a poor polypyrimidine tract it would be very interesting to investigate if U2AF35 binding is dispensable for *TJP1* intron 19 splicing. The in parts hypothetical mechanism of *TJP1* exon 20 splicing repression is illustrated in Fig. 4.4 Without hnRNP L the 3' splice site of exon 20 is recognised by U2AF which recruits the U2 snRNP to the branch site leading to splicing of the intron. The presence of hnRNP L blocks U2AF65 binding to the polypyrimidine tract and likely also recognition of the 3' splice site by U2AF35 resulting in exon 20 skipping.

4.4.2.1 Short *TJP1* RNA transcripts form different secondary structures

Surprisingly, free RNAs transcribed from wildtype and mutant *TJP1* constructs did not show the same migration behaviour in band shift experiments although they were of the same length (Fig. 3.14). An implication of this finding was that the transcripts form different secondary structures.

I used the RNAfold web server of the Vienna RNA Websuite which can be used to predict the minimum free energy (MFE) secondary structure and pair probabilities of single stranded RNA sequences (Gruber *et al.*, 2008). Structure drawings of the predicted intra-molecular base-pairing interactions are shown in Fig. 4.5 Surprisingly, the predicted secondary structures for *TJP1* wildtype and the initial mutant construct show a higher similarity than the two new mutants mutTG and mutG. Apparently, the secondary structure cannot fully explain the differences observed in the migration behaviour of the short *TJP1* RNA transcripts.

One has to consider the possibility that the different hnRNP L binding abilities of the *TJP1* RNA transcripts are due to their distinct RNA folding. To circumvent this

issue UV crosslinking and immunoprecipitation experiments were carried out with *TJP1* wildtype RNA transcripts alone using hnRNP L-depleted nuclear extract.

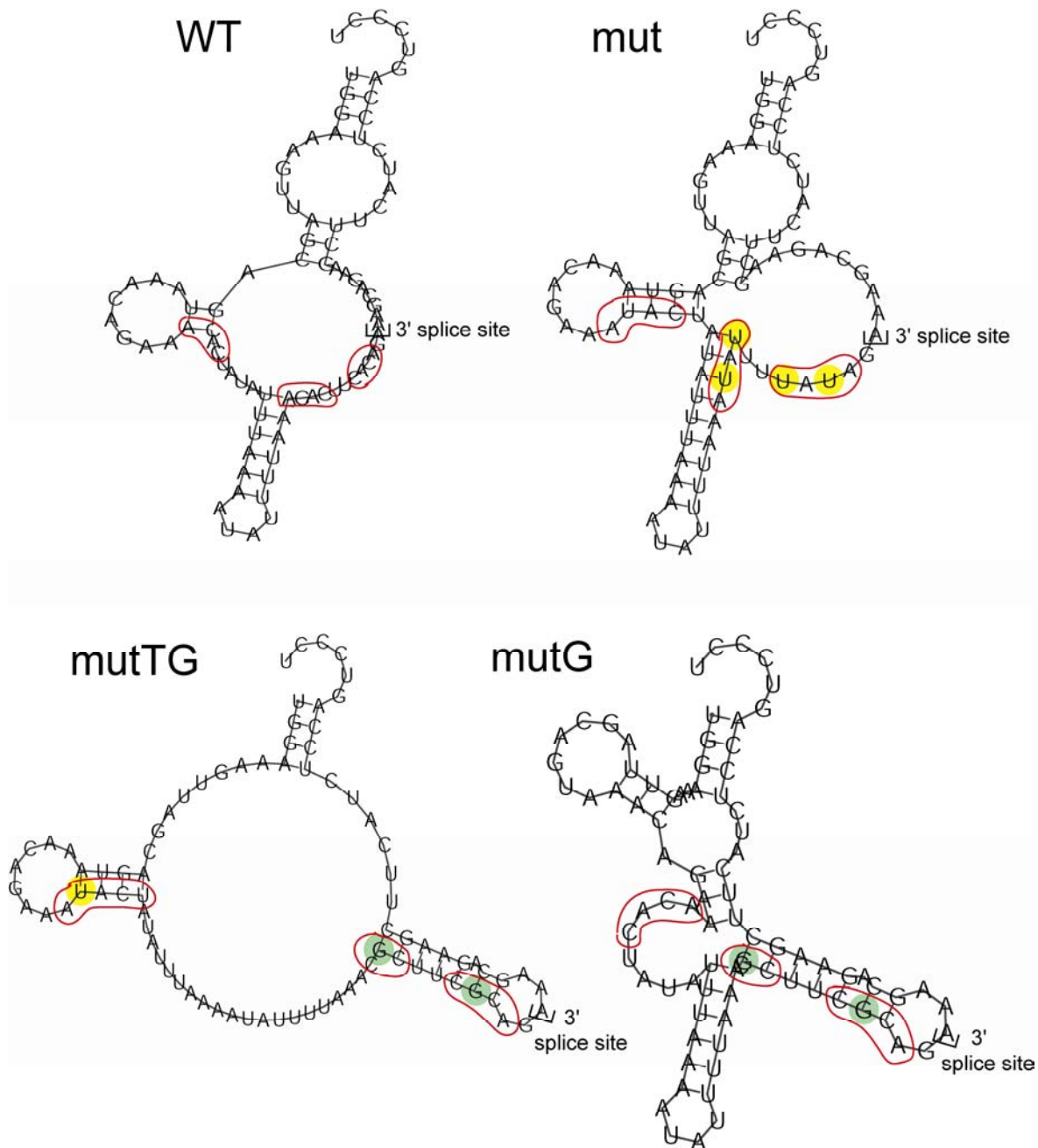


Figure 4.5
Secondary structure of *TJP1* short RNA transcripts. Secondary structure and base pair probabilities predicted for the *TJP1* short RNA transcripts (84 nt) with the RNAfold web server of the Vienna RNA Websuite (<http://rna.tbi.univie.ac.at>). Point mutations (C→T in yellow; A→G in green) are highlighted. The hnRNP L binding motifs are encircled. The 3' splice site is indicated for each transcript.

4.5 HnRNP L regulates alternative poly(A) site selection

Almost all eukaryotic mRNAs undergo 3' end processing namely polyadenylation which is responsible for the addition of poly(A) tails (Colgan & Manley, 1997).

Polyadenylation of mRNAs occurs in a two-step reaction involving endonucleolytic cleavage at the polyadenylation site (poly(A) site) followed by synthesis of an adenosine tail (poly(A) tail). In mammals, several factors are known to participate in the polyadenylation reaction. These proteins include cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors CF I and CF II, and poly(A) polymerase (PAP). The mRNA cleavage and polyadenylation site is defined by conserved sequence elements, such as the polyadenylation signal AAUAAA, which are located 10 to 30 nucleotides upstream of the cleavage site. 3' end formation plays an important role in gene expression since improperly processed mRNAs are not transported out of the nucleus (Lutz, 2008). Poly(A) tails have been shown to influence mRNA stability and translation.

Recent genome-wide studies concluded that the previously underestimated alternative polyadenylation represents a common mechanism contributing to protein diversity (Tian *et al.*, 2005). Internal polyadenylation sites are usually located in intronic regions which lead to the conversion of an internal exon to a 3' terminal exon. So far, more than half of mammalian genes were found to have at least one internal polyadenylation event that potentially results in an mRNA variant. Moreover, a number of studies have shown that alternative polyadenylation is often coupled to an alternative splicing event (Le Texier *et al.*, 2006; Tian *et al.*, 2007).

We have identified an alternative polyadenylation event in the *ASAH1* gene which is regulated by hnRNP L (Fig. 3.24). Upon knockdown of hnRNP L, use of an internal polyadenylation site in intron 5 of the gene was increased. RT-PCR analysis with an oligo(dT) primer confirmed that this poly(A) site is genuine. Alternative polyadenylation of *ASAH1* results in a shortened mRNA transcript due to conversion of an elongated exon 5 into the 3' terminal exon. Until now the mechanism of poly(A) site selection by hnRNP L is unclear although potential hnRNP L binding motifs were found in the target region (Hung *et al.*, 2008). It remains to be investigated if hnRNP L interacts with the polyadenylation

machinery or if the mechanism of alternative polyadenylation is different from constitutive polyadenylation. Since alternative polyadenylation was shown to be a common mechanism contributing to protein diversity it seems likely that other genes besides *ASAH1* will be identified in further studies.

4.6 HnRNP L represses inclusion of cryptic exons

Recently, it was reported that a CA repeat sequence in the mouse *ITGA2* gene functions as a length-dependent splicing activator mediated by hnRNP L (Cheli & Kunicki, 2006). These findings correlated with our splicing analysis of the human *eNOS* gene where hnRNP L activates splicing by binding to intronic variable-length CA-repeats (Hui *et al.*, 2003b). Sequence analysis of the human *ITGA2* gene revealed that it also contains a CA-rich region similar to the mouse gene (Fig. 3.19). Mutational analysis of *ITGA2* minigene splicing led to the identification of a cryptic exon within intron 1 of the gene (ITGA2a; Fig. 3.20). The presence of the CA-rich sequence within the cryptic exon suppressed its inclusion into the mRNA. Surprisingly, additional *ITGA2* minigene constructs showed differing splicing patterns (Fig. 3.23). Substitution of the CA-rich region in intron 1 resulted in either decreased splicing efficiency (ITGA2b) or inclusion of the cryptic exon (ITGA2c). Why was inclusion of the cryptic exon observed for only one of the two constructs? The constructs differ in parts of their intronic sequence but are of similar size and both contain the potential cryptic exon splice sites (Fig. 3.22). Recognition of the cryptic exon seems to be independent to a certain amount on distance between the two cryptic splice sites but rather depend on the sequence of the cryptic exon. It remains to be investigated which sequence element in ITGA2b prevents recognition of the cryptic exon. Furthermore, it is very unlikely that the cryptic splice sites would be used in the same manner in the natural context since they would be separated by approximately 37 kb. This does not exclude the possibility, however, that other cryptic exons may be recognised within *ITGA2* intron 1 with another combination of cryptic splice sites. Moreover, it was demonstrated in GST-pulldown experiments that hnRNP L directly binds to the CA-rich sequence in *ITGA2* intron 1 with high affinity but not to an unrelated control sequence (data not shown). This finding implied that regulation of *ITGA2* splicing is mediated by hnRNP L binding to the intronic CA-rich element.

Recently, we have shown for two genes, *ARGBP2* and *LIFR*, that hnRNP L suppresses recognition of multiple exons in long introns which represents a novel mechanism of splicing regulation by hnRNP L (Hung *et al.*, 2008). A typical human protein-coding gene contains many sequences that match the splice site consensus but are normally not used (Sironi *et al.*, 2004; Sun & Chasin, 2000). Such splicing signals are referred to as pseudo splice sites. Additional signals are necessary for the splicing machinery to distinguish between real and pseudo splice sites. Part of the pseudo exons may rely on silencer elements for splicing repression. It has been shown that disease-causing mutations can affect pseudo exons by leading to aberrant pseudo exon inclusion (Pagani *et al.*, 2002; Sylvie Tuffery-Giraud, 2003). These findings suggest that hnRNP L may be involved in the repression of pseudoexons, most likely by binding to silencer elements located within the pseudoexon. The results from the mutational analysis of *ITGA2* splicing suggest that this gene also falls into this category. It will be necessary to confirm experimentally that regulation of *ITGA2* splicing is mediated by hnRNP L and furthermore if cryptic exons can be identified in the natural context of *ITGA2* intron 1.

In the case of the *ITGA2b* minigene constructs, the CA-rich sequence functions as a splicing activator in correlation with the mouse data (Cheli & Kunicki, 2006). This finding also corresponds to our findings regarding the splicing regulation of the human *eNOS* gene (Hui *et al.*, 2003b).

4.7 Perspectives

The identification of new hnRNP L target genes led to the discovery of novel modes of hnRNP L-dependent splicing regulation but considering the high abundance of CA-rich regions in the human genome, the target genes we identified so far may only be the tip of the iceberg. New methods are now available which can be used for the detection of hnRNP L target genes in living cells. UV crosslinking and immunoprecipitation (CLIP) in combination with the recently developed high-throughput sequencing methods represent a powerful new tool for the identification of *in vivo* targets of RNA-binding proteins such as hnRNP L (Ule *et al.*, 2005; Wang *et al.*, 2008). The knowledge of further target genes may help us to reveal common principles of splicing regulation mediated by hnRNP L.

Little is known so far about hnRNP L-interacting proteins in connection with splicing regulation. HnRNP L was shown in yeast two-hybrid screens to interact with PTB (hnRNP I), hnRNP D/AUF1, hnRNP E2, and hnRNP K (Hahm *et al.*, 1998a; Kim *et al.*, 2000; Park *et al.*, 2007). The protein also forms homodimers and potentially also heterodimers with hnRNP LL (Kim *et al.*, 2000, and data not shown). Rothrock and co-workers found hnRNP L, E2 and PTB associated with an exonic splicing silencer (ESS1) of the *CD45* gene but only hnRNP L levels were decreased by mutation of the silencer sequence (Rothrock *et al.*, 2005). Cooperative binding of the three proteins to the ESS1 could not be detected but the authors suggested that the interaction between these proteins may increase the overall stability of the ESS1-bound regulatory complex. To investigate if hnRNP L cooperates with other splicing factors and/or with components of the spliceosome machinery will be a major challenge for the better understanding of regulatory networks.

Another main topic in splicing regulation by hnRNP L concerns post-translational modifications. Changes in the phosphorylation status are a common feature for the modulation of the splicing activity of SR proteins (Blaustein *et al.*, 2005; Feng *et al.*, 2008; Hagopian *et al.*, 2008; Patel *et al.*, 2005). However, our knowledge on post-translational modifications of hnRNPs in splicing regulation is currently very limited. For hnRNP A1 it has been shown that cytoplasmic accumulation of the protein upon osmotic shock is due to an increase in its phosphorylation state (Allemand *et al.*, 2005; van der Houven van Oordt *et al.*, 2000). The decrease in nuclear levels of hnRNP A1 led to alternative splicing changes in an E1A minigene reporter.

It will be an interesting task to investigate whether the activity and/or subcellular localisation of hnRNP L is also regulated by post-translational modifications. Moreover, since hnRNP A1 and L are known shuttling proteins between nucleus and cytoplasm and hnRNP L was shown to enhance nucleoplasmic export of intronless mRNAs (Guang *et al.*, 2005; Kim *et al.*, 2000).

5. References

- Abelson, J., Trotta, C. R. & Li, H. (1998).** tRNA Splicing. *J Biol Chem* **273**, 12685-12688.
- Allemand, E., Guil, S., Myers, M., Moscat, J., Cáceres, J. F. & Krainer, A. R. (2005).** Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3605-3610.
- Ast, G. (2004).** How did alternative splicing evolve? *Nat Rev Genet* **5**, 773-782.
- Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G. & Reed, R. (1992).** Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol Cell Biol* **12**, 3165-3175.
- Black, D. L. (2000).** Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology. *Cell* **103**, 367-370.
- Black, D. L. (2003).** Mechanism of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* **72**, 291-336.
- Blaustein, M., Pelisch, F., Tanos, T. & other authors (2005).** Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* **12**, 1037-1044.
- Blencowe, B. J. (2006).** Alternative Splicing: New Insights from Global Analyses. *Cell* **126**, 37-47.
- Brow, D. A. (2002).** Allosteric cascade of spliceosome activation. *Annual Review of Genetics* **36**, 333-360.
- Cartegni, L., Chew, S. L. & Krainer, A. R. (2002).** Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298.
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. (2003).** ESEfinder: a web resource to identify exonic splicing enhancers. *Nucl Acids Res* **31**, 3568-3571.
- Cartegni, L., Hastings, M. L., Calarco, J. A., de Stanchina, E. & Krainer, A. R. (2006).** Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, SMN1 and SMN2. *The American Journal of Human Genetics* **78**, 63-77.
- Cech, T. R. (1990).** Self-Splicing of Group I Introns. *Annual Review of Biochemistry* **59**, 543-568.
- Cheli, Y. & Kunicki, T. J. (2006).** hnRNP L regulates differences in expression of mouse integrin $\alpha 2\beta 1$. *Blood* **107**, 4391-4398.

Cheung, H., Baggerly, K., Tsavachidis, S., Bachinski, L., Neubauer, V., Nixon, T., Aldape, K., Cote, G. & Krahe, R. (2008). Global analysis of aberrant pre-mRNA splicing in glioblastoma using exon expression arrays. *BMC Genomics* **9**, 216.

Clark, T., Schweitzer, A., Chen, T., Staples, M., Lu, G., Wang, H., Williams, A. & Blume, J. (2007). Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biology* **8**, R64.

Colgan, D. F. & Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes & Development* **11**, 2755-2766.

Coulter, L., Landree, M. & Cooper, T. (1997). Identification of a new class of exonic splicing enhancers by in vivo selection [published erratum appears in Mol Cell Biol 1997 Jun;17(6):3468]. *Mol Cell Biol* **17**, 2143-2150.

Das, D., Clark, T. A., Schweitzer, A. & other authors (2007). A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. *Nucl Acids Res* **35**, 4845-4857.

Das, R. & Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA* **5**, 1504-1508.

Deutsch, M. & Long, M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucl Acids Res* **27**, 3219-3228.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. (1993). hnRNP Proteins and the Biogenesis of mRNA. *Annual Review of Biochemistry* **62**, 289-321.

Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**, 195-205.

Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J. P. & Marie, J. (2004). hnRNP A1 and the SR Proteins ASF/SF2 and SC35 Have Antagonistic Functions in Splicing of β -Tropomyosin Exon 6B. *J Biol Chem* **279**, 38249-38259.

Faustino, N. A. & Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes Dev* **17**, 419-437.

Feng, Y., Chen, M. & Manley, J. L. (2008). Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat Struct Mol Biol* **15**, 1040-1048.

Frugier, T., Nicole, S., Cifuentes-Diaz, C. & Melki, J. (2002). The molecular bases of spinal muscular atrophy. *Current Opinion in Genetics & Development* **12**, 294-298.

- Fukumoto, H., Seino, S., Imura, H., Seino, Y., Eddy, R. L., Fukushima, Y., Byers, M. G., Shows, T. B. & Bell, G. I. (1988).** Sequence, tissue distribution, and chromosomal localization of mRNA encoding a human glucose transporter-like protein. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 5434-5438.
- Gardina, P., Clark, T., Shimada, B. & other authors (2006).** Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**, 325.
- Gershon, D. (2002).** Microarray technology: An array of opportunities. *Nature* **416**, 885-891.
- Graveley, B. R. (2000).** Sorting out the complexity of SR protein functions. *RNA* **6**, 1197-1211.
- Green, M. R. (1986).** Pre-mRNA Splicing. *Annual Review of Genetics* **20**, 671-708.
- Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. (2008).** The Vienna RNA Websuite. *Nucl Acids Res* **36 (Web Server Issue)**, W70-W74.
- Guang, S., Felthaus, A. M. & Mertz, J. E. (2005).** Binding of hnRNP L to the Pre-mRNA Processing Enhancer of the Herpes Simplex Virus Thymidine Kinase Gene Enhances both Polyadenylation and Nucleocytoplasmic Export of Intronless mRNAs. *Mol Cell Biol* **25**, 6303-6313.
- Hagopian, J. C., Ma, C. T., Meade, B. R., Albuquerque, C. P., Ngo, J. C., Ghosh, G., Jennings, P. A., Fu, X. D. & Adams, J. A. (2008).** Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1. *J Mol Biol* **382**, 894-909.
- Hahm, B., Cho, O. H., Kim, J.-E., Kim, Y. K., Kim, J. H., Oh, Y. L. & Jang, S. K. (1998a).** Polypyrimidine tract-binding protein interacts with HnRNP L. *FEBS Letters* **425**, 401-406.
- Hahm, B., Kim, Y. K., Kim, J. H., Kim, T. Y. & Jang, S. K. (1998b).** Heterogeneous Nuclear Ribonucleoprotein L Interacts with the 3' Border of the Internal Ribosomal Entry Site of Hepatitis C Virus. *J Virol* **72**, 8782-8788.
- Hastings, M. L. & Krainer, A. R. (2001).** Pre-mRNA splicing in the new millennium. *Current Opinion in Cell Biology* **13**, 302-309.
- Heller, M. J. (2002).** DNA microarray technology: Devices, Systems, and Applications. *Annual Review of Biomedical Engineering* **4**, 129-153.
- House, A. E. & Lynch, K. W. (2006).** An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat Struct Mol Biol* **13**, 937-944.

Hui, J., Reither, G. & Bindereif, A. (2003a). Novel functional role of CA repeats and hnRNP L in RNA stability. *RNA* **9**, 931-936.

Hui, J., Stangl, K., Lane, W. S. & Bindereif, A. (2003b). HnRNP L stimulates splicing of the eNOS gene by binding to variable-length CA repeats. *Nat Struct Biol* **10**, 33-37.

Hui, J., Hung, L. H., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S. A. & Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* **24**, 1988-1998.

Hung, L. H., Heiner, M., Hui, J., Schreiner, S., Benes, V. & Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* **14**, 284-296.

Hwang, B., Lim, J. H., Hahm, B., Jang, S. K. & Lee, S. W. (2008). hnRNP L is required for the translation mediated by HCV IRES. *Biochem Biophys Res Commun*.

Inoue, O., Suzuki-Inoue, K., Dean, W. L., Frampton, J. & Watson, S. P. (2003). Integrin $\alpha 2\beta 1$ mediates outside-in regulation of platelet spreading on collagen through activation of Src kinases and PLC $\gamma 2$. *J Cell Biol* **160**, 769-780.

Janssen, R., Bogardus, C., Takeda, J., Knowler, W. & Thompson, D. (1994). Linkage analysis of acute insulin secretion with GLUT2 and glucokinase in Pima Indians and the identification of a missense mutation in GLUT2. *Diabetes* **43**, 558-563.

Jenkins, J. L., Shen, H., Green, M. R. & Kielkopf, C. L. (2008). Solution Conformation and Thermodynamic Characteristics of RNA Binding by the Splicing Factor U2AF65. *J Biol Chem* **283**, 33641-33649.

Jhavar, S., Reid, A., Clark, J. & other authors (2008). Detection of TMPRSS2-ERG Translocations in Human Prostate Cancer by Expression Profiling Using GeneChip Human Exon 1.0 ST Arrays. *J Mol Diagn* **10**, 50-57.

Jurica, M. S. & Moore, M. J. (2003). Pre-mRNA Splicing: Awash in a Sea of Proteins. *Molecular Cell* **12**, 5-14.

Kim, J. H., Hahm, B., Kim, Y. K., Choi, M. & Jang, S. K. (2000). Protein-protein interaction among hnRNPs shuttling between nucleus and cytoplasm. *Journal of Molecular Biology* **298**, 395-405.

Krämer, A. (1996). The Structure and Function of Proteins Involved in Mammalian Pre-mRNA Splicing. *Annual Review of Biochemistry* **65**, 367-409.

Krämer, A., Keller, W., Appel, B. & Lührmann, R. (1984). The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell* **38**, 299-307.

- Krawczak, M., Reiss, J. & Cooper, D. N. (1992).** The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Human Genetics* **90**, 41-54.
- Kurihara, H., Anderson, J. M. & Farquhar, M. G. (1992).** Diversity among tight junctions in rat kidney: glomerular slit diaphragms and endothelial junctions express only one isoform of the tight junction protein ZO-1. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 7075-7079.
- Kuyumcu-Martinez, N. M., Wang, G.-S. & Cooper, T. A. (2007).** Increased Steady-State Levels of CUGBP1 in Myotonic Dystrophy 1 Are Due to PKC-Mediated Hyperphosphorylation. *Molecular Cell* **28**, 68-78.
- Lam, B. J., Bakshi, A., Ekinci, F. Y., Webb, J., Graveley, B. R. & Hertel, K. J. (2003).** Enhancer-dependent 5'-Splice Site Control of fruitless Pre-mRNA Splicing. *J Biol Chem* **278**, 22740-22747.
- Lander, E. S., Linton, L. M., Birren, B. & authors, o. (2001).** Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- Le Texier, V., Riethoven, J.-J., Kumanduri, V., Gopalakrishnan, C., Lopez, F., Gautheret, D. & Thanaraj, T. A. (2006).** AltTrans: Transcript pattern variants annotated for both alternative splicing and alternative polyadenylation. *BMC Bioinformatics* **7**, 169.
- Lee, C. & Roy, M. (2004).** Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol* **5**, 231.
- Lee, C. & Wang, Q. (2005).** Bioinformatics analysis of alternative splicing. *Brief Bioinform* **6**, 23-33.
- Li, C.-M., Park, J.-H., He, X. & other authors (1999).** The Human Acid Ceramidase Gene (ASAH): Structure, Chromosomal Location, Mutation Analysis, and Expression. *Genomics* **62**, 223-231.
- Liu, X. & Mertz, J. E. (1995).** HnRNP L binds a cis-acting RNA sequence element that enables intron-independent gene expression. *Genes Dev* **9**, 1766-1780.
- Lukong, K. E., Chang, K.-w., Khandjian, E. W. & Richard, S. (2008).** RNA-binding proteins in human genetic disease. *Trends in Genetics* **24**, 416-425.
- Lutz, C. S. (2008).** Alternative Polyadenylation: A Twist on mRNA 3' End Formation. *ACS Chemical Biology* **3**, 609-617.
- Maniatis, T. & Tasic, B. (2002).** Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236-243.
- Matlin, A. J., Clark, F. & Smith, C. W. J. (2005).** Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**, 386-398.

- Matlin, A. J., Southby, J., Gooding, C. & Smith, C. W. J. (2007).** Repression of α -actinin SM exon splicing by assisted binding of PTB to the polypyrimidine tract. *RNA* **13**, 1214-1223.
- Matsutani, A., Hing, A., Steinbrueck, T., Janssen, R., Weber, J., Permutt, M. A. & Donis-Keller, H. (1992).** Mapping the human liver/islet glucose transporter (GLUT2) gene within a genetic linkage map of chromosome 3q using a (CA)_n dinucleotide repeat polymorphism and characterization of the polymorphism in three racial groups. *Genomics* **13**, 495-501.
- McGlinchy, N. J. & Smith, C. W. J. (2008).** Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences* **33**, 385-393.
- McKee, A., Neretti, N., Carvalho, L., Meyer, C., Fox, E., Brodsky, A. & Silver, P. (2007).** Exon expression profiling reveals stimulus-mediated exon use in neural cells. *Genome Biology* **8**, R159.
- McManus, M. T. & Sharp, P. A. (2002).** Gene silencing in mammals by small interfering RNAs. *Nat Rev Genet* **3**, 737-747.
- Meister, G., Eggert, C. & Fischer, U. (2002).** SMN-mediated assembly of RNPs: a complex story. *Trends in Cell Biology* **12**, 472-478.
- Michaud, S. & Reed, R. (1993).** A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* **7**, 1008-1020.
- Modrek, B. & Lee, C. (2002).** A genomic view of alternative splicing. *Nat Genet* **30**, 13-19.
- Nilsen, T. W. (1994).** RNA-RNA interactions in the spliceosome: Unraveling the ties that bind. *Cell* **78**, 1-4.
- Oberdoerffer, S., Moita, L. F., Neems, D., Freitas, R. P., Hacohen, N. & Rao, A. (2008).** Regulation of CD45 Alternative Splicing by Heterogeneous Ribonucleoprotein, hnRNPLL. *Science* **321**, 686-691.
- Okoniewski, M. J. & Miller, C. J. (2008).** Comprehensive Analysis of Affymetrix Exon Arrays Using BioConductor. *PLoS Computational Biology* **4**, e6.
- Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G. W., Yeakley, J. M., Rio, D. C. & Graveley, B. R. (2007).** A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol* **14**, 1134-1140.
- Pacheco, T. R., Coelho, M. B., Desterro, J. M. P., Mollet, I. & Carmo-Fonseca, M. (2006).** In Vivo Requirement of the Small Subunit of U2AF for Recognition of a Weak 3' Splice Site. *Mol Cell Biol* **26**, 8183-8190.
- Pagani, F., Buratti, E., Stuani, C., Bendix, R., Dork, T. & Baralle, F. E. (2002).** A new type of mutation causes a splicing defect in ATM. *Nat Genet* **30**, 426-429.

- Park, H. G., Yoon, J. Y. & Choi, M. (2007).** Heterogeneous nuclear ribonucleoprotein D/AUF1 interacts with heterogeneous nuclear ribonucleoprotein L. *J Biosci* **32**, 1263-1272.
- Patel, N. A., Kaneko, S., Apostolatos, H. S. & other authors (2005).** Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase Cbeta11 alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40. *J Biol Chem* **280**, 14302-14309.
- Perez, I., Lin, C. H., McAfee, J. G. & Patton, J. G. (1997).** Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA* **3**, 764-778.
- Philips, A. V. & Cooper, T. A. (2000).** RNA processing and human disease. *Cellular and Molecular Life Sciences (CMLS)* **57**, 235-249.
- Pinol-Roma, S., Swanson, M., Gall, J. & Dreyfuss, G. (1989).** A novel heterogeneous nuclear RNP protein with a unique distribution on nascent transcripts. *J Cell Biol* **109**, 2575-2587.
- Pozzoli, U., Sironi, M., Cagliani, R., Comi, G. P., Bardoni, A. & Bresolin, N. (2002).** Comparative Analysis of the Human Dystrophin and Utrophin Gene Structures. *Genetics* **160**, 793-798.
- Raghuathan, P. L. & Guthrie, C. (1998).** A Spliceosomal Recycling Factor That Reanneals U4 and U6 Small Nuclear Ribonucleoprotein Particles. *Science* **279**, 857-860.
- Raker, V. A., Hartmuth, K., Kastner, B. & Luhrmann, R. (1999).** Spliceosomal U snRNP Core Assembly: Sm Proteins Assemble onto an Sm Site RNA Nonanucleotide in a Specific and Thermodynamically Stable Manner. *Mol Cell Biol* **19**, 6554-6565.
- Robida, M. D., Rahn, A. & Singh, R. (2007).** Genome-Wide Identification of Alternatively Spliced mRNA Targets of Specific RNA-Binding Proteins. *PLoS ONE* **2**, e520.
- Rosbach, O., Hung, L. H., Schreiner, S., Grishina, I., Heiner, M., Hui, J. & Bindereif, A. (2009).** Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol Cell Biol* **29**, 1442-1451.
- Rothrock, C. R., House, A. E. & Lynch, K. W. (2005).** HnRNP L represses exon splicing via a regulated exonic splicing silencer. *EMBO J* **24**, 2792-2802.
- Ruskin, B., Zamore, P. D. & Green, M. R. (1988).** A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207-219.
- Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J. & Wasserman, S. A. (1996).** Control of Male Sexual Behavior and Sexual Orientation in Drosophila by the fruitless Gene. *Cell* **87**, 1079-1089.

Santer, R., Schneppenheim, R., Dombrowski, A., Gotze, H., Steinmann, B. & Schaub, J. (1997). Mutations in GLUT2, the gene for the liver-type glucose transporter, in patients with Fanconi-Bickel syndrome. *Nat Genet* **17**, 324-326.

Sauliere, J., Sureau, A., Expert-Bezancon, A. & Marie, J. (2006). The Polypyrimidine Tract Binding Protein (PTB) Represses Splicing of Exon 6B from the β -Tropomyosin Pre-mRNA by Directly Interfering with the Binding of the U2AF65 Subunit. *Mol Cell Biol* **26**, 8755-8769.

Schaal, T. D. & Maniatis, T. (1999). Selection and Characterization of Pre-mRNA Splicing Enhancers: Identification of Novel SR Protein-Specific Enhancer Sequences. *Mol Cell Biol* **19**, 1705-1719.

Schellenberg, M. J., Ritchie, D. B. & MacMillan, A. M. (2008). Pre-mRNA splicing: a complex picture in higher definition. *Trends in Biochemical Sciences* **33**, 243-246.

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R. & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucl Acids Res* **34**, 3955-3967.

Shih, S.-C. & Claffey, K. P. (1999). Regulation of Human Vascular Endothelial Growth Factor mRNA Stability in Hypoxia by Heterogeneous Nuclear Ribonucleoprotein L. *J Biol Chem* **274**, 1359-1365.

Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., Giorda, R. & Pozzoli, U. (2004). Silencer elements as possible inhibitors of pseudoexon splicing. *Nucl Acids Res* **32**, 1783-1791.

Smith, P. J., Zhang, C., Wang, J., Chew, S. L., Zhang, M. Q. & Krainer, A. R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* **15**, 2490-2508.

Spellman, R., Rideau, A., Matlin, A. & other authors (2005). Regulation of alternative splicing by PTB and associated factors. *Biochem Soc Trans* **33**, 457-460.

Staley, J. P. & Guthrie, C. (1998). Mechanical Devices of the Spliceosome: Motors, Clocks, Springs, and Things. *Cell* **92**, 315-326.

Stangl, K., Cascorbi, I., Laule, M. & other authors (2000). High CA repeat numbers in intron 13 of the endothelial nitric oxide synthase gene and increased risk of coronary artery disease. *Pharmacogenetics* **10**, 133-140.

Sun, H. & Chasin, L. A. (2000). Multiple Splicing Defects in an Intronic False Exon. *Mol Cell Biol* **20**, 6414-6425.

Sylvie Tuffery-Giraud, C. S., Sylvie Chambert, Mireille Claustres, (2003). Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. *Human Mutation* **21**, 608-614.

- Takada, Y. & Hemler, M. (1989).** The primary structure of the VLA-2/collagen receptor alpha 2 subunit (platelet GPIa): homology to other integrins and the presence of a possible collagen-binding domain. *J Cell Biol* **109**, 397-407.
- Thorens, B., Sarkar, H. K., Kaback, H. R. & Lodish, H. F. (1988).** Cloning and functional expression in bacteria of a novel glucose transporter present in liver, intestine, kidney, and β -pancreatic islet cells. *Cell* **55**, 281-290.
- Thorsen, K., Sorensen, K. D., Brems-Eskildsen, A. S. & other authors (2008).** Alternative Splicing in Colon, Bladder, and Prostate Cancer Identified by Exon Array Analysis. *Mol Cell Proteomics* **7**, 1214-1224.
- Tian, B., Hu, J., Zhang, H. & Lutz, C. S. (2005).** A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucl Acids Res* **33**, 201-212.
- Tian, B., Pan, Z. & Lee, J. Y. (2007).** Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Research* **17**, 156-165.
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. (2008).** Crystal Structure of a Self-Spliced Group II Intron. *Science* **320**, 77-82.
- Topp, J. D., Jackson, J., Melton, A. A. & Lynch, K. W. (2008).** A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA* **14**, 2038-2049.
- Ule, J., Jensen, K. B., Ruggiu, M., Mele, A., Ule, A. & Darnell, R. B. (2003).** CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science* **302**, 1212-1215.
- Ule, J., Jensen, K., Mele, A. & Darnell, R. B. (2005).** CLIP: A method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376-386.
- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J. & Darnell, R. B. (2006).** An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580-586.
- Valadkhan, S. & Manley, J. L. (2001).** Splicing-related catalysis by protein-free snRNAs. *Nature* **413**, 701-707.
- van der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J. & Caceres, J. F. (2000).** The MKK3/6-p38-signaling Cascade Alters the Subcellular Distribution of hnRNP A1 and Modulates Alternative Splicing Regulation. *J Cell Biol* **149**, 307-316.
- Venter, J. C., Adams, M. D., Myers, E. W. & other authors (2001).** The Sequence of the Human Genome. *Science* **291**, 1304-1351.
- Wagner, E. J. & Garcia-Blanco, M. A. (2001).** Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Mol Cell Biol* **21**, 3281-3288.

5. References

- Wang, J., Smith, P. J., Krainer, A. R. & Zhang, M. Q. (2005).** Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucl Acids Res* **33**, 5053-5062.
- Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M. & Burge, C. B. (2004).** Systematic Identification and Analysis of Exonic Splicing Silencers. *Cell* **119**, 831-845.
- Wang, Z. & Burge, C. B. (2008).** Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802-813.
- Wang, Z., Gerstein, M. & Snyder, M. (2008).** RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **advanced online publication**.
- Wassarman, D. A. (1993).** Psoralen crosslinking of small RNAs in vitro. *Molecular Biology Reports* **17**, 143-151.
- Waterston, R. H., Lindblad-Toh, K. & Birney, E. (2002).** Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562.
- Will, C. L. & Lührmann, R. (2001).** Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology* **13**, 290-301.
- Will, C. L. & Lührmann, R. (2005).** Splicing of a rare class of introns by the U12-dependent spliceosome. *Biological Chemistry* **386**, 713.
- Willott, E., Balda, M. S., Heintzelman, M., Jameson, B. & Anderson, J. M. (1992).** Localization and differential expression of two isoforms of the tight junction protein ZO-1. *Am J Physiol Cell Physiol* **262**, C1119-1124.
- Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. J. (2004).** Autoregulation of Polypyrimidine Tract Binding Protein by Alternative Splicing Leading to Nonsense-Mediated Decay. *Molecular Cell* **13**, 91-100.
- Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. (1999).** Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**, 832-835.
- Yeo, G. & Burge, C. B. (2004).** Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology* **11**, 377-394.
- Yeo, G. W., Nostrand, E. L. V. & Liang, T. Y. (2007).** Discovery and Analysis of Evolutionarily Conserved Intronic Splicing Regulatory Elements. *PLoS Genetics* **3**, e85.
- Zamore, P. D. & Green, M. R. (1989).** Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 9243-9247.

Zamore, P. D., Patton, J. G. & Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-614.

Zillmann, M., Zapp, M. L. & Berget, S. M. (1988). Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol* **8**, 814-821.

Zuo, P. & Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev* **10**, 1356-1368.

6. Appendices

Abbreviations

°C	centigrade	hnRNP LL	hnRNP L-like
µg	microgram	HUVEC	human umbilical vein endothelial cell
µl	microliter		
µM	micromolar	IgG	immunoglobulin G
A	adenosine	IPTG	isopropyl-1-thio-β-D-galactoside
aa	amino acid(s)	ISE	intronic splicing enhancer
APS	ammonium persulfate	ISRE	intronic splicing regulatory element
ASAH1	N-acylshingosine amidohydrolase 1	ISS	intronic splicing silencer
ATP	adenosine triphosphate	ITGA2	integrin alpha-2
bp	base pair(s)	kb	kilobasepair
BSA	bovine serum albumin	K _D	dissociation constant
C	cytidine	kDa	kilodalton
Ci	Curie	M	molar
cm	centimeter	mg	milligram
CMV	cytomegalovirus	min	minute(s)
cpm	counts per minute	mJ	millijoule
CTP	cytidine triphosphate	ml	milliliter
DMEM	Dulbecco's Modified Eagle's medium	mM	millimolar
DMPC	dimethyl pyrocarbonate	mmol	millimole
DNA	deoxyribonucleic acid	mRNA	messenger RNA
DNase	deoxyribonuclease	mut	mutant
dNTP	deoxynucleoside triphosphate	N	any nucleotide
DTT	dithiothreitol	ng	nanogram
<i>E.coli</i>	<i>Escherichia coli</i>	Ni-NTA	nickel-nitrilotriacetic acid
e.g.	exempli gratia (=for example)	nm	nanometer
ECL	enhanced chemiluminescence	nM	nanomolar
EDTA	ethylenediamine tetraacetic acid	NMD	nonsense-mediated decay
eNOS	endothelial nitric oxide synthase	NP-40	nonidet P-40
ESE	exonic splicing enhancer	nt	nucleotide(s)
ESS	exonic splicing silencer	NTP	ribonucleoside triphosphate
EST	expressed sequence tag	OD	optical density
<i>et al.</i>	<i>et alii</i> (=and others)	PAGE	polyacrylamide gel electrophoresis
FCS	fetal calf serum	PBS	phosphate-buffered saline
g	gram	PCR	polymerase chain reaction
g	acceleration of gravity	PK	proteinase K
G	guanosine	pmol	picomol
GST	glutathione sulfate transferase	PMSF	phenylmethylsulfonyl fluoride
GTP	guanosine triphosphate	Poly(A)	polyadenylic acid
h	hour(s)	pre-mRNA	precursor messenger RNA
HBS	HEPES-buffered saline	PTB	polypyrimidine tract binding protein
HEPES	N-2-hydroxyethylpiperazine	PTC	premature termination codon
His	histidine	RNA	ribonucleic acid
hnRNA	heterogeneous nuclear RNA	RNAi	RNA interference
hnRNP	heterogeneous nuclear ribonucleoprotein	RNase	ribonuclease
		rpm	rounds per minute

RRM	RNA recognition motif
RT	reverse transcription
RT	reverse transcriptase
s	second(s)
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
SELEX	systematic evolution of ligands by exponential enrichment
SF1	splicing factor 1
siRNA	small interfering RNA
SLC2A2	solute carrier family 2
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SR	serine-arginine-rich
sub	substitution
T	thymidine
TEMED	N, N, N', N'- tetramethylethylenediamine
TJP1	tight junction protein
Tris	tris- hydroxymethylaminomethane
tRNA	transfer RNA
Tween20	polyoxyethylenesorbiten monolaurate
U	uridine
U	unit
U2AF	U2 auxiliary factor
U2AF35	U2AF 35 kDa subunit
U2AF65	U2AF 65 kDa subunit
UTP	uridine triphosphate
UV	ultraviolet
V	volt
v/v	volume per volume
W	watt
w/v	weight per volume
WT	wildtype
Y	pyrimidine base

Curriculum vitae

Name: Heiner
Vorname: Monika
Geburtsdatum: 1. Juli 1980
Geburtsort: Marburg-Wehrda
Staatsangehörigkeit: deutsch
Familienstand: ledig

Schulbildung: 1986-1992 Mittelpunktschule Perftal
1992-1999 Lahntalschule Biedenkopf
Juni 1999 Allgemeine Hochschulreife
Lahntalschule Biedenkopf

Studium: 1999-2004 Diplomstudiengang in Biologie
Justus-Liebig-Universität Gießen
Hauptfächer: Biochemie, Mikrobiologie,
Zellbiologie, Nebenfächer: Chemie, Physik
August 2004 Abschluss: Diplom-Biologin
Justus-Liebig-Universität Gießen, Institut für
Biochemie

Diplomarbeit: Identifizierung und
Funktionsanalyse von intronischen
spleißregulatorischen Sequenzen

Promotion: 2004-2008 Justus-Liebig-Universität Gießen, Institut für
Biochemie
Thema: Alternative splicing regulation in the
human system: Mechanistic studies of
hnRNP L and CA-rich elements

Weiterbildung: 2007 Staatlich anerkannter Grundkurs
„Sicherheit bei Arbeiten in gentechnischen
Anlagen“ zur Bestellung als Projektleiter
oder Beauftragter für die Biologische
Sicherheit
2008 Teilnahme am Media Workshop „Public
Understanding of Science“

- Publications:** Rajan, P., Dalglish, C., Bourgeois, C.F., Heiner, M., Emami, K., Clark, E.L., Bindereif, A., Stevenin, J., Robson, C.N., Leung, H.Y., and Elliott, D.J. (2009). Proteomics identifies an interaction between Sam68 and hnRNP L with a role in alternative splicing. Submitted for publication
- Schindler, S., Heiner, M., Platzer, M., and Szafranski, K. (2009). Comparison of methods for splice variant quantification. Submitted for publication
- Rosbach, O., Hung, L. H., Schreiner, S., Grishina, I., Heiner, M., Hui, J. & Bindereif, A. (2009). Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Mol Cell Biol* 29, 1442-1451.
- Hung, L. H.* , Heiner, M.* , Hui, J.* , Schreiner, S., Benes, V. & Bindereif, A. (2008). Diverse roles of hnRNP L in mammalian mRNA processing: a combined microarray and RNAi analysis. *RNA* 14, 284-296.*joint first authorship
- Hui, J., Hung, L. H., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S. A. & Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J* 24, 1988-1998.
- Meeting abstracts:** Heiner, M. & Bindereif, A. Diverse roles of heterogeneous ribonucleoprotein L: How hnRNP L regulates alternative splicing of target genes. First international EURASNET conference on alternative splicing. Krakow, 2008 (poster presentation)
- Hung, L.H., Schreiner, S., Hui, J., Heiner, M. & Bindereif, A. Analysis of hnRNP L/LL targets in alternative splicing. EURASNET annual meeting. Ile de Bandor, 2007 (poster presentation)
- Heiner, M., Hui, J., Hung, L.H., Schreiner, S. & Bindereif, A. Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. Annual meeting of the RNA society. Banff, 2005 (poster presentation)

Acknowledgements

Zuerst möchte ich mich bei Prof. Dr. Albrecht Bindereif herzlich bedanken für alle Hilfe und Unterstützung.

Vielen Dank auch an Prof. Rainer Renkawitz für seine Arbeit als zweiter Gutachter meiner Dissertation.

Meinen Kollegen Inna Grishina, Lee-Hsueh Hung, Nicolas Jaé, Konstantin Licht, Dr. Zsafia Palfi, Christian Preußner, Marco Preußner, Tanja Rösel, Oliver Roßbach und Silke Schreiner möchte ich für alle Hilfe und Unterstützung danken. Danke an die „Jungs“, dass sie immer für die nötige Ablenkung sorgen. Vielen Dank an Judith, Inna und Tanja für die vielen „lunchtime“ Gespräche. Danke Tanja für das Korrektur lesen der Arbeit und das du immer so ehrlich bist. Ganz besonderen Dank an Silke, die mir immer mit Rat, Tat und Schokolade zur Seite steht.

I would also like to thank my former colleague Dr. Jingy Hui from whom I learned so much during my diploma thesis and at the beginning of my PhD. We still miss you in the lab.

Vielen Dank an die “ehemalige Gießener Runde” Anita, Henriette und ganz besonders Grit, die mich, wenn es nötig ist, wieder aufbauen. Es ist immer schön, wenn man sich auf etwas freuen kann.

Meinen Eltern möchte ich ganz herzlich danken, dass sie mir meine Ausbildung ermöglicht haben.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig durchgeführt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Gießen, den 23.12.2008

Monika Heiner