

Characterization of YT521-B, a novel RNA binding protein regulated by tyrosine phosphorylation

Den Naturwissenschaftlichen Fakultäten
der Friedrich-Alexander-Universität Erlangen-Nürnberg
zur
Erlangung des Doktorgrades

vorgelegt von
Zhaiyi Zhang
aus Tianjin, China

2007

Als Dissertation genehmigt von den Naturwissen-
schaftlichen Fakultäten der Universität Erlangen-Nürnberg.

Tag der mündlichen Prüfung:

20 September 2007

Vorsitzender der Prüfungskommission:

Prof. Dr. Eberhard Bänsch

Erstberichterstatte:

Prof. Dr. Andreas Burkovski

Zweitberichterstatte:

Prof. Dr. Michael Wegner

We shall not cease from exploration.

And the end of all our exploring will be to arrive where we started,

and know the place for the first time.

Thomas. S. Eliot
"Collected Poems 1962"

To my parents
To my sister

ACKNOWLEDGMENTS

This work presented here was performed in the Institute of Biochemistry at Friedrich-Alexander-University Erlangen-Nürnberg. There are many people to whom I owe the fulfillment of this dissertation.

My sincerest gratitude goes to Prof. Stefan Stamm for his scientific guide, his confidence in me all these years and his strong support in helping me work freely.

I would also like to thank the present and former members of the Stamms lab, Bettina Heinrich, Tatyana Novoyatleva, Shivendra Kishore, Natalya Benderska, Yesheng Tang, Jinyi Hui, Amit Khanna, David Smith, Ilona Rafalska and Dominique Olbert for creating such a friendly and comfortable working atmosphere.

I am also thankful to all the members of the Institute of Biochemistry, in particular, Prof. Wegner and his group for sharing their resources with us.

Many thanks are due to Dr. Horst Wolff and Prof. Ruth Brack-Werner from the Institute of Molecular Virology, GSF in Neuherberg for performing protein shuttling assay. I'm thankful to Professor Janusz Bujnicki from the international Institute of Molecular and Cell Biology in Warsaw for YTH model generation. I also would like to thank Michael Hiller for nice collaboration.

I appreciate all my friends for their cheerful company and mental support.

In the end, I would like to thank my parents and my sister for their unconditional love, for their patience and for their encouragement in times when I needed most. Without their support, I couldn't be able to accomplish this work.

Many many thanks!

PUBLICATIONS

Parts of this work are included in the following publications:

Zhang, Z., Rafalska, I., Hiller, M., de la Grange, P., Pudimat, R., Heinrich, B. and Stamm, S. the YTH domain is a novel RNA binding domain (submitted)

Hiller, M., **Zhang, Z.**, Backofen, R. and Stamm, S. Pre-mRNA secondary structures influence exon recognition (accepted)

Rafalska, I.¹, **Zhang, Z.**¹, Benderska, N., Wolff, H., Hartmann, A.M., Brack-Werner, R. and Stamm, S. (2004) The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum. Mol. Genet.*, 13, 1535-1549

¹: equal contribution to the manuscript

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., **Zhang, Z.**, Toiber, D., Thanaraj, T.A. and Soreq, H. (2005) Function of alternative splicing. *Gene*, 344, 1-20

Wilkinson, F.L., Holaska, J. M., **Zhang, Z.**, Sharma, A., Manilal, S., Holt, I., Stamm, S., Wilson, K.L. and Morris G.E. (2003) Emerin interacts in vitro with the splicing-associated factor, YT521-B. *Eur. J. Biochem.* 270, 2459–2466

Heinrich, B., **Zhang, Z.**, Novoyatleva, T. and Stamm, S. (2005) Aberrant pre-mRNA splicing as a cause of human disease. *Journal of Clinical Ligand Assay*, 7, 68-74.

CONTENTS

| | |
|--|------------|
| CONTENTS | I |
| ABBREVIATIONS..... | IV |
| ZUSAMENFASSUNG | VI |
| ABSTRACT | VII |
| | |
| 1. INTRODUCTION | 1 |
| 1.1. RNA binding domains | 1 |
| 1.1.1. RRM | 2 |
| 1.1.2. KH domain | 3 |
| 1.1.3. PAZ..... | 4 |
| 1.1.4. OB folds..... | 4 |
| 1.2. Splicing..... | 5 |
| 1.2.1. Splicing passway and spliceosome assembly | 6 |
| 1.2.2. Exon recognition and intron bridging | 9 |
| 1.3. Alternative splicing | 11 |
| 1.3.1. Regulation of alternative splicing | 13 |
| 1.3.1.1. Phosphorylation-dependent control of the splicing machinery | 14 |
| 1.3.1.2. Signaling to the splicing machinery | 14 |
| 1.3.2. Function of alternative splicing | 15 |
| 1.3.2.1. Introduction of stop codons..... | 16 |
| 1.3.2.2. Changes of the protein structure | 17 |
| 1.3.2.3. Changes in the untranslated region | 18 |
| 1.4. Emery-Dreifuss muscular dystrophy | 18 |
| 1.5. pre-mRNA secondary structure influence exon recognition..... | 19 |
| | |
| 2. RESEARCH OVERVIEW | 21 |
| | |
| 3. MATERIALS AND METHODS..... | 22 |
| 3.1. Materials | 22 |
| 3.1.1. Chemicals | 22 |
| 3.1.2. Enzymes and enzyme inhibitor..... | 23 |
| 3.1.3. Cell lines and media..... | 24 |
| 3.1.4. Bacterial stains and media | 24 |
| 3.1.4.1. Bacterial stains | 24 |
| 3.1.4.2. Media | 24 |
| 3.1.5. Antibiotics | 25 |
| 3.1.6. Antibodies..... | 25 |
| 3.1.7. Plasmids..... | 25 |
| 3.1.7.1. Clones from the Stamm's lab collection or outside sources..... | 25 |
| 3.1.7.2. Newly made clones | 26 |
| 3.1.8. Primers..... | 27 |
| 3.1.8.1. Primers used for cloning | 27 |
| 3.1.8.2. Primers used for site-directed mutagenesis..... | 28 |
| 3.1.8.3. Primers used for overlap extension | 28 |

Contents

| | |
|--|-----------|
| 3.1.8.4. Primers used for RT-PCR | 29 |
| 3.1.8.5. Primers used for SELEX | 29 |
| 3.1.8.6. Primers used for microarray data validation | 29 |
| 3.1.9. siRNA sequence..... | 29 |
| 3.2. Methods | 29 |
| 3.2.1. Plasmid DNA isolation | 29 |
| 3.2.2. Determination of nucleic acids concentration..... | 30 |
| 3.2.3. Restriction endonuclease digestion..... | 30 |
| 3.2.4. Electrophoresis of DNA..... | 31 |
| 3.2.5. Elution of DNA from agarose gels | 31 |
| 3.2.6. PCR amplification of DNA..... | 31 |
| 3.2.7. DNA ligation | 31 |
| 3.2.8. Preparation of competent <i>E. coli</i> cells | 32 |
| 3.2.9. Transformation of <i>E. coli</i> cells | 32 |
| 3.2.10. Site-direction mutagenesis of DNA | 33 |
| 3.2.11. Domain deletion by overlap extension | 33 |
| 3.2.12. Expression and purification of HIS-tagged proteins in Baculovirus Expression System | 35 |
| 3.2.12.1. Expression of recombinant protein | 35 |
| 3.2.12.2. Purification of recombinant protein | 36 |
| 3.2.13. Determination of protein concentration | 37 |
| 3.2.14. Systematic evolution of ligands by exponential enrichment (SELEX)..... | 37 |
| 3.2.14.1. <i>In vitro</i> transcription..... | 38 |
| 3.2.15. Minigene construction | 39 |
| 3.2.16. RNA electrophoretic mobility shift assay (gel shift) | 39 |
| 3.2.16.1. Labeling of RNA probe..... | 40 |
| 3.2.17. Light scattering assay | 40 |
| 3.2.18. Freezing, thawing and subculturing of eukaryotic cells..... | 41 |
| 3.2.19. Transfection of eukaryotic cells..... | 41 |
| 3.2.19.1. Plasmid transfection..... | 41 |
| 3.2.19.2. siRNA transfection..... | 42 |
| 3.2.19.3. Plasmid and siRNA transfection | 42 |
| 3.2.20. Fixing attached eukaryotic cells on cover slips | 42 |
| 3.2.21. Immunostaining | 42 |
| 3.2.22. Immunoprecipitation of proteins..... | 43 |
| 3.2.23. Electrophoresis of proteins | 44 |
| 3.2.24. Staining of protein gels | 44 |
| 3.2.25. Western blotting..... | 44 |
| 3.2.26. Solubility assay | 45 |
| 3.2.27. <i>In vivo</i> splicing assay | 46 |
| 3.2.28. Isolation of total RNA | 47 |
| 3.2.29. RT-PCR | 47 |
| 3.2.30. Affymetrix Mouse Exon 1.0 ST array analysis..... | 47 |
| 3.3. Databases and computational tools | 47 |
| 4. RESULTS | 49 |
| 4.1. Modular structure of YT521-B | 49 |
| 4.2. The YTH domain is a novel RNA recognition domain | 50 |
| 4.3. Influence of YT521-B binding motifs on alternative splicing in vivo | 58 |
| 4.4. Effect of decreasing YT521-B concentration on splice site selection | 60 |
| 4.5. Three of the conserved residues in the YTH domain are critical for splice site selection | 61 |
| 4.6. Identification of endogenous target exons of YT521-B by microarray | 63 |
| 4.7. YT521-B is phosphorylated on tyrosine residue | 64 |

| | |
|---|-----------|
| 4.8. YT521-B localizes with c-Abl kinase in the nucleus..... | 66 |
| 4.9. YT521-B binds to membrane bound tyrosine kinase c-Src..... | 66 |
| 4.10. YT521-B shuttles between nucleus and cytosol..... | 67 |
| 4.11. Tyrosine phosphorylation regulates YT521-B solubility..... | 68 |
| 4.12. Tyrosine phosphorylation changes the ability of YT521-B to influence SRp20 alternative pre-mRNA splicing..... | 70 |
| 4.13. Interaction with Emerin influences YT-521-B on splice site selection of CD44 minigene..... | 71 |
| 4.14. The activity of splicing regulatory sequences depends on their conformation in the premRNA..... | 72 |
| 5. DISCUSSION..... | 75 |
| 5.1. The YTH domain is a novel RNA binding domain..... | 75 |
| 5.2. YT521-B is a vertebrate specific splicing factor..... | 77 |
| 5.3. YT521-B nuclear localization and binding ability are regulated by tyrosine phosphorylation..... | 78 |
| 5.4. The influence of YT521-B on splice site selection is regulated by tyrosine phosphorylation..... | 79 |
| 5.5. Emerin binds to YT521-B and abolishes its influence on splice site selection..... | 81 |
| 5.6. YT521-B working model..... | 82 |
| REFERENCES..... | 84 |

ABBREVIATIONS

| | |
|-------------------|---|
| 9G8 | splicing factor, arginine/serine-rich 7 |
| ActD | actinomycin D |
| ASD | alternative splicing database |
| ASF | alternative splicing factor |
| ATP | adenosine 5'-triphosphate |
| BAF | barrier-to-autointegration factor |
| bp | base pairs |
| BSA | bovine serum albumin |
| CBP | CREB binding protein |
| CBs | Cajal bodies |
| cDNA | complementary DNA |
| CLK | CDC2-like kinase |
| CT | chromosome territory |
| CTD | carboxyterminal domain (of RNA polymerase II) |
| DFC | dense fibrillar component |
| dH ₂ O | distilled water |
| DMEM | dulbecco's modified eagle medium |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| dNTP | deoxyribonucleotidtriphosphate |
| Dscam | Down syndrome cell adhesion molecule |
| dsRBD | double-stranded RNA-binding domain |
| DTT | dithiothreitol |
| ECL | enhanced chemiluminiscence |
| EDMD | Emery-Dreifuss muscular dystrophy |
| EDTA | ethylenediaminetetraacetic acid |
| EGFP | enhanced green fluorescent protein |
| EMSA | electrophoretic mobility shift assay |
| ESE | exonic splicing enhancer |
| EST | expressed sequence tag |
| FCS | fetal calf serum |
| FGFR | fibroblast growth factor receptor |
| GC | granular component |
| GH | growth hormone |
| HEK | human embryonic kidney |
| hnRNP | heterogenous nuclear ribonucleoprotein |
| IL-4 | interleukin-4 |
| IPTG | isopropyl β -D-1-thiogalactopyranoside |
| kDa | kilodalton |
| KH domain | hnRNP K homology domain |
| LAP | lamina-associated polypeptide |
| LEM | LAP-Emerin-MAN1 |
| mGluR | metabotropic glutamate receptor |
| mRNA | messenger RNA |
| NMD | nonsense-mediated decay |
| NOR | nucleolar organizer region |
| NOVA | neuro-oncological ventral antigen |
| NPC | nuclear pore complex |

Abbreviations

| | |
|------------|---|
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PKC | protein kinase C |
| PML | promyelocytic leukemia |
| PMSF | phenylmethanesulfonyl fluoride |
| POD | PML oncogenic domain |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| rpm | revolutions per minute |
| RRM | RNA recognition motif |
| RS domain | arginine/serine rich domain |
| RT | room temperature |
| RT-PCR | reverse transcription followed by polymerase chain reaction |
| RUST | regulated unproductive splicing and translation |
| SAF | scaffold attachment factor (A or B) |
| Sam68 | Src associated in mitosis 68kDa |
| SC35 | splicing component, 35 kDa; splicing factor, arginine/serine-rich 2 |
| SDS | sodium dodecyl sulfate |
| SF | splicing factor (1 or 2) |
| SFRS14 | splicing factor, arginine/serine-rich 14 |
| SH | Src homology domain (2 or 3) |
| SLM | Sam68 like molecule (1 or 2) |
| SMA | spinal muscular atrophy |
| SMN | survival motor neuron gene (1 or 2) |
| SNB | Sam68 nuclear body |
| snoRNP | small nucleolar ribonucleoprotein |
| snRNP | small nuclear ribonucleoprotein particle |
| SR-protein | serine-arginine- rich protein |
| STAR | signal transduction and activation of RNA |
| TBE | tris-borate-EDTA buffer |
| TCA | trichloroacetic acid |
| TE | tris-EDTA |
| TEMED | N,N,N',N'-tetramethylethylenediamine |
| Tra2 | transformer 2 |
| tRNA | transfer RNA |
| TSH | thyroid stimulating hormone |
| U2AF | U2 snRNP auxiliary factor (35 or 65 kDa) |
| UTR | untranslated region |
| YTH domain | YT521-B homology domain |

ZUSAMENFASSUNG

Alternatives Spleißen ist einer der wichtigsten Mechanismen, um die Expression der genetischen Information in Metazoen zu regulieren. Die Herstellung von mehreren Spleißvarianten aus einem einzelnen Gen ermöglicht es, eine Vielzahl von Proteinen aus der relativ geringen Anzahl an Genen zu bilden. DNA Microarray Experimente zeigen, dass mindestens 74% aller menschlichen Gene alternativ gespleißt werden. Alternatives Spleißen beeinflusst Zelldifferenzierung, Entwicklung und die genetischen Ursachen von Krankheiten. Dieser Mechanismus ist somit eine grundlegende Komponente der Genregulation. Die Mechanismen der Spleiß-Regulation, insbesondere die damit verbundenen Signaltransduktionswege sind immer noch nicht gut verstanden. Deshalb wurde in dieser Arbeit die Regulation alternativen Spleißens unter Verwendung des Wirbeltier-spezifischen Spleißfaktors YT521-B als Modell untersucht. Das ubiquitär exprimierte Protein YT521-B wurde als Bindungspartner einiger anderer Spleißfaktoren identifiziert. YT521-B kann die Auswahl alternativer Spleißstellen konzentrationsabhängig verändern. Es wurde ausserdem gezeigt, dass YT521-B in einem dynamischen neuartigen nukleären Kompartiment lokalisiert ist, dem so genannten „YT body“.

YT521-B gehört keiner der bekannten Proteinfamilien an und enthält keine bereits bekannten RNA Bindedomänen in seiner Struktur. Im ersten Teil dieser Arbeit wurde gezeigt, dass die bereits durch Vergleich von Proteinsequenzen identifizierte YTH (YT521-B Homologie) Domäne eine neuartige RNA Bindedomäne ist. Diese wurde in 174 Proteinen in Eukaryonten gefunden und ist charakterisiert durch 14 unveränderliche Reste mit einer gemischten alpha-Helix Beta-Faltblatt Faltung. Die YTH Domäne bindet an kurze, degenerierte, einzelsträngige RNA Sequenzmotive. Bindung an RNA verursacht Aggregation des YT521-B Proteins. Die Ergebnisse zeigen, dass YT521-B die Auswahl von alternativen Spleißstellen direkt beeinflussen kann, wenn das Bindungsmotiv in alternativen Exons vorhanden ist. Drei unveränderliche Reste innerhalb der YTH Domäne sind unabdingbar für die Fähigkeit von YT521-B, die Spleißstellenauswahl beeinflussen zu können. Microarray Analysen zeigten, dass YT521-B vorwiegend Wirbeltier-spezifische Exons reguliert.

Im Zweiten Teil der Arbeit wurde gezeigt, dass die subzelluläre Lokalisation und Funktion von YT521-B durch Tyrosin Phosphorylierung reguliert wird. YT521-B wird durch c-Abl im Zellkern phosphoryliert. Das Protein bewegt sich zwischen Zellkern und Cytosol („shuttling“), wo es durch c-Src oder p59^{l^yn} phosphoryliert werden kann. Tyrosin Phosphorylierung verursacht die Verteilung von YT521-B aus den „YT bodies“ in das Kernplasma und verschiebt das phosphorylierte Protein in die unlösliche Kernfraktion. Tyrosin Phosphorylierung von YT521-B verändert auch den Effekt auf die alternative Spleißstellenauswahl. Zusammenfassend wurde gezeigt, dass YT521-B ein Spleißfaktor mit einer neuen Bindedomäne ist, der über Tyrosin Phosphorylierung durch Kinasen der Src Familie reguliert wird.

ABSTRACT

Alternative pre-mRNA splicing is a central mode of genetic regulation in metazoans. Variability in splicing pattern is a major source of generating protein diversity from the relatively limited number of genes. DNA microarray experiments indicate that at least 74% of all human genes are alternatively spliced. Together with its impact on cell differentiation, development and disease, alternative splicing emerges as a fundamental component of gene regulation. The mechanisms of splicing regulation, especially the associated signal-transduction pathways are still not very well understood. Therefore the focus of this project is to understand the mechanism of the alternative splicing regulation, using the vertebrate-specific splicing factor YT521-B as a model. It has been shown that the ubiquitously expressed protein YT521-B was identified as an interactor with several other splicing factors that are implicated in RNA metabolism. YT521-B can change alternative splice site usage in a concentration dependent manner. It has been also shown that YT521-B is located in a dynamic novel nuclear compartment, the YT body.

YT521-B does not belong to any of the known protein families and does not have any previously known RNA binding domains in its structure. One part of this work shows that the YTH (YT521-B Homology) domain previously identified by comparing protein sequences with rat YT521-B is a novel RNA binding domain. This novel RNA binding domain is found in 174 proteins expressed in eukaryotes and is characterized by 14 invariant residues with a mixed alpha-helix beta-sheet fold. The YTH domain binds to a short, degenerated, single-stranded RNA sequence motif. Binding to the RNA causes aggregation of YT521-B protein. The results show that the presence of the binding motif in alternative exons is necessary for YT521-B to directly influence splice site selection *in vivo*. Three invariant residues within the YTH domain are indispensable for the ability of YT521-B to influence splice site selection. Microarray analyses demonstrate that YT521-B predominantly regulates vertebrate-specific exons. The other part of this work showed that subcellular localization and function of YT521-B is regulated by tyrosine phosphorylation. YT521-B is tyrosine phosphorylated by c-Abl in the nucleus. The protein shuttles between nucleus and cytosol, where it can be phosphorylated by c-Src or p59^{fyn}. Tyrosine phosphorylation causes dispersion of YT521-B from YT bodies to the nucleoplasm and moves the phosphorylated protein into insoluble nuclear fraction. Tyrosine phosphorylation of YT521-B can also change the protein effect on alternative splice site selection. In summary, YT521-B was shown to be a splicing factor with a new binding domain that is regulated by tyrosine phosphorylation emanating from Src family kinases.

1. INTRODUCTION

The flow of genetic information from the genome to the proteome occurs in several steps, such as transcription, RNA processing and translation. The eukaryotic messenger RNA (mRNA) processing involves several major events, such as 5' capping, 3' polyadenylation (Shatkin and Manley, 2000), splicing (Kornblihtt et al., 2004) and RNA editing (Wedekind et al., 2003). The 5' capping reaction replaces the triphosphate group at the 5' end of mRNA chain with a modified GTP (7-methylguanosine or m⁷G) which protects it from ribonucleases and also helps recognition of mRNA by ribosome during translation. Splicing is a post transcription modification, in which the intervening sequences (introns) of precursor messenger RNA (pre-mRNA) are removed and the sequences that code for messenger RNA (exons) are joined. Alternations in splice site choice can cause pre-mRNA to be spliced in different ways (alternative splicing). Therefore, the alternative splicing is a major contribution to protein diversity. Polyadenylation at the 3' end increases the stability of the mRNA transcript. RNA editing changes the nucleotide compositions of the mRNA which alters its protein coding information.

These processes are crucial for eukaryotic gene expression. Regulation of gene expression at the posttranscriptional level is mainly achieved by RNA binding proteins (Lunde et al., 2007). This chapter introduces few RNA binding domains, the process of splicing, and regulation and function of the alternative splicing. In addition, the fundamentals of Emery-Dreifuss muscular dystrophy and the influence of pre-mRNA secondary structure on exon recognition are presented.

1.1. RNA binding domains

RNA is rarely “naked”; as soon as RNA is transcribed, ribonucleoproteins (RNPs) wrap and organize nascent RNA transcripts into groups in order to percolate them together down the chain of splicing, nuclear export, stability, translation and degradation so that proteins are efficiently produced to meet the needs of the organism (Keene, 2001; Dreyfuss et al., 2002). Each of these steps is processed and regulated by specific RNA binding proteins (RBPs) (Kornblihtt et al., 2004; Zorio and Bentley, 2004). The diversity of functions of RNA binding proteins suggests a correspondingly large diversity in the structures that are responsible for RNA recognition. More than hundreds of nucleic acid

binding domains were identified. Here some well characterized nucleic acid binding domains are described as examples.

1.1.1. RRM

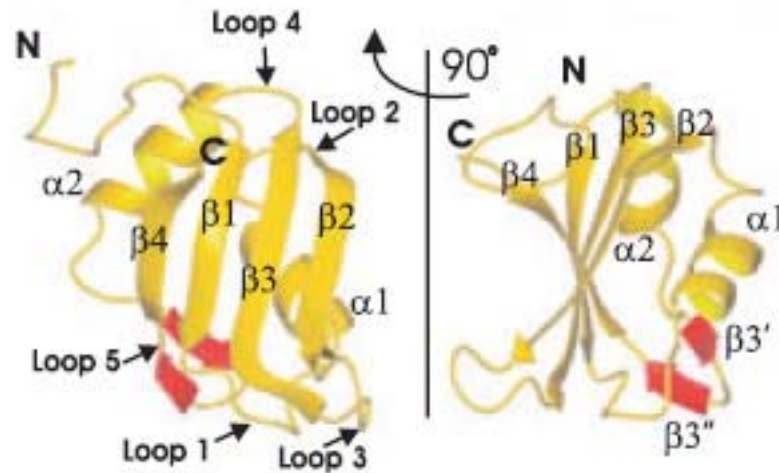


Figure1: Ribbon model of a typical RRM fold, from hnRNP A1. The fold is composed of one four-stranded antiparallel beta-sheet spacially arranged in the order beta4-beta1-beta3-beta2 from left to right when facing the sheet and two helices alpha 1 and alpha 2 packed against the beta-sheet (Maris et al., 2005).

The RNA recognition motif (RRM), also known as ribonucleoprotein domain (RNP) or the RNA binding domain (RBD), identified in the late 1980s, is one of the most abundant protein domain in eukaryotes (Dreyfuss et al., 1988). To date, a total of 6064 RRM motifs have been identified in 3541 different proteins. In humans, about 2% of gene products contain the RRM, and many eukaryotic proteins have multiple copies of RRMs within a protein (Bateman et al, 2002). The abundance of RRM domain reflects the biological importance of this domain in cellular functions. Eukaryotic RRM proteins are involved in all post-transcriptional events, such as pre-mRNA splicing, alternative splicing, mRNA stability, RNA editing, mRNA export, translation regulation and degradation (Birney et al., 1993).

The RRM domain is 90-100 amino acids long and consists of a beta1-alpha1-beta2-beta3-alpha2-beta4 fold with a four-stranded antiparallel beta sheets forming a surface that displays two highly conserved RNP 1 and RNP 2 motifs (Figure 1). The RNP 1 and RNP 2 motifs are localized in the central strands of the beta-sheets, beta3 and beta1, and four conserved residues namely RNP 1 position 1, 3 and 5 and RNP 2 position 2 contribute to sequence-specific single-stranded RNA binding (Maris et al., 2005).

1.1.2. KH domain

The hnRNP-K-homology (KH) domain was initially defined in hnRNP K as a conserved region of 65-70 amino acid residues that is repeated three times in this protein (Musco et al., 1999). A canonical KH domain consists of a stable three-stranded antiparallel beta-sheet, orientated against three helices. This stable beta1-alpha1-alpha2-beta2-beta3-alpha3 fold exposes an RNA-binding surface: the invariable tetrapeptide Gly-X-X-Gly (where X present lysine, arginine or glycine) between the helices alpha1 and alpha2 (Figure 2; Musco et al., 1996; Lewis et al., 1999). The variable loop connecting beta2 and beta3 strands play an essential role in RNA recognition (Lewis et al., 2000).

So far, a large number of KH domain-containing proteins have been identified in a wide variety of species ranging from bacteria to human. These KH domain proteins assume a wide spectrum of biological functions, including transcriptional and translational controls, mRNA stabilization, and mRNA splicing. Different KH domains possess quite different nucleic acid binding specificities. For examples, the neuron specific splicing factor Nova binds to single-strand RNA via its KH3 domain (Lewis et al., 2000); KH domain of poly(C) binding proteins (PCBPs) is able to bind to both CU-rich single-stranded RNA and C-rich single-stranded DNA (Dejgaard and Leffers., 1996; Du et al., 2005); by binding to the repetitive CU-rich sequences on the 3' UTR of LOX mRNA, hnRNP K mediates translational silencing of erythroid 15-lipoxygenase (Ostareck et al., 1997)

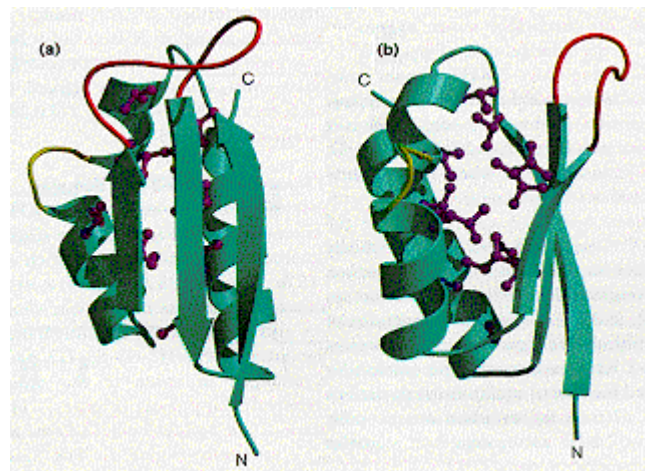


Figure 2: Structure of Nova KH domain. Ribbon diagram of Nova KH3 with conserved aliphatic residues comprising the hydrophobic core illustrated in ball and stick format. The invariant Gly-X-X-Gly is shown in yellow, and the invariant loop in red. (A) A view of the beta-sheet face of the KH domain. (B) A 90 degree rotation from the view in (A) (Lewis et al., 1999).

1.1.3. PAZ

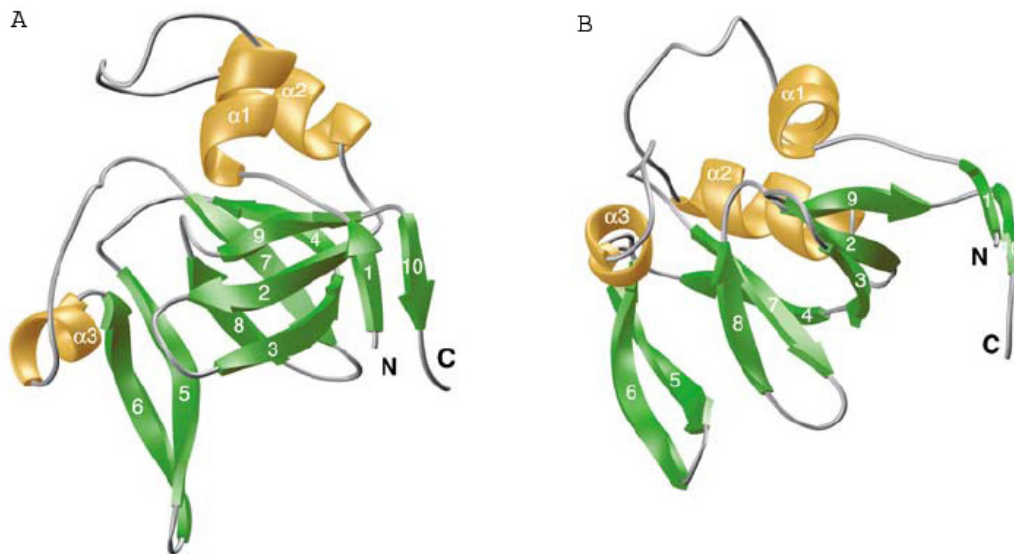


Figure 3: Three-dimensional structure of the *D. melanogaster* Ago1 PAZ domain. The PAZ domain consists of a lefthanded, six-stranded beta-barrel capped at one end by two alpha-helices and wrapped on one side by a distinctive appendage, which comprises a long beta-hairpin and a short alpha-helix. (A) sideview (B) Topview (Yan et al., 2003)

The evolutionarily conserved PAZ (Piwi/Argonaute/Zwille) domain was found in the Dicer and Argonaute protein families that are essential components of the RNA-mediated gene-silencing pathways (Williams and Rubin, 2002; Knight and Bass, 2001). The three-dimensional structure of PAZ consists of a left-handed, six-stranded beta-barrel capped at one end by two alpha-helices and wrapped on one side by a distinctive appendage comprised a long beta-hairpin and a short alpha-helix (Figure 3). This six-stranded beta-barrel core consists of amino acid residues conserved within the PAZ family, and binds to 5' to 3' orientated single-stranded RNA (Yan et al., 2003).

Though the exact biological function of PAZ domain remains unknown, NMR analysis, mutagenesis and RNA mobility shift assay data together define this module as an RNA binding domain. The PAZ domain might be involved in the binding of 5' phosphorylated short regulatory RNA, whose 5' phosphorylation is required for incorporation into the RNA-induced silencing complex (RISC) in RNA interference (Nykanen et al., 2001; Martinez et al., 2002; Yan et al., 2003).

1.1.4. OB folds

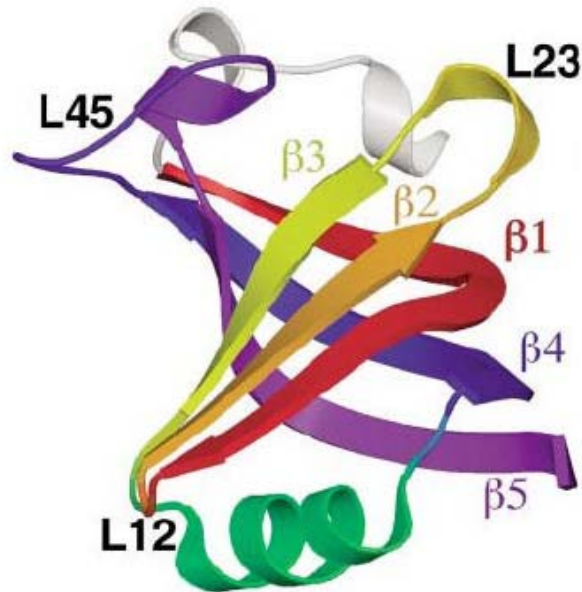


Figure 4: The canonical OB-fold domain. The OB-fold from AspRS is shown in stereo as representative of the ideal OB-fold domain. From the N terminus to the C terminus, strand $\beta 1$ is shown in red, $\beta 2$ in orange, $\beta 3$ in yellow, the helix between $\beta 3$ and $\beta 4$ in green, $\beta 4$ in blue, and $\beta 5$ in violet. An α -helix, which is found in half of the OB-folds in these complexes, is shown in white at the top of the figure, just N-terminal to strand $\beta 1$. Variable loops between strands are indicated in black text (Theobald et al., 2003).

The OB (oligonucleotide/oligosaccharide-binding) fold is found in all three kingdoms and is well characterized in both sequence and structural databases. The OB-fold is a small structure motif that ranges between 70 to 150 amino acid in length. Interestingly, although there is no strong sequence relationship between disparate OB-fold family members, this particular structure can be easily recognized based on its distinct topology (Theobald et al., 2003). As shown in Figure 4, the OB-fold is a five-stranded mixed beta-barrel which was often described as a Greek key motif, and one end of the barrel is capped by an alpha-helix (Murzin, 1993). The OB-fold structure seems to support a binding interface that is easily adapted to the binding of a range of different biological molecules. This interface has at its centre beta strands 2 and 3, and is enclosed at the bottom left by loop 12, at the top by loop 45 and at the right by loop 23. In different structures, loops 23 and 45 show wide variation in both length and sequence (Figure 4). Different OB-fold proteins also use this interface to bind to RNA, single-stranded DNA, oligosaccharides and proteins (Arcus, 2002). Therefore, the proteins containing this motif are involved in almost every aspect of DNA and RNA processing.

1.2. Splicing

For the expression of most protein-coding genes in eukaryotes RNA splicing is an essential step. In eukaryotes the nascent RNA transcript or so-called precursor RNA,

undergoes post-transcriptional modifications, such as 5' capping, 3' cleavage/polyadenylation, and RNA splicing to produce mature RNA for translation. Therefore, the task for the RNA splicing machinery is to recognize exons in the precursor RNA sequences and to accurately remove introns. Like the other steps in gene expression, RNA splicing can be regulated. It is an important mechanism in controlling the expression of many genes. Especially after the release of human genome sequences (Lander et al., 2001, Venter et al., 2001), alternative splicing is recognized as one of the fundamental mechanisms to increase protein diversity from the limited number of human genes.

1.2.1. Splicing passway and spliceosome assembly

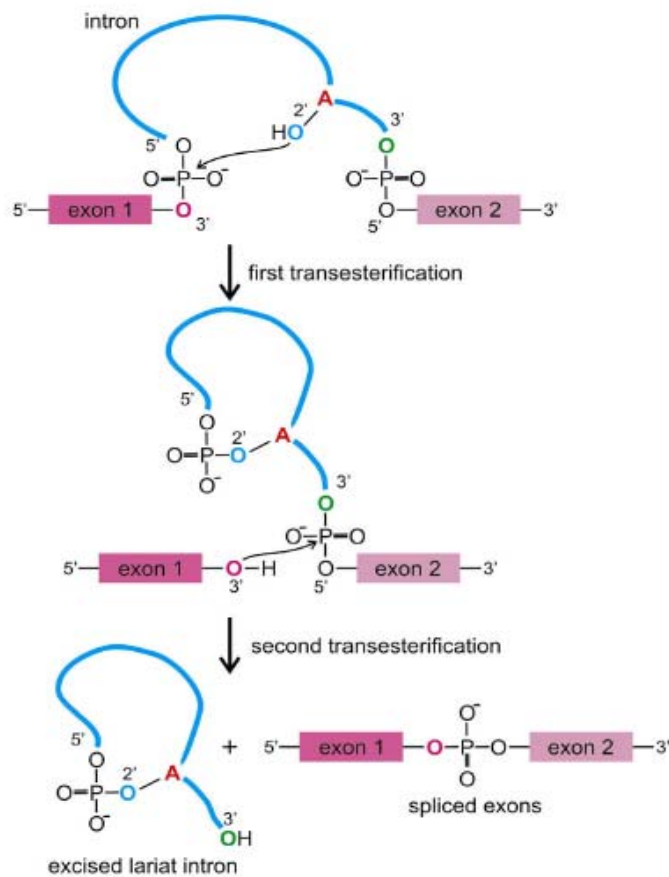


Figure 5: Pre-mRNA splicing pathway. Removal of the intron (blue line) between its flanking exons (purple boxes) proceeds via two transesterification reactions. The 3' oxygens of exon 1 and intron are indicated in purple and green, respectively. The 2' oxygen of branch point adenosine (in red) is shown in blue (Figure adapted from Lodish *et al.*, 2000).

The excision of the introns and the joining of the exons from a pre-mRNA are directed by special sequences at the intron/exon junction termed splice site. The 5' splice site includes a GU dinucleotide at the 5' end of the intron encompassed within a larger,

less conserved consensus sequence (Burge et al., 1999). The 3' splice site region has three conserved sequence elements, the branch point, followed by a polypyrimidine tract, followed by a terminal AG at the extreme 3' end of the intron.

Introns in the nuclear pre-mRNAs are removed by a two-step splicing pathway (Figure 5). In the first step, the 2' hydroxyl group of the branch site adenosine attacks the 5' phosphate of the intron, and the phosphodiester bond between the 5' phosphate of the intron and the 3' oxygen of the first exon (purple O, Figure 5) is broken. This generates two splicing intermediates: the first exon intermediate with a free 3' hydroxyl group and the lariat intermediate containing the second exon and the intron. The lariat structure is the result of forming an unusual 2'-5' phosphodiester bond between the 5' phosphate of the intron and the 2' oxygen (blue O, Figure 5) of the adenosine at the branch site. In the second step, the nucleophilic attack of the free 3' hydroxyl group of the first exon on the 5' phosphate of the second exon cleaves the 3'-5' phosphodiester bond at the 3' splice site. This yields two spliced products: the spliced mRNA with two ligated exons and the excised intron with a 3' hydroxyl group in a lariat structure. As described above, pre-mRNA splicing pathway proceeds through a two step transesterification reactions which lead to exon ligation and release of intron lariat. Since the number of phosphodiester bonds in either reaction is not changed, no energy is consumed.

Splicing is carried out by spliceosome. Generally each spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) and more than 100 accessory proteins (Zhou *et al.*, 2002). Each snRNP is a tight complex composed of several proteins and a short RNA molecule. snRNPs U1, U2, U4, U5 and U6 assemble in major class U2 type spliceosome, and snRNPs U11, U12, U4, U5 and U6 are in minor class U12 type spliceosome.

The spliceosome is a highly dynamic and complex machine. Numerous and ordered interactions between RNA-RNA, protein-RNA, and protein-protein contribute to the assembly of active spliceosomes (Figure 6; Moore et al., 1993; Burge et al., 1999; Collin and Guthrie, 2000; Brow, 2002). Spliceosome assembly is initiated by the ATP-independent formation of the early (E) complex in mammals, in which the 5' splice site and the 3' splice site are initially recognized by the U1 snRNP and the splicing factor U2 snRNP auxiliary factor (U2AF), respectively. In the E complex, the 5' end of U1 snRNA base-pairs with the 5' splice site at positions +1 to +6 of the intron. In the E complex, the 65 kDa subunit of U2 snRNP auxiliary factor (U2AF65) binds to the polypyrimidine tract upstream of the 3' splice site, and the 35 kDa subunit of U2AF (U2AF35) binds to the

essential AG dinucleotide at the 3' splice site. Subsequent to E complex formation, the pre-spliceosome A complex is assembled by the ATP-dependent and stable interaction of the U2 snRNP with the branch site. The branch site sequence is first recognized by SF1 protein. The KH domain of SF1 specifically binds to the branch point adenosine residue and its flanking nucleotides. Base-pairing between the U2 snRNA and the branch site sequence forms a duplex with a bulged adenosine, facilitated by cooperative binding of U2AF65 and SF1 proteins. Entering of the pre-assembled U4/U5/U6 tri-snRNP complex into the spliceosome cycle contributes to the conversion of pre-spliceosome A complex to mature spliceosome B complex. In the tri-snRNP, U4 and U6 snRNAs extensively base-pair with each other. The U5 snRNA is not paired with any other snRNAs, but is in close proximity to the U4/U6 snRNP. The tri-snRNP associates with the pre-spliceosome in an ATP-dependent manner. Two important RNA rearrangements take place at this stage. The U1 snRNA dissociates from the 5' splice site, and instead, the U6 snRNA binds to the 5' splice site. In addition, the U6 snRNA dissociates from the U4 snRNA and forms a new duplex with the U2 snRNA. The U5 snRNA base-pairs with poorly conserved exon sequences at the 5' and 3' splice sites. The two-step splicing reaction takes place in the active spliceosome, the C complex. There is also a minor class of spliceosome (U12 type) that excises a small family of introns that use different consensus sequences. Only the major class (U2 type) of spliceosome is discussed here.

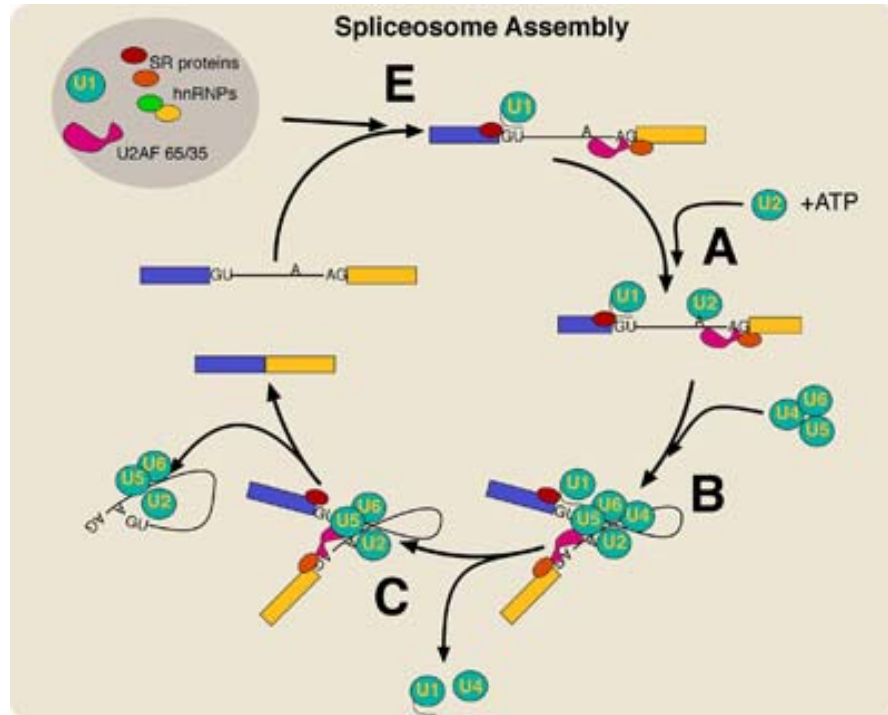


Figure 6: Spliceosome assembly. The spliceosome contains five small nuclear ribonucleoproteins that assemble onto intron. The early (E) complex contains the U1 snRNP bound to the 5' splice site. Each element of the 3' splice site is bound by a specific protein, the branch point by SF1, the polypyrimidine tract by U2AF65, and the AG dinucleotide by U2AF35. The A complex forms when U2 snRNP engages the branch point via RNA/RNA base-pairing. This complex is joined by U4/U5/U6 tri-snRNP to form the B complex. The B complex is then extensively rearranged to form the C complex that catalyzes the two chemical steps of splicing. During this rearrangement the interactions of the U1 and U4 snRNPs are lost and the U6 snRNP is brought into contact with the 5' splice site. The spliceosome contains many additional proteins, such as SR proteins and hnRNPs. After splicing, the spliceosome dissociates, and is re-assembled to take part in a new round of splicing cycle (Figure adapted from: <http://npd.hgu.mrc.ac.uk/compartments/speckles.html>).

1.2.2. Exon recognition and intron bridging

The splice site consensus sequences are generally not sufficient information to determine whether a site will assemble a spliceosome and function in splicing. Other information and interaction are necessary to activate their use. An average human gene contains a mean of 8.8 exons, with a mean size of 145 nucleotides (nt), the mean intron length is 3365 nt, and the 5' and 3' UTR are 770 and 300 nt, respectively. After pre-mRNA processing, the average mRNA exported into the cytosol consists of 1340 nt coding sequence, 1070 nt untranslated regions and a poly (A) tail (Lander et al., 2001). This shows that more than 90% of the pre-mRNA is removed as introns and only about 10% of the average pre-mRNA is joined as exonic sequences by pre-mRNA splicing. Human cells are not only capable of accurately recognizing the small exons within the larger intron context, but are also able to recognize exons alternatively. The specificity can be generated by the presence of additional non-splice site regulatory elements within the exon or introns. RNA elements that act positively to stimulate spliceosome assembly

are called splicing enhancers. Exonic splicing enhancers are commonly found even in constitutive exons. Other exonic or intronic RNA sequences act as splicing silencers or repressors to block spliceosome assembly and certain splicing choices. These elements are characterized by the loose consensus sequences and this degeneracy prevents from interfering with the coding capacity of the exons.

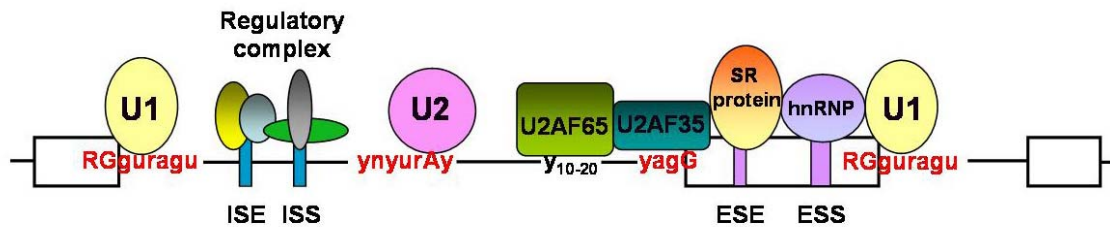


Figure 7: Classical and auxiliary splicing elements and splice factors. Exons are indicated as boxes, the intron as a thick line. Splicing regulator elements (enhancers or silencers) are shown as purple boxes in exons or as blue boxes in introns. 5' Splice site (RGguragu), 3' splice site $(y)_{10}ncagG$ and branch point (ynyurAy) are indicated ($y=c$ or u , $n=a$, t , c or g). Two major groups of proteins, hnRNPs and SR or SR-related proteins bind to splicing regulator elements.

Many of these elements are bound by known RNA-binding proteins (RBPs), such as the SR proteins and members of the hnRNP group of proteins. Splicing regulation is sometimes associated with both enhancers and silencers; which splice sites are recognized is often the result of antagonistic effects generated by bound regulators. Moreover, many elements are not strict silencers or enhancers, rather the position of an element relative to an alternative exon can determine whether it acts positively or negatively (Hui et al., 2005).

The SR proteins constitute the best-studied family of non-snRNP proteins required for pre-mRNA splicing. SR proteins have a characteristic structural organization, which consists of one or two N-terminal RNA recognition motifs (RRMs or RBDs for RNA binding domains), that function in sequence-specific RNA binding, and a variable-length C-terminal arginine/serine-rich (RS) domain required for protein-protein interaction with other RS domains. The serines in an RS domain can be highly phosphorylated. The SR protein family members include so far ASF/SF2, SC35, SRp20, SRp30c, 9G8, SRp40, SRp55, and SRp70 (Graveley, 2000). A number of additional splicing factors containing RS domains are structurally and functionally related to SR proteins and are collectively referred to as SR-related proteins or SRrps. Although SR proteins or SR-related proteins do recognize specific RNA sequences, their consensus sequences are rather degenerate. As the interaction is weak and not highly specific between individual splicing factor and

the the regulatory sequence, different SR and SR-related proteins can act through the same regulatory elements and influence the same splice sites. Higher specificity is also achieved by protein–protein interactions that allow simultaneous binding of multiple proteins to RNA. It is well known that SR or SR-related proteins can promote and stabilize the formation of E complexes containing U1 snRNP bound to the 5' splice site and U2 snRNP bound to the pre-mRNA branch site. They can also facilitate the recruitment of U4/U6 and U5 snRNPs. In addition, SR and SR-like proteins can bridge the introns by interacting with themselves and the core spliceosomal components.

In antagonizing the positive effects of SR proteins, hnRNPs can influence splice site selection by binding to splicing silencers or repressors. The hnRNP proteins are among the most abundant proteins in the nucleus and are identified by their association with unspliced mRNA precursors (hnRNAs). The nascent hnRNAs are immediately bound by hnRNPs, resulting in the formation of hnRNP complexes. All hnRNPs share a common structure containing RNA binding domains and auxiliary domains, which are composed of clusters of certain amino acids, and might mediate protein-protein interaction or facilitate protein localization (Krecic and Swanson, 1999). Two mechanisms of splicing repression have been proposed. First, binding of these regulatory factors to the silencer sequences could interfere directly with the assembly of spliceosomal components. It could block the exon bridging interactions that occur during exon recognition, or it could block splicing activation by SR proteins binding to adjacent ESEs. Second, dimerization of inhibitory proteins surrounding the exon causes the alternative exon to loop out and to be skipped by the splicing machinery. Interestingly, hnRNP H could act as a splicing repressor when bound to an ESS in beta-tropomyosin, but as an activator when bound to a similar element in HIV Tat exon 2 (Chen et al., 1999; Caputi and Zahler, 2002).

In summary exon recognition and splice site selection is a combinatorial process involving a complex interplay between positive and negative regulators that function through cognate enhancers and silencers (Figure 7).

1.3. Alternative splicing

Most exons are constitutively spliced and included in the mRNA. In contrast to constitutive splicing, alternative splicing is the process to produce distinct mRNAs from a single pre-mRNA, by using various combinations of the 5' and 3' splice sites of different exons, a process which is called alternative splicing. Alternative splicing is an important

mechanism for increasing protein diversity during post-transcriptional processing.

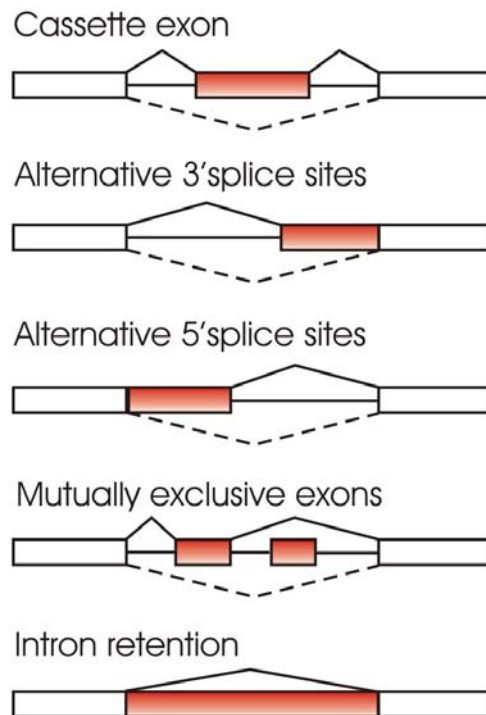


Figure 8: Types of alternative splicing. In all five examples of alternative splicing, flanking constitutive exons are indicated as white boxes, and alternative spliced regions in red, introns are presented by solid lines, and dashed lines indicate splicing activities

It was unexpected that the human genome contains only about 20,000 to 25,000 genes (International Human Genome Sequencing Consortium, 2004), and the genome of *Drosophila melanogaster* (14,000) contains fewer genes than the simpler organism *Caenorhabditis elegans* (19,000). The apparent discrepancy between gene number and complexity of the organism implicated that the number of functional proteins was much higher than the number of genes, as the number of human proteins is estimated to be more than 90,000 (Woodley and Valcarcel 2002). Bioinformatics analysis indicates that 74 % of human genes are involved in alternative splicing (Modrek et al., 2001), which contributes significantly to human proteome complexity. Given that the human expressed sequence tags (ESTs) cover only a portion of all possible transcripts; this number is likely to be an underestimate. On average, a human gene generates two to three transcripts. However, extreme cases exist: The human neurexin3 gene can potentially form 1728 transcripts due to alternative splicing at four different sites. In *Drosophila*, the Down syndrome cell adhesion molecule (DSCAM), which is involved in the specificity of neuronal connectivity can potentially generate 38016 different protein isoforms due to the combinatorial use of alternative exons (Celotto and Graveley, 2001). The significance of

alternative splicing is based not only on increasing protein diversity, but also in modulating the level of alternatively spliced isoforms in a tissue- or developmental stage-specific manner.

There are five basic forms of alternative splicing. As shown in Figure 8, exons can be either skipped or included, either extended or shortened, or included in a mutually exclusive manner, and introns can be either removed or retained. Cassette exon (or exon skipping) accounts for 38% of the alternative splicing events conserved between human and mouse genomes. Alternatively 3' splice sites and 5' splice sites account for 18% and 8% of the conserved events, respectively. Intron retention is responsible for less than 3% of the alternative splice events conserved between human and mouse genomes. There are more complex events, like mutually exclusive events, alternative transcription start sites, and multiple polyadenylation sites account for the remaining 33% of the alternatively spliced events (Ast G. 2004). An estimated 75% of all alternative splicing patterns change the coding sequence (Zavolan et al., 2003), indicating that alternative splicing is a major mechanism enhancing protein diversity.

1.3.1. Regulation of alternative splicing

As discussed above, alternative splicing of pre-mRNA is a major source of protein diversity and the regulation of this process is crucial for diverse cellular functions in both physiological and pathological situations. Alternative splicing is often regulated in a tissue- or developmental stage-specific manner. In some cases, the presence or absence of an individual mammalian tissue-specific regulator is sufficient to determine splicing patterns. For example, the brain-specific RNA binding Nova proteins are involved in determining a number of neuronal synapse-specific alternative splicing events through binding to intronic or exonic regulatory elements that contain UCAAY motifs (Ule et al., 2003). Those Nova target genes are highly related in function. They were associated with the function of inhibitory synapses, post-synaptic and pre-synaptic structures, as well as signaling and protein synthesis, suggesting that a single splicing factor regulates isoform expression of different genes in inhibitory neurons.

However, most tissue-specific alternative splicing events seem to be regulated by a more complex network of synergistic and antagonistic influences between groups of regulators. The splicing patterns can be also altered by extracellular stimuli such as hormones, immune response, neuronal depolarization, and cellular stress. In most cases, these changes are reversible, indicating that they are part of a normal physiological response (Stamm, 2002). In several cases, the mechanism leading to changes in alternative

splicing involves changes in the phosphorylation of splicing factors, which influence their ability to bind to RNA or other splicing factors. Since splicing factors appear to regulate coherent biological functions (Ule and Darnell, 2003), changing their activity will most likely result in a coordinated response to the stimulus that triggered the initial phosphorylation signal. Using this mechanism, the mRNA expression of different, seemingly unrelated genes can be coordinated.

1.3.1.1. Phosphorylation-dependent control of the splicing machinery

It is known that the phosphorylation status of snRNPs and spliceosome-associated proteins (SAPs) affects the assembly of the spliceosome (Mermound et al., 1994; Tazi et al., 1992; Tazi et al., 1994). Phosphorylation has also been implicated in the regulation of splice site selection, as protein phosphorylation has also been shown to modulate the alternative splicing of a number of exons both *in vivo* (Hartmann et al., 2001) and *in vitro* (Prasad et al., 1999). Several splicing factors, such as serine/arginine-rich (SR) proteins are known to be phosphorylated *in vivo* (Graveley 2000). Changes in the phosphorylation state of the splicing factors can influence their ability to interact with RNA (Chen et al., 2001) and with other proteins (Hartmann et al., 1999). For instance, the SR proteins, a family of splicing factors that are highly conserved in Metazoa, contain RS domain in which the serine residues can be highly phosphorylated and subsequently affect its protein and RNA binding ability. The phosphorylated SR protein SF2/ASF shows a decreased binding ability to other SR protein, and the non-phosphorylated RS domain of SF2/ASF can interact with non-specific RNA, whereas phosphorylation prevents such interaction (Xiao and Manley 1997). Moreover changes in the phosphorylation level of SR protein, leading to hyper or hypophosphorylation, have been shown to inhibit splicing (Prasad et al., 1999).

Phosphorylation of SR proteins can influence their subcellular localization as well (Misteli, 1998). Protein phosphorylation controls the release of the SR-proteins from the storage compartments in the nucleus (Wang et al., 1998) and causes relocalization of hnRNP proteins to the cytoplasm (van der Houven van Oordt et al., 2000). In this way phosphorylation can alter the active concentration of the splice factors. Since the splicing factors can sometimes regulate coherent biological functions, it is likely that changes in their activity can influence the expression of different seemingly unrelated genes.

1.3.1.2. Signaling to the splicing machinery

Though the transcripts of most metazoan protein-coding genes are alternatively spliced, the mechanisms that regulate splicing, in particular the associated signal-transduction pathways are poorly understood. However, several studies in recent years have emerged that link cell signalling and splicing control.

In living cells, the splicing pattern of many primary transcripts can be altered in response to numerous stimuli, such as growth factors, cytokines, hormones and depolarization. The change of cell surface molecule CD44 isoforms is an excellent example of signal-dependent regulation of alternative splicing. CD44 can form multiple isoforms due to alternative splicing of internal cassette exons. Most tissues express the smallest and most common isoform, CD44s, however in primary T-cells, the alternative splicing process creates another transcript that contains the variant exon-5 (v5). The activation of T-cells receptor switches on Ras signalling pathway which involves several kinase cascades that ultimately deliver signals to the cell nucleus. Although the Ras signalling pathway is complicated, key kinases involved include the serine/threonine kinase Raf which interacts directly with Ras. Raf, in turn, activates the mitogen-activated protein kinase (MAPK) cascade, which leads to the activation and nuclear translocation of the extracellular signal-regulated kinase (ERK). Activated ERK, then translocates into the cell nucleus and phosphorylates splicing factors that induce the v5 inclusion. (Chang, L. and Karin, M., 2001; Weg-Remers et al., 2001; König et al., 1998) Which factors influence exon v5 splicing phosphorylated by ERK remain unclear. However, *in vitro* and *in vivo* evidences show that the nuclear RNA-binding protein SAM68 (Src-associated in mitosis 68 kD), which participates in various distinct cellular processes including splicing (Taylor et al., 2004), activates exon v5 inclusion in response to phosphorylation by ERK (Matter et al., 2002). The regulation of CD44 exon v5 which does not require protein synthesis, demonstrates that signal-induced phosphorylation of regulatory splice factors is the regulatory principle.

1.3.2. Function of alternative splicing

Gene regulation through alternative splicing is more versatile than regulation through promoter activity. Variant transcripts generated through alternative splicing are often tissue and/or developmental specific, resulting in effects seen only in certain cells or developmental stages. Changes in alternative splicing can modulate transcript expression levels by subjecting mRNAs to nonsense-mediated decay (NMD) and alter the structure of the gene product by inserting, or deleting, novel protein parts. The structural changes fall into three categories: introduction of stop codons, changes of the protein structure and

changes in the 5' or 3' untranslated region. The effects caused by alternative splicing range from a complete loss of function to subtle effects that are difficult to detect.

1.3.2.1. Introduction of stop codons

mRNAs that contain premature stop codons (PCTs) can be degraded by nonsense-mediated decay (NMD). NMD in mammalian cells generally occurs when a nonsense codon resides more than 50–55 nucleotides upstream of a splicing-generated exon–exon junction. mRNAs from intronless genes are immune to NMD (Maquat and Li 2001; Brocke et al., 2002), therefore the NMD is dependent on pre-mRNA splicing. This dependence of NMD reflects the need for an exon junction complex (EJC) of proteins deposited 20–24 nucleotides downstream of a nonsense codon (Figure 9). The EJC consists not only of the Upf NMD factors but also of proteins involved in pre-mRNA splicing, such as RNPS1, UAP56, SRm160 and Pnn/DRS (Maquat, 2004).

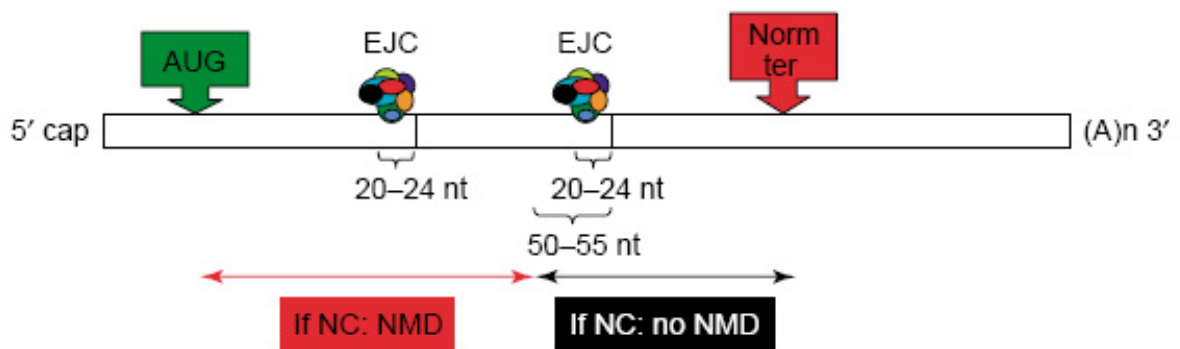


Figure 9: Mechanisms by which cells distinguish between nonsense codons that elicit NMD and nonsense codons that do not. Nonsense codons (NCs) in mammalian cells are recognized after splicing and generally elicit NMD provided there is at least one splicing-generated exon–exon junction more than 50–55 nucleotides downstream. Dependence on a junction actually reflects dependence on an exon junction complex (EJC) of proteins that is deposited about 20–24 nucleotides upstream of junctions. The EJC consists of at least 12 proteins, including the NMD factors Upf3 or Upf3X and Upf2. Data indicate that Upf1 also comprises the EJC but only transiently and in a way that depends on Upf3 or Upf3X and Upf2. Norm ter: normal termination codon. (Figure adapted from Lejeune and Maquat 2005)

Alternative pre-mRNA splicing appears to be a main source of nonsense codons that elicit NMD. About 25–35% of alternative exons introduce frameshifts or stop codons into the pre-mRNA (Stamm et al., 2000; Lewis et al., 2003). Since approximately 75% of these exons are predicted to be subject to nonsense-mediated decay, more than 30% of mRNAs that derive from alternative splicing have been estimated to contain a nonsense codon that elicits NMD (Lewis et al., 2003). NMD is present in all eukaryotic organisms; however seems to be most prevalent in mammals in using alternative splicing as a process by which many nonsense codons are generated as well as a process that is required for

their elimination. This process, which has been termed RUST, for regulated unproductive splicing and translation, currently represents the function of alternative splicing with the most obvious biological consequences.

1.3.2.2. Changes of the protein structure

Approximately 75% of alternative splicing events occur in the translated regions of mRNAs and will affect the protein-coding region (Okazaki et al., 2002; Zavolan et al., 2003). Changes in the protein primary structure can alter the binding properties of proteins, influence their intracellular localization and modify their enzymatic activity and by diverse mechanisms.

Alternative splicing can generate protein isoforms that differ in their binding properties by deleting binding domains or inserting peptide sequences. As an example, latency-associated peptide-binding protein binds TGF-beta in an isoform-dependent manner (Koli et al., 2001) Alternative splicing can further determine the ligand specificity of a receptor. For instance, the fibroblast growth factor receptor gene FGFR-2 generates two isoforms that differ by 49 amino acids in the extracellular domain. Depending on the presence of this domain, the receptor binds to both fibroblast and keratinocyte growth factor or only to fibroblast growth factor (Miki et al., 1992). Similar to the ligands-protein interaction, protein-protein interaction can be regulated by alternative splicing through deletion or insertion of binding domains or by controlling the number of multiple binding motifs to modulate protein binding affinities. Alternative splicing can also modify the DNA-transcription factors interaction by abolishing the DNA binding domain of transcription factors, which contributes to transcriptional regulation.

Alternative splicing can influence the intracellular localization of numerous proteins, usually by regulating the interaction of proteins with membranes or influencing localization signals. Alternative splicing can generate non-membrane-bound protein isoforms by deleting or interrupting transmembrane or membrane-association domains. These soluble isoforms can be released from the cell or translocate into a different intercellular compartment. They can lose the ability of signal transduction (Tone et al., 2001), can be less stable (Garrison et al., 2001) or have a different effect on immune system modulation (Riteau et al., 2001). Alternative splicing can also regulate the localization of proteins in various subcellular sites and organelles. Proteins can be sequestered into compartments, where they perform no function. This mechanism is widely used for receptor molecules and alternative splicing can regulate their retention in membrane-enclosed compartments. For example, one splice variant of the dopamine D2

receptor is retained more efficiently in the endoplasmic reticulum than the other (Prou et al., 2001), which influences the overall dopamine D2 activity. Similarly, sublocalization of proteins within organelle can be regulated by alternative splicing. In the nucleus, proteins can be present in different nuclear substructures, such as in the nucleoplasm and speckles, due to alternative splicing (Nishizawa et al., 2001).

1.3.2.3. Changes in the untranslated region

Some examples have demonstrated that alternative splicing functions by changing the properties of the mRNA. Alternative splicing events occurring in 5' and 3' UTRs may influence the stability of the RNA. For example, alternative exons in the 5' UTR of the HIV-1 virus can either promote or inhibit the nuclear degradation of their mRNA, which regulates HIV-1 gene expression (Krummheuer et al., 2001). The splicing factor SC35 autoregulates by activating splicing events within the 3'UTR of SC35 pre-mRNA (Sureau et al., 2001).

1.4. Emery-Dreifuss muscular dystrophy

Emery-Dreifuss muscular dystrophy (EDMD) is a neuromuscular degenerative condition with an associated dilated cardiomyopathy and cardiac conduction defect (Ellis 2006). EDMD is associated with a substantial risk of sudden cardiac death (Sylvius and Tesson, 2006). It can be inherited in either an X-linked or autosomal manner. The mutation or loss of emerin, an integral protein of inner nuclear membrane, causes the X-linked EDMD (Bione et al., 1995). Autosomally inherited EDMD is caused by mutation in lamin A/C, the intermediate filament protein associated with inner nuclear membrane (Bonne et al., 1999). Emerin and lamins colocalize at the nuclear rim (Manilal et al., 1996), and later lamin A/C was shown to co-immunoprecipitate as part of a novel nuclear envelope protein complex that also included the proteins emerin, lamin B and F-actin (Fairley et al., 1999). Subsequently, a direct interaction between lamin A and recombinant emerin has been demonstrated (Clements et al., 2000; Lee et al., 2001). A homologous domain, LEM (LAP-Emerin-MAN1) at the emerin N-terminus (amino acids 6–44) is shared by other nuclear membrane proteins, such as LAP2b (Furukawa, 1999) and MAN1 (Lin et al., 2000). The LEM domain interacts with BAF (barrier-to-autointegration factor), a DNA bridging protein. BAF recruits LEM domain containing proteins to chromatin during nuclear assembly and thus provides a physical link between the nuclear envelope and chromatin (Lee et al., 2001). Yeast two-hybrid studies have shown that the tail region of lamins A/C is responsible for binding emerin (Sakaki et al., 2001). The lamin A/C knockout mouse displays muscular dystrophy (Sullivan et al., 1999), suggesting that

EDMD is caused by functional defects in the lamin A/C–emerin complex. It seems that emerin has a long list of binding partners, which include the transcriptional repressors GCL (germ cell less) and Btf (death promoting transcriptional repressor). Therefore the yeast two-hybrid method was used to screen a human heart cDNA library for novel emerin binding partners and YT521-B, an RNA alternative splicing factor was identified as an interacting protein. This discovery increases the functional range of proteins that interact with emerin. Small disruptions of the normal balance of gene expression could lead to the slow degeneration of cardiac and skeletal muscle tissues associated with EDMD.

1.5. pre-mRNA secondary structure influence exon recognition

RNA molecules can adopt an essentially limitless number of structures in solution by base pairings and hydrophobic interactions. Most proteins regulating splice site selection recognize single-stranded, not base-paired RNA. For example, KH (hnRNP K homology) domains, and RRM (RNA recognition domains) bind to 2 to 10 nucleotides of single-stranded RNA (Auweter et al., 2006). Frequently, the RNA part that binds to a protein is in a hairpin loop (Buckanovich and Darnell, 1997; Skrisovska et al., 2007). The large majority of human genes is alternatively spliced, which was supported by EST-based database analysis indicating that 35–60% of all human gene products are alternatively spliced (Modrek and Lee, 2002). In higher organisms, the splice sites do not contain all the information that is required for accurate intron recognition (Lim and Burge, 2001). Additional regulatory motifs, such as exonic or intronic enhancer and silencer sequences are essential for the alternative and constitutive splicing (Blencowe, 2000). Moreover, there is emerging evidence that pre-mRNA secondary structure plays a role in alternative splicing. For example, a deletion in the mouse fibronectin EDA exon leads to a shift of a critical ESE from single- into doublestranded conformation, which causes exon skipping (Buratti et al., 2004). The skipping of exon 7 of the *SMN2* gene is correlated with the stability of a stem structure that sequesters the donor splice site (Singh et al., 2007). The splicing of mutually exclusive exons in the rat *FGFR2* and *Drosophila DSCAM* gene is regulated by conserved secondary structures (Graveley, 2005; Muh et al., 2002).

To understand whether the structural context of exonic or intronic regulatory motifs has a general importance, a large set of experimentally verified splicing enhancer and silencer sequences with their natural sequence context was compiled and analyzed. The result shows that the splicing motifs are located in a structural context that favors their sequence to be single-stranded. The computational analysis was confirmed by

experimental results using minigenes to compare the effect of known splicing motifs in single- and double-stranded conformation. These data suggest that pre-mRNA secondary structures are an integral part of the splice site recognition (Buratti, E. and Baralle, F. E., 2004).

2. RESEARCH OVERVIEW

As described in the section “exon recognition and intron bridging” (section 1.2.2), most of the trans-acting factors involved in splicing include at least one RNA binding domain in their structures. Previous studies showed that YT521-B was able to change alternative splice site usage, but does not belong to any of known RNA binding protein families.

First, in the result section 4.1 to 4.6, both computational analysis and experimental data demonstrate that the YTH domain is a novel RNA binding domain. These sections also include the identification of the degenerated and single-strand RNA binding motif of YTH domain. The results also show that the presence of the binding motif in alternative exons is necessary for YT521-B to directly influence splice site selection *in vivo*.

Second, as the mechanisms regulating the alternative splice site selection, especially the associated signal-transduction pathways are still not very well understood, section 4.7 to 4.13 were focused on the determination of the effect of tyrosine phosphorylation on vertebrate-specific splicing factor YT521-B. The results demonstrate that YT521-B can interact and subsequently phosphorylated on tyrosine residues by several non-receptor tyrosine kinases. The tyrosine phosphorylation can influence on both subnuclear localization and biological activity of YT521-B.

Finally, a small section 4.14 was dedicated to the study of the influence of pre-mRNA secondary structures on exon recognition.

Together these data show that YT521-B is a splicing factor with a new binding domain that is regulated by tyrosine phosphorylation emanating from Src family kinases.

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Chemicals

| Product | Supplier | Product | Supplier |
|---|-------------------|--|------------|
| BAC-to-BAC baculovirus expression system | Invitrogen | 2-Propanol | Roth |
| Lysozyme | Sigma | di-Sodiumhydrogen phosphate | Merck |
| Agarose ultra pure | Invitrogen | dNTPs | Invitrogen |
| Chloroform: Isoamyl alcohol | Sigma | Magnesium chloride (MgCl ₂ x 4H ₂ O) | Merck |
| Glycine | Roth | Magnesium sulfate | Sigma |
| Benzonase | Sigma | Methanol | Roth |
| Agar | GibcoBRL | Microcystin | Axxora |
| Acetic acid | Roth | N,N,N',N'-tetramethylethylenediamine (TEMED) | Sigma |
| Ammonium persulfate (APS) | Sigma | Ni-NTA Agarose | Qiagen |
| Ampicilin | Sigma | Nonidet P-40 (NP-40) | Sigma |
| Glycogen | Roche | Paraformaldehyde | Merck |
| Aprotinin | Sigma | PEG 3500 | Sigma |
| [α - ³² P]-CTP (800 Ci/mmol, 10 μ Ci/ μ l) | Hartmann Analytic | Perhydrol 30% H ₂ O ₂ | Merck |
| Boric acid | Roth | Phenol: Chloroform: Isoamyl alcohol | Sigma |
| Bovine serum albumin (BSA) | Sigma | Phenylmethylsulfonyl fluoride (PMSF) | Sigma |
| Bradford reagent (BioRad Protein Assay) | BioRad | Plasmid maxi kit | Qiagen |
| Brilliant Blue R 250 | Sigma | Potassium chloride (KCl) | Merck |
| Bromophenol blue | Merck | Protease Inhibitor Cocktail | Sigma |
| Calcium chloride dehydrate (CaCl ₂ 2H ₂ O) | Merck | Protein A Sepharose | Amersham |
| Cellfectin | Invitrogen | QIAEX II gel extraction kit | Qiagen |
| Deoxycholic acid | Sigma | Qiagen M13 kit | Qiagen |
| N-2-hydroxyethylpiperazine (HEPES) | Sigma | Sepharose CL-4B | Pharmacia |

| Product | Supplier | Product | Supplier |
|---|-------------------|---|-----------------|
| Dextrose | Sigma | Sodium acetate | Merck |
| 30% Acrylamide/Bis solution | Sigma | Sodium chloride (NaCl) | Roth |
| Ethanol | Roth | Sodium dedecyl sulfate (SDS) | Sigma |
| Ethidium bromide | Sigma | Sodium dihydrogen phosphate (NaH ₂ PO ₄) | Merck |
| Gelatin | Sigma | Sodium fluoride | Sigma |
| Dithiothreitol (DTT) | Merck | Sodium hydroxide | Merck |
| Ethylenediaminetetraacetic acid (EDTA) | Merck | Sodium orthovanadate | Sigma |
| Ficoll 400 | Fluka | Sodium pyrophosphate | Merck |
| Glutathione –Sepharose 4B | Amersham | TNT Reticulocyte kit | Promega |
| Glycerol | Sigma | TOPO TA cloning kit | invitrogen |
| β-Mercaptoethanol | Merck | Trichloroacetic acid (TCA) | Sigma |
| Glycerol 2-phosphate | Sigma | Tris-hydroxymethylaminomethane | Sigma |
| HiperFect | Qiagen | Triton X-100 | Sigma |
| Imidazole | Roth | tRNA from yeast | Sigma |
| IPTG | Sigma | Trypsin-EDTA | Invitrogen |
| Dimethylformamide (DMSO) | Sigma | Tryptone | Sigma |
| Kanamycin | Sigma | Tween 20 | Sigma |
| Chloramphenicol | Sigma | Unsupplemented Grace's Medium | GibcoBRL |
| [γ- ³² P]-ATP (3,000 Ci/mmol, 5μCi/μl) | Hartmann Analytic | X-Gal | Sigma |
| Luminol | Sigma | Xylene cyanole FF | Sigma |
| p-Iodophenol | Sigma | Yeast Extract | Sigma |

3.1.2. Enzymes and enzyme inhibitor

| Product | Supplier | Product | Supplier |
|-----------------------------|-----------------|--------------------------|-----------------|
| BamHI | NEB | Proteinase K | Sigma |
| EcoRI | NEB | RNase inhibitor | Roche |
| DpnI | NEB | RQ1 RNase free DNase | Promega |
| NotI | NEB | SuperScript II | Invitrogen |
| SacII | NEB | T4 DNA Ligase | NEB |
| Sall | NEB | T4 polynucleotide kinase | NEB |
| NheI | NEB | T7 DNA Polymerase | NEB |
| Calf intestinal phosphatase | Boehringer | T7 RNA polymerase | Promega |

| Product | Supplier | Product | Supplier |
|-------------------------|------------|--------------------|----------|
| Platinum Pfx polymerase | Invitrogen | Taq DNA polymerase | Peqlab |

3.1.3. Cell lines and media

HEK293 cells and Neuro-2a cells were maintained in DMEM supplemented with 10% Fetal Calf Serum (both from GibcoBRL) in a 37°C, 5 % CO₂ humidified incubator (Heraeus). Each 2 days, cells were trypsinized by 1 x Trypsin-EDTA and reseeded into new dishes. Sf9 cells were cultured in TNM-FH insect medium (BD) at 27°C. When cells become 90% confluent, they were displaced from the flask's surface by tapping the flask sharply against the hand 3 or 4 times and then reseeded into fresh flask.

| Name | Description | Supplier |
|----------|--|-------------------|
| HEK293 | Human embryonic kidney transformed with adenovirus 5 DNA | ATCC No. CRL-1573 |
| Neuro-2a | Mouse neuroblastoma | ATCC No. CCL-131 |
| COS-7 | African green monkey kidney SV40 transformed | ATCC No. CRL-1651 |
| Sf9 | Fall armyworm <i>Spodoptera frugiperda</i> pupal ovarian tissue cell | Invitrogen |

3.1.4. Bacterial stains and media

3.1.4.1. Bacterial stains

| Strain | Genotype | Reference |
|-----------------------|--|------------------------------|
| BL21(DE3)-RIL | <i>ompT hsdS</i> (r _B m _B) <i>dcm</i> ⁺ Tet ^r <i>gal</i> λ(DE3) <i>endA Hte</i> [<i>argU ileY leuW Cam</i> ^r] | (Studier, F.W. et al., 1990) |
| XL1-Blue MRF' | Δ(<i>mcrA</i>)183 Δ(<i>mcrCB-hsdSMR-mrr</i>) 173 <i>endA1 supE44 thi-1 recA1 gyrA96 relA1 lac</i> [F' <i>proAB lacI</i> ^q ZΔM15 Tn10 (Tet ^r)] | (Bullock W.O. et al., 1987) |
| DH10Bac TM | F' <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>)Φ80/ <i>lacZ</i> ΔM15 Δ <i>lacX74 deo RrecA1 endA 1ara D139Δ (ara,leu) 7697 galK λ-rpsLnup G / bMON 14272 / pMON7124</i> | (Hanahan D., 1983) |

3.1.4.2. Media

| Component | LB medium(1L) | LB Agar (1L) |
|---------------|---------------|--------------|
| NaCl | 10g | 10g |
| Tryptone | 10g | 10g |
| Yeast extract | 5g | 5g |
| Agar | 0 | 20g |

3.1.5. Antibiotics

| Antibiotic | Working concentration | |
|-----------------|-----------------------|-------------|
| | Liquid culture | Agar plates |
| Ampicilin | 100µg/ml | 100µg/ml |
| Chloramphenicol | 15µg/ml | 30µg/ml |
| Kanamycin | 20µg/ml | 20µg/ml |
| Gentamycin | 10mg/ml | 7 µg/ml |
| Tetracyclin | 10mg/ml | 10µg/ml |

Ampicilin and kanamycin were stored at 4°C, chloramphenicol and tetracyclin at -20° and Gentamycin was stored at RT.

3.1.6. Antibodies

| Antibody | Dilution | Organism | Source |
|-----------------|----------|----------|---------------|
| anti pTyr(PY20) | 1: 5000 | Mouse | Santa Cruz |
| anti Flag M2 | 1: 1000 | Mouse | Sigma |
| anti-actin | 1:2000 | Mouse | Amersham |
| MANEM5 mAb | 1:1000 | Mouse | Custom made** |
| anti-GFP | 1: 3000 | Mouse | Roche |
| anti-YT521-B | 1:3000 | Rabbit | Custom made* |
| Anti-Abl | 1:2000 | Mouse | Santa Cruz |
| Cy3 | 1:1500 | Mouse | Dianova |
| anti- mouse Ig | 1:10000 | Sheep | Amersham |
| anti- rabbit Ig | 1:10000 | Donkey | Amersham |

*Antiserum was raised against a mixture of two YT521-B peptides: P1 RSARSVILIFSVMRESGKFCQCG and P2 KDGELNVLDDILTEVPEQDDECG (Nayler et al., 2000)

**Antiserum was provided by Dr. Glenn E. Morris

3.1.7. Plasmids

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

| Name | Backbone | Description | Reference |
|---------------|----------|---|---------------------------|
| pEGFP-C2 | pEGFP-C2 | CMV promoter, Kan ^r /Neo ^r , fl ori | Clontech |
| pht6-Fl-FLAG | pcDNA | YT521-B FLAG-tagged | (Nayler O. et al., 1998) |
| c-src wt | pcDNA3.1 | c-Src kinase | (Wong B.R. et al., 1999) |
| pRK5-abl | pRK5 | c-Abl kinase | (Nayler O. et al., 1998) |
| pRK5-fyn | pRK5 | Fyn kinase | (Nayler O. et al., 1998) |
| pRK5-fyn-KA | pRK5 | Catalytic inactive Fyn kinase | (Nayler O. et al., 1998) |
| pUHG10-3(FER) | pUHD10-3 | FerH kinase | (Hao Q.L. et al., 1991) |
| Sik YF pRK5- | pcDNA3 | Constitutively active Sik | (Derry J.J. et al., 2000) |

| Name | Backbone | Description | Reference |
|-------------------|--------------|---|---------------------------|
| fyn | | kinase | |
| pSVL-Syk | pSVL | pSVL Syk kinase | (Zhang J. et al., 1996) |
| CSK | pcDNA3 | CSK kinase | (Nayler O. et al., 1998) |
| AUG1(pcDN A3-Rlk) | pcDNA3 | Rlk kinase | (Debnath J. et al., 1999) |
| pCR3.1 MGTra | pCR3.1TA | Tra2-beta minigene | (Stoilov P. et al., 2004) |
| SV9/10L/11 | Exontrap | Tau minigene | (Gao Q.-S. et al., 2000) |
| WT SMN MG | pCI | SMN2 minigene | (Lorson C. et al., 1999) |
| Pht6-HTa | pFastBac-HTa | Full length YT521-B was cloned into <i>Drosophila</i> expression vector | None |
| pEGFP-HGRG8 | pEGFP-C2 | Homo sapiens YTH domain family, member 2 | None |

3.1.7.2. Newly made clones

| Name | Backbone | Description | Tag |
|----------------|-------------|--|------|
| HTa-YTHdel | pFastBacHTa | YTH domain (from residue 356 to 499) deleted | HIS |
| pEGFP-YTHdel | pEGFP-C2 | YTH domain (from residue 356 to 499) deleted | EGFP |
| pEGFP-DACA1 | pEGFP-C2 | YTH domain protein 1 (Dermatomyositis associated with cancer putative autoantigen 1) | EGFP |
| pEGFP-YTHDF3 | pEGFP-C2 | Homo sapiens YTH domain family, member 3 | EGFP |
| W431DYTH | pEGFP-C2 | YT521-B mutant (W->D) within YTH domain | EGFP |
| R478NYTH | pEGFP-C2 | YT521-B mutant (R->N) within YTH domain | EGFP |
| MG-YT1 | DUP33 | One YTH binding motif | None |
| MG-YT2 | DUP33 | Two YTH binding motif | None |
| MG-YT3 | DUP33 | Three YTH binding motif | None |
| SXN-ESEm-loop1 | DUP33 | CD44 (enhancer) | None |
| SXN-ESEm-stem1 | DUP33 | CD44 (enhancer) | None |
| SXN-ESSm-loop1 | DUP33 | Fibronectin EDA (Comp. Enh.) | None |
| SXN-ESSm-stem1 | DUP33 | Fibronectin EDA (Comp. Enh.) | None |
| SXN-ESS-loop1 | DUP33 | hnRNP A1 (silencer) | None |
| SXN-ESS-stem1 | DUP33 | hnRNP A1 (silencer) | None |
| SXN-ESSn-loop1 | DUP33 | Serotonin (Comp. Sil.) | None |
| SXN-ESSn-stem2 | DUP33 | Serotonin (Comp. Sil.) | None |

3.1.8. Primers

3.1.8.1. Primers used for cloning

| Name | Orientation | Sequence | Clone name |
|---------------------|-------------|---|----------------|
| DACA1-EcoRI | sense | GAATTCATGTCGGCCACCAGCGTG | pEGFP-DACA1 |
| DACA1-SacII | antisense | CCGCGGTCATTGTTTGTTCGACTCTGCC | |
| YTHDF3-EcoRI | sense | GAATTCATGTCAGCCACTAGCGTGG A | pEGFP-YTHDF3 |
| YTHDF3-SacII | antisense | CCGCGGTTATTGTTTGTTCCTATTCTCTCCC | |
| YT(1)SalI | sense | TCGACAGAGTCCAGTCTGTCAGTCAG | MG-YT1 |
| YT(1)BamHI | antisense | GATCCTGACTGACAGACTGGACTCTG | |
| YT(2)SalI | sense | TCGACGATGCATGCAATGGATGCGG G | MG-YT2 |
| YT(2)BamHI | antisense | GATCCCCGCATCCATTGCATGCATCG | |
| YT(3)SalI | sense | TCGACGAATCCAGAATCCTGAATCC G | MG-YT3 |
| YT(3)BamHI | antisense | GATCCGGATTCAGGATTCTGGATTCCG | |
| ESEm loop1 SalI | sense | TCGACATCCATGGGGCTGGATGTGACGTACAACAACAATACGTCACATATTCCTCTCATGAG | SXN-ESEm-loop1 |
| ESEm loop1 BamHI | antisense | GATCCTCATGAGAGGAAGTATGTGACGTATTGTGGTTGTACGTCACATCCAGCCCCATGGATG | |
| ESEm stem1 SalI | sense | TCGACATGATGGGTATGTGCGTTGCTTCGGCAACAACAACCTCATCGCATACTTCCTCTCATGAG | SXN-ESEm-stem1 |
| ESEm stem1 BamHI | antisense | GATCCTCATGAGAGGAAGTATGCGATGAGTTGTGGTTGCCGAAGCAACGCACATACCCATCATG | |
| ESSm loop1 SalI | sense | TCGACATCCATGGGGCTGGATGTGACGTAAACAAGGCATACGTCACATAGCTTCCTCTCATGAG | SXN-ESSm-loop1 |
| ESSm loop1 BamHI | antisense | GATCCTCATGAGAGGAAGCTATGTGACGTATGCCTTGTTACGTCACATCCAGCCCCATGGATG | |
| ESSm stem1 SalI | sense | TCGACCTACCTTGCGCATGATACGCA TGCGCAAGGTAGCACTGCATGAGCTTCCTCACGTTTG | SXN-ESSm-stem1 |
| ESSm stem1 BamHI | antisense | GATCCAAACGTGAGGAAGCTCATGCAGTGCTACCTTGCGCATGCGTATCATGCGCAAGGTAGG | |

| Name | Orientation | Sequence | Clone name |
|------------------|-------------|---|--------------------|
| ESS loop1 Sall | sense | TCGACATCCATGGGGCTGGATGTGA CGTAGTAGGGTATACGTCACATAGCT TCCTCTCATGAG | SXN-ESS- loop1 |
| ESS loop1 BamHI | antisense | GATCCTCATGAGAGGAAGCTATGTG ACGTATACCCTACTACGTCACATCCA GCCCCATGGATG | |
| ESS stem1 Sall | sense | TCGACCTACCCTACGCATGATACGCA TG CGTAGGGTAGCACTGCATGAGCTTCC TCACGTTTG | SXN-ESS- stem1 |
| ESS stem1 BamHI | antisense | GATCCAAACGTGAGGAAGCTCATGC AGTGCTACCCTACGCATGCGTATCAT GCGTAGGGTAGG | |
| ESSn loop1 Sall | sense | TCGACATCCATGGGGCTGGATGTGA CGTAGTAAGTGAATACGTCATATCTT ACCTCTCATGAG | SXN-ESSn- loop1 |
| ESSn loop1 BamHI | antisense | GATCCTCATGAGAGGTAAGATATGA CGTATTCACTTACTACGTCACATCCA GCCCCATGGATG | |
| ESSn stem2 Sall | sense | TCGACATCCAGTAAGCTACGCTCCGA TGC GTAAGTGAGTCCGCTCACTTACG CATCTCATGAG | SXN-ESSn- stem2 |
| ESSn stem2 BamHI | antisense | GATCCTCATGAGATGCGTAAGTGAG CGGACTCACTTACGCATCGGAGCGT AGCTTACTGGATG | |

3.1.8.2. Primers used for site-directed mutagenesis

| Name | Sequence | Description | Clone name |
|-------|---------------------------------|-------------------------------|------------|
| W431D | TCTCCTATACATGATGTGCT TCCAGCA | Introduced mutant: tgg -> gat | W431D YTH |
| R478N | GTAAAGATTGGAAATGATG GACAGGAA | Introduced mutant: cgt -> aat | R478N YTH |

3.1.8.3. Primers used for overlap extension

| Name | Orientation | Sequence | Clone name |
|---------------|-------------|---|------------------------|
| YTHdel For | sense | ATCTGTCCTTgctagcAGTATTGA CTTGTATCAGCTCATTCAT | pFastBacHTa- YTHdel |
| YTHdel Rev | antisense | AGTCAATACTgctagcAAGGACA GATTTGAGTTTACTGGTTTG | |
| EcoRIpht6-HTa | sense | GAATTCGCCACCATGGCGG | |
| pht6-NotIHTa | antisense | GCGGCCGCTTATCTTCGATAA CGACCTCTTTCCCC | |

3.1.8.4. Primers used for RT-PCR

| Name | Orientation | Sequence | Target |
|-----------|-------------|------------------------------------|-----------------|
| N5Ins | sense | GAGGGATCCGCTTCCTGCCCC | CD44v5 minigene |
| N3Ins | antisense | CTCCCGGGCCACCTCCAGTGCC | |
| T7 | sense | TAATACGACTCACTATAGGG | SRp20 minigene |
| X16R | antisense | CCTGGTCGACACTCTAGATTCCTTTCATTTGACC | |
| SXNglob1F | sense | CACTCCTGATGCTGTTATGG | SXN minigene |
| Glob2AS | antisense | CCATTTGACCATTACCCACA | |
| pClfor | sense | GGTGTCCTACTCCCAGTTCAA | SMN2 minigene |
| SMNex8rev | antisense | GCCTCACCACCGTGCTGG | |

3.1.8.5. Primers used for SELEX

| Name | Orientation | Sequence |
|--------|-------------|--------------------------------------|
| T7 pro | sense | TAATACGACTCACTATAGGGATCCGAATTCCCGACT |
| RT | antisense | GGAAGCTTCTCGAGACGC |

3.1.8.6. Primers used for microarray data validation

| Name | Orientation | Sequence | Reference | Description |
|---------------|-------------|--------------------------|-----------------------|--|
| Anubl1(exon2) | sense | CAAAGCTGATTTGGGTAGAATTGG | BC048508 AK089032 | Mus musculus AN1, ubiquitin-like, homolog |
| Anubl1(exon5) | antisense | GCCAATACCAGCTTCAAGGTACA | | |
| Cxcl10(exon3) | sense | TCCGGAATCTAAGACCATCAAGAA | NM_021274 AK152234 | Mus musculus chemokine (C-X-C motif) ligand 10 |
| Cxcl10(exon4) | antisense | AGACCAAGGGCAATTAGGACTAGC | | |
| Sec24c(exon4) | sense | CTCAGCAGTTTGGTCCTCCATT | NM_172596 | Mus musculus SEC24 related gene family, member C |
| Sec24c(exon6) | antisense | CTGACTGCCAAAGGTAGGTGCT | | |
| Rhot2(exon12) | sense | CGGATGTTTCGAGAAGCATGA | NM_145999 AK142183 | Mus musculus ras homolog gene family, member T2 |
| Rhot2(exon14) | antisense | GCAAGGCATTGCTGGACATC | | |
| Zfp687(exon5) | sense | TCTCCTCACACTTTGACCAGCAC | NM_030074 AK148239 | Mus musculus zinc finger protein 687 |
| Zfp687(exon6) | antisense | CAGTTCTCCTCGTCGGGCTCT | | |
| GSK3β(exon8) | sense | GAACTCCAACAAGGGAGCAA | NM_019827 | Mus musculus |
| GSK3β(exon10) | antisense | TGGTGAAGTTGAAGAGTGCGAG | | |

3.1.9. siRNA sequence

TGGATTTGCAGGCGTGAATTA

3.2. Methods

3.2.1. Plasmid DNA isolation

Large amounts of plasmid DNA were isolated using QIAGEN Plasmid Maxi kit according to the manufacturer's protocol.

Smaller amounts of plasmid DNA were isolated from the alkaline lysis method first described by Birnboim and Doly (Birnboim and Doly, 1979). In brief, a single bacterial colony carrying the desired plasmid was picked using a sterile toothpick and cultured overnight at 37°C with vigorous shaking (220rpm) in 3ml LB medium containing the appropriate antibiotics. The cells were harvested by centrifugation for 5 minutes at 5,000 rpm. The pellet was resuspended in 200µl buffer P1. Equal volume of lysis buffer P2 was then added and the solution mixed gently by inversion. The cells were allowed to lyse for 5 minutes, followed by adding 200µl of the neutralization buffer P3. The tube was mixed gently by inversion and the solution was maintained on ice for 20 minutes. After centrifugation for 10 minutes at 12,000 rpm, the resulting supernatant was precipitated by adding 1 volume of isopropanol. Plasmid DNA was pelleted by centrifugation at 12,000 rpm for 10 minutes, washed with 70 % ethanol, air-dried and dissolved in 30µl of TE buffer.

| BUFFER P1 | BUFFER P2 | BUFFER P3 | BUFFER TE |
|-------------------|------------------|----------------------|------------------|
| 50 mM Tris-HCl | 200 mM NaOH | 3M Potassium acetate | 10 mM Tris-HCl |
| 10 mM EDTA | 1% SDS | pH 5.5 | 1 mM EDTA |
| 100 µg/ml RNase A | | | pH 8.0 |
| pH 8.0 | | | |

3.2.2. Determination of nucleic acids concentration

The DNA and RNA concentrations in solution were estimated using a spectrophotometer (Eppendorf BioPhotometer 6131). Plastic cuvettes were used for visible spectrophotometry. The absorbance of the solution was measured at 260 nm and concentration was calculated using following formulas:

$$1 A_{260}=40 \mu\text{g/ml for RNA}$$

$$1 A_{260}=50 \mu\text{g/ml for double stranded DNA}$$

$$1 A_{260}=37 \mu\text{g/ml for single stranded DNA}$$

3.2.3. Restriction endonuclease digestion

Double strand DNA was mixed with restriction endonucleases, appropriate restriction buffer, restriction enzymes, and distilled and deionized water and then incubated at the recommended temperature. Normally, 1 to 5 U of restriction enzyme was used to digest 1 µg of DNA. Restriction endonucleases were inactivated by either heating at 65°C for 15 min.

3.2.4. Electrophoresis of DNA

DNA was resolved on 0.8-2% agarose gels prepared in 1 x TBE buffer. The electrophoresis was run for 80 min at 100 V. The gels were stained for 30 min in 0.5 mg/ml ethidium bromide and visualized under UV light, $\lambda = 260$ nm.

1X TBE

90 mM Tris-borate

20 mM EDTA

6 X GEL-LOADING BUFFER

0.25% bromophenol blue

0.25% xylene cyanol FF

15% Ficoll 400 in dH₂O

3.2.5. Elution of DNA from agarose gels

DNA was purified from agarose gels where crystal violet was added to a final concentration of 2 μ g per ml to detect DNA under visible light. Individual bands were excised and DNA was extracted using the Qiagen QIAEX II gel extraction kit according to the manufacturer's protocol.

6 X CRYSTAL VIOLET GEL-LOADING BUFFER

0.25% crystal violet

15% Ficoll 400 in dH₂O

3.2.6. PCR amplification of DNA

A standard PCR reaction to amplify DNA from a plasmid template contained 1-80 ng of plasmid DNA, forward and reverse primers (0.4 μ M each), dNTPs (200 μ M), 1 x Taq polymerase buffer, 1.5 mM MgCl₂ and 1 U Taq polymerase in total volume of 25 μ l. When the amplification was made for cloning purposes, a high-fidelity polymerase, i.e. Platinum Pfx polymerase was used instead of Taq polymerase. The amplification was carried out in a Perkin Elmer GeneAmp PCR System 9700 thermocycler under the following conditions: initial denaturation for 2-4 min at 94°C; 25-35 cycles of 15-30 sec at 94°C, annealing at the T_m of the primers pair, extension of 1 min per 1 kb at 72°C (or 68°C for Pfx polymerase). After the last cycle the reaction was held for 5 min at the extension temperature to complete the amplification of all products.

3.2.7. DNA ligation

When the vector ends were blunt or compatible with each other, the vector was dephosphorylated prior to ligation to prevent self-ligation. To remove 5' phosphates from the

vector, 2 U of Calf intestinal phosphatase (CIP, Boehringer) was added to 5 µg of linearised vector in 1 x CIP buffer in 20 µl. The reaction was incubated for 1 hour at 37°C. CIP was subsequently inactivated by heating the reaction to 68°C for 20 min. A typical ligation reaction contained vector and insert at a ratio of about 1:3 (500-1000 ng total DNA), 1 x ligase buffer, 1 mM ATP and 200-400 U T4 DNA Ligase (New England Biolabs) in 10 µl. The incubation was carried out for 1 hour at room temperature for cohesive-end ligation or overnight at 16°C for blunt-end ligation. After that, one third to one half of the ligation mixture was then transformed in *E coli* cells.

3.2.8. Preparation of competent *E. coli* cells

5 ml of LB medium were inoculated with a single bacterial colony and grown overnight at 37°C with vigorous shaking. 4 ml of this culture were transferred to 250 ml LB and grown to early logarithmic phase (OD₆₀₀ = 0.3-0.6). The culture was transferred into a sterile 50 ml falcon tube and centrifuged for 10 min at 2500 rpm at 4°C. The bacterial pellet was sequentially resuspended in 1/10 volume of cold TSB buffer and incubated on ice for 10 min. Cells were aliquoted into cold Eppendorf tubes and frozen in liquid nitrogen. Competent bacterial cells could then be stored at -80°C for several months.

TSB BUFFER

10% PEG 3500

5% DMSO

10 mM MgCl₂

10 mM MgSO₄

in LB medium

pH 6.1

3.2.9. Transformation of *E. coli* cells

1-10 ng of plasmid DNA or a ligation reaction were added to 20 µl of 5 x KCM buffer and then the volume was brought with water up to 100 µl. Equal volume of competent cells was added. The reaction mixture was incubated on ice for 20 min followed by incubation at room temperature for 10 min. Then 1 ml of LB medium was added and the bacteria were incubated for 1 h at 37°C with vigorous shaking. Finally cells were plated on LB Agar plates containing appropriate antibiotic. Plates were incubated at 37°C until colonies were visible.

5 X KCM BUFFER

500 mM KCl

150 mM CaCl₂

250 mM MgCl₂

3.2.10. Site-direction mutagenesis of DNA

Site-directed mutagenesis was performed according to the method described by Kunkel (Kunkel T.A. et al., 1985). The DNA of interest was cloned into a vector carrying the f1 phage origin of replication and thus capable of existing in both single- and double-stranded forms. The recombinant plasmid was transformed into *E.coli* strain CJ236 deficient in dUTPase (*dut*) and uracil N-glycosylase (*ung*). These mutations result in a number of uracils being substituted for thymine in the nascent DNA. After transformation, bacteria were grown on plates containing chloramphenicol in addition to the plasmid specific antibiotic, to ensure the presence of the F' episome necessary for production of helper phage. To isolate single-stranded DNA from the plasmid of interest, colonies were grown in 5 ml of LB medium for 90 min and then 5108 x pfu of helper phage M13KO7 (New England BioLabs) was added. The culture was grown for overnight at 37 °C and single-stranded DNA was isolated with the Qiagen M13 kit according to the manufacturer's protocol. This uracil containing DNA was used as a template in the *in vitro* mutagenesis reaction. Phosphorylated oligonucleotides containing desired mutations were annealed to the template at a molar ratio of 20:1 in 10 µl of 1 x T7 DNA polymerase buffer. The DNA was denatured for 5 min at 94 °C and then the temperature was gradually decreased from 70 °C to 37 °C at a rate of 1 °C per minute. The extension of the annealed primer was carried out in 20 µl by adding to the same tube 1 µl of 10 x T7 DNA Polymerase buffer, 0.8 µl of 10 mM dNTPs, 1.5 µl of 10 mM ATP, 3 U T7 DNA Polymerase and 2 U FastLink T4 DNA Ligase. The reaction was incubated at 37 °C for 45 min. The ligase was inactivated by incubation at 65 °C for 20 min. The mutagenesis reaction was transformed into competent XL1Blue *E.coli* cells. Replication of the plasmid in this strain leads to repair of the template strand and consequently to production of plasmid carrying the desired mutation. All mutant plasmids were verified by sequencing.

3.2.11. Domain deletion by overlap extension

Four primers were designed to introduce mutations by this method which was first described by Higuchi et al., 1989. One set of forward F and reverse R primer was complementary to the extreme ends of the DNA template. The other set of forward MF and reverse MR primer was designed such that it was complementary to the desired mutant sequence across the deleted domain. For proper amplification, it was ensured that the complementarity to

the mutant template in the 3' region of the primers was more than that at the 5' end (see Figure 10). The first PCR was carried out with Proofreading Taq polymerase to avoid any overhang of adenosine residue. Individual PCRs were carried out to amplify fragments with primer sets F/MR and with MF/R respectively.

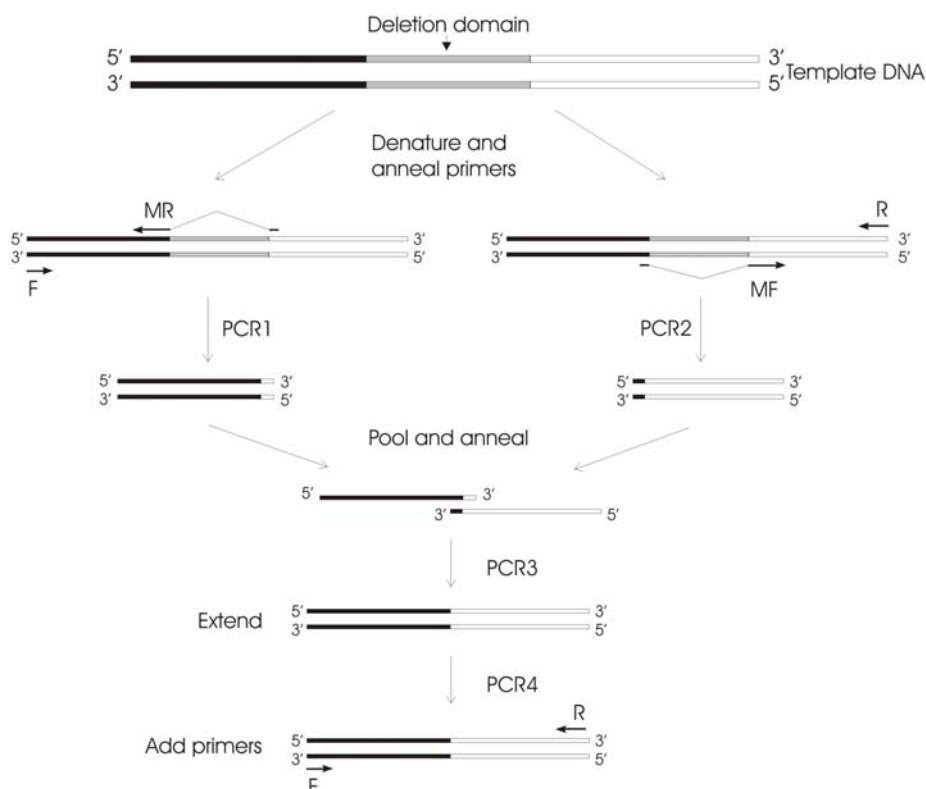


Figure 10: Domain deletion by overlap extension. In separate PCR amplification reactions 1 (Primers F and MR) and 2 (Primers MF and R), two partially overlapping fragments of the target gene are amplified. In PCR 3, the denatured products from PCR 1 and PCR 2 anneal at the region of overlap and extend to form full length double – stranded mutant DNA. In PCR 4, the full length mutant DNA is amplified using primers F and R. (Modified from *Molecular Cloning: A Laboratory Manual*, Sambrook and Russel, third edition, 2001).

The amplified fragments were gel eluted to free them from any contaminating DNA template. 200 ng of the individual purified fragments were pooled together and allowed to anneal and extend without any addition of primer with dNTPs (200 μ M), 1 x Taq polymerase buffer, 1.5 mM MgCl₂ and 1 U Taq polymerase in total volume of 25 μ l. The amplification was carried out in a Perkin Elmer GeneAmp PCR System 9700 thermocycler under the following conditions: initial denaturation for 5 min at 94°C; 10 cycles of 30 sec at 94°C, annealing at 50°C, extension of 1 min per 1 kb at 72°C. After the last cycle the reaction was held for 5 min at the extension temperature to complete the amplification of all products. External primers (F and R) were then added and the reaction was again supplemented with 1 U of Taq polymerase. Final PCR was performed with the following conditions: initial denaturation for 5 min at 94°C; 30 cycles of 30

sec at 94°C, annealing at 60°C and extension of 1 min per 1 kb at 72°C. The last cycle was followed by another 5 min of extension at 72°C. A part of the amplified fragment was run on the Agarose gel and the other subcloned into pCR4 TOPO (Invitrogen) for sequencing.

3.2.12. Expression and purification of HIS-tagged proteins in Baculovirus Expression System

3.2.12.1. Expression of recombinant protein

The Bac-to-Bac[®] Baculovirus Expression System was developed based on site-specific transposition of an expression cassette into a baculovirus shuttle vector propagated in *E. coli* (Luckow V.A. et al., 1993).

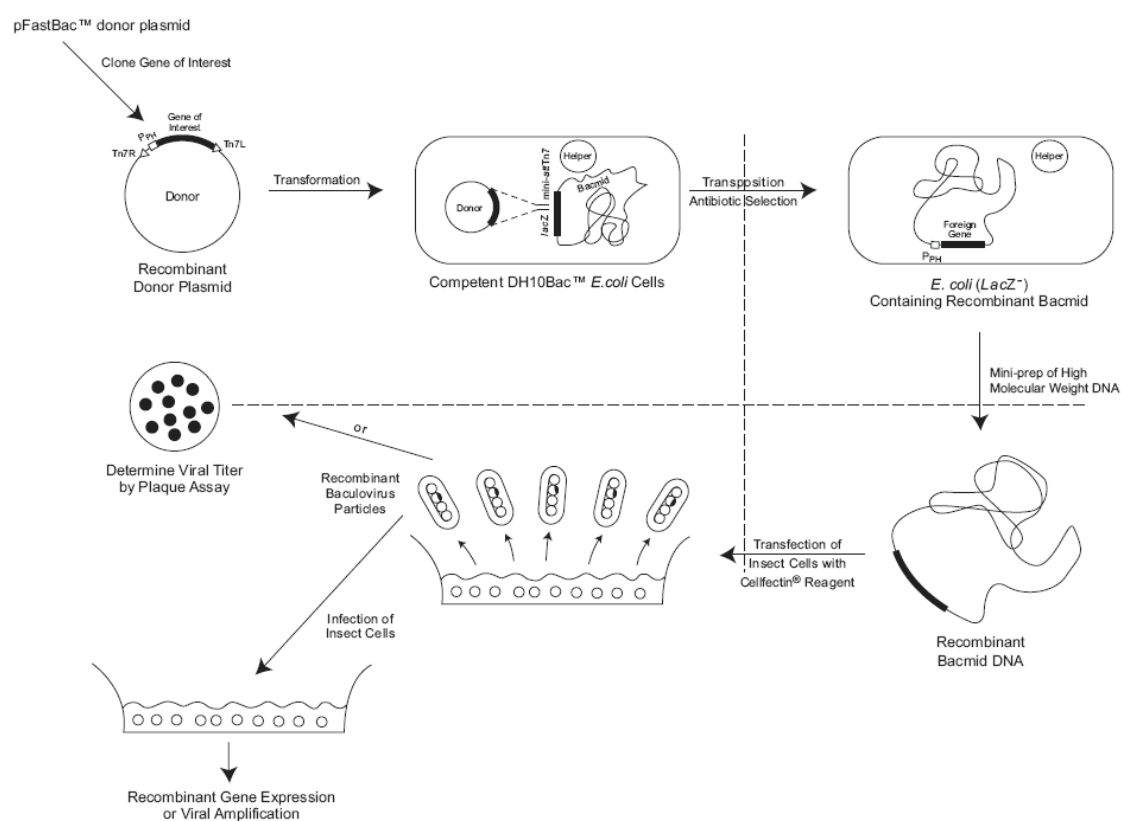


Figure 11: Generation of recombinant baculovirus and the expression of the gene of interest using the Bac-to-Bac Baculovirus Expression system. The expression of the gene of interest is controlled by polyhedron (PH) promoter of *Autographa californica multiple nuclear polyhedrosis virus* (AcMNPV) for a high-level expression in insect cells. The expression cassette is flanked by the left and right arms of transposon 7 elements (Tn7) Tn7R and Tn7L, and mini-attTn7 refers to a short segment containing the attachment site for Tn7. DH10Bac cells contain a baculovirus shuttle vector (bacmid) with a mini-attTn7 target site and a helper plasmid. Transposition occurs between the mini-attTn7 element on the pFastBac vector and the mini-attTn7 target site on the bacmid to generate a recombinant bacmid (Adapted from Invitrogen instruction manual).

The DNA of interest was cloned into pFastBac HTa vector. Purified plasmid construct (200 pg / μ l in TE, pH 8) was gently mixed with 100 μ l of the DH10Bac cells, and incubated on ice for 30 min, followed by heat-shock at 42°C for 45 seconds. Then 900 μ l of LB medium was added and followed by 4 h incubation at 37 °C with shaking (225rpm). The transformation

product was diluted from 10 to 1000 times, and 100 μ l of each dilution was plated on an LB agar plates containing 50 μ g/ml kanamycin, 7 μ g/ml gentamicin, 10 μ g/ml tetracycline, 100 μ g/ml Bluo-gal, and 40 μ g/ml IPTG. After incubation at 37 °C for 48 h, white colonies were picked for bacmid isolation. The recombinant bacmid was verified by PCR using gene specific primers.

Before transfection, 9×10^5 / 8cm^2 Sf9 cells were seeded and allowed to attach for at least 1 h at 27°C. 1 μ g of purified bacmid DNA was diluted in 100 μ l of unsupplemented Grace's Medium and 6 μ l of Cellfectin reagent was diluted in 100 μ l of unsupplemented Grace's Medium in parallel. The two dilutions were combined and mixed gently followed by 15 to 45 min incubation at RT. The DNA:lipid complexes were gently added to the cells for 5 h and then replaced by growth media.

Once the transfected cells demonstrate signs of late stage infection, the medium was collected for viral plaque assay to determine the titer of baculoviral stock. The baculoviral stock with a suitable titer was used to infect insect cells for recombinant protein expression.

3.2.12.2. Purification of recombinant protein

The purification system used for HIS-tagged recombinant protein is based on the selectivity and affinity of nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography matrices for biomolecules which have been tagged with 6 consecutive histidine residues.

The infected cells were washed by PBS and collected by centrifugation for 5 min at 1000g. 4 ml lysis buffer were used per $1-2 \times 10^7$ cells. After incubation on ice for 10 min, the lysate was centrifuged at 10,000g for 10 min at 4°C to pellet cellular debris and DNA. 200 μ l of 50% Ni-NTA slurry (equilibrated with PBS and lysis buffer) were added to the cleared lysate and binding for 2 h at 4°C with gentle shaking (200 rpm on a rotary shaker). After binding, Ni-NTA agarose beads were pelleted by centrifugation at 800g for 5 min, followed by two times washing with 800 μ l wash buffer. HIS-tagged recombinant protein was then eluted with 100-200 μ l elution buffer. The eluates were collected by centrifug 800g for 5 min at 4°C. The purified recombinant protein was analyzed by SDS-PAGE.

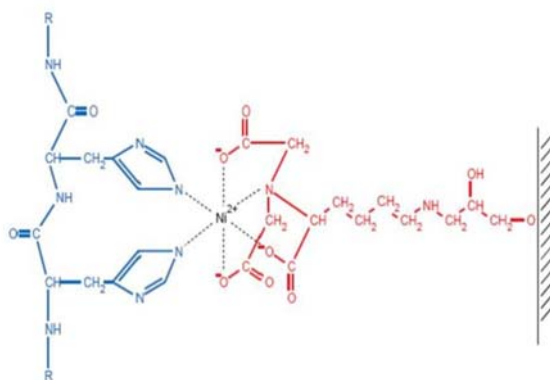


Figure 12: Interaction between neighboring residues in the 6xHis tag and Ni-NTA matrix. Nitrilotriacetic acid (NTA) is a tetradentate chelating adsorbent. NTA occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the 6xHis-tagged protein (Adapted from Qiagen instruction manual, 2003).

Buffers used for purification of HIS-tagged recombinant proteins from insect cells:

| Lysis buffer | Wash buffer | Elution buffer |
|------------------------|------------------------|------------------------|
| 50mM NaPO ₄ | 50mM NaPO ₄ | 50mM NaPO ₄ |
| 300mM NaCl | 300mM NaCl | 550mM NaCl |
| 15mM imidazole | 15mM imidazole | 400mM imidazole |
| <u>Freshly add</u> | <u>Freshly add</u> | <u>Freshly add</u> |
| 1% NP40 | 1mM PMSF | 1mM PMSF |
| 100U/ml Benzonase | pH 8.0 | pH 8.0 |
| 1μg/ml Aprotinine | | |
| 0.5μg/ml leupeptine | | |
| 1mM PMSF | | |
| pH 8.0 | | |

3.2.13. Determination of protein concentration

Protein concentration was estimated using BioRad Protein Assay Kit based on Bradford method. Protein in 800 μl of distilled water was mixed with 200 μl of 1 x Dye Reagent and incubated for 5 min at RT. Absorbance of the solution was measured in a spectrophotometer at $\lambda=595$ nm. Concentration of samples was read from the standard curve where OD₅₉₅ was plotted versus concentration of BSA standards.

3.2.14. Systematic evolution of ligands by exponential enrichment (SELEX)

The technique of *in vitro* selection or SELEX was independently developed by Szostak and L. Gold in 1990. SELEX allows the simultaneous screening of highly diverse pools of different RNA or DNA (dsDNA or ssDNA) molecules for a particular feature.

3.2.14.1. *In vitro* transcription

The initial DNA pool for generating 20 nucleotides degenerate RNA random sequence was amplified by PCR using 10 pM of DNA pool, 10 pM of T7 pro primer and 10 pM of RT primer in a standard 50µl PCR reaction. The PCR condition is: initial denaturation 4 min at 94°C; 25 cycles of 30 sec at 96°C, annealing at the 55°C for 30 sec, extension of 30 sec at 72°C. After the last cycle the reaction was held for 5 min at the extension temperature.

The PCR product was gel purified and transcribed with T7 RNA polymerase (20U/µl), rNTP (10mM), DTT (0.1M), MgCl₂ (1M) and RNasin (40U/µl) 3 h at 37°C followed by DNase I treatment for 30 min at 37°C. The RNA random sequence pool was further purified by phenol/chloroform extraction, and precipitated by ethanol. The RNA pellet was washed with 70% ethanol and resuspended in 70 µl binding buffer (see 3.2.14.2, below).

3.2.14.2. *In vitro* selection

The RNA pool was then pre-incubated with binding buffer (without tRNA) washed Ni-NTA agarose beads for 15 min at 4°C to avoid non-specific binding. After pre-incubation the Ni-NTA beads was removed from the system by centrifugation. The supernatant containing RNA pool was transferred to a fresh tube. For *in vitro* selection, 2µg HIS-YT521-B fusion protein was added to the RNA pool for 30 min at RT with shaking, followed by addition of 25 µl pre-washed Ni-NTA beads to the binding system for 30 min at 4°C. Afterwards Ni-NTA beads were washed twice with binding buffer and resuspend in 50µl of 2 x proteinase K buffer and incubated with 5µg of proteinase K at 37°C for 30 min. Selected RNA was purified by phenol/chloroform extraction and ethanol precipitation. The selected RNA was resuspended in 1x RT buffer and incubated at 65°C for 10 min then put on ice for 2 min. Afterwards, 0.5µl of 0.1M DTT, 1.0µl RNasin and 2.0µl Superscript II (Invitrogen) were added to RT reaction which was incubated at 42°C for 2 hours, then heated 15 min at 72°C. The RT mixture was subsequently used for PCR amplification (96°C 30', 55°C 30' 72°C 30', 28 cycles) The RNA transcript for the next round of selection was prepared as described above. After 7 rounds of selection, PCR products was cloned into pCRTPO vector followed by DNA sequencing.

Buffer used in SELEX:

| Binding buffer | 2 x Proteinase K buffer | RT buffer |
|-------------------------|--------------------------------|--------------------------|
| 10mM Tris | 200mM Tris | 10µl first strand buffer |
| 100mM KCl | 25mM EDTA | 4µl RT primer |
| 2.5mM MgCl ₂ | 300mM NaCl | 5µl 10mM dNTPs |
| 0.1% TritonX-100 | 2% SDS | 31µl H ₂ O |
| 0.1mg/ml tRNA | pH7.5 | |
| pH7.5 | | |

3.2.15. Minigene construction

Forward and reverse primers used as the second exon of SXN minigene were designed with a 5' *Sall* site and a 3' *BamHI* site. 10 μ l of each oligo (0.37 μ g/ μ l) were phosphorylated by T4 PNK in the presence of 2 μ l 10mM ATP for 1 h at 37°C. After phosphorylation, forward and reverse primers were annealed and inserted into *Sall* and *BamHI* linerized vector.

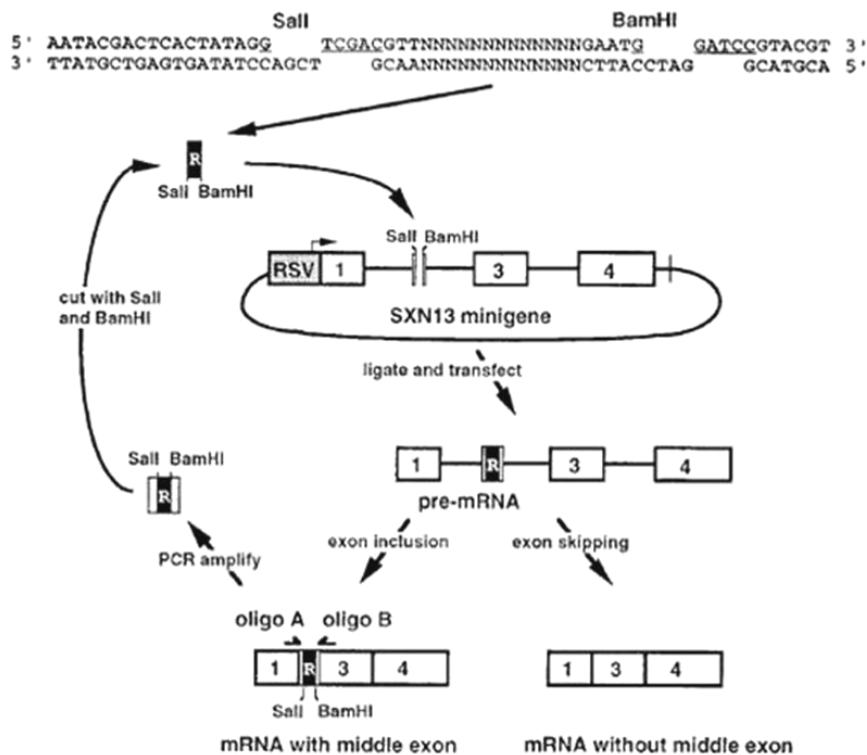


Figure 13: Minigene constructs. Primers designed with either a 5' *Sall* site and or a 3' *BamHI* site used as alternative exon were inserted into SXN minigene. This alternative exon flanked by duplicated intron 1 from human β -globin such that the first and third exons of the minigene are globin exons 1 and 2. Introns 3 and exon 4 of the minigene are derived from cTNT intron 6 and exon 18. The level of alternative exon inclusion can be modulated by changing exon sequence and size (Adapted from Coulter L. et al., 1997).

3.2.16. RNA electrophoretic mobility shift assay (gel shift)

RNA gel shift assays were performed according to the protocols described previously (Black et al., 1998; Thomson et al., 1999). 1 μ g of purified YT521-B or YTHdel recombinant proteins were pre-incubated for 10 minutes at room temperature in 1x binding buffer (150 μ g/ml BSA, 5mM DTT and 0.8 μ l of 80 mM MgAc) in a final volume of 20 μ l. After radiolabeled RNA oligonucleotides were added, the mixture was incubated for 10 more minutes. 20 μ l of 10% glycerol as a sample buffer were added and RNA:protein complexes were separated on 3.5 % native polyacrylamide gel in 1 x TrisAc MgAc buffer at 4°C. The gel was pre-run for 30 minutes at 4°C before loading the samples.

Buffer used in gel shift:

| 10 x TrisAc MgAc | 3.5% NATIVE PA GEL (30 ml): | 5 x binding buffer |
|-------------------------|------------------------------------|---------------------------|
| 250 mM TrisAc | 3.5 ml 30% AA/Bis 37.5:1 | 60 mM Hepes-KOH |
| 50mM MgAc | 3 ml 10 x TrisAc MgAc | 60% glycerol |
| pH 7.5 | 1.5 ml 50% Glycerol | 240 mM KCl |
| | 240 µl 10% APS | 0.6 mM EDTA |
| | 15 µl TEMED | |
| | 21.8 ml dH2O | |

3.2.16.1. Labeling of RNA probe

The probes used for gel shift were labeled by *in vitro* transcription. 1 µg of DNA template were mixed with 5 µl of 800 Ci/mmol [$\alpha^{32}\text{P}$]-CTP in 1 x transcription optimized buffer, 0.5 mM of rATP, rGTP, rUTP, 0.01 mM of rCTP, 10 mM DTT, 10 units of T7 RNA polymerase were added and the reaction was carried out at 37°C. 2 h after incubation, DNase I treatment was done by adding 1 µl of RQ1 DNase and incubating for another 15 min. Radiolabeled RNA was purified by phenol/chloroform extraction and ethanol precipitation.

3.2.17. Light scattering assay

Fluorimeter SFM25 (Kontron Analytical) was employed in the light scattering experiment. Recombinant protein (final concentration is 0.06µg/µl) was pre-incubated with binding buffer for 10 minutes at room temperature. The scattered wavelength at 600nm was captured and recorded. After pre-incubation, *in vitro* transcribed RNA (5nM) was added to the system, binding for 10 minutes then 10% TCA was added to precipitate proteins.

5 x Binding buffer

50 mM HEPES
 500mM KCl
 0.1% Triton X-100
 50% glycerol
 25mM DTT (freshly add)
 5mM PMSF(freshly add)
 pH7.5

3.2.18. Freezing, thawing and subculturing of eukaryotic cells

To freeze, cells were grown to mid logarithmic phase (about 75% of confluence) in 10 cm Petri dishes. They were collected by trypsinization with 1 x Trypsin/EDTA, resuspended in 1 ml of the freezing medium (90% of the growth medium and 10% of DMSO). Vials were placed in Nalge Nunc Cooler giving a cooling rate of $\sim 1^{\circ}\text{C}/\text{min}$ while at -80°C . Cells were stored later in liquid nitrogen.

To thaw, cells were incubated at 37°C . The entire content of the tube was transferred to a 10 cm Petri dish and 10 ml of the growth medium were added. The dish was placed in the incubator at 37°C and 5% CO_2 . When cells were attached to the plastic surface, the medium was removed and replaced with fresh one. The cells were maintained in the incubator until ready for the subculturing.

Cells were subcultured after reaching confluence. The monolayer was detached by adding 1 X Trypsin /EDTA and incubating at 37°C until single cell suspension was formed. 1/5 – 1/10 of this suspension was transferred to a new dish and mixed with the growth medium. Cells were maintained in the incubator at 37°C and 5% CO_2 .

3.2.19. Transfection of eukaryotic cells

3.2.19.1. Plasmid transfection

The procedure used for HEK293 cells was based on the one published by Chen and Okayama (Chen and Okayama, 1987). Exponentially growing cells were replated at a density of about 3×10^5 cells / 8 cm^2 . Growth medium was added and the cells were incubated at 37°C , 5% CO_2 for about 24 h, to reach 60-70% of confluence. For most applications cells were grown in 6-well plates, with 2 ml of growth medium per well. The transfection reaction for one well was made the following way. 1 to 5 μg of expression construct were mixed with 25 μl of 1 M CaCl_2 in final volume of 100 μl . Equal volumes of 2 x HBS buffer was added drop by drop, with constant mixing. In order to form a precipitate, the solution was allowed to stay at room temperature for 20 min. After that, it was added to the growth medium. To express the transfected plasmid, cells were grown for additional 24 h at 37°C , 3 % CO_2 .

2 X HBS buffer

280 mM NaCl

10 mM KCl

1.5 mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$

12 mM Dextrose

50 mM Hepes

pH 6.95

3.2.19.2. siRNA transfection

HEK293 cells were cultured in DMEM supplemented with 10% fetal calf serum and the optimal confluency of adherent cells should be 50-80% on the day of transfection. Optimal On the day before transfection, $2-8 \times 10^4$ cells were seeded into a single well of a 24-well plate in 0.5ml of medium. On the day of transfection, 37.5 ng siRNA were diluted in 100 μ l of DMEM medium without serum to reach a final siRNA concentration of 5 nM. 3 μ l of HiPerfect transfection reagent was added to the dilution and then mixed by vortexing. The samples were incubated for 5-10 min at RT before the formed complexes were added drop-wise onto the cells. The cells with the transfection complexes were maintained under normal growth condition. After an appropriate time, the cells were harvested and used for experiment.

3.2.19.3. Plasmid and siRNA transfection

One day before transfection, appropriate amount of HEK293 cells were seeded into a 24-well plate. The optimal confluency for transfection is 50-80%. On the day of transfection, 100ng of minigene and 34.5ng siRNA were diluted in 25 μ l of HBS buffer. 3 μ l of Hiperfect Transfection Reagent was added to 22 μ l of HBS buffer. After incubating for 5 minutes at RT, the two dilutions were mixed, vortexed and incubated for another 5 minutes at RT. The transfection complexes were added onto the cells. 6-72 h after transfection cells were harvested for RNA and protein isolation.

HBS buffer

20mM Hepes

150mM NaCl

pH7.4

3.2.20. Fixing attached eukaryotic cells on cover slips

Cells grown on cover slips and transfected with pEGFP-C2 constructs were fixed with 4% formaldehyde in 1 x PBS, pH 7.4 for 20 min at 4 °C. Cells were washed 3 times in PBS prior to mounting on microscope slides with Gel/Mount (Biomed). Cells were examined by confocal laser scanning microscopy (Leica).

3.2.21. Immunostaining

Cells were fixed in 4% formaldehyde and PBS for 20 min at 4°C, washed three times in PBS and 0.1% Triton X-100, and blocked in PBS, 0.1% Triton X-100, and 3% NGS for 2 hours at room temperature. Cells were then incubated with primary antibody (diluted in PBS, 0.1% Triton X-100, and 3% bovine serum albumin) overnight at 4°C and washed three times in PBS and 0.1% Triton X-100. Cells were then incubated with fluorescent molecule -coupled secondary

antibody (Dianova), diluted in PBS and 0.1% Triton X-100, for 2 hours at room temperature, washed three times in PBS and 0.1% Triton X-100, and mounted on glass slides with Gel/Mount. Finally the cells were examined by confocal laser scanning microscopy.

3.2.22. Immunoprecipitation of proteins

20-24 hours after plating transfection, cells were washed with 1x cold PBS and lysed for 20 min on ice in 200 μ l of RIPA buffer. The lysates were collected and cleared by centrifugation for about 1 min at 12800 rpm. The supernatant was transferred to a fresh Eppendorf tube and diluted with 800 μ l of RIPA rescue buffer. The antibody recognizing the tag of the expressed protein was added to the lysates and incubated at 4 °C overnight on a rotating wheel. On the next day, 50 μ l of Protein A Sepharose / Sepharose CL-4B (1:1) was added and the incubation continued for 3 to 6 h. The Sepharose beads were pelleted by centrifugation for 1 min at 1000 rpm in a microcentrifuge followed by 3 washes with 500 μ l of 1 x HNTG buffer. 30 μ l of 3 x SDS sample buffer were finally added to the washed pellet and boiled for 5 min at 95 °C. The beads were spun down and the supernatant loaded on SDS-polyacrylamide gel. After proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), then transferred to nitrocellulose membrane and followed by Western blot analysis.

Preparation of Protein A Sepharose / Sepharose CL-4B:

Protein A Sepharose beads were washed in 15 ml of distilled H₂O and pelleted at 500 rpm for 2 min at 4 °C. After a second wash with dH₂O equal volume of Sepharose CL-4B was added and the beads were washed two more times in RIPA rescue buffer and kept at 4 °C.

Buffer used for immunoprecipitation:

| RIPA buffer | RIPA RESCUE buffer | HNTG wash buffer |
|----------------------------------|----------------------------------|-------------------------|
| 1% NP40 | 20 mM NaCl | 150 mM NaCl |
| 1% Sodium deoxycholate | 10 mM NaPO ₄ , pH 7.2 | 50 mM HEPES, pH 7.5 |
| 0.1% SDS | 1 mM NaF | 1 mM EDTA |
| 150 mM NaCl | 5 mM β -glycerophosphate | 10% Glycerol |
| 10 mM NaPO ₄ , pH 7.2 | <u>freshly added</u> | 0.1% Triton X-100 |
| 2 mM EDTA | 1 mM DTT | <u>freshly added</u> |
| 50 mM NaF | 2 mM Sodium | 2 mM Sodium |
| 5 mM β -glycerophosphate | orthovanadate | orthovanadate |
| <u>freshly added</u> | 1 mM PMSF | 100 mM NaF |
| 4 mM Sodium orthovanadate | 20 μ g/ml Aprotinin | 1 mM PMSF |
| 1 mM DTT | | 20 μ g/ml Aprotinin |
| 1 mM PMSF | | |
| 20 μ g/ml Aprotinin | | |
| 100 U/ml Benzonase | | |

3.2.23. Electrophoresis of proteins

Proteins bands were resolved on denaturing SDS polyacrylamide gels, using the BioRad gel electrophoresis system (with standards: 10 cm × 7.5 cm × 0.5 cm gels). The separating gel was 7.5-15%, depending on the molecular weight of the proteins, and the stacking gel was 4%. The proteins were mixed with sample loading buffer, denatured at 95 °C for 5 min and loaded on the gel. Electrophoresis was carried out at 100 V for 2 hours in SDS gel running buffer.

Solution used in SDS PAGE:

| Separating gel (10ml) | 7.5% | Stacking gel (10ml) | 4% |
|------------------------------|-------------|----------------------------|-----------|
| dH ₂ O | 4.85 ml | dH ₂ O | 6.1 ml |
| 1.5 M Tris-HCl, pH 8.8 | 2.5 ml | 0.5 M Tris-HCl, pH 8.8 | 2.5 ml |
| 10% SDS | 100 µl | 10% SDS | 100 µl |
| 30% Acrylamide/Bis | 2.5 ml | 30% Acrylamide/Bis | 1.3 ml |
| 10% Ammonium Persulfate | 100 µl | 10% Ammonium Persulfate | 100 µl |
| TEMED | 10 µl | TEMED | 10 µl |

3 × SDS SAMPLE BUFFER

150 mM Tris-HCl, pH 6.8
 6% SDS
 30% Glycerol
 3% β-Mercaptoethanol
 0.3% Bromophenol blue

SDS GEL RUNNING BUFFER

250 mM Glycine, pH 8.3
 25 mM Tris
 0.1% SDS

3.2.24. Staining of protein gels

After disassembling SDS-PAGE, the gel was incubated in Coomassie blue staining solution (2.5% Coomassie Brilliant Blue R250, 45% Methanol, and 10% Acetic acid) for 2-3 h at RT with shaking. The gel was then destained for 20 min in 50% Methanol/10% Acetic acid and 2-3 more times in 20% Methanol/10% Acetic acid until the background became clear. The gel was dried using a gel dryer (Biorad).

3.2.25. Western blotting

Proteins resolved on SDS polyacrylamide gels were transferred to nitrocellulose membrane (Schleicher and Schuell) in transfer buffer, for 45 min at 120 V. Before the transfer, membrane and the gel were equilibrated for 5 min in the protein transfer buffer. After transferring the membrane was blocked for 1 hour in 1 x NET-gelatine buffer at RT. Primary antibody was then added and the incubation was allowed to proceed overnight at 4 °C or at RT

for 2 hours. The membrane was washed three times for 15-20 min in 1 x NET-gelatine and incubated with a secondary antibody coupled to horseradish peroxidase for 1 hour. The membrane was subsequently washed three times for 20 min in 1 x NET-gelatine and the bound antibodies were detected by the ECL system. Equal amounts of solutions ECL1 and ECL2 were mixed and added to the membrane for 5 min. The membrane was then exposed to an X-ray film (Fuji SuperRX) and developed in a Kodak developing machine.

Buffer used in Western blot:

TRANSFER BUFFER

192 mM Glycine

25 mM Tris

20% Methanol

NET-GELATINE

150 mM NaCl

5 mM EDTA

50 mM Tris-HCl, pH 7.5

0.05% Triton X-100

0.25% Gelatine

ECL1:

4.5 mM Luminol

4.3 mM p-Iodophenol

100 mM Tris, pH 9.5

ECL2:

0.003% H₂O₂

100 mM Tris, pH 9.5

3.2.26. Solubility assay

20-24 hours after the transfection, cells were lysed for 20 min on ice in 200 µl of HNTG buffer. The lysates were centrifuged at 13000g for 15 min at 4°C, 100 µl 3 x SDS sample buffer was added to the supernatant, and 300 µl 3x SDS sample buffer was added to the remaining pellet. The probes were mixed, boiled for 5 min, and 30 µl of each fraction loaded onto 7.5% SDS-polyacrylamide gels. Alternatively, cells were lysed for 20 min on ice in 200 µl of RIPA buffer. Lysates were centrifuged for 15 min at 4°C, 100 µl 3 x SDS sample buffer was added to the supernatant, and 300 µl 3 x SDS sample buffer was added to the pellet. 30 µl of the fractions was loaded in each lane and analyzed on 7.5% SDS-polyacrylamide gels. The proteins were analyzed by Western blotting and ECL using anti GFP and PK2 antibodies.

HNTG LYSIS BUFFER

50 mM Hepes, pH 7.5
 150 mM NaCl
 1% Triton X-100
 10% Glycerol
 1 mM EDTA
 20 mM Sodium pyrophosphate
 2 mM Sodium orthovanadate
 100 mM NaF
 5 mM β -glycerophosphate
Freshly add
 1 mM PMSF
 1 μ g/ml Aprotinin
 100 U/ml Benzonase

RIPA LYSIS BUFFER

0.01 M Sodium phosphate
 1% NP-40
 1% Sodium deoxycholate
 0.1% SDS
 2 mM EDTA
 0.15 M NaCl
 4 mM Sodium orthovanadate
 1 mM NaF
 5 mM β -glycerolphosphate
Freshly add
 1 mM PMSF
 1 μ g/ml Aprotinin
 100 U/ml Benzonase

3.2.27. *In vivo* splicing assay

To determine the influence of a protein on the splicing of selected minigenes, *in vivo* splicing was performed as described (Stoss et al., 1999; Tang et al., 2005). 1 to 2 μ g of the minigene plasmid were transfected in eukaryotic cells together with an expression construct for the protein. Usually a concentration-dependent effect was assessed. The protein was transfected in increasing amounts, in the range of 0 to 3 μ g. To avoid ‘squenching’ effects, the ‘empty’ parental expression plasmid containing the promoter was added in decreasing amounts, to ensure a constant amount of transfected DNA. Cells were seeded in 6-well plates and transfection was done 24 hours after plating. After incubation for 14-17 hours at 3% CO₂ total RNA was isolated from the cells.

400 ng of RNA were used in a reverse transcription reaction (section 3.2.24.). The reverse primer used for RT was specific for the vector in which the minigene was cloned, to suppress reverse transcription of the endogenous RNA. To avoid the problem of the amplification of minigene DNA, DpnI restriction enzyme was added into the reverse transcription reaction. DpnI cuts GATC sequence in double-stranded DNA when the adenosine is methylated but does not cut non-methylated single-stranded DNA or cDNA. A control reaction with dH₂O instead of RNA was included.

1/8 of the reverse transcription reactions were used for PCR with minigene-specific primers (section 3.2.24.). The primers were selected to amplify alternatively spliced minigene

products. A control reaction with no template (RNA instead of cDNA) was included in the PCR. The PCR programs were optimized for each minigene in trial experiments.

PCR reactions were resolved on a 0.3-0.4 cm thick 2 % agarose TBE gel and the image was analyzed using ImageJ analysis software (<http://rsb.info.nih.gov/ij/>).

3.2.28. Isolation of total RNA

Total RNA was isolated from eukaryotic cells grown in 6-well plates. Cells were washed with 1 x PBS and the RNeasy Mini kit was used according to the manufacturer's protocol. RNA was eluted from the column in 30 μ l of RNase-free dH₂O. However this procedure was applied only when the RNA of interest was greater than 200 bases.

3.2.29. RT-PCR

400 ng of total RNA (200 ng/ μ l), 5 pmol of reverse primer, 40 U of SuperScript II reverse transcriptase, and optionally 4 U of DpnI restriction endonuclease were mixed in 5 μ l of RT buffer. To reverse transcribe the RNA, the reaction was incubated at 42°C for 45 min.

1/8 of a typical reverse transcription reaction was used to amplify cDNA. The reaction was held in 25 μ l and contained 10 pmol of specific forward and reverse primers, 200 mM dNTPs, 1 x Taq polymerase buffer and 1 U of Taq DNA polymerase. The conditions of the PCR cycles were dependent on the template to be amplified (see section 3.2.22. for conditions of amplifying minigene products from *in vivo* splicing assays).

RT BUFFER

300 μ l 5 X First strand synthesis buffer (Invitrogen)

150 μ l 0.1 M DTT (Invitrogen)

75 μ l 10 mM dNTPs

475 μ l dH₂O

3.2.30. Affymetrix Mouse Exon 1.0 ST array analysis

Each 5 μ g of total RNA at a concentration of 1 μ g/ μ l (OD_{260/280} : Between 1.9-2.1) from three independent experiment were provided to the company Affymetrix for Mouse Exon 1.0 ST Array hybridization. The microarray result was analyzed by using Arrayassist software.

3.3. Databases and computational tools

| Database/software | URL | Description | Reference |
|-------------------|---|---|-----------------------------|
| ASD | http://www.ebi.ac.uk/asd | The alternative splicing database | (Thanaraj A. et al., 2004) |
| ASePCR | http://genome.ewha.ac.kr/ASePCR/ | Web-based application emulating the RT-PCR in various tissues | (Thanaraj A. et al., 2004) |

| Database/software | URL | Description | Reference |
|--------------------------|---|---|---|
| ClustalW | http://www.ebi.ac.uk/clustalw/index.html | Multiple sequence alignment program for DNA or proteins | (Thompson J.D. et al, 1994) |
| ESE | http://rulai.cshl.edu/tools/ESE/ | Finds putative binding regions for several splice factors | (Cartegni L. et al., 2003) |
| Human BLAT Search | http://www.genome.ucsc.edu/cgi-bin/hgBlat | Sequence alignment tool similar to BLAST | (Kent W.J., 2002) |
| NCBI BLAST and PSI-BLAST | http://www3.ncbi.nlm.nih.gov/BLAST/ | Finds regions of sequence similarity | (Altschul S.F. et al., 1990); (Altschul S.F. et al., 1997) |

4. RESULTS

YT521-B was identified by yeast two-hybrid as interactor of the splicing factors Sam68/p62, rSLM-1, rSLM-2, hnRNPG, and rSAF-B that are implicated in RNA metabolism. Previous studies showed that YT521-B is localized in a novel nuclear compartment and is characteristically concentrated in 5-20 evenly distributed dots named YT bodies (Hartmann et al., 1999; Nayler et al., 2000). Those bodies that contain focal sites of transcription are dynamic compartments which first appear at the beginning of S-phase in the cell cycle and disperse during mitosis. YT521-B is known to be a target of non-receptor tyrosine kinases and is also involved in splice site selection. Studies presented in this chapter show that YTH (YT521-B homology) domain, localized in the middle of YT521-B protein, is a novel RNA binding domain and binds to a six nucleotides degenerate motif. Microarray analyses demonstrate that YT521-B predominantly regulates vertebrate-specific exons. Pull-down assay shows that YT521-B can be phosphorylated on tyrosine residue by several non-receptor tyrosine kinases; therefore its intranuclear localization and function are altered. Together the results suggest that YT521-B acts as a bridge that links both signal transduction pathway and pre-mRNA splicing regulation.

4.1. Modular structure of YT521-B

The MyHits Motif Scan program (Pagni et al., 2004) was used in order to inspect the modular structure of YT521-B, which is important for its function study. The amino acid sequence of YT521-B reveals four putative bipartite nuclear localization signals (marked in green); a glutamic acid-rich region of unknown function (E-rich, shaded in blue) at the N-terminus; a putative nucleic acid binding YTH domain, located in the center of the protein (shaded in red); a proline-rich stretch (P-rich, shaded in green) and a glutamic acid/arginine-rich region (ER-rich, shaded in yellow) at the C-terminus. The P-rich region is possibly responsible for binding to SH3 domain containing proteins (Kay et al., 2000), and to WW domain containing proteins (Ilsley et al., 2002). The ER-rich domain is involved in the possible protein:protein interactions (Hartmann et al., 1999; Yanagisawa et al., 2000).

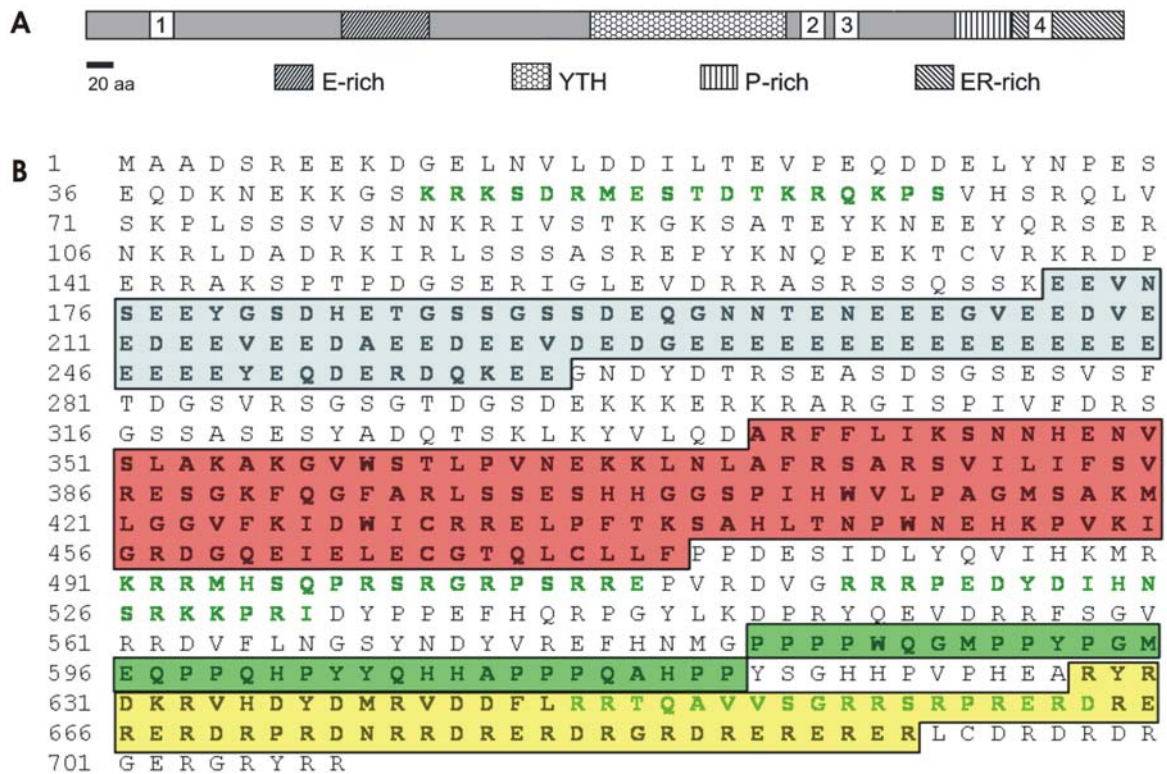


Figure 14: Modular structure of YT521-B. (A) Domain structure of YT521-B. The amino-terminal glutamic acid-rich (E-rich) region, the carboxy-terminal proline-rich (P-rich) region, the glutamic acid/arginine-rich (ER-rich) region and YTH homology (YTH) domain are indicated by different shaded boxes. Boxes with numbers indicate the four nuclear localization signals (NLS). (B) Protein sequence of YT521-B. The four NLS are highlighted in green. The E-rich region, YTH domain, P-rich and ER-rich regions are shown respectively as blue, red, green and yellow shadowed boxes.

4.2. The YTH domain is a novel RNA recognition domain

During BLAST searches for YT521-B homologs, a conserved part between residues 356 and 499 of the rat YT521-B protein was identified. Then a PSI-BLAST (Altschul et al., 1997) search was performed with this part of the YT521-B protein. The inclusion threshold used in the PSI-BLAST search was 0.005. After four iterations, multiple proteins from *Plasmodium falciparum*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, *Mus musculus*, *Homo sapiens*, and *Arabidopsis thaliana*, *Oryza sativa* were found to have a similar region, with E (EXPECT) values in the range 10^{-11} - 10^{-60} . The E value is a statistical significance threshold for reporting matches against database sequences and used to judge the reliability of the scores. Many of these protein sequences were derived from automated gene prediction and were not confirmed by mRNA or EST sequences. To identify sequences of existing proteins, BLAST searches with the conserved region against the translated nonredundant nucleotide database at NCBI (filtered using “biomol_mrna[PROP]” as a keyword) and the translated EST database was performed. The sequences from these searches, aligned to the full length of the query and with E values below 10^{-6} , were aligned using ClustalW (Thompson et al., 1994), after the redundancies had been

removed (Figure 15). Additional BLAST searches against the genome databases confirmed that the conserved region is present exclusively in eukaryotic genomes.

The conserved region which defines a novel domain in these proteins was named the YT521-B homology (YTH) domain (Stoilov et al., 2002b). This domain is usually located in the middle or at the Carboxy-terminus of the protein. The YTH domain shows remarkable conservation across a wide range of species with 14 invariant and 19 highly conserved residues. The proteins present in the alignment do not share significant similarity outside the YTH domain, with the exception of the closely related vertebrate orthologs of YT521-B.

SMART and PFAM databases were searched using SMART sequence analysis for additional known domains in the proteins containing YTH domain. The PHD program (Rost, 1996) was used to determine the putative secondary structure of the YTH domain. The domain is predicted to have a mixed alpha-helix beta-sheet fold, with four alpha-helices and six beta-strands. The conservation pattern follows the predicted secondary structure, with three blocks of conserved sequence separated by loops of variable size. Notable features of the domain are the highly conserved aromatic residues located within the beta-sheet and being reminiscent of the aromatic residues conservation within the RNA recognition motif (RRM) domain. In the RRM domain, conserved aromatic residues located in the beta-sheet are crucial for RNA binding (Hoffman et al., 1991). Based on this observation, the YTH domain is predicted to be an RNA binding domain.

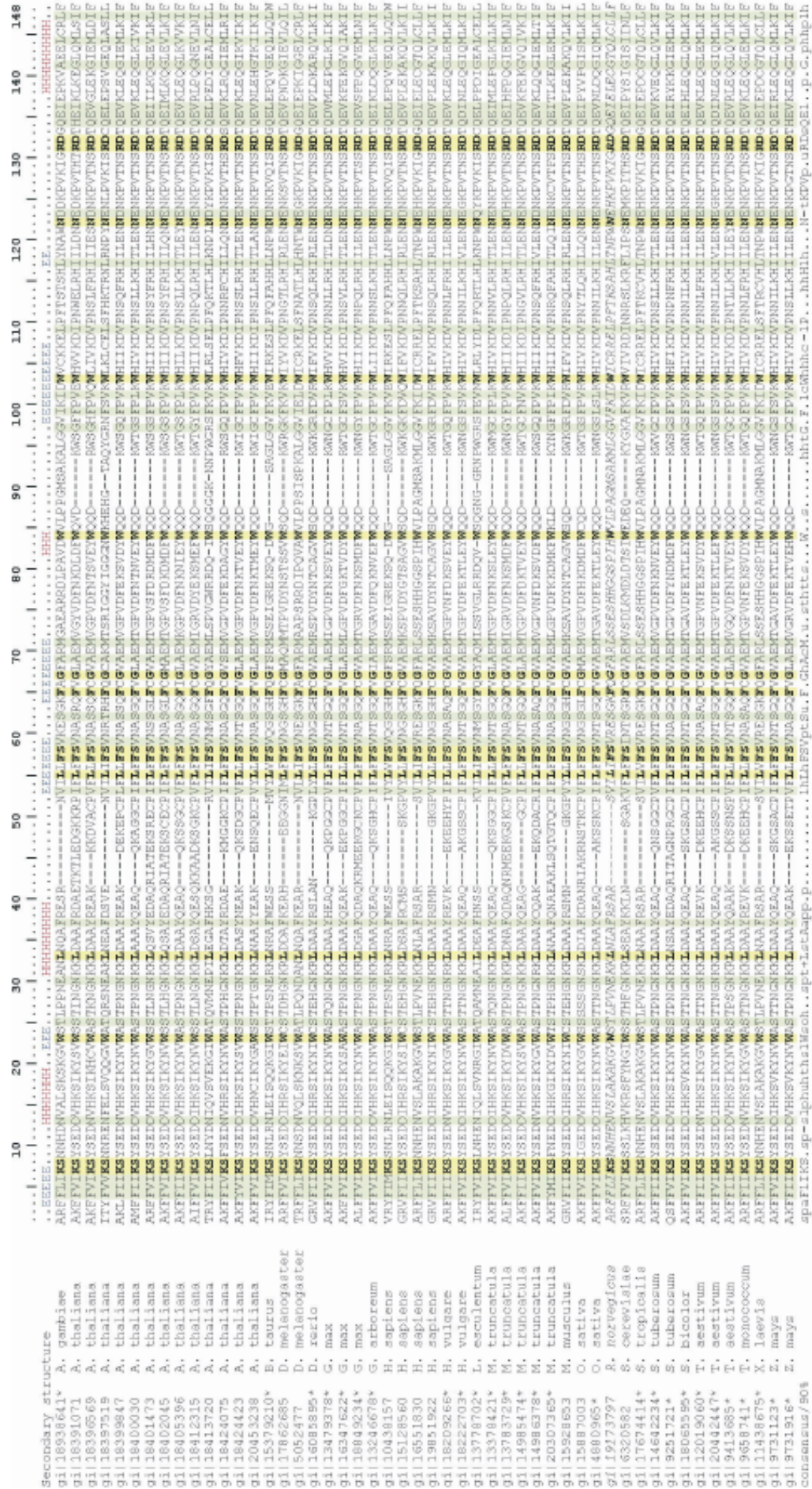


Figure 15: Multiple alignment of sequences containing YTH homology (YTH) domains. The sequences are denoted with the GenBank identifier of the corresponding nucleotide sequence followed by the species name. Translations of partial EST sequences are indicated with an asterisk. The sequences were identified using searches of the translated EST and the non-redundant nucleotide databases at NCBI using rat YT521-B (indicated with an arrow). The similar residues that are present in all sequences are shaded (green) using the PAM250 matrix. The identical residues are shaded in yellow. The secondary structure, as predicted by the PHD program, is shown on top. E denotes extended (β-strands) structure and H denotes the predicted α-helices.

The YTH domain family currently has 174 members which have been identified in various eukaryotic species. The molecularly characterized human YTH domain family members are YTHDF1/DACA1, YTHDF2/HGRG8 and YTHDF3. These three members are highly similar at cDNA and protein sequence level (Figure and table). YTHDF2 mRNA was detected in 16 different human tissues. A (TG)_n microsatellite polymorphism in the fourth intron of the YTHDF2 gene is found to be associated with human longevity (Cardelli et al., 2006), although its molecular function is unclear. The biological function of the remaining 171 YTH-family members that share no common sequence motifs outside the YTH domain is unknown.

```

YT521-B  ARFFLIKSNHENVSLAKAKGVWSTLPVNEKKNLAFRSARS---VILIFSVRESGKFQG 415
DACA1    GRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDSAFRCMSSKGPVYLLFSVNGSGHFCCG 448
HGRG8    GRVFIIKSYSEDDIHRSIKYNIW CSTEHGNKRLDAA YRSMNGKGPVYLLFSVNGSGHFCCG 469
YTHDF3   GRVFIIKSYSEDDIHRSIKYSIWCSTEHGNKRLDAA YRSLNGKGPLYLLFSVNGSGHFCCG 475
.*:*:*** .::: : .:*. : .:***: *:* . : *:*:**. **:* *
YT521-B  FARLSSESHHGSPHWWLPAGMSAKMLGGVFKIDWICRRELPTKSAHLTNPWNEHKPV 475
DACA1    VAEMKSPVDYGTSA GVWSQDK-----WKGRFDVQWIFVKDVPNNQLRHIRLENNDNKPV 502
HGRG8    VAEMKSAVDYNTCAGVWSQDK-----WKGRFDVRWIFVKDVPNSQLRHIRLENNENKPV 523
YTHDF3   VAEMKSVVDYNAYAGVWSQDK-----WKGRFVVKWIFVKDVPNNQLRHIRLENNDNKPV 529
.*:.* .: . * * * .: * * * : * : * : * : * : *
YT521-B  KIGRDGQEIIELEGTQLCLLF----- 495
DACA1    TNSRDTQEVPLEKAKQVLKIISSYKHTTTSIFDDFAHYEKRQEEEEVVRKERQSRNKQ 559
HGRG8    TNSRDTQEVPLEKAKQVLKIIAS YKHTTTSIFDDF SHYEKRQEEEE SVKKERQGRGK- 579
YTHDF3   TNSRDTQEVPLEKAKQVLKIIATFKHTTTSIFDDFAHYEKRQEEEEAMRREERNRKQ- 585
.* ** ** : * : * :

```

Figure 16: Alignment of YTH domains from different human proteins. Asterisk marks the identical residues, colon marks the conserved substitutions, and dot marks the semi-conserved substitutions.

Table 1: Sequence identity among human YTH domains.

| Sequence name | Sequence name | Sequence identity % |
|---------------|---------------|---------------------|
| YT521-B | DACA1 | 32 |
| YT521-B | HGRG8 | 30 |
| YT521-B | YTHDF3 | 29 |
| DACA1 | HGRG8 | 66 |
| DACA1 | YTHDF3 | 67 |
| HGRG8 | YTHDF3 | 66 |

To test whether the YTH domain influences the localization of other YTH domain containing proteins, the intracellular localization of all known human YTH domain family protein members were determined. Equal amounts of EGFP-YT521-B, EGFP-YT521-del (YTH), EGFP-DACA1, EGFP-HGRG8 and EGFP-YTHDF3 were transfected into HEK293 cells. As shown in Figure 17, both wild type YT521-B and YTH domain deletion mutant (YTHdel) are localized in the nucleoplasm and concentrated in nuclear dots. This result suggests that the deletion of the YTH domain did not influence the localization of YT521-B in YT bodies. Furthermore, the three other human YTH domain containing proteins, YTHDF1/DACA1,

YTHDF2/HGRG8, and YTHDF3 are all localized in the cytosol. Together, these data indicate that the YTH domain does not dictate the localization of the YTH domain family proteins.

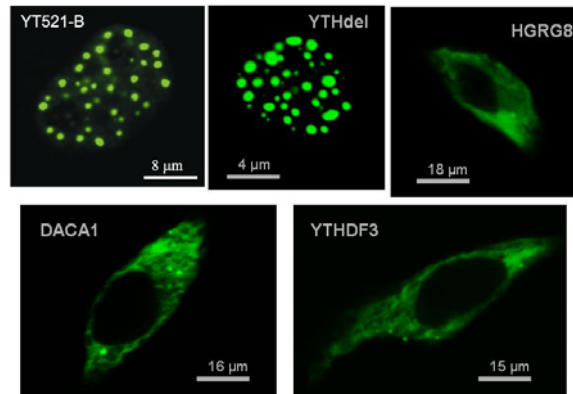


Figure 17: human YTH-domain containing proteins localization. Proteins with carboxy-terminal tagged EGFP were expressed in HEK293 cells. The pictures show representative cells.

The mixed alpha-helix beta-sheet fold structure prediction reveals that YT521-B protein can directly interact with RNA via its YTH domain. To test this hypothesis, several experimental approaches were employed. First the genes coding for full length YT521-B and YT521-B lacking the YTH domain were cloned into the Baculovirus expression vector pFastBacHT A. This vector introduces an N-terminal HIS-tag that permits the efficient purification through the binding to chelating resin Ni-NTA. Full length recombinant protein YT521-B and the protein lacking YTH domain (YTHdel) were expressed in *Drosophila* cells and purified under native condition in the presence of Ni-NTA agarose. To avoid the binding of contaminating proteins to Ni-NTA beads, 15mM of imidazole were present in the lysis buffer and washing buffer. Recombinant proteins were finally eluted from Ni-NTA resin with 400mM imidazole (figure 18).

To identify the sequence requirements of the RNA motif that bind to YT521-B, *in vitro* SELEX experiment were then performed. Systematic evolution of ligands by exponential enrichment (SELEX) is a technique that allows the simultaneous screening of highly diverse pools of different RNA or DNA (dsDNA or ssDNA) molecules for a particular feature. In order to determine the nucleotide feature of YTH domain, the baculovirus generated recombinant proteins and a random 20-mer RNA pool were used. After seven rounds of SELEX, RNA sequences that bind to recombinant YT521-B were selected and cloned into pCR-TOPO vector for sequencing. The results are shown in Figure 19 A. Similar to other RNA binding proteins; YT521-B binds to highly degenerate sequences that can be only described by a weight matrix, which is shown in Figure 19 B. SELEX experiment result shows that YT521-B binds to short RNA motifs that loosely follow a consensus sequence.

Results

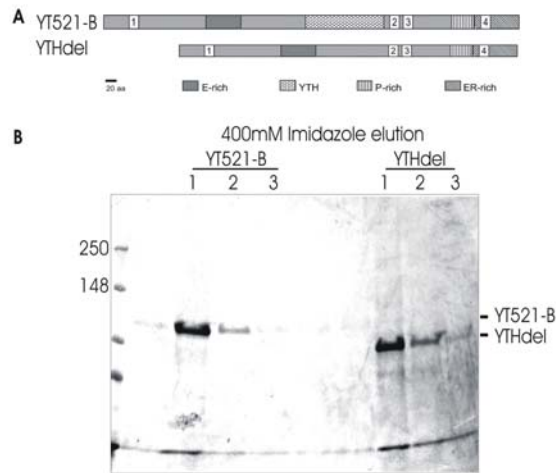


Figure 18: Expression and purification of YT521-B and deletion mutant YTHdel. (A) Domain structure of full length YT521-B and YTH domain deletion mutant YTHdel. Boxes with numbers indicate the four NLS. The E-rich region, P-rich region, ER-rich region and YTH domain are indicated by different shaded boxes. (B) Full length protein YT521-B and YTH domain deletion mutant YTHdel were purified on Ni-NTA agarose beads after the expression in insect cells Sf9. Purified recombinant protein was visualized on polyacrylamide gels by Coomassie blue staining. The localization of YT521-B and YTHdel band is indicated on the right.

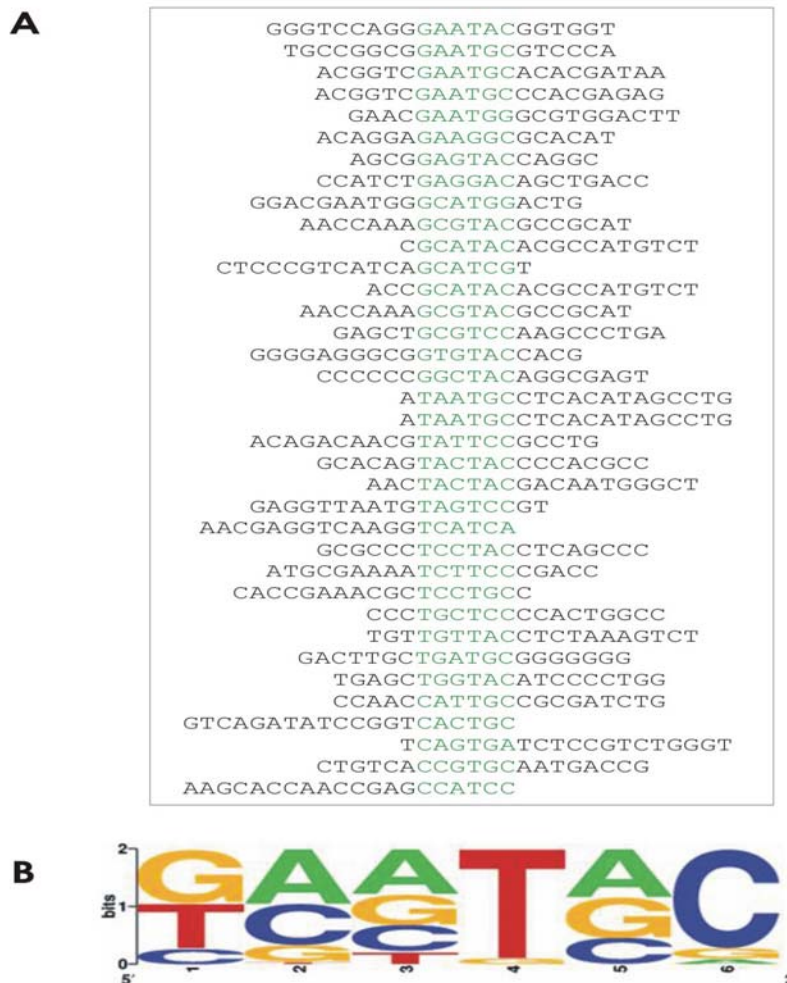


Figure 19: RNA motif binding to YT521-B determined by *in vitro* SELEX. (A) Representative SELEX sequences. The common sequence motif is highlighted in green. (B) Weight matrix, generated by WebLogo program (<http://weblogo.berkeley.edu>) describes the degenerate sequence element that present in all SELEX sequences.

In order to confirm the binding between YT521-B and the RNA motif obtained from SELEX in a different experimental system, RNA electrophoretic mobility shift assay (RNA gel shift) was performed. Two SELEX sequences containing the YTH domain binding motif (probe 1 and 2) were radioactively labeled. One short TA-rich sequence (probe 3), that lacks the YTH domain binding motif, but binds to the STAR protein rSLM-2 *in vitro* (Stoss et al., 2001) was also labeled and used as an experimental control. These probes were incubated *in vitro* with recombinant YT521-B and a change in the probe's mobility was detected using non-denaturing polyacrylamid gel electrophoresis. As a negative control, we used the YTH protein (YTHdel, Figure 20) that lacks the YTH domain, but contains all other protein parts. HeLa nuclear extract was used as a protein source as a positive control. As shown in Figure 20, an accumulation of the signal in the wells of the native polyacrylamid gel was observed when YT521-B was present, but no band shift was observed within the gel. This phenomenon was observed under various experimental conditions. The change in RNA mobility was due to the YTH domain, as the YTHdel protein did not cause any signal accumulation. In contrast, when the TA-rich RNA probe (probe 3) was used in gel shift assay, a band shift was observed when using Hela nuclear extract as a protein source, but no signal accumulation when using YT521-B or YTHdel protein as a protein source (Figure 20). Together these data suggest that binding to RNA causes an aggregation of YT521-B proteins. To test this hypothesis, a fluorimeter that measures the change of protein size in solution (Lomakin et al., 2005) was employed for light scattering experiment. After the protein solution was placed in the sample chamber, the emission light that was scattered by protein particles was detected by the photodetector positioned at 90° angle to the incident beam. As shown in Figure 20, recombinant YT521-B protein solution showed no light scattering, indicating that the protein molecules or possible protein aggregates are smaller than the wavelength of the light source. However, when an RNA probe containing the YT521-B binding motif is added to the solution, the light scattered by protein aggregate was observed (Figure 20 C, left). When a solution of YTHdel protein was used in the experiment with the addition of the same RNA probe, no change could be detected (Figure 20 C, middle). Finally, the TA-rich RNA probe was tested and did not cause any protein accumulation when added to a solution of full length YT521-B protein (Figure 20 C, right). In all cases, the maximum light scattering was observed when the protein was precipitated by TCA addition. These data suggest that RNA binding causes YT521-B protein aggregation *in vitro*.

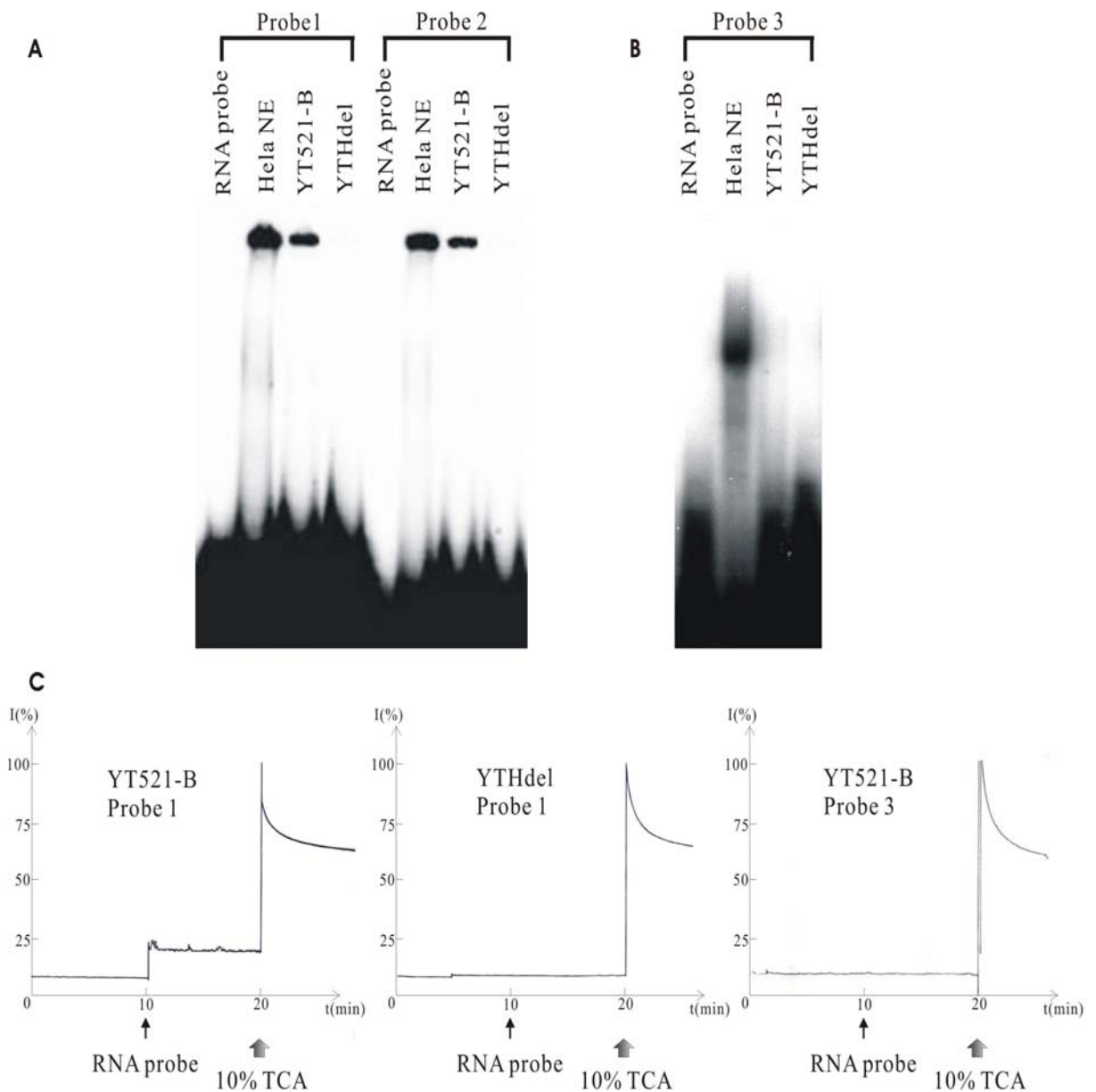


Figure 20: Interaction of YT521-B with RNA. (A) Gel mobility shift analysis of YT521-B using *in vitro* transcribed ^{32}P -labelled RNA probes containing the YTH-binding motif. Recombinant protein YT521-B and YT521-B lacking the YTH domain (YTHdel) were incubated with RNA probes 1 and 2. HeLa nuclear extract (HeLa NE) serving as a positive control. The sequences of RNA were: Probe 1: CGCATACACGCCATGTCT; probe 2: GAGCTGCGTCCAAAGCCCTGA. The match to the consensus YT521-B motif in the probes is underlined. (B) Gel mobility shift analysis of YT521-B using RNA probes without YTH-binding motif. The sequence of probe 3: TAATAAAGAACAACCCCTC lacks the YTH-domain binding motif. (C) Light scattering of the RNA:YT521-B complex. 100 ng/ μl recombinant YT521-B proteins (left, right) were used for light scattering. After pre-incubation at RT for 10 min, 5nM RNA solutions containing either RNA probe 1 (left, middle) or RNA probe 3 lacking the YTH recognition sequence (right) were added to the protein solution. Subsequently, TCA was added to a final concentration of 10% to precipitate all the proteins. The time point for the addition of RNA probes and TCA are indicated by narrow arrow and TCA broad arrow, respectively. The light scattering intensity relative to the intensity of fully TCA-precipitated protein is plotted versus the time.

4.3. Influence of YT521-B binding motifs on alternative splicing *in vivo*

It was previously described that increasing the YT521-B concentration influences alternative splice site selection on a concentration dependant manner (Hartmann et al., 1999). Therefore a reporter minigene was employed to determine whether the direct binding between YT521-B and RNA sequences has an effect on alternative splice site selection *in vivo*. 20-mer sequences which contain either, one, two or three repeats of YTH domain binding motifs were introduced into SXN-minigene (Coulter et al., 1997) as an alternatively spliced exon (Figure 21 A). This reporter minigene system has been widely adopted for analyzing the impact of sequences on pre-mRNA processing. In all experiments, the amount of YT521-B was increased by cotransfecting 0-3 μg of expression constructs with 1 μg of reporter minigene construct. An identical protein expression construct lacking the YTH domain (YTHdel, Figure 21 A) was served as a negative control. As shown in Figure 21 B-D, YT521-B promoted inclusion of the alternative exon in a concentration-dependent manner. In contrast, no effect could be detected when comparable amounts of YT521-B lacking the YTH domain were cotransfected (YTHdel, Figure 21 B-D, F). A negative control reporter SXN-minigene containing no YTH domain binding motif in the alternative exon showed no change in its splicing pattern (Figure 21 E). The total amount of exon inclusion was dependent on the number of YTH binding motifs in the alternative exon. It increased from about 40 to 90% when one or two binding motifs were present, respectively. These data therefore suggest that the YTH binding motif causes dependency of alternative exons on YT521-B concentration.

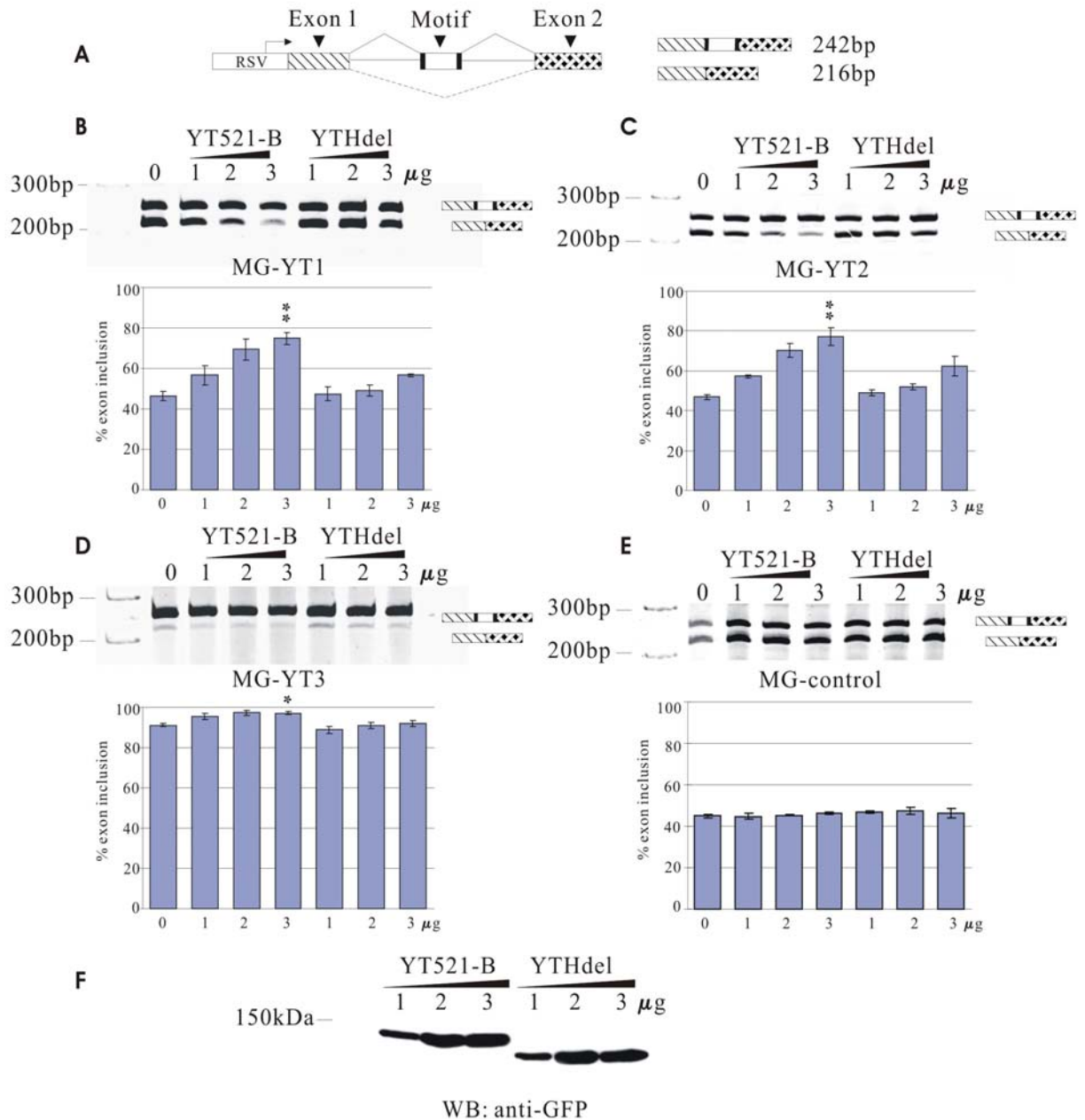


Figure 21: Influence of YT521-B binding motifs on alternative splicing *in vivo*. (A) Structure of the SXN minigene (see section 3.2.15. for detail of minigene construction). The globin exon 1 and 2 (shaded) flank a central artificial exon in which the YT521-B binding sequence was introduced, indicated by “motif”. At least three independent experiments were evaluated using the Student’s t-test. (B) A minigene MG-YT1 containing one YT521-B binding motif sequence, which is underlined (AGAGTCCAGTCTGTCAGTCA), was cotransfected with an increasing amount of YT521-B or YT521-B lacking YTH domain (YTHdel) for *in vivo* splicing assay. The amount of transfected EGFP-YT521-B or EGFP-YT521-B-YTHdel is indicated and normalized using pEGFP-C2. HEK293 cells were used for transient transfection and the RNA was analyzed by RT-PCR. The structure of amplified product is indicated on the right. Stars show statistical significant differences with $P < 0.01$. (C) Cotransfection assay using a minigene MG-YT2 with the sequence that contains two YTH binding motifs (GATGCATGCAATGGATGCGG), $p < 0.01$. (D) Cotransfection assay using a minigene MG-YT3 with the sequence that contains three YTH binding motifs (CGAATCCAGAATCCTGAATCC), $p = 0.02$. (E) Cotransfection assay using a minigene containing a control sequence lacking the YTH binding motif (GGCGATAATGTGTAAATGCC). (F) The increase of EGFP-YT521-B and EGFP-YT521-B-YTHdel in the transfection assay was detected by Western blot detecting using antibody against EGFP.

4.4. Effect of decreasing YT521-B concentration on splice site selection

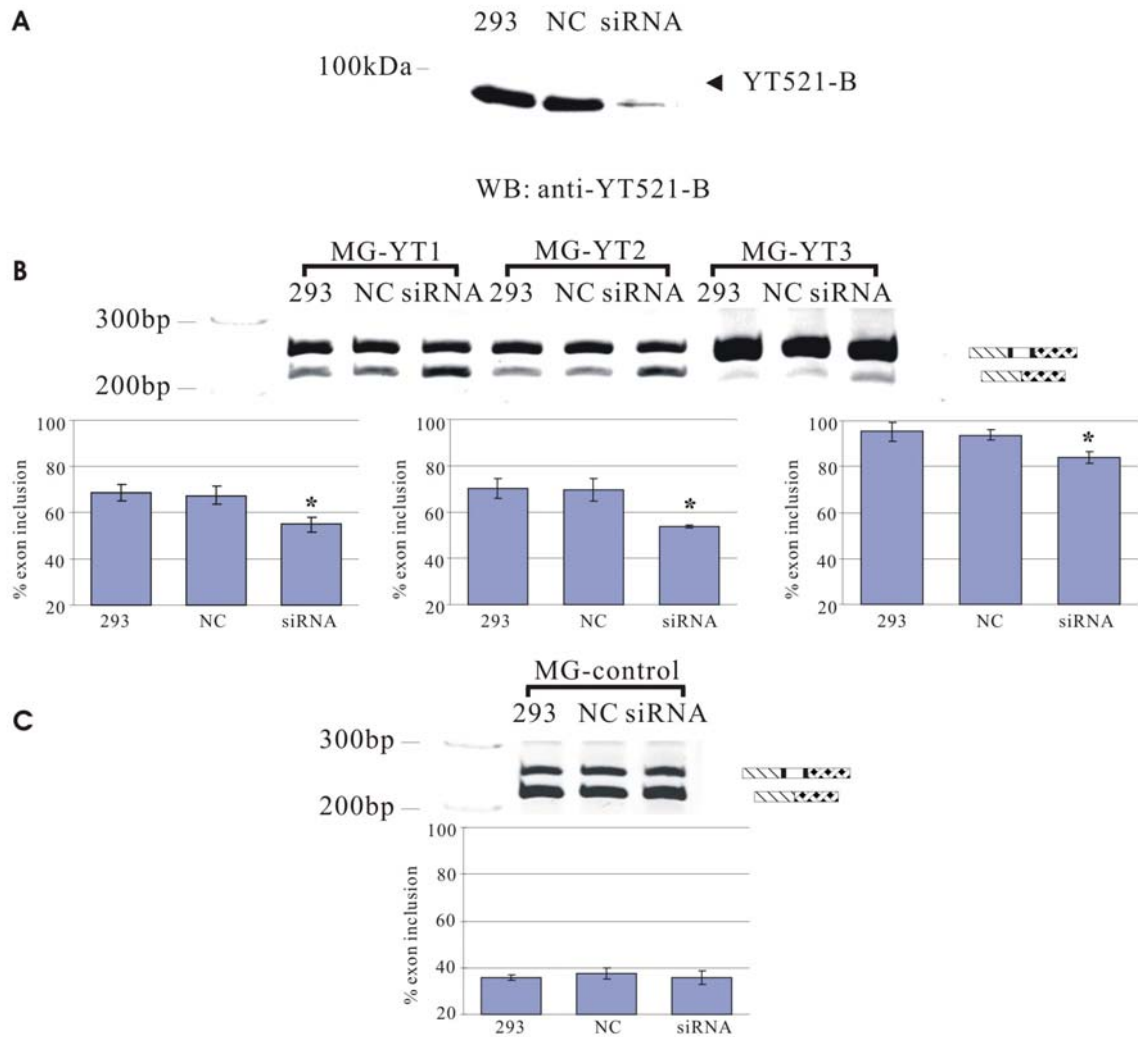


Figure 22: Influence of decreasing YT521-B concentration by siRNA on splice site selection. (A) 42 hours after siRNA transfection, cellular lysates were analyzed by Western blot. The detection was with an antiserum against YT521-B. 293: HEK293 cellular lysates without siRNA treatment; NC: HEK293 cells were transfected with negative control siRNA; siRNA: HEK293 cells were transfected with siRNA against YT521-B. (B) 1 μ g of the minigenes containing 1, 2, or 3 YT521-B binding motifs were cotransfected with the indicated siRNAs in HEK293 cells. C: control, untreated HEK293 cells, NC: siRNA against pBluescript, siRNA: removal of YT521-B by siRNA. 42 hours after siRNA treatment, the RNA was isolated and analyzed by RT-PCR. Representative ethidium bromide stained polyacrylamide gels showing the effect of YT521-B reduction by siRNA on the MG-YT1, MG-YT2, and MG-YT3 minigenes. The statistical evaluation of three independent experiments is shown underneath. (C) 1 μ g control minigene, which lacks YT521-B binding motif, was cotransfected with indicated siRNAs. 42 hours after siRNA treatment, the RNA was isolated and analyzed by RT-PCR. The statistical evaluation is shown underneath the gel.

Small interfering RNA (siRNA) has been widely adopted as a powerful tool for selectively interfering with gene expression (McManus and Sharp, 2002). Since the endogenous YT521-B protein is abundantly expressed in HEK293 cells, siRNA was used to reduce the gene expression level of YT521-B. A synthetic siRNA duplex of 21-nt, paired in a manner to have a 3' 2-nt AA overhang, was cotransfected into HEK293 cells with the same reporter minigene

system employed previously. As shown in Figure 22 A, the endogenous YT521-B was depleted by siRNA. The effect of YT521-B depletion on splice site selection of reporter minigenes containing one, two or three YT521-B binding motifs was detected by RT-PCR. Alternative exon skipping caused by YT521-B depletion in the reporter minigenes was observed (Figure 22), when the YTH domain binding motif was present. In contrast, the minigene with no YTH domain binding motif in its alternative exon showed no splicing pattern changes (Figure 22 C). These data together showed that the interaction between YT521-B and pre-mRNA is necessary to influence alternative splice site selection in an YT521-B concentration-dependent manner.

4.5. Three of the conserved residues in the YTH domain are critical for splice site selection

It is well known that conserved aromatic residues localized in the beta1 and beta3 strands of RRM domain provide staking interaction with RNA (Maris et al., 2005). A remarkable feature of the YTH domain is the full evolutionary conservation of 14 residues between its members. Furthermore, the conserved aromatic residues located within the beta-strand are similar to the RRM domain (Stoilov et al., 2002b). Mutagenesis is a useful tool that provides more information of molecular contact between proteins and their target RNA. Therefore this method was used to investigate the function of the conserved residues located inside the YTH domain. All the 14 invariant residues in the YTH domain were mutated (Figure 23 A). To study their role in splice site selection, the EGFP-tagged mutants were cotransfected with the reporter minigene containing two YTH domain binding motifs (Figure 23 B). 18-20 hours after transfection proteins and RNAs were isolated. Proteins were analyzed by Western blot to check the protein expression level and the result (Figure 23 C) shows that all mutants expressed similar amount of protein when normalized with endogenous YT521-B, indicating that any activity change is due to the mutation of conserved residue. RNAs were analyzed by RT-PCR and the result (Figure 23 B) shows that changing the conserved residues W380, F412 and G414 abolished the effect on splice site selection. In order to confirm that these three residues are crucial for splice site selection, a wider mutant protein concentration range was used. 1-3 μ g of EGFP-tagged W380, F412 and G414 mutants were cotransfection with 1 μ g of the same reporter minigene employed in the precedent experiment. The Western blot shows the increasing amount of mutant proteins (Figure 23), and RT-PCR result shows even at the highest concentration, none of the mutants was able to alter splice site selection (Figure 23). As shown in the section 4.3 (Figure 21), the YTH domain is indispensable for the regulation of alternative splice site selection. Therefore YT521-B exerts its effect by direct binding to the pre-mRNA and not by sequestering other splicing factors. Together the mutagenesis analyses show that three residues W380, F412, G414 located in the beta strand of YTH domain are necessary for the ability of the protein to modulate splice site selection.

Results

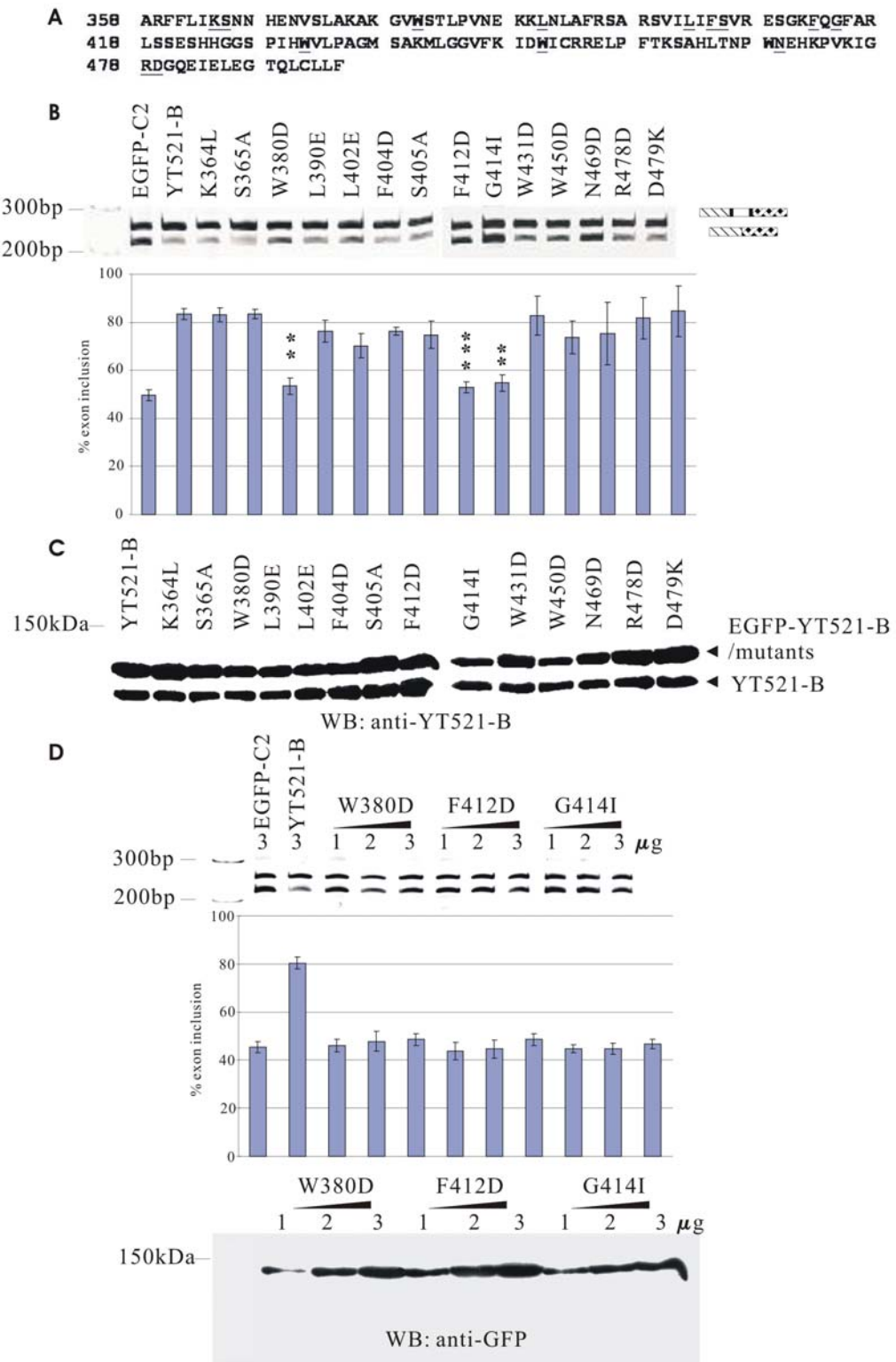


Figure 23: Mutational analysis of the YTH domain. (A) Sequence of the YTH domain found in YT521-B. The phylogenetically conserved residues are underlined. (B) The introduced mutation within YTH domain influences splice site selection of MG-YT2 minigene. HEK293 cells were transiently transfected with YT521-B or YTH domain mutations that are indicated by the amino acids changed. P values are 0.0092 for W380D, 0.0038 for F412D, and 0.007 for G414I. (C) Crude lysates of transfected HEK293 cells were analyzed for YTH domain mutations expression by Western blot, using an antiserum against YT521-B (D) Analysis of three YTH domain mutations most severely affecting the splice site selection of MG-YT2 minigene. 1 to 3 μ g of expression plasmids for each construct were cotransfected with the 1 μ g of reporter minigene and pEGFP-C2 plasmid was used for normalization. RNA was isolated 18 hours later and analyzed by RT-PCR. The statistical evaluation of three independent experiments is shown underneath. (E) Western blot shows the increase of expression of each mutant, detected by anti-EGFP antibody.

4.6. Identification of endogenous target exons of YT521-B by microarray

So far, the study of influence of YT521-B on splice site selection was only done with model constructs. In order to determine the effect of YT521-B on splice site selection of endogenous targets the Affymetrix GeneChip Mouse Exon 1.0 ST microarray was used. This new exon-centric array allows genome-wide identification of differential splicing variation and concurrently provides a flexible and inclusive analysis of gene expression (Gardina et al., 2006). The splicing patterns of all known or predicted mouse genes were compared in the presence or absence of overexpressed EGFP-YT521-B. The mRNAs isolated from transfected or non-transfected neuro 2A cells were labeled and hybridized to the GeneChip. The chip result was analyzed by commercial software and validated by RT-PCR. The construct YT521-B lacking the YTH domain (YTHdel) was used as a control in order to show that the influence of YT521-B on splice site selection is due to direct interaction between YTH domain and target sequence. High-scoring events which have a fold change over 1.5 when comparing YT521-B overexpression group and control group were verified by RT-PCR using primers located in the constitutive exons that flank the alternative one. Several regulated events were detected by GeneChip analysis. As shown in Figure 24, exons of *Anub11*, *Cxcl10*, *Rhot2*, *Sec24c* and *Zfp687* were representatively verified by semi-quantitative RT-PCR. In these events, only YT521-B, but not the YTH domain deletion mutant YTHdel, effects splice site selection. Since those exons were regulated by YT521-B, the presence of YTH domain motif was subsequently inspected. The RNA analyzer program was used to scan for high-score (>3.04) YTH binding motif sequence in the regulated exons and their surrounding sequences. The presence of YTH-binding site clusters was observed in all the regulated events but not in the *GSK3B* sequence, which was not influenced by YT521-B. An interesting aspect of the regulated genes is that similar to YT521-B they are all vertebrate specific. *Anub11*, *Cxcl10* and *Zfp687* genes are only expressed in vertebrates. In contrast, *Rhot2* and *Sec24c* are found in all phyla, but the YT521-B-dependent exon is specific for vertebrates. These data suggest that YT521-B regulates vertebrate-specific RNAs that it binds to via its YTH domain.

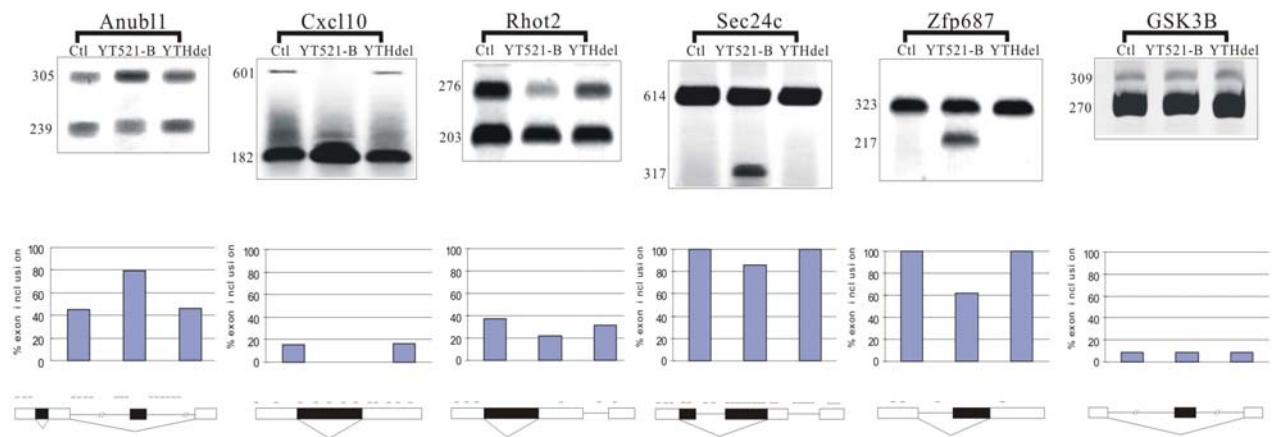


Figure 24: Endogenous target genes of YT521-B. YT521-B-dependent exons were identified by Mouse Exon 1.0 ST Array (Affymetrix) analysis. High-scoring events were analyzed by RT-PCR using primers in the flanking exons. YT521-B: overexpression of EGFP-YT521-B, YTHdel: overexpression of EGFP-YT521-B-YTH del. The picture on the top shows a representative RT-PCR analysis, the graph in the middle shows the statistical evaluation of each experiment and the cartoon on the bottom shows the localization of YTH binding signatures on top of the annotated transcript structure. The processing of GSK3B was shown to be YT521-B independent both by Exon Array analysis and by RT-PCR.

4.7. YT521-B is phosphorylated on tyrosine residue

As shown in Figure 14, the nuclear protein YT521-B is composed of several domains: the RNA binding domain YTH, a proline rich domain containing putative SH3 domain binding sites and a glutamic acid/arginine-rich region involved in protein–protein interaction. The properties of YT521-B make it a good candidate for linking signal-transduction pathways and mRNA splicing. Previous studies showed that the Src family kinase p59^{fyn} tyrosine phosphorylated both YT521-B and a signal transduction and activation of RNA (STAR) family member Sam68, and subsequently regulated their interaction (Hartmann et al., 1999).

Since YT521-B is detected only in the nucleus the question whether other non-receptor tyrosine kinases whose expression is often not only confined to the cell membrane can cause its phosphorylation was addressed. The sequencing of the human genome demonstrated that humans possess 90 unique tyrosine kinases. Thirty-two of them are non-receptor type kinases which can be subdivided into 10 subfamilies (Robinson et al., 2000). In unstimulated cells, the activity of the endogenous kinases can only be detected using phosphorylation-specific antibodies against specific target sites. To test the result of kinase activation, their activity was increased by transfecting cDNAs expressing the kinases. A member of each subfamily was cotransfected with EGFP-YT521-B. After immunoprecipitation, tyrosine phosphorylation of YT521-B was detected by western blot using the phosphotyrosine specific antibody Py20. As shown in Figure 25 A, the kinases c-Abl, Rlk, p59^{fyn} and c-Src, as members of the ABL, TEC and SRC families, phosphorylate YT521-B, whereas the related kinases Csk, FerH, Sik, Lyn and Syk, representing the CSK, FES, FRK and SYK families, had no effect. Reblotting (Figure 25 B)

of the immunoprecipitates demonstrated the presence of YT521-B in all experiments, suggesting a specific action of c-Abl, Rlk, p59fyn and c-Src. The kinases that did not phosphorylate YT521-B were active, as demonstrated by their ability to phosphorylate unknown proteins in cellular lysates (Figure 25 C). Together, the data show that YT521-B is phosphorylated by specific nonreceptor tyrosine kinase.

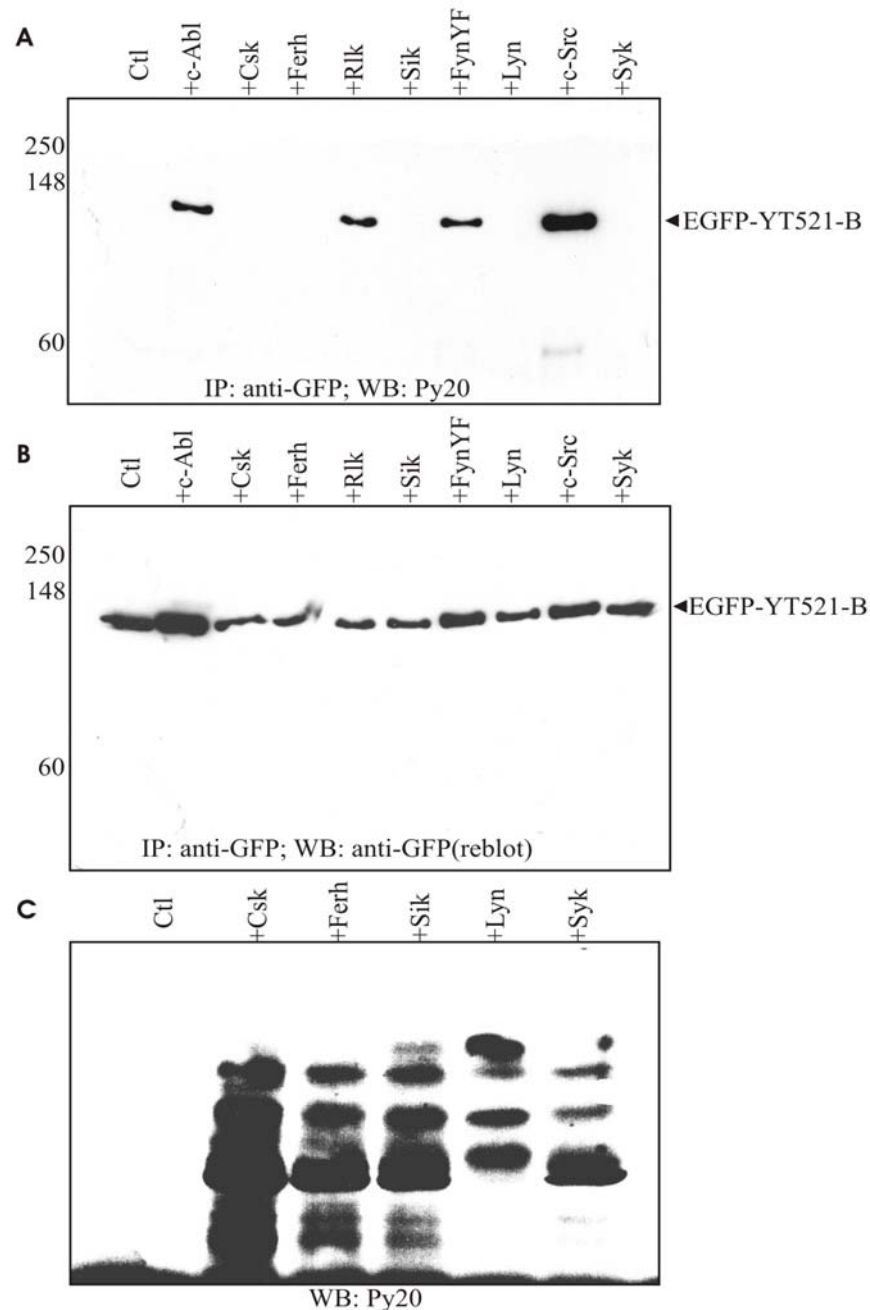


Figure 25: Phosphorylation of YT521-B by several non-receptor tyrosine kinases. (A) EGFP-YT521-B was cotransfected with indicated non-receptor tyrosine kinases in HEK293 cells. For immunoprecipitation (IP), GFP-tagged YT521-B protein was precipitated with anti-GFP antibody and phosphorylated tyrosine residues were detected using anti-phosphotyrosine antibody Py20. The reblot in (B) was analyzed for EGFP-YT521-B expression using anti-GFP antibody. (C) Crude lysates of HEK293 cells transfected with tyrosine kinases that do not phosphorylate YT521-B were analyzed for tyrosine phosphorylation with Py20 antibody. The Western blot shows the activity of the kinases employed. The position of EGFP-YT521-B is indicated on the right and the molecular weight is given in kilodaltons on the left.

4.8. YT521-B localizes with c-Abl kinase in the nucleus

Tyrosine kinase c-Abl was coimmunoprecipitated with YT521-B (Rafalska et al., 2004). Therefore it was subsequently tested whether the two proteins colocalize in the cell. Previous study showed YT521-B is present in a distinct nuclear substructure, termed the YT bodies (Nayler et al., 2000). These YT bodies can be detected using an antiserum that was previously developed (Nayer et al., 2000). Without the presence of overexpressed c-Abl kinase, endogenous YT521-B is present in 20–30 nuclear dots (Figure 26 A). After c-Abl transfection, these dots dissolve and colocalize with c-Abl in the nucleus (Figure 26 C and D). As expected, c-Abl is also present in the cytosol (Figure 26 B) (Taagepera et al., 1998). However, YT521-B is completely absent in the cytosol, demonstrating that overexpression of c-Abl does not induce a translocation of YT521-B into the cytosol (Figure 26 D). Together with the coimmunoprecipitation (Figure 25 A), these data suggest that YT521-B is a target of c-Abl in the nucleus.

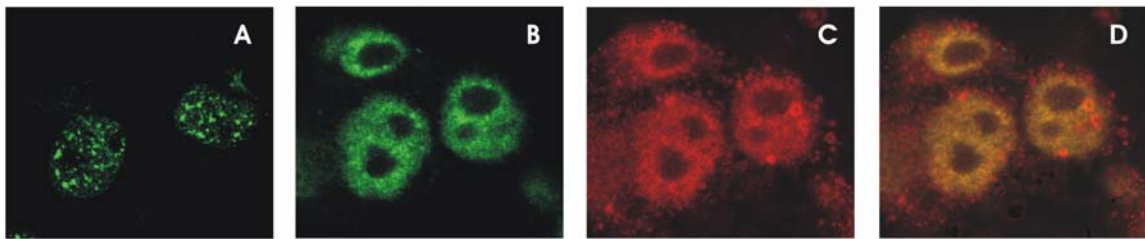


Figure 26: Endogenous YT521-B colocalized with c-Abl in the nucleus. Endogenous YT521-B is stained in green and c-Abl is shown in red (A) Endogenous YT521-B that forms characteristic YT bodies was detected in Cos-7 cells nucleus. (B) Cos-7 cells were transfected with c-Abl. In the presence of c-Abl kinase, YT bodies are dispersed and YT521-B signal is spread throughout the nucleoplasm. (C) Transfected c-Abl signal can be detected in both nucleus and cytosol. (D) Superimposition of images in B and C that shows co-localization between endogenous YT521-B and c-Abl.

4.9. YT521-B binds to membrane bound tyrosine kinase c-Src

The c-Src kinase which phosphorylates YT521-B on tyrosine residue is ubiquitously expressed and located at the plasma membrane. The coimmunoprecipitation experiment showed that in the presence of c-Src kinase, two signals were detected using phosphotyrosine specific antibody Py20 (Figure 25 A). The upper band corresponding to tyrosine phosphorylated EGFP-YT521-B was verified by reblotting with anti-GFP antibody. The lower band, with a molecular weight around 60 kDa, might be the autophosphorylated c-Src kinase that bound to YT521-B. To test whether there is a direct molecular interaction between YT521-B and membrane bound tyrosine kinase, EGFP-YT521-B was expressed in HEK293 cells in the presence of c-Src and immunoprecipitated using its GFP-tag (Figure 27, left). As in previous experiment, in the presence of c-Src, two bands are tyrosine phosphorylated (Figure 27, middle). The comparison with the western blot probed against YT521-B demonstrates that the upper band corresponds to

EGFP-YT521-B (Figure 27, left). The reblot using an antibody against c-Src shows that the lower band represents phosphorylated c-Src (Figure 27, right). Together these experiments demonstrated that YT521-B can bind to c-Src in immunoprecipitation under *in vivo* condition.

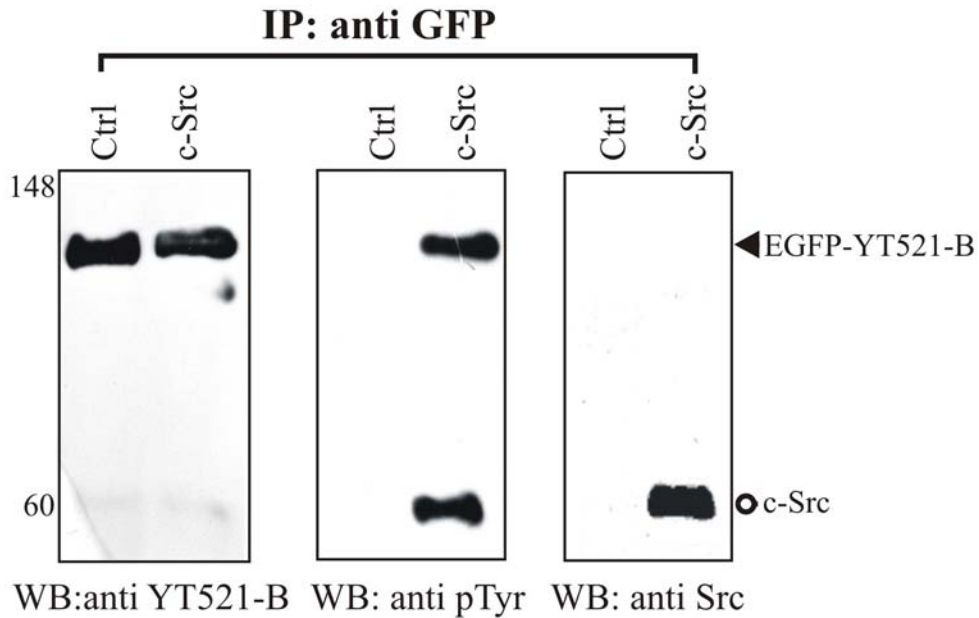


Figure 27: YT521-B binds to non-receptor tyrosine kinase c-Src. EGFP-YT521-B was co-expressed with or without a c-Src expression construct in HEK293 cells. Protein was immunoprecipitated with anti-GFP antibody. Immunoprecipitates were analyzed with antisera against YT521-B (left), phosphotyrosine (middle), and Src (right). The bands of EGFP-YT521-B and c-Src are indicated on the right.

4.10. YT521-B shuttles between nucleus and cytosol

YT521-B was phosphorylated by the membrane bound kinases c-Src (Figure 25) and p59^{fyn} (Hartmann et al, 1999), to which it binds in immunoprecipitation (Figure 25) (Hartmann et al., 1999). Since YT521-B is detected only in the nucleus of cells, a phosphorylation by membrane bound kinases would only be possible during mitosis when the nuclear structure disintegrates or would require shuttling of YT521-B between nucleus and cytosol that has been reported for several proteins implicated in splice site selection (Caceres et al., 1998). To test these possibilities, cell fusion assay was employed to assess the capability of EGFP-YT521-B to shuttle between nucleus and cytoplasm. The assay monitors accumulation of a fluorescent protein in acceptor nuclei of a newly formed polykaryon (Lee et al., 1999; Neumann et al., 2001). This experiment was performed in collaboration with Dr. Ruth Brack-Werner. Transfected HeLa cells expressing EGFP-YT521-B were fused with an excess of untransfected HeLa cells and accumulation of EGFP-YT521-B in acceptor nuclei monitored by time-lapse imaging. The donor cells were marked with red fluorescent protein to allow distinction from acceptor cells. As shown in Figure 28, EGFP-YT521-B fluorescence is visible in the acceptor nuclei 28 min after cell fusion and increases by 308 min. Concomitantly, fluorescence intensity of the donor nucleus

decreases. No shuttling is observed for GFP-tagged B23 (Neumann et al., 2001) or histone 2B (data not shown) within 360 and 720 min, respectively, in this assay. These results indicate that YT521-B shuttles between nucleus and cytosol, where it could interact with membrane bound src-family kinases.

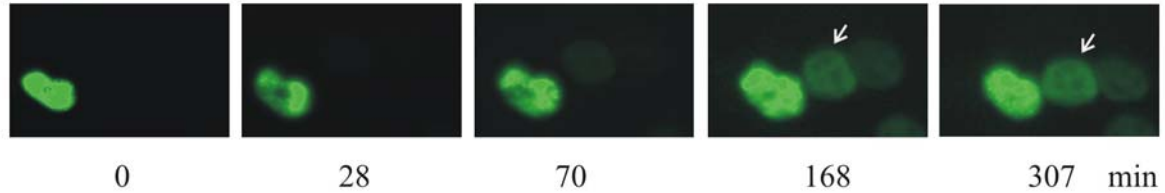


Figure 28: YT521-B shuttles between nucleus and cytosol. HeLa cells transfected with EGFP-YT521-B were fused to an excess of untransfected cells. EGFP fluorescence at the time points in minutes after fusion indicated at the bottom. An arrow indicates acceptor nuclei showing accumulation of EGFP-YT521-B.

4.11. Tyrosine phosphorylation regulates YT521-B solubility

In the nucleus, YT521-B is present in two biochemically distinguishable compartments (Nayler et al., 2000). The first compartment is YT bodies, which are soluble under non-denaturing conditions in Triton X-100 containing buffers. The remaining YT521-B protein resides diffusely in the nucleoplasm and is insoluble in Triton X-100 containing buffers. Previous study demonstrated that YT521-B can translocate during the cell cycle or after actinomycin D treatment between these two compartments (Nayler et al., 2000). High concentrations of actinomycin D (50 mg/ml) that block all three RNA polymerases cause complete dispersion of YT bodies and the accumulation of YT521-B in the Triton X-100 insoluble nuclear fraction (Nayler et al., 2000). Therefore whether tyrosine phosphorylation of EGFP-YT521-B would also result in the accumulation of the protein in an insoluble nuclear fraction was tested. EGFP-YT521-B was transfected into HEK293 cells and analyzed its solubility in a 1% Triton X-100-based cell lysis buffer (HNTG). As shown in Figure 29, in the presence of c-Abl, c-Src or p59^{fyn}, EGFP-YT521-B was present in the Triton insoluble pellet (P) fraction (Figure 29 A). In contrast, when we used a RIPA/benzonase buffer, the protein was present in the soluble (S) fraction (Figure 29 B), demonstrating that the protein is not covalently bound. This effect is dependent on YT521-B phosphorylation, as the presence of Csk that does not phosphorylate YT521-B, but does phosphorylate other proteins, had no effect (Figure 29 E and F). Next, we determined whether endogenous YT521-B showed a similar response to tyrosine phosphorylation and analyzed its solubility in the presence of c-Abl, p59^{fyn} and c-Src. As shown in Figure 29 C, phosphorylation by these kinases resulted in YT521-B accumulation in a Triton X-100 insoluble compartment. Again, the phosphorylated protein was soluble in RIPA/benzonase containing buffer (Figure 29 D). As shown in Figure 29 G and H, the presence

of Csk had no effect on endogenous YT521-B solubility. These data suggest that the tyrosine phosphorylation of YT521-B causes its association with insoluble nuclear structures.

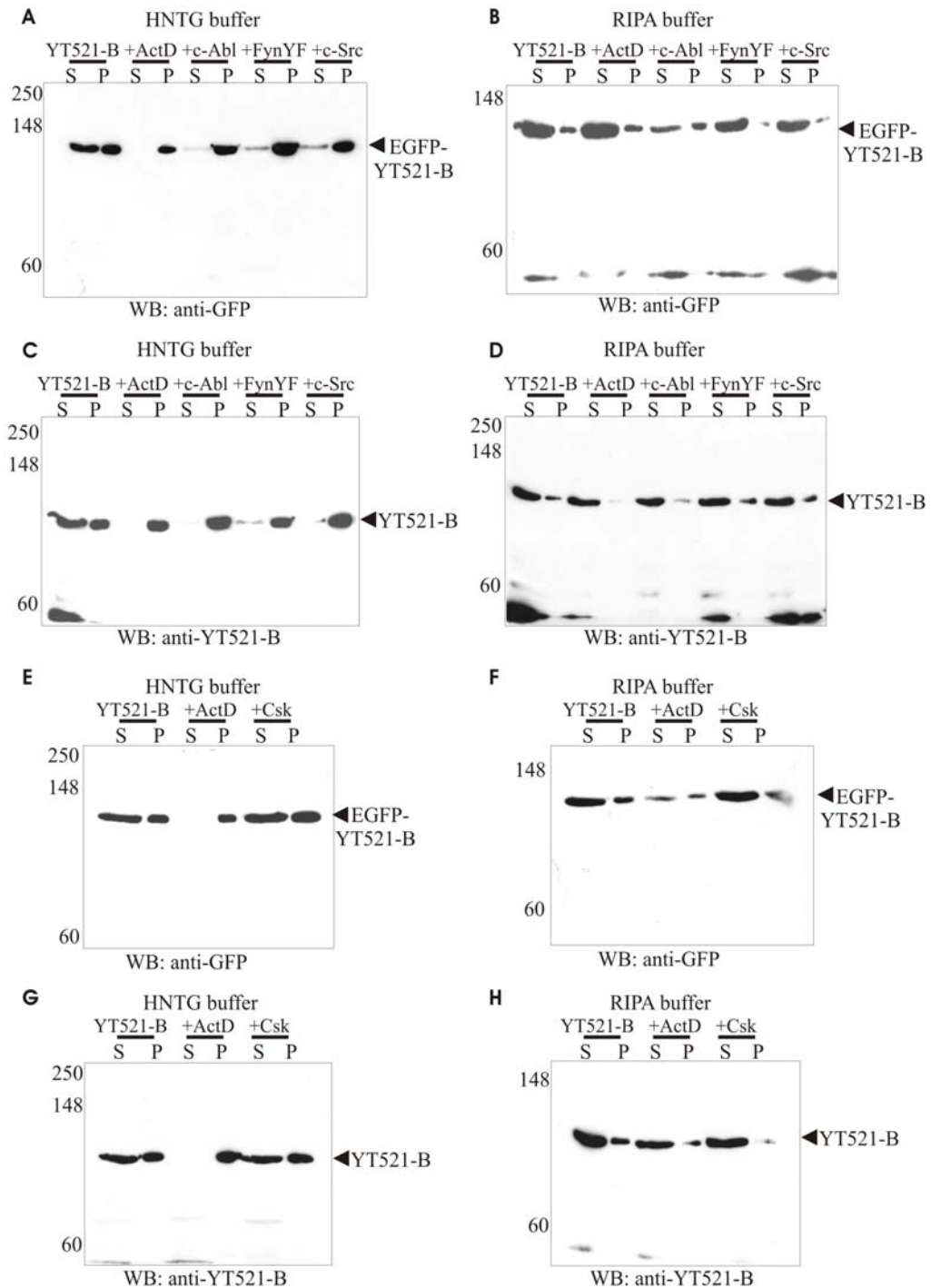


Figure 29: Phosphorylation of YT521-B influences its accumulation in a Triton X-100 insoluble nuclear compartment. The solubility of endogenous YT521-B was determined in the presence of actinomycin D, c-Abl, Fyn, c-Src, and Csk in HNTG (**A and E**) and RIPA (**B and F**) buffers. HEK293 cells were transfected with c-Abl, p59fyn, c-Src and Csk tyrosine kinases or treated with 50 μ g/ml actinomycin D for 3 h. Cells were lysed in HNTG buffer or RIPA buffer containing benzonase. Equal amounts of the soluble fraction (S) or redissolved pellet fraction (P) were analyzed in 10% SDS-PAGE gels. Proteins were analyzed by western blot and ECL using the antisera anti-YT521-B. The solubility of overexpressed EGFP-YT521-B was determined in the presence of actinomycin D, c-Abl, Fyn, c-Src and Csk in HNTG (**C and G**) and RIPA (**D and H**) buffers. EGFP-YT521-B was co-expressed with p59fyn, c-Abl, c-Src and Csk tyrosine kinases or EGFP-YT521-B transfected cells were treated with 50 μ g/ml actinomycin D for 3 h. Cells were lysed in HNTG buffer or RIPA buffer containing benzonase. Lysates were analyzed by western blot by anti GFP antibody.

4.12. Tyrosine phosphorylation changes the ability of YT521-B to influence SRp20 alternative pre-mRNA splicing

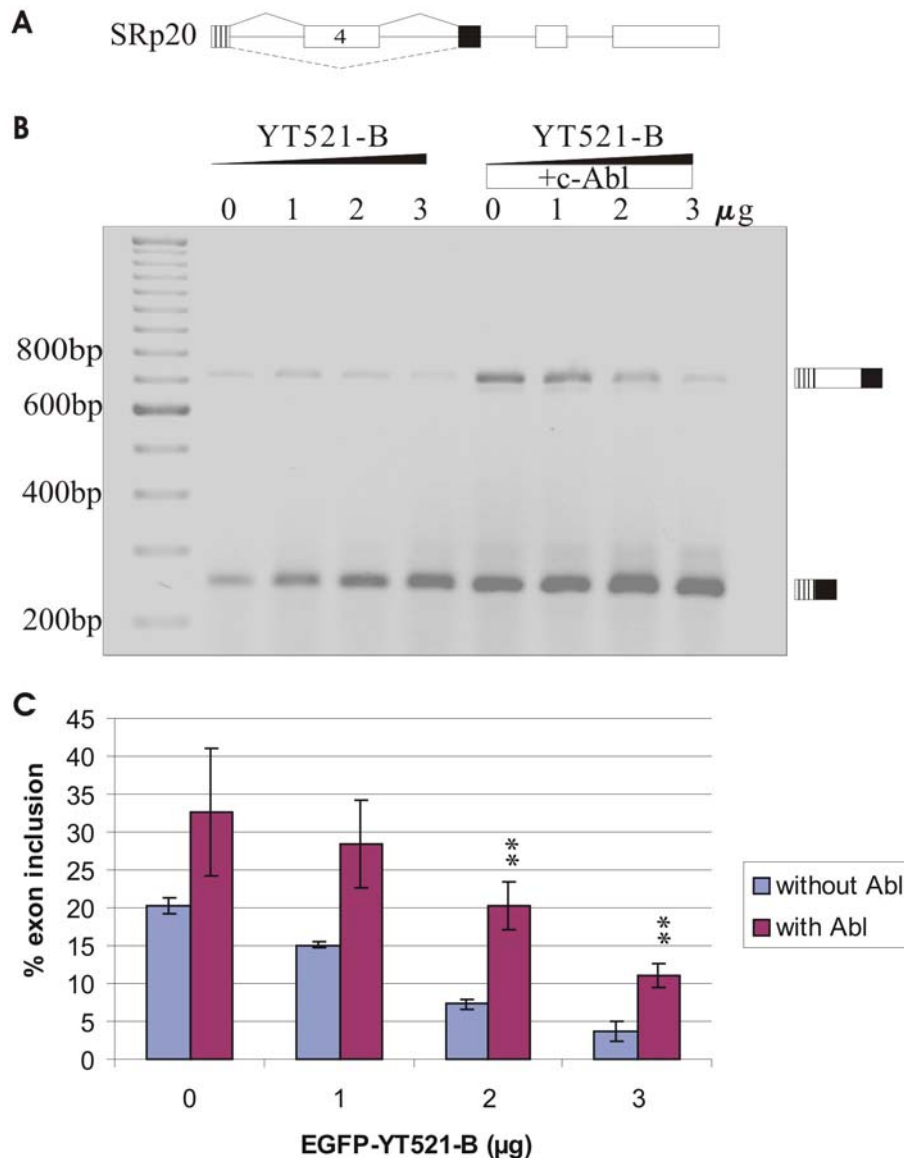


Figure 30: Phosphorylation of YT521-B changes splice site selection of SRp20, exon 4. (A) Structure of the SRp20 minigene used (Jumaa et al., 1997). Exons are shown as boxes, introns as lines, and the alternative exon 4 is indicated with number. The splicing patterns are shown by different line styles. (B) Influence of c-Abl on YT521-B mediated changes in SRp20 exon 4 splice site selection. HEK293 cells were transiently transfected with increasing amounts of EGFP-YT521-B in the presence of the SRp20 minigene (lanes 1–4). The amount of transfected EGFP-YT521-B is indicated and normalized using pEGFP-C2. An additional 0.5 μg EGFP-C2 was present in lanes 1–4 to allow comparison with the c-Abl cotransfections in lanes 5–8. In lanes 5–8, 0.5 μg of c-Abl expression plasmid was cotransfected with an increasing amount of EGFP-YT521-B. The RNA was analyzed by RT-PCR. The structure of the amplified products is indicated on the right. (C) Statistical evaluation of the RT-PCR results. The ratio between the signal corresponding to exon inclusion and all products was determined from at least three different independent experiments. Stars indicate statistical significant differences ($P=0.0007$ and $P=0.001$, for 2 and 3 μg , respectively).

To test whether the dependency of YT521-B function on its phosphorylation status is present in other systems, SRp20 minigene pXB MG that is regulated by YT521-B (Hartmann et al., 1999) was tested (Figure 30). As shown in Figure 30 B, an increase of YT521-B

concentration reduces inclusion of the alternative exon 4 in a concentration dependent manner. This dependency is partially abolished by c-abl induced phosphorylation (Figure 30 B). In the presence of c-abl, exon 4 inclusion is increased, even when no additional YT521-B is transfected (Figure 30 B). It is possible that c-abl acts here on endogenous YT521-B. The presence of c-abl reduces the ability of YT521-B to decrease exon 4 usage 2-to 4-fold (Figure 30 B and C). Since the exon 4 of SRp20 was regulated by YT521-B, the presence of YTH domain binding motif was subsequently inspected in the alternative exon 4 and surrounding sequence. High-score (>3.04) YTH binding motif clusters were found in those sequence using RNA analyzer program. These results show that YT521-B regulates RNAs by direct binding via its YTH domain and tyrosine phosphorylation modulates the influence of YT521-B on splice site selection of SRp20 pre-mRNAs.

4.13. Interaction with Emerin influences YT-521-B on splice site selection of CD44 minigene

Emerin is a nuclear membrane protein that interacts with lamin A/C at the nuclear envelope (Manilal et al., 1996). Mutations in emerin can cause X-linked Emery–Dreifuss muscular dystrophy (EDMD) (Bione et al., 1995). The functions of emerin are poorly understood, but emerin has several binding partners. YT521-B was identified as an interactor of emerin from a high-stringency yeast two-hybrid screen of a human heart cDNA library, with a full length emerin as bait (Wilkinson et al., 2003). Specific binding between emerin and the functional C-terminal domain of YT521-B was confirmed by pull-down assays (Wilkinson et al., 2003). YT521-B has been shown to reduce inclusion of the alternative exon v5 of CD44 minigene in a concentration dependent manner (Rafalska et al., 2004). To test whether emerin can influence YT521-B on splice site selection, YT521-B was cotransfected with emerin in HEK293 cells in the presence of CD44 reporter minigene. Cells transfected with 1 μ g YT521-B had inclusion of exon v5 reduced significantly ($P = 0.006$; t-test) from 34% to 17% (Figure 31 B, lane 1-2). In the presence of an increase amount of emerin, the effect of YT521-B on splice site selection of exon v5 was partially reversed from 17% to 24% (Figure 31 B, lane 3-6). The Western blot shows the increasing amount of mutant emerin (Figure 31 C). These data suggest that emerin mediates the influence of YT521-B dependent splice site selection by its binding to the C-terminal of YT521-B.

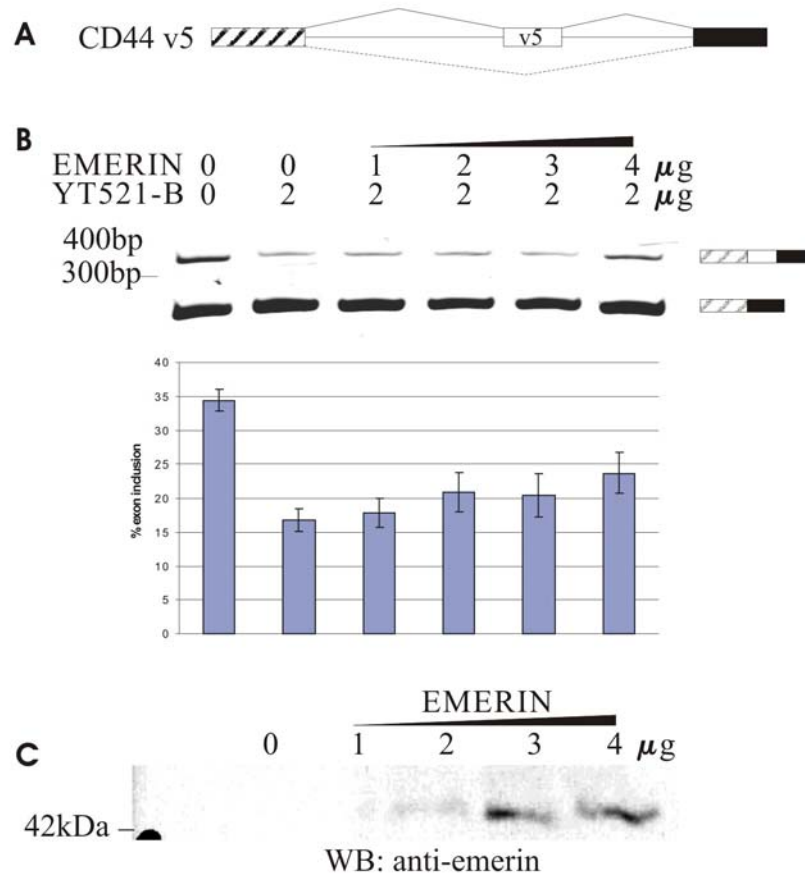


Figure 31: Emerin reverses the influence of YT521-B on splice site selection. (A) Structure of the CD44 v5 minigene used (König et al., 1998). Exons are shown as boxes, introns as lines, and the alternative CD44 exon v5 is indicated. The splicing patterns are shown by different line styles. (B) Influence of emerlin and YT521-B on CD44 exon v5 splice site selection. HEK293 cells were cotransfected with an increasing amount of emerlin in the presence of EGFP-C2 and 1 μg CD44 exon v5 minigene. The amount of transfected EGFP-YT521-B and emerlin is indicated and normalized using pEGFP-C2. The RNA was analyzed by RT-PCR. The structure of the amplified products is indicated on the right. The graph shows the statistical evaluation of the RT-PCR results. The ratio between the signal corresponding to exon inclusion and all products was determined from at least three different independent experiments. (C) The expression level of emerlin in the experiment was determined by Western blot using an antisera MANEM5 mAb.

4.14. The activity of splicing regulatory sequences depends on their conformation in the premRNA

Emerin is a nuclear membrane protein that interacts with lamin A/C at the nuclear envelope (Manilal et al., 1996). Mutations in emerlin can cause X-linked Emery–Dreifuss muscular dystrophy (EDMD) (Bione et al., 1995). The functions of emerlin are poorly understood, but emerlin has several binding partners. YT521-B was identified as an interactor of emerlin from a high-stringency yeast two-hybrid method of a human heart cDNA library, with a full length emerlin as bait (Wilkinson et al., 2003). Specific binding between emerlin and the functional C-terminal domain of YT521-B was confirmed by pull-down assays (Wilkinson et al., 2003). YT521-B has been shown to reduce inclusion of the alternative exon v5 of CD44 minigene in a concentration dependent manner (Rafalska et al., 2004). To test whether emerlin can influence

YT521-B on splice site selection, YT521-B was cotransfected with emerlin in HEK293 cells in the presence of CD44 reporter minigene. Cells transfected with 1 μ g YT521-B had inclusion of exon v5 reduced significantly ($P = 0.006$; t-test) from 34% to 17% (Figure 31 B, lane 1-2). In the presence of an increase amount of emerlin, the effect of YT521-B on splice site selection of exon v5 was partially reversed from 17% to 24% (Figure 31 B, lane 3-6). The Western blot shows the increasing amount of mutant emerlin (Figure 31 C). These data suggest that emerlin mediates the influence of YT521-B dependent splice site selection by its binding to the C-terminal of YT521-B.

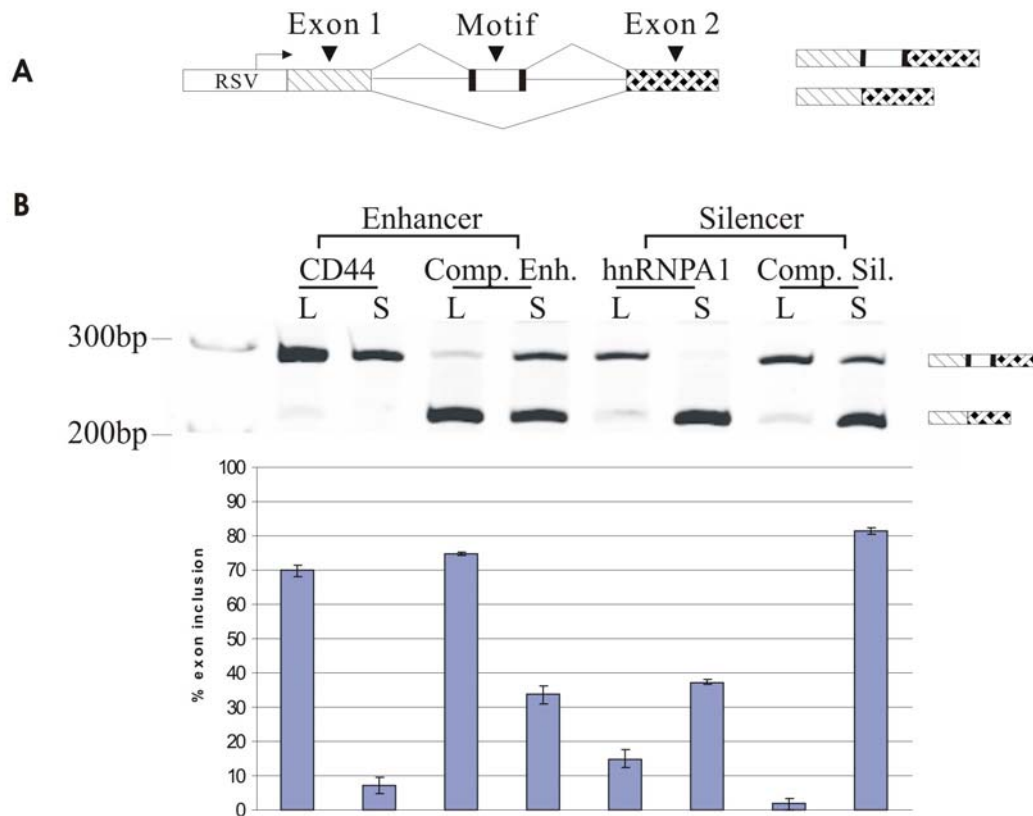


Figure 32: Influence of mRNA conformation on splice site selection. (A) The structure of the SXN-derived minigenes (see section 3.2.15. for detail of minigene construction). The shaded globin exon 1 and 2 flank a central artificial exon, indicated by “motif”, in which the YT521-B binding sequence was introduced. At least three independent experiments were evaluated using the Student’s t-test. The arrow indicates the RSV promoter. (B) Analysis of the reporter constructs *in vivo*. L: motif is in a loop structure, S: motif is in a stem structure. 1 μ g of each construct was expressed in HEK293 cells and its RNA analyzed after 24 hours of transfection. The sequences containing the motifs (underlined) were: CD44: loop structure: ATCCATGGGGCTGGATGTGACGTACAACCAC AATACGTCACATACTTCCTCTCATGA, and stem structure: ATGATGGGTATGTGCGTTGCTTCGGCAACC ACAACTCATCGCATACTTCCTCTCATGA; Comp. Enh. loop structure: ATCCATGGGGCTGGATGTGACGT ACAAAGGCATACGTCACATAGCTTCCTCTCATGA, and stem structure: CTACCTTGCGCATGATACGCAT GCGCAAGGTAGCACTGCATGAGCTTCCTCACGTTT; HnRNPA1: loop structure: ATCCATGGGGCTGGAT GTGACGTAGTAGGGTAATACGTCACATAGCTTCCTCTCATGA, and stem structure: CTACCCTACGCATGA TACGCATGCGTAGGGTAGCACTGCATGAGCTTCCTCACGTTT; Comp. Sil.: loop structure: ATCCATGGG GCTGGATGTGACGTAGTAAAGTGAAATACGTCATATCTTACCTCTCATGA, and stem structure ATCCAGTAA GCTACGCTCCGATGCGTAAAGTGAGTCGCTCACTTACGCATCTCATGA (C) Statistical analysis of three independent experiments.

In higher organisms, the splice sites do not contain all the information that is required for accurate intron recognition (Lim et al., 2001). Additional exonic and intronic splicing elements

such as ESE, ESS, ISE, and ISS are essential for the alternative and constitutive splicing. These splicing regulatory motifs bind to RNA binding proteins (Auweter et al., 2006). The secondary structure of a pre-mRNA influences a number of processing steps including alternative splicing (Buratti et al., 2004). Since most splicing regulatory proteins bind to single-stranded RNA, the sequestration of RNA into double strands could prevent their binding. Whether the localization of a splicing regulatory element in a single- or double-stranded RNA structure could influence splice site selection was analyzed by experimental approach. In order to test the hypothesis, splicing enhancer and silencer motifs located either in a single-stranded loop or in a double-stranded stem were inserted into the alternative exon of SXN reporter minigene which was widely adopted to analyze the impact of sequences on pre-mRNA processing (Coulter et al., 1997). The single- and double-stranded motifs are located at similar positions in the exon to avoid positional effects (Goren et al., 2006). As enhancer sequences, the experimentally well-characterized enhancer of the CD44 (CAACCACAA) (Stickeler et al., 2001) and a pentamer (CAAGG), which is the core of many computational predicted enhancers (Zhang et al., 2004) were used. As silencers, the experimentally characterized hnRNP A1 (TAGGGT) silencer and a computationally predicted silencer (GTAAGTGA) (Wang et al., 2004), which was previously experimentally verified in the HTR2C pre-mRNA (Kishore and Stamm 2006) were employed. HEK293 cells were transfected with different constructs and 18-22 hours later the RNAs were isolated and analyzed by RT-PCR. As shown in Figure 32 B and C, the function of these regulatory motifs strongly depends on their localization within an RNA conformation. Enhancers result in stronger exon inclusion when they are located in a loop compared with the location in the stem structure. Likewise, silencers located in a loop lead to stronger exon skipping compared to the loop structure. These data suggest that the pre-mRNA conformation influences the action of a splicing regulatory element.

5. DISCUSSION

5.1. The YTH domain is a novel RNA binding domain

Most splicing contains at least one RNA recognition motif which is necessary for the regulation of gene expression. The predicted secondary structure of YTH domain is a mixture of α -helices and β -sheets. With the conserved aromatic residues found in the β -strand, the mixed $\alpha\beta$ -fold structure resembles the RNA recognition motif (RRM) domain (Stoilov et al., 2002b). One major aim of the project presented here was to analyze the role of the YTH domain identified in the founding member of the YTH domain family of proteins, YT521-B. The data show that the function of the YTH domain is to bind to single stranded, non-structured RNA. We identified a six nucleotide YTH domain binding motif by *in vitro* SELEX (Figure 19). A striking feature of RNA sequences recognized by YT521-B is their high degeneracy. Therefore they can only be described by a degenerate weight matrix. This degeneracy of the RNAs binding to the YTH domain was also observed for the other biochemically characterized member of the YTH domain proteins, Mmi1 (Harigaya et al., 2006). Mmi1 shows 24 % identity and 45 % similarity to human YT521-B and could bind specifically to a cis-acting region of *mei4*, *ssm4*, *spo5* and *rec8* mRNAs. This cis-acting element was called determinant of selective removal (DSR). DSRs from *mei4*, *ssm4*, *spo5* and *rec8* share no obvious similarity in either the primary or the secondary structure, suggesting that the YTH domain of Mmi1 also recognizes highly degenerate RNA sequences. The exon microarray analysis shows that the YTH domain regulates distinct pre-mRNAs, which shows high-specific recognition *in vivo*. Since all *in vivo* targets of YT521-B contain clusters of YTH domain recognition motif (Figure 24), the high specificity observed *in vivo* is most likely due to binding of multiple YT521-B molecules to a single RNA. When the RNA molecules contained the YTH binding signature, the formation of complexes between YT521-B and RNAs in solution *in vitro* was observed in both gel shift assays and light scattering experiment (Figure 25). The formation of these complexes is facilitated by the ability of YT521-B C-termini to bind to each other (Hartmann et al., 1999). It is therefore likely that the high specificity of YT521-B *in vivo* is achieved by combining interactions between different proteins and binding of their YTH domains to several degenerate motifs on the RNA. The other well-characterized YTH family member, Mmi1, also shows aggregation upon target RNA binding (Harigaya et al., 2006). Another striking feature of YT521-B and Mmi1 is their localization in dynamic nuclear foci. YT521-B is localized in a novel nuclear compartment, the YT body, whose formation is regulated by phosphorylation (Figure 26). Mmi1 is found in several scattered foci during mitosis, but due to its interaction with Mei2 converges into one single dot in meiosis

(Harigaya et al., 2006). The comparison of all YTH domain protein family members shows that in addition to the YTH domain they contain numerous other protein domains that mostly enable protein interaction. Earlier, yeast-two hybrid analysis of YT521-B demonstrated that the YTH domain is not necessary for binding to other proteins (Hartmann et al., 1999). A possible general function of the YTH domain could therefore be the recruitment of interacting proteins to RNAs. It is therefore likely that YTH protein family members function in different aspects of cellular RNA metabolism. Finally, the YTH domain is not a cellular targeting signal, as we demonstrated that the human YTH protein family members can reside both in the nucleus and cytosol (Figure 27).

There are several similarities between the YTH domain and the RNA recognition motif (RRM), which is the most frequent RNA binding domain. Both domains contain an array of α -helices and β -sheets with evolutionary conserved aromatic residues in the β -sheets that are important for binding. RNA molecules binding to both domains can be only described by degenerate motifs. The major difference between the domains is the striking conservation of 14 invariant residues in the YTH domain (Stoilov et al., 2002b) which is in contrast to the only three invariant residues in the RRM (Maris et al., 2005). Both RRM and YTH domain fulfill functions in the cytosol and nucleus. However, on the proteins level, the YTH domain is combined with a larger set of interaction domains than the RRM.

The newly generated three dimension (3D) human YTH domain model suggests that the predicted YTH domain folds into an alpha-beta sandwich structure with a beta1-alpha1-beta2-alpha2-beta3-beta4- beta5-beta6- alpha3-beta7-alpha4 topology (Figure 33 A and B). The fold of YTH domain model is similar to the OB-fold, a very common nucleic acid binding module that is found in some RNA-binding protein, including the transcriptional terminator Rho and aminoacyl-tRNA synthetase (Theobald et al., 2003). The typical OB-fold consists of two three-stranded antiparallel beta-sheets, where beta1 was shared by both sheets (Figure 33 C). These two large antiparallel beta sheets, beta1-beta2-beta3 and beta4-beta5-beta6 twist and coil to form a closed structure, a beta barrel. OB-folds tend to use the cleft that runs across the surface of the OB-fold perpendicular to the axis of the beta-barrel for nucleic acid binding. However, the way in which the OB-fold recognizes the RNA is quite variable, and even the number of nucleotides per OB-fold can span a wide range. The 3D model of YTH domain showed a three-stranded beta sheet formed by beta-strand 3, 4 and 6, while beta-strand 1, 2 and 7 were arranged in an antiparalle fasion. Though beta-strands 1, 2 and 7 do not form a perfect three-stranded beta sheet, they build an enclosed structure together with the three-stranded beta sheet, beta3-beta4-beta6. Furthermore, this structure formed by these six beta-strands might serve as the RNA binding center as the cleft formed by the beta-barrel in OB-fold. Interestingly,

the conserved residues W380, F412 and G414 of YTH domain which were necessary for the ability of the protein YT521-B to modulate splice site selection, are located on the beta-strand 2 and 4 respectively. Therefore these three conserved residues might play an essential role in protein-RNA binding.

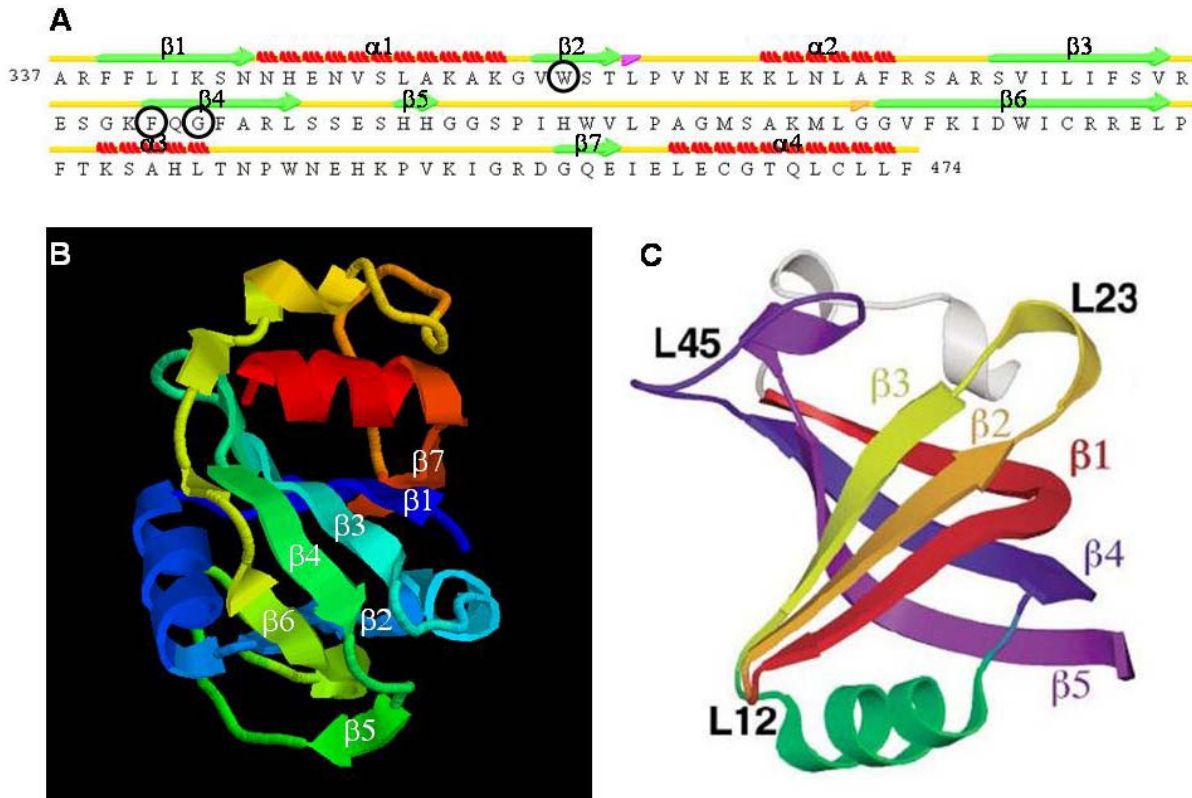


Figure 33: Theoretical prediction structure of YTH domain. (A) Predicted two dimension structure of human YTH domain. The predicted alpha-helices and beta-strands were marked in red and green respectively, indicated with numbers. The three conserved residues within YTH domain were marked by black circle. (B) Ribbon model of three dimension structure of human YTH domain. The beta-strands were indicated with numbers. The 2D and 3D YTH domain model were generated by Janusz Bujnicki (International Institute of Molecular and Cell Biology, Warsaw). (C) **The canonical OB-fold domain.** The OB-fold from AspRS is shown in stereo as representative of the ideal OB-fold domain (Theobald et al., 2003).

5.2. YT521-B is a vertebrate specific splicing factor

BLAST searches revealed that multiple proteins from *Saccharomyces cerevisiae*, *Plasmodium falciparum*, *Drosophila melanogaster*, *Mus musculus*, *Homo sapiens*, and *Arabidopsis thaliana*, *Oryza sativa* share a conserved region, the YTH domain. Database analysis result shows that all known YT521-B orthologs are found only in vertebrates from zebrafish to primates. YT521-B orthologs were present neither in bacterial nor in yeast and furthermore there was no YT521-B orthologs present in *C. elegans* during the searches. Similar to other splicing factors like SF2/ASF, SRp30c and tra2-beta1 (Stoilov et al., 2004), YT521-B may influence splice site selection indirectly by sequestration of other proteins, suggesting that numerous genes might be affected without direct involvement of the YTH domain. In order to

determine direct, RNA-dependent influences on splice site selection, exon microarray analysis were performed to compare overexpression of YT521-B with a mutant lacking the YTH domain. This experiment demonstrates that either the pre-mRNAs or the alternative exons that are directly regulated by YT521-B are also vertebrate-specific. Five representative exons regulated by YT521-B, not the mutant that lacks the YTH domain, were identified in this work. With the exception of *Zfp687* and *Rhot2* where the alternative exons keep the frame, all other exons regulated by YT521-B introduce stop codons or generate frameshifts, suggesting that YT521-B can influence the generation of proteins from different genes. A detailed analysis of NOVA target genes demonstrated that this splicing regulatory protein influences a set of functionally related genes (Ule et al., 2006). The example of YT521-B shows that a splicing factor not only can regulate functionally related pre-mRNAs, but also evolutionary related RNAs. All YT521-B target genes share a cluster of YTH domain binding sites on the pre-mRNA, which is probably recognized by multiple YT521-B molecules that also interact with each other. The formation of this vertebrate-specific complex generates then a vertebrate-specific readout of the genome.

5.3. YT521-B nuclear localization and binding ability are regulated by tyrosine phosphorylation

Many nuclear activities are concentrated in subnuclear foci called nuclear bodies, which include nuclear speckles, Cajal bodies, and PML bodies (Lamond and Sleeman, 2003; Spector, 2003). These nuclear bodies are highly dynamic intranuclear structures involved in cell cycle change, RNA transcription and splicing, RNA and DNA modification, malignant transformation or tumor suppression (Lamond and Spector, 2003; Trotman et al., 2006). YT521-B defined a novel subnuclear structure, the YT bodies. YT bodies are dynamic compartment, which first appear at the S-phase in the cell cycle and disperse during mitosis.

The experiments described here show that the intranuclear distribution of YT521-B is dependent on its tyrosine-phosphorylation status. Non-phosphorylated YT521-B is predominantly located in YT bodies that contain focal sites of transcription and partially overlap with SC35 nuclear speckles (Nayler et al., 2000). YT bodies are soluble in non-denaturing buffers. Tyrosine phosphorylation of YT521-B changes that distribution (Figure 26) and solubility (Figure 29). It causes diffuse nuclear appearance of YT521-B and its tight association with an insoluble nuclear fraction. The release of YT521-B from YT bodies after tyrosine phosphorylation is reminiscent of the dispersion of nuclear speckles after extensive phosphorylation on serine residue of their proteins through Clk/Sty kinases (Colwill et al., 1996; Gui et al., 1994). The serine phosphorylation status of splicing factors affects protein-protein interaction and regulates protein activity. *In vitro* assays have shown that Clk/Sty kinases can influence the splicing activity of SR proteins by altering their phosphorylation status (Prasad et

al., 1999). Since c-Abl causes YT521-B phosphorylation (Figure 25), colocalizes with it (Figure 26), YT521-B may be a novel nuclear target for c-Abl. It is possible that the first association between c-Abl and YT521-B is mediated by the proline-rich region of YT521-B and the SH3 domain of c-Abl. Tyrosine phosphorylation would then stabilize the binding by phosphotyrosine-SH2 domain interaction. Several other non-receptor tyrosine kinases do not phosphorylate YT521-B, indicating a specific interaction between c-Abl and YT521-B. It is well established that RNA polIII CTD is phosphorylated by c-Abl (Baskaran et al., 1993). Since YT bodies contain transcriptional start sites, it is possible that c-Abl phosphorylates YT521-B and RNA polIII in the same nuclear compartment. The activity of nuclear c-Abl is controlled during the cell cycle via interaction with pRB (Welch and Wang, 1993). The cyclin dependent hyperphosphorylation of pRB releases binding to c-Abl and activates the kinase at the S-phase. This correlates with the finding that YT bodies cannot be detected during S-phase and form at the entrance of G1 (Van Etten, 1999). It also explains why YT bodies are markers for differentiated cells and are less pronounced in rapidly dividing transformed cells (Nayler et al., 2000). However, in contrast to the CTD (Baskaran et al., 1993), YT521-B is phosphorylated by src family kinases. YT521-B shuttles between nucleus and the cytosol (Figure 28), which shows that the protein can contact the membrane bound src-family kinases c-Src and p59^{lyn} that phosphorylate YT521-B and bind to it tightly. It is therefore likely that these kinases can directly phosphorylate YT521-B and influence its properties. The Scansite Motif Scan program predicts 22 possible tyrosine phosphorylation sites in YT521-B. It remains to be determined which sites are responsible for the localization effects.

After tyrosine phosphorylation, YT521-B changes its binding properties and associates with an insoluble nuclear structure. There are several parallels between the behavior of YT521-B and membrane bound receptor tyrosine kinases that form large signaling complexes after phosphorylation using scaffolding adaptors. First, in both cases tyrosine phosphorylation induces binding to other proteins. After phosphorylation YT521-B binds tightly to c-Abl, most likely to its SH2 domain. Secondly, the complex formation induced by phosphorylation changes the function of the molecules. Membrane bound receptor tyrosine kinases gain signaling function in the activated complexes, whereas YT521-B seems to lose its function in splice site regulation. It is not clear however, whether there are nuclear scaffolding adaptor proteins in non-transformed cells. One possibility is that RNA present in the nuclear matrix serves as such an adaptor, since YTH domain of YT521-B recognizes a degenerate motif of single strand RNAs.

5.4. The influence of YT521-B on splice site selection is regulated by tyrosine phosphorylation

Similar to SR proteins where the phosphorylation status mediates their ability on splicing regulation, a functional consequence of YT521-B phosphorylation is to modulate the influence of YT521-B on splice site selection. YT521-B binds to several proteins implicated in splice site selection, which it can change in a concentration-dependent manner (Hartmann et al., 1999).

Phosphorylation on tyrosine residue changes the ability of YT521-B to regulate the cassette exon in SRp20 (Figure 30). Non-phosphorylated YT521-B colocalizes with transcriptional start sites and shows overlapping localization with nuclear speckles (Nayler et al., 2000). A similar influence of phosphorylation on the ability of YT521-B to regulate the alternative cassette exon v5 of the CD44 minigene is observed (Rafalska et al., 2004). In the IL-4 receptor system, YT521-B causes intron retention of IL-4R minigene. The retention of the intron between exon 9 and 10 can be detected in lymphocytes, indicating that it is a physiological event (Rafalska et al., 2004). These results show that YT521-B regulates not only cassette exons, but also other mode of alternative splicing. The function of YT521-B in this case is due to the binding to nascent mRNA via its YTH domain and the sequestering of other splicing factors via the protein interaction domain in its C-termini. The interactions between YT521-B and other proteins functioning in pre-RNA processing are intrinsically weak (Hartmann et al., 1999). This allows for dynamic association and reassociation to the nascent pre-mRNA-protein complexes. Such dynamic behavior is only possible when the protein is soluble. As a consequence, non-phosphorylated YT521-B can regulate splice sites in a concentration-dependent manner. Phosphorylation results in a strong association between YT521-B and nuclear structures (Figure 26). Furthermore, phosphorylated YT521-B is dispersed throughout the nucleus and removed from the vicinity of actively transcribed genes (Figure 26). Due to its insolubility and spatial distance, phosphorylated YT521-B is effectively removed from pre-mRNA processing events in the cell and can no longer influence splice site selection. A similar phosphorylation-dependent redistribution of nuclear splicing factors has been demonstrated for SF2/ASF (Caceres et al., 1998), where serine/threonine phosphorylation induced by the Clk/Sty kinases causes accumulation of SF2/ASF in the cytosol. Interestingly, the phosphorylation affects mainly the localization of proteins, but not their ability to change splice sites *in vitro* (Cazalla et al., 2002). Serine/threonine phosphorylation under stress conditions (van der Houven van Oordt et al., 2000) or hypoxia (Daoud et al., 2002) causes hyperphosphorylation and accumulation of SR-proteins or hnRNPs in the cytosol. This could serve the similar purpose of removing factors from pre-mRNA processing events. In fact, *in vivo* splice site selection is altered after application of osmotic stress or hypoxia (van der Houven van Oordt et al., 2000; Daoud et al., 2002). Using cotransfection assays, it was previously demonstrated that sequestration of splicing factors can regulate splice site selection (Stoilov et al., 2004). Although this mechanism can explain

differences in splice site selection between tissues with different concentration of regulatory proteins, it cannot account for rapid changes in splice site selection after cellular stimulation (Stamm, 2002). It is possible that phosphorylation-dependent sequestration of YT521-B in an insoluble state is a mechanism to regulate splice site selection in response to a kinase signal. After phosphorylation, YT521-B is localized in nuclear regions distant from the areas where transcription and pre-mRNA processing occurs. The phosphorylated protein is insoluble and most likely not able to participate in the dynamic rearrangement of protein complexes necessary for splice site recognition. This mechanism allows the cell to temporarily lower the active concentration of YT521-B without destroying the protein. The mechanism partially explains the frequently observed changes of splice site selection in malignant transformation (Xu and Lee, 2003) and during development (Black, 2003), as these processes are triggered or are concomitant with an increase in receptor tyrosine phosphorylation. It implies that in addition to the combination of splicing factors (Smith and Valcarcel, 2000) the tyrosine phosphorylation status of splicing factors can contribute to cell-specific alternative splicing.

5.5. Emerin binds to YT521-B and abolishes its influence on splice site selection

Emery-Dreifuss muscular dystrophy is predominantly a disease of cardiac and skeletal muscle and tendons, other tissues being unaffected (Morris 2001). In 1994, Daniela Toniolo's group identified the X-linked EDMD gene locus at Xq28 (*EMD* or *STA* gene) (Bione et al., 1994). The *EMD* gene encodes a nuclear envelope protein, which was named emerin. The functions of emerin are not very well understood, but many studies show that the emerin loss or mutation in emerin cause X-linked Emery-Dreifuss muscular dystrophy (Manilal et al., 1996; Morris, 2001). YT521-B was identified as a novel emerin-binding protein (Wilkinson et al., 2003). The interacting region of YT521-B comprises part of a proline rich domain, plus the intact C-terminal glutamic acid/arginine-rich (ER-rich) domain, which is required for binding to several RNA splicing factors (Hartmann et al., 1999). If the function of YT521-B is regulated by interactions with emerin or the emerin–lamin A/C complex, some of the cardiac and skeletal muscle symptoms of EDMD may result from inappropriate splicing of tissue-specific mRNAs. This could explain how changes in widely expressed proteins like YT521-B and emerin might produce the tissue-specific effects in EDMD (Morris 2001).

The results of *in vivo* splicing suggest that emerin binding to YT521-B might sequester or inactivate YT521-B. Alternatively, the emerin–lamin A/C complex might help to organize transcription factors, splicing factors and chromatin to ensure that specific RNA transcripts are correctly spliced. Alternative splicing can regulate gene expression directly by changing the polypeptides encoded by primary RNA transcripts. However, transcription factors are themselves the products of alternative splicing, and defects in splice site selection may have

downstream indirect effects on gene expression due to inappropriately expressed transcription factors. Although the functional effects of emerin on YT521-B may appear modest, it must be remembered that effects of complete absence of emerin in X-linked EDMD are also quite modest, with clinical features appearing many years after birth. Splicing defects are often associated with late onset disease or predisposition to disease (Stoilov et al., 2002a).

5.6. YT521-B working model

To summarize data presented in this work, the current working model of YT521-B function is proposed in Figure 34. As shown there, YT521-B contains a RNA binding domain YTH, a Proline-rich and glutamic acid/arginine-rich region at carboxy-terminal. The data presented in this work show that YTH domain binds to single stranded RNA which contains multiple copies of a six nucleotides degenerate motif. Also the data demonstrate that YT21-B is capable to associate with the oncogenic kinase c-Abl and c-Src, and can be phosphorylated on tyrosine residues by them. Therefore YT521-B might participate in various distinct cellular processes, including mRNA-processing reactions such as splicing. The properties of YT521-B make it an excellent candidate for linking signal transduction pathways and RNA processing.

YT521-B is localized in YT bodies that mostly overlap with focal sites of transcription (Nayler et al., 2000). YT521-B is also a part of a protein complex assembling around DNA binding scaffold attachment factor B (SAF-B) (Hartmann et al., 1999). In YT bodies, different chromosomal loops are combined through interaction between DNA and SAF-B. Speckles which are often found in close contact with YT bodies serve as accessory domains supplying the transcriptionally active site with the necessary helper factors. Without external stimuli (Figure 34 A), YT521-B is not phosphorylated and can form transient complexes with the processed RNA that has a cluster of YTH domain recognition motifs and its interacting splicing factors. It is likely that several YT521-B molecules bind to a single RNA due to its multiple YTH domain binding site. The formation of these complexes is facilitated by the ability of YT521-B C-termini to bind to each other (Hartmann et al., 1999). Due to the specific complex formed, an alternative exon is included.

The function of YT521-B is regulated by phosphorylation emanating from both cytosolic and nuclear non-receptor tyrosine kinases. This phosphorylation regulates the interaction between YT521-B and various proteins involved in splice site selection and nuclear scaffolding. In response to extracellular signalling, YT521-B is phosphorylated on tyrosine residues by nuclear c-abl. Phosphorylated YT521-B associates with nuclear components, probably the nuclear matrix, in an insoluble fraction. Due to this sequestration its influence on splice site selection is abolished. In this example, different complexes form on the pre-mRNA, resulting in exon skipping. This working model explains the dependency of alternative splicing events on the

tyrosine phosphorylation state that has been observed in several model systems (Stamm, 2002), such as cancer and embryonic development.

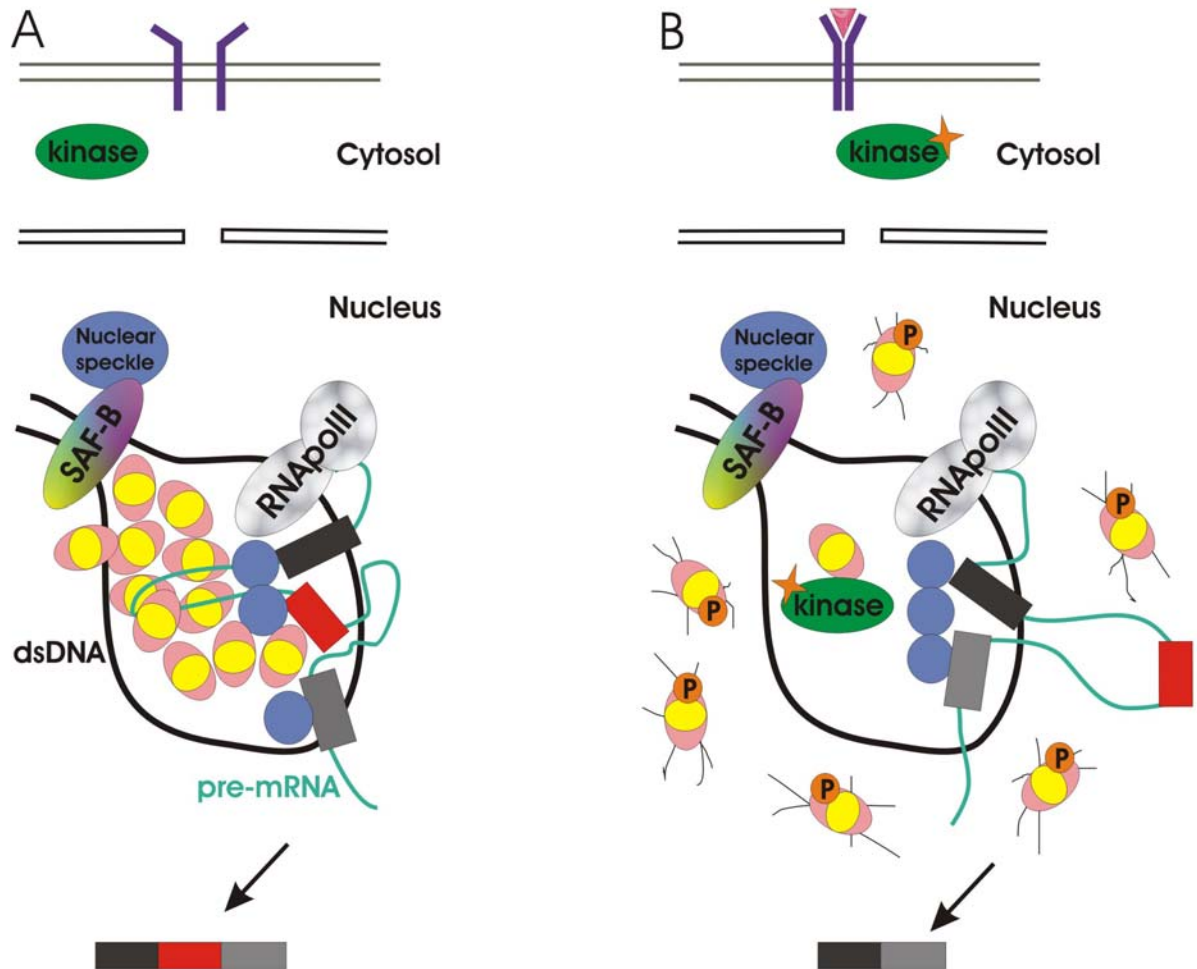


Figure 34: Working model for YTH521-B. The nuclear protein YTH521-B (pink) contains a novel RNA binding domain (yellow oval inside the pink circle). The YTH domain binds to single-strand non structure RNA molecule. Several YTH521-B proteins interact with a single RNA molecule due to the multiple YTH domain binding motifs presented in the RNA. YTH521-B (pink) is located in YT bodies that are often associated with SC35-speckles, which consists of SR-proteins and hnRNPs (blue). YTH521-B binds to scaffold-attachment factor B (SAF-B), a protein implicated in organizing DNA (black line) into transcriptionally active loops. YT-bodies contain transcriptional active sites where pre-mRNA consisting of constitutive exons (dark gray and light gray boxes), alternative exons (red box) and intron (green line) is made. The pre-mRNA forms complexes with hnRNPs and SR-proteins (blue) that bind to YTH521-B (pink). (A) A protein complex is formed that recognizes alternative exons and alternative introns. As a result, this exon is included. (B) In the presence of extracellular stimuli, non-receptor tyrosine kinases are activated. After phosphorylation by nuclear tyrosine kinases, YTH521-B binds to a nuclear structure (gray lines) and becomes insoluble. The complex of proteins and pre-mRNA that defines exons changes and as a result, the alternative exon (red) is excised. Due to this sequestration mechanism, tyrosine phosphorylation can change splice site selection.

REFERENCES

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.*, 25, 3389-3402.
- Arcus, V. (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* 12, 794-801
- Ast G. (2004) How did alternative splicing evolve? *Nat Rev Genet.*, Oct, 5(10), 773-82.
- Auweter, S. D., Oberstrass, F. C. & Allain, F. H. (2006) Sequence-specific binding of singlestranded RNA: is there a code for recognition? *Nucleic Acids Res* 34, 4943-59.
- Baskaran, R., Dahmus, M.E. and Wang, J.Y.J. (1993) Tyrosine Phosphorylation of Mammalian RNA Polymerase II Carboxyl-Terminal Domain. *Proc. Natl Acad. Sci. U. S. A.*, 90, 11167-11171.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res* 30, 276–280.
- Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G. and Toniolo, D. (1994) Identification of a novel Xlinked gene responsible for Emery-Dreifuss muscular dystrophy. *Nat. Genet.* 8, 323–327.
- Bione, S., Small, K., Aksmanovic, V.M., D’Urso, M., Ciccodicola, A., Merlini, L., Morandi, L., Kress, W., Yates, J.R., Warren, S.T. et al. (1995) Identification of new mutations in the Emery–Dreifuss muscular dystrophy gene and evidence for heterogeneity of the disease. *Hum. Mol. Genet.* 4, 1859–1863.
- Birney, E., Kumar, S., and 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.
- Blencowe, B. J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* 25, 106-10
- Bonne, G., DiBarletta, M.R., Varnous, S., Becane, H.M., Hammouda, E.H., Merlini, L., Muntoni, F., Greenberg, C.R., Gary, F., Urtizbera, J.A., Duboc, D., Fardeau, M., Toniolo, D. and Schwartz, K. (1999) Mutations in the gene encoding lamin A/C cause autosomal dominant Emery–Dreifuss muscular dystrophy. *Nature Genet.* 21, 285–288.
- Brocke K.S., Neu-Yilik G., Gehring N.H., Hentze M.W., and Kulozik A.E.(2002) The human intronless melanocortin 4-receptor gene is NMDinsensitive. *Hum Mol Genet* , 11:331-335.
- Brow D.A. (2002) Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* 36, 333-360
- Burge C.B., Tuschl T., and Sharp P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. *The RNA World*, 2nd ed, Cold Spring Harbor Laboratory Press, 525-560
- Buckanovich, R. J. and Darnell, R. B. (1997) The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol* 17, 3194-201
- Buratti, E. and Baralle, F. E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol Cell Biol* 24, 10505-14.
- Caceres, J.F., Sreaton, G.R. and Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.*, 12, 55–66.
- Caceres, J.F., Sreaton, G.R. and Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.*, 12, 55-66.
- Caputi M. and Zahler A.M. (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev specific exon 6D. *EMBO J.* 21, 845-855
- Cardelli, M., Marchegiani, F., Cavallone, L., Olivieri, F., Giovagnetti, S., Mugianesi, E., Moresi, R., Lisa, R., and Franceschi, C. (2006) A Polymorphism of the YTHDF2 Gene (1p35) Located in

- an Alu-Rich Genomic Domain Is Associated With Human Longevity. *Journal of Gerontology: BIOLOGICAL SCIENCES*, 61A, 547–556
- Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R. and Caceres, J.F. (2002) Nuclear Export and Retention Signals in the RS Domain of SR Proteins. *Mol. Cell. Biol.*, 22, 6871-6882.
- Celotto A.M. and Graveley B.R. (2001) Alternative splicing of the Drosophila Dscam pre-mRNA is both temporally and spatially regulated. *Genetics*, Oct, 159(2), 599-608.
- Chang, L. and Karin, M. (2001) Mammalian MAP kinase signalling cascades. *Nature* 410, 37–40.
- Chen C.D., Kobayashi R., and Helfman D.M. (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.* 13, 593-606
- Ciccarone, V.C., Polayes, D., and Luckow V.A. (1997). Generation of Recombinant Baculovirus DNA in E. coli Using Baculovirus Shuttle Vector, Volume 13, U. Reisch, ed. (Totowa, NJ: Humana Press Inc.).
- Clements, L., Manilal, S., Love, D.R. and Morris, G.E. (2000) Direct interaction between emerin and Lamin A. *Biochem. Biophys. Res. Commun.* 267, 709–714.
- Collins C.A. and Guthrie C. (2000) The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol.* 7, 850-854
- Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C. and Duncan, P. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.*, 15, 265-275.
- Coulter, L. R., Landree, M. A. & Cooper, T. A. (1997) Identification of a New Class of Exonic Splicing Enhancers by In Vivo Selection. *Mol Cell Biol* 17, 2143-50

- Daoud, R., Mies, G., Smialowska, A., Olah, L., Hossmann, K.-A. and Stamm, S. (2002) Ischemia Induces a Translocation of the Splicing Factor tra2-beta 1 and Changes Alternative Splicing Patterns in the Brain. *J. Neurosci.*, 22, 5889-5899.
- Darnell, J.E., Jr. (1997) Phosphotyrosine signaling and the single cell:metazoan boundary. *Proc. Natl Acad. Sci. U. S. A.*, 94, 11767-11769.
- Debnath, J., Chamorro, M., Czar, M.J., Schaeffer, E.M., Lenardo, M.J., Varmus, H.E. and Schwartzberg, P.L. (1999) rlk/TXK encodes two forms of a novel cysteine string tyrosine kinase activated by Src family kinases. *Mol. Cell. Biol.*, 19, 1498–1507.
- Dejgaard, K., and Leffers, H. (1996). Characterisation of the nucleicacid-binding activity of KH domains. Different properties of difAcknowledgments ferent domains. *Eur. J. Biochem.* 241, 425–431.
- Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell Biol.* 3, 195-205.
- Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci.* 13, 86–91.
- Du, Z., Lee, J.K., Tjhen, R., Li, S., Pan, H., Stroud R.M., and James, T.L. (2005) Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 Å. *J. Biol. Chem.* 280,38823-38830
- Ellis J. A. (2006) Emery-Dreifuss muscular dystrophy at the nuclear envelope: 10 years onCell. *Mol. Life Sci.* 63: 2702–270
- Fairley, E. A., Kendrick-Jones, J. and Ellis, J. A. (1999) The Emery-Dreifuss muscular dystrophy phenotype arises from aberrant targeting and binding of emerin at the inner nuclear membrane. *J. Cell Sci.* 112, 2571–2582.
- Famulok, M.; Szostak, J. W., (1993) Selection of Functional RNA and DNA Molecules from Randomized Sequences. *Nucleic Acids and Molecular Biology*, 7: 271.

- Furukawa, K. (1999) LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2–chromatin interaction. *J. Cell Sci.* 112, 2485–2492.
- Garrison S., Hojgaard A., Patillo D., Weis J.J. and Weis J.H. (2001) Functional Characterization of Pactolus, a beta -Integrin-like Protein Preferentially Expressed by Neutrophils. *J. Biol. Chem.*, 276, 35500-35511.
- Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T., and Ast, G. (2006) Comparative analysis identifies exonic splicing regulatory sequences-- The complex definition of enhancers and silencers. *Mol Cell* 22, 769-81.
- Graveley, B. R. (2000) Sorting out the complexity of SR protein functions. *RNA* 6, 1197-1211.
- Graveley, B. R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65-73.
- Gui, J. F., Lane, W. S. & Fu, X. D. (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* 369, 678–682.
- Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A. and Yamamoto, M. (2006) Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* 442, 45-50.
- Hartmann, A. M., Nayler, O., Schwaiger, F. W., Obermeier, A. and Stamm, S. (1999) The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59(fyn). *Mol Biol Cell* 10, 3909-26.
- Hoffman, D.W., Query, C.C., Golden, B.L., White, S.W. and Keene, J.D. (1991) RNABinding Domain of the A Protein Component of the U1 Small Nuclear Ribonucleoprotein Analyzed by NMR Spectroscopy is Structurally Similar to Ribosomal Proteins. *Proc. Natl Acad. Sci. U.S.A.*, 88, 2495-2499.
- Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S.A., and Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing *EMBO J.* 24, 1988–1998.

- Ilsley J.L., Sudol M., Winder S.J. (2002) The WW domain: linking cell signaling to the membrane cytoskeleton. *Cellular signaling* 14, 183-189.
- Jumaa, H. and Nielsen, P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.*, 16, 5077–5085.
- Keene, J.D. (2001) Ribonucleoprotein infrastructure regulating the flow of genetic information between the genome and the proteome. *Proc. Natl. Acad. Sci. USA* 98, 7018-7024.
- Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311, 230-2.
- Knight, S.W., and Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293, 2269–2271.
- Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J. and Nogues, G. (2004) Multiple links between transcription and splicing. *RNA*, 10, 1489-1498.
- Krecic A.M. and Swanson M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* 11, 363-371
- Koli, K., Saharinen, J., Karkkainen, M., and Keski-Oja, J., (2001) Novel non-TGF-beta-binding splice variant of LTBP-4 in human cells and tissues provides means to decrease TGF-beta deposition. *J. Cell. Sci.* 114, 2869–2878.
- König, H., Ponta, H. and Herrlich, P. (1998) Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.*, 17, 2904–2913.
- Lamond A.I., Sleeman J.E.. (2003) Nuclear substructure and dynamics. *Curr. Biol.* 13, R825–28
- Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.*, 4, 605-612.

- Lander E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921
- Lee, K.K., Haraguchi, T., Lee, R.S., Koujin, T., Hiraoka, Y. and Wilson, K.L. (2001) Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF. *J. Cell Sci.* 114, 4567–4573.
- Lee, S., Neumann, M., Stearman, R., Stauber, R., Pause, A., Pavlakis, G.N. and Klausner, R.D. (1999) Transcription-dependent nuclear-cytoplasmic trafficking is required for the function of the von Hippel-Lindau tumor suppressor protein. *Mol. Cell. Biol.*, 19, 1486–1497.
- Lewis B.P., Green R.E. and Brenner S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. U. S. A.*, 100, 189-192.
- Lewis, H.A., Chen, H., Edo, C., Buckanovich, R.J., Yang, Y.Y., Musunuru, K., Zhong, R., Darnell, R.B. and Burley, S.K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold Des.*, 7, 191-203.
- Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B. and Burley, S.K. (2000) Sequence-Specific RNA Binding by a Nova KH Domain: Implications for Paraneoplastic Disease and the Fragile X Syndrome. *Cell*, 100, 323-332.
- Lim, L. P. & Burge, C. B. (2001) A computational analysis of sequence features involved in recognition of short introns. *Proc Natl Acad Sci* 98, 11193-8.
- Lin, F., Blake, D.L., Callebaut, I., Skerjanc, I.S., Holmer, L., McBurney, M.W., Paulin-Levasseur, M. and Worman, H.J. (2000) MAN1, an inner nuclear membrane protein that shares the LEM domain with lamina-associated polypeptide 2 and emerin. *J. Biol. Chem.* 275, 4840–4847.
- Lomakin, A., Teplow, D. B., and Benedek, G. B. (2005) Quasielastic light scattering for protein assembly studies. *Methods Mol Biol* 299, 153-74

- Luckow, V. A., Lee, C. S., Barry, G. F., and Olins, P. O. (1993). Efficient Generation of Infectious Recombinant Baculoviruses by Site-Specific Transposon-Mediated Insertion of Foreign Genes into a Baculovirus Genome Propagated in *Escherichia coli*. *J. Virol.* 67, 4566-4579.
- Lunde, B.M., Moore, C., and Varani, G. (2007) RNA-binding proteins: modular design for efficient function. *Mol. Cell Biol.* 8, 479-490
- Manilal, S., Nguyen, T.M., Sewry, C.A. & Morris, G.E. (1996) The Emery–Dreifuss muscular dystrophy protein, emerin, is a nuclear membrane protein. *Hum. Mol. Genet.* 5, 801–806.
- Maris C., Dominguez C., and Allain F. H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS journal*, 272, 2118-2131.
- Maquat L.E., and Li X., (2001) Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA*, 7,445-456.
- Maquat L.E., (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol*, 5,89-99.
- Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110, 563–574
- Matter, N., Herrlich, P. and König, H. (2002) Signal-dependent regulation of splicing via phosphorylation of SAM68. *Nature* 420, 691–695.
- Mermond J.E., Cohen P.T., and Lamond A.I. (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* 13, 5679-5688
- Miki T, Bottaro DP, Fleming TP, Smith CL, Burgess WH, Chan AM, and Aaronson SA. (1992) Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. *Proc. Natl. Acad. Sci.* 89, 246–250.
- Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F. and Spector, D.L. (1998) Serine Phosphorylation of SR Proteins Is Required for Their Recruitment to Sites of Transcription In Vivo. *J. Cell Biol.*, 143, 297-307.

- McManus M.T. and Sharp P.A. Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* 3, 737-747 (2002).
- Modrek, B., and Lee, C. (2002) A genomic view of alternative splicing. *Nat. Genet.* 30, 13– 19.
- Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes *Nucl. Acids Res.*, 29, 2850–2859.
- Moore M.J., Query C.C., and Sharp P.A. (1993) Splicing of precursors to mRNA by the spliceosome. *The RNA World*, 1st ed, Cold Spring Harbor Laboratory Press, 303-357
- Morris, G.E. (2001) The role of the nuclear envelope in Emery–Dreifuss muscular dystrophy. *Trends Mol. Med.* 7, 572–577.
- Muh, S. J., Hovhannisyan, R. H. and Carstens, R. P. (2002) A Non-sequence-specific doublestranded RNA structural element regulates splicing of two mutually exclusive exons of fibroblast growth factor receptor 2 (FGFR2). *J Biol Chem* 277, 50143-54
- Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J. and Pastore, A. (1996) Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, 85, 237-245.
- Murzin, A.G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* 12, 861-867.
- Nayler, O., Hartmann, A. M. & Stamm, S. (2000) The ER repeat protein YT521-B localizes to a novel subnuclear compartment *J. Cell Biol.* 150, 949-961
- Neumann, M., Afonina, E., Ceccherini-Silberstein, F., Schlicht, S., Erfle, V., Pavlakis, G.N. and Brack-Werner, R. (2001) Nucleocytoplasmic transport in human astrocytes: decreased nuclear uptake of the HIV Rev shuttle protein. *J. Cell Sci.*, 114, 1717–1729.

- Nishizawa Y., Usukura J., Singh D.P., Chylack L.T. and Shinohara T. (2001) Spatial and temporal dynamics of two alternatively spliced regulatory factors, lens epithelium-derived growth factor (ledgf/p75) and p52, in the nucleus. *Cell Tissue Res.*, 305.
- Nykanen, A., Haley, B., and Zamore, P. D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* 107, 309–321
- Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* 89, 597–606.
- Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372, 432-438.
- Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C.V. and Falquet, L. (2004) MyHits: a new interactive resource for protein annotation and domain identification. *Nucl. Acids Res.*, 32, W332-335.
- Paul J Gardina, Tyson A Clark, Brian Shimada, Michelle K Staples, Qing Yang, James Veitch, Anthony Schweitzer, Tarif Awad, Charles Sugnet, Suzanne Dee, Christopher Davies, Alan Williams and Yaron Turpaz. (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* 2006, 7:325 doi:10.1186/1471-2164-7-325
- Prasad, J., Colwill, K., Pawson, T. & Manley, J. L. (1999) The protein kinase Clk/Sty directly modulates SR protein activity: both hyper- and hypophosphorylation inhibit splicing. *Mol. Cell Biol.* 19, 6991–7000
- Prou, D., Gu, W.J., Le Crom, S., Vincent, J.D., Salamero, J., and Vernier, P., (2001) Intracellular retention of the two isoforms of the D(2) dopamine receptor promotes endoplasmic reticulum disruption. *J. Cell. Sci.* 114, 3517– 3527.

- Rafalska, I., Zhang, Z., Benderska, N., Wolff, H., Hartmann, A. M., Brack-Werner, R. & Stamm, S. (2004) The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum Mol Genet* 13, 1535-49.
- Riteau, B., Rouas-Freiss, N., Menier, C., Paul, P., Dausset, J., Carosella, E.D., (2001) HLA-G2, -G3, and -G4 isoforms expressed as nonmature cell surface glycoproteins inhibit NK and antigen-specific CTL cytotoxicity. *J. Immunol.* 166, 5018– 5026.
- Robinson, D.R., Wu, Y.-M. and Lin, S.-F. (2000) The protein tyrosine kinase family of the human genome. *Oncogene*, 19, 5548-5557.
- Rongish, B.J. and Kinsey, W.H. (2000) Transient nuclear localization of Fyn kinase during development in zebrafish. *Anat. Rec.*, 260, 115–123.
- Rost, B. (1996) PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol.*, 266, 525-539.
- Sakaki, M., Koike, H., Takahashi, N., Sasagawa, N., Tomioka, S., Arahata, K. and Ishiura, S. (2001) Interaction between emerin and nuclear lamins. *J. Biochem.* 129, 321–327.
- Shatkin, A.J. and Manley, J.L. (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.*, 7, 838-842.
- Singh, N. N., Singh, R. N. and Androphy, E. J. (2007) Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* 35, 371-89
- Skrisovska, L., Bourgeois CF, Stefl R, Grellscheid SN, Kister L, Wenter P, Elliott DJ, Stevenin J., and Allain FH (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep* 8, 372-9
- Smith, C.W.J. and Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.*, 25, 381-388.

- Spector DL. (2003) The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* 72:573–608
- Stamm, S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.* 11, 2409-2416.
- Stickeler, E., Fraser SD, Honig A, Chen AL, Berget SM, Cooper TA. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *Embo J* 20, 3821-30.
- Stoilov, P., Daoud, R., Nayler, O. & Stamm, S. (2004) Human tra2-beta1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum Mol Genet* 13, 509-524
- Stoilov, P., Meshorer, E., Gencheva, M., Glick, D., Soreq, H. and Stamm, S. (2002a) Defects in pre-mRNA processing as causes of and predisposition to diseases. *DNA Cell Biol.* 21, 803–818.
- Stoilov, P., Rafalska, I. & Stamm, S. (2002b) YTH: a new domain in nuclear proteins. *Trends Biochem. Sci.* 27, 495-496.
- Stoss, O., Olbrich, M., Hartmann, A. M., Konig, H., Memmott, J., Andreadis, A. & Stamm, S. (2001) The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites. *J Biol Chem* 276, 8665-73.
- Stoss, O., Stoilov, P., Hartmann, A. M., Nayler, O. & Stamm, S. (1999) The in vivo minigene approach to analyze tissue-specific splicing. *Brain Research Protocols* 4, 383-394.
- Sullivan, T., Escalante-Alcalde, D., Bhatt, H., Anver, M., Bhat, N., Nagashima, K., Stewart, C.L. and Burke, B. (1999) Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* 147, 913–920.
- Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* 20, 1785– 1796.

- Sylvius, N. and Tesson, F. (2006) Lamin A/C and cardiac diseases. *Curr. Opin. Cardiol.* 21, 159–165.
- Taagepera, S., McDonald, D., Loeb, J.E., Whitaker, L.L., McElroy, A.K., Wang, J.Y.J. and Hope, T.J. (1998) Nuclear-cytoplasmic shuttling of C-ABL tyrosine kinase. *Proc. Natl Acad. Sci. U. S. A.*, 95, 7457–7462.
- Taylor, S. J., Resnick, R. J. and Shalloway, D. (2004) SAM68 exerts separable effects on cell cycle progression and apoptosis. *BMC Cell Biol.* 5, 5.
- Tazi, J., Daugeron, M.C., Cathala, G., Brunel, C., and Jeanteur, P. (1992) Adenosine phosphorothioates (ATPaS and ATPgS) differentially affect the two steps of mammalian pre-mRNA splicing. *J. Bio. Chem.* 267, 4322-4326.
- Tazi, J., Kornstädt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C., and Luhrmann, R. (1993) Thiophosphoylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* 363, 283-286.
- Theobald, D.L., Mitton-Fry, R.M., and Wuttke D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct.*, 32: 115–133.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22, 4673–4680.
- Tone M., Tone Y., Fairchild P.J., Wykes M. and Waldmann H. (2001) Regulation of CD40 function by its isoforms generated through alternative splicing. *Proc. Natl Acad. Sci. U. S. A.*, 98, 1751-1756.
- Trotman, L.C., Alimonti, A., Scaglioni, P.P., Koutcher, J.A., Cordon-Cardo, C., and Pandolfi, P.P. (2006) Identification of a tumor suppressor network opposing nuclear Akt function. *Nature*, 441,523-527.
- Tuerk C, Gold L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science.* 249, 505-10.

- Ule, J., and Darnell, R.B., (2003) RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin.Neurobiol.* 16, 102-110
- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J. & Darnell, R. B. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature.* 30;444 (7119):580-6.
- van der Houven van Oordt, W., Diaz-Meco, M.T., Lozano, J., Krainer, A.R., Moscat, J. and Caceres, J.F. (2000) The MKK3/6-p38-signaling Cascade Alters the Subcellular Distribution of hnRNP A1 and Modulates Alternative Splicing Regulation. *J. Cell Biol.*, 149, 307-316.
- Van Etten, R.A. (1999) Cycling, stressed-out and nervous: cellular functions of c-Abl. *Trends Cell Biol.*, 9, 179–186.
- Venter J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304-1351
- Wang, Z., Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell* 119, 831-45.
- Wedekind, J.E., Dance, G.S.C., Sowden, M.P. and Smith, H.C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet.*, 19, 207-216.
- Weg-Remers, S., Ponta, H., Herrlich, P. and König, H. (2001) Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. *EMBO J.* 20, 4194–4203.
- Welch, P. and Wang, J. (1993) A C-terminal protein-binding domain in the retinoblastoma protein regulates nuclear c-Abl tyrosine kinase in the cell cycle. *Cell*, 75, 779-790.
- Williams, R.W., and Rubin, G.M. (2002) ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos. *Proc. Natl Acad. Sci.* 99, 6889–6894.
- Wilkinson, F.L., Holaska, J.M., Zhang, Z., Sharma, A., Manilal, S., Holt, I, Stamm, S., Wilson, K.L. and Morris G.E. (2003) Emerin interacts in vitro with the splicing-associated factor, YT521-B. *Eur J Biochem.* 270(11):2459-66.

- Woodley, L., and Valcarcel, J. (2002) Regulation of alternative pre-mRNA splicing. *Brief. Funct. Genomic. Proteomic.* 1, 266-277
- Xiao, S.H. and Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.*, 11, 334-344.
- Xu, Q. and Lee, C. (2003) Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. *Nucl. Acids Res.*, 31, 5635-5643.
- Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou M.M. (2003) Structure and conserved RNA binding of the PAZ domain. *Nature* 426: 469-474
- Zavolan M., Kondo S., Schonbach C., Adachi J., Hume D.A., Hayashizaki Y. and Gaasterland T. (2003) Impact of Alternative Initiation, Splicing, and Termination on the Diversity of the mRNA Transcripts Encoded by the Mouse Transcriptome. *Genome Res.*, 13, 1290-1300.
- Zhang, X. H. & Chasin, L. A. (2004) Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* 18, 1241-50.
- Zhou Z., Licklider L.J., Gygi S.P., and Reed R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182-185
- Zorio, D.A.R. and Bentley, D.L. (2004) The link between mRNA processing and transcription: communication works both ways. *Exp. Cell Res.*, 296, 91-97.

Curriculum Vitae

Name: Zhaiyi Zhang
Date of Birth: 11th October, 1975.
Place of Birth: Tianjin, China.

Education:

11/2002-09/2007 Friedrich-Alexander-University, Erlangen, Germany
Graduate student at Stamm's lab,
09/2001-07/2002 University of Franch-Comté, Besançon, France
Master of Biochemistry, Molecular and cellular Biology
M. Sc.: "Studying the expression of sulfhydryl oxidase 3 gene"
09/1994-07/1998 The East China University of Science and Technology, Shanghai, P.R. China
Bachelor of Science, Biochemistry
B.Sc. Thesis: "Gene therapy of Parkinson disease"
09/1991-07/1994 High School of Haihe River, Tianjin, P.R. China
High School Diploma

Professional Experience:

09/2001-07/2002 Research for Master's dissertation in Biochemistry department,
University of Franch-Comté, Besançon, France
08/1998-01/2000 Working in the Laboratory of Phamacology,
Tianjin Pharmaceutical Research Institute, Tianjin, P.R.China
11/1997-06/1998 Research for Bachelor's dissertation in the Institute of Biochemistry, Cell and
Biology, Chinese Academy of Sciences, Shanghai, P.R.China

Honors and awards:

11/1994-07/1998 The East China University of Science and Technology scholarship for
outstanding student (four times)

Publications:

Zhang, Z., Rafalska, I., Hiller, M., de la Grange, P., Pudimat, R., Heinrich, B. and Stamm, S. the YTH domain is a novel RNA binding domain (submitted)
Hiller, M., **Zhang, Z.**, Backofen, R. and Stamm, S. Pre-mRNA secondary structures influence exon recognition (submitted)
Heinrich, B., **Zhang, Z.**, Novoyatleva, T. and Stamm, S. Aberrant pre-mRNA splicing as a cause of human disease (2005). *Journal of Clinical Ligand Assay*, 7, 68-74.
Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., **Zhang, Z.**, Toiber, D., Thanaraj, T.A., Soreq, H. (2005) Function of alternative splicing. *Gene*, 344, 1-20
Rafalska, I., **Zhang, Z.**, Benderska, N., Wolff, H., Hartmann, A.M., Brack-Werner, R., Stamm, S. (2004) The intranuclear localization and function of YT521-B is regulated by tyrosine phosphorylation. *Hum. Mol. Genet.*, 13, 1535-1549
Wilkinson, F.L., Holaska, J.M., **Zhang, Z.**, Sharma, A., Manilal, S., Holt, I., Stamm, S., Wilson, K.L., Morris, G.E. (2003) Emerin interacts in vitro with the splicing-associated factor, YT521-B. *Eur. J. Biochem.* 270, 2459–2466
Zhang, Z. and Stamm, S., pre-mRNAs splicing regulation by YTH, a novel RNA binding domain, EMBO 1st meeting on "pre-mRNA processing and diseases", Cortina d'Ampezzo, Italy January 14-17, 2007 (conference materials)
Zhang, Z. and Stamm, S., YTH: a novel RNA binding domain. 2nd Symposium on alternative transcript diversity II biology and therapeutics, EMBL Heidelberg, Germany March 21-23, 2006 (conference materials)
Zhang, Z., Dognin-Bergeret, M., Jouvenot, M., Radom, J., Studying of the expression of the sulfhydryl oxidase gene in normal cell cultures. 29th National Young Researchers' Conference, France 2002 (conference materials)