

DISS. ETH NO. 24567

**STRUCTURAL AND FUNCTIONAL INVESTIGATION OF TWO  
SPLICING FACTORS: HUMAN HNRNP G AND YEAST NPL3**

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

AHMED MOURSY

Master of Science

ETH Zürich

born on *01. 09. 1987*

citizen of Egypt

accepted on the recommendation of

Prof. Dr. Frédéric Allain

Prof. Dr. Vikram Panse

Prof. Dr. Constance Ciaudo

Dr. Antoine Cléry

**2017**

## Summary

The genetic information stored in the DNA is transferred by a first step of transcription to RNA. The transcribed pre-messenger RNA (pre-mRNA) is subjected to a series of maturation steps before it is translated into functional protein. One of the major maturation steps is constitutive splicing, which consists in the removal of non-coding sequences (introns) within the RNA and the ligation of coding sequences (exons). Interestingly, the pre-mRNA can be spliced differently by the inclusion or exclusion of specific exonic or intronic elements resulting in different mRNAs from a single gene. This process known as alternative splicing allows the modulation of the message conveyed by each gene. The resulting diversity of messages provides eukaryotic cells some flexibility to react to variations in intra and extracellular conditions. Splicing is subjected to very tight regulation steps to ensure that the correct pieces of spliced RNA are ligated together and give a meaningful mRNA. This control is ensured by two major classes of proteins, namely serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs). The splicing outcome will depend on their competition to bind the pre-mRNA at a precise position that will allow them to either stimulate or inhibit the splicing reaction. This action is achieved by influencing the recruitment of the nuclear machinery responsible for the splicing catalysis, the spliceosome.

In the Spinal Muscular Atrophy (SMA) disease, patients lack a functional copy of Survival of Motor Neurons 1 (*SMN1*) gene and retain only a closely related gene *SMN2* that does not allow the splicing of exon 7 leading to the production of truncated, and non-functional SMN proteins. One of the main activators of *SMN2* exon 7 splicing is hnRNP G. We solved the NMR structure of the RNA Recognition Motif (RRM) of this splicing factor bound to an *SMN2* derived RNA and showed, in combination with *in vivo* splicing assays, that contrary to what was previously proposed, this protein binds *SMN2* at a specific site in exon 7. This discovery was important because it allowed us to propose a model of how hnRNP G and its interacting partner Tra2- $\beta$ 1 are co-recruited on *SMN2* exon 7. Moreover, it allows the investigation of novel structure-based therapeutic strategies targeting those two strong splicing activators of *SMN2* exon 7.

In contrast to humans, only ~4% of *Saccharomyces cerevisiae* genes contain introns, and very few among them undergo alternative splicing. *S. cerevisiae* lacks classical SR proteins but

contains 3 SR-like proteins. Npl3 is the only SR-like protein in budding yeast that was reported to influence splicing. We identified the minimal RNA motifs recognized by the 2 RRMs of Npl3 and determined the structure of their individual complexes using NMR. The structures explained the specific interaction of Npl3 with RNA and allowed us to design point mutations affecting the binding of either RRM1 or RRM2 to RNA. Using those Npl3 variants, we showed that the interaction of both domains with RNA is required for the splicing function of Npl3 in yeast. However, the binding of RRM1 but not RRM2 to RNA is required for the interaction with chromatin remodeling complexes. Finally, we used a combined CRAC and iCLIP approach to show that the motifs bound by each RRM *in vitro* were enriched in the Npl3 binding sites *in vivo*. Based on all those results, we proposed a model explaining the binding of both RRMs to the natural RNA targets of Npl3.

In summary, we investigated both structurally and functionally how two important splicing factors belonging to the two main families of splicing regulators bind specifically their RNA target sequences and how this binding contributes to their functions *in vivo*.

## Zusammenfassung

Die genetische Information welche in der DNS gespeichert ist wird in einem ersten Schritt in RNS umgeschrieben. Die transkribierte pre-Boten-RNS (pre-mRNA für engl.: *pre-messenger RNA*) durchläuft eine Serie von Reifungsschritten bevor sie in funktionales Protein übersetzt wird. Einer der Hauptschritte ist das konstitutive Spleissen bei dem nicht-kodierende Sequenzen (Introns) entfernt werden und kodierende Sequenzen (Exons) zusammengefügt werden. Dabei kann die pre-mRNA unterschiedlich gespleisst werden und exonische oder intronische Elemente können ein- und ausgeschlossen werden. Dies resultiert in unterschiedlichen mRNAs kodiert von demselben Gen. Dieser als “alternatives Spleissen” bekannte Prozess ermöglicht eine Modulation der Information die von den einzelnen Genen überbracht wird. Die resultierende Diversität der Boten-RNS erlaubt der eukaryotischen Zellen eine gewisse Flexibilität um auf Änderungen von intrazellulären und extrazellulären Bedingungen zu reagieren. Das Spleissen wird sehr genau kontrolliert um sicher zu stellen, dass die richtigen RNS-Stücke zusammengefügt werden und eine sinnvolle Boten-RNS ergeben. Diese Kontrolle wird hauptsächlich von zwei Protein-Klassen durchgeführt, namentlich die SR Proteine und heterogene nukleäre Ribonucleoproteine (hnRNPs). Das Spleissergebnis hängt dabei von der kompetitiven Bindung dieser Proteine an der exakten Bindestelle auf der pre-mRNA ab, was wiederum das Spleissens durch die Rekrutierung der nukleären Spleiss-Maschinerie, dem Spleissosom, inhibiert oder stimuliert.

Patienten, die an spinaler Muskelatrophie erkrankt sind, haben keine funktionale Kopie des Survival of Motor Neurons 1 (SMN1) Genes und nur das nahverwandte SMN2 Gen, welches nicht das Spleissen von Exon 7 erlaubt und somit nur zur Produktion eines verkürzten und minderfunktionalem SMN Proteins führt. Einer der wichtigsten Aktivatoren des Spleissens von SMN2 Exon 7 ist hnRNP G. Wir haben die NMR-Struktur des RNA Erkennungsmotivs (RRM, engl.: *RNA Recognition Motif*) dieses Spleissfaktors gebunden und eine von SMN2 abgeleitete RNS gelöst und konnten in Kombination mit *in vivo* Spleissassays zeigen, dass – anders als zuvor vorgeschlagen – dieses Protein die SMN2 pre-mRNA an Exon 7 spezifisch bindet. Diese Entdeckung ist wichtig, da sie uns erlaubt ein Model vorzuschlagen wie hnRNP G und sein Interaktionspartner Tra2-β1 zusammen an SMN2 Exon 7 rekrutiert werden. Zudem erlaubt dies

neue strukturbasierte Therapieansätze welche auf die beiden potenten Spleissaktivatoren von SMN2 Exon 7 abzielen.

Im Gegensatz zum Mensch enthalten nur ca. 4 % der Gene von *Saccharomyces cerevisiae* Introns und nur wenige von diesen werden alternativ gespleisst. *S. cerevisiae* hat selbst keine klassischen SR Proteine dafür aber 3 SR-ähnliche Proteine. Npl3 ist bisher das einzige SR-ähnliche Protein in knospender Hefe von dem bekannt ist, dass es das Spleissen beeinflusst. Wir identifizierten die minimalen Erkennungs-RNS-sequenzen der zwei RRM's von Npl3 und lösten die Strukturen ihrer Komplexe mit RNS mittels NMR. Die Strukturen erklären die Spezifität von Npl3 und ermöglichten das Design von Punktmutationen welche die Bindung an die RNS beeinflussen. Mit Hilfe dieser Mutanten konnten wir zeigen, dass die Bindung von beiden Npl3-RRM's an die RNS für die Spleissfunktion von Npl3 in Hefe notwendig ist. Dies ist jedoch nicht der Fall für die Interaktion mit dem Chromatin Remodelierungskomplexen (engl.: chromatin remodeling complexes). Hierbei ist nur die Bindung von RRM1 aber nicht RRM2 notwendig. Schließlich konnten wir mittels einer kombinierten CRAC und iCLIP Anwendung zeigen, dass die *in vitro* identifizierten Bindestellen auf der RNS auch *in vivo* in den identifizierten Npl3 Bindestellen vermehrt detektiert werden. Aufgrund unserer Daten konnten wir Model vorschlagen, dass erklärt wie beide RRM's und Npl3 Zielsequenzen bindet.

Zusammengefasst haben wir strukturell und funktionell erforscht wie zwei wichtige Spleissfaktoren, welche jeweils eine der Hauptfamilien vertreten, spezifisch ihre RNS-Zielsequenzen binden und wie diese Bindung zur Funktion *in vivo* beiträgt.