Structure - function analysis of heterogeneous nuclear ribonucleoproteins L and LL

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
Inna Grishina
aus Tula, Russland

Gießen, Juli 2010

Die vorliegende Arbeit wurde am Institut für Biochemie des Fachbereichs 08 der Justus-Liebig-Universität Gießen in der Zeit von Juli 2007 bis Juli 2010 unter der Leitung von Prof. Dr. Albrecht Bindereif angefertigt.

Dekan:	Prof. Dr. Volkmar Wolters Institut für Tierökologie Justus-Liebig-Universität Gießen
1. Gutachter:	Prof. Dr. Albrecht Bindereif Institut für Biochemie Justus-Liebig-Universität Gießen
2. Gutachter:	Prof. Dr. Michael Niepmann Biochemisches Institut, Fachbereich Medizin Justus-Liebig-Universität Gießen

Contents

Zusammenfassung

Das abundante kernlokalieserte RNA-Bindeprotein hnRNP L (*heterogeneous nuclear ribonucleoprotein L*) erfüllt eine Reihe von verschiedenen Aufgaben, die im Cytoplasma beziehungsweise im Zellkern stattfinden, u.a. Export von Intron-freien mRNAs, IRES-vermittelte Translation, mRNA-Stabilität und Regulation von alternativem Spleißen. Es ist bekannt, dass hnRNP L seine eigene Expression auf der Ebene des alternativen Spleißens reguliert, ein Prozess, der als Autoregulation bezeichnet wird. HnRNP L bindet neben CA-Wiederholungssequenzen auch CA-reiche Sequenzen. Ein bekanntes Paralog von hnRNP L, hnRNP L-like (LL), wird gewebespezifisch exprimiert. Bisher konnte gezeigt werden, dass hnRNP LL in Abhängigkeit von der T-Zell-Aktivierung das alternative Spleißmuster von CD45 reguliert.

Die beiden Proteine, hnRNP L und hnRNP LL, zeigen bei etwa gleicher Größe (L: 558 Aminosäuren vs. LL: 542 Aminosäuren) 58% Übereinstimmung in ihrer Aminosäuresequenz. Beide besitzen vier klassische RNA-Erkennungsmotive, sogenannte RRMs, aber die Glycin-reiche Region von hnRNP L ist in hnRNP LL weniger stark ausgeprägt, und die Prolin-reiche Region fehlt vollständig.

Um die Funktion der einzelnen Domänen beider Proteine im Detail zu untersuchen, habe ich eine Reihe von Proteinvarianten kloniert. Bei einigen wurden Teile der Proteinsequenz entfernt, während bei anderen einzelne oder mehrere Aminosäuren substitutiert wurden.

Zunächst habe ich die Bindungseigenschaften von hnRNP L und hnRNP LL sowie ihrer Varianten an CA-Wiederholungssequenzen bzw. CA-reichen Sequenzen mittels EMSA (*electrophoretic mobility shift assay*) und Filterbindungsexperimenten untersucht. Ich konnte zeigen, dass für hnRNP L eine Kombination von zwei RNA-Bindedomänen, entweder RRMs 1 und 2 oder RRMs 2 und 3, notwendig und ausreichend ist, um CA-Wiederholungssequenzen mit hoher Affinität zu binden. Im Gegensatz dazu werden alle vier RRMs für eine hoch-affine Bindung von CA-reichen Sequenzen benötigt. Im Falle von hnRNP LL sind allerdings für die Bindung sowohl von CA-Wiederholungssequenzen als auch von CA-reichen Sequenzen alle vier RRMs nötig. Mutationsanalysen für hnRNP L haben zudem gezeigt, dass RRM2 hauptverantwortlich für die RNA-Bindungsspezifität sowie Bindungsaffinität ist.

EMSA und Gelfitrationsexperimente zeigen, dass hnRNP L mindestens zwei *high-score* Bindemotive benötigt, welche durch einen kurzen Abschnitt von 7 bis 10 Nukleotiden Länge getrennt sein müssen, um die RNA fest binden zu können.

Des Weiteren wurden die unterschiedlichen hnRNP L Varianten auf ihre regulatorische Aktivität in *in vitro* Spleißkomplementierung getestet. Als Substrat diente ein *SLC2A2* Minigen-Konstrukt, welches in hnRNP L-depletiertem Kernextrakt inkubiert wurde. Durch Zugabe der verschiedenen Proteinvarianten konnte ich zeigen, dass nur das komplette Wildtyp-Protein und keine der verkürzten Mutanten eine Repressorfunktion bei der Regulation des alternativen Spleißens ausübt.

Außerdem konnte gezeigt werden, dass die Interaktion zwischen den RRMs 3 und 4 des hnRNP L Proteins für die Bindung CA-reicher Sequenzen und die Spleißregulation wichtig ist.

Zusammenfassend ist zu sagen, dass aufgrund der vorliegenden Ergebnisse alle vier RRMs von hnRNP L nötig sind, um die volle Repressor-Funktion des Proteins bei der Regulation von alternativem Spleißen zu entfalten, während die beiden N-terminalen RRMs für die stabile Bindung zu CA-repetitiver RNA ausreichend sind.

Summary

The heterogeneous nuclear ribonucleoprotein L (hnRNP L), an abundant nuclear RNA-binding protein, plays both nuclear and cytoplasmic roles in mRNA export of intronless genes, IRES-mediated translation, mRNA stability and alternative splicing regulation. Recently, it was reported that hnRNP L autoregulates its own expression on the level of alternative splicing. HnRNP L protein recognises CA-repeat as well as CA-rich clusters. HnRNP L-like (LL) protein is a paralog of hnRNP L, whose expression is upregulated in a tissue-specific manner. It was shown that hnRNP LL regulates alternative splicing of *CD45* exon 4 upon T-cell activation.

HnRNP L and hnRNP LL share 58% overall amino acid identity and have similar sizes (558 vs. 542 amino acids). Both proteins contain four classical RNA recognition motifs (RRMs). The glycine-rich region of hnRNP L is less pronounced in hnRNP LL, and the proline-rich region of hnRNP L is absent in hnRNP LL.

To investigate the role of individual domains in hnRNP L and hnRNP LL protein function, I created a series of deletion derivatives and mutants with one or several amino acid substitutions in individual RNA-binding domains.

First, I analysed the RNA-binding properties of the full-length and deletion constructs of hnRNP L and LL by EMSA (electrophoretic mobility shift assay) and filter binding assay. Two substrates were used: CA-repeat and CA-rich RNAs. I demonstrated that the combination of two RNA-binding domains of hnRNP L (RRMs 1/2 and 2/3) is both necessary and sufficient for high-affinity binding to CA-repeat RNA. In contrast, the high-affinity binding of hnRNP L to CA-rich RNA requires all four RRMs. In the case of hnRNP LL all four RRM domains are required for high-affinity binding to both substrates. Mutational analysis revealed that hnRNP L RRM2 is a major determinant for RNA-binding specificity and affinity.

EMSA in combination with gel filtration of protein-RNA complexes indicated that hnRNP L requires at least two high-score binding motifs, separated by a short spacer (7-10 nucleotides) to bind tightly to the RNA.

Second, hnRNP L mutant derivatives were tested for alternative splicing activity, using a *SLC2A2* minigene construct and hnRNP L depleted nuclear extract. I demonstrated that only full-length protein and not the truncated mutant proteins could function as a repressor in regulation of alternative splicing.

In addition, a specific role of inter-domain interaction between RRMs 3 and 4 in CA-rich RNA binding and function of hnRNP L as a splicing repressor was uncovered.

In sum, my results suggest that the presence of all four RRMs is essential for splicing repressor activity of hnRNP L, whereas the two N-terminal RRMs are sufficient for tight association with CA-repeat RNA.

1. Introduction

The studying of several eukaryotic genomes demonstrated that the large proteomic complexity is achieved with a limited number of protein-coding genes. These findings reveal the importance of post-transcriptional mechanisms in generation of protein diversity. Eukaryotic messenger RNA (mRNA) undergoes a series of processing events: capping of the 5' end, polyadenylation of the 3' end, splicing, and editing of individual nucleotides in the RNA.

Splicing plays one of the most important roles in the generation of protein isoforms from a limited number of genes. Alternative splicing is the inclusion of alternative exons or introns from the pre-mRNAs into the mature mRNA. A recent study of the human genome reveals that 95% of all multi-exon genes undergo alternative splicing (Wang *et al*., 2008). Therefore, alternative splicing allows the existence of large proteomic complexity based on a limited number of genes.

The correct post-transcriptional RNA processing steps are regulated by RNA-binding proteins (RBPs). Eukaryotic cells encode a large number of RBPs (thousands in vertebrates), and each protein has unique RNA-binding activity and protein-protein interaction characteristics (Glisovic *et al*., 2008). The activity of RBPs is mediated by a relatively small number of RNA-binding scaffolds whose properties are further modulated by auxiliary domains.

1.1 Splicing of pre-mRNA

Eukaryotic genes are composed of non-coding sequences (introns) and coding regions (exons). The average human gene contains 8 introns with an average length of 3,4 kb, interspersed by exons that average less than 300 bp in length (Sakharkar *et al*., 2004). The largest known gene is the human dystrophin gene, which has 79 exons spanning at least 2,300 kilobases (kb) (Pozzoli *et al*., 2002). Splicing is a process of removing the introns from pre-mRNA. One of the main challenges during pre-mRNA splicing is the reliable determination of the exon/intron boundaries.

There are several conserved intronic sequence elements essential for exon definition: 5' splice site (AG/GURAGU), 3' splice site (CAG/G), branchpoint sequence (YNYURAC), which is typically located within 30 to 50 nucleotides upstream of the 3' splice site, and an immediately adjacent stretch of pyrimidines

termed polypyrimidine tract (Green *et al*., 1986; Lim and Burge, 2001; Sheth *et al*., 2006) (Figure 1.1).

Introns with GT-AG splice sites are called U2-type or major introns. A novel class of eukaryotic nuclear pre-mRNA introns was found on the basis of their unusual splice sites (Hall and Padgett, 1994). These introns contain AT and AC dinucleotides at the 5’ and 3’ splice sites, respectively. This type of introns was named U12-type or minor introns. U12 introns are recognised by a different spliceosome and excised through identical chemical pathway (Hall and Padgett, 1996; Tarn and Steitz, 1996a, 1996b; Tarn and Steitz, 1997; König *et al*., 2007). U12 introns occur in the total population of introns at a frequency of about 1/5000 to 1/10000 (Burge *et al*., 1998; Levine and Durbin, 2001). All known examples of U12-dependent introns occur in genes containing multiple U2-dependent introns.

Chemically, splicing proceeds via two successive transesterification reactions (Fig. 1.1). The branch point A residue plays a critical role in the enzymatic reaction. The first step is a nucleophilic attack. The 2' hydroxyl group of the conserved adenosine within the branching site attacks the the 5' splice site at the exon1-intron junction. An unusual 2'-5' phosphodiester bond is made between both residues and the exon1-intron junction is cleaved. The products are a 2'-5' phosphodiester RNA lariat structure and a free 3'-OH (leaving group) at the upstream exon. The second step is another nucleophilic attack. The 3'-OH end of the released exon1 then attacks the scissile phosphodiester bond of the conserved guanosine of the 3' splice site at the intron-exon2 junction. This reaction liberates the 3'-OH of the intron resulting in a free lariat and spliced exons. The two exon sequences are joined together, while the intron sequence is released as a lariat structure (Moore and Sharp, 1993). The spliceosome, which catalyses the two transesterification steps, is described in more detail in the next chapter.

1.2 Spliceosome assembly

The spliceosome is a macromolecular machinery that catalyses the removal of introns from eukaryotic pre-mRNA (Staley and Guthrie, 1998). It is formed from five small nuclear ribonucleoprotein particles (U snRNPs) together with an additional group of spliceosome-associated splicing factors. This complex is highly dynamic and changes its structure and composition during splicing.

Each U snRNP particle consist of a U snRNA (uridine-rich small nuclear RNA) complexes with a common set of 7 core proteins (Sm or Sm-like (LSm) proteins)

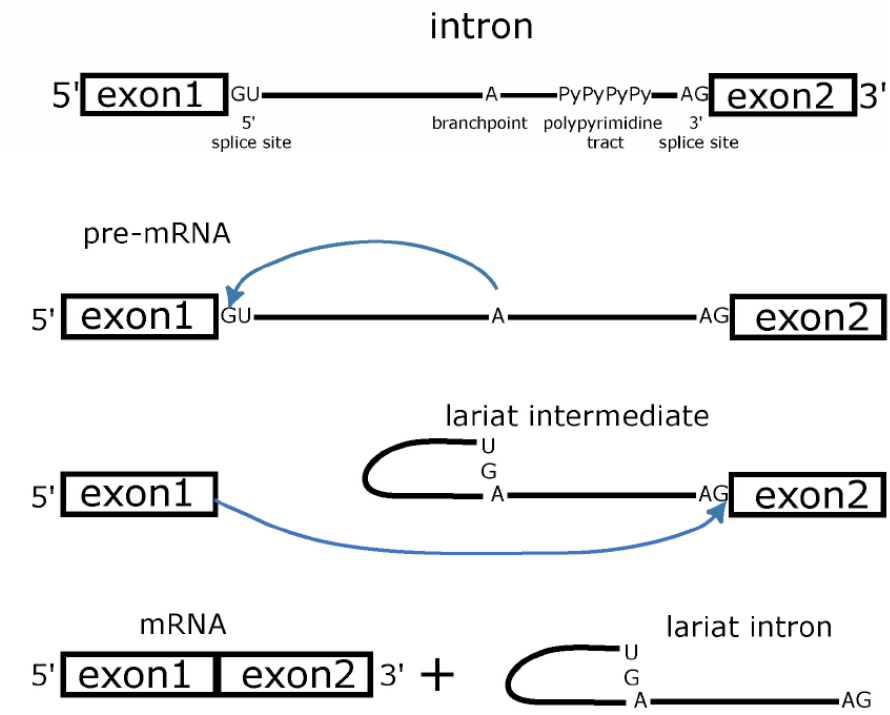


Figure 1.1 Two-step chemical mechanism of pre-mRNA splicing.
Exons are depicted as boxes, intron as a black line. The branch point adenosine (A), the 5' and 3' splice sites and polypyrimidine tract are indicated. The first and second steps of transesterification reaction proceed from top to bottom showing the end products (ligated exons and released lariat intron) at the bottom panel.

and a set of specific proteins unique for each snRNP. U1, U2, U4, U5 and U6 are the major spliceosomal snRNPs. In addition, the human spliceosome contains numerous non-snRNP proteins. Initial mass spectrometric analysis of mixed population of affinity-purified spliceosomal complexes indicated that between 150 (Zhou *et al*., 2002) and 300 distinct proteins (Rappsilber *et al*., 2002) co-purify with the spliceosome.

The spliceosome assembles *de novo* on each intron through a sequential and highly coordinated pathway (Fig. 1.2). Initiation of pre-spliceosome assembly or E complex formation begins with ATP-independent binding of the U1 snRNP through base-pairing interactions of the 5' end of the U1 snRNA with the 5' splice site of the intron. This interaction in higher eukaryotes is stabilised by members of the serine-arginine-rich (SR) protein family and proteins of the U1 snRNP (Valcarcel and Green, 1996). In addition to the U1-5' splice site interaction, the earliest assembly phase of the spliceosome also involves the binding of the SF1/BBP protein and the U2 auxiliary factor (U2AF) to the BPS and the polypyrimidine tract just downstream of the BPS, respectively (Black, 2003). The U2 snRNP is then recruited to the branch point sequence via base-pairing between the pre-mRNA and the U2 snRNA

in an ATP-dependent manner, generating the A complex. The B complex is formed by the recruitment of the U4/U6•U5 tri-snRNP to the pre-spliceosome. The spliceosome becomes catalytically active upon rearrangement and destabilisation of U1 and U4 snRNPs, which leads to the recognition of part of the 5' splice site by U6 snRNA (Staley and Guthrie, 1998; Nilsen, 2003). The activated spliceosome then carries out the first catalytic step of splicing, generating the C complex. Prior to the second catalytic step, additional rearrangements occur in the spliceosomal RNP network (Konarska *et al*., 2006). After the second catalytic step, the spliceosome dissociates, releasing the mRNA. It also releases the U2, U5, and U6 snRNPs to be recycled for new rounds of splicing.

There are two distinct types of spliceosomes, the major (or U2-type) spliceosome and the minor (or U12-type) spliceosome. The spliceosome discussed above is the major spliceosome.

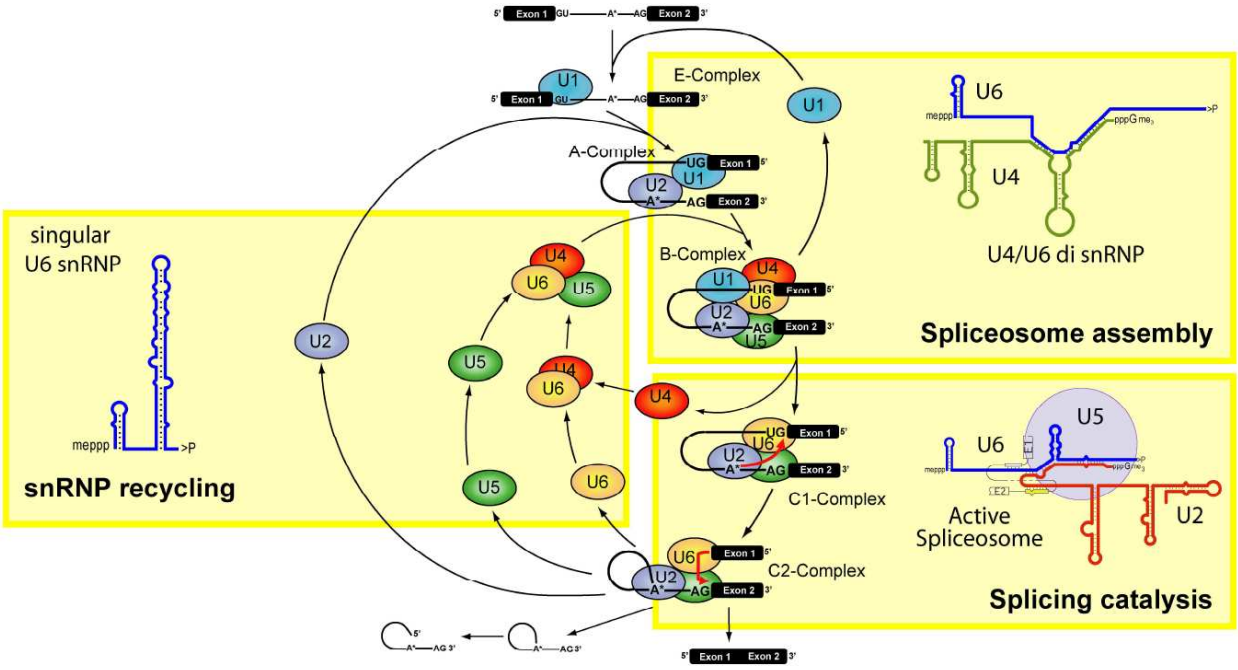


Figure 1.2 Spliceosome assembly cycle.
Pre-mRNA with two exons and one intron (black boxes and black line, respectively) is shown on the top. The branch point adenosine (A*), the 5' (GU) and 3' (AG) splice sites are indicated. Spliceosomal U1, U2, U4, U5 and U6 snRNPs (coloured cycles) associate with pre-mRNA in a stepwise manner. Three main steps: spliceosome assembly, splicing catalysis and snRNP recycling are marked by yellow boxes. Secondary structures of U6 snRNP, U4/U6 di-snRNP and base-pairing of U2/U6 in the active spliceosome are indicated in the yellow boxes. The two transesterification reactions taking place in catalytic center of the spliceosome are indicated by red arrows. After splicing and spliceosome dissociation snRNPs take part in a new round of the splicing cycle. (provided by Dr. J.Medenbach)

The two types of the spliceosome share the U5 snRNP, but each has four other specific snRNPs: U1, U2 and U4/U6 in the U2-type spliceosome, and their low-abundant functional analogues, namely U11, U12 and U4atac/U6atac in the U12-type spliceosome (Tarn and Steitz, 1997; Burge *et al*., 1998). Each snRNA differs from its analogues in the primary sequence but they share a remarkable similarity in the secondary structure (Tarn and Steitz, 1997). Moreover, the two spliceosomes contain a large common set of protein components, and the intricate network of the RNA–RNA interactions is strikingly similar in each spliceosome (Will and Luhrmann, 2005; Patel *et al*., 2003). Nevertheless, the individual spliceosomes can only catalyze the removal of their cognate introns.

1.3 Alternative splicing

Sequencing of the human genome a few years ago revealed a great surprise. Instead of supporting the expected number of 100,000-140,000 genes, nowadays only around 23,000 genes are assumed (MCkernan *et al*., 2009). This number is not much bigger compared to the primitive nematode *C.elegans* (19,000 genes). However, the diversity of the human proteins is clearly more complex compared to invertebrates. Alternative splicing is a crucial mechanism, allowing individual genes to express multiple mRNAs that encode proteins with diverse and even antagonistic function (Fig. 1.3). Notable example is the *Drosophila Dscam* gene, which codes for a cell surface protein involved in neuronal connectivity. The combinatorial alternative splicing of the exons can potentially generate up to 38,000 different protein isoforms, more then twice the number of genes in entire *Drosophila* genome (Graveley *et al*., 2001). Bioinformatics analysis based on aligning ESTs to the genomic sequence shows that 59% of 245 genes present on chromosome 22 are alternatively spliced (Modrek *et al*., 2001). The high frequency of alternative splicing in humans is also supported by EST based database analysis, indicating that 35–60% of all human gene products are alternatively spliced (Mironov *et al*., 1999; Kan *et al*., 2001; Modrek *et al*., 2001). DNA microarray experiments indicate that 95% of all human genes are alternatively spliced (Wang *et al*., 2008).

Alternative splicing in protein-coding regions generates segments of mRNA variability that can insert or remove amino acids, shift the reading frame, or introduce the termination codon (Fig. 1.3). Thus, changes in the protein sequence can influence almost all aspects of protein function, such as binding properties, enzymatic activity, intracellular localisation, phosphorylation and glycosylation patterns. Alternative splicing in 5’ and 3’ untranslated regions (UTRs) affects gene

expression by removing or inserting regulator elements controlling translation, mRNA stability, or localisation.

Some protein isoforms might be generated by alternative splicing in a tissue- or developmental stage-specific manner. In this case, alternative splicing undergoes cell-specific regulation in which splicing pathway are modulated according to cell type, developmental stage, gender or in response to external stimuli.

There are five major patterns into which alternative splicing events can be classified: cassette exons, mutually exclusive exons, retained introns, the use of alternative 5' splice sites and alternative 3' splice sites (Fig. 1.4). Recently, it has been found that the usage of alternative promoters and poly (A) sites further contributes to isoform diversity.

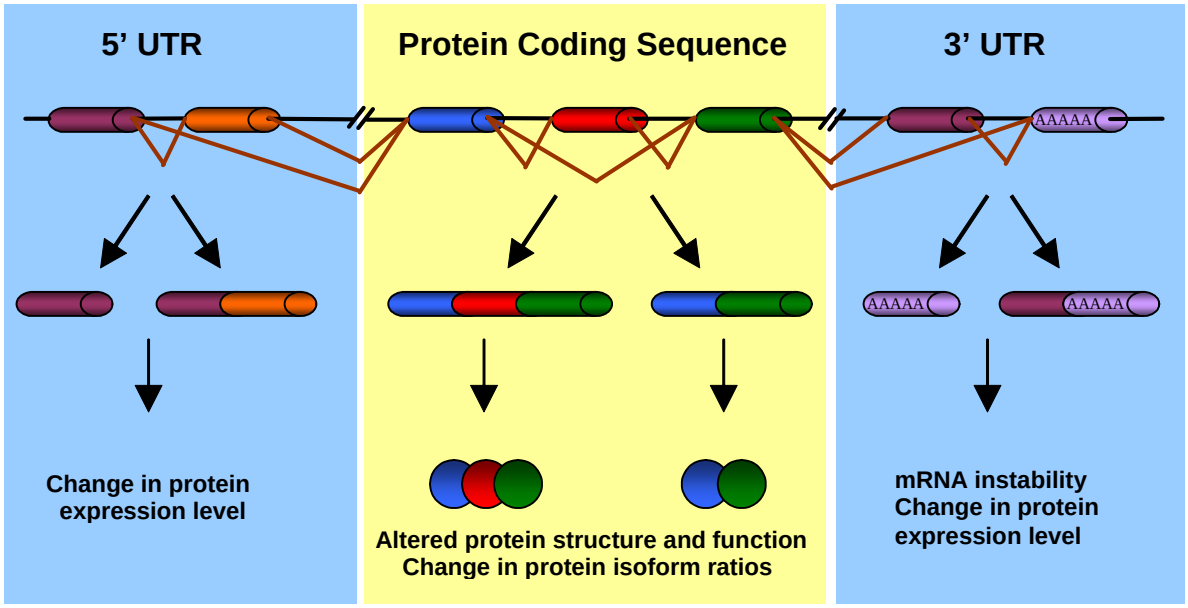


Figure 1.3 The impact of alternative RNA splicing.
Alternative splicing can occur in any region of nascent mRNA. The 5'UTR sequence contains regulatory regions that control protein expression. Insertion or deletion of these regions will have a consequence on protein expression. The 3'UTR region contains mRNA stability domains. Insertion or deletion of these domains will have a consequence on mRNA stability and therefore protein expression. Alternative splicing within the protein coding sequence results in altered protein structure and function. Exons and introns are represented by coloured cylinders and black lines, respectively. Alternative splicing patterns are indicated by brown lines.

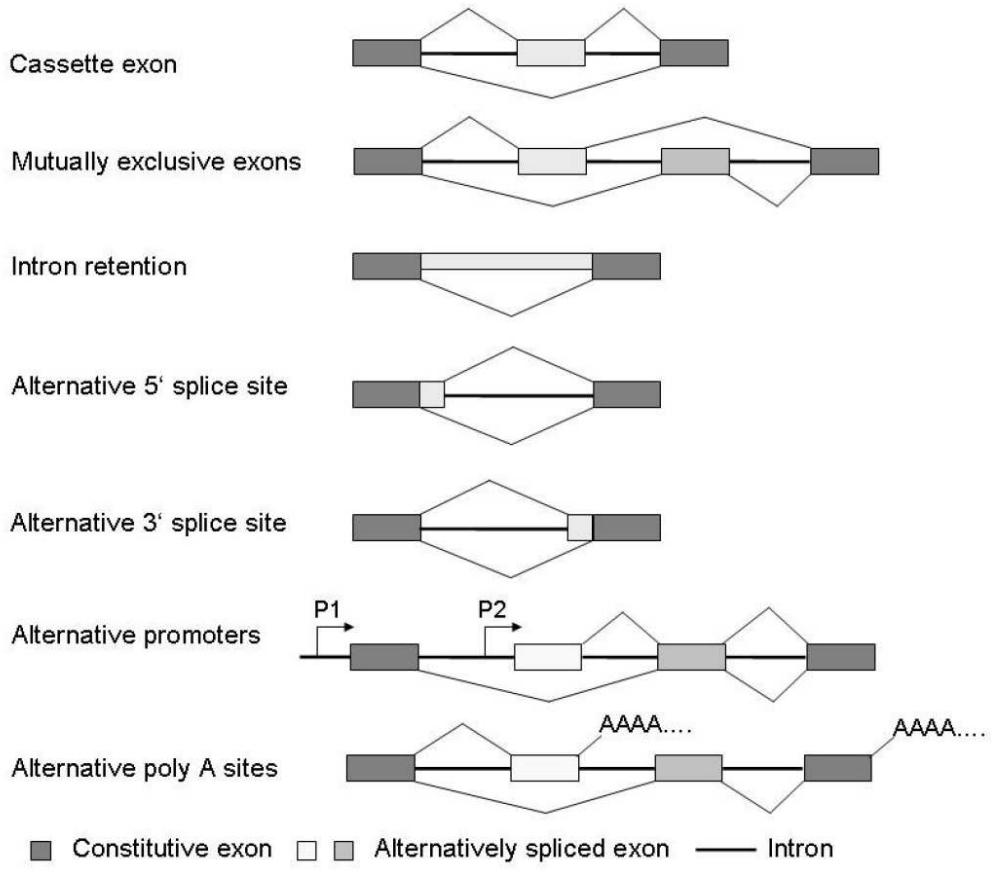


Figure 1.4 Patterns of alternative splicing.
The exons are indicated as boxes. Introns are indicated as thick black lines. Alternative splicing patterns are indicated as thin lines. P1/P2: alternative promoters; AAAA…: poly A sites.

1.4 Mechanism of splicing regulation

1.4.1 Splicing enhancers and silencers

The most remarkable features of the mammalian pre-mRNA splicing machinery are its ability to select precisely correct pairs of splice sites, reliably distinguish authentic exons from pseudo-exons, and modulate the selection of alternative splice sites. The weakly conserved consensus sequences marking exon/intron borders are loosely defined, very short and degenerated, and provide less than 50% of the necessary information for accurate removal of introns in human (Lim and Burge, 2001). The spliceosome machinery must rely on additional auxiliary sequence features. Additional RNA sequence elements (*cis*-acting elements)

provide the necessary information for either activation or repression of constitutive and alternative splicing. Motifs that promote splicing are called enhancers, while those that inhibit splicing are named silencers. Depending on the position and function of the *cis*-regulatory elements, they are divided into four groups: exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE) and intronic splicing silencer (ISS). These elements have a loosely defined, very short (6-8nt), degenerate and partially overlapping consensus sequences. Several types of enhancer/silencer elements have been identified by functional SELEX (Systematic Evolution of Ligands by EXponential enrichment) experiments *in vivo* and *in vitro* (reviewed in Zheng, 2004): the purine-rich and non-purine-rich sequences, and adenosine/cytosine-rich elements (ACEs) (Coulter *et al*., 1997). Recent global studies have discovered that the relative enrichment in exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs) helps distinguishing between authentic and pseudoexons (Zhang and Chasin 2004; Zhang *et al*., 2005; Wang *et al*., 2004).

The exonic splicing enhancers (ESEs) are the most studied. They are common in both alternative and constitutive exons. ESEs are specific short nucleotide sequences that are targeted essentially by SR proteins which then promote exon definition (Blenowe, 2000). Conversely, the exonic splicing silencers (ESSs) help the spliceosome to ignore pseudoexons and decoy splice sites. They act as binding sites for proteins promoting exon exclusion (mainly hnRNP proteins) (Zhu *et al*., 2001). Besides their role in constitutive splicing, ESEs and ESSs play an important role in the regulation of alternative splicing (Black, 2003). The mutation in ESEs can result in skipping the mutant exon, with dramatic effect on the structure of gene product (Cartegni *et al*., 2002).

Intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs) are intronic *cis*-elements that play similar roles as ESEs and ESSs. Recently, intronic CA-rich and CA-repeat elements were describe as *cis*-regulatory elements that play an important role in regulation of alternative splicing (Hui *et al*., 2005; Hung *et al*., 2008; Heiner *et al*., 2010).

Splicing enhancers are located close to the splice sites that they activate. However, the action of splicing enhancers is position dependent. Changing the location of splicing enhancers alters their dependence on particular *trans*-acting factors (Tian and Maniatis, 1994), and determines whether they activate 5’ or 3’ splice sites (Heinrichs *et al*., 1998). It can even transform them into negative regulatory elements (Kanopka *et al*., 1996).

Proteins binding to enhancer or silencer sequences and modulating the alternative splice site selection can be subdivided into two major groups: members of the SR

family of proteins and hnRNPs. These groups are described in more detail in the next chapters.

1.4.2 SR and SR-related proteins

SR proteins are a family of phylogenetically conserved, structurally related, essential pre-mRNA splicing factors. Members of the SR family have a modular structure consisting of one or two copies of an N-terminal RNA recognition motifs [RRMs or so-called RNA-binding domain (RBD)] followed by a variable-length C-terminal domain enriched in serine and arginine dipeptides, known as the RS domain (Fig. 1.5A). The RRMs determine RNA-binding specificity, whereas the RS domain functions as a protein–protein interaction module by recruiting components of the core splicing apparatus to promote splice site pairing (Wu and Maniatis, 1993). More recently, it has also been demonstrated that the RS domain can directly contact both the pre-mRNA branch point and the 5' splice site (Shen *et al*., 2004). The extensive serine phosphorylation of the RS domain plays an important role in regulating the activities and localisation of SR proteins (Sanford and Longman, 2003).

SR-related proteins are a group of additional splicing factors containing RS domains and structurally and functionally related to SR family proteins (Fig. 1.5B) (Fu, 1995, Blencowe *et al*., 1999). Some SR-related proteins, such as U2AF65, U2AF35, U1-70K are directly involved in constitutive and alternative splicing. Interactions between SR and SR-related proteins allow mediating communication between the 5′ and 3′ splice sites during the early stages of spliceosome assembly (Blencowe, 2000).

The presence of different SR proteins in mammalian cells suggests that they regulate RNA splicing specifically. The RNA-binding specificity of each SR protein was identified by functional SELEX approach. This approach led to identification of both purine- and non-purine-rich SR protein-specific ESEs (Schaal and Maniatis, 1999). Those studies also established that SR proteins are sequence-specific RNA-binding proteins with distinct RNA-binding specificity.

The activity of SR proteins in constitutive and alternative splicing regulation relies principally on their interactions with RNA regulatory sequences. SR-protein-binding sites within exons (ESEs) exert a positive effect on splice site selection, whereas intronic binding sites for SR proteins repress splice site selection (Kanopka *et al*., 1996).

For example, SR proteins, bound to ESEs, recruit and stabilise the binding of U1 snRNP and U2AF to the 5' and 3' splice sites, respectively, in a process known as exon definition (Fig. 1.6A) (Boukis *et al*., 2004). SR proteins have been shown to facilitate the recruitment of the U4/U6•U5 tri-snRNP to the pre-spliceosome (Fig. 1.6B) (Roscigno and Carcia-Blanco, 1995). The binding of SR proteins to intronic *cis*-element prevents binding of U2AF65 and U2AF35 to the 3' splice site (Fig. 1.6C).

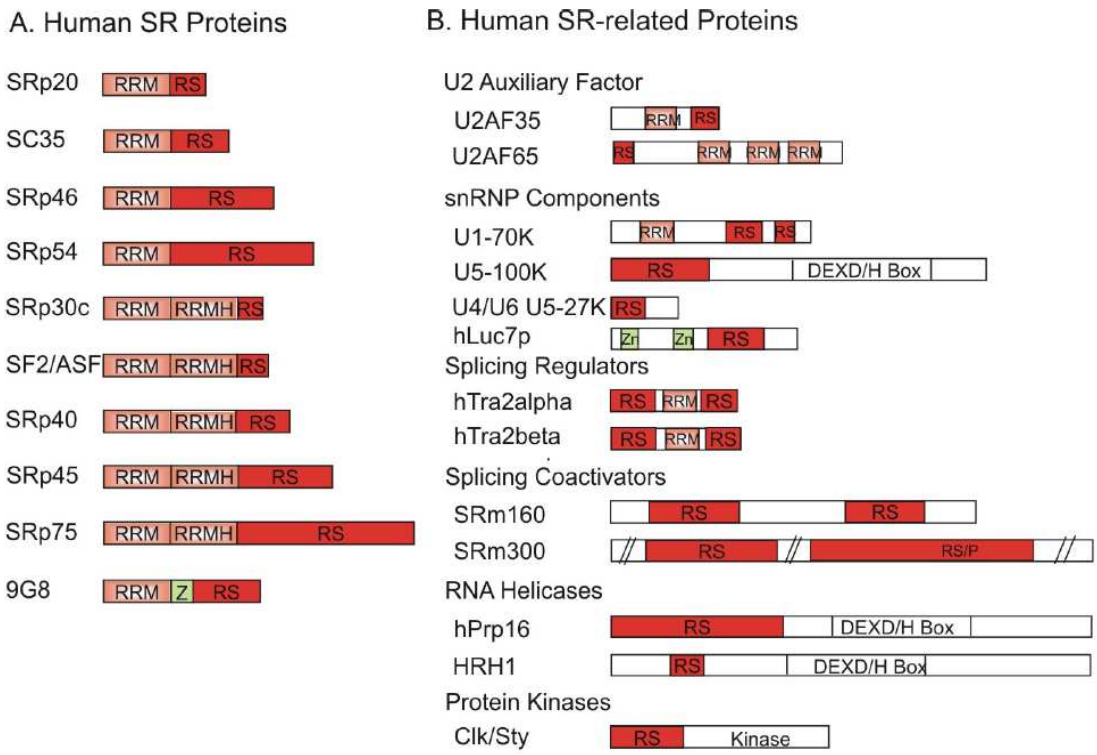


Figure 1.5 Schematic diagram of human SR and SR-related proteins.
Domain structures of known members of the human SR protein family **(A)** and SR-related proteins **(B)** are depicted. RRM (RNA recognition motif) and RRMH (RRM homology) are shown as light red boxes, Zinc knuckle (Z) as a green box, RS domain as a dark red box, DEXD/H Box: motif characteristic for RNA helicases (adapted from Graveley *et al*., 2000).

SR proteins are key players in the regulation of alternative splicing. Many alternative exons carry weak splicing signals; for instance, many alternative 3' splice site contain poor polypyrimidine tracts, which are recognised inefficiently by U2AF, resulting in exon skipping. However, SR proteins, bound to ESEs, can compensate for a weak polypyrimidine tract by recruiting U2AF (Wang and Smith, 2005). Alternatively, ESE-associated SR proteins may promote alternative exon inclusion by antagonising the negative activity of hnRNPs (heterogeneous nuclear ribonucleoproteins), such as hnRNP A1, bound to exonic splicing silencer elements (Fig. 1.6D) (Blencowe, 2000).

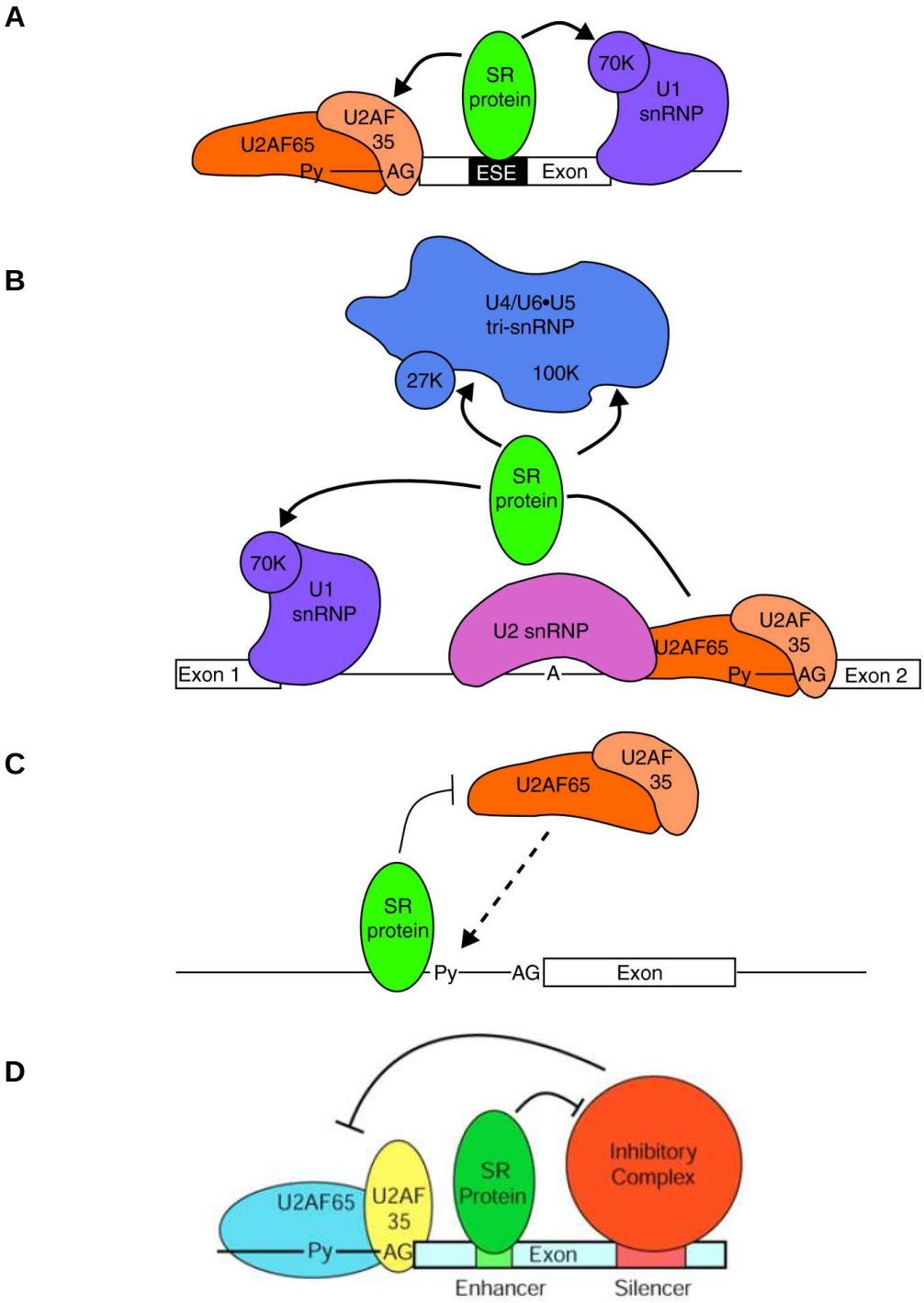


Figure 1.6 Splicing functions of SR proteins.
(A) SR proteins (green), bound to an ESE, may function in constitutive splicing by interacting with the splicing factors U2AF bound at the upstream 3' splice site and U1 snRNP bound to the downstream 5' splice site. Py represents the polypyrimidine tract, the binding site for U2AF. **(B)** Two exon-independent functions of SR proteins. SR proteins facilitate splice-site pairing by simultaneously interacting with U1 snRNP and U2AF across the intron. SR proteins also assist in recruiting the U4/U6U5 tri-snRNP. **(C)** Splicing repression is mediated when SR proteins associate with intronic sequences close to the splice sites (adapted from Shepard and Hertel, 2009). **(D)** SR protein, bound to an ESE, antagonises the negative effect on splicing of an inhibitory protein bound to a juxtaposed ESS.

1.4.3 HnRNP proteins

The nuclear protein-coding transcripts that are produced by RNA polymerase II are termed pre-mRNAs or hnRNAs (heterogeneous nuclear RNA, historical term that described their size heterogeneity and cellular location). Nascent pre-mRNAs are associated with specific proteins with which they form complexes termed heterogeneous nuclear ribonucleoprotein (hnRNP) particles. The collective term for the proteins that bind hnRNAs is hnRNP proteins (Dreyfuss *et al*., 1993, Krecic and Swanson, 1999). HnRNP proteins are a diverse family of abundant nuclear RNA-binding proteins which are involved in many RNA-related biological processes such as transcriptional regulation, telomere-length maintenance, pre-mRNA processing, mRNA stability and export to cytoplasm, and translation (Kim *et al*., 2000). Approximately, 20 major hnRNPs with molecular mass range from 34 kDa to 120 kDa have been identified, and are designated from hnRNP A1 to U (Dreyfuss *et al*., 2003). All hnRNP proteins share a common structure containing various types of RNA-binding domains (RRM, KH, RGG) and auxiliary domains (rich in glycine and other amino acids), which might mediate protein-protein interaction or act in a protein localisation. Many of the arginine residues within the RGG box are potential methylation sites, which could play a role in regulation of the RNA-binding activity. Arginine methylation and serine/threonine phosphorylation are common modifications for several hnRNP proteins (Liu and Dreyfuss, 1995; Schutt *et al*., 2001).

Initially it was thought that hnRNP proteins bind non-specifically to hnRNA. Later the binding specificities for most of hnRNP proteins were identified by SELEX and CLIP assay. For example, the sequence GGGA is recognised by all hnRNP H family proteins (Caputi and Zahler, 2001), hnRNP L protein recognises CACA motifs (Hui *et al*., 2005), and hnRNP I (PTB) protein specifically binds to UCUU motif (Perez *et al*., 1997).

HnRNP proteins, similar as SR proteins, participate in the regulation of alternative splicing. In contrast to SR proteins, most hnRNPs tend to interact with splicing silencer sequences and function as splicing repressors (Matlin *et al*., 2005). Some hnRNP proteins prevent binding of other splicing factors to the splicing control elements within pre-mRNA by forming complexes with RNA and/or other proteins (Fig. 1.7A) (Zhu *et al*., 2001). HnRNP proteins also mediate long-range interactions between distant RNA regions flanking alternative exons, thus looping out the intervening region of the pre-mRNA and preventing splicing of the excluded RNA region (Fig. 1.7B) (Blanchette and Chabot, 1999; Martinez-Contreras *et al*., 2006). In several cases, hnRNP proteins have been found to antagonise, both *in vitro* and

in vivo, the activity of SR proteins (Mayeda *et al*., 1993; Zhu *et al*., 2001). Several proteins, such as hnRNP L, hnRNP LL, hnRNP F and hnRNP H have been shown to act either as a repressor or an activator depending on the location of their binding sites (Hui *et al*., 2005; Mauger *et al*., 2008).

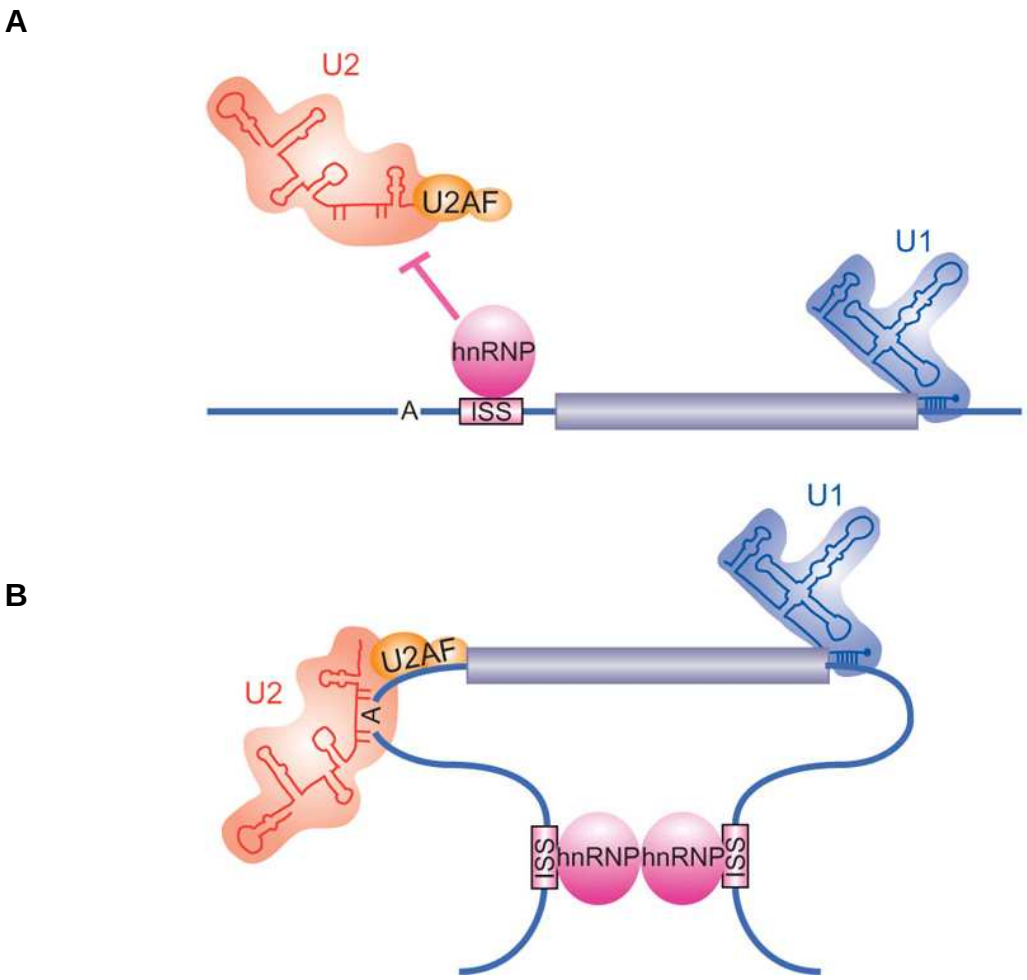


Figure 1.7 Splicing functions of hnRNP proteins.
(A) Splicing repressors such as hnRNP proteins can inhibit splicing by binding to intronic splicing silencers (ISSs), where they interfere with the binding of U2AF to the 3' splice site. **(B)** Alternatively, hnRNP proteins can bind to ISSs in the introns flanking an exon, where they engage in protein-protein interactions and loop out the intervening exon (adapted from Graveley, 2009).

1.5 RNA-binding motifs

RNA-binding proteins play key roles in post-transcriptional control of pre-mRNA, which can occur at many different steps in RNA metabolism, including splicing, polyadenylation, mRNA stability, mRNA localisation and translation. The diversity of

function of RNA-binding proteins would suggest a correspondingly large diversity in the structures that are responsible for RNA recognition. However, most RNA-binding proteins are built from few RNA-binding domains. Some well-characterised RNA-binding domains include the following: RNA-binding domain (RBD, also known as RNA recognition motif, RRM); K-homology (KH) domain; RGG (Arg-Gly-Gly) box; Zinc finger (ZnF, mostly C-x8-C-x5-C-x3-H type); cold-shock domain (Y-box proteins); DEAD/DEAH box; Pumilio/FBF (PUF) domain; double-stranded RNA-binding domain (dsRBD); Piwi/Argonaute/Zwille (PAZ) domain; Sm domain, etc.
Many RBPs have one or more copies of the same RBD while others have two or more distinct domains. Individual RBDs are separated by linker sequences of highly variable length. These linkers provide a critical determinant of binding affinity. In many cases, individual RBDs within the same protein have different binding specificities, which suggest they may allow a single protein to bridge multiple RNAs (*trans*) or multiple RBDs may interact with non-specific RNA lattice to increase binding affinity (*cis*) (Shamoo *et al.*, 1995).

1.5.1 RNA recognition motif

The RNA recognition motif, also known as RNA-binding domain or ribonucleoprotein domain (RNP) is one of the most abundant protein domains in eukaryotes. In human, 497 proteins containing at least one RRM have been identified (Venter, 2001). In eukaryotic proteins, RRMs are often present as multiple copies within a protein (44%, two to six RRMs) and/or together with other domains (21%) (Maris *et al*., 2005). RRM-domain-containing proteins are involved in many cellular functions, particularly messenger RNA and ribosomal RNA processing, splicing and translation regulation, RNA export and RNA stability (Dreyfuss *et al*., 2002).
Typically, an RRM is approximately 90 amino acids long with a typical βαββαβ topology that forms a four-stranded β-sheet packed against two α–helices (exemplified here by RRM3 and RRM4 of PTB; Fig. 1.8). The main protein surface of the RRM involved in the interaction with the RNA is the four-stranded β-sheet, which usually contacts a variable number of nucleotides, ranging from a minimum of two in the case of CBP20 RRM (Calero *et al*., 2002; Mazza *et al*., 2002) to a maximum of eight for U2B'' RRM1 (Price *et al*., 1998).
The most conserved RRM signature sequence is an eight-residue motif called RNP1 (in β3-sheet), which has the consensus [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY]. A second six-residue region of homology, called RNP2 (in β1-sheet), is typically

located 30 residues N-terminal to RNP1, and has the consensus [ILV]-[FY]-[ILV]-X-N-L. Additional conserved amino acids define an 80-residue domain that encompasses the RNA-binding function (Scherly *et al*., 1989).

Recently structural studies emphasise that not only the β-sheet surface but also the loops connecting β-strands and α–helices can be crucial for nucleic acid recognition. For example, one loop for ASF/SF2 RRM2 (Tintaru *et al*., 2007), two loops for Fox-1 (Auwert *et al*., 2006) and three loops for hnRNP F (Dominguez and Allain, 2006) are crucial for nucleic acid interactions.

Some individual RRMs can bind to RNA with great specificity, but multiple domains are often needed to define specificity because number of nucleotides that are recognised by an individual RRM is generally too small to define a unique binding sequence (Auwert *et al*., 2006).

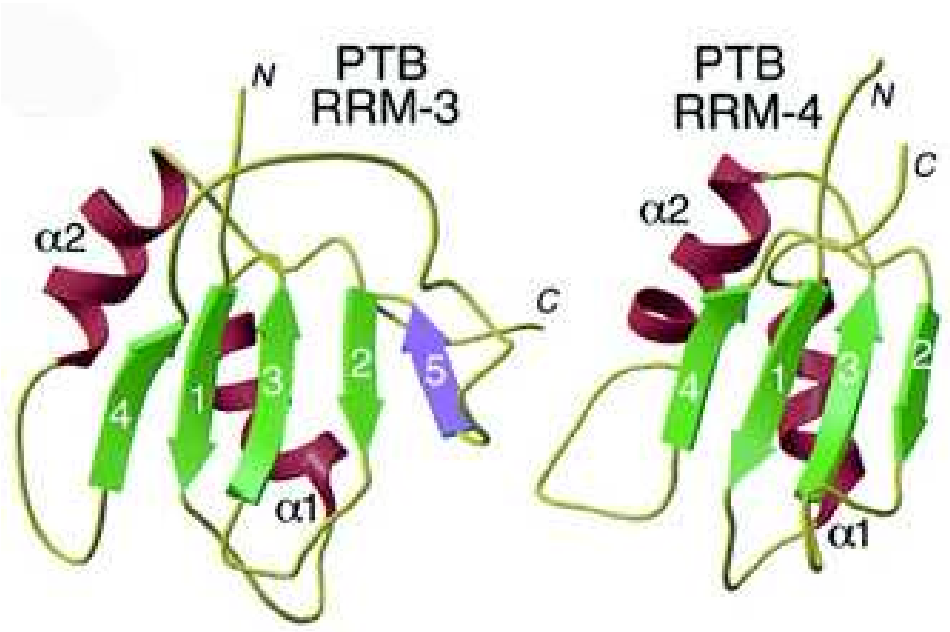


Figure 1.8 Ribbon diagrams for RRMs 3 and 4 of PTB.
The fold is composed of one four-stranded antiparallel beta-sheet specially arranged in the order β4-β1-β3-β2 from left to right (in green) and two helices α1 and α2 (in purple) packed against the beta-sheet. PTB RRM3 contains an additional strand (β5) (in blue) on the side of the RNA-binding surface. (Conte *et al*., 2000).

1.6 HnRNP L and LL

The heterogeneous nuclear ribonucleoprotein (hnRNP) L is a very abundant nuclear RNA-binding protein of 64 kDa (Pinol-Roma *et al*., 1989). Several diverse functions for hnRNP L have been reported, including its binding in a sequence-

specific manner to *cis*-acting RNA sequence elements that enables intron-independent gene expression and facilitate nuclear export of intronless genes (Guang *et al*., 2005; Liu and Mertz, 1995). It has been shown that hnRNP L binds to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase intronless mRNA and enhances its cytoplasmic export. The binding of hnRNP L to the HCV IRES correlates with the translation efficiency of HCV mRNAs (Hahm *et al*., 1998b). More recently, hnRNP L has been shown to regulate the stability of human vascular endothelial growth factor (VEGF) mRNA in response to hypoxia (Shin and Claffey, 1999; Ray *et al*., 2008). Another example of function hnRNP L in mRNA stability is the human endothelial nitric synthase (*eNOS*) gene (Hui *et al*., 2003a). In addition to its role as a sequence-specific RNA protection factor, hnRNP L can act as a splicing regulator. It has been shown that differential binding of hnRNP L to intronic CA-rich sequences of a different length is the mechanism responsible for regulation of mRNA splicing in the human *eNOS* gene. The largest intron 13 in *eNOS* carries a polymorphic CA-rich region (14-44 repeats), and hnRNP L binds preferentially to the longer CA-rich sequence. Moreover, the length of the polymorphic CA-rich region in the *eNOS* gene has been correlated with an independent risk factor for coronary artery disease (Stangl *et al*., 2000).

With an *in vitro* SELEX approach, the binding specificity of hnRNP L has been determined (Fig. 1.9). The sequences obtained showed an enrichment of CA dinucleotides with ACAC and CACA representing the minimal high-score binding motifs for hnRNP L. In addition, hnRNP L recognises with high affinity certain CA-rich clusters (Hui *et al*., 2005).

Recently, hnRNP L protein was characterised as a global regulator of alternative splicing. Many other target genes were identified besides *eNOS*. It was reported that skipping of the *CD45* variable exons 4 and 5 is regulated by binding of hnRNP

Figure 1.9 Sequence motif recognised by hnRNP L.
The 10-nucleotide consensus sequence defined by SELEX approach. The frequency of each of the four nucleotides at any position in the consensus is represented by the letter height. Boxes mark the two highly conserved core tetranucleotides (Hui *et al*., 2005).

L to an activation-responsive sequences (ARSs) that is located within each exon (Rothrock *et al*., 2005; Tong *et al*., 2005; Motta-Mena *et al*., 2010). 11 more target genes of hnRNP L were identified on a global level by combination of splice-sensitive microarray analysis and an RNAi-knockdown approach (Hung *et al*., 2008). Among them *SLC2A2* and *TJP1*, whose mechanistic basis of hnRNP L-regulated alternative splicing were investigated in detail (Heiner *et al*., 2010).

Recently, hnRNP L-like (hnRNP LL), a closely related paralog of hnRNP L, was also characterised as a global regulator of alternative splicing in activated T cells (Oberdoerffer *et al*., 2008; Wu *et al*., 2008). It was reported that hnRNP LL plays an important role in regulation of alternative splicing of *CD45* exon 4 (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 its paralog hnRNP LL share 58% of overall amino acid identity and similar size (558 vs. 542 amino acids) (Fig. 1.10). Both proteins contain four classical RNA recognition motifs (RRMs). The alignment of each individual RRM of hnRNP L and hnRNP LL shows that all RRMs are highly conserved, with RRM2 being most conserved. The N-terminal glycine-rich regions of hnRNP L is less pronounced in the LL paralog, and the proline-rich regions between RRMs 2 and 3 of hnRNP L are absent in 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.

Previously, it has been shown that hnRNP L and hnRNP LL intarects *in vivo* with each other in JSL1 cells (Oberdoerffer *at al*., 2008)

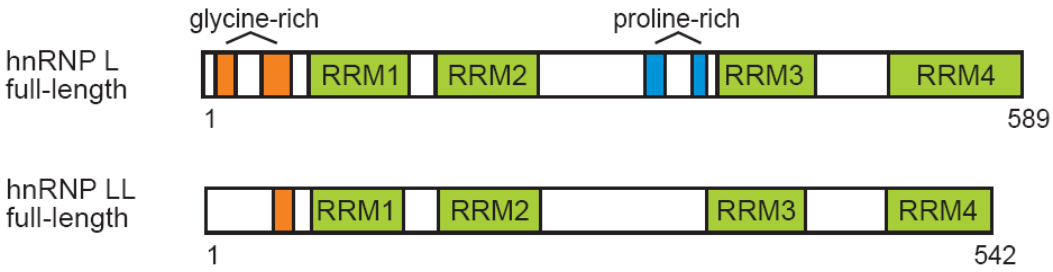


Figure 1.10 Domain structure of the hnRNP L and hnRNP LL proteins.
The domain structure of hnRNP L (P14866; 589 amino acids) and hnRNP LL (Q53T80; 542 amino acids) are schematically represented. Four classical RRM motifs are shown as green boxes, the glycine- and proline-rich regions are shown as orange and blue boxes, respectively.

1.7 Alternative splicing and disease

As alternative splicing affects numerous genes, it is not surprising that changes in alternative splicing are frequently associated with human diseases and cancer. It has been estimated that 15% of all point mutations causing human genetic diseases (Krawczak *et al*., 1992). This is likely to be an underestimate because the analysis was limited to mutations in the classical splice-site sequences. It is known that widespread aberrant splicing is also caused by mutations that disrupt exonic splicing elements (ESEs and ESSs) (Faustino and Cooper, 2003). Thereby, the mutations can be subdivided into two types. Type I mutations affect invariant positions of splice sites resulting in disruption of canonical splice sites. These mutations are responsible for 9-10% of the genetic diseases that are caused by point mutations (Wang and Cooper, 2007). Type II mutations occur in variant positions of splice sites, in enhancer or silencer regions. For some genes, up to 50% of point mutations within exon affected splicing; and it has been hypothesised that more than half of known disease-causing mutations disrupt splicing (Lopez-Bigas *et al*., 2005).

Familial dysautonomia (FD) is an example for a disease caused by a mutation in the 5' splice site (Maayan *et al*., 1987; Carmel *et al*., 2004). FD is a recessive disease that caused by loss of function of the i-kappa-B kinase complex associated proteins (IKBKAP). Affected children show abnormal development of the nervous system that is associated with demyelination in various regions. U1 snRNA interacts with nine nucleotides at 5' splice site (three nucleotides of the exon and first six nucleotides of the downstream intron). The majority of 5' splice sites show complementarity to seven base pairs of U1 snRNA. This means there are usually three mismatches between U1 snRNA and the 5' splice site, which are not randomly distributed. Usually either the exonic portion of the 5' splice site is weakened due to some mismatches, or the intronic portion. In both cases the weaker portion is compensated by strong intronic or exonic portion, respectively (Siliciano and Guthrie, 1988). In exon 20 of the IKBKAP gene, the exonic part of the splice site is weak, due to an A at position -1. The mutation T→C in position 6 of intron 20 weakens the intronic part of the 5' splice site. This interrupts base paring with U1 snRNP, which causes exon skipping.

Spinal muscular atrophy (SMA) is an example of the detrimental effect of a single nucleotide substitution in exonic splicing signal in the survival of motor neuron 2 (*SMN2*) gene. 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). In humans, there are two *SMN* genes, *SMN1* and *SMN2*, both of which encode the same open reading frame. SMN protein plays a critical role in snRNP biogenesis. The vast majority of SMA patients have deletions of the *SMN1* gene, but retain *SMN2*. But single nucleotide substitution C→T at position 6 in exon 7 of *SMN2* gene significantly alters the splicing pattern of *SMN2* pre-mRNA, causing frequent skipping of exon 7, probably due to the disruption of an ESE element within exon. Thus, *SMN2* gene, producing a truncated and non-functional protein, is not able to completely compensate for the loss of SMN1 protein.

1.8 Specific aims

As described above, hnRNP L and its paralog hnRNP LL are the RNA-binding proteins playing the important roles in alternative splicing regulation. Both proteins as typical RNA-binding proteins contain four classical RRMs. The presence of multiple RNA-binding domains immediately raises the question about the contribution of individual domains in RNA binding and alternative splicing regulation.

In the first part of my thesis I aimed to study RNA-binding properties of hnRNP L and hnRNP LL proteins and the role of the individual domains in binding to two different substrates: CA-repeat and CA-rich RNAs. Moreover, I intended to test the structural feature of hnRNP L, namely inter-domain interaction between RRM3 and RRM4, for its role in RNA binding.

In the second part of my thesis I investigated how hnRNP L binds to unusually long CA-rich cluster within its own pre-mRNA. The results obtained provide further insights how the cell can autoregulate its hnRNP L level.

In the third part of my thesis I aimed to determine the important features within CA-rich RNA (number of high-score binding motifs and the length of the spacers between them) required for high-affinity binding of hnRNP L.

In the fourth part of my thesis I aimed to study the contributions of individual RRMs as well as the glycine-rich region in the splicing repression function of hnRNP L.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and reagents

Acetic acid	Roth
Aceton	Roth
Acrylamide	Bio-Rad
Acrylamide/bisacrylamide 30, 30%, 37,5:1	Roth
Acrylamide/bisacrylamide 40, 40%, 19:1	Roth
Agarose ultra pure	Roth
Ammonium persulfate (APS)	Bio-Rad
Ampicilin	Roth
Bacto-agar	Roth
Bacto-trypton	Roth
Bacto-yeast extract	Roth
Bisacrylamide	Bio-Rad
Boric acid	Roth
Bovine serum albumin, RNase free	Roche
Bromophenol blue	Merck
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal)	Roche
Calcium chloride	Merck
Chloroform	Roth
Coomassie brilliant blue R250	Merck
Creatine phosphate	Roche
Dimethyl pyrocarbonate (DMPC)	Sigma
Di-sodium hydrogenphosphate dehydrate ($Na_2HPO_4 \cdot 2H_20$)	Merck
Dithioreitol (DTT)	Roche
Ethanol absolute	Roth
Ethidium bromide	Roth
Ethylendiaminetetraacetic acid (EDTA)	Roth
Ethylen glycol	Roth
Formamide	Roth

Glucose	Sigma
Glycerol	Roth
Glycin	Roth
Glycogen	PeqLab
Heparin	Sigma
N-2-hydroxyethylpiperazine (HEPES)	Sigma
Imidazole	Roth
Isoamyl alcohol	Roth
Isopropanol	Roth
Isopropyl-1-thio-β-D-galactoside (IPTG)	Roche
Kanamycin	Sigma
Lysozyme	Sigma
Magnesium chloride	Merck
Magnesium sulfate	Sigma
2- mercaptoethanol	Roth
Methanol	Roth
Nonidet P-40 /Igepal CA-630	Sigma
Phenylmethylsulfonyl fluoride (PMSF)	Roth
Polyoxyethyleneorbiten monolaurate (Tween20)	Sigma
Polyvinylalcohol	Merck
Potassium chloride (KCl)	Roth
Roti-phenol	Roth
Roti-phenol/chloroform	Roth
Sodium acetic acid (NaAc)	Merck
Sodium chloride (NaCl)	Roth
Sodium citrate	Roth
Sodium dihydrogen phosphate monohydrate ($NaH_2PO_4 \cdot 2H_20$)	Merck
Sodium dodecyl sulfate (SDS)	Roth
Tetracycline	Sigma
N,N,N´,N´-tetramethylenediamine (TEMED)	Bio-Rad
Tris-hydroxymethylaminomethane (Tris)	Roth
Triton X-100	Merck
tRNA from yeast	Roche
urea	Roth
Xylene cyanole FF	Sigma

2.1.2 Nucleotides

deoxynucleoside triphosphate set (dNTP, 100 mM)	Roche
m^7GpppG cap analog	Biozym
ribonucleoside triphosphate set (NTP, 100 mM)	Roche
[α-^{32}P]ATP (3,000 Ci/mmol, 5 µCi/µl)	Hartmann Analytic
[α-^{32}P]CTP (800 Ci/mmol, 10 µCi/µl)	Hartmann Analytic

2.1.3 Enzymes and enzyme inhibitors

AcTEV Protease (10 U/ µl)	Invitrogen
Protease inhibitor cocktail tablets	Roth
Poteinase K (PK, 10 µg/µl)	Roche
Pwo DNA polymerase	Roth
Restriction endonucleases	New England Biolabs
RNase A (100 mg/ml)	Qiagen
RNase inhibitor (RNaseOUT, 40 U/µl)	Invitrogen
RQ1 RNase free DNase (1 U/µl)	Promega
Shrimp alkaline phosphatase (SAP, 1 U/µl)	Roche
Taq DNA polymerase (5 U/µl)	Own purification
T4 DNA ligase, (400 U/µl)	New England Biolabs
T7 RNA polymerase (20 U/µl)	Fermentas
qScript reverse transcriptase	Quanta Biosciences

2.1.4 Reaction buffers

10x *Pwo* Polymerase buffer	Roth
10× restriction enzyme buffer 1, 2, 3, and 4	New England Biolabs
10× RQ1 DNase buffer	Promega
10× SAP buffer	Roche
10× *Taq* polymerase buffer	Promega
5× T4 DNA ligase buffer	New England Biolabs
5× transcription optimised buffer	Fermentas
5× qScript RT buffer	Invitrogen

2.1.5 Molecular weight markers

GeneRuler DNA ladder mix	Fermentas
Molecular weights 14.000-500.000 marker	Sigma
peqGold protein marker IV	PeqLab

2.1.6 Kits

BAC-to-BAC baculovirus expression system	Invitrogen
GC-rich PCR system	Roth
MEGAshortscript kit	Ambion
TOPO TA cloning kit	Invitrogen
QIAGEN RNA/DNA mini kit	QIAGEN
QIAGEN plasmid maxi kit	QIAGEN
QIAprep spin miniprep kit	QIAGEN
QIAquick gel extraction kit	QIAGEN
Silver stain kit	Bio-Rad

2.1.7 Plasmids

pcDNA3	Invitrogen
pcDNA3-GSTZ1	described in Hui *et al*., 2005
pcDNA3-SLC2A2-Wt, -sub	described in Hui *et al*., 2005
pComp-CA32	describe in Hui *et al*., 2003
pComp-control	describe in Hui *et al*., 2003
pFAST-BAC Htb-hnRNP L	described in Hui *et al*., 2005
pGEM3Zf(-)	Promega
pGEX-5x-2	Amersham Bioscience
pRSET-hnRNP L	provided by J. Kim, Pohang University of Science and Technology, Korea
pSP65-MINX	described in Zillman *et al*., 1988
pQE30	QIAGEN

2.1.8 *E.coli* strains and cell lines

BL21 Star (DE3) F- ompT hsdSB (rB-mB-) gal dcm rne131 (DE3) Invitrogen

JM109 {genotype: endA1, gyrA96, thi, hsdR17 (rk-, mk+), relA1, supE44, ≅(lacproAB), [F´, traD36, proAB, laqlqZ≅M15]}	Promega
M15 [pREP4] {genotype: NalS strS rifS thi- lac- ara- gal+ mtl- F- recA+ uvr+ lon+ [pREP4 KanR]}	Qiagen
Sf21 (Spodopetera frugiperda cell) primary HUVEC cell line	Invitrogen
TOP 10 high-competent cells	Invitrogen

2.1.9 Antibodies

Anti-FLAG monoclonal antibody	Sigma
Anti-GAPDH monoclonal antibody	Sigma
Anti-His monoclonal antibody	Qiagen
Anti-hnRNP L monoclonal antibody (4D11)	Sigma
Anti-hnRNP L peptide polyclonal antibody (D-17)	Santa Cruz Biotechnology
Anti-hnRNP LL polyclonal antibody	home made
Anti-mouse Immunoglobulin-Peroxidase	Sigma
Anti-rabbit Immunoglobulin-Peroxidase	Sigma

2.1.10 DNA oligonucleotides

DNA oligonucleotides were ordered from MWG Biotech, Germany

FP cluster A	5’-AATTGGGCCCAACCATACTTCAGCGGCTCT-3’
RP cluster A	5’-AATTGGGCCCTTCAAGTTTGCTTGGATTTTGACC-3’
FP cluster B	5’-AATTGGGCCCCTGCAGTCACCGGC-3’
RP cluster B	5’-AATTGGGCCCGGTGTATGAAGTGGGTAGTGTCC-3’
FP cluster Δ	5’-AATTGGGCCCGGAAGCGGAACACGTAGAAA-3’
RP cluster Δ	5’-AATTGGGCCCAATCATGCGAAACGATCCTC-3’
FP-T7-cluster A	5’-AATACGACTCACTATAGGGAACCATACTTCAGCGGCTCT-3’
FP-T7-cluster B	5’-AATACGACTCACTATAGGGCTGCAGTCACCGGCTGCCA-3’
FP-T7-cluster Δ	5’-AATACGACTCACTATAGGGGGAAGCGGAACACGTAGAAA-3’
5’-KpnI-FP	5’-GGTGGGTACCACAGCCATT-3’
FLAG-L-RP	5’-TTAACTCGAGTCTTGTCGTCATCGTCTTTGTAGTCGG AGGCGTGCTGAGC-3’
hypo-1 (fwd)	5’-CTCGGTAATTGAGAGGAGCG-3’
hypo-2 (rev)	5’- CAAGGCAACATGAGATCAACC-3’

hypo-3 (fwd)	5’-TTAAGAATTCCATCCTCCTCCTCTTCCTCC- 3’
hypo-4 (rev)	5’-TTAAGTCGACTTATAAATGGGATGATGTAGAAAA 3’
LL-KpnI-FP	5’-GGATGGTACCAGCAGCTACAAA-3’
LL-FLAG-RP	5’-AATTGCGGCCGCTCACTTGTCGTCATCGTCTTTGTA GTCTAAATGGGATGATGTAGAAA-3’
2-142 for	5’- TTAAGAATTCACTACTCTACCAGCCAGAAGA-3’
V225I rev	5’-TTCCTGAAAATGATAATTCTCT-3’
V225A rev	5’-TTCCTGAAAATGGCAATTCT-3’
SeqI for	5’-AGAGAATTATCATTTTCAGGAAGAATGGAG-3’
SeqA for	5’-AGAGAATTGCCATTTTCAGGAAGAATGGAG-3’
FLAG-3-471	5’-TTAACTCGAGTCTTGTCGTCATCGTCTTTGTAG TCCACGTTGCTGGGGTGCTG-3’
L-Full-FP	5’-TTAAGAATTCACATGGTGAAGATGGCGGC-3’
L-KpnI-RP	5’-TTAAGGTACCCACCGTGGGGCCCTCCATA-3’
L-Full-screen-FP	5’-TGGGGAGAACTACGATGACCC-3’
T7-CA1	5’-TAATACGACTCACTATAGGGAGA$(CA)_{20}$ -3’
T7-CA2	5’-$(TG)_{20}$TCTCCCTATAGTGAGTCGTATTA-3’
T7-CA3	5’-TAATACGACTCACTATAGGGAGA$(CA)_{10}$ -3’
T7-CA4	5’-$(TG)_{10}$TCTCCCTATAGTGAGTCGTATTA-3’
T7-UCUU1	5’-TAATACGACTCACTATAGGGAG$(TCTT)_6$ -3’
T7-UCUU2	5’-$(AAGA)_6$CTCCCTATAGTGAGTCGTATTA-3’
T7-CU3	5’-TAATACGACTCACTATAGGGAGT$(CT)_{10}$-3’
T7-CU4	5’-$(AG)_{10}$ACTCCCTATAGTGAGTCGTATTA-3’
T7-control-1	5’-TAATACGACTCACTATAGGGAGAcTcTGAAAC TGTGCTGAGTT-3’
T7-control-2	5’-AACTCAGCACAGTTTCAGAGTCTCCCTATA GTGAGTCGTATTA-3’
GSTZ1-U1	5’-TAATACGACTCACTATAGGGACACTTGCACCCTTGCACA-3’
GSTZ1-U2	5’-TGTGTGAGTGTAAGAGTGT-3’
GSTZ1-U3	5’-TAATACGACTCACTATAGGGCTAGTAACGGCCGCCAGT-3’
GSTZ1-U4	5’-GATGGATATCTGCAGAATTC-3’
SLC2A2 fwd	5’-GGGCTGAGGAAGAGACTG-3’
SLC2A2 rev	5’-ACTAATAAGAATGCCCGTGACG-3’
MINX E1 fwd	5’-GAATACACGGAATTCGAGCTC-3’
MINX E2 rev	5’-GATCCCCACTGGAAAGACC-3’
BamHI-TEV	5’-GATCCTTGAGAATTTGTATTTTCAGGGTGG-3’
TEV-EcoRI	5’-AATTCTACCCTGAAAATACAAATTCTCAAG-3’
GST-fwd	5'-ATATAGCATGGCCTTTGCAG-3'
TEV-rev	5'-CCTGAAAATACAAATTCTCAAG-3'

KpnI-GST-fwd	5'-AATTGGTACCTCCCCTATACTAGGTTATTG-3'
G231A-M235V-rev	5’-ATTCCACCACCGCCTGAACTGCATTCTTCCT-3’
G231A-M235V-fwd	5’-AGGAAGAATGCAGTTCAGGCGGTGGTGGAAT-3’
V104A-rev	5’-CCCCTGATGTGGGCAACTGGGGA-3’
V104A-fwd	5’-TCCCCAGTTGCCCACATCAGGGG-3’
A6.1	5’-AATACGACTCACTATAGGGGAATGTACACAG AAGTAGGAAACACAGAGA-3’
A6.2	5’-TCTCTGTGTTTCCTACTTCTGTGTACATTCCCCT ATAGTGAGTCGTATT-3’
B1.1	5’-AATACGACTCACTATAGGGTATTTTTTCTGCAGT CACCGGCTGCCAACA-3’
B1.2	5’-TGTTGGCAGCCGGTGACTGCAGAAAAAATAC CCTATAGTGAGTCGTATT-3’
B2.1	5’-AATACGACTCACTATAGGGTAATCTGCACCAA CTGGAGATCAGAAACCA-3’
B2.2	5’-TGGTTTCTGATCTCCAGTTGGTGCAGATTACCC TATAGTGAGTCGTATT-3’
B3.1	5’-AATACGACTCACTATAGGGAAAAAATTTTAACC ACAACAAAAAAAAAAT-3’
B3.2	5’-ATTTTTTTTTTGTTGTGGTTAAAATTTTTTCCCTA TAGTGAGTCGTATT3’
B4.1	5’-AATACGACTCACTATAGGGCAGCTCAGAGGCA CACAGTTACCTGCTGCC-3’
B4.2	5’-GGCAGCAGGTAACTGTGTGCCTCTGAGCTGCC CTATAGTGAGTCGTATT-3’
B5.1	5’-AATACGACTCACTATAGGGCGTAGTGGGGCCA CTGGCACACCTCATACC3’
B5.2	5’-GGTATGAGGTGTGCCAGTGGCCCCACTACGC CCTATAGTGAGTCGTATT-3’
B6.1	5’-AATACGACTCACTATAGGGGGAGACACGCACAC GGAGCGCTACACAGCC-3’
B6.2	5’-GGCTGTGTAGCGCTCCGTGTGCGTGTCTCCC CCTATAGTGAGTCGTATT-3’
B7.1	5’-AATACGACTCACTATAGGGAAGCACCGCACCAC CGCACTGCAGCACATA-3’
B7.2	5’-TATGTGCTGCAGTGCGGTGGTGCGGTGCTTC CCTATAGTGAGTCGTATT-3’

B8.1	5’-AATACGACTCACTATAGGGGGAGGGACACTAC CCACTTCATACACCTTT-3’
B8.2	5’-AAAGGTGTATGAAGTGGGTAGTGTCCCTCCCCC TATAGTGAGTCGTATT-3’
B3.1Mut	5’-AATACGACTCACTATAGGGAAAAAATTTTAACCG CGACAAAAAAAAAAT-3’
B3.2Mut	5’-ATTTTTTTTTTGTCGCGGTTAAAATTTTTTCCC TATAGTGAGTCGTATT-3’
Mut 1 B7.1	5’-AATACGACTCACTATAGGGAAGCGCGGCACC ACCGCACTGCAGCACATA-3’
Mut 1 B7.2	5’-TATGTGCTGCAGTGCGGTGGTGCCGCGCTTCC CTATAGTGAGTCGTATT-3’
Mut 1-2 B7.1	5’-AATACGACTCACTATAGGGAAGCGCGGCGCC GCCGCACTGCAGCACATA-3’
Mut 1-2 B7.2	5’-TATGTGCTGCAGTGCGGCGGCGCCGCGCTTCC CTATAGTGAGTCGTATT-3’
Mut 3 B7.1	5’-AATACGACTCACTATAGGGAAGCACCGCACCAC CGCACTGCAGCGCGTA-3’
Mut 3 B7.2	5’-TACGCGCTGCAGTGCGGTGGTGCGGTGCTTC CCTATAGTGAGTCGTATT-3’
Mut 2 B7.1	5’-AATACGACTCACTATAGGGAAGCACCGCGCCGC CGCACTGCAGCACATA3’
Mut 2 B7.2	5’-TATGTGCTGCAGTGCGGCGGCGCGGTGCTTCCC TATAGTGAGTCGTATT-3’
B6B7.1	5’-AATACGACTCACTATAGGGGGAGACACGCACA CGGAGCGCTACACAGCCAGAAGCACCGCACCACC GCACTGCAGCACATA-3’
B6B7.2	5’-TATGTGCTGCAGTGCGGTGGTGCGGTGCTTCTG GCTGTGTAGCGCTCCGTGTGCGTGTCTCCCCCTAT AGTGAGTCGTATT-3’
B7.1	5’-AATACGACTCACTATAGGGAAGCACCGCACCACC GCACTGCAGCACATA-3’
B7.2	5’-TATGTGCTGCAGTGCGGTGGTGCGGTGCTTCCCT ATAGTGAGTCGTATT-3’
B7(4).1	5’-AATACGACTCACTATAGGGAAGCACCGCACCACCC TGCCACATA-3’
B7(4).2	5’-TATGTGGCAGGGTGGTGCGGTGCTTCCCTATAGTG AGTCGTATT-3’
B7(13).1	5’-AATACGACTCACTATAGGGAAGCACCGCACCACC

	TGGCACTGCAGTGCACATA-3’
B7(13).2	5’-TATGTGCACTGCAGTGCCAGGTGGTGCGGTGCT TCCCTATAGTGAGTCGTATT-3’

2.1.11 RNA oligonucleotides

5’-Biotin-$(CA)_{32}$-3’	Xeragon

2.1.12 Other materials

Eppendorf tube, 1.5 ml, 2 ml	Eppendorf
Falcon tube	Greiner
FuGene transfection reagent	Roche
Glutathione agarose 4B	Macherey-Nagel
HeLa cell nuclear extract	4C Biotech
Hybond ECL nitrocellulose membrane	GE Healthcare
Lumi-Light western blotting substrate	Roche
Neutravedin agarose resin	Thermo Scientific
Nickel-nitrilotriacetic acid (Ni-NTA) agarose	QIAGEN
Pasteur pipet	Roth
Roti-block	Roth
Superdex 200	Sigma
X-ray film	Kodak X-OMAT

2.2 Methods

2.2.1 DNA cloning

2.2.1.1 Preparation of plasmid DNA

Plasmid DNA was isolated from overnight culture using the QIAprep Spin Miniprep Kit (QIAGENE) according to the manufacturer’s instruction. 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

The separation of DNA fragments according to their size was performed using gels with 1-2 % agarose in TBE buffer (100 mM boric acid; 100 mM Tris; 2 mM EDTA pH 8,8) with ethidium bromibe (1:20.000 dilution). Samples were mixed with appropriate amount of 6x DNA loading buffer (30% (v/v) glycerol; 0,025% (w/v) bromphenol blue) before loading. Gels were run at 130V. The bands were visualised with gel documentation system (SynGene).

2.2.1.3 DNA extraction from agarose gels

DNA bands were excised with a scalpel, transferred to sterile Eppendorf vials, weighed and purified with QIAprep Gel Extraction Kit columns (QIAGEN) following the instructions of the manufacturer.

2.2.1.4 DNA cleavage with restriction enzymes

Restriction digests were performed using the buffer system and temperature recommended by the manufacturer (New England Biolabs). 1-5 units of enzyme per 1 µg DNA were used. Incubation time was at least 1 hour. Completion of the digests was analysed on agarose gels (2.2.1.2). DNA was purified by gel extraction or phenolisation.

2.2.1.5 Dephosphorylation of DNA

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

2.2.1.6 Ligation

DNA fragments were ligated with T4 DNA ligase (New England Biolabs) in a volume of 10 µl at room temperature for 2 hours or at 16ºC overnight using the buffer system supplied by the manufacturer.

2.2.1.7 Transformation of *E.coli* cells

For transformation, chemically competent cells were thawed on ice. 100 µl of the cells were mixed with 100 ng of plasmid DNA or 10 µl of ligation mixture and incubated on ice for 30 min. The mixture was placed at 42ºC for 40 s, then on ice for 5 min. After addition of 500 µl LB medium and 60 min incubation at 37^{o}C, the mixture was plated on pre-warmed LB agar plates with 100 µg/ml ampicillin. The plates were incubated overnight at 37ºC. If blue/white selection was carried out for selecting recombinants, 40 µl of X-gal stock solution and 40 µl of IPTG solution were spread on LB plates before the transformation cultures were plated out.

2.2.1.8 PCR amplification of DNA

A standard PCR reaction to amplify DNA from a plasmid template contained 1-10 ng of plasmid DNA, forward and reverse primers (0,5 µM each), dNTPs (200 µM), 1x *Taq* polymerase buffer, 1 mM $MgCl_2$ and 1U *Taq* polymerase in total volume of 25 µl. When the amplification was made for cloning purposes, a high-fidelity polymerase, e.g. *Pwo* DNA polymerase was used instead of *Taq* polymerase. The amplification was carried out in a GeneAmp PCR System 9700 thermocycler under the following conditions: initial denaturation for 2-4 min at 94ºC; 25-35 cycles of 15-30 sec at 94ºC, annealing at the T_m of the primers pair, extension of 1 min per 1 kb at 72ºC. After the last cycle the reaction was held for 5-10 min at the extension temperature to allow completion of amplification of all products.

2.2.2 Generation of hnRNP L and hnRNP LL mutants

pGEX-5x-2-hnRNP L-FLAG. The *EcoRI-XhoI* restriction fragment containing the full-length cDNA for hnRNP L was released from pFAST-BAC Htb-hnRNP L and cloned into pGEX-5x-2 expression vector. To introduce the FLAG-tag the PCR fragment amplified form pGEX-5x-2-hnRNP L vector with 5'-KpnI and FLAG-L

primers was digested with *KpnI* and *EcoRI* restriction enzymes and subcloned into pGEX-5x-2-hnRNP L vector.

pGEX-5x-2-hnRNP LL-FLAG. Total RNA was isolated from HeLa cells and reverse transcribed with oligo(dT) primers. The hnRNP LL ORF amplified with specific hypo-1 and hypo-2 primers was used as a template for amplification of protein coding region with hypo-3 and hypo-4 primers. After *EcoRI* and *SalI* restriction the PCR fragment was cloned into pGEX-5x-2 expression vector. The insertion of FLAG-tag was done as for pGEX-5x-2-hnRNP L-FLAG vector with specific primers: LL-KpnI and LL-FLAG.

The deletion mutants of hnRNP L and hnRNP LL were generated by replacement of restriction fragments of cloned wild-type hnRNP L and hnRNP LL cDNA with PCR amplified fragments.

TEV cleavage site was introduced in pGEX-5x-2-L-FLAG construct and its deletion derivatives by replacement of the DNA fragment between *EcoRI* and *BamHI* restriction sites of the initial constructs with annealed and *EcoRI-BamHI* cut oligos containing TEV cleavage site (TEV-EcoRI / BamHI-TEV).

For V104A, V225A, G231A-M235V mutants, two steps of PCR amplification were employed. First, wt hnRNP L cDNA was amplified in separate reactions with oligo pairs for V104A: L-Full-FP/ V104A-rev and V104A-fwd/ L-KpnI-RP, for V225A: L-Full-FP/ V225A-rev and V225A-fwd/ L-KpnI-RP, for G231A-M235V: L-Full-FP/ G231A-M235A-rev and G231A-V235M-fwd/ L-KpnI-RP. Each combination of PCR products, which have a 20 bp overlap, were gel-purified and used in equimolar amounts for second PCR amplification with oligos L-Full-FP/ L-KpnI-RP for all constructs. The PCR products were purified and digested with *EcoRI* and *KpnI* restriction enzymes and subcloned into corresponding sites of pGEX-5x-2-TEV-hnRNP L-FLAG.

All subcloning steps were carried out using *E.coli* J109M. All constructions were confirmed by sequence analysis (SeqLab, Göttingen).

2.2.3 Expression and purification of proteins in *E.coli*

2.2.3.1 GST-tagged hnRNP L, hnRNP LL and their derivatives

To express GST-tagged hnRNP L, hnRNP LL and threir mutant derivates, the expression constructs were transformed into *E.coli* BL21 Star cells. Bacteria were grown in 400 ml of LB medium, containing ampicillin (100 µg/ml) to OD_{600}=0,6. The expression of recombinant proteins was induced by 0,1 mM IPTG overnight.

Harvested bacteria were subjected to one freeze/thaw cycle at -80ºC and resuspended in 10 ml Extraction Buffer (50 mM Tris-HCl, pH 8,5; 100 mM NaCl; 1 mM EDTA; 1 mM DTT; protease inhibitor cocktail, Roche). After incubation on ice for 20 min with 1 mg/ml of lysozyme, the cell suspension was sonicated 3 times by pulses of 30 sec at 50% amplitude (Branson sonifier model B-15), and cellular debris was removed by centrifugation for 30 min at 14.000 rpm at 4ºC. The supernatant was incubated 30 min on a rotating wheel at 4ºC, with glutathione beads (200 µl of 50% bead slurry, GE Healthcare) pre-equilibrated in a 1:1 ratio in Extraction Buffer. Beads were washed 4 times for 5 min on a rotating wheel at RT with Extraction Buffer, followed by elution of recombinant GST-proteins with 200 µl of Elution Buffer (50 mM Tris-HCl pH 8,5; 100 mM NaCl; 1 mM EDTA; 1 mM DTT; protease inhibitor cocktail (Roche); 30 mM reduced glutathione). Aliquots of each elution fraction were separated on 10% SDS-PAGE and stained by Coomassie Blue. Fractions containing recombinant GST-proteins were collected, dialysed against Buffer D (20 mM HEPES pH 8,0; 100 mM KCl; 0,2 mM EDTA pH 8,0; 20% glycerol; 15 mM $MgCl_2$) and protein concentration was estimated on Coomassie Blue stained SDS-PAGE (2.2.8 and 2.2.9) , using BSA (Biolabs) as a standard.
In order to remove GST tag the TEV cleavage was performed. GST-tagged proteins bound to glutathione beads were incubated with TEV protease at 4ºC overnight in TEV cleavage buffer (10 mM Tris-HCl pH 8,0; 150 mM NaCl; 0,1% NP-40; 0,5 mM EDTA; 1 mM DTT). The supernatant was collected and dialysed against buffer D.

2.2.3.2 His-tagged hnRNP L

To express His-tagged hnRNP L pQE30-hnRNP L construct was transformed into BL21 Star *E.coli* strain. Single colony was then inoculated into 4 ml of LB medium and grown overnight. The overnight culture was inoculated into 400 ml of fresh LB medium containing ampicillin (100 µg/ml) and incubated at RT to OD_{600}=0,6. For induction of the protein expression 1 mM of IPTG was added and cultured overnight at RT. After the induction cells were harvested by centrifugation for 15 min at 4000 rpm. The recombinant His-tagged hnRNP L protein was purified over Ni-NTA beads according to the manufacturer's instruction (QIAexpressionist, QIAGEN).

2.2.4 Generation and purification of recombinant baculovirus-expressed His-tagged hnRNP L

Recombinant baculoviruses are widely used for the production of high levels of properly post-translationally modified, biologically active and functional recombinant

proteins. Baculovirus expression system based upon the ability to propagate AcMNPV (Autographa californica multiple nuclear polyhedrosis virus) in insect cells. The heterologous genes are placed under transcriptional control of the strong AcNPV polyhedron promoter, thus the recombinant protein is expressed in place of the naturally occurred polyhedron protein. Since baculoviruses are non-infectious to vertebrates, they are safer to work with than other mammalian viruses (Inceoglu *et al*., 2001). The recombinant baculovirus expressing hnRNP L was made using the Bac-to-Bac baculovirus expression system (Invitrogen, USA) according to the manufacturer's instructions. It is a rapid and efficient method based on site-specific transposition.

A full-length cDNA encoding hnRNP L was cloned into a pFASTBAC vector, and the recombinant plasmid was transformed into DH10BAC competent cells which contain the bacmid with a mini-*att*Tn7 target site and the helper plasmid. The mini-Tn7 element on the pFAST-BAC plasmid can transpose to the mini-*att*Tn7 target site on the bacmid in the presence of transposition proteins provided by the helper plasmid. Colonies containing recombinant bacmids were identified by antibiotic selection and blue/white screening, since the transposition results in disruption of the lacZα gene. High molecular weight mini-prep DNA was prepared from selected *E.coli* clones containing the recombinant bacmid, and this DNA was then used to transfect insect cells. The steps to generate a recombinant baculovirus by site-specific transposition using the BAC-TO-BAC Baculovirus Expression System are outlined in figure 2.1.

2.2.4.1 Infection of insect cells

SF21 cells (2×10^7) were seeded in a 150 cm^2 flask with SF-900 II medium with L-glutamine and 10% FCS and incubated at 27ºC for 30 min. After the cells were attached to the flask, they were infected by 10 µl of viral stock [2×10^9 plaque forming units (pfu)/ml] at a multiplicity of infection (MOI) of 10. Four days after infection, infected cells were harvested by centrifugation for 10 min at 500 g, 4ºC.

2.2.4.2 Purification of His-tagged hnRNP L from baculovirus-infected insect cells

The cell pellet from one 150 cm^2 flask (2×10^7cells) was resuspended in 4 ml of lysis buffer (50 mM NaH_2PO_4; 300 mM NaCl; 10 mM imidazole; pH 8,0) and

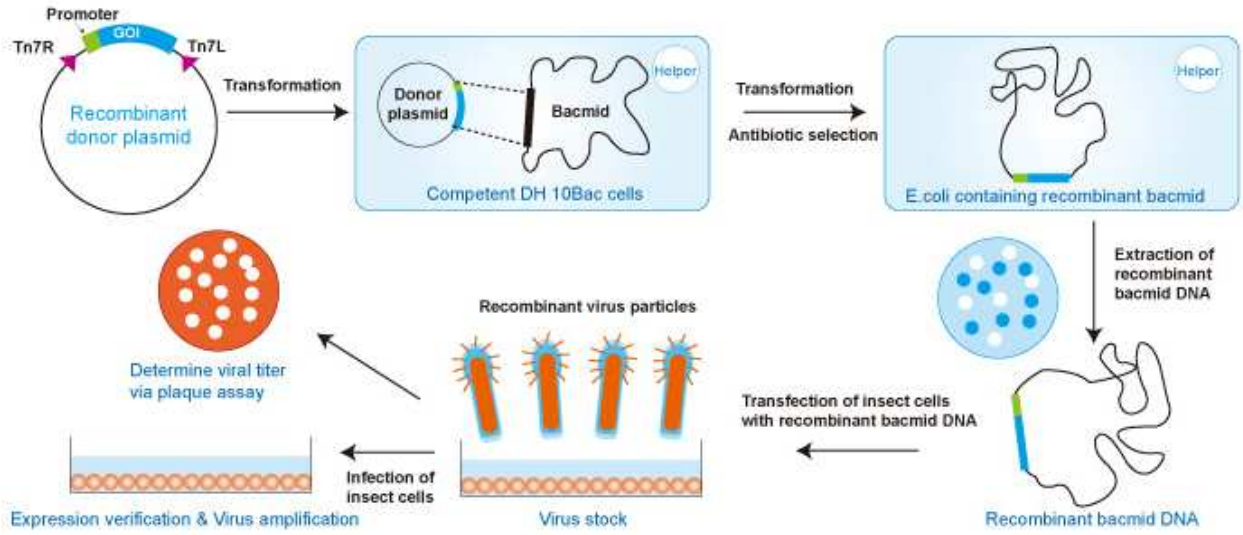


Figure 2.1 Generation of recombinant baculoviruses and gene expression with the BAC-TO-BAC expression system.
Promotor (green box) - AcNPV polyhedron promoter, Tn7R and Tn7L indicate transposon 7 elements, GOI – foreign gene of interest (Invitrogen instruction manual).

incubated on ice for 10 min. Then the cell lysate was centrifuged at 10.000 x g at 4ºC for 10 min to pellet cellular debris and DNA. 50 µl of packed Ni-NTA beads was added to cleared lasyte and incubated at 4ºC for 2 h with rotating. The bound Ni-NTA beads were pelleted by centrifugation at 6,000 rpm for 1 min and washed with 1 ml of wash buffer (50 mM NaH_2PO_4; 300 mM NaCl; 20 mM imidazole; pH 8,0) four times. The His-tagged protein was eluted with 100 µl of elution buffer (50 mM NaH_2PO_4; 300 mM NaCl; 250 mM imidazole; pH 8.0) four times. The eluates were collected and analysed by SDS-PAGE.

2.2.5 *In vitro* transcription

2.2.5.1 Annealing of DNA oligos

Two complementary oligonucleotides were mixed at the same molar concentration in 2x Annealing Buffer (20 mM Tris-HCl, pH 8; 100 mM NaCl) at a final volume of 100 µl. The tube was placed in a standard heatblock for 2-5 min at 95ºC. The heatblock was turned off and the sample was slowly cooled down.

2.2.5.2 Transcription of ^{32}P-labeled RNA

Internal radioactive labeling of RNAs was performed by T7 *in vitro* transcription. Either PCR products or annealed oligos were used as templates for transcription. Only in the case of MINX, RNA was transcribed from a linearised plasmid. 2 µl of template DNA (100 ng/µl) or 1 µl of plasmid (1 µg/µl) DNA were mixed with 5 µl 5x transcription buffer, 2,5 µl 100 mM DTT; 1,25 µl 10 mM ATP; 1,25 µl 1 mM CTP; 1,25 µl 10 mM UTP; 1,25 µl 10 mM GTP; 1 µl RNase inhibitor; 1 µl [α-^{32}P]CTP (800 Ci/mmol) and 1 µl T7 RNA polymerase (20 U/µl). DMPC-H_2O was added to the final volume of 25 µl. For MINX transcription, SP6 RNA polymerase was used. Transcriptions were carried out at 37ºC for 2 h. 2 µl RQ1 DNase was added to each reaction and incubation continued for 30 min at 37ºC.
Unincorporated nucleotides were removed from the transcription reactions using RNA spin columns following the manufacturer's instructions (Roche).
Transcribed RNAs were precipitated with 600 µl ethanol; 20 µl 3M NaAc pH 5,2; and 1 µl (20 mg/ml) glycogen. After pelleting, washing and drying, the amount of RNA was measured using a scintillation counter and calculated using the following formula:

$$\text{RNA [ng]} = \text{CTP cold [µM]} \times \text{volume of reaction [µl]} \times \%\text{incorporation} \times 0{,}0132$$

The transcripts were dissolved in an appropriate volume of DMPC-H_2O.

2.2.5.3 Gel purification of ^{32}P-labeled RNA

RNA transcripts were separated by denaturing polyacrylamide gel (10% (v/v) Acrylamide/Bis (1x TBE, 50% urea); 10% (v/v) APS; 0,001 % (v/v) TEMED) electrophoresis. The RNA bands were cut out of the gel and sliced into small pieces. RNAs were eluted from the gel by 300 µl of 2× PK buffer overnight at room temperature with rotating. The eluate was phenolised with Roti-phenol/chloroform and precipitated with 1 µl of glycogen (20 mg/ml) and 900 µl of ethanol. After washing and air-drying, the purified RNAs were dissolved in DMPC-treated H_2O.

2.2.5.4 Transcription without ^{32}P-label

Transcription of RNA was carried out as described above. Only 1,25 µl of 10 mM CTP was added since [α-^{32}P]CTP 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-H_2O.

2.2.5.5 Biotin attachment

3'-end biotinylation of transcribed RNA was done according to the protocol (Hartmann *et al*., 2005).

2.2.6 *In vitro* splicing of pre-mRNAs

2.2.6.1 Splicing reaction

Approximately 10 ng of unlabeled, capped pre-mRNA substrate was incubated with 30% HeLa nuclear extract in a total volume of 12,5 µl containing 3,2 mM $MgCl_2$; 20 mM phosphocreatine; 0,5 mM ATP; 2,66 % polyvinyl alcohol; 1,6 U/µl RNaseOUT (Invitrogene). Reaction mixtures were incubated for 4 h at 30ºC.

2.2.6.2 Proteinase K treatment

Aliquots of the splicing reaction were mixed with 83,5 µl DMPC-H_2O; 100 µl 2x PK buffer (200 mM Tris-HCl; 300 mM NaCl; 25 mM EDTA; 2% SDS) and 4 µl PK (10 mg/ml). Reactions were incubated for 30 min at 37ºC. The RNA was extracted with phenol, precipitated, washed, dried and dissolved in 5 µl DMPC-H_2O.

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

In the RT-PCR assay, 1 µl of RNA was reverse-transcribed in a volume of 15 µl with qScript Reverse Transcriptase (Quanta Biosciences) and random hexamer primer according to the manufacturer's protocol to generate DNA. 5 µl of the RT reaction was used as template in the PCR assay (2.2.1.8). 10 µl of each PCR

reaction was analysed by 2% agarose gel electrophoresis and ethidium bromide staining (2.2.1.2).

2.2.7 Depletion of hnRNP L from HeLa nuclear extract

HeLa cell nuclear extract was depleted of hnRNP L with a 5'-biotinylated $(CA)_{32}$ RNA oligonucleotide and neutravidin-agarose resin. 200 µl neutravidin-agarose resin was blocked in 500 ml blocking solution (4 mM HEPES pH 8,0; 0,2 mM DTT; 2 mM $MgCl_2$; 20 mM KCl; 0,002% (v/v) NP-40; 0,2 mg/ml tRNA; 1 mg/ml BSA; 0,2 mg/ml glycogen) at 4ºC O/N followed by washing four times with 1 ml of WB 400 (20 mM HEPES pH 8,0; 1 mM DTT; 10 mM $MgCl_2$; 400 mM KCl; 0,01% (v/v) NP-40). For each depletion reaction, 20 µl of packed beads were then incubated with 6 µg of the 5'-biotinylated $(CA)_{32}$ RNA oligonucleotide in 200 µl of WB 400 for 4 h at 4ºC 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 µl of HeLa nuclear extract for 30 min at 30ºC with rotation. Then, the KCl concentration was increased up to 600 mM and the incubation continued for 20 min at 4ºC. After removal of the neutravidin beads, depleted nuclear extracts were dialysed against buffer D for 2 h at 4ºC.

2.2.8 Electrophoresis of proteins

Proteins were resolved on denaturing SDS polyacrylamide gels. Two buffers were used: 4x stacking gel buffer (0,5 M Tris-HCl pH 6,8; 0,4% (w/v) SDS) and 4x separating buffer (1,5 M Tris-HCl pH 8,8; 0,4% (w/v) SDS) to prepare either stacking gel (5% acrylamide/bisacrylamide 37,5:1; 1x stacking gel buffer; 200 µl APS; 20 µl TEMED in 20 ml) or separating gels (10-12% Acrylamide/bisacrylamide 37,5:1; 1x separating gel buffer; 100 µl APS; 10 µl TEMED in 10 ml) respectively. The proteins were mixed with sample loading buffer (2% SDS; 10% glycerol; 50 mM Tris-HCl pH 6,8; 0,005% BPB), denatured at 96ºC for 5 min and loaded on the gel. Electrophoresis was carried out at 100-130 V in SDS gel running buffer (25 mM Tris; 250 mM glycine pH 8,3; 0,1% (w/v) SDS). Gels were subjected to either Western blot analysis or Coomassie blue staining.

2.2.9 Coomassie staining

To detect the proteins in SDS polyacrylamide gels, coomassie staining was used. After electrophoresis the gel was placed in staining solution (0,25% (w/v) Coomassie Brilliant Blue R250; 50% (v/v) methanol; 10% (v/v) acetic acid) for 20 - 30 min at RT. After the incubation the gel was distained with Coomassie blue distaining solution (50% (v/v) methanol; 10% (v/v) acetic acid) until background became clear. The gel was dried using a vacuum gel dryer.

2.2.10 Western blotting

Proteins were resolved on 10 % SDS-PAGE and transferred to PVDF nitrocellulose membrane (GE Healthcare) in transfer buffer (50 mM Tris; 380 mM glycine; 20% (v/v) methanol; 0,02% (w/v) SDS) for 30 min at 300 mA using a semi-dry transfer cell (Bio-Rad). The membrane was blocked for 1 hour in a blocking buffer (1x TBS; 1x Roti-Block and 0,05% (v/v) Tween) at RT. The primary antibodies were diluted in fresh blocking buffer as follows: monoclonal anti-hnRNP L (4D11) AB 1:10.000; peptide polyclonal anti-hnRNP L (D-17) AB 1:500; polyclonal anti-hnRNP LL AB 1:200; monoclonal anti-GAPDH AB 1:10.000 and added to mambrane for incubation for 1h at RT. The membrane was washed three times for 10 min in 1 x TBST (50 mM Tris-HCl pH 8,0; 150 mM NaCl; 0,05% Tween) and incubated with a secondary peroxidase conjugated antibodies diluted in blocking buffer 1:10.000 for 1 hour at RT. The membrane was subsequently washed three times for 10 min in TBST and the bound antibodies were detected by the Lumi-Light ECL system according to supplemented protocol. The membrane was then exposed to an X-ray film (Kodak X-OMAT) and developed.

2.2.11 Electromobility shift assay (band shift)

The different amount of hnRNP L and its mutant derivates were incubated with 4 ng of ^{32}P-labeled gel-purified RNA (GSTZ1; $(CA)_{10}$; $(CA)_{20}$) in a binding buffer containing 20 mM HEPES pH 8,0; 100 mM KCl; 15mM $MgCl_2$; 0,2 mM EDTA pH 8,0; 20% glycerol; 5 mg tRNA. The total reaction volume was 25 µl. The mixtures were incubated at 4ºC for 20 min. 10µl of aliquots was transferred to a new tube containing 1 µl of heparin (4 mg/µl). After 5 min heparin treatment at room temperature, 1 µl of native RNA gel loading buffer (0.025% (w/v) bromophenol blue; 30% (v/v) glycerol) was added and the samples were fractionated on a 6% native

RNA gel (acrylamide/bisacrylamide 80:1; 1xTBE; 400 µl 10% APS; 40 µl TEMED filled up with H_2O to 50 ml) for 1h at 23W. Samples were visualised by autoradiography.

2.2.12 Filter binding assay

Filter binding assay was carried out as described (Hui *et al*., 2005). Approximately 90 fmol of *in vitro* transcribed ^{32}P-labeled RNA was incubated with different amount of hnRNP L and its mutants in a final volume of 50 µl of binding buffer (10 mM Tris-HCl pH 8,0; 100 mM KCl; 2,5 mM $MgCl_2$; 0,01% NP-40; 10% glycerol; 0,2 mg/ml BSA) at RT for 20 min. The reactions were applied onto the pre-incubated with washing buffer (10 mM Tris-HCl pH 8,0; 100 mM KCl; 2,5 mM $MgCl_2$) for 30 min 0,45 µm supported nitrocellulose membrane (Bio-Rad). After washing and air-drying, the filter was quantified on a PhosphorImager.

2.2.13 Gel-filtration of RNA-protein complexes

^{32}P-labeled RNA was denatured for 2 min at 95ºC followed by cooling for 5 min on ice. Appropriate amount of His-tagged hnRNP L in buffer D was added to RNA. RNA and protein were incubated for 15 on ice. Ger filtration of RNA-protein complex was performed using a 1 x 30-cm glass column (Bio-Rad) filled up to 28 cm with Superdex 200 gel filtration matrix (Sigma) and gel filtration buffer (10 mM Tris-HCl pH 8; 100 mM KCl; 2,5 mM $MgCl_2$) at a flow rate of 100 µl/min using a peristaltic pump. The eluted fractions (250 µl each) were analysed by denaturing RNA gel and autoradiography.

2.2.14 Glycerol gradient

^{32}P-labeled 5'-CA cluster or 3'- CA cluster RNAs were incubated with different amount of His-tagged hnRNP L at 4ºC for 20 min in Buffer G (20 mM HEPES; 150 mM KCl; 1,5 mM $MgCl_2$) in the presence of 2,3 µg/µl tRNA and 1,6 U/µl RNasOUT. Samples were then applied onto a 10–30% glycerol gradient in 2 ml and centrifuged at 44,000 rpm at 4ºC for 4 h, using TLS 50 rotor (Beckman). Fractions were taken from the top of tube by using a pipette (10 fractions, 200 µl per fraction). RNAs from each fraction were isolated and analysed by denaturing gel and autoradiography.

2.2.15 *In vitro* protein-binding assay

Purified GST-tagged hnRNP L and its mutant derivates were incubated with 100 µl packed glutathione Sepharose 4B (GE Healthcare) in TIF buffer (150 mM NaCl; 20 mM Tris-HCl pH 8,0; 1 mM $MgCl_2$; 0,1% NP40; 10% Glycerol) for 30 min at RT. The beads were pelleted and washed four times with TIF buffer followed by incubation with 2 µg of recombinant His-tagged hnRNP LL in 500 µl of TIF buffer for 1h at 4ºC. The beads were washed four times with TIF buffer and resuspended in SDS sample buffer (2% SDS; 80 mM Tris-HCl; 5% β-mercaptoethanol; 15% glycerol; 0,05% bromphenol blue; pH 6.8) and boiled. The proteins eluted from the beads were then fractionated on a 10% SDS-PAGE, transferred to a nitrocellulose membrane, and visualised by immunoblot with an anti-His monoclonal antibody (Qiagen).

2.2.16 Databases and computational tools

Database / software	URL	Description	Reference
ClustalW	http://www.ebi.ac.uk/clustalw/index.html	Multiple sequence alignment program for DNA or proteins	(Thompson *et al*., 1994)
Human BLAT Search	http://www.genome.ucsc.edu/cgi-bin/hgBlat	Sequence alignment tool similar to BLAST	(Kent, 2002; Kent *et al*., 2002)
NCBI HomoloGene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene	Automated detection of homologs among the annotated genes of several completely sequenced eukaryotic genomes	(Wheeler *et al*., 2005)
SMART	http://smart.embl-heidelberg.de	Allows identification of genetically mobile domains and analysis of domain architectures	(Schultz *et al*., 1998; Letunic *et al*., 2002)
Pfam	http://pfam.sanger.ac.uk	Database of protein families that includes their annotations and multiple sequence alignments	(Finn e*t al*., 2008)
SMS (sequence manipulation suite)	http://www.bioinformatics.org/sms2/index.	JavaScript programs for generating, formatting, and analysing short DNA and protein sequences	(Stothard, 2000)
RNAfold	http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi	Prediction of secondary structures of single stranded RNA or DNA sequences	(Hofacker, 2003)

3. Results

3.1 HnRNP L and hnRNP LL domain structures

SMART and PFAM databases (Schultz, 1998; Finn *et al*., 2008) were used in order to inspect the domain structure of hnRNP L and its paralog hnRNP LL. The three classical RNA recognition motifs (RRM) were identified in both proteins (Fig. 3.1), notably RRM4 of both proteins was not identified as an individual domain by the SMART computational sequence analysis tools, but it was included into the domain structure of hnRNP L and hnRNP LL as a loosely conserved RRM. The amino acid sequence analysis also revealed that hnRNP L has an additional glycine-rich region at the N-terminal end of the protein and a proline-rich region between RRM2 and RRM3. In contrast, hnRNP LL has only a less pronounced glycine-rich region and does not have a proline-rich region.

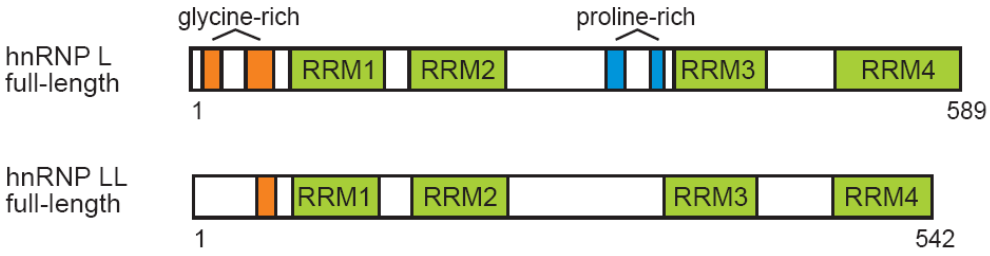


Figure 3.1 Domain structures of the hnRNP L and hnRNP LL proteins.
Schematic representation of the domain structures of hnRNP L (P14866; 589 amino acids) and the closely related hnRNP LL protein (Q53T80; 542 amino acids). Four RNA recognition motifs (RRM) are represented by the green boxes. The glycine- and proline-rich regions are shown in orange and blue, respectively

A comprehensive alignment of the RNA recognition motifs of hnRNP L and hnRNP LL with all known RRM containing proteins (Fig. 3.2, only several proteins are shown) confirmed that three RRMs, namely 1, 2 and 3 of hnRNP L as well as hnRNP LL match the criteria for identifying bonafide RRMs with RRM2 being most conserved, whereas RRM4 domains only partially satisfy the criteria. Distinctive features of most RRMs are βαββαβ structure and the solvent-exposed aromatic residues in the β-3 strand (RNP1) and β1 strand (RNP2), which are conserved in all of these RRMs. Although in the RRMs of hnRNP L and hnRNP LL several conserved phenylalanine aromatic residues are substituted by leucine, the

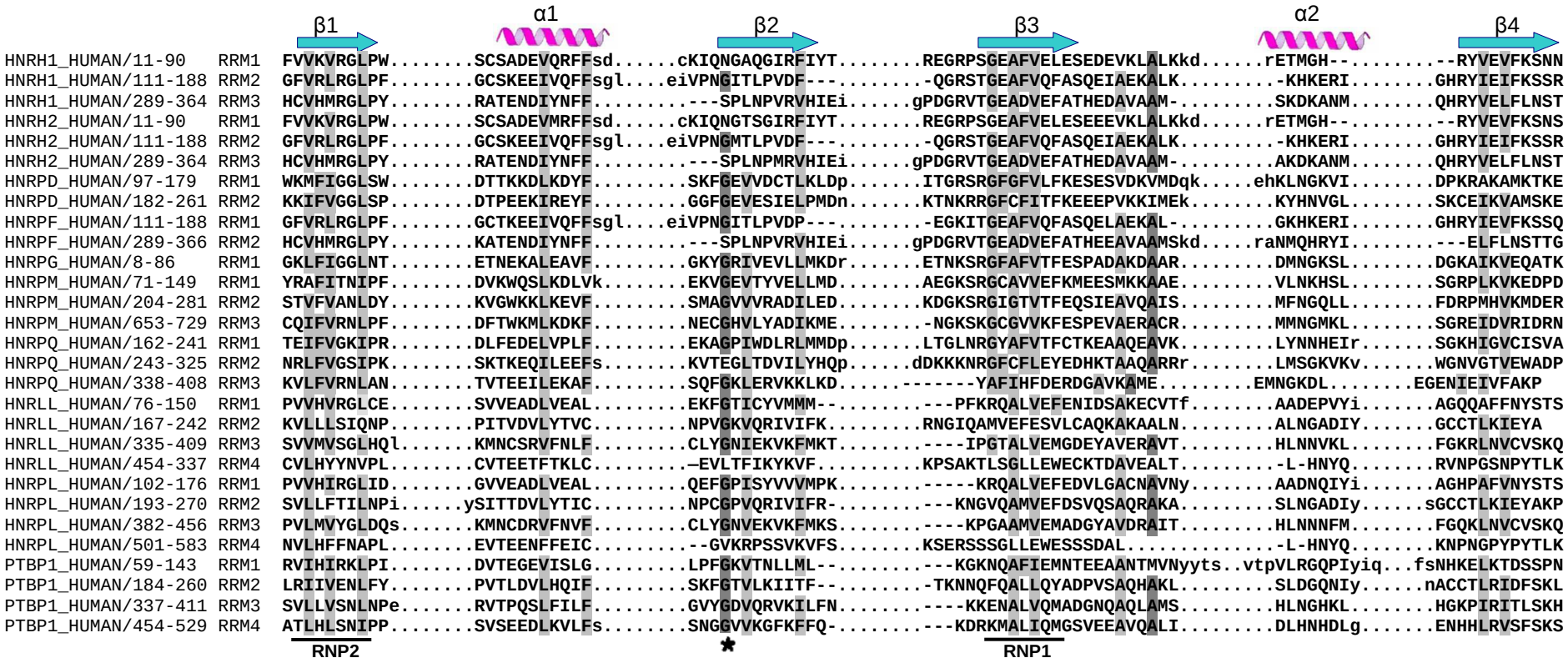


Figure 3.2 Multiple alignment of selected RRMs of human hnRNP proteins.
Sequences were aligned using the ClusterW program. Amino acids identical or similar are shaded by dark grey or light grey, respectively. Conserved secondary structure elements are indicated above the alignment. RNP1 and RNP2 motifs are marked below the alignment. The asterisk indicates the position of a conserved glycine residue important for secondary structure formation. The numbers on the left of the alignment indicate the first and last residues of the respective RRM.

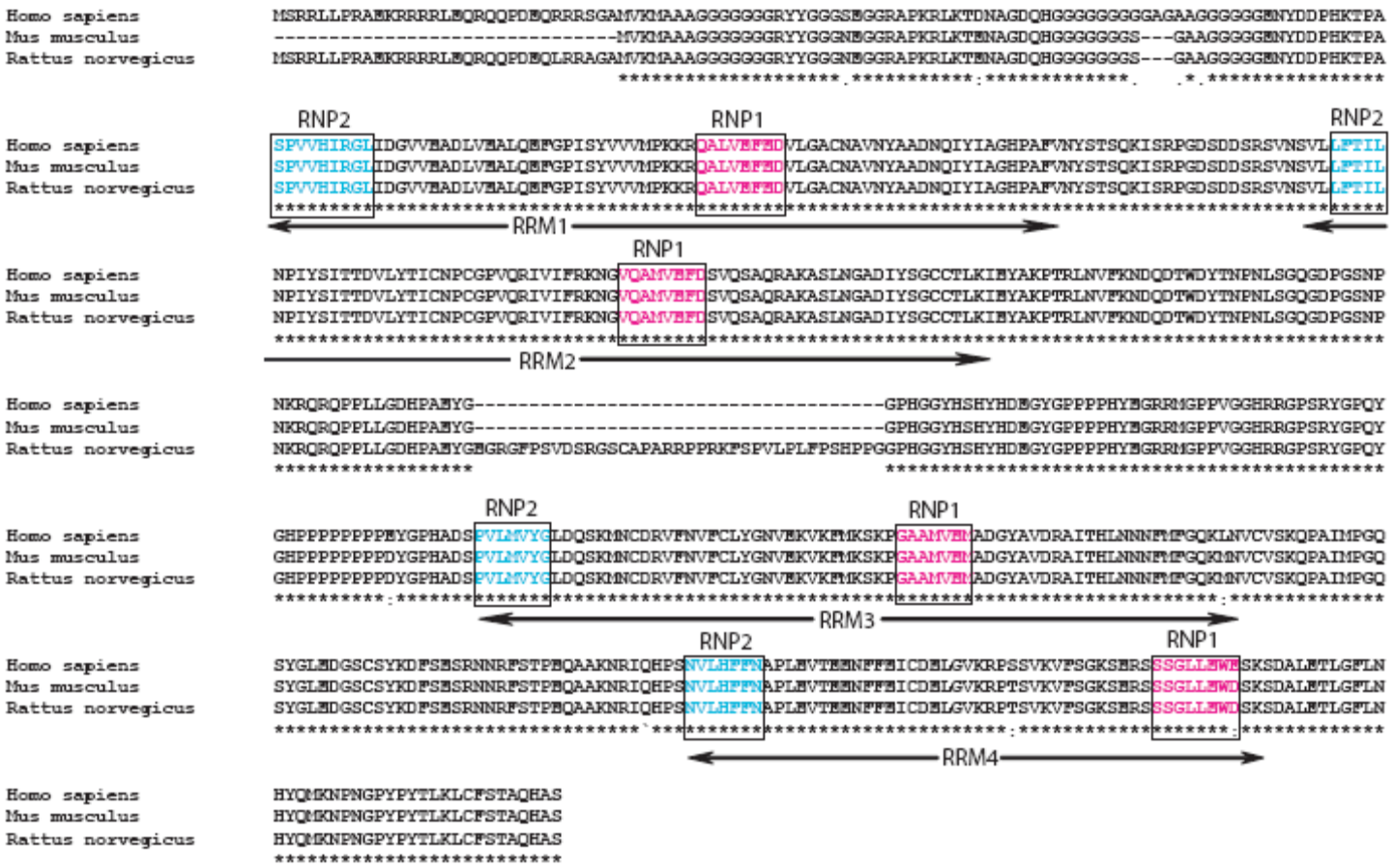


Figure 3.3 Amino acid sequence alignment of hnRNP L orthologues.
Full-length hnRNP L orthologues from three different organisms were aligned. The accession numbers of each sequence is given in the following: heterogeneous nuclear ribonucleoprotein L from Homo sapiens: NP_001524.2; Mus musculus: NP_796275.2; Rattus norvegicus: XP_001068144.1. Each RNA recognition motif is labeled by arrows below the alignment. The RNP1 and RNP2 motifs are indicated by black boxes, conserved amino acids in RNP1 and RNP2 are marked by light pink and blue, respectively. Conserved amino acids are indicated by asterixes.

hydrophobic interactions which are predominant at these positions are not disrupted.
The RRMs of both proteins, except RRM4, have another important hallmark of RNA-recognition motifs – the conserved glycine residues at position 25 which seems to be required for the turn into β2 sheet.
In collaboration with the group of Dr. J. Bujnicki (Warsaw, Poland) secondary structures of hnRNP L and LL were predicted. The prediction was based on a series of NMR structures of individual RRM domains of PTB in complex with RNA (Oberstrass *et al*., 2005). The model revealed that RRM domains of hnRNP L and hnRNP LL have a classical βαββα secondary structure. In addition, both proteins have extra β5 sheets in their RRM2 and RRM4.
Alignment of amino acid sequences of the full-length hnRNP L protein with its orthologues from mouse and rat shows strong sequence conservation (97% identity of human vs. mouse and 87% identity of human vs. rat) (Fig. 3.3). RRM1 and RRM2 show 100% identity, whereas RRMs 2 and 4 have several amino acid substitutions in comparison to the mouse and rat domains. The murine hnRNP L protein is shorter lacking 31 amino acids at the N-terminus, whereas in the rat protein the linker region between RRM2 and RRM3 is 38 amino acids longer than in the human and mouse hnRNP L which can result in a different spatial arrangement of the two RRMs.

3.2 Design of hnRNP L and hnRNP LL mutant derivatives

HnRNP L and hnRNP LL proteins are composed of several domains including four RRMs, the glycine-rich and proline-rich regions. In order to study the contribution of each individual domain to the protein's functions I designed a series of domain deletion derivatives based on structurally defined domain boundaries. In addition, several mutations were introduced in each individual RNA recognition motif in order to disrupt either the RNA-binding ability of a particular domain or inter-domain interaction between RRM3 and RRM4 (Fig. 3.4 A, B).

3.2.1 Deletion derivatives

Four deletion mutant derivatives were generated for hnRNP L and three for hnRNP LL. Either the N-terminal domains RRMs 1 and 2 or the C-terminal domains RRMs 3 and 4, or RRM1 and RRM4 were deleted to create truncated versions (L/LLΔN, L/LLΔC, L-173-502, LL-162-429). In addition, I deleted the glycine-rich region (LΔG) of the hnRNP L protein.

All constructs contain an N-terminal GST-tag and a C-terminal FLAG-tag allowing the usage of both tags for protein purification. In addition, a TEV cleavage site was inserted between GST and hnRNP L ORF to remove the GST-tag.

3.2.2 Point mutation derivatives

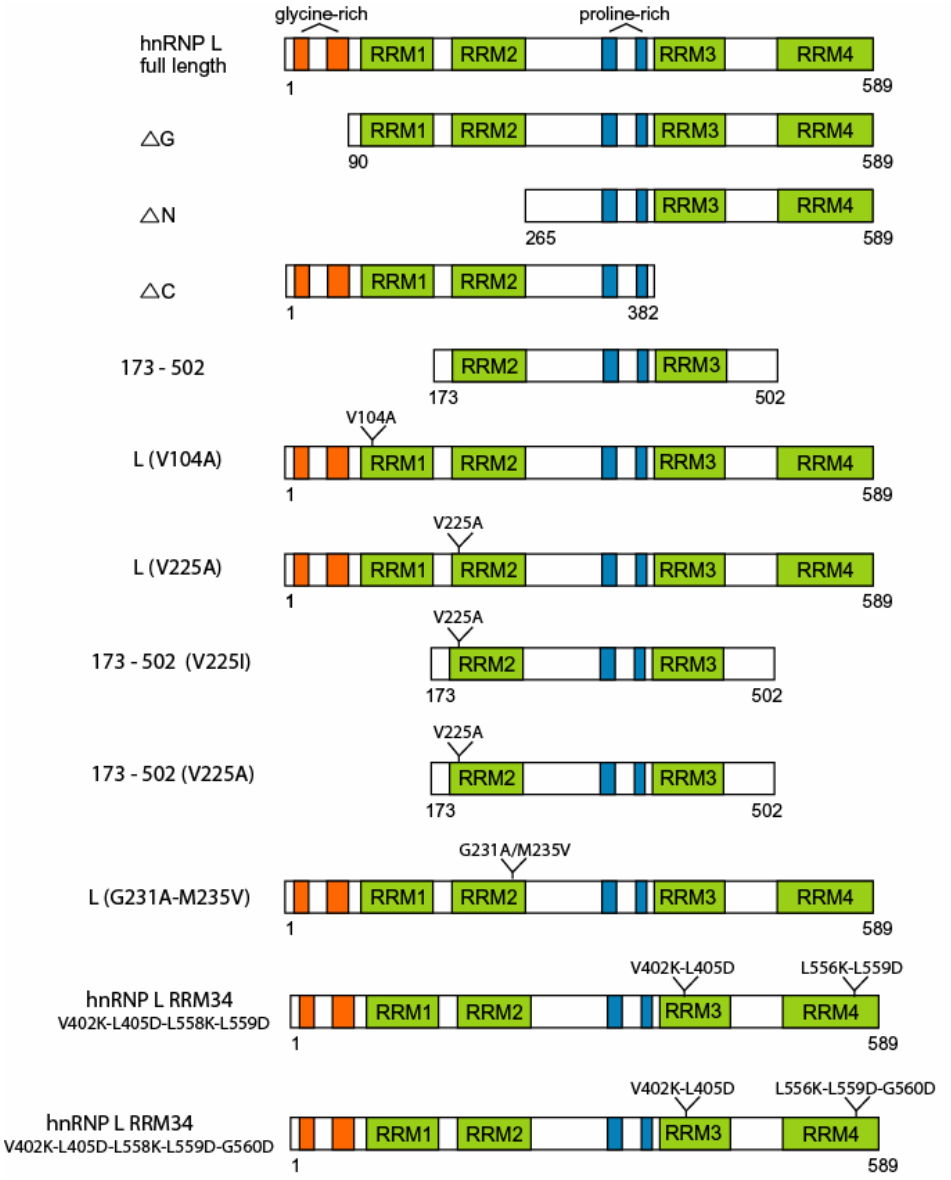
In an attempt to inactivate the RNA-binding properties of RRM1 and RRM2 of hnRNP L, single and double point mutations were introduced either in the RNP2 and RNP1 motifs itself or in the surrounding regions (Fig. 3.4A). Either conserved valine at position 104, or valine at position 225, or glycine and methionine at positions 231 and 235, respectively, were changed for A or V. The mutants are named according to the positions of the introduced point substitutions. Only one amino acid substitution, V225A, was made in two different contexts: in the full-length protein and in the deletion mutant containing RRMs 2 and 3; all other mutations were made only in the full-length protein context. All mutants are GST-tagged at the N-terminus and FLAG-tagged at the C-terminus and carry the TEV cleavage site for removal of the GST-tag.
Our collaborator Prof. F. Allain (Zurich, Switzerland) has shown that RRM3 and RRM4 interact with each other despite the long linker between the two RRMs (Skrisovska and Allain, 2008). In order to disrupt the inter-domain interaction charged amino acid residues were replaced by several hydrophobic side chains. The positions of the substitutions are indicated in the Figure 3.4A. These two constructs carry a His-tag at the N-terminus.

3.3 RNA-binding activity of hnRNP L, hnRNP LL and their deletion derivatives (EMSA)

3.3.1 RNA substrates

Previous experiments including an *in vitro* SELEX approach and various biochemical assays determined the binding specificity of hnRNP L and hnRNP LL (Hui *et al*., 2005; Topp *et al*., 2008). It has been shown that both proteins bind with high affinity to two types of RNA substrates: CA-repeat and CA-rich sequences.
In binding assays I used CA-repeat RNAs of different length: $(CA)_{10}$ and $(CA)_{20}$ RNAs as well as CA-rich RNA: GSTZ1. As a negative control I used an unspecific RNA substrate derived from pcDNA3 vector: GSTZ1-substitution carrying no high-score binding motifs for hnRNP L.

A



B

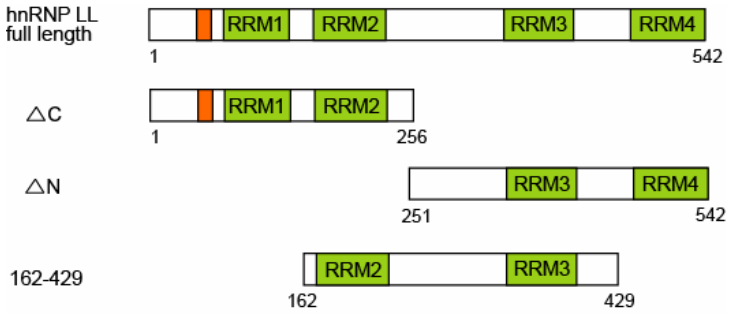


Figure 3.4 Generation of hnRNP L and hnRNP LL mutant proteins.
Schematic diagram of the domain structure of hnRNP L **(A)**, hnRNP LL **(B)** derivatives. The numbers below each construct depict amino acid positions; numbers above indicate amino acid substitutions. RNA recognition motifs are marked as green boxes, the glycine-rich and proline-rich regions are shown as orange and blue boxes, respectively.

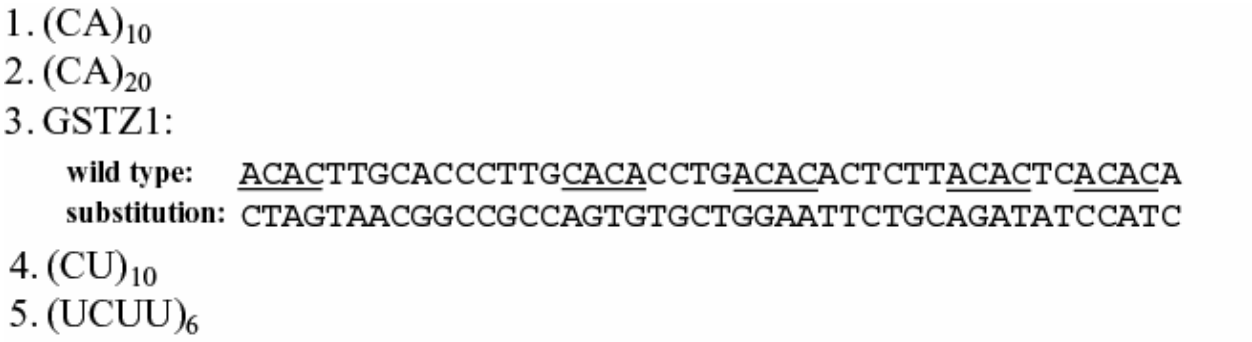


Figure 3.5 RNA binding substrates.
List of RNA substrates used for EMSA and filter binding assays. Number 1 and 2 are CA-repeat RNAs of 20 and 40 nucleotides in length, respectively. Number 3 is a CA-rich RNA GSTZ1 of 44 nucleotides and an unspecific RNA of the same length. Numbers 4 and 5 are CU- and UCUU-repeat RNAs of 20 and 24 nucleotides, respectively. The underlined sequences represent hnRNP L high-score binding motifs as determined by SELEX.

The $(CU)_{10}$ and $(UCUU)_6$ repeat RNAs were used to test the change in substrate specificity of hnRNP L upon the introduction of several amino acid substitutions in one of the RNA-binding motifs (Fig. 3.5)

3.3.2 Identification of the domains of hnRNP L and hnRNP LL critical for CA-repeat RNA-binding activity

In order to test the requirement of individual RRMs of hnRNP L and hnRNP LL for binding to CA-rich RNA, I performed Electromobility shift assays (EMSA) with full-length proteins and their truncated derivatives (Fig. 3.6).
The $(CA)_{10}$ and $(CA)_{20}$ substrates were internally labeled with [α^{32}P] CTP via T7 *in vitro* transcription. A constant concentration of RNA substrate was incubated with increasing amounts of recombinant GST-tagged hnRNP L or its deletion mutant derivatives. After incubation the samples were treated with heparin to prevent non-specific association. Formed complexes were resolved on a native polyacrylamide gel and detected by autoradiography.
As shown in Fig 3.6A the full-length protein binds to $(CA)_{10}$ repeat RNA very efficiently (lanes 1-5). Results with deletion mutants demonstrated that the binding affinities of full-length hnRNP L and hnRNP LΔC (lanes 6-10) are comparable, indicating that RRM1 and RRM2 are most important for binding of CA-repeat RNA. Mutant protein hnRNP L173-502 consisting of RRM2 and RRM3 showed a reduced binding affinity to $(CA)_{10}$ repeat RNA (lanes 16-19). Interestingly, hnRNP LΔN

mutant which includes only RRM3 and RRM4 showed no binding to $(CA)_{10}$ repeat RNA substrate (lanes 11-15). EMSA with the individual RRMs showed no detectable binding (data not shown), suggesting the importance of the combined action of at least two RRMs for efficient binding to CA-repeat RNA.

The EMSA with $(CA)_{20}$ repeat RNA and hnRNP L deletion derivatives (Fig. 3.5B) confirmed the observed results with $(CA)_{10}$ repeat RNA, indicating that two-domain combinations (RRM1/2, RRM2/3) are sufficient for efficient binding to CA-repeat-containing RNA. In contrast, the C-terminal domains (RRM3/4) did not show any RNA-binding activity (data not shown).

The band shift assay with hnRNP L173-502 mutant and $(CA)_{20}$ RNA (Fig. 3.6B, lanes 13-18) indicated that deletion of RRM1 and RRM4 slightly reduced binding and produced an aberrantly shifted complex that remained in the gel slot. This might be due to abnormal aggregation of this particular mutant.

Interestingly, both mutant proteins hnRNP LΔC and hnRNP L173-502 consisting of RRM2 and RRM1 or RRM3, respectively, showed a comparable reduction of binding affinity to $(CA)_{20}$ repeat RNA (lanes 7-18), suggesting that RRM1 and RRM3 might play similar roles in binding, probably in stabilisation of the RNA-protein complex.

The results of EMSA with full-length hnRNP L and two different CA-repeat RNAs [$(CA)_{10}$ vs. $(CA)_{20}$] indicated that two CA-repeat sequences of the different length are bound with different affinity by hnRNP L (Fig. 3.6 panel A: lanes 1-5; panel B: lanes 1-6), demonstrating that a CA-repeat element with the length of 20 nucleotides might be too short to interact optimally. Indeed, increase of the substrate length up to 40 nucleotides optimised the binding affinity, but an increase up to 64 nucleotides (data not shown) did not enhance the binding affinity further. Thus, I conclude that the minimal target site required for maximal affinity of a single molecule of hnRNP L is around 40 nucleotides.

In the case of hnRNP LL and its deletion derivatives EMSA (Fig. 3.6C) demonstrated that full-length protein binds to $(CA)_{10}$ repeat RNA with high binding affinity (lanes 1-3), comparable with hnRNP L, but deletion mutant derivatives completely lost their binding activity. No complexes could be detected (lanes 3-12) for the mutants, except hnRNP LL162-429 where binding affinity was very low.

The band shift assay with GST protein (Fig. 3.6C, lanes 13-14) confirmed that the shifts observed were not due to unspecific association of the GST-tag of all tested proteins with the RNA substrates.

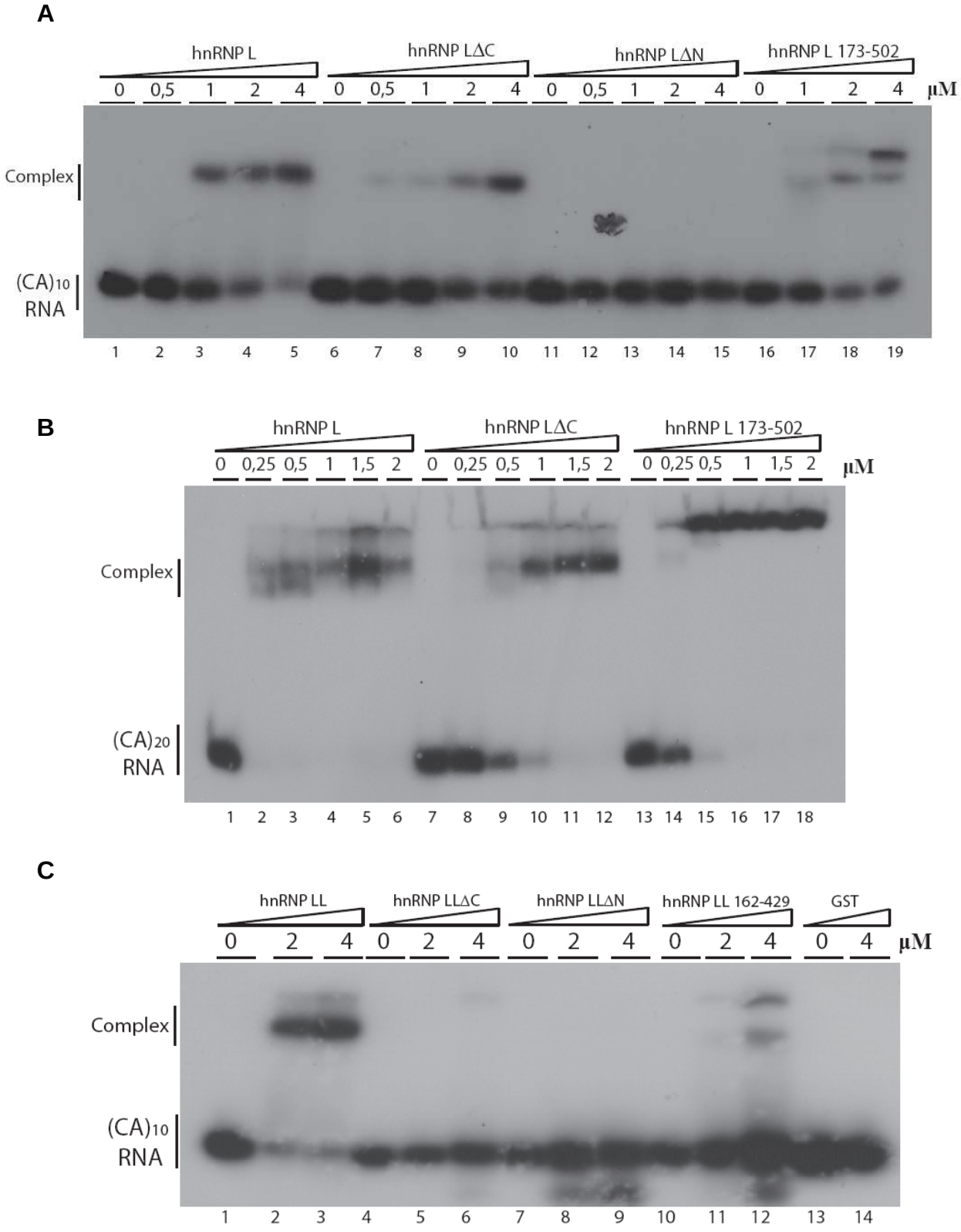


Figure 3.6 Analysis of binding of hnRNP L and hnRNP LL deletion derivatives to CA-repeat RNAs by EMSA.
Increasing amounts of hnRNP L and its deletion mutants **(A, B)** as well as hnRNP LL and its deletion mutants **(C)** were incubated with ^{32}P-labeled $(CA)_{10}$ **(A,C)** or $(CA)_{20}$ RNA **(B)**, analysed on a native polyacrylamide gel and visualised by autoradiography. The protein concentration is given above each gel. Positions of resulting RNA-protein complexes and free RNA are indicated on the left. The EMSA with GST protein serves as a negative control (**C**, lanes 13-14).

3.3.3 Identification of the domains of hnRNP L and hnRNP LL critical for binding activity to CA-rich RNA

In order to investigate the ability of hnRNP L and hnRNP LL mutant proteins lacking several RRMs to bind to a CA-rich substrate I performed electromobility shift assay with CA-rich RNA GSTZ1 labeled by *in vitro* transcription with [α^{32}P] CTP (Fig. 3.7).
I determined that full-length hnRNP L bound to GSTZ1 RNA with high affinity (Fig. 3.7A, lanes 1-4), although it was slightly reduced compared with the binding to CA-repeat RNAs. The removal of the N-terminal RRMs 3 and 4 domains caused a significant reduction in binding affinity to GSTZ1 RNA (Fig. 3.7A, lanes 5-8), suggesting that the presence of only two domains RRM1 and 2 is not enough for high-affinity binding to CA-rich elements. The mutant LΔN with RRMs 1 and 2 deleted completely lost the ability to bind GSTZ1 RNA (Fig. 3.7A, lanes 9-12). The deletion of RRMs 1 and 4 resulted in a decrease of binding to GSTZ1 RNA (Fig. 3.7A, lanes 13-18) compared to the LΔC mutant.
EMSA with hnRNP LL deletion derivatives and CA-rich GSTZ1 RNA (Fig. 3.7B) indicated that the full-length protein bound to GSTZ1 (Fig. 3.7B, lanes 1-4) with low binding affinity and deletion mutant proteins did not form detectable complexes with GSTZ1 RNA (Fig. 3.7B, lanes 5-16).

3.3.4 RRM2 of hnRNP L is primarily responsible for high-affinity RNA binding

My previous band shift experiments indicated that the high-affinity RNA binding by hnRNP L appears to require a combination of two RRMs (1/2 or 2/3), with RRM2 being present in both variants. I decided to test the function of hnRNP L RRM2 in RNA binding by introducing a single point mutation in RRM2 in a truncated two-domain context hnRNP L173-502 (Fig. 3.4A). The amino acid valine at position 225 was predicted to be responsible for the recognition of high-score binding motif by RRM2 although it is located neither in RNP1 nor in RNP2, but upstream of RNP1. Taking in to account that the predicted secondary structure of hnRNP L was based on PTB structure and PTB has isoleucine at the same position, was decided to change V225 of hnRNP L to either isoleucine or alanine. The binding to three different substrates was investigated using band shift assays. Specific RNA binding was assessed by utilising an $(CA)_{10}$ repeat RNA substrate (Fig. 3.8A). Altered binding specificity was measured with $(CU)_{10}$ and $(UCUU)_6$ repeat RNAs (Fig. 3.8B, C, respectively).

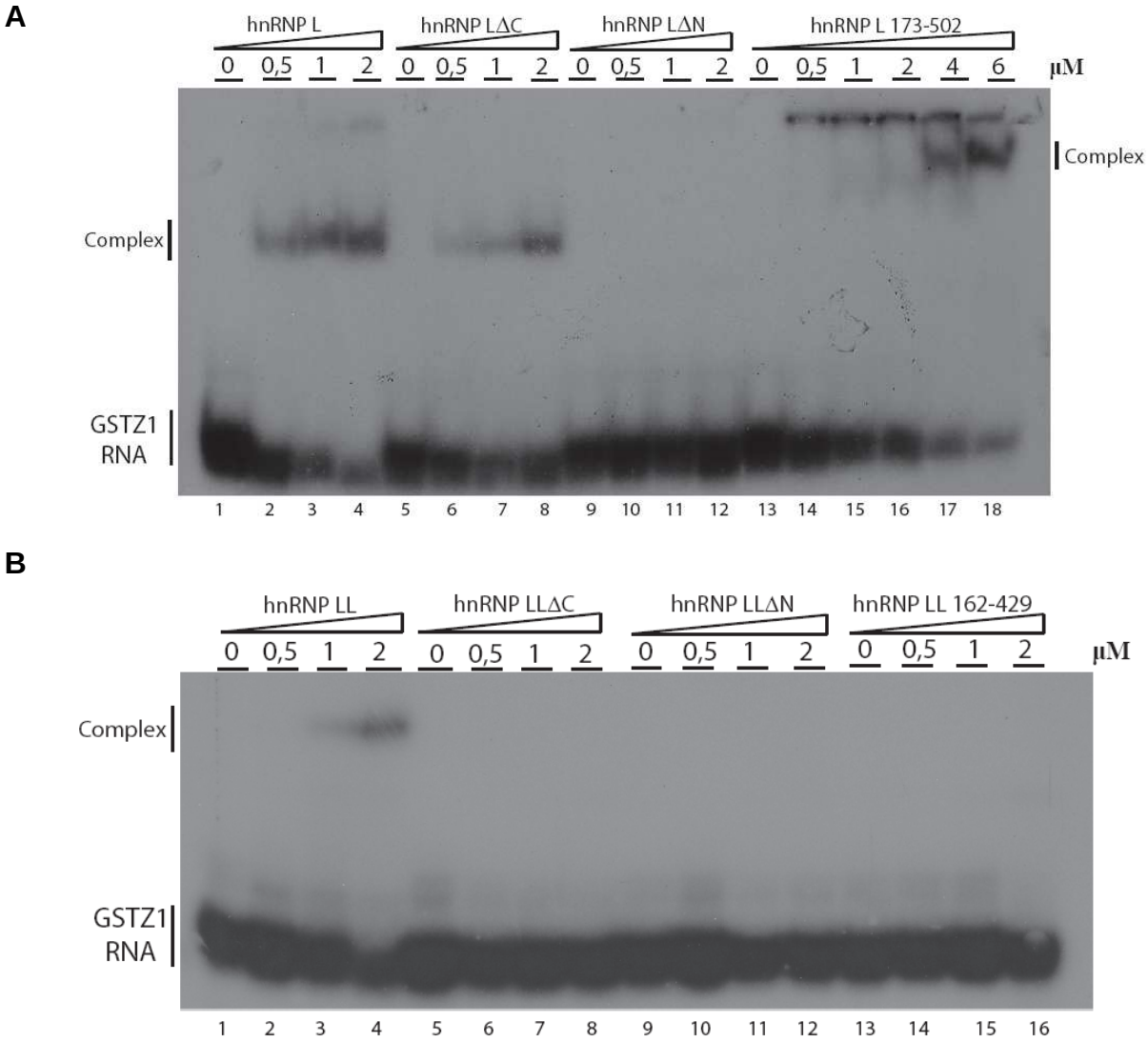


Figure 3.7 Analysis of binding of hnRNP L and hnRNP LL deletion derivatives to CA-rich RNA by EMSA.
Constant amount of ^{32}P-labeled GSTZ1 RNA was incubated with increasing amounts of hnRNP L deletion mutants **(A)** and hnRNP LL deletion mutants **(B)**. Protein-RNA complexes and unbound RNA were separated by electrophoresis on a native polyacrylamide gel and visualised by autoradiography. The protein concentrations are given above each gel. Positions of resulting RNA-protein complexes and free RNAs are indicated on the left and right.

In this experiment, a fixed amount of radiolabeled RNA was incubated with increasing amounts of hnRNP L173-502, hnRNP L173-502(V225I), hnRNP L173-502(V225A) as well as with full-length hnRNP L and PTB. Protein/RNA complexes were resolved by gel electrophoresis.

HnRNP L173-502 bound to CA-repeat RNA with high affinity which corresponds to results previously observed (Fig. 3.6A, lanes 1-6). The association of two proteins bearing amino acid substitutions (V225I and V225A) with $(CA)_{10}$ RNA was completely abolished (Fig. 3.8A, lanes 7-17), indicating the important role of valine

at position 225 for CACA motif recognition by RRM2 and the essential role of the whole RRM2 domain for high-affinity binding to RNA.

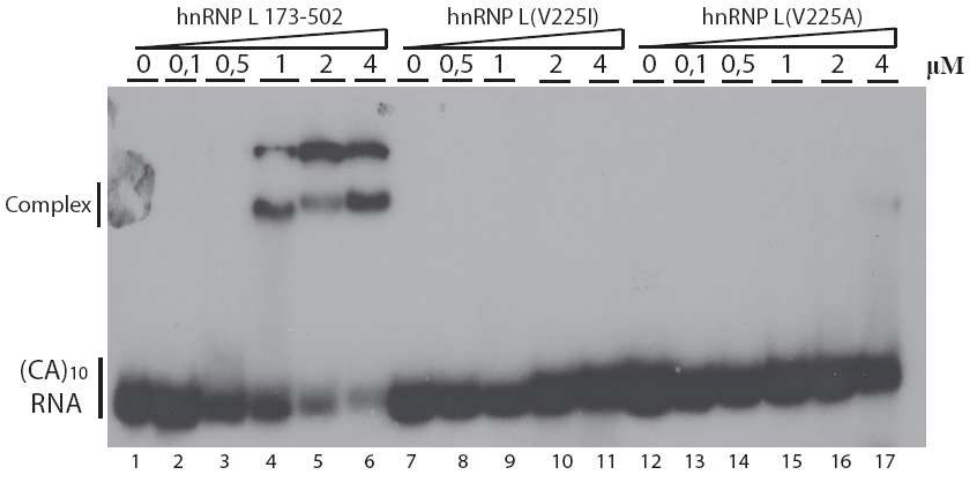
Since in hnRNP L173-502(V225I) the substitution was made based on PTB alignment and Val was changed to Ile, we suggested that this mutation might possibly change the substrate specificity of the mutated protein towards the recognition of PTB specific motifs: CU and UCUU.

In order to test this hypothesis, I performed EMSA with ^{32}P-labeled $(CU)_{10}$ (Fig 3.8B) and $(UCUU)_6$ (Fig. 3.8C) RNAs and mutant proteins as described above.

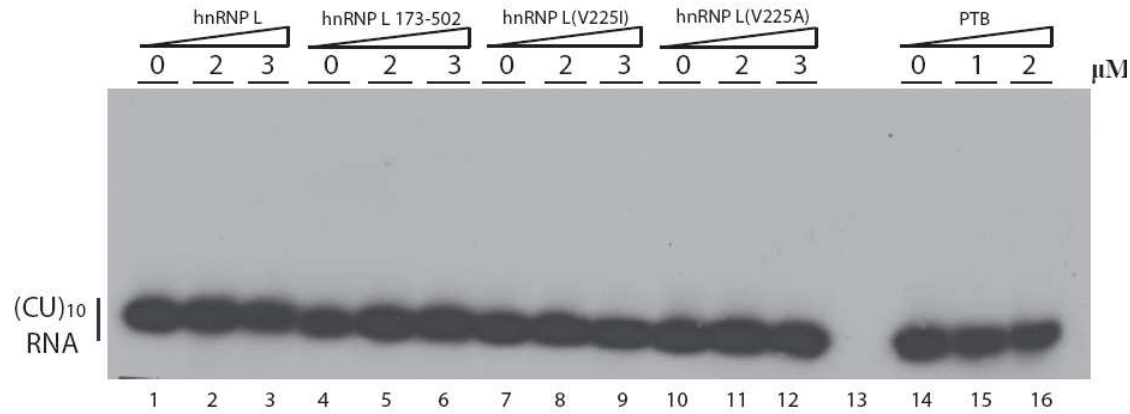
Expectedly, the full-length hnRNP L and hnRNP L173-502 did not show any binding activity with either RNA substrates (Fig. 3.8B and C, lanes 1-6), confirming the protein's preference for CACA motifs as a substrate. HnRNP L173-502(V225I) and hnRNP L173-502(V225A) (Fig. 3.8B and C, lanes 7-12) did not associate with CU and UCUU repeats either, indicating that the substitution of Val at position 225 to Ile or Ala did not change the substrate specificity of the protein.

Interestingly, positive control PTB bound with high affinity only to $(UCUU)_6$ RNA (Fig. 3.8B, lanes 14-16 vs. panel C, lanes 13-15), confirming previous observations pointing out the UCUU sequence as the preferred RNA binding site of PTB (Perez *et al*., 1997).

A



B



C

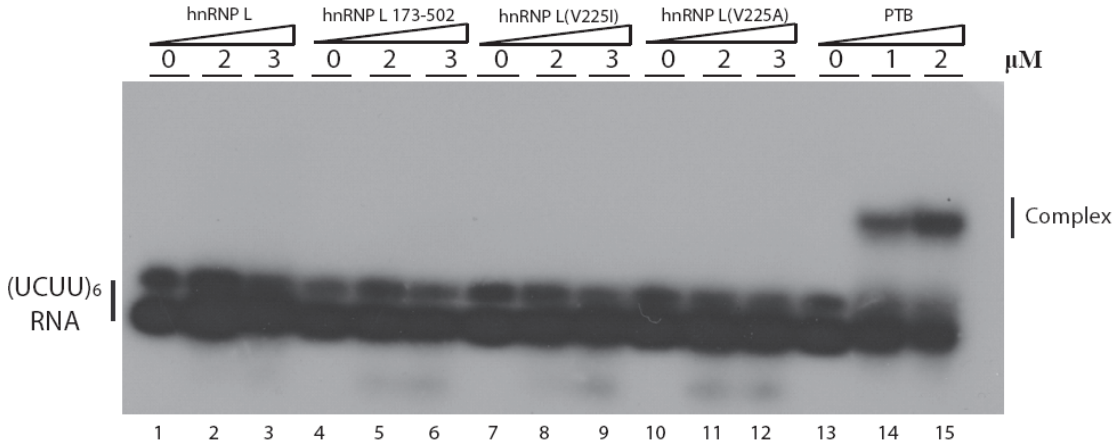


Figure 3.8 RRM2 of hnRNP L is an essential domain for binding activity.
A fixed amount of ^{32}P-labeled $(CA)_{10}$ **(A)**, $(CU)_{10}$ **(B)** and $(UCUU)_6$ **(C)** RNAs was incubated with the indicated concentration of full-length or mutant hnRNP L proteins and PTB. Protein-RNA complexes and unbound RNA were separated on a non-denaturing polyacrylamide gel. Autoradiograms of the gels are shown.

3.3.5 Unspecific binding activity of hnRNP L and hnRNP LL mutant proteins

Non-specific RNA-binding activity of all hnRNP L and hnRNP LL deletion mutants was tested by EMSA with unspecific radiolabeled GSTZ1-substitution RNA. No complexes were observed with any of the tested proteins (Fig. 3.9).

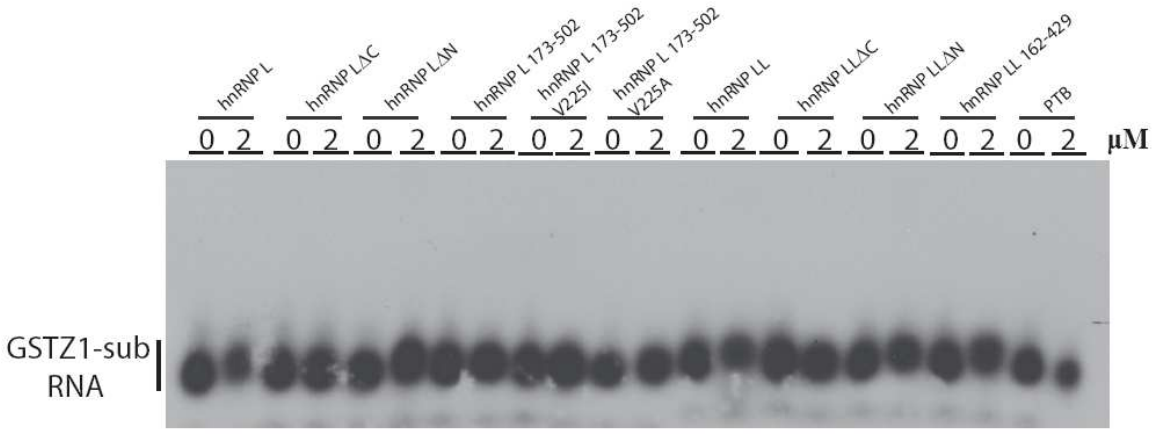


Figure 3.9 HnRNP L and hnRNP LL mutant derivatives show no unspecific binding activity.
At a concentration of 2µM each protein (wild-type and mutant) was incubated with radiolabeled non-specific GSTZ1-substitution RNA and then separated on a 6% non-denaturing polyacrylamide gel. The position of unbound RNA is indicated on the left. Autoradiogram of the gel is shown.

3.4 RNA-binding activity of hnRNP L, hnRNP LL and their mutant derivatives (filter binding assays)

Filter binding assays were used to measure the equilibrium dissociation constant of hnRNP L mutants (Fig. 3.10) for CA-repeat and CA-rich RNAs in order to assess quantitatively the contributions of the individual domains and the mutated residues to RNA binding.

In the experiment, a fixed amount of radiolabeled $(CA)_{20}$ or GSTZ1 RNAs were pre-incubated with increasing amounts of hnRNP L mutants and stable binding was measured as the radioactivity retained on nitrocellulose membranes. The amount of radioactivity in each dot was quantified using a Molecular Dynamics Phosphorimager. The dissociation constant (K_D) values were estimated by the half-maximal binding observed in filter binding assay (Fig. 3.10).

The full-length hnRNP L showed comparable high binding efficiencies to both substrates: K_D 72nM for $(CA)_{20}$ vs. K_D 119nM for GSTZ1. The binding of the ΔG mutant lacking only the N-terminal glycine-rich region to both RNAs was as efficient as for the wild-type protein, suggesting that this region does not play an important role in CA-repeat or CA-rich RNA binding. The ΔN protein showed no detectable binding to either RNAs, thus the K_D value could not be calculated. The binding affinity of the ΔC for $(CA)_{20}$ RNA was slightly reduced compared to full-length protein (K_D for LΔC 180nM), but binding to GSTZ1 was significantly decreased and K_D value could not be determined.

The single substitution of valine at position 104 in RNP2 of RRM1 [L(V104A)] slightly reduced binding to $(CA)_{20}$ and GSTZ1 RNA binding (K_D 119nM and 140nM, respectively), suggesting that this mutation does not directly influence RNA binding of the RRM1, or it is compensated by the other RRMs. In contrast, substitution of valine at position 225 in the β2-sheet of RRM2 yielded a protein [L(V225A)] that bound to CA-repeat RNA with reduced affinity (K_D 213nM) and to CA-rich RNA with even lower affinity (K_D 529nM). Another protein with a double amino acid substitution in RNP1 of RRM2 [L(G231A-M235V)] exhibited the same slightly reduced affinity for $(CA)_{20}$ (K_D 213nM) and GSTZ1 (K_D 169nM) compared to L(V104A) protein.

Interestingly, the amino acid mutations introduced to disrupt inter-domain interactions between RRM3 and RRM4 [hnRNP L (V402K-L405D-L556K-L559D) and hnRNP L (V402K-L405D-L556K-L559D-G560D)] reduced binding of the proteins to $(CA)_{20}$ (K_D 265nM and 230nM, respectively) and strongly reduced the binding to GSTZ1 (K_D is too low or not measurable).

The data of the filter binding assays are consistent with the EMSA results and demonstrate synergy between the RRM1 and RRM2 in binding to CA repeats.

HnRNP L, ΔN and ΔC proteins showed consistent binding to $(CA)_{20}$ and GSTZ1 RNA in both assays.

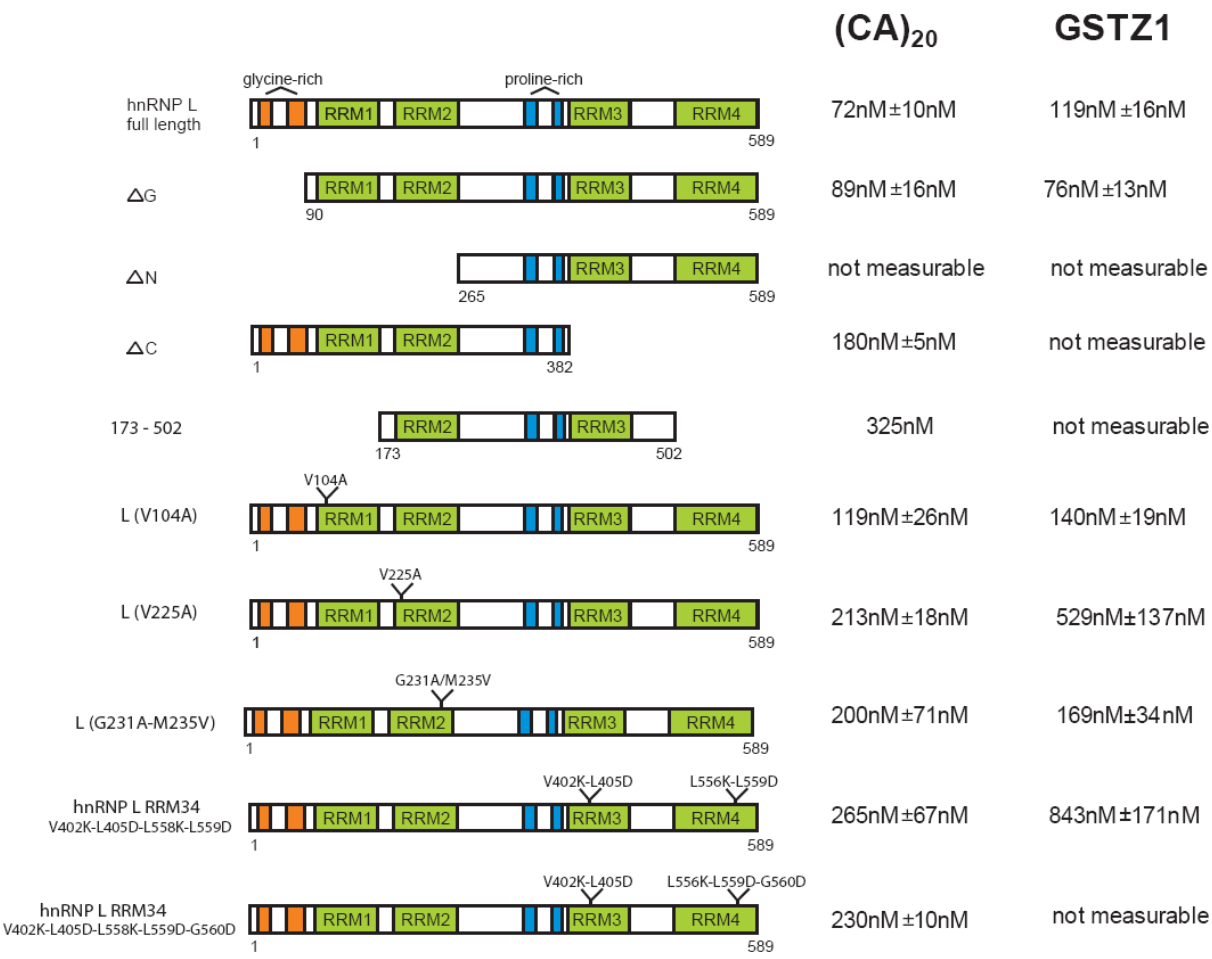


Figure 3.10 Dissociation constants (K_D) of hnRNP L mutant proteins for binding to CA-repeat and CA-rich RNAs.
To compare binding affinities, 90 fmol of labeled $(CA)_{20}$ or GSTZ1 RNA was incubated with 50, 90, 180, 360, 720 nM of hnRNP L mutants. All reactions contained 0,2 mg/ml BSA. Bound and free RNA were quantified and K_D values calculated based on the mean of three independent experiments. The GST-tag of mutant proteins used for filter binding had been removed by TEV protease cleavage.

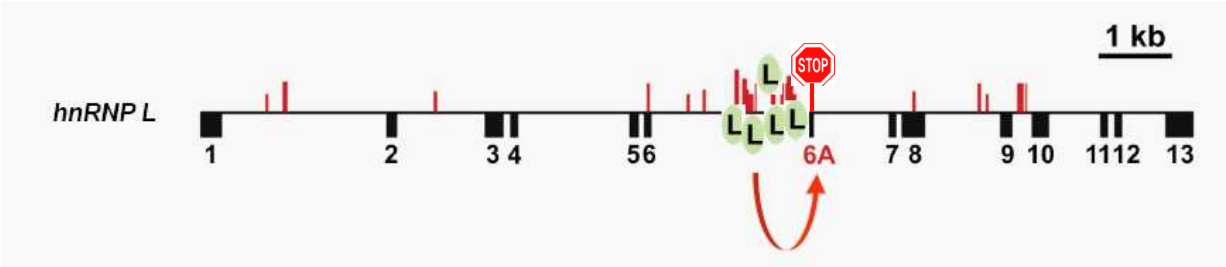
3.5 HnRNP L and CA-rich cluster: two elements of autoregulation

3.5.1 CA-rich cluster

In our laboratory we discovered an autoregulation mechanism for hnRNP L protein (Rossbach *et al*., 2009). By binding to a very unusually long CA-rich cluster within

its own pre-mRNA, hnRNP L activates the poison exon 6A inclusion, leading to the introduction of a premature termination codon and resulting in NMD (Fig. 3.11A). The entire region of 2 kb in intron 6 is highly conserved, close to ultraconservation in the strict sense (Lareau *et al*., 2007). The CA-rich cluster is a dense sequence of twenty five high-score hnRNP L binding motifs extending over more then 800 nt within this highly conserved intronic region (Fig. 3.11B).

A



B

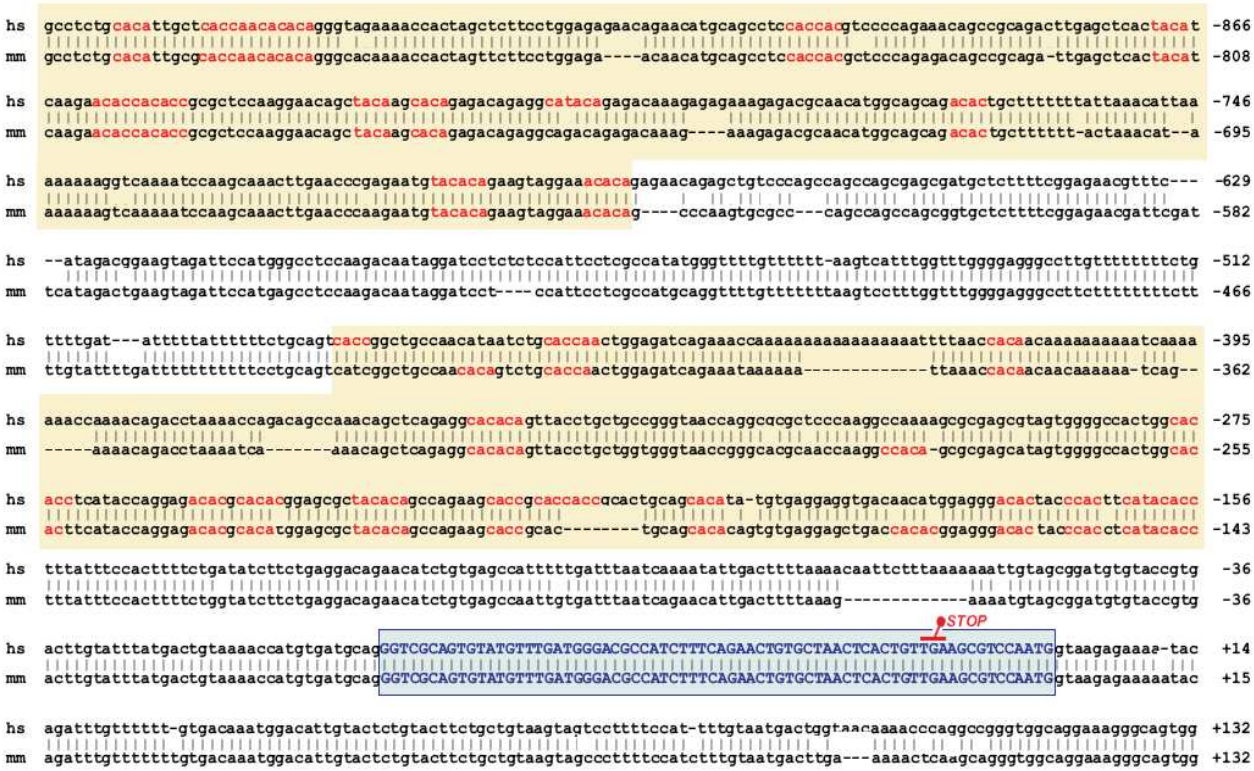


Figure 3.11 CA-rich cluster within a highly conserved region in the hnRNP L gene.
(A) Exon-intron structure of the human hnRNP L gene. Red bars above the line represent hnRNP L binding motifs, with their height corresponding to their score (for how these scores were derived, see Hung *et al*., 2008). **(B)** Alignment of CA-rich cluster and intron 6 sequences (human (hs) vs. mouse (mm)). Numbering refers to nucleotide positions relative to the exon 6A. Exon 6A boxed in blue; the two parts of the CA cluster are highlighted by light violet shading, with the CA motifs in red. “STOP” codon indicated in red (adapted from Rossbach *et al*., 2009).

3.5.2 HnRNP L binds to CA-rich cluster

Based on the SELEX-derived RNA-binding specificity of hnRNP L, we suggested that hnRNP L binds to the CA-rich cluster of its own pre-mRNA with high affinity.
To test this hypothesis, I performed *in vitro* pull-down assays (Fig. 3.12). The CA-rich cluster RNA was *in vitro* transcribed in two parts: cluster A – the 5'-terminal portion (300 nt) and cluster B – 3'-terminal portion (346 nt). Both RNAs were 3'-biotinylated and incubated in HeLa nuclear extract. HnRNP L binding was then assessed by pulling down the biotinylated RNAs, using neutravidin-agarose followed by Western blot analysis of the pull-down material with the hnRNP L-specific monoclonal antibody, 4D11. As specificity controls, a 5'-biotinylated $(CA)_{32}$ RNA, cluster Δ (an unrelated 3'-biotinylated RNA) and beads alone were used.
As shown in Fig 3.12 both parts of the CA-rich cluster RNA efficiently pulled down hnRNP L from HeLa nuclear extract (between 5 and 10% of total hnRNP L at 70 mM of KCl, lanes 1-3) as well as $(CA)_{32}$ RNA (lanes 4), whereas pull-down with the control cluster Δ RNA and beads alone did not recover significant hnRNP L amount (lanes 5, 6). Together, the data indicate that hnRNP L protein binds specifically to the conserved CA cluster.
In parallel, I tested the binding of the paralog protein hnRNP LL to the CA-rich cluster (Fig. 3.12). HnRNP LL could also be specifically pulled down from HeLa nuclear extract by both parts of hnRNP L CA cluster (lanes 2, 3) and by $(CA)_{32}$ RNA (lane 4) at 70 mM KCl.
Next, the salt dependency of CA cluster recognition by hnRNP L and hnRNP LL was tested (Fig. 3.12). I incubated biotinylated cluster A, cluster B, cluster Δ, and $(CA)_{32}$ RNAs in HeLa nuclear extract under conditions of different stringency (KCl was adjusted to 70, 200, 300 or 400 mM), followed by neutravidin-agarose selection, washing and Western blot analysis with monoclonal anti-hnRNP L (4D11) and anti-hnRNP LL antibodies.
The data presented in Fig. 3.12 illustrated the differences in affinity of hnRNP L and hnRNP LL for CA cluster RNA within the salt concentration range mentioned above. The strongest association of hnRNP L and hnRNP LL with both parts of the CA cluster was detected at 70 mM KCl. At 200 mM KCl the binding of hnRNP L to cluster A and B was slightly affected, whereas for hnRNP LL binding was already completely abolished. Increase of salt concentration up to 300 mM KCl prevented the association of hnRNP L with cluster B RNA and significantly reduced association with cluster A RNA. At 400 mM KCl hnRNP L failed to bind of the CA-rich cluster. Interestingly, binding of hnRNP L and hnRNP LL to $(CA)_{32}$ RNA was stable up to 400 mM KCl and above (data not shown).

In summery, biotin pull-down assay of endogenous hnRNP L and hnRNP LL revealed salt-dependent binding of both proteins to CA cluster RNA, with hnRNP L being more strongly associated with CA cluster.

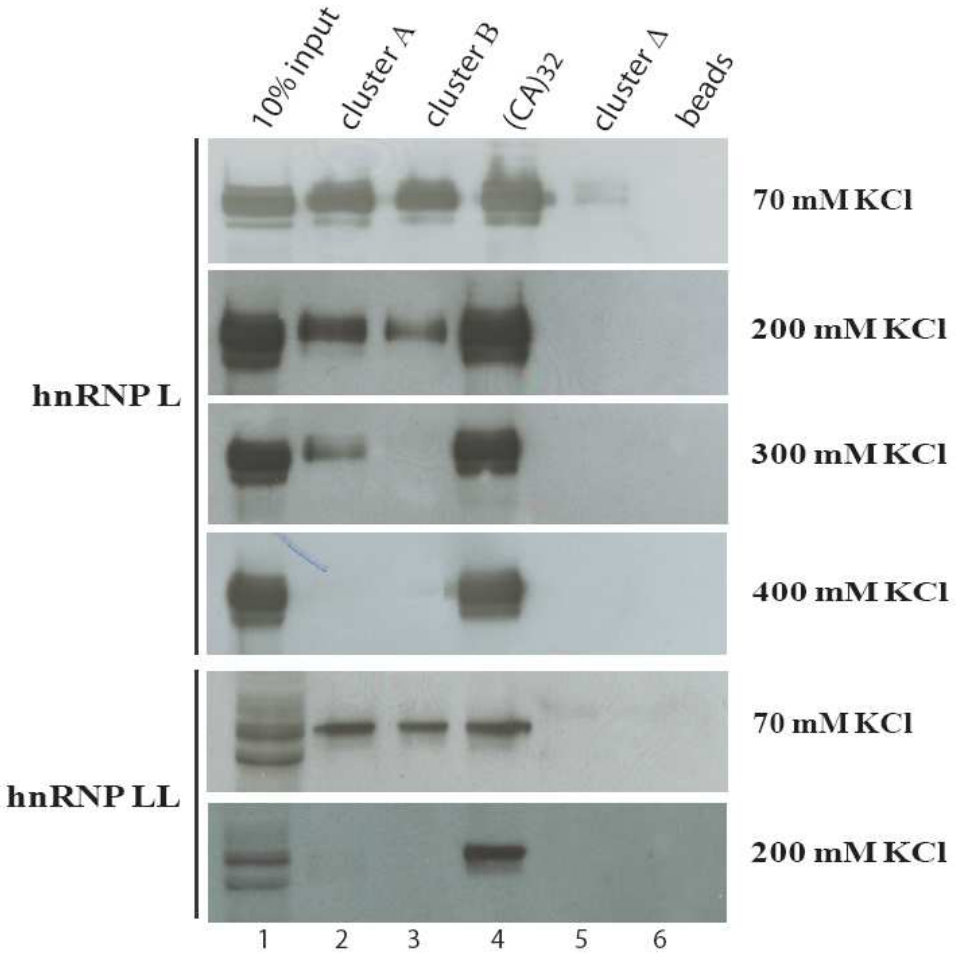


Figure 3.12 Salt-dependent binding of hnRNP L and hnRNP LL to CA-rich cluster.
Biotinylated cluster A and B RNAs (lanes 2 and 3, respectively), $(CA)_{32}$ RNA (lane 4) and cluster Δ RNA (lane 5) were incubated in HeLa nuclear extract under conditions of different stringency (indicated on the right), followed by neutravidin-agarose pull-down. The bound proteins were analysed by Western blot with anti-hnRNP L monoclonal antibody (4D11) or anti-hnRNP LL antibody. As input 10% of the HeLa nuclear extract was loaded (lane 1). As control, beads were incubated with HeLa nuclear extract without RNA (lane 6).

3.5.3 Cooperative binding of hnRNP L to CA-rich cluster

The results described above indicate that both parts of the CA cluster bind hnRNP L protein with high affinity. Taking into account that cluster A has eleven and cluster B fourteenth high-score binding sites (Fig. 3.11B in red) we postulated that each single high-score binding motif could bind one molecule of hnRNP L, and that the entire CA cluster may bind several hnRNP L molecules, in a cooperative fashion.

To investigate this hypothesis, I performed EMSA with fixed amount of ^{32}P-labeled cluster A or cluster B RNAs and increasing amounts of baculovirus-expressed His-tagged recombinant hnRNP L.

As shown in Fig. 3.13, a shift for both RNA substrates could be detected with 10 – 20-fold molar excess of recombinant hnRNP L protein (lanes 2-3, 10-11). The three bands with different migration behaviour in lane 1 probably represent different secondary structures of cluster A RNA. Binding at higher protein concentration resulted in the formation of complexes that did not enter the gel. It appeared at the lowest protein concentration and became more pronounced as the protein concentration was increased. Lack of discrete intermediate complexes suggested a high degree of cooperativity.

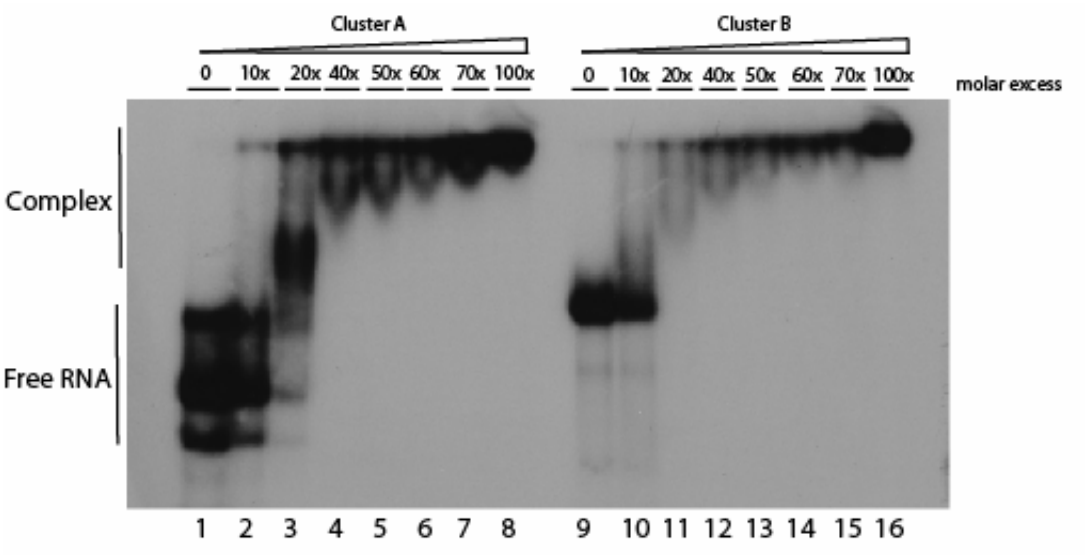


Figure 3.13 Binding of hnRNP L to CA cluster: evidence for cooperative interactions.
Radiolabeled cluster A RNA (lanes 1-8) and cluster B RNA (lanes 9-16) were incubated with increasing amounts of recombinant hnRNP L. Protein-RNA complexes and unbound RNA were analysed on a native acrylamide gel and visualised by autoradiography. The molar excess of protein over RNA is given above the gel. Positions of resulting RNA-protein complexes and free RNAs are indicated on the left.

Being aware that the lack of intermediated complex could result from "disproportionation" artefacts (Kleinschmidt *et al*., 1991), in which there is preferential loss of singly occupied complex compared with fully occupied complex during electrophoresis and that this could also affect the quantitation of the binding affinity, I also analysed protein-RNA complex formation, using glycerol gradient centrifugation (Fig. 3.14).

The ^{32}P-labeled cluster A or cluster B RNAs were pre-incubated with different amounts of recombinant hnRNP L and then separated by centrifugation on a glycerol gradient.

The RNA analysis of the gradient showed that free RNA of cluster A (Fig. 3.14A) and free RNA of cluster B (Fig. 3.14B) peaked in the same fractions (2 to 4). Only one peak in fractions 5-6 or 6-7 for cluster A and B, respectively, was observed in the presence of large excess of recombinant hnRNP L (1:50), indicating the sedimentation rate of protein-RNA complex. The further increase in excess of the protein did not shift the peak (data not shown). In the presence of small excess of recombinant hnRNP L (1:5), two peaks could be detected: first peak in the fractions 3 or 3-4 for cluster A and B, respectively, corresponding to sedimentation rate of free RNA; second peak in the fractions 5-6 or 6-7 for cluster A and B, respectively, corresponding to protein-RNA complex. No intermediate complexes were detected.

The glycerol gradient centrifugation of unrelated cluster Δ RNA and recombinant hnRNP L (1:50 molar excess of the protein) (Fig. 3.14(C)) showed a peak in the fractions 2-4, which is similar to the distribution of free cluster A and B RNAs, indicating that hnRNP L does not interact specifically with cluster Δ RNA and that the shifts observed for both cluster RNAs were specific.

Taking together the results from biotin pull-down assay, EMSA and glycerol gradient centrifugation analysis I conclude that hnRNP L protein binds to the CA-rich cluster in a highly cooperative manner.

3.5.4 Binding of hnRNP L to short RNAs

In order to characterise hnRNP L binding to cluster B in more detail, I divided this region into transcripts of 30 nucleotides (Fig. 3.15A). Each short RNA carries either one or two or three high-score binding sites and assayed binding with recombinant His-tagged hnRNP L.

The short RNAs (B1-B8, A6) were internally labeled with [α^{32}P]CTP via T7 *in vitro* transcription. A constant amount of RNA substrate was incubated with increasing amounts of recombinant hnRNP L. After incubation the resulting complexes were resolved on a native polyacrylamide gel and detected by autoradiography.

As shown in Fig. 3.15(B) the short RNAs containing a single high-score binding site (B3, B4, B5 are shown) failed to form a detectable complex with hnRNP L protein, whereas short RNAs with either two (A6) or three (B6-B8) high-score binding motifs bound hnRNP L with high affinity.

EMSA with short RNAs revealed a general correlation between high-affinity binding and amount of binding sites, indicating that high-affinity binding of hnRNP L requires at least two high-score motifs within the RNA substrate.

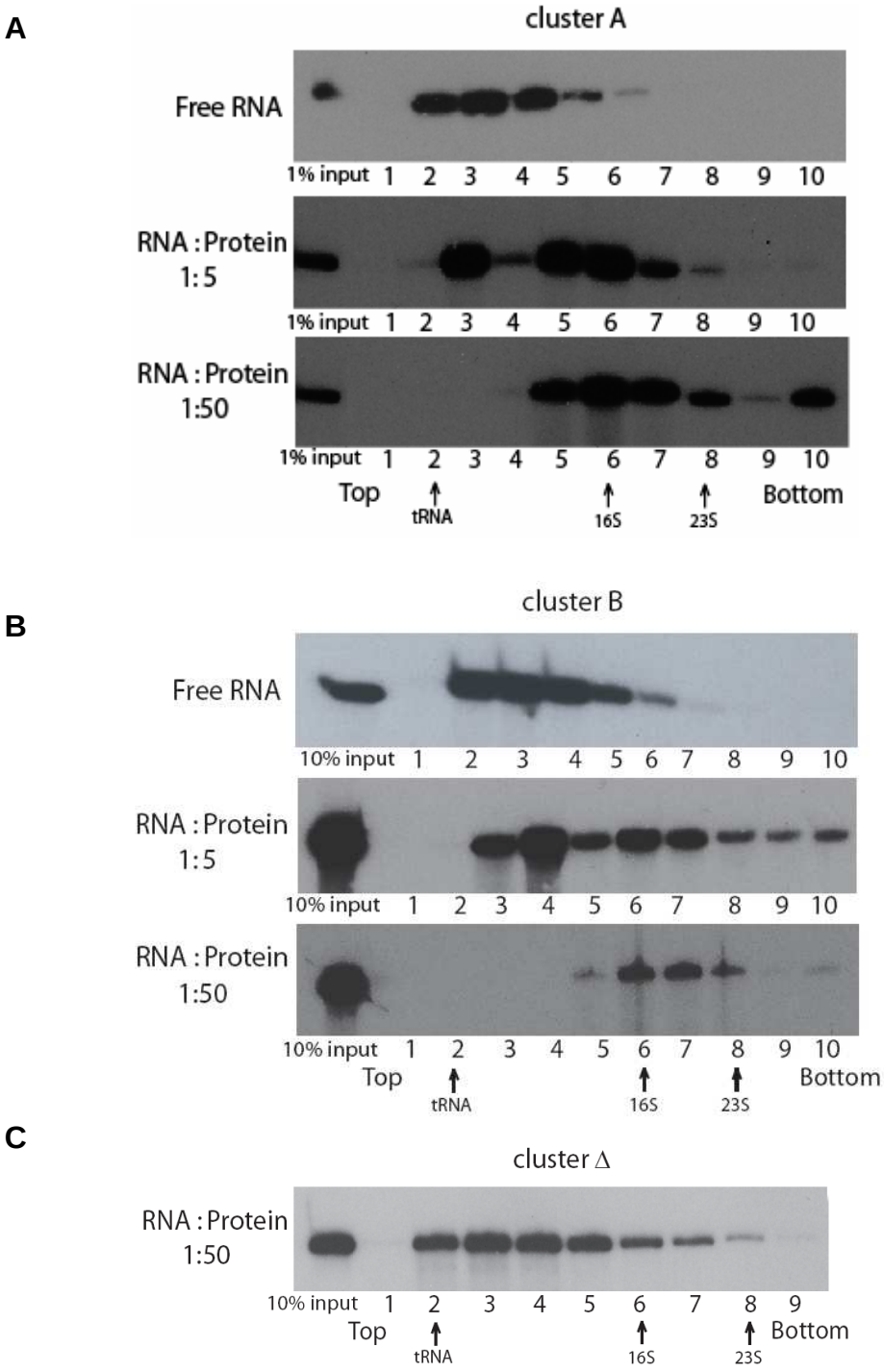


Figure 3.14 Glycerol gradient fractionation of protein-RNA complexes.
^{32}P-labeled cluster A RNA **(A)** or cluster B RNA **(B)** was incubated with recombinant hnRNP L protein in a molar ratio as indicated on the left. The resulting complexes were separated on a 10-30% glycerol gradient. 10 fractions were collected from the top to bottom. RNA was recovered from each fraction, separated on a 10% denaturing gel and visualised by autoradiography. Arrows indicate the sedimentation positions of markers from a gradient run in parallel. Gradient centrifugation of cluster Δ RNA and hnRNP L is shown on panel **(C)**.

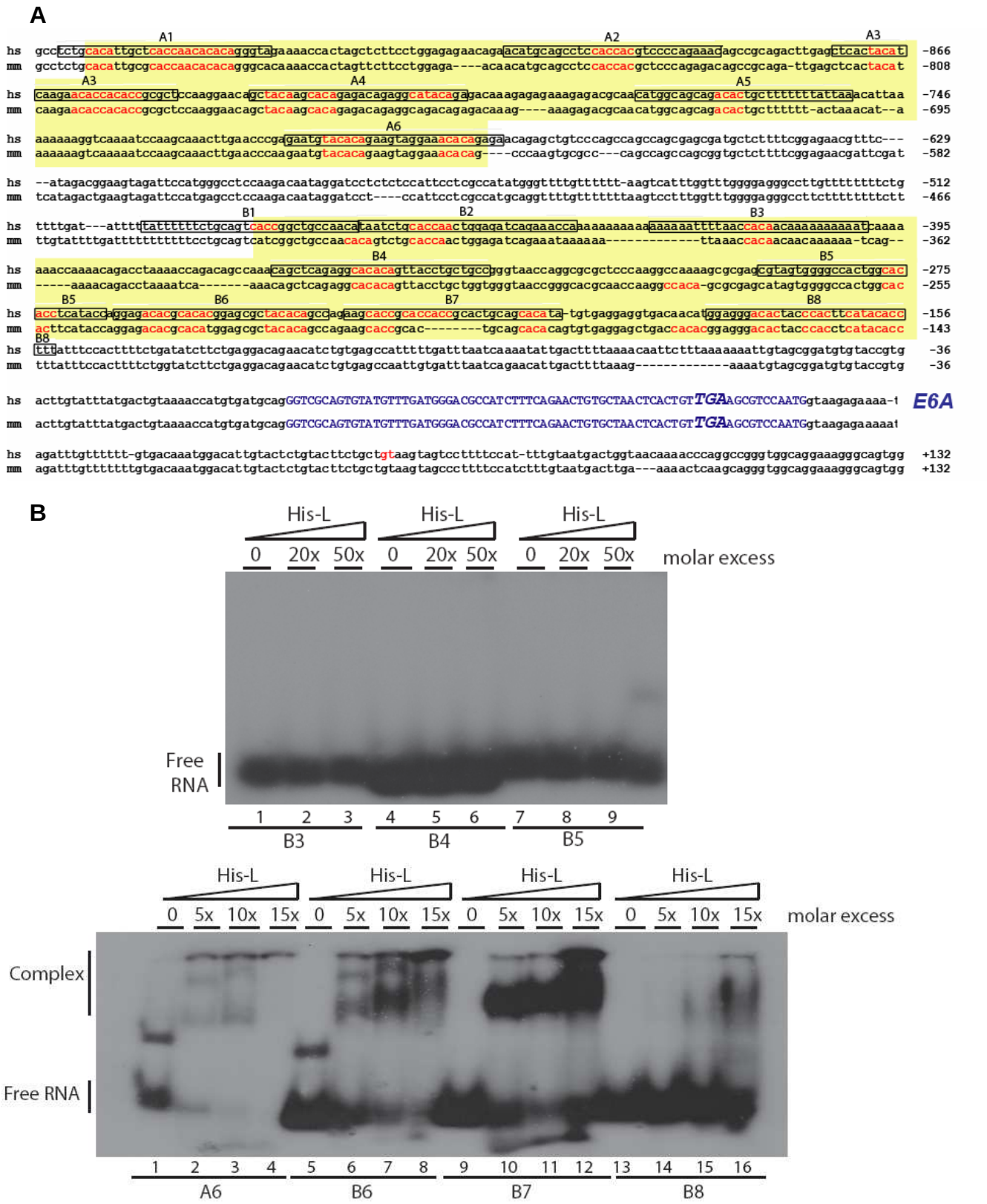


Figure 3.15 EMSA with hnRNP L and short RNAs.
(A) Alignment of the CA-rich cluster in intron 6 sequences (human (hs) vs. mouse (mm)). Sequences of the human 30mer RNAs used for the binding assay are boxed in black, name of each short RNA is indicated above the box. High-score binding motifs coloured in red. **(B)** ^{32}P-labeled short RNAs were incubated with recombinant His-tagged hnRNP L (molar excess indicated above each gel). Protein-RNA complexes and unbound RNA were separated on a non-denaturing polyacrylamide gel. Autoradiograms of the gels are shown. Position of unbound RNA is indicated on the left.

To demonstrate in a more controlled assay that the formation of protein-RNA complexes depends on the number of high-score binding motifs and not on the context of the substrate RNA, I performed EMSA with B7 short RNA and several mutant derivatives (Fig. 3.16A). The mutant derivatives were generated by introduction of nucleotide substitutions (A or C to G) in either one or two high-score binding motifs.

EMSA was done with ^{32}P-labeled RNA and recombinant hnRNP L protein (Fig. 3.16).

The results indicated that wild-type B7 RNA efficiently associated with hnRNP L protein (Fig. 3.16B lanes 1-5), whereas the mutations introduced in one of the high-score binding motifs either strongly diminished (B7M1, lanes 6-10) or completely abolished (B7M12, lanes 11-14 and B7M3, lanes 15-19) binding of hnRNP L to RNA.

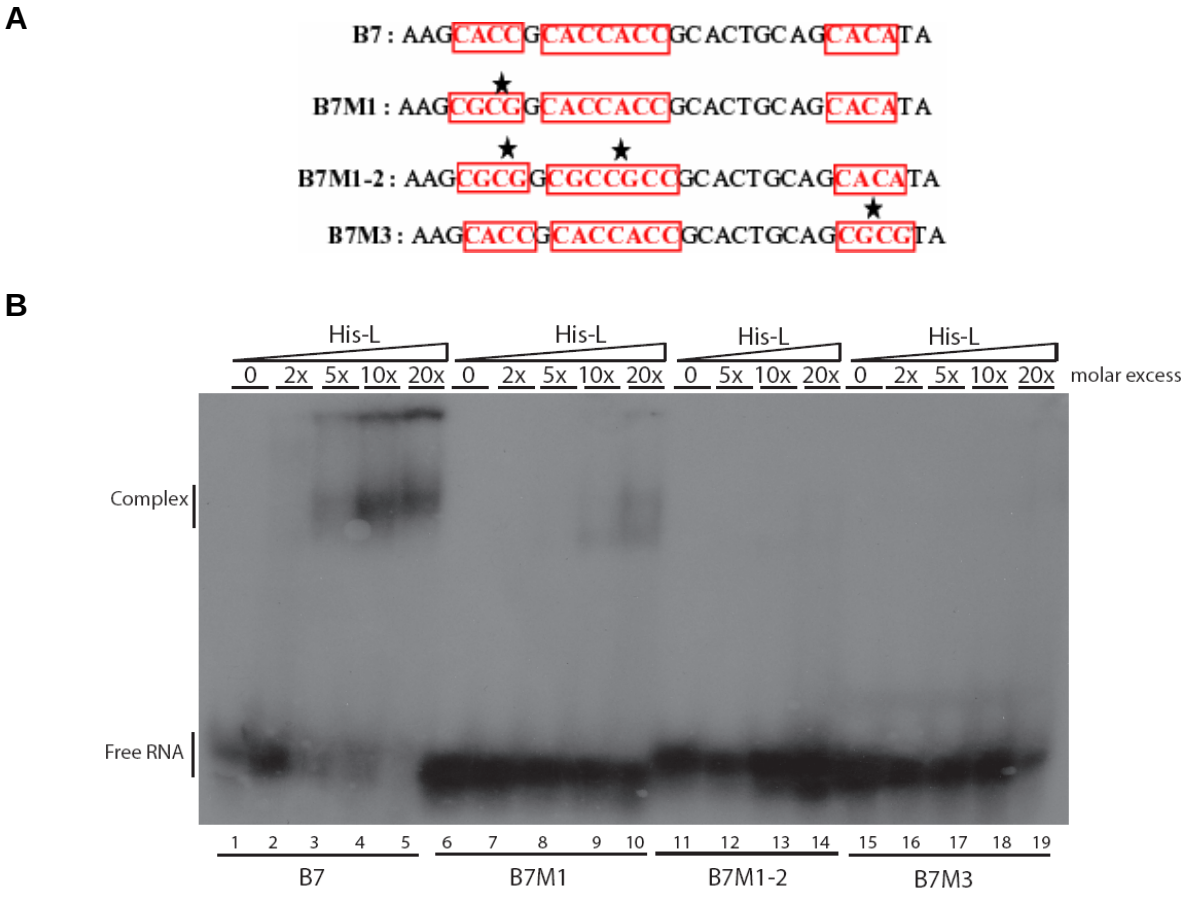


Figure 3.16 Mutational analysis of the B7 element.
(A) The complete sequence of 30-nucleotide B7 RNA and its mutant derivatives. The high-score binding motifs are boxed in red, mutated motifs are marked with the star. **(B)** Band shift experiment with B7 mutant derivatives. A fixed amount of ^{32}P-labeled RNA (indicated below the gel) was incubated with recombinant hnRNP L protein in molar excess as indicated above the gel. Protein-RNA complexes and unbound RNA were separated by non-denaturing polyacrylamide gel electrophoresis. Autoradiogram of the gel is shown.

A new inspection of the sequence of the B7 and its mutants revealed that a single high-score binding motif was insufficient for high-affinity RNA binding of hnRNP L. RNA with two motifs was bound with very low affinity, demonstrating that the presence of two high-score motifs in B7 RNA context were also not optimal for tight binding. Only all three motifs allowed strong binding of hnRNP L to the RNA substrate.

Since the binding of hnRNP L to CA cluster occurs in a cooperative manner, we propose that by mutation of high-score motifs we disrupted the cooperativity between several hnRNP L molecules.

To understand the complex assembly of hnRNP L and B7 RNA, I performed a gel filtration analysis of protein-RNA complexes.

The ^{32}P-labeled B7 RNA pre-incubated with different amounts of the recombinant His-tagged hnRNP L was loaded onto the Superdex 200 gel filtration column and eluted under a flow rate of 0,1 ml/min.

The results (Fig. 3.17) demonstrated that B7 RNA migrated with a retention time corresponding to a protein of a molecular mass of approximately 20 kDa (Fig. 3.17, fraction 27). However, when B7 RNA was pre-incubated with a 10-fold molar excess of hnRNP L for 20 min on ice, the peak was shifted to fraction 20-21 with a corresponding Mr of 80 kDa. Under these conditions, hnRNP L and B7 RNA formed a complex composed of one molecule of hnRNP L protein with one molecule of B7 RNA. A further increase of protein concentration did not shift the peak towards higher molecular weights.

Next, I decided to increase the length of RNA and used B6B7 RNA, which is 60 nucleotides in length, composed of B6 and B7 short RNAs with six high-score binding motifs in total. Gel filtration analysis revealed that hnRNP L formed a complex with B6B7 RNA that migrated on a gel filtration column with a retention time corresponding to a protein of approximately Mr 140 kDa in size (fraction 16). Since B6B7 RNA has a Mr 20 kDa, this behaviour in consistent with a complex consisting of two hnRNP L molecules bound to one B6B7 RNA. Free B6B7 RNA was detected in fractions 19-20.

The results from the mutational analysis of B7 RNA and gel filtration indicate that a single high-score binding motif might be recognised by a single RRM of hnRNP L, and the more RRMs are involved in RNA binding the more stable is the association of hnRNP L with the RNA.

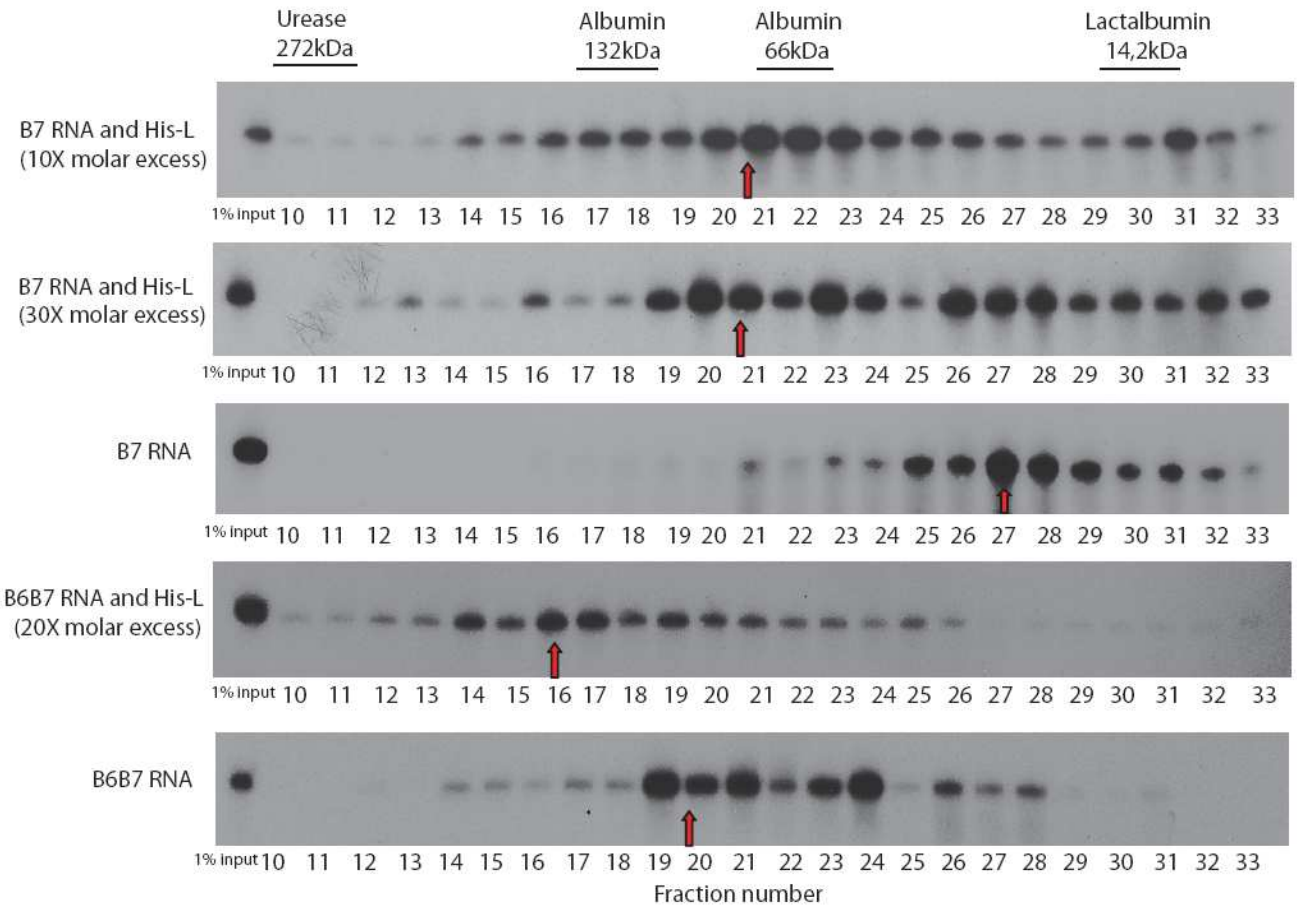


Figure 3.17 Gel filtration analysis of hnRNP L-B7 and hnRNP L-B6B7 RNA complex formation. ^{32}P-labeled B7 and B6B7 RNA were incubated with hnRNP L protein in a molar excess indicated on the left. The complex assembly reactions were performed in a total volume of 150 µl. All reactions contained 2 mg/ml tRNA. The samples were loaded onto Superdex-200 gel filtration column and eluted under a flow rate of 0,1 ml/min. 33 fractions were collected. Numbers of the fraction are indicated below each gel. RNA was recovered from each fraction, separated on a 10% denaturing gel and visualised by autoradiography. Positions of the molecular mass standards are indicated on the top. Peaks are marked with the arrows.

In order to obtain further insight into RNA sequence requirement for stable binding of hnRNP L to RNA, I made a new set of mutants to investigate the flexibility of the distances between high-score binding motifs.

Comparison of the distances between high-score binding motifs in A6, B6, B7 and B8 short RNAs of the CA cluster revealed that the average length of the spacer is 9 nucleotides. I chose B7 RNA with a nine-nucleotide spacer and either decreased the length of spacer down to four nucleotides [B7(4)] or increased it up to thirteen nucleotides [B7(13)] (Fig. 3.18A).

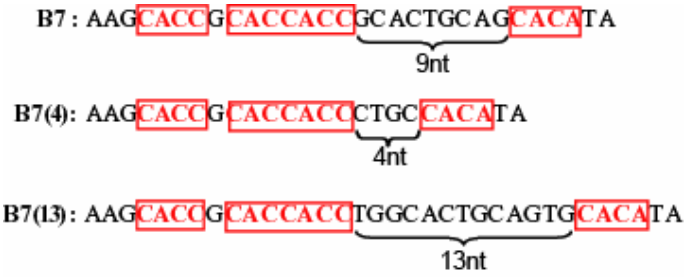
The ^{32}P-labeled RNAs described above were tested for binding of GST-tagged hnRNP L and hnRNP LL using gel shift analysis (Fig. 3.18B).

The GST-tagged hnRNP L (lanes 4, 5) efficiently bound to B7 RNA, whereas binding of GST-tagged hnRNP LL was much weaker (lanes 2, 3), indicating that B7 short RNA is not an optimal binding element for hnRNP LL, unlike the full-length CA-rich cluster (data not shown).

Although, some complex formation was seen with both mutant RNAs and GST-hnRNP L (lanes 9, 10, 14, 15), binding was very weak, indicating that there are tight restriction on the length of spacer resulting for optimal binding.

Since all RRMs of hnRNP L are separated by the linkers of a certain length, the high-affinity binding of RRMs to high-score binding motifs within the same RNA depend on the distances separating these motifs.

A



B

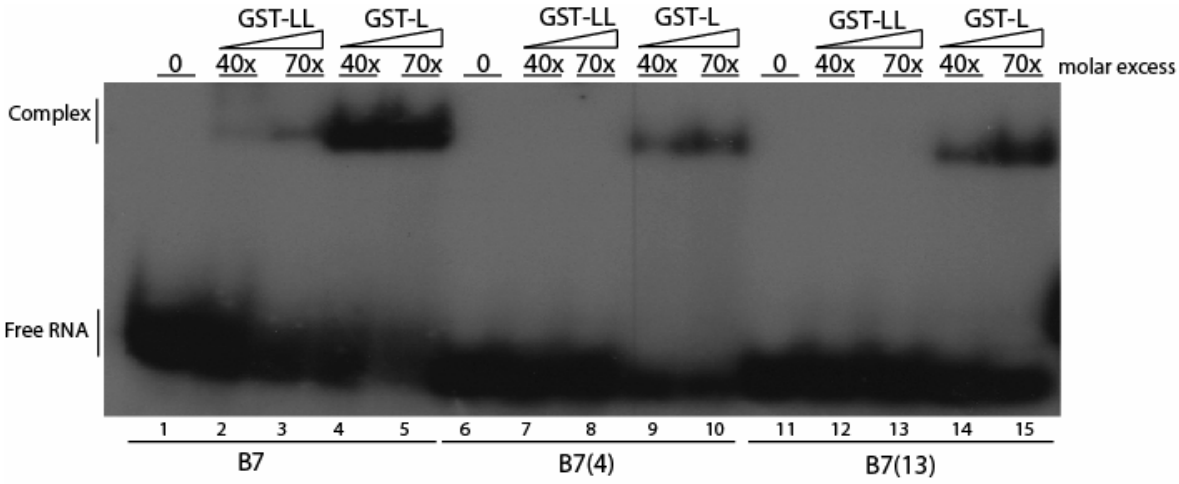


Figure 3.18 Required length of the spacer for high-affinity binding.
(A) Sequence of 30-nucleotide B7 RNA and its spacer mutants. The high-score binding motifs are boxed in red. **(B)** EMSA with B7 mutant derivatives. A fixed amount of ^{32}P-labeled RNA (indicated below the gel) was incubated with recombinant hnRNP L or hnRNP LL proteins in molar excess indicated above the gel. Protein-RNA complexes and unbound RNA were separated by non-denaturing polyacrylamide gel electrophoresis. Autoradiogram of the gel is shown. Positions of resulting RNA-protein complexes and free RNAs are indicated on the left.

3.6 HnRNP L and hnRNP LL protein-protein interaction

3.6.1 Oligomerisation state of hnRNP L

To examine the oligomeric state of hnRNP L, recombinant His-tagged hnRNP L was subjected to gel filtration analysis under non-denaturing conditions (Fig. 3.19). This analysis showed that under these conditions, recombinant hnRNP L eluted with an apparent molecular mass of 66 kDa (fraction 11) and 132 kDa (fraction 9) which corresponds to the expected mass of the monomer and the dimer, respectively. Similarly, yeast two-hybrid system and *in vitro* co-precipitation assays indicated that hnRNP L could interact with itself (Kim *et al*., 2000). Thus, hnRNP L exists in solution as a monomer and as a dimer, raising the question of the contribution of the homodimerisation to the overall function of hnRNP L.

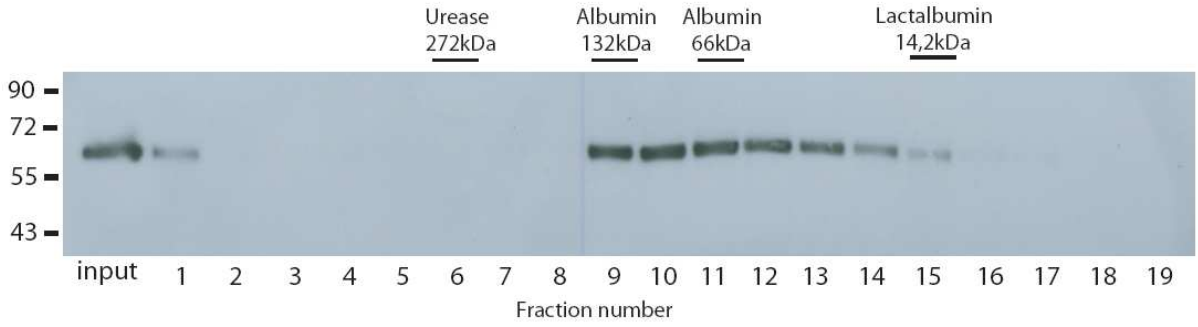


Figure 3.19 Gel filtration analysis of hnRNP L.
Recombinant His-tagged hnRNP L was applied to a Superdex-200 gel filtration column and eluted under a flow rate of 0,1 ml/min. 20 fractions were collected. The proteins from each fraction were precipitated and analysed by SDS-PAGE. The SDS molecular weight markers (M, kDa) and the elution pattern of four marker proteins and their sizes are as indicated. Proteins were transferred to a PVDF membrane and Western blot was performed to visualise hnRNP L using anti-His antibody. Input lane (0,04% of recombinant hnRNP L) is on the left.

3.6.2 Mapping of the interaction region of hnRNP L and hnRNP LL

In order to investigate whether hnRNP L and its paralog hnRNP LL are interacting with each other I performed GST pull-down experiments (Fig. 3.20). Immobilised GST-hnRNP LL was incubated with HeLa nuclear extract (Fig. 3.20A), or immobilised GST-hnRNP L was incubated with recombinant His-tagged hnRNP LL (Fig. 3.20B).

Pull-down experiments revealed strong interaction between hnRNP L and hnRNP LL (A, B, lane 2). Neither hnRNP L no hnRNP LL were pulled-down either by GST alone or by beads (A, B, lanes 6, 7). Because both hnRNP L and LL are known RNA-binding proteins, I next tested whether the interaction between these proteins is mediated by RNA. Binding experiments in the presence of RNase A were performed. No change in the strength of the binding of hnRNP L to LL and vice versa was found (data not shown). This indicates that hnRNP L and hnRNP LL interact directly with each other; however the possibility that RNA further regulates this interaction cannot be excluded.

In order to determine the hnRNP L and hnRNP LL domains required for protein-protein interaction, a series of GST-tagged deletion mutants were tested for the ability to interact either with full-length hnRNP L or hnRNP LL (Fig. 3.20).

As shown in Fig. 3.20B most of the hnRNP L protein was required for the interaction with hnRNP LL (lanes 3-5), indicating the presence of multiple interacting regions. In the case of hnRNP LL the N-terminal part of the protein spanning RRM1 and RRM2 failed to interact with hnRNP L, indicating that the C-terminal half of hnRNP LL (amino acid 162-542) is involved in the interaction with hnRNP L.

The results for hnRNP L are in agreement with previously observed data, where hnRNP L structure was mapped in terms of interaction with other proteins (hnRNP E2, K, I, L), demonstrating that the junction sequences between RRM2 and RRM3 as well as between RRM3 and RRM4 of hnRNP L participate in protein-protein interaction (Kim *et al*., 2000).

3.7 Alternative splicing regulation by hnRNP L and its mutant derivatives

3.7.1 *SLC2A2* minigene construct

SLC2A2 was identified as an hnRNP L alternative splicing target gene by a genome-wide database search (Hui *et al*., 2005). The gene contains a CA-repeat sequence in intron 4 close to the alternatively spliced exon 4 (Fig. 3.21). To investigate the influence of these CA repeats on splicing of exon 4, a *SLC2A2* minigene construct was used. It carries exons 3 to 5 with the alternatively spliced exon 4 in between. The length of introns 3 and 4 was shortened. In a substitution derivative, the CA-repeat element was replaced by a non-specific control sequence. It was shown that binding of hnRNP L to the CA-repeat sequence interferes with recognition of the 5'-splice site by U1 snRNP, resulting in exon 4 skipping (Heiner *et al*., 2010).

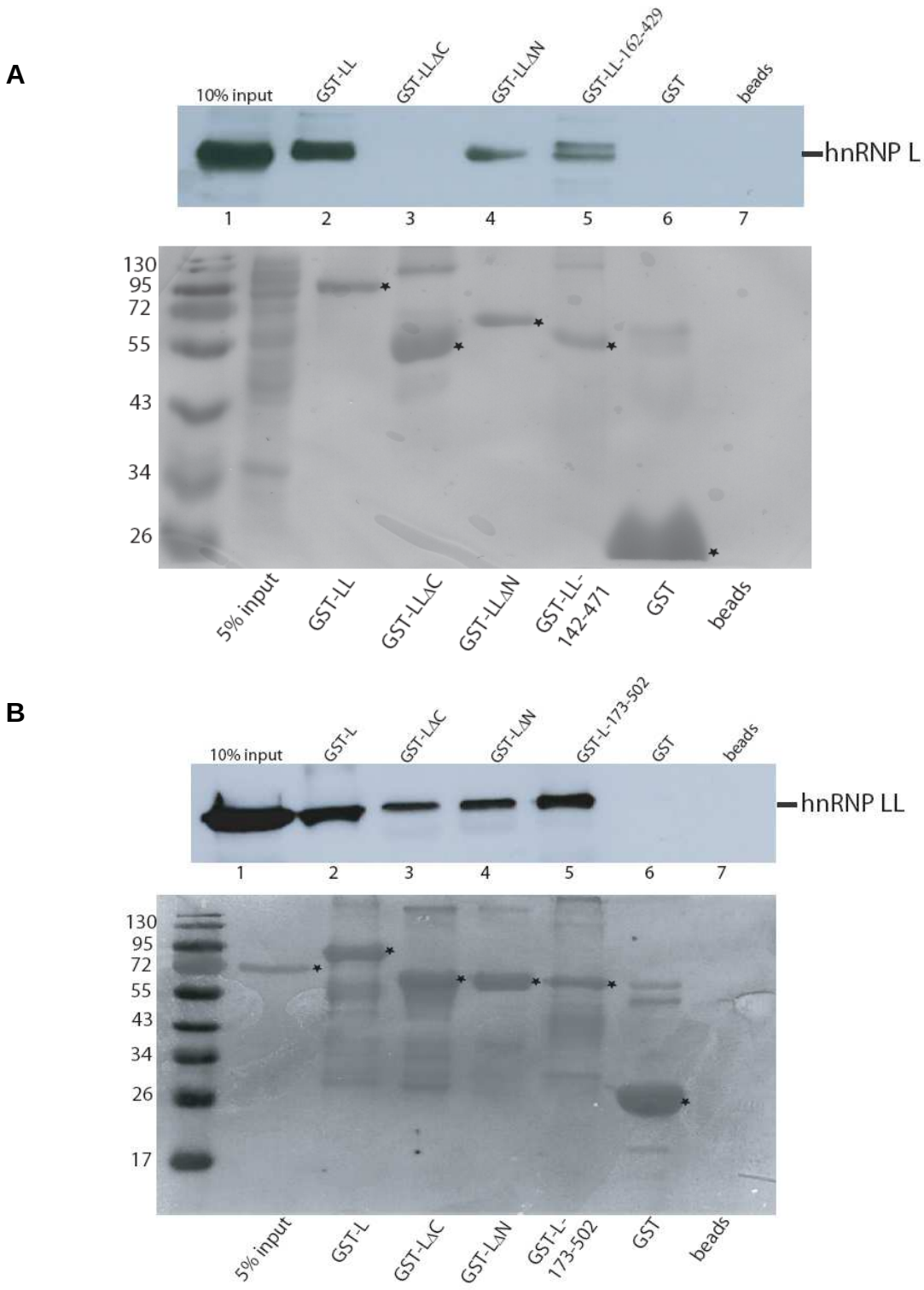


Figure 3.20 Mapping of hnRNP L and hnRNP LL domains responsible for protein-protein binding.
Pull-down of endogenous hnRNP L with glutathione-Sepharose beads pre-coated with recombinant GST, GST-hnRNP LL or its deletion mutants **(A)** or His-hnRNP LL with glutathione-Sepharose beads pre-coated with recombinant GST, GST-hnRNP L or its deletion mutants **(B)**. Input and bound fractions were analysed by SDS-PAGE followed by Western blotting with specific antibodies either against hnRNP L (4D11) **(A)** or anti-His **(B)**. Ponceau S staining of the membranes served as a loading control. The SDS molecular weight markers (M, kDa) are indicated on the left, the specific protein bands are marked by asterisks.

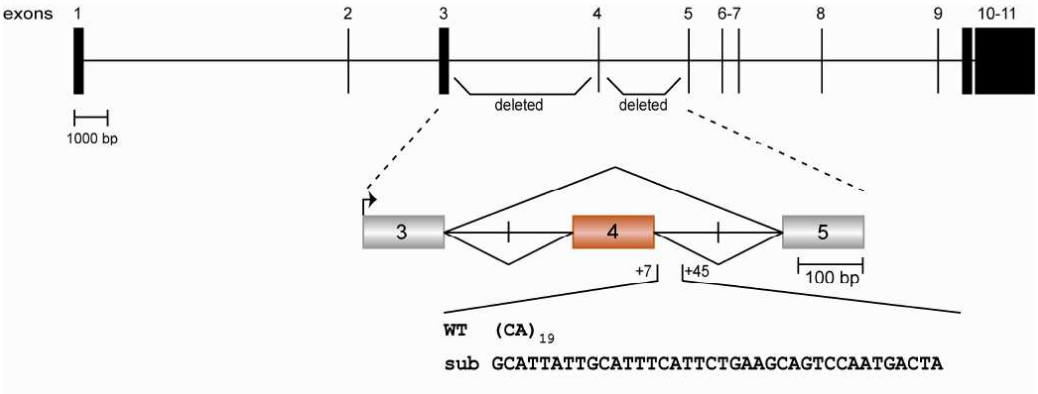


Figure 3.21 Schematic representation of the *SCL2A2* gene structure and minigene construct. Exons are indicated as black boxes, introns as lines. Exon numbers are labeled. The *SLC2A2* minigene construct carrying constitutive exons 3 and 5 (shown as grey boxes) and alternative exon 4 (shown as a red box). The length of the introns 3 and 4 was shortened as shown. Alternative splicing patterns are indicated by lines below and above the minigene construct. Wild-type (WT) and substituted (sub) sequence elements are given with boundaries relative to the 5' splice site (adopted from Heiner *et al*., 2010).

3.7.2 Regulation of exon 4 skipping in vitro by hnRNP L and its mutant derivatives: four RRMs required for repressor activity

Previously it was shown that depletion of hnRNP L protein from HeLa nuclear extract can affect splicing of *SLC2A2* exon 4 by significantly increasing the level of its inclusion. Add-back of recombinant hnRNP L protein reversed this effect (Heiner *et al*., 2010).

The domains of hnRNP L required for splicing repression were examined by testing several mutant proteins in the splicing complementation assay.

To deplete HeLa nuclear extract of hnRNP L protein I took advantage of the tight binding of hnRNP L to $(CA)_{32}$ RNA (Hui *et al*., 2003b). HeLa nuclear extract was incubated with 5'-biotinylated $(CA)_{32}$ RNA oligonucleotide pre-bound to neutravidin-agarose beads. After isolation of biotin RNA-bound proteins, the depleted nuclear extract was used for *in vitro* splicing assay. As control, a mock depletion was done in the absence of RNA oligonucleotide.

To test for the efficiency of depletion I performed Western blot analysis with specific anti-hnRNP L antibody (4D11) (Fig. 3.22A). The efficiency of hnRNP L depletion was 80-90% (compared lanes 1 and 2). GAPDH served as a loading control. Mock depletion did not result in a substantial depletion of hnRNP L from HeLa nuclear extract (data not shown).

Next, I confirmed the repressive function of hnRNP L protein on *SLC2A2* exon 4 -inclusion. *In vitro* transcribed wild-type pre-mRNA was spliced either in hnRNP L-

depleted or mock-depleted nuclear extract (Fig. 3.22B). Alternative splicing was analysed by semi-quantitative RT-PCR.

Depletion of hnRNP L but not mock depletion activated splicing of alternative exon 4. I detected 62% exon inclusion in mock-depleted extract whereas over 90% exon inclusion was observed in hnRNP L-depleted extract (Fig. 3.22B, lanes 1, 2). Addition of recombinant hnRNP L to the depleted extract repressed splicing to the same extent observed in mock-depleted extract (lanes 3-5). I detected 60% of exon inclusion at 300 ng of recombinant hnRNP L per 12,5-µl splicing reaction.

Add-back of hnRNP L lacking the N-terminal glycine-rich region (LΔG) restored the inhibitory effect on exon 4 inclusion in a dose-dependent manner (Fig. 3.22B, lanes 6-8), indicating that the N-terminal glycine-rich region plays no essential role in alternative splicing regulation.

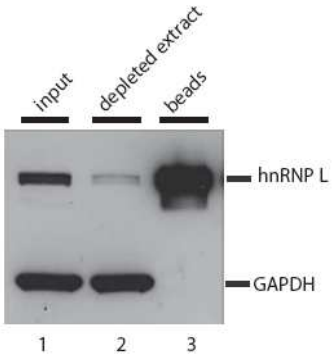
Two hnRNP L deletion mutants (LΔC, LΔN) failed to restore splicing repression (Fig. 3.22C, lanes 3-6). Over 90% of exon 4 inclusion compared to hnRNP L-depleted nuclear extract (lane 2) was observed with addition of 200 ng of each protein.

Add-back of the proteins with amino acid substitutions either in RRM1 or RRM2 restored the inhibitory effect on exon 4 inclusion, except L-V225A (Fig. 3.22D). The level of exon inclusion was decreased to 57 and 62% for L-V104A and L-G231A-M235V, respectively (lanes 3, 4 and 5, 6). Interestingly, the protein with V225A mutation showed no inhibitory effect on exon 4 inclusion (lanes 7, 8).

The proteins with mutations introduced in RRM3 and RRM4 (L-V402K-L405D-L558K-L559D and L-V402K-L405D-L558K-L559D) did not repress splicing of exon 4 (Fig. 3.22E). Addition of 300 ng of each protein did not change the level of exon inclusion (lanes 4, 6), 88% of exon inclusion for both proteins compared to hnRNP L-depleted nuclear extract was detected (lane 2).

Taken together, these data clearly demonstrate that all four RRMs are necessary for hnRNP L function in alternative splicing regulation.

A



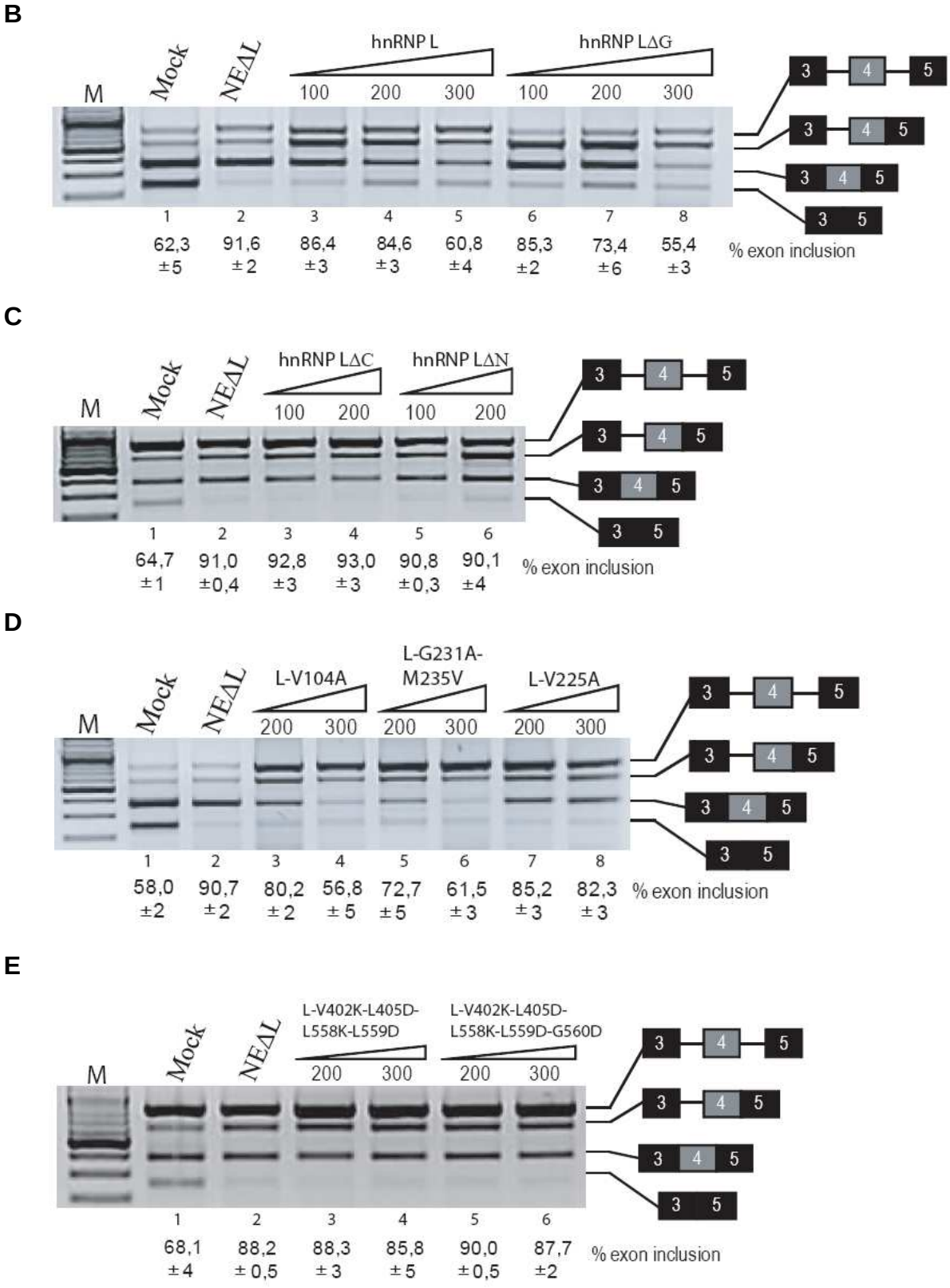


Figure 3.22 *SLC2A2* pre-mRNA splicing modulated by hnRNP L and its mutant derivatives. **(A)** HnRNP L protein was selectively depleted from HeLa nuclear extract via biotin pull-down with $(CA)_{32}$ RNA oligonucleotide, the depletion was analysed by Western blotting with an hnRNP L-specific monoclonal antibody (4D11). Lane 1, untreated extract; lane 2, hnRNP L-depleted extract; lane 3, protein eluted from neutravidin beads. GAPDH detection was used as internal standard. **(B, C, D, E)** *In vitro* splicing of *SLC2A2* pre-mRNA in mock-depleted nuclear extract (lane 1), hnRNP L-depleted nuclear extract (lane 2) or in the presence of increasing amounts (indicated above the gels in nanograms) of indicated recombinant hnRNP L wild-type or mutant proteins.

Figure 3.23 summarises the RNA binding and alternative splicing activities obtained with the different proteins.

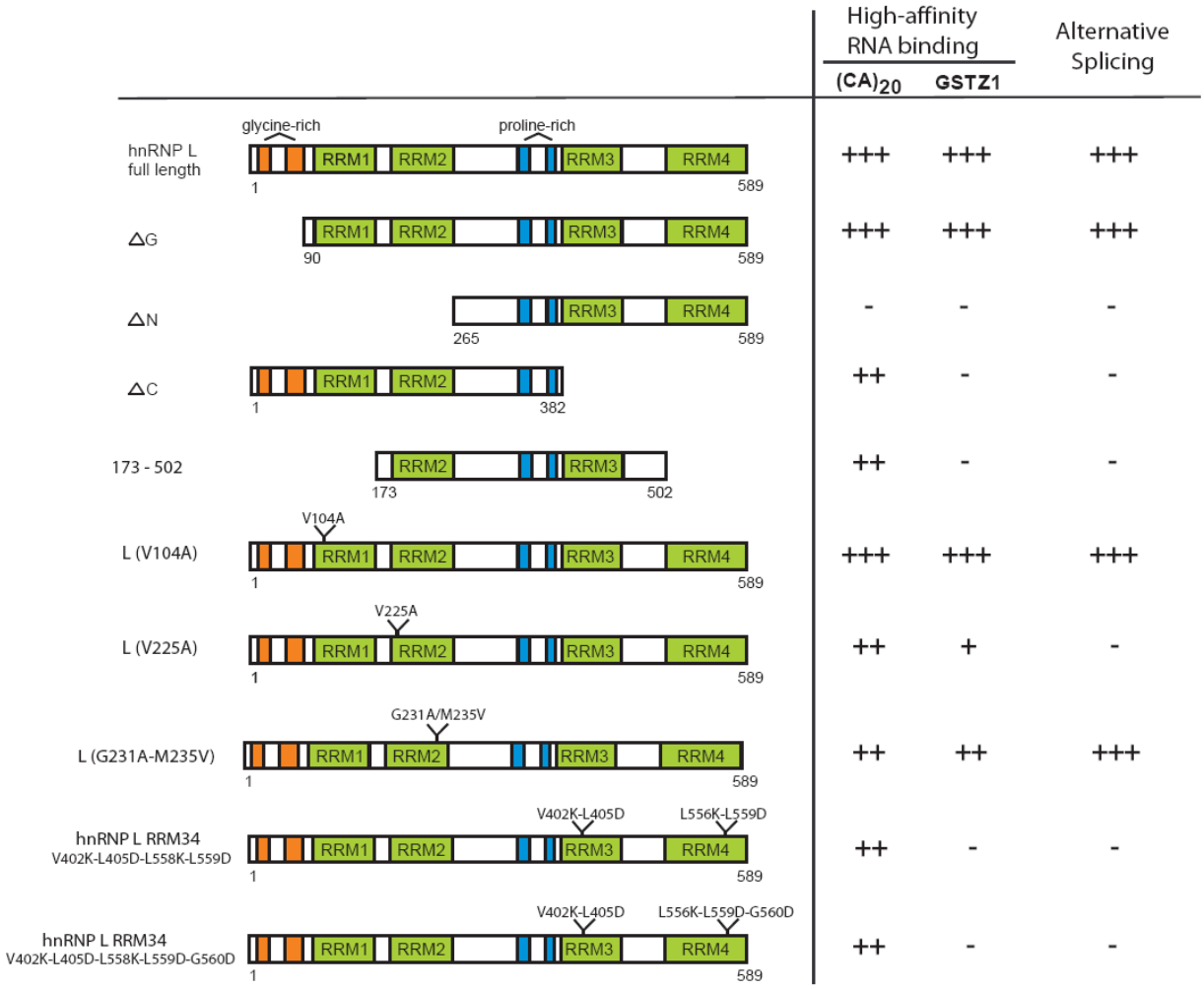


Figure 3.23 Summary of the activities of hnRNP L mutant proteins.
Structures of the full-length and variant proteins are shown schematically at left. The RRMs, the glycine- and proline-rich regions are indicated by green, orange and blue boxes, respectively. Relative specific activities for high-affinity binding to different substrates and repression activity in alternative splicing are indicated by + and – signs. See Figures 3.10 and 3.22 for data used in scoring these values.

4. Discussion

A common feature of many RNA-binding proteins in the RRM family as well as in the family of double-stranded RNA-binding proteins and RNA-binding proteins containing RGG repeats and KH domains is the presence of multiple RNA-binding domains (Kiledjian *et al*., 1999; Birney *et al*., 1993; Burd and Dreyfuss, 1994; Kharrat *et al*., 1995; Musco *et al*., 1996; Maris *et al*., 2005). Since proteins contain one to four RNA-binding domains, an intriguing question is what is the function of multiple RNA-binding domains?

My PhD work was aimed towards understanding the contribution of individual RRM domains of hnRNP L and hnRNP LL proteins to their functions.

For this purpose, I generated hnRNP L and hnRNP LL deletion derivatives and variants with amino acid substitutions to study their activities in RNA binding and alternative splicing regulation.

4.1 Function of multiple RRMs of the hnRNP L and hnRNP LL proteins in RNA binding

In some RNA-binding proteins with two or more RRMs, a subset of one or more RRM-containing regions is generally sufficient for RNA binding (Burd *et al*., 1991; Chung *et al*., 1996; Dember *et al*., 1996; Sachs *et al*., 1987). Some other RNA-binding proteins do require all of their RRMs for RNA-binding activity (Green *et al*., 1992; Zamore *et al*., 1992).

EMSA and filter binding assay results with hnRNP L deletion derivatives (summarised in Fig. 3.10) demonstrate that the combination of at least two RRMs (1/2 or 2/3) is necessary and sufficient for high-affinity binding of hnRNP L to CA-repeat RNA. In comparison, the C-terminal RRMs 3 and 4 do not show detectable association with CA-repeat RNA.

Interestingly, the binding of hnRNP L to CA-repeat RNA with a nanomolar affinity is achieved only in the presence of RRM2 with at least one additional RRM. A similar phenomenon is observed with many other multi-RRM proteins. For example, in Sxl1 (Samuels *et al*., 1998), hnRNP A1 (Shamoo *et al*., 1994), poly(A)-binding protein (Burd *et al*., 1991), nucleolin (Serin *et al*., 1997), ASF/SF2 (Caceres and Kreiner, 1993), and U2AF (Zamore *et al*., 1992), binding by a single RRM is much weaker and/or less specific than binding by a combination of two or more RRMs.

In contrast to CA-repeat RNA, the high-affinity recognition of CA-rich RNA required the entire hnRNP L protein and deletion mutant proteins containing only two RRMs fail to stably associate with the RNA.

RNA-binding proteins in general, besides their RNA-binding domains, contain other domains that are termed auxiliary domains. The functional significance of auxiliary domains is a relatively unexplored area; for example, these regions may mediate protein-protein interactions and localise the protein within the cell. HnRNP L has the glycine- and proline-rich regions. Deletion of the N-terminal glycine-rich region had no effect on RNA binding to CA-repeat and CA-rich RNAs, indicating that the glycine-rich region is not necessary for RNA binding. It might have some other function, for example, nuclear localisation of hnRNP L, since the nuclear localisation signal locates within a glycine-rich region (Taguchi *et al*., 2004).

Substitution of Val104 in the RRM1, Val225 and Gly231/Met235 in the RRM2 of hnRNP L did not significantly reduce the binding activity of the full-length protein to CA-repeat and CA-rich RNAs. These amino acid residues are not the major determinants of general RRM-mediated RNA binding or the introduced mutations are compensated by binding of other intact RRMs.

Interestingly, the inter-domain interaction between hnRNP L RRM3 and RRM4 (Skrisovska and Allain, 2008) which has been disrupted by several amino acid substitutions affected only binding to CA-rich RNA, indicating that tight interaction of hnRNP L RRM3 and RRM4 is important for recognition of high-score binding motifs spread over the RNA substrate. The structure analysis of two PTB domains RRM3 and RRM4 (RRM3-4) revealed a very unusual interaction between these two RRMs and also a new role for the inter-domain linker (Vitali *et al*., 2006). The structure determination of PTB RRM3-4 in its free and RNA bound states revealed that RRM3 and RRM4 are interacting in both states, but a single difference in structures lies in a small region of the inter-domain linker. This region is unstructured in the free RRM3-4 and becomes structured upon RNA binding, forming a short α-helix, which is significantly involved in RNA binding (Vitali *et al*., 2006). Sequence alignment of human PTB and hnRNP L shows that the inter-domain contact and the resulting structure of RRM3 and RRM4 are very likely conserved between these two proteins.

EMSA results for hnRNP LL protein indicate that for high-affinity binding to CA-repeat and CA-rich RNAs all four RRMs are required, since deletion derivatives LLΔC (RRMs 1 and 2), LLΔN (RRMs 3 and 4) and LL162-429 (RRMs 2 and 3) demonstrate no binding activity.

HnRNP L and hnRNP LL have 58% overall protein sequence identity and are similar in size. The RNA-binding domains in these proteins are particularly conserved, however, there are some divergent residues in these domains that may account for the differential sensitivity of hnRNP L and hnRNP LL to the tested

CA-repeat and CA-rich RNAs. Furthermore, the glycine- and proline-rich regions, which are less pronounced and absent in hnRNP LL, respectively, may be important for function and/or fine tuning of the binding specificity.

4.2 Role of RRM2 in RNA-binding activity of hnRNP L

From the analysis of the binding behaviour of the two hnRNP L deletion constructs, LΔC (RRMs 1 and 2) and L173-502 (RRMs 2 and 3), it might be concluded that RRM2 represents the domain required for sequence specific RNA recognition, whereas RRM1 and RRM 3 probably contribute to stabilisation of the binding (Fig. 3.6). The important role of hnRNP L RRM2 in RNA binding was confirmed by the introduction of point mutations in RRM2 in the context of two-RRM derivatives (Fig. 3.4). Mutant proteins with inactivated RNA-binding activity of RRM2 did not associate with CA-repeat RNA, confirming that RRM2 is the major determinant for RNA-binding affinity and specificity. However, other domains of hnRNP L besides RRM2 are also needed for sequence specific recognition since it appears that RRM2 alone can not efficiently bind to RNA. Thus, RRM2 may be sufficient to generate RNA-binding sequence-specificity, but additional RRMs are required for stabilisation of the formed complex. This is consistent with the fact that the full-length protein shows the highest binding affinity, and may provide a mechanistic explanation for the observation that deletion mutant proteins show reduced binding affinity compared to full-length hnRNP L.

Sequence comparison revealed that human hnRNP L shares significant homology (55% similarity, 29% identity) with PTB. Moreover, both proteins contain four RRMs in which sequence identity reaches 32%. However, the substrate specificity of hnRNP L was not changed upon mutation of important for substrate specificity amino acid residue valine in RRM2 to isoleucine, present in PTB at the same position (Fig. 3.8).

4.3 CA-rich elements: important features for high-affinity binding of hnRNP L

My data demonstrate that hnRNP L binds tightly and specifically to CA-rich RNA, such as GSTZ1 or CA-rich cluster. The binding of hnRNP L to CA-rich cluster within its own pre-mRNA has been shown to be important for autoregulation of hnRNP L protein expression in HeLa cells (Rossbach *et al*., 2009). A highly conserved CA-rich cluster is a dense cluster of high-score hnRNP L binding motifs extending over 800 nucleotides and bound by more then one molecule of hnRNP L in a

cooperative manner. Detailed study of CA-rich cluster indicates that for high-affinity binding of hnRNP L several high-score binding motifs are necessary. Moreover they have to be separated by a spacer of 7-10 nucleotides (Fig. 3.15 and 3.18). These data demonstrate that hnRNP L, probably, binds RNA via multiple RRM domains, each of which independently recognise different high-score binding motifs on the RNA. This observation is in agreement with previously obtained results indicating that the presence of all four RRMs provides highest binding affinity.

Structure analysis of RRM domains in complex with RNA have demonstrated that a single RRM can accommodate between two to eight nucleotides (Mazza *et al*., 2002; Price *et al*., 1998; Calero *et al*., 2002). In addition, the structures of several tandem RRMs bound to RNA have been determined. In most cases (Deo *et al*., 1999; Handa *et al*., 1999; Johansson *et al*., 2004; Wang and Tanaka-Hall, 2001; Allain *et al*., 2000), both RRM domains are separated by a small linker and recognise two adjacent stretches within the same RNA molecule. This topology provides a large RNA-binding surface. However, there are exceptions, for example RRMs 3 and 4 of PTB (Oberstrass *et al*., 2005; Vitali *et al*., 2006). The tandem RRMs make intensive inter-domain contact and their RNA-binding surfaces point away from each other. This topology prevents the two domains from binding adjacent sequences within a single RNA, but instead favours looping of RNA if the two binding sites are separated by at least 15 nucleotides (Oberstrass *et al*., 2005).

4.4 HnRNP L working model

To summarise the RNA-binding data presented in this work, the current working model of hnRNP L binding activity is proposed in Figure 4.1. As shown, the hnRNP L protein contains four RRMs, where RRM1 and RRM2 act, probably, independently of each other, and RRM3 and RRM4 are tightly associated with each other. Two types of RNA substrates might be recognised by hnRNP L: CA-repeat and CA-rich RNAs. High-affinity binding to CA-repeat RNA requires the combination of two RRMs (1/2 and 2/3), however, presence of all four RRMs stabilise the interaction to a greater extend. For high-affinity recognition of CA-rich RNA the full-length protein is necessary.

The RRMs of hnRNP L are connected by linkers of variable length. The ability of the protein to recognise a specific RNA depends on the linker length and its rigidity. CA-repeat RNA is more adoptable for recognition by hnRNP L and the lengths of the inter-domain linkers are not very important. CA-rich RNA carries high-score binding motifs separated by spacers. A long linker sequence between two RRMs (e.g. between 2 and 3) allows the two domains to recognise binding motifs

separated by a spacer of a varying length, whereas a short linker (e.g. between 1 and 2) pre-disposes the domains to bind to adjacent motifs.

The binding of RRM4 to RNA was determined by NMR spectroscopy study of RRM4-RNA complex (personal communication with Prof. Allain, Zurich). Thus, the inter-domain interaction between RRM3 and RRM4 might induce the formation of RNA loop or might be involved in functionally important homodimerisation or interaction with other splicing factors.

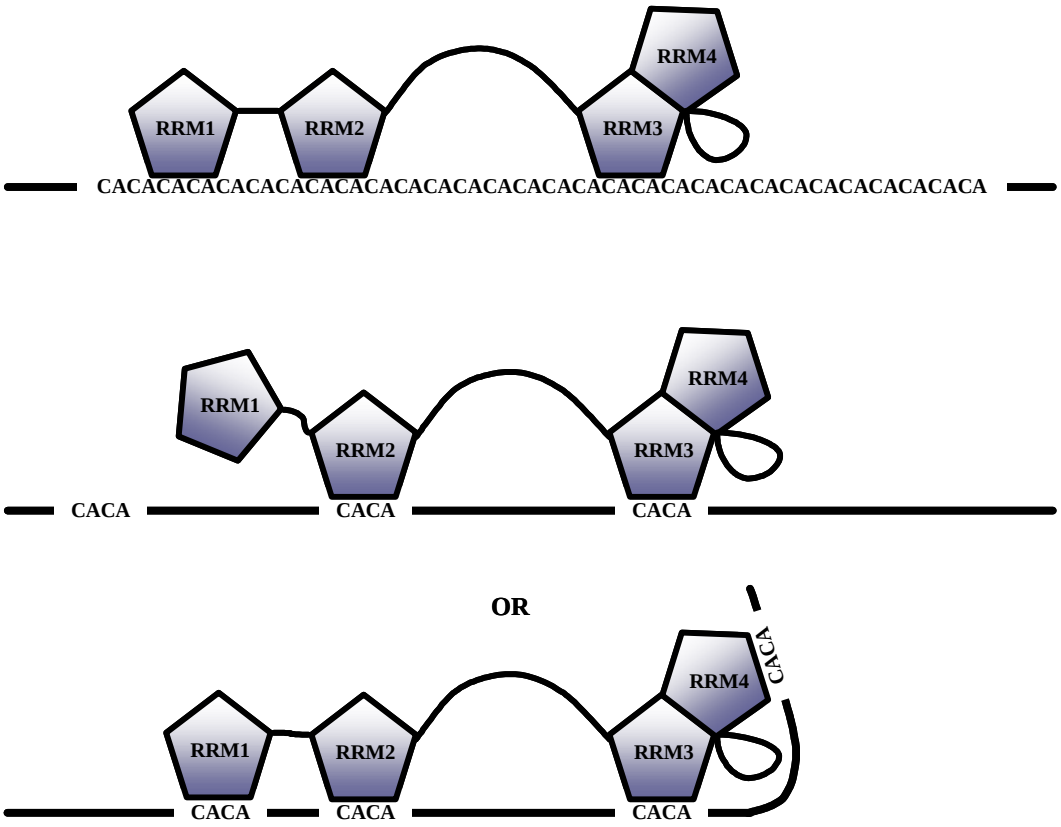


Figure 4.1 Model illustrating the interaction of hnRNP L with different RNA substrates.
The RNA-binding protein hnRNP L contains four RRMs (blue pentagon). Each RRM interacts with a single high-score binding motif (CACA) within the same RNA. The high-affinity binding to CA-repeat RNA is not restricted by the length of the inter-domain linkers due to the presence of a large number of possible binding sites. In the case of CA-rich sequences, the RNA should meet the spatial requirements such as the length of the spacers between high-score binding sites to allow high-affinity binding of hnRNP L. The C-terminal domains RRM3 and RRM4 might be involved either in RNA looping or interaction with other splicing factors or homodimerisation of hnRNP L.

4.5 HnRNP L interacts with hnRNP LL

Heterogeneous nuclear ribonucleoproteins (hnRNPs), being very abundant and avid RNA-binding proteins, play an important role in several RNA-related biological processes such as transcription, pre-mRNA processing, mature mRNA transport to cytoplasm and translation. These diverse functional properties can be achieved not only through RNA binding, but can also be powerfully modified by a protein-protein

interaction with other factors (Kim *et al*., 2000; Hahm *et al*., 1998a; Scherly *et al*., 1990; D'Ambrogio *et al*., 2009).

Many interaction partners defining the hnRNP L function were identified in the past. The interaction with human AP-endonuclease 1 identified a novel function of hnRNP L in transcriptional regulation (Kuninger *et al*., 2002), interaction of hnRNP L with hnRNP D/AUF1 function in the export of mRNA from nucleus to cytoplasm, mRNA turnover and the translation of a certain mRNA (Park *et al*., 2007). As a component of SLM/Sam68 nuclear bodies hnRNP L is involved in coupling signalling and splicing (Rajan *et al*., 2009), as an adopter for TAP and Aly/REF proteins binding of hnRNP L enhances nucleocytoplasmic export of PPE-containing intronless RNAs (Guang *et al*., 2005; Liu and Mertz 1995), interaction between hnRNP L and PTB plays an important role in positive regulation of Cat-1 translation via IRES that causes a global decrease in protein synthesis under stress conditions (Majumder *et al*., 2009). HhRNP L as a component of human KMT3a complex contributes to H3K36 trimethylation indicating a new function of hnRNP L in chromatin modification (Yuan *et al*., 2009).

In my work I identified that hnRNP L specifically interacts with its paralog hnRNP LL *in vitro* (Fig. 3.20). The interaction was mapped, indicating the regions of both proteins involved in protein-protein interaction. HnRNP L shows presence of multiple interactions sites, suggesting that protein-protein interaction can occur through different parts of the protein. In contrast, for hnRNP LL I could demonstrate that only the N-terminal domains (RRMs 3 and 4) are required for interaction with hnRNP L.

Considering my results and reported functions of hnRNP L and hnRNP LL, it is possible to speculate about the role for the hnRNP L-LL interaction. Previously, only splicing regulatory function of hnRNP LL has been reported: it is an essential regulator of CD45 splicing in activated T-cells (Topp *et al*., 2008; Wu *et al*., 2008; Oberdoerffer *et al*., 2008), whereas several functions have been documented for hnRNP L, including pre-mRNA splicing, mRNA polyadenylation, export, translation, and stability (Hahm *et al*., 1998b; Shih and Claffey 1999; Hiu *et al*., 2003; Hung *et al*., 2008; Tong *et al*., 2005). The hnRNP L-LL complex may combine these activities and perform different functions. Taking into account that hnRNP LL is a tissue-specific protein, its expression is up-regulated in blood cells and testis, and the hnRNP L-LL interaction may have a tissue-specific role. Due to the presence of multiple interaction domains the hnRNP L-LL complex may simultaneously interact with other factors to form multiple protein or protein-RNA complexes. It is further plausible that the interaction of hnRNP L-LL with one of their partners (protein and RNA) may serve to regulate the interaction with another partner.

4.6 HnRNP L cooperatively binds to CA-rich cluster to activate hnRNP L exon 6a inclusion

It was discovered that hnRNP L autoregulates its own expression on the level of alternative splicing, using a highly conserved CA-rich cluster spread over 800 nucleotides of intron 6. If included, exon 6A introduces a premature termination codon, resulting in NMD and downregulation of hnRNP L mRNA level (Rossbach *et al*., 2009). I demonstrated here that hnRNP L binds with high affinity and specificity to both parts of the CA cluster (Fig. 3.12). Moreover, EMSA and glycerol gradient centrifugation results show that the binding of hnRNP L molecules to the CA-rich cluster occurs in a strong cooperative manner (Fig. 3.13; 3.14).

Cooperative binding of hnRNP L to multiple high-score binding sites on its own pre-mRNA raises several possibilities as how the autoregulation of splicing might be achieved. This is a very critical biological step, since autoregulation may be important to prevent abnormally high levels of hnRNP L, which could lead to inappropriate processing of multiple targets. First, cooperative multiple binding could provide additive binding strength, which might simply be required for specific interaction between hnRNP L and CA-rich cluster of intron 6. Second, cooperative interaction of multiple hnRNP L molecules may facilitate binding to low-score binding motifs between high-score binding motifs to make such a region more accessible for splicing factors and spliceosome assembly. Third, cooperativity could also result in a faster response to increased level of hnRNP L proteins. In this case, the switch between two splicing modes could be set in a reliable way such that a threshold level of hnRNP L protein is both necessary and sufficient for selection of exon 6A inclusion. The autoregulatory feedback mechanism would contribute to hnRNP L active homeostasis.

My gel filtration results demonstrate that hnRNP L exists in solution under RNA-free conditions in monomeric and dimeric forms (Fig. 3.19). Thus, hnRNP L monomers or dimers could bind to each pair of CA elements. These initial interactions may then assemble into a stable higher order structure, resulting in an exon 6A exposed for spliceosome assembly.

In addition, I could show that hnRNP L's paralog hnRNP LL also binds to the CA-rich cluster of hnRNP L pre-mRNA with high affinity and specificity (Fig. 3.12). Previously, we could demonstrate a cross-regulatory mechanism explaining the reciprocal regulation of the two proteins (Rossbach *et al*., 2009). It was shown that hnRNP L regulates not only its own expression but also the expression of hnRNP LL protein by inclusion of a similar "poison exon" into the pre-mRNA of hnRNP LL. The binding of hnRNP LL to hnRNP L pre-mRNA raises the question whether hnRNP LL could cross-regulate the expression of hnRNP L. Since nuclear hnRNP L

protein levels are more than 10-fold higher than those of hnRNP LL (Hung *et al*., 2008), it is more meaningful to test this hypothesis in cell lines or in certain tissues where the protein level of hnRNP LL is up-regulated, for example T lymphocytes (Oberdoerffer *et al*., 2008; Topp *et al*., 2008).

4.7 Function of multiple RRMs in alternative splicing regulation

HnRNP L was identified as a global and versatile regulatory protein in the human system with roles in alterative splicing (Hung *et al*., 2008). HnRNP L has been shown to act either as a repressor or as an activator of alternative splicing depending on the location of its binding sites (Hui *et al*., 2005). One of the well-characterised target genes of hnRNP L is the *SLC2A2* gene (Heiner *et al*., 2010). It was demonstrated that if hnRNP L binds to the CA-repeat sequence close to the 5' splice site of the alternative exon 4 recruitment of the U1 snRNP is impaired resulting in exon 4 skipping.

Here I have investigated the function of the different domains of hnRNP L in the regulation of alternative splicing, namely repression of alternative exon 4 inclusion of a *SLC2A2* minigene construct. The aim of this study was to determine whether specific domains of hnRNP L are necessary for the splicing repressor activity of the protein.

Splicing complementation assays confirmed that alternative splicing of *SLC2A2* is effectively regulated by full-length hnRNP L as well as by LΔG protein (Fig. 3.22B), indicating that the glycine-rich region is not required for splicing repression. Deletion mutant proteins LΔC (RRMs 1 and 2) and LΔN (RRMs 3 and 4) did not restore the inhibitory function on exon inclusion (Fig. 3.22C). These findings suggest that high-affinity binding of LΔC protein to CA-repeat RNA is not sufficient for regulation of alternative splicing. Due to the absence of two N-terminal RRMs the functionally important homodimerisation or interaction with other splicing factors might fail and the LΔC protein is unable to regulate exon repression.

The contribution of all four RRMs in alternative splicing regulation was confirmed using mutant proteins with amino acid substitutions in RRM1 and RRM2 (Fig. 3.22D) in splicing complementation assays. These proteins did not abolish the alternative splicing activity of hnRNP L, except for the mutation at position 225 in RRM2, although the point mutations resulted in reduced binding affinity to CA-repeat RNA.

It was shown before that RRM3 and RRM4 of hnRNP L make intensive inter-domain contacts (Skrisovska and Allain, 2008) and that this interaction is important for their RNA-binding activity (Fig. 3.10). Analysis of alternative splicing activity of

hnRNP L indicates that the repressor function of the protein was abolished upon the disruption of the inter-domain interaction between these two domains (Fig. 3.22E).
Previously, the important role of the inter-domain interaction between RRM3 and RRM4 in splicing regulation was shown for PTB protein (homolog of hnRNP L). Deletion of the entire RRM4 or the 12 amino acid region of RRM4 abolished the repression activity of PTB (Haiyng *et al*., 2002). Later the 12 amino acid region was mapped by NMR spectroscopy as a part of an unusually large inter-domain interface between RRM3 and RRM4 (Vitali *et al*., 2006), demonstrating that disruption of the interaction had an effect on splicing activity of PTB.

4.8 Perspectives

The structural study of hnRNP L and hnRNP LL in their free and RNA bound states will help to gain insight into the protein's three-dimensional structure, providing a detailed knowledge about the organisation of the proteins' RRMs. The structural analysis of the RNA-protein complex will reveal the basis for sequence-specific recognition of singe-stranded RNA, high-affinity binding of hnRNP L and hnRNP LL to certain RNA sequences, and functional synergy between the RRMs.
The sequence specificity of hnRNP L appears to be similar for RNA and ssDNA (data not shown). The three-dimensional study of hnRNP L in complex with either RNA or ssDNA will give information about differences in binding to these substrates.
The inter-RRM linker is highly conserved in length and sequence and has been implicated in RNA binding and alternative splicing regulation (Burd *at al*., 1994; Mayeda *et al*. 1998). The structure analysis of hnRNP L will demonstrate whether the linker segment carrying the glycine- and proline-rich regions are directly involved in both protein–protein and protein–RNA interactions.
Further studies should aim to identify the interaction partners for hnRNP L that are required for regulation of splicing activity of the protein. The study with the deletion mutants should be continued to identify the mechanism and the regions of the hnRNP L protein involved in homodimerisation and cooperative binding. For many proteins, for example PTB, multimerisation and cooperative binding to target RNAs are functionally important. The zone of repression formed by PTB through multimerisation across an alternative exon or between binding sites flanking an alternative exon is one of the mechanisms of splicing regulation (Wagner *at al*., 2001; Auweter *at al*., 2008).
Number of studies should be done to identify the post-transcriptional modifications of hnRNP L and hnRNP LL proteins and the role of such modifications in splicing

regulation. It has been shown that pre-mRNA splicing, as well as other biological processes, can be regulated both positively and negatively by reversible protein phosphorylation (Mermoud *et al*., 1994). For example, the function of hnRNP C proteins in pre-spliceosome assembly is coupled to dynamic cycle of their phosphorylation and dephosphorylation (Mayrand *et al*., 1993). It has been shown that hnRNP A1 accumulates in cytoplasm upon its phosphorylation (Allemand *et al*., 2005). The decrease in nuclear levels of hnRNP A1 led to alternative splicing changes in an E1A minigene reporter.

5. References

Allain, F.H-T., Bouvet, P., Dieckmann, T., Feigon, J. (2000). Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **19**, 6870-6881.

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. *Proc. Nat. Acad. Sci.USA* **102**, 3605-3610.

Auweter, S.D., Fasan, R., Reymond, L., Underwood, J.G., Black, D.L., Pitsch, S., Allain, F.H-T. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* **25**, 163-173.

Auweter, S.D., Allain, F.H-T. (2008). Structure-function relationships of the polypyrimidine tract binding protein. *Cell. Mol. Life Sci.* **65**, 516-527.

Birney, E., Kumar, S., Krainer, A.R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803-5816.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291-336.

Blanchette, M., Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939 -1952.

Blencowe, B.J., Bowman, J.A., McCracken, S.. Rosonina, E. (1999). SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.* **77**, 277-291.

Blencowe, J. (2000). Exonic splicing enhancer: mechanism of action, diversity and role in human genetic diseases. *Elsevier Science Ltd.* 106-110.

Boukis, L.A., Liu, N., Furuyama, S., Bruzik, J.P. (2004). Ser/Arg-rich protein-mediated communication between U1 and U2 small nuclear ribonucleoprotein particles. *J. Biol. Chem.* **279**, 29647-29653.

Burd, C.G., Matunis, E.L., Dreyfuss, G. (1991). The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. *Mol. Cell. Biol.* **11**, 3419-3424.

Burd, C.G., Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-621.

Burge, C.B., Padgett, R.A., Sharp, P.A. (1998). Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**, 773-785.

Caceres, J.F., Kreiner, A.R. (1993). Functional analysis of pre-mRNA splicing factor SF/ASF structural domains. *EMBO J.* **12**, 4715-4726.

Calero, G., Wilson, K.F., Ly, T., Rios-Stener, J.L., Clardy, J.C., Cerione, R.A. (2002). Structural basis of m7GpppG binding to the nuclear cap-biding protein complex. *Nature Struct. Biol.* **9**, 912-917.

Caputi, M., Zahler, A.M. (2001). Determination of the RNA Binding Specificity of the Heterogeneous Nuclear Ribonucleoprotein (hnRNP) H/H′/F/2H9 Family. *J. Biol. Chem.* **276**, 43850-43859.

Carmel, I., Tal, S., Vig, I., Ast, G. (2004). Comparative anylysis defects dependencies among the 5' splice-site positions. *RNA* **10**, 828-840.

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., 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.

Chung, S., Jiang, L., Cheng, S., Furneaux, H. (1996). Purification and Properties of HuD, a Neuronal RNA-binding Protein. *J. Biol. Chem.* **271**, 11518-11524.

Coggill, P., Finn, R.D., Bateman, A. (2008). Identifying protein domains with the Pfam database. *Curr. Protoc. Bioinformatics.*

Conte, M.R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132-3141.

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

D'Ambrogio, A., Buratti, E., Stuani, C., Guarnaccia, C., Romano, M., Ayala, Y.M., Baralle, F.E. (2009). Functional mapping of the interaction between TDP-43 and hnRNP A2 *in vivo*. *Nucleic Acids Res.* **37**, 4116-4126.

Dember, L.M., Kim, N.D., Lui, K-Q., Anderson, P. (1996). Individual RNA Recognition Motifs of TIA-1 and TIAR Have Different RNA Binding Specificities. *J. Biol. Chem.* **271**, 2783-2788.

Deo, R.C., Bonanno, J.B., Sonenberg, N., Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835-845.

Dominguez, C., Allain, F.H-T. (2006). NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res.* **34**, 3634-3645.

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

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

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.

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

Glisovic, T., Bachorik, J.L., Yong, J., Dreyfuss, G. (2008). RNA-binding proteins and posttranscriptional gene regulation. *FEBS Letters* **582**, 1977-1986.

Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100-107.

Graveley, B.R. (2009). Alternative splicing: regulation without regulators. *Nature Struct. Mol. Biol.* **16**, 13-15.

Green, M. R. (1986). Pre-mRNA Splicing. *Ann. Rev. Genet.* **20**, 671-708.

Green, S.R., Manche, L., Mathews, M.B. (1995). Two functionally distinct RNA-binding motifs in the regulatory domain of the protein kinase DAI. *Mol. Cell. Biol.* **15**, 358-364.

Guang, S., Felthauser, 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.

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. Virology* **72**, 8782-8788.

Haiying, L., Zhang, W., Reed, R.B., Liu, W., Grabowski, P.J. (2002). Mutation in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein. *RNA* **8**, 137-149.

Hall, S.L., Padgett, R.A. (1994). Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357-365.

Hall, S.L., Padgett, R.A. (1996). Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**, 1716-1718.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., Yokoyama, S. (1999). Structural dasis for recognition of trs mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.

Hartmann, K., Bindereif, A., Schön, A., Westhof, E. (2005). Handbook of RNA Biochemistry. *Wiley-VCH Verlag GmbH & Co. KGaA*, Chapret 6, 86-93.

Heiner, M., Hui, J., Schreiner, S., Hung, L.H., Bindereif, A. (2010). HnRNP L – mediated regulation of mammalian alternative splicing by interference with splice site recognition. *RNA Biol.* **7**, 56-64.

Heinrichs, V., Ryner, L.C., Baker, B.C. (1997). Regulation of sex-specific selection of fruitless 5' splice sites by transformer and transformer-2. *Mol. Cell. Biol.* **18**, 450-458.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429-3431.

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. *Nature 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.

Inceoglu, A.B., Kamita, S.G., Hinton, A.C., Huang, Q., Severson, T.F., Kang, K., Hammock, B.D. (2001). Recombinant baculoviruses for insect control. *Pest. Management Science* **57**, 981-987.

Jahansson, C., Finger, L.D., Trantirek, L., Mueller, T.D., Kim, S., Laird-Offringa, I.A., Feigon, J. (2004). Solution structure of the complex formed by the two N-terminal RNA-binding domains of Nucleolin and pre-rRNA target. *J. Mol. Biol.* **337**, 799-816.

Kan, Z., Rouchka, E.C., Gish, W.R., State, D.J. (2001). Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.* **11**, 889-900.

Kanopka, A., Mühlemann, O., Akusjarvi, G. (1996). SR proteins that are essential for generic pre-mRNA splicing inhibit splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535-538.

Kent, W.J. (2002). BLAT-The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656-664.

Kharrat, A., Macias, M.J., Gibson, T.J., Nilges, M., Pastore, A. (1995). Structure of the dsRNA binding domain of E. coli RNase III. *EMBO J.* **14**, 3572-3584.

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

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

Kleinschmidt, C., Tovar, K., Hillen, W. (1991). Computer simulations and experimental studies of gel mobility patterns for weak and strong non-cooperative protein binding to two targets on the same DNA: application to binding of Tet repressor variants to multiple and single tet operator sites. *Nucleic Acids Res.* **19**, 1021-1028.

Koenig, H., Matter, N., Bader, R., Thiele, T., Müller, F. (2007). Splicing Segregation: The Minor Spliceosome Acts outside the Nucleus and Controls Cell Proliferation. *Cell* **131**, 718-729.

Konarska, M.M., Vilardell, J., Query, C.C. (2006). Repositioning of the reaction intermediate within the catalytic center of the spliceosome. *Mol. Cell* **21**, 543-553.

Kramer, A. (1996). The Structure and Function of Proteins Involved in Mammalian Pre-mRNA Splicing. *Annu. Rev. Biochem.* **65**, 367-409.

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 Genet.* **90**, 41-54.

Krecic, M.K., Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. *Current Opinion in Cell Biology* **11**, 363-371.

Kuninger, D.T., Izumi, T., Papaconstantinou, J., Mitra, S. (2002). Human AP-endonuclease 1 and hnRNP-L interact with a nCaRE-like repressor element in the AP-endonuclease 1 promoter. *Nucleic Acids Res.* **30**, 823-829.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., Brenner, S.E. (2007). Unproductive splicing of SR genes associated with conserved and ultraconserved DNA elements, *Nature* **446**, 926-929.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F.D., Copley, R.R., Schmidt, S., Ponting, C.P., Bork, P. (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30**, 242-244.

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

Lim, L.P., Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. *Proc. Nat. Acad.Sci. USA* **98**, 11193-11198.

Liu, Q., Dreyfuss, G. (1995). *In vivo* and *in vitro* arginine methylation of RNA-binding proteins. *Mol. Cell. Biol.* **15**, 2800-2808.

Liu, X., Mertz, J. E. (1995). HnRNP L binds a cis-acting RNA sequence element that enables intron-independent gene expression. *Genes & Development* **9**, 1766-1780.

Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary diseases? *FEBS Letters* **579**, 1900-1903.

Maayan, C., Kaplan, E., Scachar, S., Peleg, O., Godfrey, S. (1987). Incidence of familial dysautonomia in Israel 1977-1981. *Clin. Genet.* **32**, 106-108.

Majumder, M., Yaman, I., Gaccioli, F., Zeenko, V.V., Wang, C., Caprara, M.G., Venema, R.C., Komar, A.A., Snider, M.D., Hatzoglou, M. (2009). The hnRNA-Binding Proteins hnRNP L and PTB Are Required for Efficient Translation of the Cat-1 Arginine/Lysine Transporter mRNA during Amino Acid Starvation. *Mol. Cell. Biol.* **29**, 2899-2912.

Maris, C., Dominguez, C., Allain, F.H-T. (2005). The RNA recognition motif, a plastic RNA binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118-2131.

Martinez-Contreras, R., Fisette, J-F., Nasim, F.H., Madden, R., Cordeau, M., Chabot, B. (2006). Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. *PLoS Biol.* **4**, 21-29.

Mauger, M.D., Lin, C., Garcia-Blanco, M.A. (2008). HnRNP H and hnRNP F complex with Fox proteins to silence FGFR2 exon IIIc. *Mol. Cell. Biol.* **28**, 5403-5419.

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

Mayeda, A., Munroe, M.H., Xu, R-M., Kreiner, A.R. (1998). Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA* **4**, 1111-1123.

Mayrand, S.H., Dwen, P., Pederson, T. (1993). Serine/threonine phosphorylation regulates binding of C hnRNP proteins to pre-mRNA. *Proc. Nat. Acad.Sci. USA* **90**, 7764-7768.

Mazza, A.C., Segref, A., Mattaj, I.W., Cusack, S. (2002). Large-scale induced fit recognition of am m7GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* **21**, 5548-5557.

McKernan, K.J., Peckham, H.E., Costa, G.L. and other authors (2009). Sequence and structural variation in human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **20**, 1527-1541.

Mermoud, J.E., Cohen, P.T., Lamond, A.I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* **13**, 5679-5688.

Modrek, B., Resch, A., Grasso, C., Lee, C. (2001). Genome-wide analysis of alternative splicing using human expressed sequence data. *Nucleic Acids Res.* **29**, 2850-2859.

Moore, M.J., Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364-368.

Motta-Mena, L.B., Heyd, F., Lynch, K. (2010). Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol. Cell* **37**, 223-234.

Mironov, A.A., Fickett, J.W., Gelfand, M.S. (1999). Frequent alternative splicing of human genes. *Genome Res.* **9**, 1288-1293.

Musco, G., Stier, G., Joseph, C., Morelli, M.A.C., Nilges, M., Gibson, T.J., Pastore, A. (1996). Sequence-Specific RNA Binding by a Nova KH Domain. *Cell* **85**, 237-245.

Nagai, K., Oubridge, C., Ito, N., Avis, J., Evans, P. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.* **20**, 235-240.

Nilsen, T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays* **25**, 1147-1149.

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.

Oberstrass, F.C., Auweter, S.D., Frat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L., Allain, F.H-T. (2005). Structure of PTB Bound to RNA: Specific Binding and Implications for Splicing Regulation. *Science* **309**, 2054-2057.

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, A.A., Steitz, J.A. (2003). Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell. Biol.* **4**, 960-970.

Perez, I., Lin, C., 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.

Pinol-Roma, S., Swanson, M.S., Gall, J.G., Dreyfuss, G. (1989). A novel heterogeneous nuclear RNP protein with a unique distribution on nascent transcripts. *J. Biol. Chem.* **6**, 2575-2587.

Rajan, P., Dalgliesh, C., Bourgeois, C.F., Heiner, M., Emami, K., Clark, E.L., Bindereif, A., Stevenin, J., Robson, C.N., Leung, H.Y., Elliott, D.J. (2009). Proteomic identification of heterogenous nuclear ribonucleoprotein L as a novel component of SLM/Sam68 Nuclear Bodies. *BMC Cell Biology* **10**, 82-92.

Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231-1245.

Ray, P.S., Jia, J., Yao, P., Majumder, M., Hatzoglou, M., Fox., P.L. (2008). A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915-919.

Roscigno, R.F., Garcia-Blanco M.A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* **1**, 692-706.

Rossbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., Bindereif, A. (2009). Auto- and crossregulation 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.

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.

Price, S.R., Evans, P.R., Nagai, K. (1998). Crystal structure of the spliceosomal U2B''-U2A' proteincomplex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645-650.

Sakharkar, M.K., Chow, V.T-K., Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biology* **4**, 387-393.

Samuels, M., Deshpande, G., Schedl, P. (1998). Activities of the Sex-lethal proteins in RNA binding and protein-protein interactions. *Nucleic Acids Res.* **26**, 2625-2637.

Sanford, J.R., Longman, D. (2003). Multiple roles of SR protein family in the splicing regulation. *Prog. Mol. Subcell. Biol.* **31**, 33-58.

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.

Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J., Mattaj, I.W. (1989). Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163-4170.

Scherly, D., Dathan, N.A., Boelens, W., van Venrooij, W.J., Mattaj, I.W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B" and their cognate RNAs. *EMBO J.* **9**, 3675-3681.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P. (1998). SMART, a simple modular architecture research tool: Identification of signalling domains. *Proc. Nat. Acad.Sci. USA* **95**, 5857-5864.

Shamoo, Y., Adbual-Manan, N., Patten, A.M., Crawford, J.K., Pellegrini, M.C., Williams, K.R. (1994). Both RNA-binding domains in heterogeneous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry* **33**, 8272-8282.

Shamoo, Y., Abdul-Manan, N., Williams, K.R. (1995). Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.* **23**, 725-728.

Shepard, P.J., Hertel, K.J. (2009). The SR protein family. *Genome Biol.* **10**, 242-251.

Siliciano, P.G., Guthrie, C. (1988). 5' Splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & Development* **2**, 1258-1267.

Serin, G., Joseph, G., Ghisolfi, L., Bauzan, M., Erard, m., Amalric, L., Bouvet, P. (1997). Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J. Biol. Chem.* **272**, 13109-13116.

Shen, H., Kan, J.L., Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote pre-spliceosome assembly. *Mol. Cell* **13**, 367-376.

Sheth, N., Roca, X., Hastings, M. L., Roeder, T., Krainer, A. R., Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic 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 *J. Biol. Chem.* **274**, 1359-1365.

Skrisovska, L., Allain, F.H-T. (2008). Improved Segmental Isotope Labeling Methods for the NMR Study of Multidomain or Large Proteins: Application to the RRMs of Npl3p and hnRNP L. *J. Mol. Biol.* **375**, 151-164.

Stangl, K., Cascorbi, I., Laule, M., and 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.

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

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **6**, 1102-1104.

Taguchi, F., Kusaba, H., Asai, A., Iwamoto, Y., Yano, K., Nakano, H., Muzukami, T., Saijo, N., Kato, H., Nishio, K. (2004). HnRNP L enhances sensitivity of the cells to KW-2189. *Int. J. Cancer* **108**, 679-685.

Tarn, W.Y., Steitz, J.A. (1996a). A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* **84**, 801-811.

Tarn, W.Y., Steitz, J.A. (1996b). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824-1832.

Tarn, W.Y., Steitz, J.A. (1997). Pre-mRNA splicing: The discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**, 132-137.

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

Tian, M., Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. *Genes & Development* **8**, 1703-1712.

Tintaru, A.M., Hautbergue, G.M., Hounslow, A.M., Hung, M-L., Lian, L-Y., Craven, C.J., Wilson, S.A. (2007). Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep.* **8**, 756-762.

Tong, A., Nguyen, J., Lynch, C.W. (2005). Differential expression of CD45 isoforms is controlled by the combined activity of basal and inducible splicing-regulatory elements in each of the variable exons. *J. Biol. Chem.* **280**, 38297-38304.

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.

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

Vitali, F., Henning, A., Oberstrass, F.C., Hargous, Y., Auweter, S.D., Erat, M., Allain, F.H-T. (2006). Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **25**, 150-162.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, C.P., Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476.

Wang, G.S., Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **8**, 749-761.

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. *Nucleic Acids Res.* **33**, 5053-5062.

Wang, X., Tanaka Hall, T.M. (2001). Structural basis for recognition of UA-rich element RNA by the Hud protein. *Nature Struct. Biol.* **7**, 141-145.

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.

Wagner, E. J., Garcia-Blanco, M. A. (2001). Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Mol. Cell. Biol.* **21**, 3281-3288.

Will, C.L., Luhrmann, R. (2005). Splicing of a rare class of introns by the U12-dependent spliceosome. *Biological Chemistry* **386**, 713-724.

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

Wu, Z., Jia, X., Cruz, L., Su, X-C., Marzolf, B., Trousch, P., Zak, D., Hamilton, A., Whittle, B., Sheahan, D., Bertram, E., Aderem, A., Otting, G., Goodnow, C.C., Hoyne, G.F. (2008). Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. *Cell Immunity* **29**, 863-875.

Yuan, W., Xie, J., Long, C., Erdjument-Bromage, H., Ding, X., Zheng, Y., Tempst, P., Chen, S., Zhu, B., Reinberg, D. (2009). Heterogeneous Nuclear Ribonucleoprotein L Is a Subunit of Human KMT3a/Set2 Complex Required for H3 Lys-36 Trimethylation Activity *in Vivo*. *J. Biol. Chem.* **284**, 15701-15707.

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

Zhang, X.T., Chasin, L. (2004). Computational definition of sequence motifs governing constitutive exon splicing. *Genes & Development* **18**, 1241-1250.

Zhang, X.T., Kangsamaksin, M., Chao, J., Chasin, L. (2005). Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol. Cell. Biol.* **25**, 7323-7332.

Zheng, Z-M. (2004). Regulation of Alternative RNA Splicing by Exon Definition and Exon Sequences in Viral and Mammalian Gene Expression. *J. Biomed*. *Sci.* **11**, 278-94.

Zhou, Z., Licklider, L.J., Gygi, S.P., Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**, 182-185.

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

6. Appendices

Abbreviations

^{o}C	centigrade
µg	microgram
µl	microliter
µM	micromolar
A	adenosine
aa	amino acid(s)
APS	ammonium persulfate
ARS	activation-responsive sequence
ATP	adenosine triphosphate
BBP	branchpoint binding protein
bp	base bair(s)
BPS	branchpoint sequence
BSA	bovine serum albumin
C	cytidine
ca.	approximately
Ci	Curie
CBP	Cap binding protein
cDNA	complementary DNA
CLIP	*in vivo* Cross-Linking and ImmunoPrecipitation
cm	centimetre
cpm	counts per minute
CTP	cytidine triphosphate
DMPC	dimetyl pyrocarbonate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
E.coli	*Escherichia coli*
e.g.	exempli gratia (=for example)
ECL	enhanced chemiluminiscence
EDTA	ethylenediamine tetraacetic acid
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
EST	expressed sequence tag
et al.	et alii (=and others)

FCS	fetal calf serum
g	gram
g	acceleration of gravity
G	guanosine
GST	glutathione sulfate transferase
GTP	guanosine triphosphate
h	hour(s)
HEPES	N-2-hydroxyethylpiperazine
His	histidine
hnRNA	heterogeneous nuclear RNA
hnRNP	heterogeneous nuclear ribonucleoprotein
hnRNP LL	hnRNP L-like
IgG	immunoglobulin G
IPTG	isopropyl-1-thio-β-Dgalactoside
ISE	intronic splicing enhancer
ISRE	intronic splicing regulatory element
ISS	intronic splicing silencer
kb	kilobasepair
K_D	dissociation constant
kDa	kilodalton
M	molar
mg	milligram
min	minute(s)
ml	milliliter
mM	millimolar
mmol	Millimole
mRNA	messenger RNA
mut	mutant
N	any nucleotide
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid
nm	nanometer
nM	nanomolar
NMD	nonsense-mediated decay
NP-40	nonidet P-40
nt	nucleotide(s)
NTP	ribonucleoside triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis

PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PK	proteinase K
pmol	picomol
PMSF	phenylmethylsulfonyl fluoride
Poly(A)	polyadenylic acid
pre-mRNA	precursor messenger RNA
PTB	polyprymidine tract binding protein
PTC	premature termination codon
RBD	RNA-binding domain
RBP	RNA-binding protein
RNA	ribonucleic acid
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
SLC2A2	solute carrier family 2
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
SR	serine-arginine-rich
T	thymidine
TJP1	tight junction protein
tRNA	transfer RNA
U	uracil
U	Unit
U2AF	U2 auxiliary factor
U2AF35	U2AF 35 kDa subunit
U2AF65	U2AF 65 kDa subunit
UTR	untranslated region
v/v	volume per volume
w/v	weight per volume
WT	wildtype
Y	pyrimidine base

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.

Publications:

Listerman I., Bledau A.S., Grishina I., Neugebauer K.M. Extragenic Accumulation of RNA Polymerase II Enhances Transcription by RNA Polymerase III. PLOS Genetics 3, 2268-2277, 2007

Rossbach O., Hung L.H., Schreiner S., Grishina I., Heiner M., Hui J., Bindereif A. Auto- and crossregulation of the hnRNP L proteins by alternative splicing. Mol. Cell. Biol. 29, 1442-1451, 2009

Sapra A.K., Änkö M-L., Grishina I., Lorenz M., Pabis M., Poser I.,Rollins J., Weiland E-M., Neugebauer K.M. SR protein family members display diverse activities in the formation of nascent and mature mRNA *in vivo*. Mol. Cell 34, 179-190, 2009

Meeting Abstracts:

Grishina I., Hui J., Bindereif A. Structure-function analysis of heterogeneous ribonucleoprotein L. Workshop “Enzymes and enzyme complexes action on nucleic acids”, Vilnius, 2010 (poster presentation)

Rossbach O., Preussner M., Grishina I., Schreiner S., Heiner M., Hung L-H., Bindereif A. HnRNP L versus hnRNP L-like: Mutational Analysis of Functional Domain Structure and Identification of Target Genes by RNA-Sequencing and CLIP. Annual meeting of the RNA society. Seattle, 2010 (conference material)

Grishina I., Hui J., Bindereif A. Structure-function analysis of heterogeneous ribonucleoprotein L. International Symposium “Control of Gene Expression and Cancer”, Moscow, 2010 (poster presentation)

Acknowledgements

I would like to thank Prof. Dr. Albrecht Bindereif for giving me the opportunity to work in his lab and for his support during my Ph.D. research.

I want to acknowledge Prof. Dr. Michael Niepmann for reviewing my thesis.

Many thanks to the former and current members of the lab Tanja Rösel, Monika Heiner, Silke Schreiner, Lee-Hsueh Hung, Nicolas Jae, Christian Preußer, Konstantin Licht, Marco Preußner, Oliver Roßbach, Zsofia Palfi for providing a good working atmosphere. Without you “Jungs”, the work here would only be half as much fun!

I would also like to thank Tanja Rösel and Monika Heiner for being such a good colleagues and friends.

Мои любимые родители и сестренка, спасибо вам за вашу бесконечную поддержку и веру в меня.

Dima, I thank you very much for your encouragement and understanding. You were always with me in my good and bad times. Thank you for always believing in me.

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 12.07.2010

Inna Grishina