



FACULTY OF SCIENCE

**Structural studies of RNA-protein and
protein-protein interactions causing
mis-splicing of CFTR exon 9**

Doctoral thesis

CHRISTELLE BEAUMONT

Supervisor: Mgr. PharmDr. Peter Lukavsky, Dr. rer. nat.

National Centre for Biomolecular Research
Program Biomolecular Chemistry and Bioinformatics

Brno 2025

M U N I
S C I

To my family,

Bibliographic record

Author: Christelle Beaumont
Faculty of Science
Masaryk University
National Centre for Biomolecular Research

Title of Thesis: Structural studies of RNA-protein and protein-protein
interactions causing mis-splicing of CFTR exon 9

Degree Program: Biomolecular Chemistry and Bioinformatics

Supervisor: Mgr. PharmDr. Peter Lukavsky, Dr. rer. nat.

Year: 2025

Number of Pages: 87

Keywords: Alternative splicing, CFTR, pre-mRNA, RNA binding protein,
RNA-binding mode, NMR spectroscopy

Abstract

RNA-protein interactions play a key role in the aberrant splicing of CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) exon 9. Exon 9 skipping leads to the production of a non-functional chloride channel associated with severe forms of cystic fibrosis. The mis-splicing depends primarily on variations in the polymorphic (TG)_mTn locus upstream of exon 9. At the pre-mRNA level, it generates an extended UG-rich binding site for TDP-43, associated with hnRNP A1 recruitment, and prevention of exon 9 3' splicing site (3'ss) recognition. While TDP-43 is the dominant inhibitor of exon 9 inclusion, the role of hnRNP A1, a protein with two RNA recognition motifs (RRM1 and RRM2) and a glycine-rich domain, remained unclear. In this work, we have studied the interaction between hnRNP A1 and the CFTR pre-mRNA using Nuclear Magnetic Resonance spectroscopy (NMR) and Isothermal Thermal Calorimetry (ITC). The affinities are sub micromolar and ITC data suggest that the separate RRM1 and RRM2 as well as tandem RRM1 and RRM2 form 1:1 complex. NMR titrations reveal that hnRNP A1 interacts with model CFTR 3'ss sequences in a fast exchange at the NMR timescale. Splicing assays finally show that this hnRNP A1 binding site represents a previously unknown exonic splicing silencer element. Together, our results shed light on the mechanism of aberrant CFTR exon 9 splicing.

Acknowledgements

I would like to express my sincere thanks to all CEITEC core facilities and administration staff from Masaryk university for their help and support. I would like to thank the committee members for their feedback.

I would like to thank my parents for their unconditional support during the four years I spent in Brno and all the years that followed for the completion of this project. Thank you for your support, which made me stronger and enabled me to overcome all the difficulties. I will never find words strong enough to express all the gratitude I feel for you, as well as my admiration for you. Cảm ơn bố.

I would like to thank my brother Fabrice for his financial support. You have been a real support along the way. Thank you, Sacha, for the comfort you gave me. You teach me that sometimes a simple smile is enough to make everything more beautiful. Thank you, Tiph-anie and Mila, for your sweetness and kindness.

Many thanks Charles for your guidance, kindness and constant support. You were always there to guide me, to encourage me and give me the motivation I needed. I am deeply grateful for your help and kindness throughout this period.

I would like to thanks Sambibi, for her encouragement which have been a precious support to me. I'm very grateful having you as a friend.

Thank you so much, Aistè. I was lucky enough to meet you nine years ago and I'm very happy that we became such good friends. Thank you for all the memories we have shared together.

I would like to thank my colleagues and friends at CEITEC-MU institute.

Table of Contents

Abstract	5
Table of contents	7
List of Figures	10
List of Tables	11
Glossary	12
List of Appendices	14
1 Introduction	15
1.1 Cystic Fibrosis	16
1.1.1 Localization of the CFTR gene	16
1.1.2 Structure and regulation of channel opening	16
1.2.3 Mutation cause cystic fibrosis	17
1.2 Mechanism and regulation of constitutive and alternative splicing	18
1.2.1 The splicing mechanism	18
1.2.2 Mode of alternative splicing	19
1.2.3 Regulation of Alternative splicing	21
1.2.3.1 Cis and Trans regulatory elements	21
1.2.4 Aberrant splicing in exon 9 causing cystic fibrosis	22
2 RNA - Protein interaction	24
2.1 RBD : RNA Recognition Protein	24
2.1.1 RRM : RNA Recognition Motif	24
2.1.1.1 The hnRNP family	24
2.1.1.2 Structural insight of hnRNP A1	25
2.1.1.3 Structural insight of TDP-43	31

2.1.2 dsRBD: double stranded RNA Binding Domain	33
2.1.2.1 Staufen interaction with the ARF1 dsRNA	33
2.1.2.2 Structural insight of STAU1	34
3 Methods used to study the hnRNP A1 interacting with the 3' ss CFTR ss RNA interaction	35
3.1 Isothermal titration microcalorimetry (ITC)	35
3.1.1 Basic thermodynamic concept of biomolecular interaction	35
3.1.1.1 Basic affinity (k_a) and Gibbs Free Energy ΔG	35
3.1.1.2 The enthalpy contribution (ΔH)	37
3.1.1.3 The entropy contribution (ΔS)	37
3.1.2 The isothermal titration microcalorimeter (ITC): principle	38
3.2 Nuclear Magnetic Resonance	41
3.2.1 Spin and magnetic resonance	41
3.2.2 Nuclear Magnetism	41
3.2.3 Macroscopic magnetization	42
3.2.4 Resonance principle	42
3.2.5 Kinetic of the interaction	43
3.2.1 NMR Titration experiments	44
4. Material et Methods	49
5. Results	54
5.1 Structural study of hnRNP A1 and 3' ss RNA CFTR exon	55
5.1.1 In vivo splicing assay identifies a novel ESS in CFTR exon	55
5.1.2 Binding studies of hnRNPA1 tandem RRM with the novel CFTR exon 9 ESS sequence	56
5.1.3 Binding studies of the individual hnRNP A1 RRM with separated ESS motifs	58
5.1.4 Assessment of the ESS- hnRNP A1 interaction by solution NMR spectroscopy	59

5.2 The extended complex of hnRNPA1 and TDP-43 interact with the UG₁₃RNA sequence	67
Small angle X-ray/neutron scattering of hnRNP A1 and TDP-43 interacting with the UG₁₃RNA sequence	67
5.3 Staufen interacting with the ARF1 dsRNA	69
5.3.1 Characterization of dsRB3 and dsRBD4 of STAU1 protein with the long dsRNA SBS ARF1	69
5.3.2 Complexes of individual dsRBDs bound to the short ARF1 SBS dsRNA	70
5.3.3 Complexes of individual dsRBD4 interacting with sARF1 SBS dsRNA	70
5.3.4 Mutation on minor groove show a low effect in vitro but has a high effect <i>in vivo</i>	71
7. Discussion	73
8. Conclusion	76
Original publications and definition of the author's contribution	79
Bibliography	81
Appendix A: hnRNP A1 induces aberrant CFTR exon 9 splicing <i>via</i> a newly discovered Exonic Splicing Silencer element	88
Appendix B: Staufen1 reads out structure and sequence features in ARF1 dsRNA for target recognition	89

List of Figures

Figure 1: Schematic representation of the CFTR (Cystic Fibrosis Transmembrane Regulator) protein	17
Figure 2: Exon and Intron boundary at the 5' and 3' splicing site	18
Figure 3: Types of alternative splicing	20
Figure 4: Structure and characteristic of hnRNP A1b and A1	27
Figure 5: Crystal structure of hnRNPA1 upon binding to d(TTAGGG) ₂	29
Figure 6: NMR structure of hnRNP A1	30
Figure 7: Structure of TDP-43 showing the two RRM	33
Figure 8: NMR structure of RRM 1 and 2 of TDP-43 binding 5'GUGUGAAUGAAU3' RNA sequence	33
Figure 9: Three most common types of interactions	39
Figure 10: Isothermal Titration Calorimetry device (ITC)	40
Figure 11: Raw microcalorimetric data integrated and fitted to the theoretical sigmoidal curve	41
Figure 12: Principles of NMR	44
Figure 13: Chemical exchange in NMR	47
Figure 14: Spectrum of different type of exchange regime in NMR timescale	48
Figure 15: Minigene splicing assay identifies a novel ESS motif in CFTR exon 9	55
Figure 16: NMR titration of hnRNP A1 RRM1 reveals specific interactions with the individual segments of the ESS motif	58
Figure 17: NMR titration of hnRNP A1 RRM2 reveals specific interactions with the individual segments of the ESS motif	59
Figure 18: NMR titration of RRM1 and RRM2 interacting with ESS motif	60
Figure 19: Comparison of NMR titration of RRM1 and RRM2 binding the RNA target and the DNA mimicking the RNA target at 308K to 323K	61
Figure 20: NMR titration of RRM1 and RRM2 with the RNA target at 323K and 315	62
Figure 21: NMR titration of hnRNP A1 RRM1 reveals specific interactions with the individual segments of the ESS motif	63
Figure 22: NMR titration of hnRNP A1 RRM2 reveals specific interactions with the individual segments of the ESS motif	65
Figure 23: The hnRNP A1 tandem RRM display a continuous binding site for the novel CFTR exon 9 ESS motif	66

List of Tables

Table 1: SAXS data collection of the hnRNPA1 and TDP-43 & 57 mer (UG13) RNA target	68
---	----

Glossary

3'ss	– 3'splicing site
5'ss	– 5'splicing site
Adenosine BP	– Adenosine Branch Point
CF	– Cystic Fibrosis
CFTR	– Cystic Fibrosis Transmembrane Conductance Regulator
CSP	– Chemical Shift Perturbation
CLIP	– Cross linking Immunoprecipitation
DNA	– Desoxyribonucleic Acid
dsRNA	– Double stranded Desoxyribonucleic Acid
R domain	– Regulation Domain
ESE	– Exon Splicing Enhancer
ESS	– Exon Splicing Silencer
hnRNP A1	– Heterogeneous Nuclear Ribonucleoprotein A1
HPLC	– High-Performance Liquid Chromatography
HSQC	– Heteronuclear Single Quantum Coherence Spectroscopy
ISE	– Intronic Splicing Enhancer
ISS	– Intronic Splicing Silencer
ITC	– Isothermal Titration Calorimetric
F508del	– Phenylalanine 508 Deletion
FA	– Fluorescence anisotropy
FPLC	– Fast Protein Liquid Chromatography
Kb	– Kilobase
mRNA	– messenger RNA
snRNA	– small nuclear RNA
MW	– Molecular Weight
NBD1	– Nucleotide Binding Domain 1

NBD2	– Nucleotide Binding Domain 2
NES	– Nuclear Export Signal
NLS	– Nuclear Localization Signal
NMD	– non-mediated decay
NMR	– Nuclear Magnetic Resonance
ppm	– parts per million
PTC	– Premature Stop Codon
PolyT	– Polypyrimidine tract
RBD	– RNA Binding Domain
RBP	– RNA Binding Protein
RRM	– RNA Recognition Motif
RNA	– Ribonucleic Acid
SELEX	– Systematic Evolution of Ligands by Exponential Enrichment
snRNP	– Small nuclear ribonucleoprotein
TDP-43	– TAR DNA-binding Protein 43
U2AF	– U2 Auxiliary Factor
UTR	– Untranslated region
SAXS	– Small Angle X-Ray Scattering
SMD	– Staufen mediated decay
STAU1	– Staufen 1
STAU2	– Staufen 2
SAXS	– Small Angle X-Ray Scattering

List of Appendices

Appendix A [hnRNP A1 induces aberrant CFTR exon 9 splicing *via* a newly discovered Exonic Splicing Silencer element]

Appendix B [Staufen1 reads our structure and sequence features in ARF1 ds RNA for target recognition]

1 Introduction

Buratti et al (1) discovered that the TDP-43 (TAR DNA-binding protein 43) is the main inhibitor of alternative splicing of exon 9 of the CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene. Upstream of exon 9, in intron 8, a mutation can produce different variants with a variable number of polyT, such as T5, T7, T9 and, in rare cases, T3 (2). The disease will be more severe depending on the number of polyTs and the ratio of mutated alleles found in the individual's genotype. The number of T and TG repeats in the polypyrimidine tract can modulate protein interactions, especially those involved in splicing such as the spliceosome, and thus affect exon 9 splicing.

Lukavsky et al (3) solved the NMR structure of TDP-43 RBD (RNA binding domain) interacting to the RNA sequence containing a UG repeat 5' GUGUGAAUGAAU 3'. The authors also perform *in vivo* splicing assay using CFTR exon 9 sequence with the TG11-T5 repeat at the 3' splicing site being identified as TDP-43 binding site. They have shown that the repression of TDP-43 splicing is determined by the interaction of the UG repeat.

Our collaborators Pr Emanuele Buratti and Dr Cristiana Stuaní have identified a new exonic silencer sequence (ESS) in the 3'ss of exon 9 of CFTR using the exon splicing assay. In addition, the protein hnRNP A1 (heterogeneous nuclear ribonucleoproteins) can bind RNA sequences involved in multiple functions such as alternative splicing. Although TDP-43 protein is the main inhibitor, we suggest that TDP-43 protein may act in concert with hnRNP A1 protein via the newly discovered ESS.

The main objective of this thesis was to characterize the interaction of both RRM (RNA recognition motif) domains of hnRNP A1 protein with the 3'ss of the recently identified CFTR exon 9 (ESS) RNA using ITC (isothermal titration calorimetry) and NMR (Nuclear magnetic resonance) experiments to reveal the mechanism of the aberrant alternative splicing of the CFTR exon 9.

This study is a part of a larger project aimed at investigating the formation of an extended complex composed of TDP-43 and the hnRNA1 proteins interacting with each other through their respective C-terminal domain (glycine-rich domain). Together, they would form an extended complex by interacting through their C termini, thus preventing access to the 3'ss and blocking the spliceosome, leading to exon skipping (4) (5).

In the first part of the thesis, cystic fibrosis will be briefly introduced. Next, the process of alternative splicing will be introduced. In the second chapter, the interaction of proteins with RNA sequences will be introduced, followed by the third chapter describing

the methods used to study RNA-protein complexes. The fourth chapter describes experimental methods. Then, the fifth chapter will present results from complexes of interest involved in the regulation of gene expression leading to cystic fibrosis, as well as results from STAU1 complexes interacting with ARF1. The results will be discussed, followed by the conclusion.

1.1 The cystic fibrosis

Cystic fibrosis is the most common genetic disease in Caucasian populations, with an autosomal recessive mode of transmission. It mainly affects the respiratory, gastrointestinal, glandular, and reproductive systems. The major clinical symptoms are airway obstruction, inflammation, and an increased risk of developing a chronic infection, leading to lung tissue destruction. Although patients initially suffer from intestinal symptoms, irreversible lung damage subsequently develops leading to life-threatening diseases.

Cystic fibrosis is estimated to affect more than 75,000 people worldwide, including 42,000 in Europe. Caucasian population mainly affected by cystic fibrosis with an incidence of approximately 1 in 3,000 – 4,000 live births (6). Cystic fibrosis is infantile disease and approximately equal numbers of men and women. Although treatments are available to maintain and improve the quality of life of patients, there is currently no cure.

1.1.1 Localization of the CFTR gene

The CFTR (Cystic Fibrosis Transmembrane Regulator) gene codes for the CFTR protein, which is a channel across the cell membrane. The channel regulates functions such as ion transport therefore ensuring the proper functioning of organs. The CFTR gene consists of 27 exons that extend over a length of 230 kb and are located in gene locus 7q31.2 on the long arm of chromosome 7 (7). The mRNA is 6,5 kb and is translated into a protein of 1480 amino acids with a molecular weight of 168 kDa (8).

1.1.2. Structure and regulation of channel opening

The CFTR protein, belongs to the ABC (adenosine triphosphate -binding cassette) superfamily characterized by a unique domain composed of 5 domains including two hydrophobic domains MSD1 and MSD2 (membrane-spanning domains) across the membrane. Each domains encompassed six alpha helixes transmembrane segments, two hydrophilic nucleotide-binding domains NBD1 and NBD2 and a regulatory domain (9).

MSDs are involved in transmembrane channel formation, while NBDs are essential for channel activity, regulating CFTR channel opening and closing. The MSD1-NBD1 and MSD2-NBD2 domains form a heterodimer crucial to CFTR protein function. The regulatory domain is a unique structure of the CFTR and contains the phosphorylation sites required for channel activity (10) (Fig. 1).

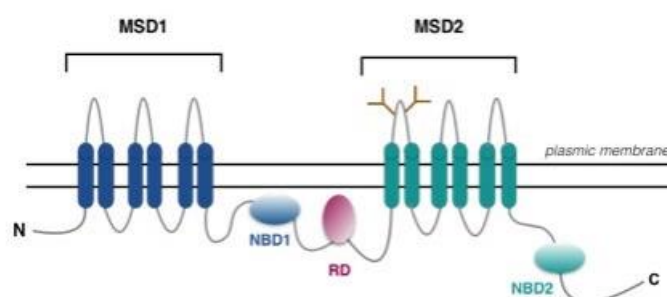


Figure 1: Schematic representation of the CFTR (Cystic Fibrosis Transmembrane Regulator) protein (11). MSD1 (dark blue) and MSD2 (dark green) are trans-membrane domain, NBD1 (light blue) and NBD2 (light green) are ATP nucleotide-binding domain and RD is a regulatory region playing a key role in the phosphorylation for the channel activity (red).

1.1.3 Mutations cause Cystic fibrosis

Multiple mutations have been identified in the CFTR gene leading to cystic fibrosis. Currently, more than 2000 CFTR mutations have been described providing a broad range of phenotypes. Most of CFTR gene mutations are missense mutations (38,9 %), followed by frameshift insertion and deletions mutations (16,1 %), splicing (11,1 %), nonsense (8,4 %), in-frame insertion and deletion (2,1 %), large insertions or deletions (2,8 %), promotor (0,9 %), sequence-specific (6,7 %) and unknown mutations (13 %) (12). These mutations can be classified into six classes of CFTR mutations, defined on the basis of impaired function.

Class I, II and III mutations have been identified as having a severe phenotype due to the absence of CFTR protein at the apical membrane of airway epithelial cells, while individuals in classes IV, V and VI have been reported to have a mild pulmonary phenotype and pancreatic sufficiency due to partial production of CFTR activity (13).

Mutations in both copies of the CFTR gene lead to the disease and result in the absence of a functional CFTR protein, or produce a dysfunctional protein that decreases the level of chloride secretion transported to the apical membrane, producing an abnormal viscous mucus and thus a risk of chronic infection. As a result, the mucus-induced

obstruction and inflammation could lead to a reduction of lung function, or damage the reproductive and gastrointestinal systems (13).

1.2 Mechanism and regulation of constitutive and alternative splicing

The concept that genes are separated by non-coding regions (introns), which are removed during RNA processing, was first introduced by Walter Gilbert in 1978 (14). Alternative splicing was then confirmed experimentally, explaining the high number of proteins translated relative to the number of genes found in the genome (15).

1.2.1 The splicing mechanism

The RNA splicing is a process that retains coding sequences and removes non-coding sequences from the RNA transcript. The retained regions are known as exons, and the removed regions are known as introns. All the exons are ligated together, and introns are removed to transcribe a mature pre-mRNA.

The splicing process involved two consecutive trans-esterification reactions. Each intron and exon junction has specific and highly conserved sequences referred as consensus sequences. The spliceosome is a complex macromolecular machine involved in the RNA splicing process and capable of recognizing these consensus sequences.

At the 5' splice site, each intron starts with the acceptor site nucleotide: GU. At the branch point, the adenine nucleotide required for the splicing process is followed by a sequence of several pyrimidines known as the polypyrimidine tract. The donor site, the dinucleotide AG, is located at the end of each intron at the 3' splice site (Fig 2).

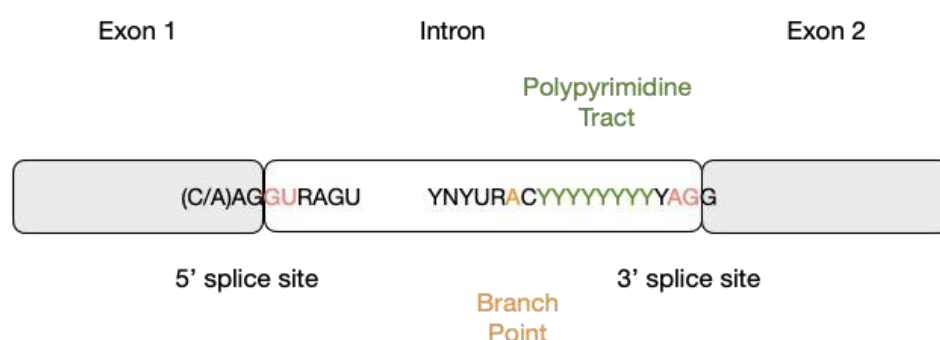


Figure 2: Exon and Intron boundary at the 5' and 3' splicing site. The junction between exons and introns is represented at the 5' and 3' splice sites. The donor site GU is represented at the 5' ss and the acceptor at the 3'ss (pink) followed by the branch point adenosine (orange) and then the YNYURAC motif (Y=Pyrimidine; R= Purine in IUPAC

ambiguous DNA notation) pattern is followed by a pyrimidine sequence named polypyrimidine tract (green).

The pre-mRNA interacts with the spliceosome made up of five small nuclear ribonucleic proteins (snRNPs) named U1, U2, U4, U5, U6, as well as splicing factors. First, the small nuclear U1 interacts with the 5' GU splicing site of the pre-mRNA, splicing factor 1 (SF1) and U2AF65 bind to the branch point site and the polypyrimidine tract. Then, U2 snRNP interacts with the "branch point" adenosine moving away SF1. Then, the tri-snRNPs, composed of U4 snRNP, U5 snRNP, and U6 snRNP, are added to form the pre-catalytic B complex. At this stage, U4 snRNP is removed, allowing U6 snRNP to replace snRNP U1 at the 5' splice site forming the activated spliceosome known as complex B. Thus, U6 snRNP and U2 snRNP interact together and bring the 5' splicing site closer to the branch point adenosine. Then, U5 snRNPs bring the two exons together to ligate both of them and release the mature mRNA as well as the lariat (16) (17).

1.2.2 Modes of alternative splicing

The splicing process binds exons and removes introns to form mature mRNA. The alternative splicing process can produce a broad range of mature RNA variants. Multiple splicing variants with different exon arrangements are transcribed and then translated into different proteins. The human genome generates approximately 90,000 different types of proteins from approximately 25,000 genes. Studies have reported that more than 95% of the human genome undergo alternative splicing (18).

There are different alternative splicing modes, as follows:

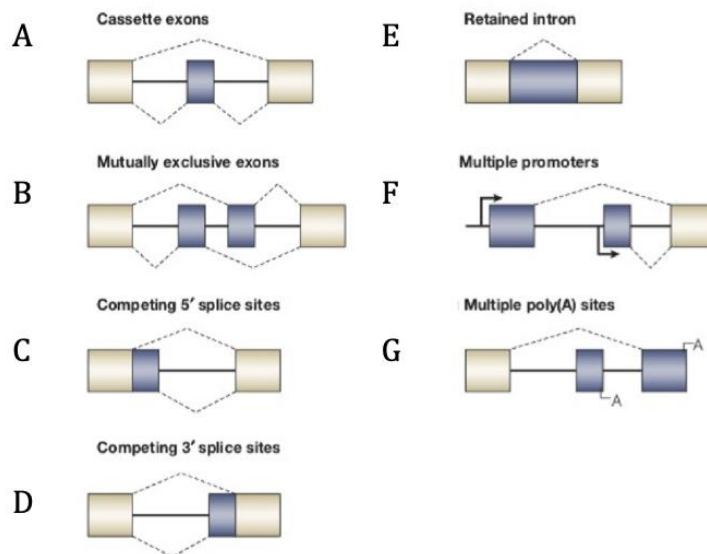


Figure 3: Types of alternative splicing (19). The exons are represented in grey and the alternative ones in blue. (A) Cassette exon (B) mutually exclusive exons (C) alternative 5' splice site (D) alternative 3' splice site (E) intron retention (F) Multiple promoters (G) Multiple poly (A) sites.

- **Cassette exons / Exon Skipping:** It is the most common alternative splicing mode. This process involves either the skipping or retention of the exon from the primary transcript in the mRNA.
- **Mutually exclusive exons:** The two exons are not included both together. In this case, one exon is retained, and the other is excluded. This process introduces additional complexity and generates exon arrangements in mature transcripts.
- **Competing 5' splice sites:** The 5' splicing-site selection is a modification of the exon flanking that creates a new splice site that can compete with the original one.
- **Competing 3' splice sites:** The 3' splicing-site selection is a modification of the exon flanking that creates a new splicing site that can compete with the original one.
- **Retained intron:** The process can conserve the pre-mRNA intron. The intron located within the coding region could code for amino acids in frameshift with the next exon, encode a stop codon, or even result in an incorrect reading frame leading to a nonfunctional protein (19).

Alternative splicing can be also combined with transcriptional modifications such as:

- **Multiple promoters:** The initiation of transcription at different promoters called alternative promoters generates different mRNAs.
- **Multiple poly(A) sites:** Splicing can lead to the use of different polyadenylation sites (19).

1.2.3 Regulation of Alternative splicing

In eukaryotes, multiple genes are alternatively spliced into several isoforms with distinct functions. Alternative splicing is the process involving the recognition of splice sites. The introns can reach 5kb on average and contain various "decoy" splicing sites and often form pseudo-exon but are rarely spliced from the pre-mRNA. In fact, the recognition of the splicing site is complex and depends on several factors such as the strength of the splicing sites, the presence of trans-acting factors, and cis-acting elements or secondary structures at the RNA level. Alternative splicing is a controlled and high-fidelity process, but a deficient regulation is usually associated with human diseases (20).

1.2.3.1 Cis and Trans regulatory elements

The two types of regulatory elements are called cis-regulatory elements and trans-regulatory elements. The motifs promoting splicing are called "activators," while those inhibiting splicing are called "silencers". Cis-regulators can act as enhancers or silencers while trans-acting splicing factors can activate or suppress the splicing process on the pre-mRNA.

Cis-regulatory elements are small variable sequences within the pre-mRNA that can define which exons are included or spliced. Cis-regulatory elements can be weak and, require multiple copies. These sequences are located either upstream or downstream of the regulated exon. Depending on the interactions with trans-acting splicing factors, these sequences may have either an activating or repressive effect. Cis-regulatory elements are named according to their location, depending on whether their function is to activate or silence exon or intron splicing: exon splicing enhancer (ESE), intronic splicing enhancer (ISE), exonic splicing silencer (ESS), and intron splicing silencer (ISS) (21) (22).

The enhancers motifs are often bound by SR (Serine Arginine rich) proteins. The SR proteins contain one or two RRM at their N terminal domain that interact with the pre-mRNA and a C-terminal domain, including various arginine and serine residues, that functions as a protein interaction domain (21). The SR protein has a positive effect on splice site recognition and activate the spliceosome assembly. In fact, these proteins can interact with the U1snRNP at 5' ss as well as with the U2AF35 subunit of the

heterodimeric factor U2AF at 3'ss. In addition, the SR domain of the second part of the U1AF65 subunit binds to both SF1 and the pyrimidine tract to stabilize and facilitate the recognition of the branch point and the polypyrimidine tract defining the boundary of exons (23).

The silencers motifs interact with negative trans regulatory elements such as hnRNP proteins (heterogeneous nuclear ribonucleoproteins). The hnRNPs proteins has one or more RRM domains and glycine-rich domains that mediate protein-protein interaction (24). The hnRNPs proteins can bind an ESS sequence ensuring the suppression of alternative exon splicing (25). For instance, a study shows the repression of tat23 splicing by hnRNP A1 is mediated by ESS3, which prevents the SR protein from promoting splicing [18]. In addition, the sequence required for the interaction of hnRNP A1 to the ESS3 has been identified as the motif UAGUGAA (26) which resembles the consensus high-affinity hnRNP A1 SELEX motif, UAGGGA/U (27).

Interaction between cis regulatory elements and trans regulatory elements at the weak splice site allows or prevents the formation of the spliceosome assembly. The alternative exon is either included or excluded, depending on the presence of multiple splicing factors, the concentration of these splicing factors in the cells, and the competitive interaction between SR activators and hnRNP inhibitors. The selection of the alternative splice site also depends on the stoichiometry and interactions with activators and inhibitors of splicing, as well as the steric conformation allowing or preventing the availability of splice sites (22).

1.2.4 Aberrant splicing in exon 9 causing cystic fibrosis

Studies report that monosymptomatic forms of cystic fibrosis often present a specific allele at the polymorphic intron 8-exon 9 junction of the CFTR gene. It has been found that the T5 allele is associated with alternative splicing of exon 9 of the CFTR gene, leading to CFTR gene transcription that excludes exon 9. Furthermore, studies show that normal populations harbor polymorphic repeats (TG)_n ranging from 9 to 13, followed by a T repeat such as T5, T7 or T9 in healthy patients. Therefore, the T5 allele is present in both healthy patients (with CF) and those with the disease. It has been suggested that the T5 allele is a pathological mutation with incomplete penetrance and could be modulated by the presence of other mutations and or by the number of polymorphic repeats (TG)_n (28).

Mutations in the CFTR exon 9 sequence encode part of the first nucleotide-binding domain, resulting in skipping of exon 9 (29). The NBD domain is dysfunctional, resulting in a non-functional CFTR protein. Besides, studies show a correlation between the number of TG repeats and the T5 allele in the appearance of the disease. Indeed, T5 CFTR genes from patients have a high number of TG repeats, whereas T5 CFTR allele

from healthy patients (with CF) have a reduced number of TG repeats resulting from the exclusion of the CFTR exon 9 of CFTR pre mRNA. However, the amount of exon skipping is not always correlated with polymorphic alleles at the 3' end of intron 9, suggesting that other factors act with the polymorphic locus to regulate the amount of exon skipping 9. In fact, the (TG)_m and (T)_n polymorphic repeats at the 3'ss end of intron 8 of the cystic fibrosis transmembrane regulator (CFTR) gene are associated with the alternative splicing of exon 9 (30).

A study identifies TDP-43 as a regulator of CFTR pre-mRNA splicing, preventing exon 9 recognition. In addition, the study reveals that TDP-43 is capable of binding a (UG)₁₂ RNA sequence with high affinity. They indicate that the length must be at least (UG)₆ repeat motif, and that binding affinity increases with the number of UG repeats. They perform a competition experiment using an oligo (TG)₁₂ repeat and found that TDP-43 can also bind the DNA sequences with a minimal number of (TG)₆ repeat (1).

In addition, a study using an *in vivo* model system performs a functional analysis of the role of the regulatory trans-acting factors hnRNPA1 and members of the SR family on the alternative splicing of CFTR exon 9. Using an *in vivo* model system, consisting of a reporter minigene, the authors have shown that (TG)_m and (T)_n polymorphic repeat are associated in exon 9 skipping as well as the sequence in intron 9 (ISS) able to modulate the alternative splicing. Moreover, the authors evaluated the role of the regulatory trans-acting factors hnRNP A1 and the SR protein on the alternative splicing of CFTR exon 9 and show that, in the case of the aberrant splicing, the SR proteins bind to an ISS element in the CFTR intron 9 and repress the splicing of CFTR exon 9. hnRNP A1 is a protein involved in the splicing and usually has an antagonist effect of the SR protein. In the case of aberrant regulation of CFTR exon 9, the authors reveal that both splicing factors, SR and hnRNP A1 proteins, repress the splicing of CFTR exon 9 splicing (31).

Another study focused on the protein interaction of the C-terminus of TDP-43 in splicing regulation. They found that TDP-43 is able to interact with the C-terminus of hnRNP A1. Furthermore, using an *in vivo* assay, they discover that the C-terminus of the TDP-43 region interacting with hnRNP A1 is required for splicing inhibitory activity (4) (5).

Further investigation report that two splicing factors, TIA-1 and PTB, play a role in alternative splicing of CFTR exon 9. TIA-1 induces exon inclusion while PTB induces exon skipping (Ref). TIA-1 binds to a rich poly pyrimidine between the 5'ss and the intronic silencer. The study also reports that TIA-1 recruits U1 to the 5'ss leading to exon inclusion and thus able to mediate alternative splicing of exon 9 (32).

2 RNA- Proteins interaction

It is essential to understand the specific binding regions within RNA-binding domains (RBDs) with RNA to elucidate the mechanisms underlying these interactions. X-ray crystallography and nuclear magnetic resonance provide insights into the structure of amino acids and nucleotides that interact in protein-RNA complexes, providing data on RBP domains in complex with RNA. Analysis of these data led to the identification of the intermolecular interaction types and preferred amino acids that characterize the specific binding between proteins and RNA.

2.1 RNA binding proteins

RNA Binding Domains (RBDs) are conserved structural motifs within RNA-binding proteins involved in specific interactions with RNA molecules. These domains, such as the RNA Recognition Motif (RRM), K Homology (KH) domain, and Zinc Finger motifs and are one of the most abundant protein domain in eukaryotes (33). These domains recognize sequence or structure specific elements in target RNAs and are essential for, splicing, processing, transport and localization (34).

2.1.1 RRM (RNA Recognition motif)

RNA recognition motifs (RRMs) constitute the largest group of RNA-binding proteins in the eukaryotic RNA recognition motif family. RRM is composed of 80 to 90 amino acids and adopt a β -sheet and α -helix structure, forming a binding interface capable of recognizing and interacting with RNA. RRM consists of four antiparallel β strands and two α helices arranged in a β - α - β - β - α - β fold with side chains that stack with the RNA bases. Those contain two conserved sequences of eight and six amino acids, known as RNP1 and RNP2(34).

2.1.1.1 The hnRNP family

Ribonucleoproteins (RNPs) assemble on pre-mRNA in the nucleus of eukaryotic cells, among this RNPs are the hnRNP proteins. These proteins play a crucial role in the regulation of messenger RNA (mRNA), including transcription, RNA maturation, transport and degradation. hnRNPs are also involved in post-transcriptional RNA processes, which are essential for the regulation of gene expression and protein production.

The hnRNPs (heterogeneous nuclear ribonucleoproteins) belong to a family of more than 20 types. The most abundant types are the hnRNP A to U, with molecular weights ranging from 34 kDa to 120 kDa (24). RRM is the most common RNA-binding

domains, containing approximately 90 amino acids and capable of interacting with nucleic acid (35).

Other common RNA-binding domains (RBDs) include RGG boxes, which are Arg-Gly-Gly repeats with interspersed aromatic amino acids, and KH domains (35) and a glycine-rich domain that can mediate protein-protein interactions as shown with the C terminus of hnRNP A/B proteins interacting with TDP-43 (5). These proteins interact with several cellular factors and play a key role in the regulation of gene expression, including alternative splicing (36), nuclear and export from the nucleus to the cytoplasm (37) telomere maintenance (38) mRNA stability and turnover (39), mRNA processing and translation regulation (24).

As for the hnRNP A/B proteins, those have a similar general structure containing two RRM domains which are highly conserved and involved in interactions with nucleic acids and a glycine-rich auxiliary domain at the carboxyl terminus that mediates protein-protein interactions (40). Although hnRNPs B2 and AB are more distantly related, they also belong to this subfamily. For instance, the RRM domains of hnRNPs A2 and B1 have around 80% of amino acids in common with the RRM domains of hnRNP A1, while the glycine-rich auxiliary domain is much less similar, sharing only 30 % (24).

The hnRNP A1 protein belongs to the hnRNP A/B subfamily and is composed of four paralogues known as A1, A2/B1, A3, A0 (24). The A/B family hnRNPs show a high degree of amino acid sequence similarity, due to the presence of their similar structural motifs. Together, hnRNP A1 and A2 have been identified as representing around 60 % of hnRNP mass and play an important role in the biological roles of hnRNPs (41). The A1 and A2 hnRNPs have a similar sequence of around 80 % and 58 % in the RRM domains and glycine-rich regions, respectively. The two RRM domains of hnRNP A3 have high sequence similarity to hnRNP A1, although its glycine rich-domain is more similar to that of A2 than A1 (40). In addition, it has been suggested that hnRNP B2 may be an alternatively spliced isoform of hnRNP A1 or A2 (42).

2.1.1.2 Structural insight of hnRNP A1

The hnRNP A1 is among the most abundant and widely expressed nuclear proteins (43). Two transcripts have been identified: hnRNP A1-B corresponding to a weight of 38 kDa and composed of 372 amino acids, and hnRNP A1-A which is the shorter version composed of 320 amino acids with a molecular weight of 34 kDa. The structure of the human hnRNP A1 consists of an N terminal proteolytic fragment containing two RRM domains connected by a linker and a flexible C terminal domain composed mostly of glycine, arginine, and serine residues (42). The two RRM domains are very similar but their function is non-redundant (44) from each other and both domains are required for the

alternative splicing activity. Using mutations of RRM1 or RRM2 it has been shown that deletion of each RRM abolishes the activity of the protein (44).

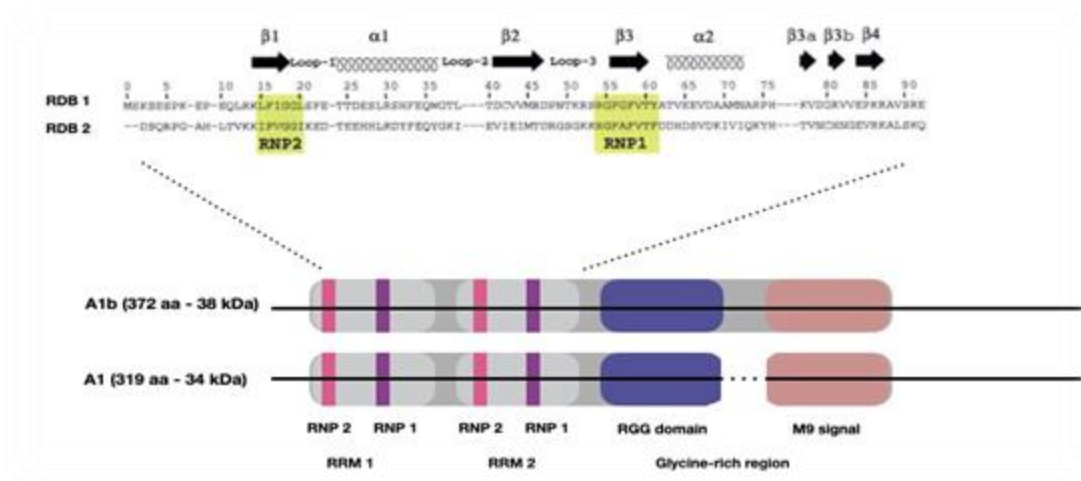


Figure 4: Structure and characteristic of hnRNP A1b and A1 (40) (45).

The C-terminal domain is composed of a rich tripeptide repeat sequence Arginine-Glycine-Glycine (RGG) including the insertion of tyrosine and phenylalanine residues, which has been identified as capable of mediating protein-protein interaction as described on the hnRNP A2 binding the C-terminus of TDP-43 (4) (5). The glycine-rich region also contains an RNA-binding domain and a nuclear shuttle/export (NS) domain called M9 (46). The latter has been found mainly in the nucleus and is capable of shuttling to the cytoplasm and aggregating in the cytoplasm (47).

Studies reveal that in two structures, obtained by X-ray crystallography at resolutions of 1.75 Å (48) and 1.9 Å (49), show that the two RRMs are monomers. UP1 has two interacting domains comprising two highly conserved domains such as RRM1 (residues 15-89) and RRM2 (residues 106-180). The RRMs interact with the DNA in an antiparallel manner, stabilized by two salt bridges. Each RRM consists of four β -strands and two α -helices, forming a $\beta 1 \alpha 1 \beta 2 \beta 3 \alpha 2 \beta 4$ motif, as shown with U1A reference by NMR (50). Both RRMs includes two highly conserved ribonucleoprotein consensus sequences, RNP1 and RNP2. The RNP-2 hexamer is located on $\beta 1$ strand, while the RNP-1 octamer is on $\beta 3$ strand. These two RNPs motifs are connected by a linker of 17 residues. Studies show helices are on the same side of the β -sheet, which stabilizes each RRM through hydrophobic interactions. Moreover, both RRM1 and RRM2 are oriented in opposite directions. Additionally, the structure features a 3_{10} helix upstream of RRM1 and a β -turn at the C-terminus of each RRM. The RNP sub-motifs within each

RRM play a crucial role in RNA binding, which explains the high conservation of these sequences.

Indeed, the four β -sheets form an extended surface with the two α -helices all located on the same side of this surface. RRM1 and RRM2 are linked by a loop that appeared flexible in the crystal structure (45). Studies show that residues in the $\alpha 2$ - $\beta 4$ - $\alpha 2$ region facilitate contacts between the two RRMs and can interact through their two pairs of arginine-aspartic acid salt bridges (Arg75 with Asp155, and Arg88 with Asp157). Both Arg75 (in RRM1) and Lys166 (in RRM2) are followed by a highly conserved tyrosine residue, common among hnRNPs A and B (40), which helps to form a turn. It has also been demonstrated that water-mediated interactions play a key role in stabilizing the packing of the two RRMs, while conserved hydrophobic residues do not significantly contribute to inter-RRM contacts. Furthermore, the linking sequence connecting the two RRMs is flexible, with six disordered residues in the middle.

Following the determination of the structure of the free hnRNP A1 protein, studies were performed to better understand the interaction between proteins and nucleic acids.

SELEX is a selection and amplification methods used to identify a specific sequence that binds to both RRM domains. The UAGGGA/U sequence was found to be a high-affinity RNA sequence recognized by both RRMs of the hnRNP A1 protein. The human telomeric repeat (TTAGGG)_n is identical to the high-affinity sequence identified using SELEX (27). hnRNP A1 probably interacts in the same manner with telomeric DNA and high-affinity RNA sequences and a subsequent study investigate the mode of interaction (48).

The authors solve a structure at a high resolution of 2.1 Å using X-ray crystallography (45). It shows that RRM1 and RRM2 crystallize as a dimer when they bind to the 12 nt ssDNA, d(TTAGGGTTAGGG)(PDB ID 1U1P). The two ssDNA sequences bind to each RRM antiparallel, with the 5' end of the DNA sequence binding to RRM1 and the 3' end binding to RRM2. The authors show that a 15° rotation changes the orientation of the two RRMs, bringing the β -sheets closer together to facilitate the interaction with DNA, compared with the crystal structure of the free RRM (45). This rotation occurs around the two pairs of Asp-Arg salt bridges, equally conserved in the free and bound structure. Here, the disordered linker becomes ordered, forming an α -helix interacting with DNA. The main difference between RRM1 and RRM2 interacting with DNA has been shown in $\beta 2$ strand.

It is worth noting that the presence of the dimer can be explained by packing forces in the crystal (51). While the interaction with DNA has been resolved by crystallography, the nature of the interaction with nucleic acids in solution remains unclear.

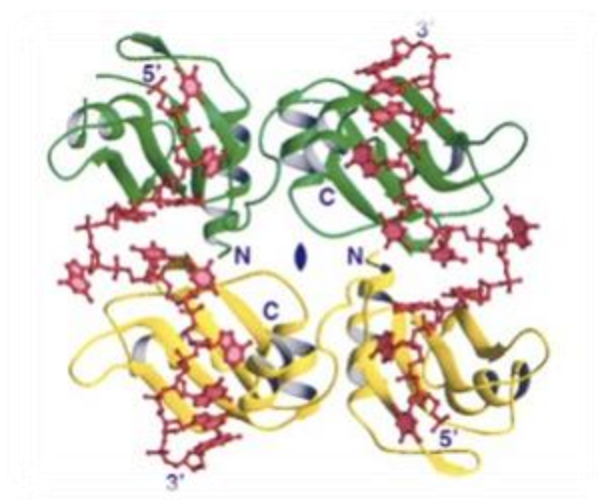


Figure 5: Crystal structure of hnRNP A1 upon binding to d(TTAGGG)₂ (**48**) (PDB ID: 1U1P). Protein molecules are shown in yellow or green as ribbons and DNA in red. RRM1 binds 3'DNA while RRM2 binds the 5' ssDNA. Both RRMs interact with the DNA in an antiparallel mode.

Solution studies using NMR spectroscopy, including a segmental labeling strategy (51), were performed to investigate potential interdomain interactions of UP1 in solution. In the study, RRM1 was ¹³C/¹⁵N-labeled, while RRM2 remained unlabeled. Several interdomain NOEs cross peaks characteristic of contacts between the two RRMs were identified. After analyzing the data, all NOEs cross peaks were converted into long-range inter-proton distances and used to determine the solution structure of UP1. Additional NOEs cross peaks comprising those interdomain NOEs were used unambiguously assign using the segmental UP1 domains. The study confirmed that UP1 contains two well-defined RRM domains connected by a flexible linker. The N and C termini and the β2-β3 loop in each RRM display a flexibility, most pronounced in RRM2. The structure of RRMs is very similar to the crystallographic structure (45), except for minor local variations of the β2-β3 loop between the solution NMR and crystal structures.

It has been shown that the two RRM domains of the hnRNP A1 protein adopt a well-defined orientation in solution and remains unchanged upon binding to DNA binding. It also explains that the domain movement observed when comparing the two crystal forms of the protein is not related to nucleic acid binding, but rather to differences in protein-protein interactions between adjacent proteins in the two distinct crystal lattices.

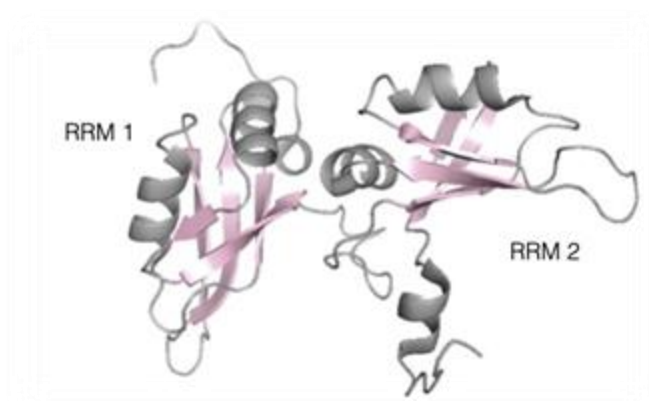


Figure 6: NMR structure of hnRNP A1 (**51**) (PDB ID 2LYV). RRM1 and RRM2 are represented (grey) including β strand (light pink) and α helices (light grey).

Recent studies have been performed using NMR spectroscopy technique. It has been shown each RRM interaction with a small RNA sequence. Another study used the NMR workflow to obtain information and then creates a model of the two RRMs interacting with an RNA sequence. A study performed by Allain and colleagues (52) has identified optimal recognition motifs that bind the individual RRMs.

The study investigates the individual RNA recognition motifs (RRMs) of hnRNP A1 through NMR titrations with short RNA sequences (6–8 nucleotides), each containing a central AG dinucleotide crucial for hnRNP A1 binding. For RRM1, binding was observed with 5′-UUAGGUC-3′, and for RRM2 with 5′-UCAGUU-3′. These RNA sequences differ slightly from the SELEX high-affinity sequence (5′-UAGGGA/U-3′) and the single-stranded DNA motif found in telomeric repeats (48).

The solution structure of hnRNP A1 RRM1 bound to 5′-UUAGGUC-3′ reveals that the β -sheet surface of RRM1 and the inter-domain linker residues bind five nucleotides (U1 to G5), with four of these specifically recognized through intermolecular hydrogen bonds. The structure is similar to crystal structures of UP1 bound to telomeric DNA repeats (48). The central A3G4 dinucleotide stacks onto Phe17 and Phe59 of the RNP1 and RNP2 motifs and forms hydrogen bond with Arg88 and backbone bond Val90. Likewise, the study shows a solution structure of hnRNP A1 RRM2 in complex with 5′-UCAGUU-3′. The β -sheet surface of RRM2 and the C-terminal residues bind three nucleotides (C2 to G4), with two nucleotides recognized sequence-specifically through intermolecular hydrogen bonds. This structure is similar to the UP1 RRM2 complex bound to (5′-TTAGGG-3′)₂ (48). The central A3G4 dinucleotide interacts with Phe108 and Phe150 of the RNP motifs and forms hydrogen bonds with Lys179 and Leu181. Nucleotides U5 and U6 at the 3′ end do not interact with RRM2 and are highly

disordered. The study suggests that the optimal RNA-binding motifs for RRM1 and RRM2 are 5'-YAGG-3' and 5'-YAGN-3', respectively, where Y is a pyrimidine and N is any nucleotide.

The investigation of the interaction of ISS-N1 RNA with UP1 and its complex using NMR spectroscopy was performed. After optimizing the complex with minimal mutations to either the RNA, protein, or both, NMR titration results revealed that RRM1 binds to the 3'-motif of ISS-N1 independently of RNA point mutations, while RRM2 binds to the 5'-motif.

Using paramagnetic relaxation enhancement (PRE) experiments, the study provides independent evidence that UP1 binds to ISS-N1 via both RRMs, with RRM1 binding to the AG 3' motif and RRM2 binding to the AG 5' motif.

The authors show a structural model that contrasts with the crystal structure of hnRNP A1 RRMs bound to telomeric DNA repeats, where hnRNP A1 dimerizes and binds to two distinct DNA oligonucleotides. In this structure, RRM1 binds to the 5' end of one DNA molecule, while RRM2 binds to the 3' end of another DNA strand (48). In their model, the two RRMs adopt an antiparallel orientation, requiring the RNA molecule to loop around both RRMs. A minimal linker length of four nucleotides is sufficient to support this binding mode. In the case of ISS-N1, the RNA linker between the two AG motifs is nine nucleotides long. In this study the authors confirm that this disruption significantly affects hnRNP A1 function, highlighting that the binding of both RRMs to RNA is crucial for the efficient repression of exon 7 inclusion.

Another study investigated the interaction between hnRNP A1 to an RNA target, namely the pri-mir-18a. This interaction involves both RRMs and a region encompassing two 5'-UAG-3' motifs (53).

Our study focuses on the RNA sequence 5' CAGGGAUUUGGGGAC 3'. The motifs highlighted in the interaction with RRMs1 and 2 are present in this sequence. Specifically, the motif 5' CAGGGAUUUGGGGAC 3' corresponds to the wild-type version, while for the mutant, the sequence is 5' CAGGGAUUUGGGCAC 3'. The preferred motif of RRM1, 5'-UAGGU-3' (52), is found in the CFTR 3' ssRNA sequence, but with an inverted orientation in 3' ssRNA 3' UAGGU 5'. The preferred motif 5'-CAG-3' (52), is also found but at both ends of CFTR 3' ssRNA sequence. The use of mutant 1 removes the preferred 3' motif, as guanine has been replaced by cytosine. In this case, the two sequences are very similar, allowing each to interact with both RRM, which raises the question of whether this binding mode is applicable to our sequence.

Additionally, a key question that arises is whether the binding mode observed in our study is consistent with or differs from previous studies, particularly due to the

repetition of GGG residues in our sequence or the redundancy present within the 3' ss CFTR RNA sequence.

We used a ssDNA sequence that mimics the ss RNA for cost reason. The impact of this modification remains unclear, especially considering that crystal structures have shown that binding to telomeric DNA induces dimerization (48). However, as demonstrated by (51), dimerization is likely due to packing forces from crystallography and should not result in large changes. We will therefore study the interaction of RNA 5' CAGGGAUUUGGGGAC 3' (or the mutant 5' CAGGGAUUUGGGCAC 3') with UP1, followed by interactions with small bipartite ss DNA fragments 5'-CAGGGAT-3' and 5'-TGGG-GAAT-3' to attempt to address these questions.

2.1.1.3 Structural insight of TDP-43

TDP-43 (TAR DNA-binding protein 43) is a member of the hnRNP family and play also a role in RNA processing and the regulation of the alternative splicing of the CFTR exon 9. TDP-43 has been identified as the main inhibitor of exon 9 skipping, interacting with UG repeats at the 3' splice site (54).

The protein TDP-43 is composed of 414 amino acids and has a molecular weight of 43 kDa. This protein is composed of four part included a N-terminal domain (NTD) from 1-76 residues and it has been reported to form a dimer or oligomer (55), two RRMs domains (RRM1: residues 106-176; RRM2: residues 191-259) and a C-terminal domain (CTD: residues 274-414) divided into four sub-regions Ser-rich (GaroS) segment, a hydrophobic segment, a Gln/Asn (Q/N) rich segment and a second (GaroS) segment. The C terminal domain is involved either in protein-protein interactions or initiates its self-aggregation (56). The protein contains also sequences like a nuclear localization signal (NLS, residues 82–98) and nuclear export signal (NES residues 239–250) (Fig 7) (57).

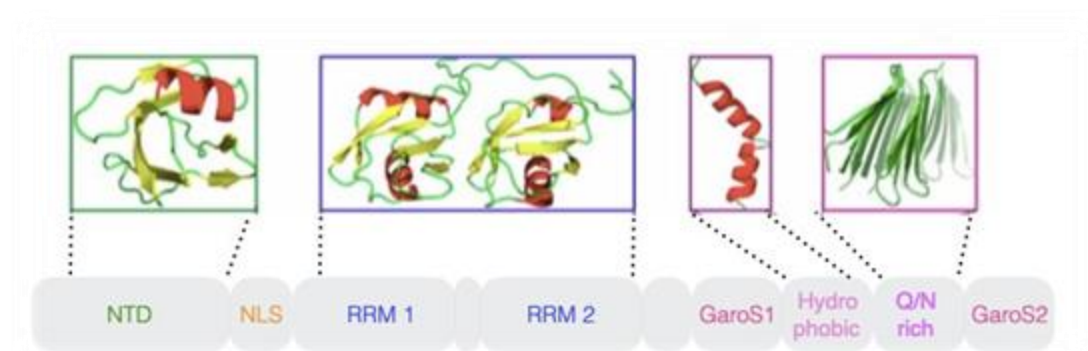


Figure 7: Structure of TDP-43 showing the variant domains (57). The two RRM s are composed of four β strands sheets and two α helices with a $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ fold organized into two highly conserved sequences named RNP1 and RNP2.

The NMR structure (PDB ID 3B82) of the two RRM s interacting with the RNA (5'-GUGUGAAUGAAU-3') shows that both RRM s contain a canonical $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ fold with an additional β -hairpin ($\beta 3'\beta 3''$) inserted in between the helix $\alpha 2$ and $\beta 4$. This additional β -hairpin extends the surface of the β sheet being the platform able to interact with RNAs. The RNA binds the protein with a specific orientation ($\beta 2\beta 4$ type). Indeed, RRM1 binds the 5' of the RNA while the RRM2 interacts with the 3' of the RNA. The interaction occurs via the positive charges from the groove of the protein that is composed of the β -sheet surface of the RRM s, the linker, and the first residues of the C-terminal (3).

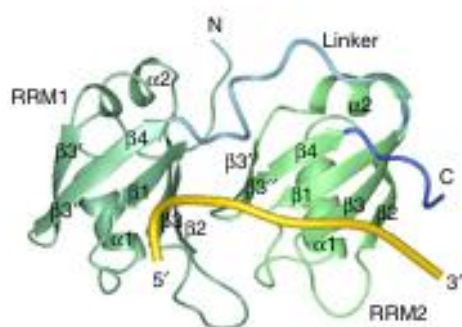


Figure 8: NMR structure of RRM1 and RRM2 of TDP-43 binding 5' GUGUGAAUGAAU 3' RNA sequence (3). RRM from the protein molecules are shown in grey and pink (PDB ID 3B82) (3).

We used the extended complex composed of two copies of hnRNP A1 and two copies of TDP-43 binding the longer UG-rich single-stranded RNA sequence with both C-termini of each protein being removed to simplify the study. Both C termini have been shown to mediate protein-protein interaction (5). Use of these constructs simplifies the study, but further investigation will be required to elucidate the potential role of the two C termini in the context of the extended complex binding to the 3' splice site region in the aberrant splicing of exon 9.

2.1.2 dsRBD: double stranded RNA Binding Domain

Double-stranded RNA-binding domains (dsRBDs) are conserved structural motifs composed of 65 to 70 amino acids that play a role in various biological processes, such as antiviral response, RNA editing, RNA processing, RNA transport, and, last but not least, RNA silencing. dsRBDs interact specifically with double-stranded RNA sequences, typically defined based on structural features rather than sequence similarities. The dsRBD domain adopts an $\alpha\beta\beta\alpha$ topology with a characteristic binding surface that interacts with the A-helix of RNA (34).

2.1.2.1 Staufen interaction with the ARF1 dsRNA

As part of my secondary project, I contributed to the Staufen project. First, I'll briefly introduce the Staufen protein, then describe the results, including a section on the mutant effect I performed alongside the postdoc.

Staufen homolog 1 (STAU1) is a dsRNA-binding protein involved in the transport and localization of mRNA in different subcellular compartments and organelles. This protein also plays a role in the control of translation and in STAU1-mediated mRNA degradation (SMD), a process of mRNA degradation in mammalian cells. Two homologs have been identified, STAU1 and STAU2, which differ in the number of RNA-binding domains (dsRBD). STAU1 lacks dsRBD1, while STAU2 has a truncated dsRBD 5. Both have the 2, 3 and 4 dsRBDs as well as a tubulin-binding domain (TBD) and an exchange domain (SSM). The dsRBDs 3 and 4 adopt a canonical $\alpha\beta\beta\alpha$ motif separated by a linker of 34 aa (58) (59). Studies showed that helix $\alpha 1$ and loop $\beta 1\beta 2$ interact at the minor groove of dsRNA target while helix $\alpha 2$, which is highly conserved, interacts with its major groove. In addition, NLS can be directed to the nucleus by STAU, where it has been associated with the regulation of alternative splicing and nuclear export. One study showed that the target dsRNA is both GC-rich and has a secondary structure in its 3'UTR region capable of interacting with the phosphodiester backbone of the target RNA. Whereas STAU1 binds to dsRNAs and facilitates the recruitment of the RNA degradation machinery, SMD enables mRNA regulation using the 3'UTR of STAU (60). This process requires the binding of STAU dsRBD4 and TBD to a key SMD factor upf1. The 19-bp loop within the (human ADP-ribosylation factor1) ARF1 3'UTR mRNA requires

the interaction with STAU1, resulting in down-regulation of ARF1 mRNA levels *in vivo* (61) Both SMD and NMD uses upf1 but are competitive pathways (60).

2.1.2.2 Structural insight of STAU1

Lazzaretti et al crystallized the protein Staufen dsRBD3 and dsRBD4 in complex with ARF1 with a 2.9 Å resolution. The authors reveals how both dsRBD interact with ARF1. Both RBDs dimerize upon binding and recognize the dsRNA A-form involved in electrostatic interactions with the RNA backbone (62). In addition, residues of both dsRBDs make direct contacts with the G and C bases of dsRNA, indicating a sequence specificity in binding. In the crystal structure, two copies of STAU1 dsRBD3/4 bind to ARF1 SBS dsRNA, forming an interface between the two dsRBDs. The structure also showed that one copy of dsRBD4 was partially disordered while the second copy was completely disordered probably due to steric clash in the crystal lattice that may affect the location of the domain alongside the RNA (62).

In order to get more structural information in solution, Yadav et al (63) determined the solution structure of the ARF1 SBS dsRNA - STAU1 dsRBD3 and dsRBD4mplex using high-resolution NMR spectroscopy. The structure show that both STAU1 dsRBD3 and dsRBD4 bind to opposite sides of ARF1 SBS dsRNA without forming a protein-protein interface unlike the crystal structure. STAU1 dsRBD3 and dsRBD4 binds the dsRNA ARF1 SBS bases in the minor groove. In addition, the authors reported that both dsRBDs are required to recognize the target. STAU1 dsRBD4 bind two pyrimidines and one purine using the helix $\alpha 1$ and loop $\beta 1$ - $\beta 2$ bound at the end of the dsRBD4 in the minor groove. Lysines from helix $\alpha 2$ bind to the phosphodiester backbone from the major groove. For dsRBD3, the authors observed the same interactions in the minor groove.

In conclusion, the mutations performed within the protein to study the effect on the binding affinity with the target have a minor effect on the binding affinity with the RNA target while having a significant impact *in vivo*.

3 Methods used to study the interaction of hnRNP A1 with the 3'ss CFTR ssRNA (or ssDNA mimicking the RNA target)

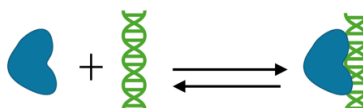
The study of interactions between a protein and RNA is essential for understanding many biological processes, such as gene expression regulation. Nuclear Magnetic Resonance can identify the residues involved in the interaction using titration experiments. Additionally, isothermal titration calorimetry (ITC) provides thermodynamic information about the interaction, such as the binding affinity (K_d), enthalpy (ΔH), and entropy (ΔS) of the complex formed. Combined, NMR and ITC provide both structural and energetic information on the recognition mechanisms between a protein and an RNA.

3.1 Isothermal titration microcalorimetry (ITC)

3.1.1 Basic thermodynamic concept of biomolecular interactions

3.1.1.1 Binding Affinity (K_a) and Gibbs Free Energy (ΔG)

The non-covalent interaction between a macromolecule M and its ligand L, leading to the formation of an ML complex, is described by the following reversible reaction:



The probability of this reaction leading to complex formation is determined by the affinity constant K_a , which is defined as the ratio between the concentrations of the complexed and free forms of the molecules. This constant is described such as:

$$K_a = \frac{[ML]}{[M][L]}$$

The energy exchanges depend on the initial state (free molecules) and final state (complex formed) of the reaction, as well as on the reaction process followed between the two states. The affinity constant K_a is related to the free energy of the interaction (Gibbs energy) which is also expressed as follows:

$$K_a = e^{-\Delta G/RT}$$

$$K_d = 1/K_a$$

T= Temperature; R=8,31 J mol⁻¹ K⁻¹

Binding affinity K_a is a measure of the intensity of the interaction between ligand and receptor. It is often expressed as an equilibrium constant (or dissociation constant K_d).

$$\Delta G = -RT \ln K_a$$

A high value of the constant lead to a higher affinity of the ligand for the protein. The reaction is considered spontaneous when ΔG is negative. This free energy arises from two energetic components: enthalpy (ΔH) and entropy (ΔS), as expressed by the equation:

$$\Delta G = \Delta H - T\Delta S$$

The Gibbs energy (ΔG) is a thermodynamic function that determines if a chemical reaction or physical process is spontaneous or not at constant temperature and pressure.

The value of ΔG can be either positive, negative or equal to 0 at equilibrium. In fact, if the Gibbs energy is:

- **$\Delta G < 0$** : The reaction is spontaneous (the reaction is performed without any external energy)
- **$\Delta G = 0$** : The system is at equilibrium. There are the same number of molecules as ligand. The initial energy is equal to the final energy of the system.
- **$\Delta G > 0$** : The reaction is not spontaneous (it requires an external energy to produce the reaction).

3.1.1.2 The enthalpy contribution (ΔH)

The enthalpy (ΔH) measures the total overall energy of the system. The interaction between a ligand such as RNA and a protein binding site can be associated with the release or absorption of heat which depends on the type of the forces involved.

The main forces contributing to enthalpy (ΔH) are:

- **Hydrogen bonds:** The interaction between polar groups (e.g. between hydrogen atoms and nitrogen or oxygen atoms) These interactions are often significant contributors to enthalpy values.
- **Electrostatic interactions:** The opposite-charge interactions (Coulomb forces) between charged groups on the receptor and ligand may lead to the release of energy.
- **Van der Waals forces:** These interactions are usually weaker but can also play a role in the heat released during the binding.
- **Hydrophobic interactions:** The hydrophobic effect (when non-polar molecules come together in an aqueous environment) often leads to decreased solvent order, which can also result in the release of heat.

The enthalpy can be positive (endothermic) or negative (exothermic), all depending on the type of the interaction. For example, interactions caused by electrostatic forces or hydrogen bonds often result in heat release (exothermic), while hydrophobic interactions can lead to heat absorption (endothermic).

3.1.1.3 The entropy contribution (ΔS)

The entropy of a protein-ligand complex is a measure of the energy disorder in the system when the protein-ligand complex is formed during the interaction.

The interaction of the ligand with the protein can lead to changes in entropy of the protein, the ligand and the solvent that surrounds each of them. The change in entropy provides information on the degree of freedom or order in which the system changes during the interaction.

-**The entropy of the ligand:** The ligand adopts different conformational configurations in solution. When interacting with the protein, the ligand has less degrees of freedom (e.g., degrees of rotational or translational freedom), which can lead to a reduction

in the entropy value of the ligand. This is due to the rigidity that occurs when the ligand binds to the binding site of the protein.

-The entropy of the protein: The conformation of the protein can change to adapt to the ligand, decreasing the flexibility or internal motion of the ligand. As a result, its entropy decreases. However, the interaction between the ligand and the protein can also induce changes in protein dynamics, and then increase entropy, leading to higher flexibility in certain regions of the protein.

-Solvent entropy: The entropy of the solvent (solvation) plays an important role in ligand-protein binding. During binding, the water molecules surrounding the ligand or protein are excluded or reorganized. This process increases the entropy of the solvent, because it is associated with a disorder in the rearrangement of water molecules, especially if the ligand is hydrophobic and drives water out of the system.

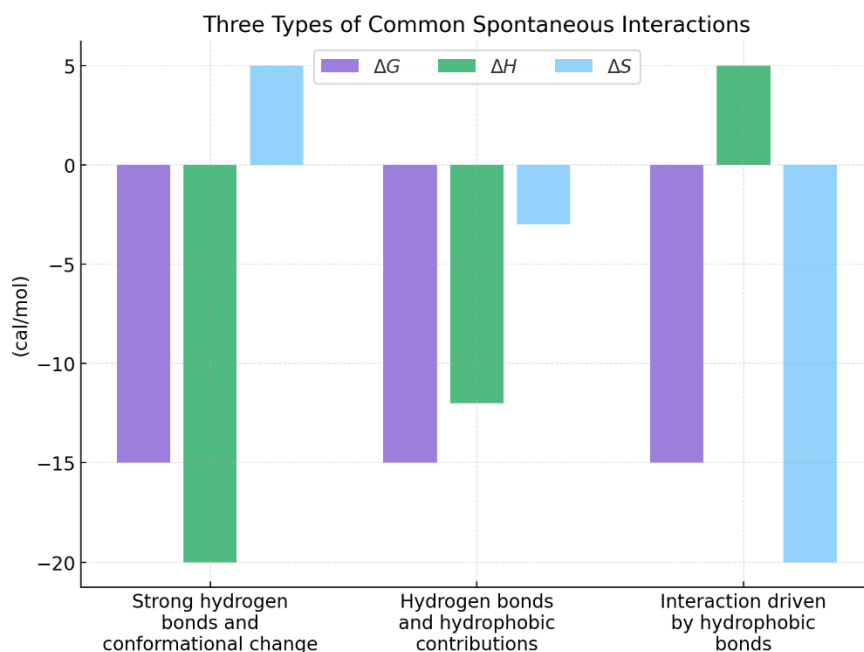


Figure 9: Three most common types of interactions. (Adapted from Isothermal Titration Calorimetry and Drug Design, ITC Application Note, Microcal) (64).

3.1.2 The isothermal titration microcalorimeter (ITC): principle

Isothermal Titration Calorimetry (ITC) has the advantage of being a direct, label-free technique to study the interaction between a protein and a nucleic acid. No chemical

modification is required, ensuring the integrity of the sample. In fact, ITC measures the heat changes associated with molecular interactions in solution which accurately represents the physiological environment. A single ITC experiment provides thermodynamic parameters such as binding affinity (K_a), enthalpy (ΔH), stoichiometry (n), and entropy ($T\Delta S$).

The microcalorimeter contains two separate cells, isolated from the external environment by an adiabatic chamber, which is maintained at a constant temperature for the duration of the titration. A reference cell contains only the buffer. The solution containing the protein, diluted in the same buffer, is loaded into the second cell (working cell). Next, a ligand (RNA) is injected into the protein at regular intervals. The interaction between the two molecules leads to the release or absorption of heat (exothermic or endothermic reaction), resulting in a temperature change in the working cell. To equilibrate the temperature between the two cells, the device applies electrical power. This signal is measured by the calorimeter, and the power supplied is measured as a function of time, then integrated (figure 10).

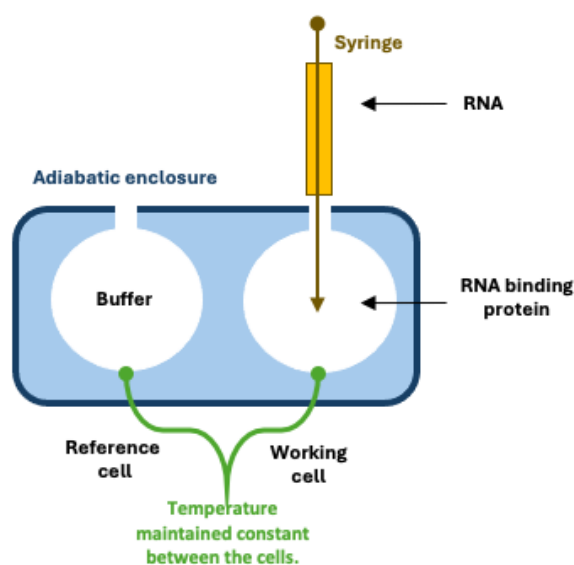


Figure 10: Isothermal Titration Calorimetry device (ITC). (Adapted from Isothermal Titration Calorimetry and Drug Design, ITC Application Note, Microcal) (64).

When the experiment starts, a large quantity of heat is released, indicating the formation of the complex. At the end of the reaction, all binding sites are saturated and no change in temperature is observed. The enthalpy variation ΔH (amplitude of the

curve), the association constant K_a (slope of the curve) and the stoichiometry (n) (inflection point of the curve) can be determined [27].

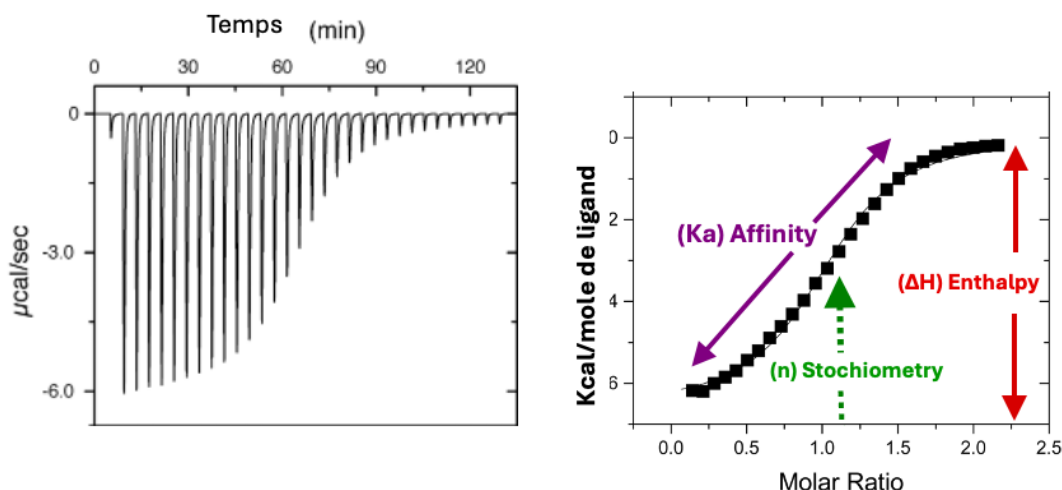


Figure 11: Raw microcalorimetric data (left), integrated and fitted to the theoretical sigmoidal curve (right). (Adapted from Isothermal Titration Calorimetry and Drug Design, ITC Application Note, Microcal) (64).

The Wiseman parameter (65) measures the precision of a titration in order to generate reliable and interpretable results. It is a useful tool for assessing the quality of experimental data obtained during an ITC experiment. The shape of the curve depends on the Wiseman parameter c as follows:

$$c = n[M_o]K_a$$

where $n[M_o]$ represents the initial concentration of binding sites in solution.

The value of c should be between 10 and 100 and the curve should correspond to a sigmoid. A value of c which is too low suggests that the concentration of the titrant (e.g. the ligand) is too low in comparison with the concentration of the protein containing the binding site. In fact, the concentration of the titrant is too diluted, and successive injections will not produce a heating signal sufficient for reliable measurement. The heat changes are too small to be detected accurately which makes the estimation of thermodynamic parameters very difficult. In addition, high values of c produce steep curves leading to unreliable measurements of K_a .

3.2 Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance spectroscopy (NMR) is based on the magnetic properties of specific atomic nuclei. This is a technique used to determine the structure, dynamics, and interactions of biomolecules. It is a powerful and non-destructive method to identify molecular interaction regions.

3.2.1 Spin and magnetic resonance

Particles that constitute matter and have mass and eventually a charge. These particles can have angular momentum named "spin" and designated by I . The spin is quantified and designed by i . The spin can be either an integer or a half integer number. The angular momentum is associated with a collinear magnetic moment characteristic of the particle.

$$\vec{\mu} = \gamma h \vec{I}$$

$\vec{\mu}$ = angular momentum

γ = gyromagnetic ratio

h = Planck constant

$\vec{I}$ = spin

A system with an angular momentum which is not equal to zero has a magnetic moment and could interact with a magnetic field, allowing the nuclear magnetic resonance to be recorded.

3.2.2 Nuclear magnetism

The proton and neutron have each a spin of $i = 1/2$. The total spin of the nucleus depends on the number of these two particles. The nuclear spin will be either zero, 1/2 integer or integer number. The nuclei that are studied are those with a spin = 1/2. We can cite the example of the proton ^1H , carbon ^{13}C , nitrogen ^{15}N , fluor ^{19}F , or phosphorus ^{31}P .

E is the interaction energy between a magnetic moment and a magnetic field, expressed by the following equation:

$$E = \vec{\mu} \cdot \vec{B}$$

When $i = 1/2$, there are two possible orientations for the spin and therefore two orientations of the magnetic moment relative to the orientation created by the magnetic field. These energy levels are known as α and β .

In the presence of a static magnetic field B_0 , the magnetic moments are along two cones aligned with the B_0 axis, corresponding to the α and β states. The ratio between N_α and N_β corresponds to each state ($i=1/2$) and is determined by the following equation:

$$N_\beta/N_\alpha = e^{-\Delta E/kT}$$

k = Constant of Boltzmann

T : Temperature

All nuclei have the same precession motion around B_0 at the same frequency ω_0 proportional to the magnetic field, known as the Larmor frequency ω_0 :

$$\omega_0 = -\gamma B_0$$

γ = gyromagnetic constant

3.2.3 Macroscopic magnetization

NMR measures a large number of identical spins providing information only about the macroscopic average. For a static magnetic field B_0 , the signal is collinear to the direction of the field B_0 and expressed as M_0 . The intensity of M_0 depends on the two spin populations in the N_α and N_β states. The gyromagnetic constant is low, which reduces the energy difference between the two states N_α and N_β , resulting in low intensity of M_0 . NMR technique has low sensitivity.

3.2.4 Resonance principle

At equilibrium, the ratio of low energy nuclei to high energy nuclei is slightly in excess, causing magnetization M_0 in the direction of the magnetic field.

During an experiment, a transverse field B_1 is applied. The result of the magnetization is no longer collinear and start to precess along the resultant of the two fields. The B_1 field is generated by a coil at the frequency being studied, during which a radio

frequency wave of a specified intensity and duration is transmitted. The resultant movement is a precession around the magnetic field at a frequency known as the Larmor frequency, with energy dissipation and a return to equilibrium. After the pulse ends, the nuclei return to their initial state and release the energy in the form of resonance signals. It corresponds to the absorption of a continuous wave at frequency ω_0 . The absorbed energy is dissipated by relaxation.

The heterogeneity of the B_0 magnetic field within the sample volume leads to loss of signal. In fact, if the applied field is not homogeneous within the sample, the rate of precession around the resulting magnetic field will be different and a loss of coherence is observed, resulting in an apparent relaxation of the T2 time constant. Simultaneously the signal corresponding to the return of the M_{xy} signal to the origine disappears, a second process is observed allowing the magnetization M_z to return along the z axis. This process is the longitudinal relaxation T1 (Fig12).

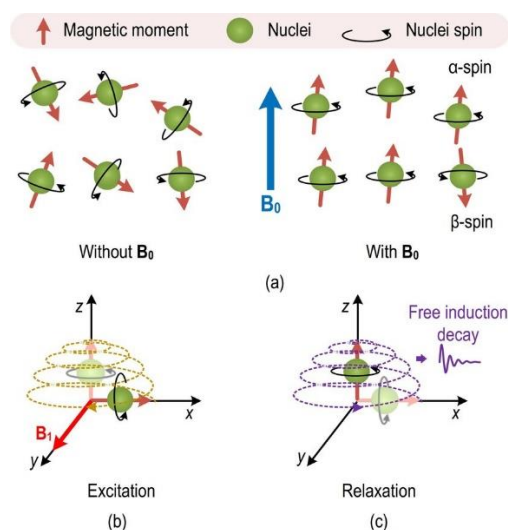
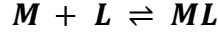


Figure 12: Principles of NMR (66). Representation of spin nuclei with or without an additional magnetic field (a). A radiofrequency excitation pulse is applied, then the magnetization is switched to the Oxy plane (b). The precession of this magnetization gives the free induction decay which is detected (c).

3.2.5 Kinetics of the interaction

The kinetics of the interaction play an important role in NMR experiments used to study protein-ligand interactions. A molecule that binds reversibly to a ligand is

characterized by the constants k_{on} and k_{off} , which are kinetic constants and determine the force of interaction between molecules.



with:

- $[M]$ is a molecule
- $[L]$ is a ligand
- $[ML]$ is the complex
- k_{on} is the association rate constant
- k_{off} is the dissociation rate constant

$$k_{on}[M][L] = k_{off}[ML] \quad ; \quad K_d = \frac{k_{off}}{k_{on}}$$

At equilibrium, the fraction of bound sites is given by:

$$\theta = \frac{[ML]}{[M]_{total}} = \frac{[L]}{K_d + [L]}$$

With:

$$[ML]_{total} = [M] + [ML]$$

Rate of Association and Dissociation:

$$\text{Association rate: } v_{assoc} = k_{on}[L][ML]$$

$$\text{Dissociation rate: } v_{dissoc} = k_{off}[ML]$$

3.2.6 NMR titration experiments

A two-dimensional 1H - ^{15}N HSQC spectrum of a ^{15}N nitrogen-labeled protein in the absence and presence of an unlabeled ligand (DNA or RNA) is used to monitor the interaction. In the case of free protein, this type of spectrum provides the chemical shift of all protons bound to nitrogen atoms and represents the fingerprint of the free protein. The chemical shifts of NH protons are especially sensitive to their chemical environment (pH, temperature and chemical environment). During NMR titration, unlabeled RNA is gradually titrated to the protein. As the ligand concentration increases, the complex is formed until the binding sites are saturated, resulting in chemical shift

perturbation (CSP). NMR titration enables the CSP to monitor the presence of complex formation, and to identify the amino acids involved and/or affected during the interaction between protein and ligand (RNA) (67).

The mode of the exchange regime can be identified by comparing the rate constant k_{off} and the difference in resonance frequencies $\Delta\nu$ of the free and bound proteins. There are three types of exchange or kinetic regime which are the slow, intermediate and fast exchange compared to NMR timescale.

In the **slow exchange regime**, the rate of exchange between the free and bound states is slower than the NMR timescale and characterized by:

$$2\pi\Delta\nu \gg k_{\text{off}}$$

$\Delta\nu$: Variation of resonance frequencies of a nuclei between the free state (ν_A) and the bound state (ν_B).

During the NMR titration experiment, when the k_{off} is slower than the difference in resonance frequencies (Hz) between the free and bound protein, two sets of peaks appear in the spectra, one corresponding to the free protein and the other to the bound form. The solution containing the protein is gradually titrated with the RNA (ligand) to form the complex. The signal from the free protein gradually disappears as the ligand is added, and the signal from the complex begins to increase. In addition, the intensities of the peaks represent the concentration of the free protein and the protein-ligand complex (67).

In the **fast exchange regime**, the rate of exchange between the free and bound states is faster than the NMR timescale and is characterized by:

$$2\pi\Delta\nu \ll k_{\text{off}}$$

When the k_{off} is higher than the difference in resonance frequencies (Hz) between the free and bound protein, a peak appears in the spectra corresponding to the weighted average of the resonance frequencies of the free and bound states. In fact, as long as the protein is slowly titrated with the RNA (ligand), the complex is formed and shifted due to the perturbation of the chemical shift resulting from the interaction between the protein and the RNA (67).

In the **intermediate exchange regime**, the exchange rate between the free and bound states is approximately equal to the NMR time scale and is characterized by:

$$2\pi\Delta\nu \approx k_{\text{off}}$$

The intermediate exchange in the NMR time scale provides a peak of the free state of the protein, which is broadened upon addition of the RNA (ligand). Although signals are likely to disappear due to signal broadening, they may reappear when the complex is formed. In some cases, the signal may disappear even when the complex is formed, preventing the complex to be observed in the NMR titration experiment (67).

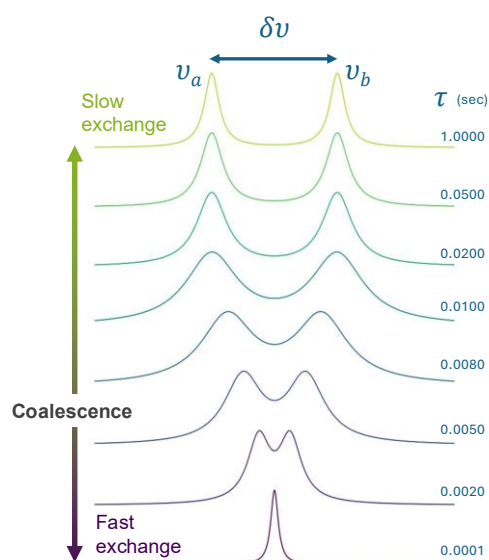


Figure 13: Exchange rate in NMR. Adapted from (68).

Additionally, the broadening of resonance frequencies has been described as mainly due to tumbling or exchange. In cases where the unlabeled titrated ligand has a very high molecular weight. During titration, signals become broader as the tumbling slows down. In such cases, all the signals of the labeled protein generally broaden and, in many cases, never reappear.

Other exchange mechanisms have been described leading to broadening, such as conformational change, structural rearrangement of the complex after binding, or dimerization of the protein before or after binding. Structural rearrangement can be determined by titration, as each titration point will produce a different value. Each titration point will produce a peak, and the latter will either be in a straight line or may be non-linear, if detectable, indicating a transient state. Other causes have been described during titration, such as broadening followed by peak sharpening, or peak splitting (69).

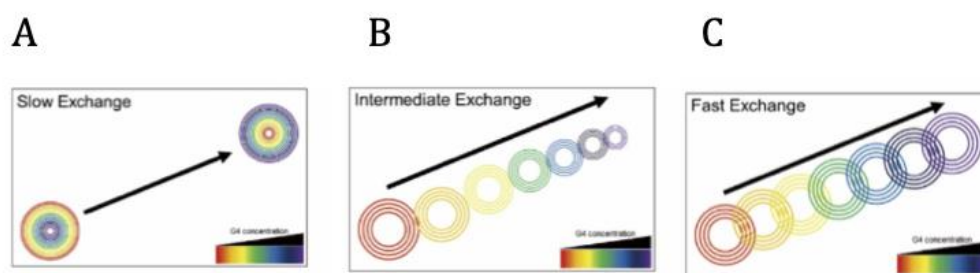


Figure 14: Spectrum of different type of exchange regime in NMR timescale. NMR titration experiments of a free protein with a ligand. The free protein is in red. The complex is in blue. Close up views, of the ^1H HSQC spectra showing a selected peak during the titration. In slow exchange (A), the two peaks correspond to the free protein and the complex. In intermediate exchange (B) the peak tends to disappear upon binding but reappear when the complex is formed. In fast exchange (C), there is the average of the peak (70).

Research projects:

1) **hnRNP A1 induces aberrant CFTR exon 9 splicing via a newly discovered Exonic Splicing Silencer element**

- Our collaborators Pr. Emanuele Buratti and Dr Cristiana Stuani recently identified a new exonic silencer sequence (ESS) located in the 3'ss of exon 9 of CFTR using a splicing assay.
- hnRNP A1 protein is capable of binding RNA sequences and is involved in multiple functions such as alternative splicing.
- TDP-43 protein is the main inhibitor and may act in concert with the hnRNP A1 protein via the ESS newly discovered.
- Interaction of both RRM domains of hnRNP A1 protein with the 3'ss of the recently identified CFTR exon 9 (ESS) RNA using isothermal titration calorimetry and NMR workflow to reveal the mechanism of the aberrant alternative splicing occurring in the CFTR gene.
- Investigating the extended complex composed of TDP-43 and hnRNP A1 proteins interacting with each other through their respective C-terminal domain (glycine-rich domain) that lead to exon skipping.

2) **Staufen1 reads our structure and sequence features in ARF1 dsRNA for target recognition**

- Mutants in STAU1 affect the binding affinity with the dsRNA target using fluorescence anisotropy measurement.
- Minor groove are not the major determinants for binding affinity *in vitro*.
- STAU1 dsRBD3/4 and sARF1 SBS dsRNA is mainly driven by electrostatic interaction *in vitro*.
- Minor groove contacts are the determinants for target selection *in vivo*.

4 Material and Methods

Cloning, expression, and purification of hnRNPA1 (RRM1&2)

The sequence of DNA encoding RRM1&2 composed of 196 amino acids was sub-cloned by PCR amplification into the pET28a vector. The construct contains a 2xHis₆ tag, a lipoyl domain to increase the solubility followed by a TEV protease cleavage site for the tag and the lipoyl domain to be removed during the first step of purification. Recombinant proteins were over-expressed in BL21(DH5- α) and RIL competent cells into *E. coli* in standard media or minimal M9 media supplemented with ¹⁵N ammonium chloride or ¹³C-labelled glucose and ¹⁵N ammonium chloride. The cells were grown at 37°C to OD600 ~0.6 and induced at OD600 ~0.9 by adding isopropyl- β -D-thiogalactopyranoside to a final concentration of 1 mM and continuing overnight at 20°C. Cells were harvested after 15h and then centrifuged at 4000 g for 20 min at 4°C. Cell pellets were resuspended in lysis buffer A (20 mM HEPES pH 7.5, 1 M NaCl, 30 mM imidazole, 5 mM BME, 10% (w/v) glycerol) supplemented with protease inhibitors, and lysed by high pressure with a french press. The cell lysate was centrifuged for 30 min at 39,000 g at 4°C. The supernatant was loaded on a Ni-NTA column on an ÄKTA Prime purification system, and after washing with buffer A the protein of interest was eluted with a 300 mM imidazole gradient. The fraction containing the protein of interest was collected then TEV protease was added according to the estimation of the amount of the protein and then dialyzed against a buffer A without imidazole overnight at room temperature to perform the cleavage of the tag. The samples were reloaded on the Ni-NTA column to remove the tag, TEV protease, and uncleaved protein. The sample containing the protein was further purified by size exclusion Chromatography on Superdex 75 column equilibrated with NMR buffer (50 mM sodium phosphate pH 6.5, 1 mM BME, 50 μ M EDTA) at 4°C and concentrated to 1 mM with a Vivaspin 10,000 MWCO.

Cloning, expression, and purification of RRM1 and RRM2 from hnRNPA1

The sequence of DNA encoding RRM1 and RRM2 individually (RRM1 from 1-105 and RRM2 from residues 91-196) was sub-cloned by PCR amplification into the pET28a vector. The construct contains a 2xHis₆ tag, a lipoyl domain to increase the solubility followed by a TEV protease cleavage site for the tag and the lipoyl domain to be removed during the first step of purification. The recombinant protein was over-expressed and purified for the construct containing both RRMs. Samples containing proteins were further purified by size exclusion Chromatography on Superdex 75 column equilibrated with NMR buffer (50 mM sodium phosphate pH 6.5, 1 mM BME, 50 μ M EDTA) at 4°C and concentrated to 1 mM with a Vivaspin 10,000 MWCO.

Cloning, expression, and purification of RRM1&2 from TDP-43

The sequence of DNA encoding RRM1&2 composed of 216 amino acids was sub-cloned by PCR amplification into the pET28a vector. The construct contains a 6Hisx2 tag, a lipoyl domain to increase the solubility followed by a TEV protease cleavage site for the tag and the lipoyl domain to be removed during the first step of purification. Recombinant proteins were over-expressed in BL21(DH5- α) and RIL competent cells into E. coli in standard media or minimal M9 media supplemented with ^{15}N ammonium chloride and/or ^{13}C -labelled glucose. The cells were grown at 37°C and induced at OD600 ~0.8 by adding isopropyl- β -D-thiogalactopyranoside to a final concentration of 1 mM and continued overnight at 16°C. Cells were harvested after 15h and then centrifuged at 4000 g for 20 min at 4°C. Cell pellets were resuspended in lysis buffer A (20 mM HEPES pH 7.5, 1 M NaCl, 30 mM imidazole, 5 mM BME, 10% (w/v) glycerol) supplemented with protease inhibitors, and lysed by high pressure with a french press. The cell lysate was centrifuged for 30 min at 39,000 g at 4°C. The supernatant was loaded on a Ni-NTA column on an ÄKTA Prime purification system, and after washing with buffer A the protein of interest was eluted with a 300 mM imidazole gradient. The fraction containing the protein of interest was collected then TEV protease was added according to the estimation of the amount of the protein and then dialyzed against a buffer A without imidazole overnight at room temperature to perform the cleavage of tag. The sample was reloaded on the Ni-NTA column to remove the tag, TEV protease, and uncleaved protein.

The sample containing the protein was further purified by size exclusion Chromatography on Superdex 75 column equilibrated with NMR buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM BME) at 4°C and concentrated to 500 mM with a Vivaspin 10,000 MWCO.

In house RNA Transcription protocol (71)

To produce single-stranded RNAs (15-mer 737 RNA: CAGGGAUUUGGGGAC and the 57 mer UG₁₃) we use a 5' cis-acting hammerhead (HH) ribozyme able to cleave itself during the T7 in vitro transcription. The in vivo transcription was optimized by preparing a small-scale reaction with $V_{\text{tot}}=25\mu\text{L}$ for each DNA template. We used a gradient of concentration (8 to 48 mM) of MgCl_2 to choose the optimal concentration. Large-scale transcription reactions (5-20 ml) contain 70 $\mu\text{g/ml}$ linearized plasmid template, 4 mM each rNTP, 1200U/ml T7 RNA polymerase, 40 mM tris-HCL (pH 8.1 at 37°C), 1 mM spermidine, 5 mM dithiothreitol, 0.1%, Triton-X 100, 3.5 $\mu\text{g/ml}$ inorganic pyrophosphatase, and MgCl_2 with the optimized concentration from the small-scale reaction. After 2 hours, we added MgCl_2 to reach 40 mM MgCl_2 and we ran the reaction for another two hours. The self-cleavage of the HH was analyzed by SDS-PAGE (20% acrylamide, 8M UREA). The RNA was visualized with 0.1 Toluidine. The reaction was stopped by adding 0,5 M EDTA, pH 8.0 to the final reaction.

In house RNA Purification and desalting protocol (71)

The RNA is purified in denaturing conditions by the ultimate HPLC system (high-performance liquid Chromatography) with an anion-exchange preparative DNA Pac PA100 Nucleic Acid Column. We performed the purification at 85°C with a flow rate of 20ml/min. The column was equilibrated with the buffer containing 6M UREA, 12.5 mM Tris/HCL, and pH 7.4. The sample was loaded into the sample loop and injected with the buffer containing 6M UREA, 12.5 mM Tris/HCL, 500 mM NaClO₄, pH 7.4 up to 50% of NaClO₄ while collecting 10ml fractions. Therefore, a steep gradient of 100% was applied to elute the uncleaved RNA and the plasmid template. The fractions were analyzed by SDS-PAGE (20% acrylamide, 8M UREA). The RNA was visualized with 0.1 Toulidine. The fractions containing the RNA were desalted by GE ÄKTA prime FPLC system (Fast protein liquid Chromatography) using 5 ml DEAE weak anion-exchange columns in series. The columns were equilibrated with the buffer containing 20 mM ammonium bicarbonate, pH 7.5 (degassed for 30 min). The sample was diluted twice, then loaded into the sample loop and injected with the buffer at a flow rate of 5 mL/min. A steep gradient containing 100% of 2.53 M ammonium bicarbonate, pH 8 was applied to collect 10 mL fractions. Fractions with the RNA were collected and then lyophilized.

RNA mutants and DNA samples

The 15-mer DNA (302 DNA: 5' CAGGGATTTGGGCAC 3') was purchased from Dharmacon company. The DNA was purified by Superdex S75 and eluted in water. The RNA Mutants (17mer RNA-a: 5' AACAGGGAUUUGGGGAC 3', 17 mer RNA-b: 5' AACAGGGAUUUGGGGAC 3', 16 mer RNA-c: 5' UCAGGGAUUUGGGUC 3', 10 mer RNA-short: 5' AGGGAUUUGG 3', 15 mer RNAUP1_C: 5' CAGGGAUUUGGGCAC 3') were purchased from Dharmacon company. Before using the RNAs were unprotected according to Dharmacon protocol purified by Superdex S75 and eluted in water. The four 7-mer DNA were purchased from Dharmacon company. The DNAs were purified by Superdex S75 and eluted in water. To produce the 7-mer single-stranded RNA (RNA: (5' CAGGGAUUUGGGGAC3')) we use a 5' cis-acting hammerhead ribozyme able to cleave itself during the T7 in vitro transcription, purified in denaturing conditions by HPLC, desalting by FPLC and then lyophilized.

NMR spectroscopy of RRM1 and RRM2 from hnRNPA1

All NMR spectra were recorded at 288K, 298K, 308K, 318K, and 313K, on Bruker AVIII-700 MHz, Bruker 800 MHz, or Avance-950 MHz spectrometers (all equipped with a cryoprobe).

Samples were measured in NMR buffer with 10% D₂O. The data were processed using TOPSPIN 2.1 and analyzed with Sparky. Protein resonances were found in the database (51). All the CSPs from NMR titration (¹H, ¹⁵N)-HSQC spectrum were calculated according to the equation $\sqrt{1/2 \{ \delta^2_{\text{H}} + (\delta_{\text{N}}/6)^2 \}}$. In addition, we recorded (¹H-¹H) TOCSY for the RNA, but the assignment was not possible since all peaks are in the same range. We recorded the F2-filtered NOESY-HSQC spectrum ($\tau_{\text{m}} = 150$ ms) measured in D₂O on the RRM1&2 ¹³C-labelled.

Isothermal titration calorimetry of RRM1&2 from hnRNPA1 with 15 mer RNA

ITC measurements were carried out at 25°C using an ITC200 calorimeter. Both protein and ligand were in the NMR buffer. Protein concentrations in the range of 4 to 5 μM were injected into the sample cell containing RNA with a concentration of 76 to 82 μM . Titrations consisted of 30 injections of 9.5 μL except for the first injection which contained 8 μL with a 10 min spacing between each injection. Data were fitted with one site binding model using the Microcal Origin software.

Isothermal titration calorimetry of RRM1&2 from hnRNPA1 with 15 mer DNA

ITC measurements were carried out at 25°C using an ITC200 calorimeter. Both protein and the ligand were in the NMR buffer. Protein concentrations in the range of 20 μM were injected into the sample cell containing RNA with a concentration of 200 μM . Titrations consisted of 30 injections of 9.5 μL except for the first injection which contained 8 μL with a 10 min spacing between each injection. Data were fitted with one site binding model using the Microcal Origin software.

Isothermal titration calorimetry of RRM1& RRM2 from hnRNPA1 with 7 mer DNA

Each RRM domain was titrated with both 7-mer DNA corresponding to the first part and the second part of the 15-mer DNA sequence. ITC measurements were carried out at 25°C using an ITC200 calorimeter. Both protein and the ligand were in the NMR buffer. Protein concentrations in the range of 200 μM were injected into the sample cell containing RNA with a concentration of 20 μM . Titrations consisted of 30 injections of 9.5 μL except for the first injection which contained 8 μL with a 10 min spacing between each injection. Data were fitted with one site binding model using the Microcal Origin software. The titration was repeated 3 times.

Formation and purification of hnRNPA1 interacting with the RNA & hnRNPA1 and TDP-43 complex interacting with the RNA containing (UG)₁₃ and the ESS motif

The complex composed of TDP-43 and RNA are prepared using an excess of TDP-43 in 20 mM HEPES, 150 mM NaCl, 0.5 mM BME, pH 7.5. The sample is diluted in the NMR buffer and mixed on a roller at low speed for 30 minutes at 25°C. The complex is purified by Superdex S75 in the same buffer. The complex of TDP-43 in 20 mM HEPES, 150 mM NaCl, 0.5 mM BME, pH 7.5 is added to an excess of hnRNPA1 in 50 mM NaP, 1 mM BME,

Small angle X-ray/neutron scattering of hnRNPA1 interacting with the RNA & hnRNPA1 and TDP-43 complex interacting with the UG₁₃RNA

The first SAXS data sets were collected using the SAXS machine BioSaxs-1000 Rigaku. Data was collected at 308K with focused (confocal Max-Flux SAXS optic Rigaku) Cu K α X-ray (1.54 Å). The sample to detector (PILATUS 100K, Dectris) distance was 0.48 m covering a scattering vector ($q=4\pi\sin(\theta)/\lambda$) range from 0.008 to 0.65 Å⁻¹. The size exclusion buffer used for the buffer measurement is (50 mM sodium phosphate+ 1 mM BME+ 50 uM EDTA, pH =6.5). The sample containing the complex YP773 and the RNA (fraction 10-11) was measured and has a concentration of 7.0 mg/ml. Six two-dimensional images were collected with an exposure time of 10 min per image for the buffer and the sample. The radial averaging, data reduction, and the buffer subtraction were performed using SAXS Lab 3.0.0r1, Rigaku. There were six individual exposures of 10 min each were done to evaluate the radiation damage and average. The integral structural parameters were determined using PRIMUS/qt ATSAS v.2.8.3

5. Results

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

Redacted due to confidentiality agreement.

RESULTS

Redacted due to confidentiality agreement.

5.3 Staufen interacting with the ARF1 dsRNA

The results presented here are from the publication titled “*Staufen1 reads out structure and sequence features in ARF1 dsRNA for target recognition*” by Deepak Kumar Yadav, Dagmar Zigackova, Maria Zlobina, Tomas Klumpler, Christelle Beaumont, Monika Kubickova, Stepanka Vanacova and Peter J. Lukavsky published in Nucleic Acids Research, 2020)

My contribution to the Staufen project was to perform the mutant to test the binding affinity. I collaborate with the postdoctoral researcher to prepare mutants, perform fluorescence anisotropy measurements, and analyze the results.

5.3.1 Characterization of dsRBD3 and dsRBD4 of STAU1 protein with the long dsRNA SBS ARF1

In this study, the dsRBD 3 and 4 binding domains of the STAU1 protein are investigated with the long dsRNA SBS ARF1 required for STAU1 binding in vivo. The sequence comprises a 19 bp stem with a canonical Watson Crick interaction and two GU base pairs formed between nucleotides 75-93 and 194-212 of the ARF1 3'UTR (73).

FA, SAXS or NMR methods were used to characterize complex interactions. FA measurements show a high binding affinity ($K_d = 26.5 \pm 2.9$ nM) for STAU1 dsRBD3/4 binding to a long ARF1 SBS dsRNA with a tetraloop UUCG. NMR titration experiments using ^1H - ^{15}N HSQC spectra with ^{15}N -labeled recombinant STAU1 dsRBD3/4 and long ARF1 SBS dsRNA provide a complex stoichiometry of 1:1. SAXS measurements confirm the 1:1 stoichiometry with a porod volume value of $52800 \text{ \AA}^3$.

Shorter ARF1 SBS constructs, representing either the upper or lower two-thirds of the long stem, were used to explore the binding mode. NMR spectra of the lower two-thirds showed minor changes, indicating a different binding mode compared to the upper two-thirds. FA measurements indicated a higher binding affinity for the upper two-thirds ($K_d = 17.8 \pm 4.3$ nM) than the lower two-thirds ($K_d = 31.0 \pm 0.1$ nM). SAXS confirmed a 1:1 stoichiometry for the STAU1 dsRBD3/4 complex with the upper two-thirds, with a Porod volume of $38900 \text{ \AA}^3$, suggesting that both dsRBDs bind to the same 15-bp region of ARF1's 3' UTR.

Smaller constructs of STAU1 dsRBD3 and dsRBD4 were produced to study the contributions of individual domains. Binding affinity tests showed that dsRBD3 had a higher affinity ($K_d = 41.4 \pm 1.5$ nM) for the upper two-thirds of ARF1 SBS compared to dsRBD4 ($K_d = 74.8 \pm 17.6$ nM). The affinity of each dsRBD was lower than the full STAU1

dsRBD3/4 complex, indicating weak cooperative binding driven by the higher affinity of dsRBD3.

5.3.2 Complexes of individuals dsRBDs bound to the short ARF1 SBS dsRNA

Complexes are then simplified to facilitate analysis of NMR data of isolated dsRBDs bound to the short sARF1.

The study shows that the ^1H - ^{15}N HSQC NMR spectrum of dsRBD4 recorded is almost identical to the one of dsRBD3 and dsRBD4 tandem, whereas dsRBD3 shows deviations and resonance broadening suggesting that dsRBD3 has different binding characteristics when isolated from the dsRBD3 and dsRBD4 tandem context. This difference is also shown in the ^1H - ^1H TOCSY spectra. It has been demonstrated that double-stranded ARF1 SBS RNA adopts the same conformation when bound to dsRBD3 and dsRBD4 or to dsRBD4, whereas dsRBD3 indicates multiple or alternative binding orientations along the double-stranded RNA, or both.

F2-filtered 3D ^{13}C F1 HSQC-NOESY shows that dsRBD3 and dsRBD4-ARF1 SBS dsRNA complexes separately, allowing intermolecular contacts to be assigned from the NOE. While dsRBD4 binds to ARF1 SBS dsRNA in a favored orientation, located in helix $\alpha 1$ in the lower part of the RNA, dsRBD3 shows multiple binding, with helix $\alpha 1$ binding to both the 5' end of ARF1 SBS dsRNA and the middle part of the stem. The data indicate that the interaction between dsRBD4 and sARF1 SBS dsRNA can be studied separately, whereas the interaction with dsRBD3 requires the tandem dsRBD3 and 4 proteins context. The labelled dsRBD3 and dsRBD4 STAU1 protein by segmental isotope NMR strategy was used where only dsRBD3 and part of the linker were labelled. Upon complex formation with sARF1 SBS dsRNA, no significant resonance broadening or deviation was observed in the ^1H - ^{15}N HSQC spectrum. Intermolecular NOEs revealed a favored orientation for dsRBD3 helix 1 in the middle part of the sARF1 SBS. The study solved the solution NMR structure of the STAU1 dsRBD4 sARF1 SBS dsRNA complexed and the STAU1 dsRBD3 dsRBD4 sARF1 SBS dsRNA complex using NMR workflow (74) (63).

5.3.3 Complexes of dsRBD4 bound interacting with sARF1 SBS dsRNA

The study shows that STAU1 dsRBD4 interacts with sARF1 in three regions. In region 3, the N-terminus of helix $\alpha 2$ of STAU1 dsRBD4 contains the conserved KKNK motif, which forms electrostatic interactions with phosphate oxygens on the major groove side of the RNA. Specifically, Lys256 forms hydrogen bonds with the phosphate oxygens of G197 and C198, while Lys257 binds to A199. On the opposite side of the sARF1 SBS dsRNA groove, Lys260 and Arg261 interact with the phosphate

oxygens of A84 and G86. Lys260 is stabilized by a hydrophobic interaction with Tyr224 of strand β 1. These interactions are crucial for recognizing the dsRNA A-form and for dsRNA binding affinity. Replacing the basic amino acids in the KKNK motif with acidic residues impairs dsRNA binding.

The study shows that in region 2, the interactions between Arg234, Arg235, and Arg236 play a role in positioning the β 1- β 2 loop in the minor groove of the UUCG tetraloop. Arg234 forms a hydrogen bond with the phosphodiester backbone through the guanidinium group and with a ribose 2' hydroxyl group through the carbonyl oxygen of its backbone. The guanidinium group of Arg 235 interact bonds with the phosphodiester backbone, while Arg236 interacts with the carbonyl oxygen (O6) of the tetraloop guanosine in the UUCG in a syn-conformation. This interaction is non-sequence-specific and could also occur with pyrimidine O2 carbonyl oxygens of pyrimidines, as seen with uracil in the ARF1 SBS 100nt loop. These interactions allow dsRBD4 to recognize the minor groove, and the hairpin shape of the RNA to be aligned with the end of the sARF1 SBS dsRNA.

The canonical region 1, contains residues 3, 4, 7, 8 and 11 of the α 1 helix. It has been shown that these residues form partially specific contacts with the minor groove of the double-stranded RNA (75).

For STAU1 dsRBD4, spacing of residues is conserved, but shifted by one amino acid, involving residues Ile207, Ser208, Ala211, Gln212 and Gln215. Ile207 establishes hydrophobic contacts with the ribose group of A85, facilitating optimal positioning of the α 1 helix in the minor groove of sARF1 SBS dsRNA. The hydroxyl of the Ser208 side chain forms a hydrogen bond with the N3 of A84, which is considered to be a non-sequence-specific interaction, since it could also occur with the N3 of guanosine or the carbonic oxygen O2 of pyrimidines. The methyl groups of residues Ala211 and Ala216 interact by stacking with the ribose motifs of A84 and G206, respectively, which further stabilizes the specific binding of helix α 1 in the minor groove.

5.3.4 Mutation on minor groove show a low effect *in vitro* but has a high effect *in vivo*

To demonstrate the relevance of the contacts involved in the minor groove, mutations of residues playing a role in the interaction with Sarf1 SBS dsRNA are mutated by alanine residues in the dsRBD3 and dsRDB4 domains. FA is measured to compare binding affinities. It was shown that mutation of Gln212 and Gln215 by an alanine decreases binding affinity by 1.6-fold, while the triple mutant involving the two preceding residues plus Ser208 decreases binding affinity by 5.2-fold. Likewise using mutation of the guanine that interacts with Ser 106 and contacts the hydrophobic sugar Ile 105 in the dsRBD3 domain results to a reduction of 3.6-fold in affinity. Similarly, the alanine

mutation of the contact with the hydrophobic sugar of Phe109 showed a 2.4-fold decrease in affinity, while the alanine mutation of the 2'hydroxyl of Glu 110 and the contact point with the hydrophobic sugar of Leu 113 in helix α 1 of the dsRBD3 domain had no effect on affinity. The same mutations were performed this time on both dsRBD3 and dsRBD4 together and the study showed that there was either no major effect or only a minor effect on affinity. The result confirms the results obtained for the individual domains and that these residues do not play a major role in the interaction. Similarly, the residues with a major effect in the isolated domains (Ser208, Ser208/Gln 212/Gln215 and Ile 105/Ser106/Ser208/Gln 212/Gln215) also show a major effect, that is, a loss of affinity due to the mutation, in the interaction with the minor groove for binding to the dsRBD target. To test the alanine mutation in the β 2 β 3 loop in STAU1 dsRBD 3 and dsRBD4 is performed on Phe109 and Arg209. The Phe 109 shows a decrease of affinity in isolated dsRBD3 or in both dsRBDs, while the Arg209 shows a minor effect on affinity. This finding indicates that the minor groove is not crucial to the interaction with STAU1 SBS dsRNA *in vitro* and that the interaction is mainly driven by electrostatic interactions along the phosphodiester backbone, as shown by the crystallographic study.

To test the relevance of minor groove interactions on STAU1 function *in vivo*, residues previously mutated *in vitro* were also mutated in cells. The study showed that *in vitro* mutation of the minor groove had no major effect, whereas mutation in cells showed a 3-to-4-fold decrease in affinity. Thus, the dsRBD3 and dsRBD4 mutations resulted in a significant decrease in binding to both mRNAs *in vivo*. The study shows that the A-form helix, hydrogen bonding and hydrophobic contact in the minor and major grooves are important for interaction and could justify the conservation of these residues in different species.

The NMR study of the interaction between STAU1 dsRBD3/4 and the ARF1 dsRNA SBS complex by segmental labeling shows a 1:1 stoichiometry. Combined with the mutation, the study provides a structure allowing the mutation to show a decrease in dsRNA binding *in vivo*.

6.Discussion

Alternative splicing is an important cornerstone of RNA-based, posttranscriptional regulation of gene expression. Production of diverse protein products through the inclusion and exclusion of certain exons from the same pre-mRNA is a major contributor to the complexity of higher organisms. But at the same time, splicing also depends on correct RNA and protein products for proper functioning, and thus mutations in both components can lead to mis-splicing and disease (76). One such mutation is found in the polymorphic (TG)_m (T)_n locus upstream of CFTR exon 9 within intron 8 which leads to recruitment of TDP43 to the pre-mRNA and subsequent TDP43-dependent recruitment hnRNP A1. The UG-rich binding site of TDP43 in the CFTR pre-mRNA is well characterized both functionally (31) (54) and structurally (77) (3). and it is known that hnRNP A1 is recruited to the 3'ss *via* a protein-protein interaction with TDP43(5) (4). The molecular origin of CFTR exon 9 splicing inhibition *via* TDP43-dependent recruitment of hnRNP A1, on the other hand, remained elusive.

Our study identifies a new ESS comprising a bipartite sequence motif at the CFTR exon 9 3'ss and the 5' end of exon 9 (5'-AGGGAUUUGGGGAU-3'). We show that hnRNP A1 is able to specifically interact with this bipartite ESS motif and that the integrity of this site is instrumental for aberrant splicing of CFTR exon 9. Our minigene splicing assay shows that the wild-type sequence around the CFTR 3'ss produces several bands corresponding to the inclusion and skipping of exon 9. Four different mutations in exon 9 which alter the 5'-GGGGA-3' motif but leave the 3'ss intact, clearly show variation of this sequence affects the inclusion of exon 9 in the mRNA. Mutating the 5'-GGGGA-3' motif to 5'-CGCGT-3' as well as its complete deletion is associated with the non-skipping of exon 9, i.e. the full inclusion in the final mRNA. Thus, these experiments reveal for the first time an ESS motif in CFTR exon 9 responsible of the exon 9 skipping.

We also investigated the RNA-binding mode of hnRNP A1 to the ESS that silences the splicing of CFTR exon 9. The ITC and NMR titration experiments strongly suggest that both RRM of hnRNPA1 interact with the bipartite ESS motif. The affinities obtained from ITC experiments using the isolated RRMs and single binding sites of the bipartite ESS motif are comparable, suggesting that there is no preferred sequence motif for the individual RRMs determining the orientation of hnRNPA1 on the RNA target. Previous studies have shown that the hnRNP A1 protein can bind specifically and with high affinity to ssRNA or ssDNA sequences because there are no qualitative differences between them (48) and high-affinity sequences have also been discovered for hnRNPA1. Indeed, SELEX experiments isolated a high-affinity RNA sequence of 5'-UAGGGA/U-3' that is recognized by hnRNPA1 (27) which is similar to the CFTR exon 9 ESS motif. The specific recognition of the motif was confirmed by an x-ray crystallography structure of hnRNPA1 tandem RRMs bound to a telomeric ssDNA sequence (5'-TTAGGGTTAGGG-3')(48). The structure revealed that both RRM1 and

RRM2 of the same monomer interact with two different strands of ssDNA, in an antiparallel manner. In this structure, RRM1 binds to four nucleotides (TAGG), while RRM2 preferably binds to five nucleotides (TTAGG). The CFTR exon 9 ESS motif also contains two binding sites with similar sequences but our NMR and ITC data suggest that a single copy of hnRNP A1 binds both motifs *via* the tandem RRMs. Thus, the dimerization in the x-ray structure might be just caused by crystal packing forces and not be relevant for CFTR exon 9 skipping.

A study was performed by Allain and colleagues has identified optimal recognition motifs that bind the individual RRMs (52). RRM1 binds to the sequence 5'-U/CAGG-3' and RRM2 to the sequence 5'-U/CAGN-3'. Binding studies revealed a strong preference for the 5'-UAG-3' sequence which can be further extended to 5'-U/CAGG-3'. They also show that both RRMs of hnRNP A1 bind to their RNA target containing the two AG-binding sites without dimerization of the protein and that hnRNP A1 binds the ISS-N1 pre-mRNA with directionality where RRM2 interacts with the 5' motif and RRM1 binds the 3' motif. Another study investigated the interaction between hnRNP A1 to an RNA target, namely the pri-mir-18a. This interaction involves both RRMs and a region encompassing two 5'-UAG-3' motifs (53).

In this study, it was demonstrated that cooperative binding of both domains to the related RNA motifs results in a strong enhancement of binding affinity and allows the unwinding of the RNA target as a stem-loop. Our ITC experiments also reveal that individual RRMs interact with the ESS in the low micromolar range while the tandem RRMs display nanomolar affinity consistent with cooperative binding of both RRMs to the bipartite ESS.

Our study using hnRNP A1 and the 7-mer DNA model mimicking the 3'ss CFTR RNA target, our experiments did not show a preference for either of the RRMs for a specific segment of the bipartite ESS sequence, both orientations of the ESS segment on the tandem RRMs interact with the same continuous protein surface. Using the larger confirm the interaction of both protein for the RNA target. As part of the SAXS result we suggest a 2:2:1 stoichiometry of the extended complex but its orientation remains unclear. In fact, any of the RRMs, or even RRM1 and RRM2 from different proteins, can bind to the RNA and form a mixture.

In addition, I contributed to the Staufen project. We will discuss the structure of the STAU1 dsRBD3/4 complex with sARF1SBSdsRNA, which shows how STAU1 selects target mRNAs for SMD. Both dsRBDs display a canonical binding. The isolated dsRBD4 binds to sARF1SBSdsRNA in a favored orientation while dsRBD3 shows non-specific binding to various regions of the dsRNA.

As part of the dsRBD3/4 tandem project, dsRBD3 selects a single binding site on dsRNA. To study this interaction, a segmentally labeled STAU1 protein was used, with only isotope-labeled dsRBD3. This enabled isotope-filtered NMR experiments to provide the determination of the dsRBD3 binding site on sARF1SBSdsRNA, and then to confirm that two dsRBD domains do not form protein-protein contacts, as shown by the crystal structure. This suggests that the dsRBD3 domain initially binds double-stranded A-form RNA non-specifically, exploring double-stranded RNA in this form, facilitating recruitment of the dsRBD4 domain for RNA binding. The dsRBD4 domain then binds selectively to a defined binding site on dsRNA, leading dsRBD3 to also bind to a specific site on the RNA. Furthermore, the interaction of the $\beta 1$ and $\beta 2$ loops of each dsRBD with the opposite ends of dsRNA could further facilitate the final selection of the dsRNA binding site.

The binding affinity of dsRBDs is determined by electrostatic interactions with the phosphodiester backbone and two hydroxyl groups, resulting in the selection of dsRNA rather than dsDNA for binding to the regular geometric shape A. Consequently, mutations in the charged basic residues in regions 2 and 3 of STAU1 dsRBD3/4 result in a 4- to 8-fold reduction in binding affinity compared with the wild-type protein. There are only a limited number of minor groove base contacts that contribute to sequence specificity and thus determine target selection. Consequently, altering minor groove recognition by mutating key residues of the protein generally has minimal impact on affinity.

The study shows that mutations, which contact the minor grooves, have a much stronger effect on the full-length protein *in vivo* and explained that can be due to the fact that the STAU1 mutant is presented in the context of the full-length protein and tested against endogenous mRNA targets. In addition, the authors suggest that the recognition of important signals by STAU1 may require not only interactions with minor grooves in dsRBD3/4 domains, but also *in vivo* dimerization (78) and the involvement of other domains, such as dsRBD2.

7.Conclusion

In conclusion of the main project, our research focuses on the interaction between hnRNPA1 and CFTR exon 9. Initially, the study was performed on the smallest complex using both RRM from hnRNPA1 and the 3' splice site CFTR RNA. DNA mimicking RNA target have been used. Several strategies were developed to solve the problem by separating the RNA recognition motif and splitting the RNA. Although stoichiometry provided an expected ratio of 1:1, the affinity was a problematic issue that could answer neither the question of stoichiometry for the whole complex nor the question of orientation. However, the formation of the extended complex comprising TDP-43, hnRNPA1, and longer CFTR RNA with 2:2:1 stoichiometry was established by SAXS methods.

This finding is a promising indicator that could confirm our research hypothesis, where TDP-43 is the main inhibitor and its recruitment, together with the hnRNPA1 protein, binds to CFTR RNA and prevents recognition of 3'splice site leading to aberrant splicing of CFTR exon 9.

Further work is needed to determine the orientation of hnRNP A1 along the CFTR exon 9. Binding this natural target with directionality might require the protein-protein interaction of full-length TDP-43 and hnRNPA1 mediated by their glycine-rich C-terminal domains (4) (5). Consistent with this idea, disrupting this interaction, via mutations or deletion in the C termini of TDP-43 as well as hnRNP A1, abolishes exon 9 skipping despite the fact that hnRNP A1 maintains its ability to bind to the target CFTR exon 9 ESS. Studying the RNA-protein and protein-protein interactions within the entire aberrant splice site using full-length TDP43 and hnRNP A1 is required to shed light on the precise mechanism of aberrant splicing occurring at CFTR exon 9.

To conclude on the STAU1 project, this is a study of the interaction between recombinant STAU1 dsRBD3 and dsRBD4 (aa 102-274) and the 19 bp human ARF1 3' UTR dsRNA SBS. STAU1 dsRBD4 binds two pyrimidines and a purine on the minor groove side via helix 1, with the $\beta 1$ $\beta 2$ loop crossing the dsRBD at the dsRNA end. The lysine in helix 2 binds to the phosphodiester backbone on the major groove, while STAU1 dsRBD3 recognizes a guanine base in the minor groove of dsRNA. Additionally, mutants interfere with ARF1 SBS minor groove recognition and abolish the binding *in vitro*, leading to reduced SMD *in vivo*. It remains unclear whether STAU1 and STAU2 fulfill their different functions in the post-transcriptional regulation of gene expression, as only a few minor groove interactions appear to determine dsRNA target selection. Depending on the complexity of the dsRNA signal, the structure of the STAU1 or

STAU2 complexes with the target RNAs may be affected. This could reveal different protein surfaces of STAU1 or STAU2, allowing their interaction with other proteins involved in RNA transport, translation regulation or GDS, and thus influence the cellular function of this interaction.

Original publications and definition of the author's contribution

The thesis is based on these manuscripts written by the author of the thesis (Christelle Beaumont).

Article 1: hnRNP A1 induces aberrant CFTR exon 9 splicing *via* a newly discovered Exonic Splicing Silencer element Appendix A to this doctoral thesis)

Author's contribution:

Christelle Beaumont 1, Ming-Yuan Chou 1, Cristiana Stuani Emanuele Buratti 2 and Peter Josef Lukavsky.

P.J.L. designed the project. C.B. and M.Y.C. expressed and purified proteins and purified ssDNA. C.B. performed all NMR measurements, ITC experiments with 15-mer ssDNA and analyses. M.Y.C. performed all ITC experiments with individual RRM and shorter ssDNAs. C.S. performed all the minigene splicing assays and analyzed the data. C.B. wrote an initial draft, P.J.L. wrote the manuscript and made proper figures, E.B. and C.S. wrote the minigene splicing assay part and made corrections to the text. All authors discussed the results and approved the manuscript.

Article 2: Stauf1 reads out structure and sequence features in ARF1 dsRNA for target recognition. Appendix B to this doctoral thesis)

Author's contribution:

Deepak Kumar Yadav, Dagmar Zigackova, Maria Zlobina, Tomas Klumpler, Christelle Beaumont, Monika Kubickova, Stepanka Vanacova and Peter J. Lukavsky.

P.J.L. and D.K.Y. designed the project. D.K.Y. expressed and purified proteins and performed in vitro RNA transcription and purification. D.K.Y. optimized the STAU1-ARF1 complexes and prepared all samples for structural studies. M.Z., M.K., C.B. and D.K.Y. prepared samples for FA measurements. D.K.Y. performed all NMR experiments with help of P.J.L. D.K.Y. and P.J.L. analyzed NMR data and performed structure calculations. M.Z., M.K., C.B. and D.K.Y. performed FA measurements and analysis. T.K., performed

and analyzed SAXS measurements. D.Z. performed all the experiments in mammalian cells, constructs, stable cell lines, mRNA quantifications by RT-qPCR, and western blotting, D.Z. and S.V analyzed the data. P.J.L. wrote the manuscript, D.Z. and S.V. wrote method and results part on the in vivo experiments in mammalian cells. All authors discussed the results and approved the manuscript.

Bibliography

1. Buratti E, Baralle FE. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem*. 28 sept 2001;276(39):36337-43.
2. Pagani F, Buratti E, Stuani C, Baralle FE. Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J Biol Chem*. 18 juill 2003;278(29):26580-8.
3. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol*. déc 2013;20(12):1443-9.
4. D'Ambrogio A, Buratti E, Stuani C, Guarnaccia C, Romano M, Ayala YM, et al. Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res*. juill 2009;37(12):4116-26.
5. Buratti E, Brindisi A, Giombi M, Tisminetzky S, Ayala YM, Baralle FE. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J Biol Chem*. 11 nov 2005;280(45):37572-84.
6. Sanders DB, Fink A. Background and Epidemiology. *Pediatr Clin North Am*. août 2016;63(4):567-84.
7. Zengerling S, Tsui LC, Grzeschik KH, Olek K, Riordan JR, Buchwald M. Mapping of DNA markers linked to the cystic fibrosis locus on the long arm of chromosome 7. *Am J Hum Genet*. mars 1987;40(3):228-36.
8. Harris A. Cystic fibrosis gene. *Br Med Bull*. oct 1992;48(4):738-53.
9. Gadsby DC, Vergani P, Csanády L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature*. 23 mars 2006;440(7083):477-83.
10. Zhang Z, Liu F, Chen J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. *Cell*. 27 juill 2017;170(3):483-491.e8.
11. Linsdell P. Mechanism of chloride permeation in the cystic fibrosis transmembrane conductance regulator chloride channel. *Exp Physiol*. 2006;91(1):123-9.
12. Lopes-Pacheco M. CFTR Modulators: The Changing Face of Cystic Fibrosis in the Era of Precision Medicine. *Front Pharmacol*. 2019;10:1662.

13. Fanen P, Wohlhuter-Haddad A, Hinzpeter A. Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. *Int J Biochem Cell Biol.* juill 2014;52:94-102.
14. Gilbert W. Why genes in pieces? *Nature.* 9 févr 1978;271(5645):501.
15. Berk AJ. Discovery of RNA splicing and genes in pieces. *Proc Natl Acad Sci.* 26 janv 2016;113(4):801-5.
16. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet.* août 2018;59(3):253-68.
17. Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem.* 2015;84:291-323.
18. WANG Y, LIU J, HUANG B, XU YM, LI J, HUANG LF, et al. Mechanism of alternative splicing and its regulation. *Biomed Rep.* mars 2015;3(2):152-8.
19. Matlin AJ, Clark F, Smith CWJ. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol.* mai 2005;6(5):386-98.
20. Tazi J, Bakkour N, Stamm S. Alternative splicing and disease. *Biochim Biophys Acta.* 2009 Jan; 1792(1): 14-26.
21. Blencowe BJ. Alternative splicing: new insights from global analyses. *Cell.* 14 juill 2006;126(1):37-47.
22. Wang Z, Xiao X, Van Nostrand E, Burge CB. General and specific functions of exonic splicing silencers in splicing control. *Mol Cell.* 7 juill 2006;23(1):61-70.
23. Zhu J, Mayeda A, Krainer AR. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell.* déc 2001;8(6):1351-61.
24. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem.* 1993;62:289-321.
25. Mayeda A, Helfman DM, Krainer AR. Modulation of Exon Skipping and Inclusion by Heterogeneous Nuclear Ribonucleoprotein A1 and Pre-mRNA Splicing Factor SF2/ASF. *MOL CELL BIOL.* 1993;13.
26. Wu YX, Kwon YJ. Aptamers: The « evolution » of SELEX. *Methods San Diego Calif.* 15 août 2016;106:21-8.

27. Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 1 mars 1994;13(5):1197-204.
28. Cuppens H, Lin W, Jaspers M, Costes B, Teng H, Vankeerberghen A, et al. Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *J Clin Invest.* 15 janv 1998;101(2):487-96.
29. Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui LC, Antonarakis SE, et al. A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. *Nature.* 26 juill 1990;346(6282):366-9.
30. Pagani F, Buratti E, Stuanı C, Romano M, Zuccato E, Niksic M, et al. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem.* 14 juill 2000;275(28):21041-7.
31. Buratti E, Dörk T, Zuccato E, Pagani F, Romano M, Baralle FE. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.* 2 avr 2001;20(7):1774-84.
32. Zuccato E, Buratti E, Stuanı C, Baralle FE, Pagani F. An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. *J Biol Chem.* 23 avr 2004;279(17):16980-8.
33. Cléry A, Blatter M, Allain FHT. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol.* juin 2008;18(3):290-8.
34. Cléry A, Allain FHT. FROM STRUCTURE TO FUNCTION OF RNA BINDING DOMAINS. In: Madame Curie Bioscience Database [Internet] [Internet]. Landes Bioscience; 2013 [cité 1 mai 2025]. Disponible sur: <https://www.ncbi.nlm.nih.gov/books/NBK63528/>
35. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science.* 29 juill 1994;265(5172):615-21.
36. Mayeda A, Krainer AR. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell.* 24 janv 1992;68(2):365-75.
37. Pollard VW, Michael WM, Nakielnı S, Siomi MC, Wang F, Dreyfuss G. A novel receptor-mediated nuclear protein import pathway. *Cell.* 20 sept 1996;86(6):985-94.

38. Moran-Jones K, Wayman L, Kennedy DD, Reddel RR, Sara S, Snee MJ, et al. hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere. *Nucleic Acids Res.* 2005;33(2):486-96.
39. Hamilton BJ, Nichols RC, Tsukamoto H, Boado RJ, Pardridge WM, Rigby WF. hnRNP A2 and hnRNP L bind the 3'UTR of glucose transporter 1 mRNA and exist as a complex in vivo. *Biochem Biophys Res Commun.* 11 août 1999;261(3):646-51.
40. He Y, Smith R. Nuclear functions of heterogeneous nuclear ribonucleoproteins A/B. *Cell Mol Life Sci CMLS.* avr 2009;66(7):1239-56.
41. Jean-Philippe J, Paz S, Caputi M. hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci.* 16 sept 2013;14(9):18999-9024.
42. Kamma H, Horiguchi H, Wan L, Matsui M, Fujiwara M, Fujimoto M, et al. Molecular characterization of the hnRNP A2/B1 proteins: tissue-specific expression and novel isoforms. *Exp Cell Res.* 1 févr 1999;246(2):399-411.
43. Kamma H, Portman DS, Dreyfuss G. Cell type-specific expression of hnRNP proteins. *Exp Cell Res.* nov 1995;221(1):187-96.
44. Mayeda A, Munroe SH, Xu RM, Krainer AR. Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA N Y N.* sept 1998;4(9):1111-23.
45. Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol.* mars 1997;4(3):215-22.
46. Nakielnny S, Dreyfuss G. Transport of proteins and RNAs in and out of the nucleus. *Cell.* 23 déc 1999;99(7):677-90.
47. Ryan VH, Dignon GL, Zerze GH, Chabata CV, Silva R, Conicella AE, et al. Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell.* 1 févr 2018;69(3):465-479.e7.
48. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* 1 mai 1999;13(9):1102-15.
49. Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR. Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Struct Lond Engl* 1993. 15 avr 1997;5(4):559-70.

50. Allain FH, Howe PW, Neuhaus D, Varani G. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* 15 sept 1997;16(18):5764-72.
51. Barraud P, Allain FHT. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM s influences the nucleic acid binding topology. *J Biomol NMR.* janv 2013;55(1):119-38.
52. Beusch I, Barraud P, Moursy A, Cléry A, Allain FHT. Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *eLife.* 26 juin 2017;6:e25736.
53. Kooshapur H, Choudhury NR, Simon B, Mühlbauer M, Jussupow A, Fernandez N, et al. Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat Commun.* 26 juin 2018;9(1):2479.
54. Ayala YM, Pagani F, Baralle FE. TDP43 depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett.* 20 févr 2006;580(5):1339-44.
55. Wright GSA, Watanabe TF, Amporn danai K, Plotkin SS, Cashman NR, Antonyuk SV, et al. Purification and Structural Characterization of Aggregation-Prone Human TDP-43 Involved in Neurodegenerative Diseases. *iScience.* 26 juin 2020;23(6):101159.
56. Jiang LL, Che MX, Zhao J, Zhou CJ, Xie MY, Li HY, et al. Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. *J Biol Chem.* 5 juill 2013;288(27):19614-24.
57. Mompeán M, Baralle M, Buratti E, Laurents DV. An Amyloid-Like Pathological Conformation of TDP-43 Is Stabilized by Hypercooperative Hydrogen Bonds. *Front Mol Neurosci.* 2016;9:125.
58. Wickham L, Duchâine T, Luo M, Nabi IR, DesGroseillers L. Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum. *Mol Cell Biol.* mars 1999;19(3):2220-30.
59. Ramos A, Grünert S, Adams J, Micklem DR, Proctor MR, Freund S, et al. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.* 1 mars 2000;19(5):997-1009.
60. Park E, Maquat LE. Staufen-mediated mRNA decay. *Wiley Interdiscip Rev RNA.* 2013;4(4):423-35.

61. Kim YK, Furic L, Parisien M, Major F, DesGroseillers L, Maquat LE. Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J*. 6 juin 2007;26(11):2670-81.
62. Lazzaretti D, Bandholz-Cajamarca L, Emmerich C, Schaaf K, Basquin C, Irion U, et al. The crystal structure of Staufen1 in complex with a physiological RNA sheds light on substrate selectivity. *Life Sci Alliance*. oct 2018;1(5):e201800187.
63. Yadav DK, Lukavsky PJ. NMR solution structure determination of large RNA-protein complexes. *Prog Nucl Magn Reson Spectrosc*. nov 2016;97:57-81.
64. Ward WH, Holdgate GA. Isothermal titration calorimetry in drug discovery. *Prog Med Chem*. 2001;38:309-76.
65. Wiseman T, Williston S, Brandts JF, Lin LN. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*. 15 mai 1989;179(1):131-7.
66. Anders J, Dreyer F, Krüger D, Schwartz I, Plenio MB, Jelezko F. Progress in miniaturization and low-field nuclear magnetic resonance. *J Magn Reson*. 1 janv 2021;322:106860.
67. Schubert Mario, Dominguez Cyril, Duss Olivier, Ravindranathan Sapna, Allain F H.-T. How to study the structure and dynamics of protein-RNA complexes by NMR spectroscopy *Biomolecular NMR spectroscopy*.
68. Leigh JS. Relaxation times in systems with chemical exchange: Some exact solutions. *J Magn Reson* 1969. 1 juin 1971;4(3):308-11.
69. Williamson MP. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*. août 2013;73:1-16.
70. Oh KI, Kim J, Park CJ, Lee JH. Dynamics Studies of DNA with Non-canonical Structure Using NMR Spectroscopy. *Int J Mol Sci*. 11 avr 2020;21(8):2673.
71. Zlobina M, Sedo O, Chou MY, Slepankova L, Lukavsky PJ. Efficient large-scale preparation and purification of short single-stranded RNA oligonucleotides. *Bio-Techniques*. févr 2016;60(2):75-83.
72. Okunola HL, Krainer AR. Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol*. oct 2009;29(20):5620-31.
73. Kim YK, Furic L, Desgroseillers L, Maquat LE. Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell*. 28 janv 2005;120(2):195-208.

74. Dominguez C, Schubert M, Duss O, Ravindranathan S, Allain FHT. Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc.* févr 2011;58(1-2):1-61.
75. Masliah G, Barraud P, Allain FHT. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol Life Sci CMLS.* juin 2013;70(11):1875-95.
76. Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell.* 20 févr 2009;136(4):777-93.
77. Kuo PH, Doudeva LG, Wang YT, Shen CKJ, Yuan HS. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* avr 2009;37(6):1799-808.
78. Gleghorn ML, Gong C, Kielkopf CL, Maquat LE. Stauf1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat Struct Mol Biol.* avr 2013;20(4):515-24.

**Appendix A: hnRNP A1 induces aberrant CFTR
exon 9 splicing *via* a newly discovered Exonic
Splicing Silencer element**

hnRNP A1 induces aberrant CFTR exon 9 splicing *via* a newly discovered Exonic Splicing Silencer element

Christelle Beaumont ¹, Ming-Yuan Chou ¹, Cristiana Stuan ², Emanuele Buratti ² and Peter Josef Lukavsky ^{1,*}

¹ CEITEC- Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic; christie.beaumont@yahoo.com (C.B.); myc237@gmail.com (M.Y.C.); peter.lukavsky@ceitec.muni.cz (P.J.L.)

² International Centre for Genetic Engineering and Biotechnology, Padriciano 99, 34012 Trieste, Italy; stuani@icgeb.org (C.S.); buratti@icgeb.org (E.B.)

* Correspondence: peter.lukavsky@ceitec.muni.cz

Abstract: RNA-protein interactions play a key role in the aberrant splicing of CFTR exon 9. Exon 9 skipping leads to the production of a non-functional chloride channel associated with severe forms of cystic fibrosis. The missplicing depends primarily on variations in the polymorphic (TG)_mT_n locus upstream of exon 9. At the pre-mRNA level, it generates an extended UG-rich binding site for TDP-43, associated with hnRNP A1 recruitment, and prevention of exon 9 3' splicing site (3'ss) recognition. While TDP-43 is the dominant inhibitor of exon 9 inclusion, the role of hnRNP A1, a protein with two RNA recognition motifs (RRM1 and RRM2) and a glycine-rich domain, remained unclear. In this work, we have studied the interaction between hnRNP A1 and the CFTR pre-mRNA using NMR spectroscopy and Isothermal Thermal Calorimetry (ITC). The affinities are submicromolar and ITC data suggest that the separate RRM1 and RRM2 form 1:1 complexes. NMR titrations reveal that hnRNP A1 interacts with model CFTR 3'ss sequences in a fast exchange regime at the NMR timescale. Splicing assays finally show that this hnRNP A1 binding site represents a previously unknown exonic splicing silencer element. Together, our results shed light on the mechanism of aberrant CFTR exon 9 splicing.

Keywords: Alternative splicing, CFTR, pre-mRNA, RNA binding protein, RNA-binding mode, RRM (RNA Recognition motif), NMR spectroscopy

1. Introduction

Cystic fibrosis is an autosomal recessive, inherited disease that affects the respiratory, digestive, and reproductive systems¹. The CFTR (Cystic Fibrosis Transmembrane conductance Regulator) gene encodes for the CFTR protein which acts as a chloride channel in epithelial cell membranes². The CFTR gene contains 27 exons. All exons included in the mRNA produces a functional protein³⁻⁵. The expression of most eukaryotic genes requires mRNA splicing which removes introns and ligates exons to produce a mature mRNA. Two types of regulating, cis-acting elements are found in pre-mRNAs which either enhance or silence splicing *via* interaction with specific trans-acting splicing factors. The cis-regulatory elements located upstream or downstream of the regulated exon within the pre-mRNA, define which exons are included or excluded from the pre-mRNA⁶. Silencer elements interact with negative trans-acting factors such as hnRNPs (heterogeneous nuclear ribonucleoproteins) and thus repress splicing. The hnRNP family is conserved in humans which highlights their importance for functioning in pre-mRNA maturation⁷.

In aberrant splicing of the CFTR gene, there is limited inclusion of exon 9 in the final mRNA due to a polymorphic sequence in the intron upstream of this exon. The result is a defect in the nucleotide binding domain of the CFTR channel which leads to CFTR channel dysfunction. Upstream of exon 9, in intron 8, polymorphisms can produce different variant numbers of the (TG)_m(T)_n element⁸⁻⁹. At the pre-mRNA level, it has been observed that polymorphic (UG)_m(U)_n repeats with low number of Us and concomitant high number of UG repeats can repress the use of the polypyrimidine tract by altering the interaction with trans-acting proteins.

In regular splicing, the small U1 snRNA of the snRNP interacts with the 5'ss of the pre-mRNA and is supported by auxiliary factors such as the serine-arginine-rich proteins (Figure 1A). Recognition of the 3'ss is conducted by a concerted interaction of the splicing factor 1, also called branch point binding protein (SF1/BBP) to the branch point sequence and the heterodimeric auxiliary factor U2AF. While U2AF35 recognizes the conserved AG dinucleotide at the 3'ss, U2AF65 binds the polypyrimidine tract located upstream¹⁰⁻¹³.

Studies showed that TDP-43 protein, displaying two RRM domains and an unstructured, glycine-rich C-terminus, has a high affinity for a (TG)₁₂T₅ sequence and binds to this site at the 3'ss of CFTR exon 9¹⁴⁻¹⁵. Since this polymorphism is associated with a very short polypyrimidine tract, binding of the canonical factor U2AF65/35 is presumably impaired (Figure 1A). TDP-43, on the other hand, has a strong affinity for the concomitant, extended UG repeat and thus acts as the principal inhibitor causing aberrant splicing of CFTR exon 9¹⁶⁻¹⁷.

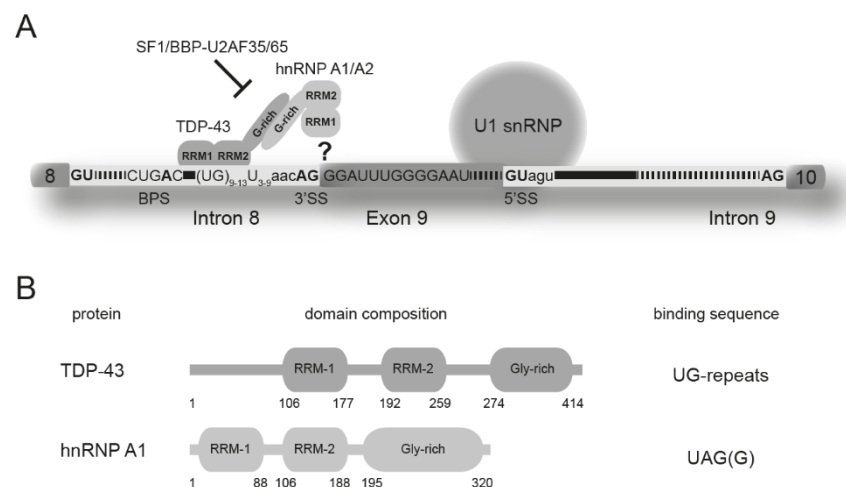


Figure 1. Aberrant splicing of CFTR exon 9. (A) Recognition of the CFTR 3'ss by TDP-43 and hnRNP A1. The relevant RNA sequences in intron 8 and exon 9 are displayed. TDP-43 binds the elongated UG-repeat and recruits hnRNP A1 via their C-terminal glycine-rich domains. Access of SF1/BBP and U2AF35/65 to the canonical binding sites (BPS, polypyrimidine tract and 3'ss AG) at the 3'ss are impaired. The pre-mRNA interaction site of hnRNP A1 is unknown. (B) Schematic domain representation of human TDP-43 RBD and human hnRNP A1. Domain boundaries of each domain are indicated according to the full length protein sequences (NP_031401 and NP_112420.1).

Previous studies have shown that beside the RRM domains, the C-terminus of TDP-43 protein is also essential for aberrant CFTR exon 9 splicing *via* an interaction with the C-terminus of hnRNP A1 (Figure 1A-B)¹⁸. Using deletion mutants, Baralle, Buratti and colleagues, mapped the minimal region (region between 321-366 of the TDP-43 C-terminus)

required for the interaction of hnRNP A1 and TDP-43¹⁹. Interestingly, similarly to TDP-43 the hnRNP A1 protein also contains two RRM domains and an unstructured, glycine-rich C-terminus (Figure 1B). The latter provides a platform for protein-protein interactions playing a key role in regulating gene expression including alternative splicing, nuclear and export from the nucleus to the cytoplasm, telomere maintenance, mRNA stability and turnover, mRNA processing and translation regulation²⁰. Although TDP-43 has been shown to be the main splicing inhibitor, hnRNP A1 is also required for aberrant CFTR exon 9 splicing¹⁸. However, until present the hnRNP A1 interaction site of its tandem RRMs at the 3'ss of exon 9 of the CFTR pre-mRNA remained elusive (Figure 1A).

The tandem RRMs of human hnRNP A1 are connected by a 17 amino acid linker and provide a platform for RNA interactions (Figure 1B). The tandem RRMs have been crystallized in the free and bound form where they dimerize upon binding with ssDNA. Two ssDNA sequences bind to each tandem RRM in an antiparallel manner where the 5' extremity of the ssDNA binds to the RRM1 of one tandem RRM domain while the 3' extremity binds to the RRM2 of the second copy²¹. The presence of this specific dimer arrangement and binding mode might be due to the packaging forces in the crystals. A study in solution using NMR spectroscopy combined with segmental isotope labeling shows that the orientation of the unbound tandem RRMs differs from the crystal structure of the free form and resembles more the one found in the ssDNA-bound form²². However, the interaction between the tandem RRMs and RNA remained controversial. Despite all the structural insights from model RNA-hnRNP A1 complexes, no interaction with a naturally occurring splice site has been characterized in detail.

Our study set out to decipher the role of hnRNP A1 in aberrant splicing of CFTR exon 9. Splicing assays reveal a previously unrecognized ESS within CFTR exon 9. NMR titration experiments and ITC measurements confirm that this element interacts with hnRNP A1 with submicromolar affinity in a specific manner forming 1:1 complexes. Our results explain how TDP-43 and hnRNP A1 work in concert to block formation of a splicing competent complex at the CFTR exon 9 3'ss and thus cause o aberrant CFTR exon 9 splicing.

2. Materials and Methods

2.1. Cloning, expression and purification of hnRNP A1 tandem RRMs

The sequence of DNA encoding the tandem RRMs composed of 196 amino-acids were subcloned by PCR amplification into the pET28a vector. The construct contains a 2xHis₆ tag, a lipoyl domain followed by a TEV protease cleavage site for the tag and the lipoyl domain to be removed after the first step of purification. Recombinant proteins were overexpressed in *E. coli* BL21(DE3) codon plus cells (Novagen) in LB rich or M9 minimal media supplemented with ¹⁵NH₄Cl or ¹⁵NH₄Cl and ¹³C-glucose. The cells were grown at 37°C to OD600 ~0.6 and then at 20°C for 1h. Protein expression was induced at OD600 ~0.9 by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and further incubated for 15h at 20°C. Cells were harvested after 15h and then centrifuged at 4000 g for 20 min at 4°C. The cell pellet was resuspended in lysis buffer A (20 mM HEPES pH 7.5, 1 M NaCl, 30 mM imidazole, 5 mM BME, 10% (w/v) glycerol) supplemented with protease inhibitors, and lysed by high pressure with a french press. The cell lysate was centrifuged 30 min at 39,000 g at 4°C. The supernatant was loaded on a Ni-NTA column on a ÄKTA Prime purification system (GE Healthcare), and after washing with buffer A the protein of interest was eluted with 300 mM imidazole gradient. The fractions containing the

protein of interest were collected, then TEV protease was added according the estimation of the amount of the protein and then dialyzed against a buffer A without imidazole overnight at room temperature to perform the cleavage of tag. Then the sample was reloaded onto a Ni-NTA column to remove the tag, TEV protease and uncleaved protein. Samples containing the protein were further purified by size exclusion chromatography on a Superdex 75 column equilibrated with NMR buffer (50 mM sodium phosphate pH 6.5, 1mM BME, 50 μ M EDTA) at 4°C and concentrated to 1 mM with a Vivaspin 10.000 MWCO.

2.2. Cloning, expression and purification of RRM1 and RRM2 of hnRNP A1

The sequence of DNA encoding RRM1 and RRM2 individually (RRM1 from residues 1-105 and RRM2 from residues 91-196) were subcloned by PCR amplification into the same modified pET28a vector as above. Recombinant proteins were overexpressed and purified as for the construct containing both RRMs. Samples containing proteins were further purified by size exclusion chromatography on Superdex 75 column equilibrated with NMR buffer (25 mM sodium phosphate pH 6.5, 1mM BME, 50 μ M EDTA) at 4°C and concentrated to 1 mM with a Vivaspin 10.000 MWCO.

2.3. ssDNA purification and desalting

The ssDNAs were purchased from Dharmacon. The 15-mer ssDNA (5'-CAGGGAT-TTGGGGAC-3'), the 7-mer ssDNAs (5'-CAGGGAT-3') and the three 8-mer ssDNAs (5'-TGGGGAAT-3', 5'-CAGGGATC-3', 5'-CTGGGCAC-3') were purified under denaturing conditions using the ultimate 3000 HPLC system with an anion-exchange preparative DNAPac PA100 Nucleic Acid Column (both Thermo Fisher Scientific) as described before²³. We performed the purification at 85°C with a flow rate of 20ml/min. The column was equilibrated with the buffer containing 6M urea, 12.5 mM Tris/HCL, pH 7.4. The sample was loaded into the sample loop and then eluted from the column with the buffer containing 6M urea, 12.5 mM Tris/HCL, 500 mM NaClO₄, pH 7.4 up to 50% of NaClO₄ while collecting 10ml fractions. The fractions were analyzed by SDS-PAGE (20% acrylamide, 8M urea). The ssDNA was visualized with 0.1% toluidine. The fractions containing the ssDNA were desalting using a ÄKTA prime FPLC system (GE Healthcare) equipped with three 5 ml DEAE weak anion-exchange columns in series²⁴. The columns were equilibrated with the buffer containing 20 mM ammonium bicarbonate, pH 7.5 (degassed for 30 min). The sample was diluted twice, then loaded into the sample loop and injected with the buffer at a flow rate of 5 mL/min. A steep gradient containing 100% of 2.53 M ammonium bicarbonate, pH 8 was applied to collect 10 mL fractions. Fractions with the ssDNA were collected and then lyophilized.

2.4. NMR experiments

All NMR experiments were carried out using Bruker Avance III HD 700, 850 and 950 MHz spectrometers each equipped with cryoprobes. Data acquisition was performed at 298K, samples were measured in NMR buffer with 10% D₂O. Data were processed using Topspin 3.2/3.5 (Bruker) and analyzed with Sparky (<http://www.cgl.ucsf.edu/home/sparky/>). Protein resonance assignments were taken from BioMagResBank under accession number 18728²². The buffer used (25 mM sodium phosphate pH 6.5, 1mM DTT) was similar to our NMR buffer (see above) so that resonance assignments could be directly transferred to our spectra. CSPs from NMR titration experiments (2D ¹H-¹⁵N HSQC spectra) were calculated using the formula: $\Delta\delta=[(\delta H^N)^2 +$

$(\delta N/6.51)^2]^{1/2}$. HSQC titrations were carried out using ^{15}N -labeled hnRNP A1 protein constructs and unlabelled ssDNA. Spectra were recorded at the following molar ratios Protein/ssDNA: 1/0.2; 1/0.4; 1/0.6; 1/0.8; 1/1; 1/2.

2.5. ITC measurements

ITC experiments were performed on a VP-ITC instrument (Microcal) at 25°C. The calorimeter was calibrated according to the manufacturer's instructions. Protein and ssDNA samples were dialyzed against the NMR buffer. Concentrations of proteins and ssDNAs were determined using optical absorbance at 280 and 260 nm, respectively. The ssDNA (200 μM) was injected into the sample cell containing the protein at a concentration of around 20 μM . Titrations consisted of 20 injections of 2 μL except for the first injection (1.5 μL) with a 2 min spacing between each injection. Data were fitted with one site binding model using Microcal Origin.

2.6. Splicing assay

Plasmid TG11-T5 has been previously described by Niksic et al.²⁵. Mutations were inserted using complementary nucleotides and Liposome-mediated transfections of 3×10^5 human HeLa cells were performed using DOTAP Liposomal Transfection Reagent (Alexis Biochemicals) according to manufacturer instructions. After 18 hours the transfection medium was replaced with fresh medium and 24 hours later the cells were washed with PBS and RNA was purified using RNAwiz (Ambion). RT-PCR reactions to specifically amplify the minigene transcripts was performed as previously reported²⁶.

3. Results

3.1. *In vivo* splicing assay identifies a novel ESS in CFTR exon

The hnRNP A1 protein binds the SELEX sequence 5'-UAGGGA/U-3'²⁷ with high affinity and iCLIP studies of functional hnRNP A1 binding sites further indicate specificity for a 5'-UAGG-3' motif²⁸. Inspection of the CFTR 3'ss (5'-AACAGGGAUUUGGGGAU-3') displays similar AG-rich motifs at the conserved 3'ss AG and within the first codons of exon 9 (Figure 1A).

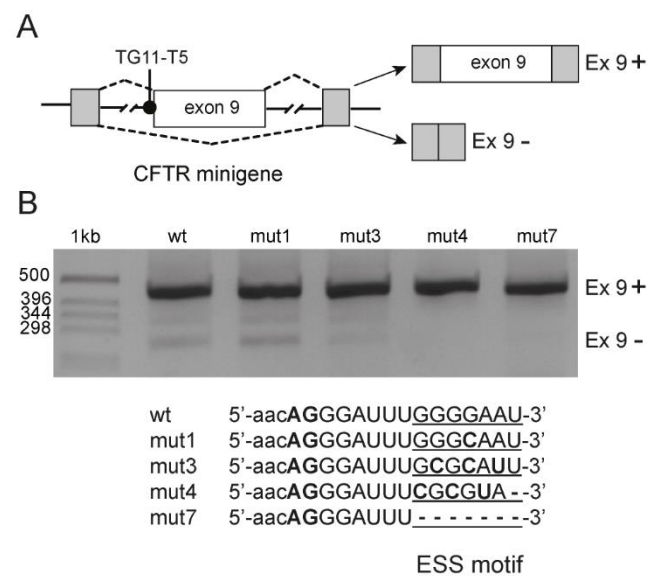


Figure 2. Minigene splicing assay identifies a novel ESS motif in CFTR exon 9. **(A)** Schematic representation of the CFTR exon 9 minigene construct used in the transfection experiments. Lines represent introns and boxes represent exons. The intronic TG11-T5 TDP-43 binding site upstream of exon 9 is also indicated. Dashed lines indicate alternative pre-mRNA processing of the minigene and the two major splice forms are shown on the right. **(B)** CFTR exon 9 inclusion upon mutation of the ESS motif in exon 9. The TG11-T5 CFTR exon 9 reporter minigenes (wt and mutants mut1, mut3, mut4, mut7) were transfected in HeLa cells. Progressive mutations of the ESS motif lead to full inclusion of exon 9.

To test the effect on splicing of these motifs, we altered the RNA sequence in exon 9 while keeping its coding potential for amino acids. The universal 3'ss **AG** was not mutated since it would delete the 3'ss but the second motif was altered progressively to delete the AG-rich motif (Figure 1A). The mutations were introduced into a CFTR exon 9 splicing minigene reporter system which contains the exon 9 sequence, the splicing junctions and part of the flanking introns with the TG11-T5 repeat at the 3'ss constituting the functional TDP-43 binding site which is indispensable for aberrant CFTR exon 9 splicing as previously described (Figure 2A)^{15, 19, 29}. In contrast to previous studies, the splicing assay was not coupled with RNAi-mediated knockdown of endogenous TDP-43 or hnRNP A1 and add-back of siRNA-resistant wild-type (wt) proteins so that the effect on splicing could be studied in a natural protein background. In addition, it was already previously shown that overexpressing hnRNP A1 in a TG11-T5 minigene context was able to increase the level of exon 9 skipping²⁶ and that the knockdown of hnRNP A1 would lead to 100% exon inclusion since both TDP-43 and hnRNP A1 are required for exon 9 skipping¹⁸⁻¹⁹. As a last consideration, we opted for the TG11-T5 repeat minigene because it reflects a borderline physiological polymorphic situation in humans. Minigenes with higher TG repeats (i.e. TG13) were not considered since the resulting increased binding of TDP-43 to these repeats could have masked the inhibitory role of hnRNP A1.

The wt 5'-GGGGAAU-3' motif and the mutant Gly to Gln codon change (mut1: GAA to CAA) had no effect on exon 9 exclusion (Figure 2B). In contrast, more severe disruption of the 5'-GGGGAAU-3' motif with a change of Gly-Glu to Ala-Asp (mut3: GGGGAA to GCGCAU) almost abolished aberrant splicing of exon 9. A change of the motif to Arg-Val (mut4: GGGGAA to CGCGUA) completely abolished aberrant exon 9 splicing and had the same effect as a complete deletion (mut7) of the ESS motif (Figure 2B). Since these mutations do not affect the UG-rich TDP-43 binding site in intron 8 and since TDP-43 binding is indispensable for aberrant splicing, we conclude that the 5'-GGGGAAU-3' motif constitutes a novel ESS in exon 9 which displays a sequence with a strong potential for hnRNP A1 binding (Figure 2B).

3.2. Binding studies of hnRNP A1 tandem RRM with the novel CFTR exon 9 ESS sequence

Next, we aimed to characterize the interaction between hnRNP A1 and the CFTR exon 9 ESS sequence identified by our splicing assay. We used a DNA sequence comprising the CFTR exon 9 ESS sequence and characterized the interaction using ITC measurements. The hnRNP A1 tandem RRM protein was titrated with 5'-CAGGGATTTGGGGAC-3' ssDNA sequence (Figure 3A). The ssDNA at a concentration of 200uM was added stepwise to a 28uM protein solution. The stoichiometry represented by the N parameter is 0.736 indicating that one protein molecule binds to one molecule of ssDNA. The value of the dissociation constant ($K_d = 233$ nM) reveals a strong affinity between the protein and the ESS motif. We note that the reaction is spontaneous with a favorable enthalpy and entropy. The enthalpy value $\Delta H = -2.37 \text{e}^4$ kcal/mol, is consistent with an exothermic reaction. The negative entropy value ($\Delta S = -56.6$ kcal/mol) indicates that the reaction weighs toward the formation of a more ordered structure such as a RNA-protein complex (Figure 3A). Next, we set out to investigate the role of the individual RRMs of hnRNP A1. RRM1 domain of the hnRNP A1 protein is titrated with the 5'-CAGGGATTTGGGGAC-3' ssDNA sequence (Figure 3B). The protein concentration in the cell is 25uM and the ssDNA concentration is 200uM. The N value of 0.496 indicative of two protein molecules binding to one ssDNA which is in line with two binding sites in the ESS motif. The K_d of 5.8uM reveals a weaker affinity as compared to the tandem RRMs. The RRM2 domain of the hnRNP A1 protein was also titrated with 5'-CAGGGATTTGGGGAC-3' ssDNA sequence (Figure 3C). The protein (25 uM) is in the cell and the ssDNA (200 uM) is again titrated to the protein solution. The N value of 0.452 also points toward two protein molecules binding to the two ESS binding sites. The $K_d = 6.8 \mu\text{M}$ also indicates a weak affinity. The fact that both isolated RRMs bind the ESS sequence with micromolar affinity as compared to the tandem RRMs which bind in the submicromolar range, is consistent with a cooperative binding mode of the tandem RRMs previously described for other ESS sites (Figures 3A-C)³⁰.

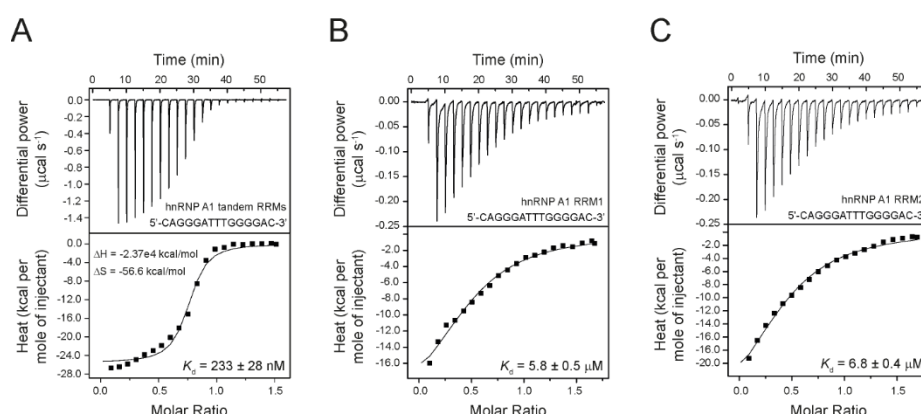


Figure 3. hnRNP A1 binds the novel bipartite ESS motif in CFTR exon 9. (A) Affinity of the hnRNP A1 tandem RRM1 for the bipartite ESS motif was determined by ITC. The determined K_d value is indicated with standard deviation as well as ΔH and ΔS . (B) Affinity of hnRNP A1 RRM1 for the bipartite ESS motif was determined by ITC. Two copies of hnRNP A1 RRM1 bind to the bipartite ESS motif. The determined K_d value is indicated with standard deviation. (C) Affinity of hnRNP A1 RRM2 for the bipartite ESS motif was determined by ITC. Two copies of hnRNP A1 RRM2 bind to the bipartite ESS motif. The determined K_d value is indicated with standard deviation.

3.3. Binding studies of the individual hnRNP A1 RRM1 with the separated ESS motifs

To further investigate the interaction of hnRNP A1 with the ESS sequence, we divided the ssDNA into two segments. We measured the affinities for both parts of the ESS motif (5'-CAGGGAT-3' and 5'-TGGGGAAT-3'). The ITC experiments shows that the binding of RRM1 for 5'-CAGGGAT-3' and 5'-TGGGGAAT-3' ssDNA occurs at a 1:1 ratio with a K_d values of 0.85 μ M and 7.25 μ M, respectively (Figures 4A-B).

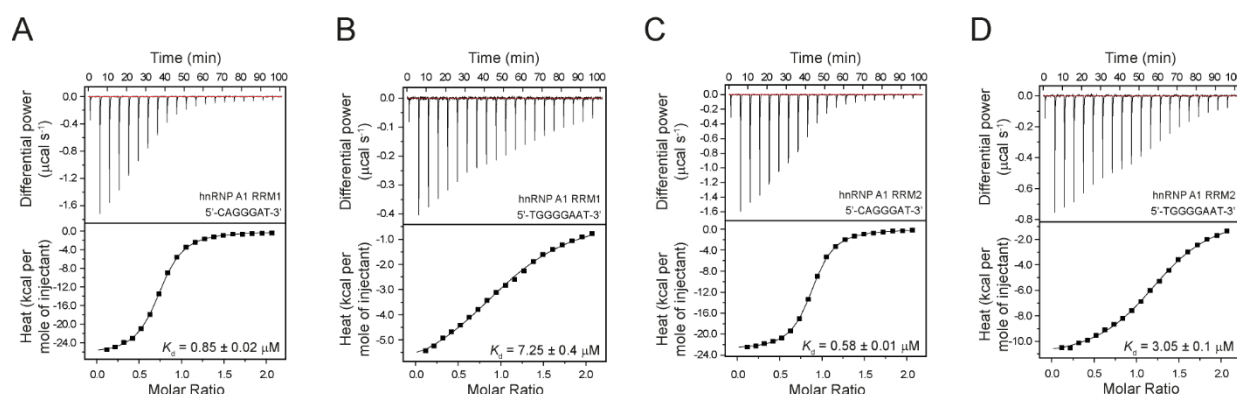


Figure 4. Both hnRNP A1 RRM1 and RRM2 bind the individual segments of the ESS motif of CFTR exon 9. (A-B) Affinity of hnRNP A1 RRM1 for the individual segments of the ESS motif were determined by ITC. 1:1 complexes are formed with the individual segments of the ESS motif. The determined K_d values are indicated with standard deviation. (C-D) Affinity of hnRNP A1 RRM2 for the individual segments of the ESS motif were determined by ITC. 1:1 complexes are formed with the individual segments of the ESS motif. The determined K_d values are indicated with standard deviation.

Thus, the RRM1 domain has an almost ten-fold higher affinity for the first part of the bipartite ESS motif. For RRM2, the ITC data show that the binding of 5'-CAGGGAT-3' and 5'-TGGGGAAT-3' DNA occurs also at a 1:1 ratio with a K_d value of 0.58 μ M and 3.05 μ M, respectively (Figures 4C-D). In conclusion, the RRM2 domain has a five-fold higher affinity for the first part of the ESS motif as well. Nonetheless, the similar binding affinities

of the individual RRM domains for the separated ESS motifs do not allow to reach a conclusion about the orientation of the ESS motif upon interaction with the tandem RRM scaffold of the hnRNP A1 protein.

3.4. Assessment of the ESS – hnRNP A1 interaction by solution NMR spectroscopy

We used the assignment of the hnRNP A1 protein published by the Allain group and superimposed them onto our ^1H - ^{15}N HSQC spectra of the amide fingerprint region of hnRNP A1²². We produced four complexes comprising each separated RRM domain with each of the separated ESS segments.

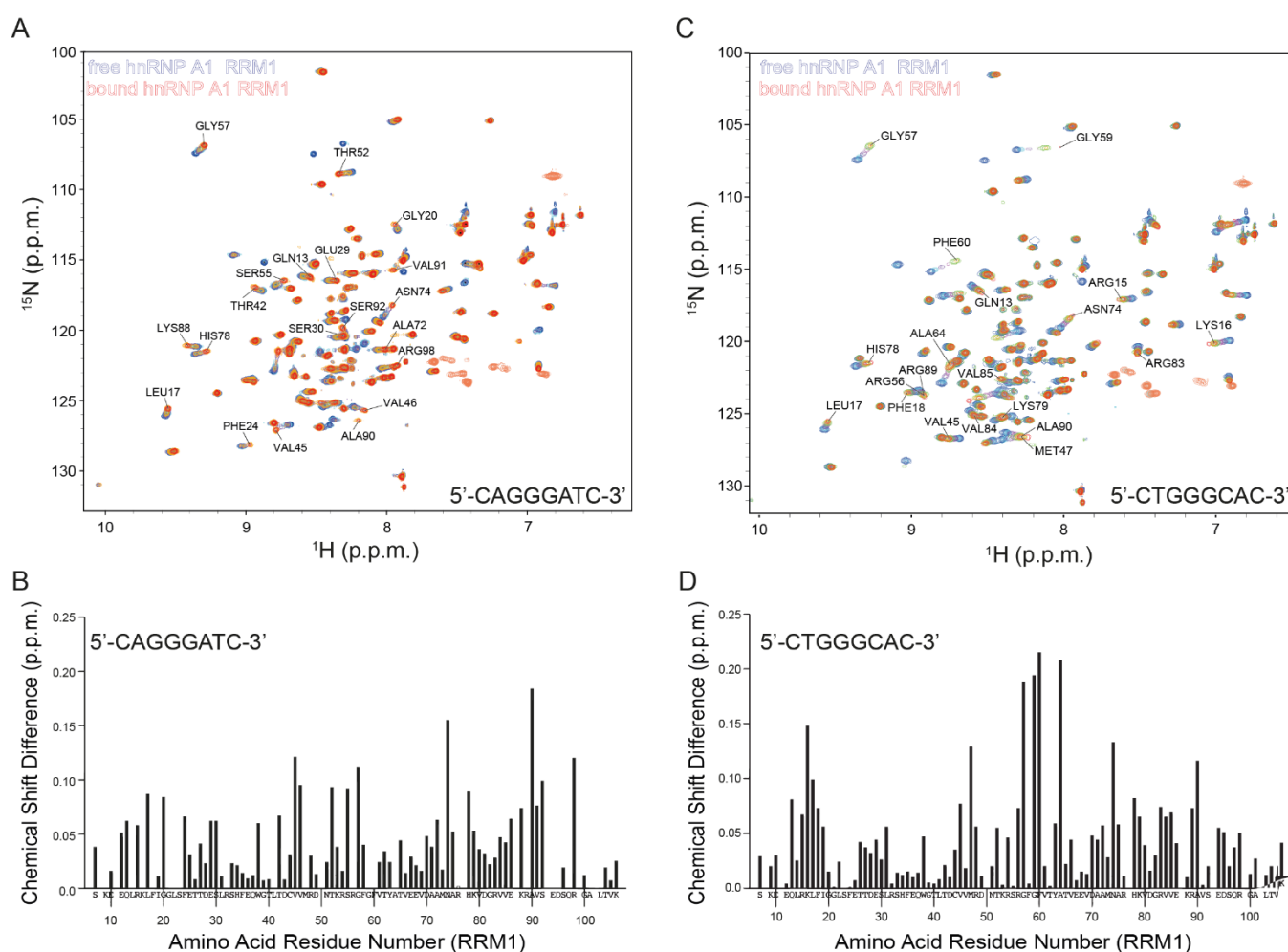


Figure 5. NMR titration of hnRNP A1 RRM1 reveals specific interactions with the individual segments of the ESS motif. (A and C) Superimposition of ^1H - ^{15}N HSQC spectra of ^{15}N -labelled hnRNP A1 RRM1 in the free (blue) and short ESS motif-bound form (red). The amide cross-peaks shifted upon interaction are labelled with amino acid type and residue number. (B and D) Combined chemical shift perturbations $\Delta\delta = [(\delta\text{H})^2 + (\delta\text{N}/6.51)^2]^{1/2}$ of hnRNP A1 RRM1 amide resonances upon short ESS motif binding versus the hnRNP A1 RRM1 amino acid sequence.

The ^1H - ^{15}N HSQC spectra of the four complexes were recorded at 298K. Since the second ESS segment contains four guanines which could form a G-quadruplex irrelevant to exon 9 splicing, and since previous studies showed that G-quadruplexes bind to the hnRNP A1 protein with high affinity³¹, we mutated the second ESS motif from 5'-

TGGGGAAT-3' to 5'-CTGGGCAC-3' according to the mut1 which had no effect on aberrant splicing (Figure 2B).

With the titration results, we could confirm the binding of the RRM1 domain with the 8-mer ssDNA model sequence (5'-CAGGGATC-3') (Figures 5A-B). The ssDNA binding of the free protein shown in blue, can be followed upon stepwise addition of the ssDNA (Figure 5A). The chemical shift perturbations (CSP) observed at each titration point indicate the potential ssDNA binding site on the protein and that the interaction occurs in a fast exchange regime in the NMR timescale. We considered all the CSPs > 0.06 more than average, indicating that the affected amino acids are either involved in the interaction with the 8-mer ssDNA model or indirectly affected by ssDNA binding. The affected amino-acids are Gln13, Leu17, Gly20, Phe24, Glu29, Ser30, Thr42, Val45, Val46, Thr52, Ser55, Gly57, Ala72, Asn74, His78, Lys88, Ala90, Val91, Ser92, and Arg98 (Figure 5B). Mapping these CSPs onto the structure shows that the interaction site of the ssDNA mostly comprises the beta sheet surface of the protein which is the canonical RNA interaction site on the RRM fold.

We subsequently recorded ¹H-¹⁵N HSQC spectra (298K) to monitor the binding of the RRM1 domain with the 8-mer DNA model (5'-CTGGGCAC-3') corresponding to the second part of the ESS motif (Figure 5C). We used the same ssDNA titration steps as for the first ESS motif to monitor ssDNA binding. Again, the nature of the CSPs indicated a fast exchange regime at the NMR time scale that the RRM1 domain binds also to the second part of the ESS motif. All the CSP > 0.06 are considered more than average and involve amino-acids Gln13, Arg15, Lys16, Leu17, Phe18, Val45, Met47, Arg56, Gly57, Gly59, Phe60, Ala64, Asn74, His78, Lys79, Arg83, Val84, Val85, Arg89, Ala90 (Figure 5D). The interaction site of the ssDNA comprises mostly the beta sheet of the protein which are again the canonical RNA interaction sites on the RRM fold.

Next, we monitored binding of RRM2 to the separate ESS motifs. Interaction occurs in a fast exchange regime in the NMR time scale. The NMR titration confirmed the interaction between RRM2 and the first ESS motif (5'-CAGGGATC-3') as well as the second ESS motif (5'-CTGGGCAC-3') (Figures 6A and 6C). Thus, our NMR titration experiments confirm the ITC results which showed that both individual RRMs interact with each part of the ESS motif. The CSPs > 0.05 observed upon binding of RRM2 to the 8-mer ssDNA (5'-CAGGGATC-3') involve amino acids Leu103, Thr104, Lys106, Lys107, Ile108, Val110, Lys131, Glu136, Thr139, Asp140, Gly142, Gly144, Phe151, Ser159, Asp161, His169, Lys180, Gln185, Met187, Ser189 (Figure 6B).

Likewise, the CSPs > 0.05 observed upon binding of RRM2 and the 8-mer ssDNA (5'-CTGGGCAC-3') involve a similar region on the protein (Arg98, Thr104, Val105, Lys106, Lys107, Val110, Gly112, Glu136, Ile137, Met138, Thr139, Asp140, Gly142, Ser143, Arg147, Phe151, Val152, Val178, Arg179, Lys180, Ala181, Leu182, Gln185, Met187, Ala188, Ser189,

Ala190 (Figure 6D). The interaction sites of both ssDNAs comprise mostly the beta sheet of the protein as well as the C-terminal helix.

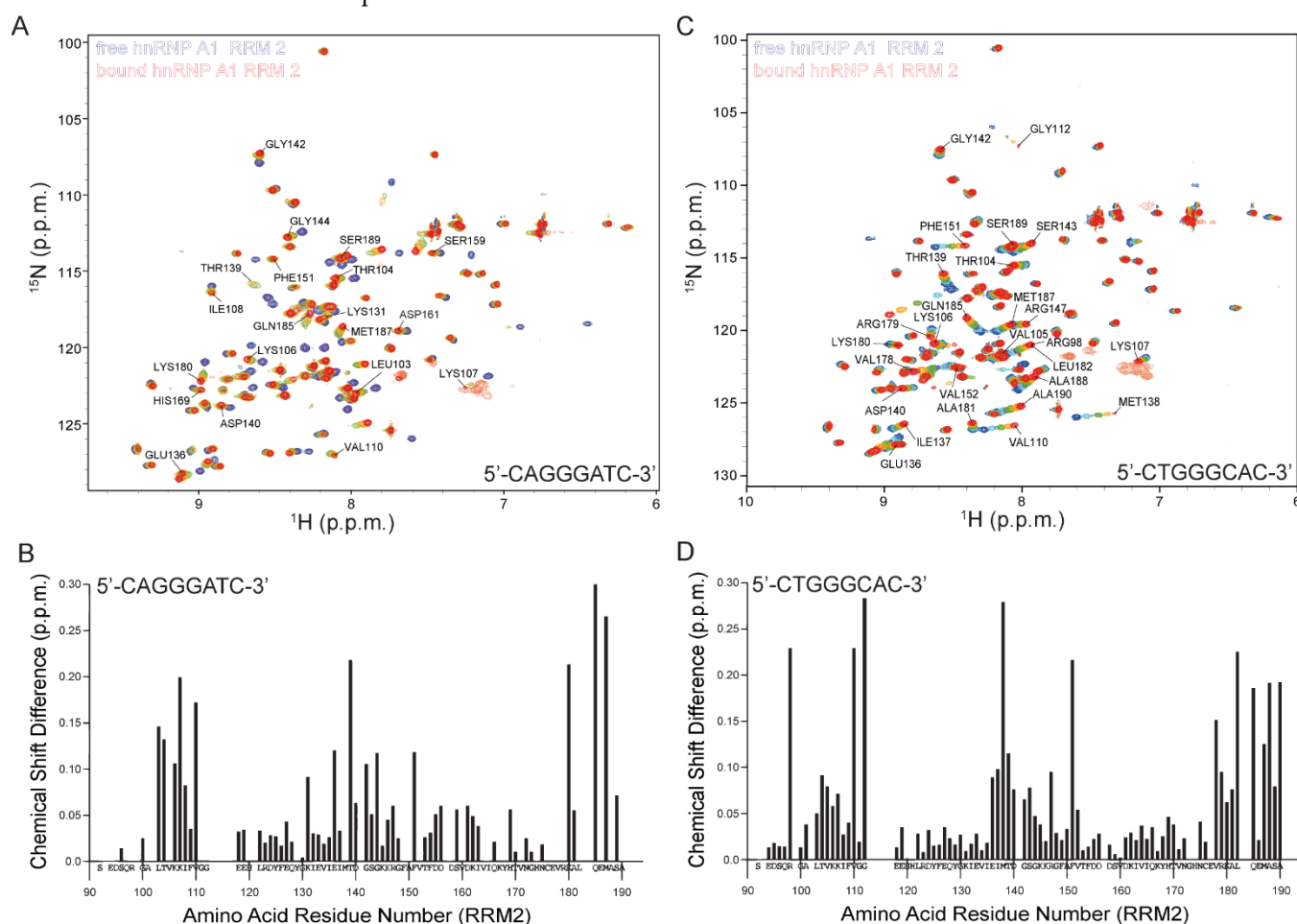


Figure 6. NMR titration of hnRNP A1 RRM2 reveals specific interactions with the individual segments of the ESS motif. (A and C) Superimposition of ^1H - ^{15}N HSQC spectra of ^{15}N -labelled hnRNP A1 RRM2 in the free (blue) and short ESS motif-bound form (red). The amide cross-peaks shifted upon interaction are labelled with amino acid type and residue number. (B and D) Combined chemical shift perturbations $\Delta\delta = [(\delta\text{H}^{\text{N}})^2 + (\delta\text{N}/6.51)^2]^{1/2}$ of hnRNP A1 RRM2 amide resonances upon short ESS motif binding versus the hnRNP A1 RRM2 amino acid sequence.

Next, we used the CSP results from the NMR titration experiments of the individual RRM2s and the separated ESS segments to visualize the potential ESS binding site on the hnRNP A1 tandem RRM2s structure determined by the Allain group²². The overall fold and arrangement of the tandem hnRNP A1 RRM2s presents a continuous binding site for the novel, bipartite CFTR exon 9 ESS segment (Figure 7). Despite that fact that our experiments did not reveal a preference of one of the RRM2s for a certain segment of the bipartite ESS sequence, both possible orientations of the ESS segment on the tandem RRM2s cover the same continuous protein surface and could thereby impair access of canonical splicing factors to the CFTR exon 3'ss and subsequently cause aberrant CFTR exon 9 splicing.

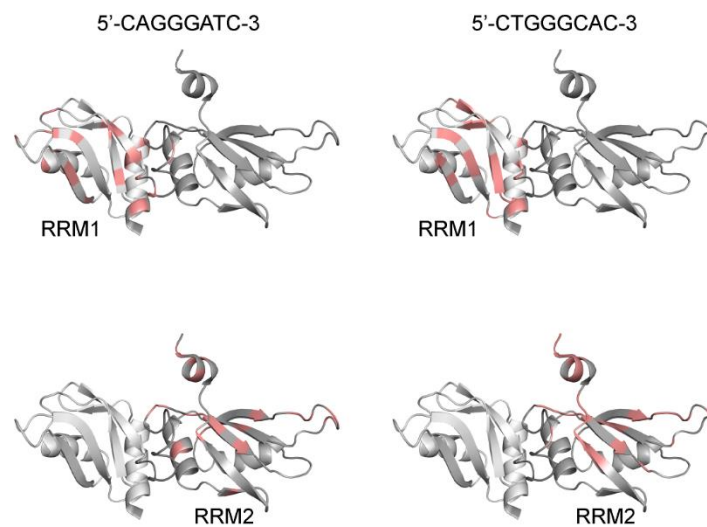


Figure 7. The hnRNP A1 tandem RRM1 and RRM2 display a continuous binding site for the novel CFTR exon 9 ESS motif. The CSPs from the titration experiments of the individual hnRNP A1 RRM1 and RRM2 and the short segments of the ESS motif are displayed onto the structure of the hnRNP A1 tandem RRM1 and RRM2 (PDB ID: 2lyv). RRM1 is coloured in light gray and RRM2 in dark gray. The CSPs are indicated onto the ribbon structure in salmon and the sequence of the corresponding ESS segments is displayed on top.

4. Discussion

Alternative splicing is an important corner stone of RNA-based, posttranscriptional regulation of gene expression. Production of diverse protein products through inclusion and exclusion of certain exons from the same pre-mRNA is a major contributor to the complexity of higher organisms. But at the same time, splicing also depends on correct RNA and protein products for proper functioning and thus variations in both components can lead to missplicing and disease³². One such variant is found in the polymorphic (TG)_m(T)_n locus upstream of CFTR exon 9 within intron 8 which leads to recruitment of TDP-43 to the pre-mRNA and subsequent TDP-43-dependent recruitment hnRNP A1. The UG-rich binding site of TDP-43 in the CFTR pre-mRNA is well characterized both functionally^{15,29} and structurally¹⁶⁻¹⁷ and it is known that hnRNP A1 is recruited to the 3'ss *via* a protein-protein interactions with TDP-43¹⁸⁻¹⁹. The molecular origin of CFTR exon 9 splicing inhibition *via* TDP-43-dependent recruitment of hnRNP A1, on the other hand, remained elusive.

Our study identifies a new ESS comprising a bipartite sequence motif at the CFTR exon 9 3'ss and the 5' end of exon 9 (5'-AGGGAUUUGGGGAAU-3'). We show that hnRNP A1 is able to specifically interact with this bipartite ESS motif and that the integrity of this site is instrumental for aberrant splicing of CFTR exon 9. Our minigene splicing assay shows that the wt sequence around the CFTR 3'ss produces several bands corresponding to the inclusion and skipping of exon 9. Four different mutations in exon 9 which alter the 5'-GGGGA-3' motif but leave the 3'ss intact, clearly show that variation of this sequence

affects the inclusion of exon 9 in the mRNA. Mutating the 5'-GGGGA-3' motif to 5'-CGCGT-3' as well as its complete deletion are associated with non-skipping of exon 9, i.e. the full inclusion into the final mRNA. Thus, these experiments reveal for the first time an ESS motif in CFTR exon 9 causative for exon 9 skipping.

We also investigated the RNA-binding mode of hnRNP A1 to the ESS that silences splicing of CFTR exon 9. The ITC and NMR titration experiments strongly suggest that both RRM1 and RRM2 of hnRNP A1 interact with the bipartite ESS motif. The affinities obtained from ITC experiments using the isolated RRM1 and single binding sites of the bipartite ESS motif are comparable, suggesting that there is no preferred sequence motif for the individual RRM1 determining the orientation of hnRNP A1 on the pre-mRNA target. Previous studies have shown that the hnRNP A1 protein can bind specifically and with high affinity to ssRNA or ssDNA sequences because there are no qualitative differences between them²¹ and several high-affinity sequences have also been discovered for hnRNP A1. Indeed, SELEX experiments isolated a high-affinity RNA sequence of 5'-UAGGGA/U-3' that is recognized by hnRNP A1²⁷ which is similar to the CFTR exon 9 ESS motif. The specific recognition of the motif was confirmed by an x-ray crystallography structure of hnRNP A1 tandem RRM1 and RRM2 bound to a telomeric ssDNA sequence (5'-TTAGGGTTAGGG-3')²¹. The structure revealed that both RRM1 and RRM2 of the same monomer interact with two different strands of ssDNA, in an antiparallel manner. In this structure, RRM1 binds to four nucleotides (5'-TAGG-3'), while RRM2 preferably binds to five nucleotides (5'-TTAGG-3'). The CFTR exon 9 ESS motif also contains two binding sites with similar sequence but our NMR and ITC data suggest that a single copy of hnRNP A1 binds both motifs *via* the tandem RRM1 and RRM2. Thus, the dimerization in the x-ray structure might be just caused by crystal packing forces and not be relevant for CFTR exon 9 skipping.

A study performed by Allain and colleagues has identified optimal recognition motifs that bind the individual RRM1 and RRM2³³. RRM1 binds to the sequence 5'-U/CAGG-3' and RRM2 to the sequence 5'-U/CAGN-3'. They also show that both RRM1 and RRM2 of hnRNP A1 bind to their RNA target containing the two AG-binding sites without dimerization of the protein and that hnRNP A1 binds the ISS-N1 pre-mRNA with directionality where RRM2 interacts with the 5' motif and RRM1 binds the 3' motif.

Another study investigated the interaction between hnRNP A1 to an RNA target, namely the pri-mir-18a. Again, this interaction involves both RRM1 and RRM2 and a region encompassing two 5'-UAG-3' motifs³⁴. In this study, it was demonstrated that cooperative binding of both domains to the related RNA motifs results in a strong enhancement of binding affinity and allows the unwinding of the RNA target as a stem loop. Our ITC experiments also reveal that individual RRM1 and RRM2 interact with the ESS in the low micromolar range while the tandem RRM1 and RRM2 display nanomolar affinity consistent with cooperative binding of both RRM1 and RRM2 to the bipartite ESS.

Further work is needed to determine the orientation of hnRNP A1 along the CFTR exon 9 ESS element. Binding this natural target with directionality might require the protein-protein interaction of full-length TDP-43 and hnRNP A1 mediated by their glycine-rich C-terminal domains¹⁸⁻¹⁹. Consistent with this idea, disrupting this interaction *via* mutations or deletion in the C-termini of TDP-43 as well as hnRNP A1, abolishes exon 9 skipping despite the fact that hnRNP A1 maintains its ability to bind to the target CFTR exon 9 ESS element. Studying the RNA-protein and protein-protein interactions within the entire aberrant splice site using full-length TDP-43 and hnRNP A1 is required to shed light on the precise mechanism of aberrant splicing occurring at CFTR exon 9.

Author Contributions:

P.J.L. designed the project. C.B. and M.Y.C. expressed and purified proteins and purified ssDNA. C.B. performed all NMR measurements, ITC experiments with 15-mer ssDNA and analyses. M.Y.C. performed all ITC experiments with individual RRM and shorter ssDNAs. C.S. performed all the minigene splicing assays and analyzed the data. C.B. wrote an initial draft, P.J.L. wrote the manuscript and made proper figures, E.B. and C.S. wrote the minigene splicing assay part and made corrections to the text. All authors discussed the results and approved the manuscript.

Funding:

P.J.L. acknowledges funding from a Marie Curie Action - Career Integration Grant (PCIG14-GA-2013-630758), an EMBO Installation Grant (3014), three Czech Science Foundation Grants (P305/15/21122S, P305/18/08153S and P305/22/20110S). E.B. is supported by AFM Telethon (project 23788).

Data Availability Statement:

The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments:

We wish to thank Prof. Radova Fiala from Josef Dadok National NMR Centre for the skillful NMR technical assistance. We acknowledge CF BIC and CF NMR, Instruct-CZ Centre, supported by MEYS CR (LM2018127) and European Regional Development Fund-Project „UP CIISB“ (No. CZ.02.1.01/0.0/0.0/18_046/0015974).

Conflicts of Interest: The authors declare no competing interests.

References

- Schmidt, B. Z.; Haaf, J. B.; Leal, T.; Noel, S., Cystic fibrosis transmembrane conductance regulator modulators in cystic fibrosis: current perspectives. *Clin Pharmacol* **2016**, *8*, 127-140.
- Zengerling, S.; Tsui, L. C.; Grzeschik, K. H.; Olek, K.; Riordan, J. R.; Buchwald, M., Mapping of DNA markers linked to the cystic fibrosis locus on the long arm of chromosome 7. *Am J Hum Genet* **1987**, *40* (3), 228-36.
- Kerem, B.; Rommens, J. M.; Buchanan, J. A.; Markiewicz, D.; Cox, T. K.; Chakravarti, A.; Buchwald, M.; Tsui, L. C., Identification of the cystic fibrosis gene: genetic analysis. *Science* **1989**, *245* (4922), 1073-80.
- Riordan, J. R.; Rommens, J. M.; Kerem, B.; Alon, N.; Rozmahel, R.; Grzelczak, Z.; Zielenski, J.; Lok, S.; Plavsic, N.; Chou, J. L.; et al., Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **1989**, *245* (4922), 1066-73.
- Rommens, J. M.; Iannuzzi, M. C.; Kerem, B.; Drumm, M. L.; Melmer, G.; Dean, M.; Rozmahel, R.; Cole, J. L.; Kennedy, D.; Hidaka, N.; et al., Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **1989**, *245* (4922), 1059-65.

6. Wahl, M. C.; Will, C. L.; Luhrmann, R., The spliceosome: design principles of a dynamic RNP machine. *Cell* **2009**, *136* (4), 701-18. 461 462
7. Wang, Z.; Burge, C. B., Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **2008**, *14* (5), 802-13. 463 464
8. Dreyfuss, G.; Matunis, M. J.; Pinol-Roma, S.; Burd, C. G., hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **1993**, *62*, 289-321. 465 466
9. Groman, J. D.; Hefferon, T. W.; Casals, T.; Bassas, L.; Estivill, X.; Des Georges, M.; Guittard, C.; Koudova, M.; Fallin, M. D.; Nemeth, K.; Fekete, G.; Kadasi, L.; Friedman, K.; Schwarz, M.; Bombieri, C.; Pignatti, P. F.; Kanavakis, E.; Tzetis, M.; Schwartz, M.; Novelli, G.; D'Apice, M. R.; Sobczynska-Tomaszewska, A.; Bal, J.; Stuhmann, M.; Macek, M., Jr.; Claustres, M.; Cutting, G. R., Variation in a repeat sequence determines whether a common variant of the cystic fibrosis transmembrane conductance regulator gene is pathogenic or benign. *Am J Hum Genet* **2004**, *74* (1), 176-9. 467 468 469 470 471
10. Castellani, C.; Cuppens, H.; Macek, M., Jr.; Cassiman, J. J.; Kerem, E.; Durie, P.; Tullis, E.; Assael, B. M.; Bombieri, C.; Brown, A.; Casals, T.; Claustres, M.; Cutting, G. R.; Dequeker, E.; Dodge, J.; Doull, I.; Farrell, P.; Ferec, C.; Girodon, E.; Johannesson, M.; Kerem, B.; Knowles, M.; Munck, A.; Pignatti, P. F.; Radojkovic, D.; Rizzotti, P.; Schwarz, M.; Stuhmann, M.; Tzetis, M.; Zielenski, J.; Elborn, J. S., Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *J Cyst Fibros* **2008**, *7* (3), 179-96. 472 473 474 475 476
11. Lee, Y.; Rio, D. C., Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **2015**, *84*, 291-323. 477
12. Matlin, A. J.; Clark, F.; Smith, C. W., Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **2005**, *6* (5), 386-98. 478 479
13. Singh, R. K.; Cooper, T. A., Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* **2012**, *18* (8), 472-82. 480
14. Buratti, E.; Baralle, F. E., Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J Biol Chem* **2001**, *276* (39), 36337-43. 481 482
15. Buratti, E.; Dork, T.; Zuccato, E.; Pagani, F.; Romano, M.; Baralle, F. E., Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* **2001**, *20* (7), 1774-84. 483 484
16. Kuo, P. H.; Doudeva, L. G.; Wang, Y. T.; Shen, C. K.; Yuan, H. S., Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res* **2009**, *37* (6), 1799-808. 485 486
17. Lukavsky, P. J.; Daujotyte, D.; Tollervey, J. R.; Ule, J.; Stuan, C.; Buratti, E.; Baralle, F. E.; Damberger, F. F.; Allain, F. H., Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat Struct Mol Biol* **2013**, *20* (12), 1443-9. 487 488
18. Buratti, E.; Brindisi, A.; Giombi, M.; Tisminetzky, S.; Ayala, Y. M.; Baralle, F. E., TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J Biol Chem* **2005**, *280* (45), 37572-84. 489 490 491
19. D'Ambrogio, A.; Buratti, E.; Stuan, C.; Guarnaccia, C.; Romano, M.; Ayala, Y. M.; Baralle, F. E., Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res* **2009**, *37* (12), 4116-26. 492 493
20. Jean-Philippe, J.; Paz, S.; Caputi, M., hnRNP A1: the Swiss army knife of gene expression. *Int J Mol Sci* **2013**, *14* (9), 18999-9024. 494
21. Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. M., Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **1999**, *13* (9), 1102-15. 495 496
22. Barraud, P.; Allain, F. H., Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM s influences the nucleic acid binding topology. *J Biomol NMR* **2013**, *55* (1), 119-38. 497 498
23. Zlobina, M.; Sedo, O.; Chou, M. Y.; Slepankova, L.; Lukavsky, P. J., Efficient large-scale preparation and purification of short single-stranded RNA oligonucleotides. *Biotechniques* **2016**, *60* (2), 75-83. 499 500
24. Easton, L. E.; Shibata, Y.; Lukavsky, P. J., Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **2010**, *16* (3), 647-53. 501 502

25. Niksic, M.; Romano, M.; Buratti, E.; Pagani, F.; Baralle, F. E., Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum Mol Genet* **1999**, *8* (13), 2339-49.
26. Pagani, F.; Buratti, E.; Stuardi, C.; Romano, M.; Zuccato, E.; Niksic, M.; Giglio, L.; Faraguna, D.; Baralle, F. E., Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *Journal of Biological Chemistry* **2000**, *275* (28), 21041-21047.
27. Burd, C. G.; Dreyfuss, G., RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* **1994**, *13* (5), 1197-204.
28. Bruun, G. H.; Doktor, T. K.; Borch-Jensen, J.; Masuda, A.; Krainer, A. R.; Ohno, K.; Andresen, B. S., Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biol* **2016**, *14*, 54.
29. Ayala, Y. M.; Pagani, F.; Baralle, F. E., TDP43 depletion rescues aberrant CFTR exon 9 skipping. *FEBS Lett* **2006**, *580* (5), 1339-44.
30. Okunola, H. L.; Krainer, A. R., Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol* **2009**, *29* (20), 5620-31.
31. Liu, X.; Ishizuka, T.; Bao, H. L.; Wada, K.; Takeda, Y.; Iida, K.; Nagasawa, K.; Yang, D.; Xu, Y., Structure-Dependent Binding of hnRNPA1 to Telomere RNA. *J Am Chem Soc* **2017**, *139* (22), 7533-7539.
32. Cooper, T. A.; Wan, L.; Dreyfuss, G., RNA and disease. *Cell* **2009**, *136* (4), 777-93.
33. Beusch, I.; Barraud, P.; Moursy, A.; Clery, A.; Allain, F. H., Tandem hnRNP A1 RNA recognition motifs act in concert to repress the splicing of survival motor neuron exon 7. *Elife* **2017**, *6*.
34. Kooshapur, H.; Choudhury, N. R.; Simon, B.; Muhlbauer, M.; Jussupow, A.; Fernandez, N.; Jones, A. N.; Dallmann, A.; Gabel, F.; Camilloni, C.; Michlewski, G.; Caceres, J. F.; Sattler, M., Structural basis for terminal loop recognition and stimulation of pri-miRNA-18a processing by hnRNP A1. *Nat Commun* **2018**, *9* (1), 2479.

Appendix B: Staufen1 reads out structure and sequence features in ARF1 dsRNA for target recognition.

Staufen1 reads out structure and sequence features in ARF1 dsRNA for target recognition

Deepak Kumar Yadav[†], Dagmar Zigáčková[†], Maria Zlobina, Tomáš Klumpler, Christelle Beaumont, Monika Kubíčková, Štěpánka Vaňáčová and Peter J. Lukavsky*

Central European Institute of Technology, Masaryk University, Kamenice 753/5, 62500, Brno, Czech Republic

Received July 30, 2019; Revised November 22, 2019; Editorial Decision November 26, 2019; Accepted December 05, 2019

ABSTRACT

Staufen1 (STAU1) is a dsRNA binding protein mediating mRNA transport and localization, translational control and STAU1-mediated mRNA decay (SMD). The STAU1 binding site (SBS) within human ADP-ribosylation factor1 (ARF1) 3'UTR binds STAU1 and this downregulates ARF1 cytoplasmic mRNA levels by SMD. However, how STAU1 recognizes specific mRNA targets is still under debate. Our structure of the ARF1 SBS–STAU1 complex uncovers target recognition by STAU1. STAU1 dsRNA binding domain (dsRBD) 4 interacts with two pyrimidines and one purine from the minor groove side via helix α 1, the β 1– β 2 loop anchors the dsRBD at the end of the dsRNA and lysines in helix α 2 bind to the phosphodiester backbone from the major groove side. STAU1 dsRBD3 displays the same binding mode with specific recognition of one guanine base. Mutants disrupting minor groove recognition of ARF1 SBS affect *in vitro* binding and reduce SMD *in vivo*. Our data thus reveal how STAU1 recognizes minor groove features in dsRNA relevant for target selection.

INTRODUCTION

STAU1 is a dsRNA-binding protein (dsRBP) mediating mRNA transport and localization as well as regulation of mRNA processing, stability and translational efficiency (1). Important functions of the protein were first observed in *Drosophila melanogaster* where it was shown to be essential for the establishment of the anterior-posterior body pattern (2–5). Two paralogs, STAU1 and STAU2, and isoforms thereof have been described in mammals. STAU1 is found in most tissues, whereas STAU2 is preferentially present in the brain (6–8). The two STAU paralogs differ mainly in the number of dsRBDs. Both contain dsRBD2, 3 and 4, of which dsRBD3 and 4 adopt the canonical α – β – β – β – α dsRBD fold with three principal dsRNA interaction modules (Figure 1A): helix α 1 and loop β 1– β 2 interacting with

dsRNA from the minor groove while conserved lysines in helix α 2 insert into the major groove (9). STAU1 lacks the dsRBD1, while STAU2 has a truncated dsRBD5 and both proteins contain a tubulin-binding domain (TBD) and a STAU-swapping motif (SSM) for homo- and heterodimerization (6,10–11). STAU proteins are mainly cytoplasmic with enrichment in the periplasmic region and at the rough endoplasmic reticulum where they associate with translating ribosomes depending on protein-protein interactions *via* dsRBD4 and TBD and on RNA-protein interactions *via* mRNA and dsRBD3 (6,12). Moreover, a bipartite NLS can target STAU1 to the nucleus where it was linked to the regulation of alternative splicing and nuclear export (13,14). Both STAU paralogs bind different, only partially overlapping subsets of mRNA substrates with protein functions in transport, transcription and cell-cycle control (15–20). The target mRNAs clearly display enrichment in GC-content and secondary structure in their 3'UTRs (Figure 1B) and the structure of *Drosophila melanogaster* dmSTAU dsRBD3 revealed only RNA phosphodiester backbone interactions with an artificial 12 base pair (bp) stem-loop (9). As a consequence, it is still unclear how STAU proteins recognize many specific mRNA targets in diverse post-transcriptional gene expression pathways.

Beside the well-established roles of STAU in mRNA transport and localization as well as translational control, SMD is the latest addition to the functional repertoire of STAU proteins. In contrast to nonsense-mediated mRNA decay (NMD) (21), a quality control pathway that degrades newly synthesized mRNAs with premature termination codons, SMD regulates mRNA abundance *via* STAU–3'UTR interactions (22). SMD is a strictly translation-dependent process which requires a double-stranded STAU binding site (SBS) within the 3'UTR downstream of the stop codon and recruitment of the NMD factor Upf1 *via* STAU dsRBD4 and its TBD (23,24). A *bona fide* SMD target is characterized by an upregulation of the mRNA with increased mRNA half-life upon depletion of STAU and Upf1 and a SBS downstream of the termination codon (24). SBS can be formed either within the 3'UTR through intramolecular base pairing of sequences up to 1kb apart (20)

*To whom correspondence should be addressed. Tel: +420 54949 5076; Email: peter.lukavsky@ceitec.muni.cz

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

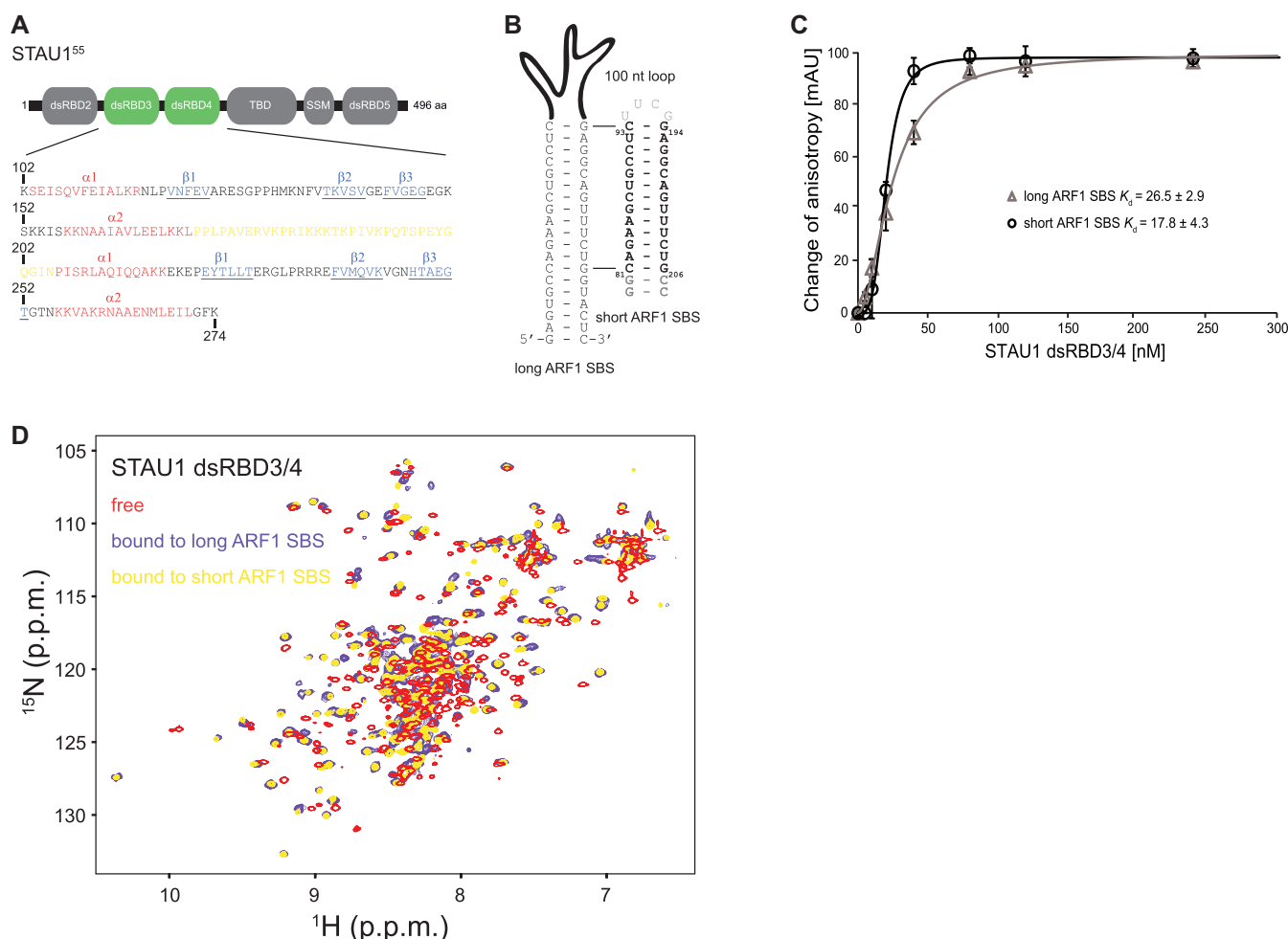


Figure 1. Interaction studies of STAU1 dsRBD3/4 with ARF1 SBS dsRNA. (A) Schematic representation of the human STAU1 domains. STAU1 dsRBD3/4 is located between dsRBD2 and the TBD, the SSM and dsRBD5. Numbering according to the full-length protein sequence (O95793–2). The sequence of recombinant STAU1 dsRBD3/4 used in this study (aa 102–274) is shown. β -strands are shown in blue, α -helices in red and linker amino acids in yellow. (B) The 19 bp SBS dsRNA of human ARF1 3'UTR is shown together with a short construct capped by a UUCG tetraloop which was used for structure determination. Numbering as in (24). (C) K_d values of STAU1 dsRBD3/4 measured by FA using 3'-fluorescein-labeled 19 bp and short ARF1 SBS dsRNA both capped with a UUCG tetraloop. The graph shows the average of three independent measurements with error bars. (D) ^1H - ^{15}N HSQC of free STAU1 dsRBD3/4 and bound to 19 bp ARF1 SBS dsRNA or the shorter construct. Large chemical shift perturbations occur in both protein domains and are virtually the same for both the 19 bp and the short ARF1 SBS dsRNA.

or by intramolecular base pairing of Alu elements within 3'UTRs with Alu elements of cytoplasmic, polyadenylated long noncoding RNAs (lncRNA) or even Alu elements in other mRNAs (25,26). The best characterized SMD target is ARF1 mRNA which contains a 19 bp stem-loop within the 3'UTR (Figure 1B) required for STAU1 binding and this sole interaction downregulates ARF1 mRNA levels *in vivo* (23,24).

The recent crystal structure of the complex of human STAU1 RBD3/4 with 19 bp ARF1 SBS at 2.9 Å resolution uncovered that the tandem dsRBDs dimerize upon RNA binding and recognize A-form dsRNA mainly through electrostatic interactions with the RNA backbone (27). The study showed that residues from both dsRBDs also make direct minor groove contacts with G and C bases in Watson–Crick base pairs of the target dsRNA, suggesting some sequence specificity. In the crystal structure, two copies of STAU1 dsRBD3/4 were bound to the ARF1 SBS

dsRNA forming an interface between the dsRBDs. However, mutations of residues at the interface did not affect co-operative binding of the dsRBDs (27). Moreover, the crystal structure contained one partially disordered dsRBD4 copy and a second completely disordered copy of dsRBD4. This could suggest steric clashes within the crystal lattice which might influence the positioning of the individual dsRBDs along the target dsRNA (27).

To resolve these issues and gain new insights into dsRNA target recognition by STAU1, we determined the solution structure of the human ARF1 SBS dsRNA - STAU1 dsRBD3/4 complex by high-resolution nuclear magnetic resonance (NMR) spectroscopy. Our structure reveals that STAU1 dsRBD3 and dsRBD4 bind to opposite sides of the ARF1 SBS dsRNA without forming a protein-protein interface. STAU1 dsRBD3/4 directly reads out RNA bases of the ARF1 SBS dsRNA via the minor groove and both dsRBDs participate in the target recognition. STAU1

dsRBD4 binds two pyrimidines and one purine on opposite strands via helix $\alpha 1$ and the $\beta 1$ - $\beta 2$ loop anchors dsRBD4 at the end of the dsRNA in the minor groove. Lysines in helix $\alpha 2$ bind to the phosphodiester backbone from the major groove. STAU1 dsRBD3 displays the same conserved binding mode and recognizes one guanine base via helix $\alpha 1$ interaction in the minor groove. STAU1 mutations of the amino acids involved in minor groove recognition of ARF1 SBS dsRNA only mildly affect RNA binding affinity *in vitro* but strongly impact SMD target mRNA binding *in vivo* and lead to upregulation of mRNA levels.

MATERIALS AND METHODS

The cDNA cloning of human STAU1 dsRBDs and PCR site-directed mutagenesis

The cDNA fragments encoding human STAU1 dsRBD3/4 (102–274) and subdomains (dsRBD3+11aa linker: 102–181, dsRBD4: 205–274) were prepared by polymerase chain reaction (PCR) from Flag-hSTAU1(20) using appropriate primers (Supplementary Table S2) and subcloned into pET28a vector (Novagen) with a TEV instead of the original thrombin cleavage site and a N-terminal His₆-Strep-His₆-tag, His₆-tag or His₆-lipo-tag (28) using the NdeI and XhoI restriction sites. Protein mutants were obtained by PCR-based site-directed mutagenesis using pET28a dsRBD3/4 (102–274), dsRBD3+11aa linker (102–181) or dsRBD4 (205–274) plasmid as a DNA template and following the QuickChange protocol (Stratagene) and efficient one-step mutagenesis protocol (29).

Purification of the recombinant STAU1 dsRBD proteins

The dsRBD3/4 (102–274) recombinant protein, subdomains (dsRBD3+11aa linker: 102–181, dsRBD4: 205–274) and mutant proteins were expressed in *Escherichia coli* BL21-CodonPlus (DE3)-RIL BL21 competent cells (Novagen) in LB rich or M9 minimal media supplemented with ¹⁵NH₄Cl or ¹⁵NH₄Cl and ¹³C-glucose. Cultures were grown to OD₆₀₀ 0.8–1.2 at 37°C, then set to 16°C for 30 min. Protein expression was induced by addition of 1 mM isopropyl β -D-thiogalactoside (IPTG) and further incubation for 14 to 18 h at 16°C. Cells were harvested by centrifugation at 4°C for 10 min at 2600 g and the cell pellet was resuspended in His A buffer (20 mM HEPES pH 7.5, 1 M NaCl, 10% glycerol (w/v), 30 mM imidazole and 3.5 mM 2-mercaptoethanol) with a tablet of protease inhibitor cocktail (cOmplete, EDTA-free, Roche). After lysis with Emulsiflex (Avestin) and centrifugation at 4°C for 30 min at 30 000 g, the supernatant was loaded onto two 5 ml HiTrap chelating columns (HiTrap, GE Healthcare) in series charged with nickel sulfate. All column chromatography was performed with an ÄKTA prime FPLC system equipped with a 50 ml superloop (GE Healthcare) at 4°C. After loading, the column was washed with 50 ml of lysis buffer followed by elution of the protein over a 90 ml gradient with His B buffer (His A buffer with 500 mM NaCl and 0.3–1 M imidazole). The pooled fractions were dialyzed against 2 l of His A buffer (500 mM NaCl and no imidazole) in the presence of TEV protease (1 mg/100 ml) to cleave the affinity-tag at RT

overnight. TEV cleavage reaction mixtures were reloaded onto HiTrap chelating columns to remove the His₆-TEV protease and the cleaved His₆-Strep-His₆-tag as well as minor contaminating proteins. The pooled fractions were dialyzed against 2 l of buffer A (50 mM potassium phosphate buffer pH 6.8, 100 mM NaCl, 3.5 mM 2-mercaptoethanol), loaded onto one 5 mL Q-sepharose column (HiTrap Q HP, GE Healthcare) to remove any remaining bound nucleic acids from the proteins. The protein fractions from the flow-through of the column were pooled, and after addition of 10% glycerol (w/v) frozen and stored at –20°C.

Design of STAU1 dsRBD3 and dsRBD4 for protein ligation with Sortase A

For protein ligation by Sortase A enzyme (30,31), the cDNA fragments were prepared by PCR from dsRBD3/4 (102–274) using appropriate primers (Supplementary Table S2) and subcloned into pET28a vector (Novagen) with a TEV cleavage site and a N-terminal His₆-Strep-His₆-tag or His₆-lipo tag. We engineered the C-terminus of dsRBD3 to end with the sequence Leu174-Pro175-Ala176-Thr-Gly-His₆ (His₆-lipo-TEV-dsRBD3_linker_LPATG-His₆) and the N-terminus of dsRBD4 to start with Gly-Val177-Glu178-Arg179 (His₆-Strep-His₆-TEV-G_linker_dsRBD4). This inserts a threonine and a glycine between Ala176 and Val177 of the linker and results in the LPTXG sequence required for ligation by Sortase A (30,31).

Preparation of segmentally isotope-labeled STAU1 dsRBD3/4

The His₆-lipo-TEV-dsRBD3_linker_LPATG-His₆ fragment was expressed in M9 minimal medium supplemented with ¹⁵NH₄Cl and ¹³C-glucose and the His₆-Strep-His₆-TEV-G_linker_dsRBD4 fragment was expressed LB rich medium as described above for STAU1 dsRBD3/4. The His₆-Strep-His₆-TEV-G_linker_dsRBD4 fragment was purified as described above. To facilitate the purification of the ligated product as described before (31), the His₆-lipo-TEV-dsRBD3_linker_LPATG-His₆ fragment was not cleaved with TEV protease and thus only purified by one run of HiTrap chelating column chromatography. Purified uncleaved His₆-lipo-TEV-dsRBD3_linker_LPATG-His₆ was subjected to anion exchange chromatography to remove remaining bound nucleic acids and protein impurities as described above.

To determine optimal conditions for the ligation reaction, a series of small scale test ligations was performed varying the concentration of fragments and Sortase A by using a dialysis approach as previously described (31). The final ligation reaction was setup in ligation buffer (50 mM Tris pH 8.0, 300 mM NaCl, 30 mM CaCl₂, 10% glycerol (w/v), 3.5 mM 2-mercaptoethanol) and the reaction mixture was dialyzed against ligation buffer for 17–20 h. The ligation process was monitored by denaturing polyacrylamide gel electrophoresis (PAGE) and the product purified by two consecutive runs of HiTrap chelating chromatography with TEV protease cleavage between the runs as described above. Purified, ligated, segmentally isotope-labeled dsRBD3/4 was incubated with SUPERase In RNase inhibitor (Ambion)

and then further purified by size-exclusion chromatography with a Sephadex S-75 gel filtration column (GE Healthcare) using gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 3.5 mM 2-mercaptoethanol) at 4°C. Fractions corresponding to the ligated protein product were pooled, concentrated and dialyzed three times against 1 l NMR buffer (50 mM potassium phosphate pH 6.5, 250 µM ethylenediaminetetraacetic acid (EDTA)) using Slide-A-Lyzer 0.5 mL dialysis cassettes with 3500 MWCO (Thermo Scientific). The mass of the final product was confirmed by MALDI-TOF.

Transcription and purification of ARF1 SBS dsRNA

For unlabeled and isotope-labeled ARF1 SBS dsRNA a cDNA fragment was prepared by PCR using three primers (Supplementary Table S2) and subcloned into pUC18 as previously described (32–34). Unlabeled and ^{13}C -labeled RNA oligonucleotides were transcribed *in vitro* from plasmid DNA linearized with BbsI/BstV2 I (SibEnzyme) using T7 RNA polymerase as previously described (32–34). Optimal magnesium chloride concentration for efficient transcription was determined by 25 µl trial reactions. After 2 to 4 h of incubation at 37°C, the crude transcription reaction was stopped upon addition of EDTA to a final concentration of 50 mM. Purification of the ARF1 SBS dsRNA was performed with three 5 ml HiTrap DEAE-sepharose FastFlow columns (GE Healthcare) connected in series as previously described (33). Fractions corresponding to pure, monomeric ARF1 SBS dsRNA were pooled, dialyzed three times against 2 l 10 mM sodium phosphate buffer (pH 6.0), frozen and stored at –20°C. If required, fractions containing pure ARF1 SBS dsRNA were concentrated using Vivaspin 20 ml centrifugal devices with 5000 MWCO (Sartorius) and passed over a Sephadex S-75 gel filtration column (GE Healthcare) with gel filtration buffer at 4°C. Before the experiments stored RNA was thawed, concentrated as before and dialyzed three times against 1 l NMR buffer using Slide-A-Lyzer 0.5 or 3 ml dialysis cassettes with 3500 MWCO (Thermo Scientific). The precise concentration was then determined based on the base composition of the hydrolyzed RNA as previously described (35).

Preparation of RNA–protein complexes

Stored, defrosted STAU1 proteins were incubated with SUPERase In RNase inhibitor (Ambion) for 30 min at RT and subjected to size-exclusion chromatography with a Sephadex S-75 gel filtration column (GE Healthcare) using gel filtration buffer at 4°C to remove aggregates. Pure, monomeric protein fractions were pooled, concentrated to up to 500 µM and dialyzed three times against 1 l NMR buffer using Slide-A-Lyzer 0.5 ml dialysis cassettes with 3000 MWCO (Thermo Scientific). Final concentrations of proteins were determined using optical absorbance at 280 nm. Assembly of RNA–protein complexes was monitored by 2D ^1H - ^{15}N HSQC experiments titrating STAU1 proteins in a stepwise manner with ARF1 SBS dsRNA until the spectra remained unchanged (ratio 1:1). All further complexes were assembled in lab in same way by mixing RNA and protein.

NMR measurement

All NMR measurements for the RNA–protein complexes were performed in NMR buffer at 308K using Bruker Avance III HD 700, 850 and 950 MHz spectrometers all equipped with cryoprobes. Data were processed using Topspin 3.2/3.5 (Bruker) and analyzed with CARA (<http://cara.nmr.ch>) or Sparky (<http://www.cgl.ucsf.edu/home/sparky/>).

Protein sequence-specific backbone and side chain assignments were achieved using unlabeled ARF1 SBS dsRNA in complex with ^{13}C , ^{15}N -labeled dsRBD4 or dsRBD3+11aa linker in 5 to 10% D_2O . The following experiments were performed: 2D ^1H - ^{15}N HSQC, 2D ^1H - ^{13}C HSQC, 3D HNCA, 3D HNCO, 3D CBCACONH, 3D HNCACB, 3D HBHACONH, 3D HC(C)H TOCSY, 3D (H)CCH TOCSY and 3D NOESY ^1H - ^{13}C HSQC aliphatic (36). Aromatic proton assignments were performed using 3D NOESY ^1H - ^{13}C HSQC aromatic experiments. 3D NOESY ^1H - ^{15}N HSQC spectra were collected on complexes of unlabeled RNA with ^{15}N -labeled protein. 3D backbone and side chain experiments were performed with non-uniform sampling (NUS) schemes and processed with NMRpipe (37). For NUS data collection and processing, Poisson disk sampling scheme and the modified CLEAN processing scheme with the conservative sparsity of about 20% were used (38,39).

Initial ARF1 SBS dsRNA resonance assignments were performed in the free form of the RNA with 2D ^1H - ^1H TOCSY, 2D ^1H - ^1H NOESY in 5–10% or 100% D_2O using unlabeled RNA. Complete ribose and aromatic proton assignments for ARF1 SBS dsRNA were achieved using ^{13}C -labeled RNA to collect 2D ^1H - ^{13}C HSQC, 3D HC(C)H TOCSY, 3D NOESY ^1H - ^{13}C HSQC ribose and 3D NOESY ^1H - ^{13}C HSQC aromatic experiments in 100% D_2O optimized for RNA (40). ARF1 SBS dsRNA resonance ($\text{H1}'$ -H6/H8 walk) assignments in the ARF1 SBS dsRNA–dsRBD3/4 complex were performed with unlabeled ARF1 SBS dsRNA in complex with segmentally isotope-labeled dsRBD3/4 or dsRBD3+11aa linker by using filtered experiments as described above for in the ARF1 SBS dsRNA–dsRBD4 complex.

ARF1 SBS dsRNA resonance ($\text{H1}'$ -H6/H8 walk) assignments in the ARF1 SBS dsRNA–dsRBD4 complex were performed with unlabeled ARF1 SBS dsRNA in complex with ^{13}C , ^{15}N -labeled protein using 2D ^{13}C 1F-filtered 2F-filtered NOESY experiments (41) in 100% D_2O and ^{13}C -labeled ARF1 SBS dsRNA in complex with ^{15}N -labeled dsRBD4 using 2D ^1H - ^{13}C HSQC and 3D NOESY ^1H - ^{13}C HSQC ribose experiments in 100% D_2O optimized for RNA (40).

Intermolecular NOEs were obtained by recording 2D ^1H - ^1H , ^{13}C 2F-filtered NOESY and 3D ^{13}C F1-filtered F2-edited HSQC-NOESY (42) in 100% D_2O using unlabeled ARF1 SBS dsRNA in complex with ^{13}C , ^{15}N -labeled dsRBD4 or segmentally isotope-labeled dsRBD3/4. Unambiguous assignments of intermolecular NOEs from protein side chains to RNA ribose and aromatic protons were also obtained from 3D NOESY ^1H - ^{13}C HSQC ribose experiments recorded on ^{13}C -labeled ARF1 SBS dsRNA in complex with ^{15}N -labeled dsRBD4 in 100% D_2O .

NOESY spectra were recorded with a mixing time of 80 or 150 ms, the 3D (H)CCH- and HC(C)H TOCSY spectra with a mixing time of 23 ms and the 2D ^1H - ^1H TOCSY spectra with mixing times of 20 or 50 ms.

Structure calculation and refinement

The AtnosCandid software (43,44) integrated into the Unio software package was used to generate preliminary structures and a list of automatically assigned NOE distance constraints for ARF1 SBS dsRNA–protein complexes. For ARF1 SBS dsRNA in complex with dsRBD4, peak peaking and NOE assignments were performed on 3D NOESY (^{15}N - and ^{13}C -edited) spectra recorded on ^{15}N -labeled or ^{13}C , ^{15}N -labeled dsRBD4 in complex with unlabeled ARF1 SBS dsRNA. Seven iterations were performed and 200 independent structures were calculated at each iteration step. Structures were calculated with CYANA by adding the manually assigned intramolecular RNA and intermolecular RNA–protein distance restraints obtained from dsRBD4–ARF1 SBS dsRNA complex. For ARF1 SBS dsRNA in complex with dsRBD3/4, peak peaking and NOE assignments were performed on 3D NOESY (^{15}N - and ^{13}C -edited) spectra recorded on ^{15}N -labeled or ^{13}C , ^{15}N -labeled dsRBD3+11aa linker in complex with unlabeled ARF1 SBS dsRNA as above. In this case, structures were calculated with CYANA by adding the manually assigned intramolecular RNA, intermolecular RNA–protein distance restraints obtained from segmentally isotope-labeled dsRBD3/4–ARF1 SBS dsRNA and dsRBD4–ARF1 SBS dsRNA complexes, respectively, and intramolecular dsRBD4 protein restraints obtained from the structure calculation of the dsRBD4–ARF1 SBS dsRNA complex. For each CYANA run, 200 independent structures were calculated. In both cases, the 50 lowest energy structures were refined with the SANDER module of AMBER 7.0 using a simulated annealing protocol (45). The 20 best structures based on energy and restraint violation energies were analyzed with PROCHECK (46). Chemical shifts for dsRBD3/4–ARF1 SBS dsRNA complex were combined from dsRBD4–ARF1 SBS dsRNA (205–274) and dsRBD3+11aa linker–ARF1 SBS dsRNA (102–181) complexes and submitted to PDB OneDep deposition.

Fluorescence anisotropy

Fluorescence anisotropy (FA) was measured on a Tecan Microplate Reader Infinite F500 (Tecan, Austria) equipped with a plate reader using 96-well plates. All measurements were performed in NMR buffer at 35°C. Measurements were performed in 10 μl reaction volume in which 10 nM 3'-end fluorescein labeled ARF1 SBS dsRNA was titrated with STAU1 dsRBD WT or mutant proteins. The data points represent the average of three measurements. The fitting was performed with Origin software (OriginLab, Northampton, MA, USA) (47). All data were normalized for adequate visualization.

Cell culture and manipulation

Mammalian cells HEK293T-REx Flip-In (Invitrogen™) and T-REx-HeLa Flip-In (Invitrogen™) were maintained

in Dulbecco's-modified Eagle's medium supplemented with 10% fetal calf serum at 37°C in the presence of 5% CO_2 .

Preparation of constructs and stable cell lines with inducible expression of STAU1 variants

The coding sequence of human STAU1 (55 kDa) was PCR amplified using appropriate primers (Supplementary Table S2) and cloned into pcDNA5 FRT/TO-N-terminal-3 $\times$ FLAG using NotI and XhoI sites. We used site-directed mutagenesis and appropriate primers (Supplementary Table S2) to obtain mutants E135A/L138A, Q212A/Q215A, E135A/L138A/Q212A/Q215A. The cloning and mutagenesis were subsequently verified by sequencing analysis with primers (Supplementary Table S2). The individual constructs for STAU1 variants were transfected together with plasmid pOG44 (Invitrogen) to HEK293T-REx Flip-In cells and T-REx-HeLa Flip-In, respectively by using Lipofectamine 3000 (Invitrogen™) reagent. Cells with stably integrated constructs were selected with Hygromycin B according to the manufacturer's protocol.

RNA immunoprecipitation (RIP)

Cells grown at 15 cm plate were induced at 70% confluency with 200 ng/ml Doxycycline for 24 h. Cells were washed with ice cold phosphate-buffered saline (PBS) and collected into 1.5 ml tube and lysed for 10 min on ice in 750 μl lysis buffer (LB) containing 150 mM NaCl, 50 mM Tris pH 8, 1% Triton X-100, EDTA-free Complete Protease Inhibitor Cocktail (Roche), 1 mM DTT. Each tube contained 120 U RNase Inhibitor (Biotechnrabbit) before the lysis. Lysates were incubated with 4.3 U Turbo DNase (Fermentas) for 15 min at 37°C and insoluble fraction was removed by 10 min centrifugation 11 000 g at 4°C. Supernatants were pre-cleared with 50 μl magnetic beads without antibodies (Dynabeads Protein G, Invitrogen) for 1 h at 4°C rotating. 1/20 volume (37.5 μl) of the lysate was mixed with TriPure reagent (Roche) for the 'input' sample RNA extraction. 50 μl of FLAG-M2 Magnetic beads (Sigma) per sample were used for immunoprecipitation. FLAG-M2 magnetic beads were first washed three times with 500 μl LB, then pre-blocked with 1 μg of yeast tRNA resuspended in 100 μl of LB for 1 h at 4°C and again washed three times with 500 μl of LB. The pre-cleared lysates were gently mixed with the pre-blocked and washed FLAG-M2 beads and incubated for 2 h at 4°C rotating. Beads were washed once with 500 μl LB for 1 min and twice with 500 μl wash buffer (350 mM NaCl, 50 mM Tris pH 8, 1% Triton X-100, EDTA-free Complete Protease Inhibitor Cocktail - Roche, 1 mM DTT) for 1 min at 4°C. In the last step, the beads were resuspended in 500 μl LB and an aliquot of 20 μl was mixed with 4 $\times$ sodium dodecyl sulphate (SDS)-PAGE loading buffer to monitor immunoprecipitation efficiency by western blot. The LB was removed and the rest of the beads were mixed with 1 ml of the TriPure reagent (Roche) to extract the coprecipitated RNA. The 'elution' RNA was isolated according to manufacturer's instructions with minor changes: 1 μl GlycoBlue coprecipitant (ThermoFisher) was added for isopropanol-based precipitation and RNA was precipitated at least one hour at -80°C . The precipitated RNA was centrifuged for 1 h 12 000 g at 4°C, the pellet

was washed with 85% ethanol and centrifuged for 20 min 12 000 *g* at 4°C. RNA was then resuspended in 10 μ l of H₂O. The 'input' RNA was isolated according to manufacturer's instructions, resuspended in 16 μ l of H₂O, treated with 1.2 U Turbo DNase (Fermentas) and frozen in liquid nitrogen to inactivate the Turbo DNase. Input RNA (2 μ g) and 1/2 volume of RNA isolated from RIP eluate (5 μ l from the total volume of 10 μ l) was used as a template for cDNA synthesis by using Superscript III reverse transcriptase (Invitrogen) and random hexamers (Thermo Scientific) according to manufacturer's instructions. The obtained cDNA was subsequently analyzed by real-time quantitative PCR (RT-qPCR).

Western blot analysis

The protein lysates from whole cell lysate (input), unbound fraction (flow-through) and purified proteins (elution) samples from the RIP experiment were mixed with SDS-PAGE sample buffer, separated by SDS-PAGE and transferred to nitrocellulose membrane by semi-dry blotting (BioRad). Proteins were detected with protein and tag-specific antibodies, respectively. The FLAG antibody (Sigma) was used at dilution 1:3000, STAU1 antibody (Proteintech) at dilution 1:1500 and α -Tubulin antibody (Sigma) at dilution 1:5000.

Real-time PCR analysis

Quantitative PCR (qPCR) was performed with the LightCycler 480 SYBR Green I Master Mix (Roche) and gene-specific primers on LightCycler 480 (Roche). Three to four biological replicates were analyzed. Transcript abundances were calculated by the $\Delta\Delta C_t$ method (48). The input data were normalized to an internal control of the housekeeping gene GAPDH mRNA, elution data were normalized to the corresponding Input samples normalized to GAPDH. *P*-values were calculated by two-tailed paired *t*-test; *P*-values < 0.1 were considered significant.

RNA steady-state levels analysis

HEK293T-REx and T-REx-HeLa control and stable-cell lines expressing WT and mutant forms of STAU1 were grown at 6-well plates to 70% confluency and induced for 24 h with doxycycline, at the concentration 200 ng/ml for HEKs and 1000 ng/ml for HeLas. Cells were washed with cold PBS and collected in 500 μ l of TriPure reagent (Roche) per one well. The total RNA was isolated according to manufacturer's instructions. Two μ g of RNA was used as a template for cDNA synthesis by using Superscript III reverse transcriptase (Invitrogen™) and random hexamers (Thermo Scientific). The cDNA was subsequently analyzed by RT-qPCR. Three to six biological replicates were analyzed. Transcript abundances were calculated by the $\Delta\Delta C_t$ method (48). The data were normalized to an internal control of the housekeeping gene GAPDH mRNA and to the control cell line. *P*-values were calculated by two-tailed paired *t*-test; *P*-values < 0.1 were considered significant.

SAXS measurements of the STAU1–ARF1 SBS dsRNA complexes

Small angle X-ray scattering (SAXS) data acquisition is summarized in Supplementary Table S1. Data were collected using BioSAXS-1000, Rigaku (CEITEC, Czech Republic) and at the B21 instrument (Diamond Light Source, UK). BioSAXS-1000 datasets were collected at 308 K, six separate two-dimensional images were collected for buffer and sample with an exposure time of 10 min per image (Pilatus 100K, Dectris). Radial averaging, data reduction and buffer subtractions were performed using SAXSLab3.0.0r1, Rigaku. B21 datasets were collected in the size-exclusion mode (SEC-SAXS) at 288 K, where two-dimensional images with an exposure time of 3 s per image (Pilatus 2M, Dectris) were continuously collected during size exclusion chromatography of the sample. Five subsequent 3 s exposures have been averaged to reduce the noise level of the data. SEC-SAXS data were analyzed using Chromixs (49). All datasets were cropped from the first data point of the Guinier region determined by PRIMUS/qt ATSAS v.2.8.3 (50) to maximum of $q = 0.3 \text{ \AA}^{-1}$ for further analysis. Integral structural parameters and MW estimate (Supplementary Table S1) were determined using PRIMUS/qt. Evaluation of the solution scattering and fitting to experimental scattering curves was performed using CRY SOL ATSAS v.2.8.3, where automatic constant subtraction was allowed, while other parameters were kept at default.

RESULTS

Characterization of STAU1 dsRBD3/4–ARF1 SBS dsRNA complexes

The principal RNA binding module of STAU1 comprises dsRBD3/4 with canonical α – β – β – β – α topology separated by a basic 34 amino acid linker (Figure 1A) (6,8). The long ARF1 SBS dsRNA required for STAU1 binding *in vivo* comprises a continuous 19 bp stem with canonical Watson–Crick and two GU base pairs formed between nt 75–93 and 194–212 of the ARF1 3'UTR (Figure 1B) (23,24). Monitoring complex formation using FA measurements reveals high affinity binding ($K_d = 26.5 \pm 2.9 \text{ nM}$) of recombinant STAU1 dsRBD3/4 to the long ARF1 SBS dsRNA capped by a UUCG tetraloop (Figure 1C). Complex formation was also assessed by NMR titration experiments monitored by recording ¹H–¹⁵N HSQC spectra. Upon step-wise addition of long ARF1 SBS dsRNA to recombinant ¹⁵N-labeled STAU1 dsRBD3/4, many protein resonances broaden beyond detection indicative of binding kinetics in the intermediate to slow exchange regime (Supplementary Figure S1A). When a 1:1 protein to dsRNA ratio is reached, protein resonances sharpen again displaying numerous chemical shift changes indicative of stable complex formation (Figure 1D and Supplementary Figure S1A). To further characterize the complex in solution, we performed SAXS measurements on the STAU1 dsRBD3/4 bound to long ARF1 SBS dsRNA. The SAXS measurements reveal a 1:1 stoichiometry of the complex as determined by a Porod volume of 52800 $\text{\AA}^3$ (Supplementary Table S1).

A single dsRBD usually displays a register variability of 8–10 bp for sequence-specific minor groove contacts with a

dsRNA (51). The long ARF1 SBS dsRNA could therefore bind the STAU1 dsRBD3/4 either in a staggered manner as previously observed for the adenosine deaminase acting on RNA 2 (ADAR 2) bound to dsRNA (52) or in a side-by-side manner on opposite sides of the long ARF1 SBS dsRNA. To test these possibilities, we designed shorter ARF1 SBS dsRNA comprising either the top or bottom two-thirds of the long, 19 bp ARF1 stem (Supplementary Figure S1B). Complex formation monitored by recording ^1H - ^{15}N HSQC and ^1H , ^1H -TOCSY spectra shows the same spectra with minor differences only when STAU1 dsRBD3/4 is bound to the long and the top two-thirds of ARF1 SBS dsRNA but the complex of dsRBD3/4 with the bottom two-thirds of the long ARF1 SBS dsRNA displays different spectra indicative of an altered binding mode (Figure 1D, Supplementary Figure S1C-D and Table 1). FA measurements using either the top or the bottom two-thirds of ARF1 SBS dsRNA shows a higher affinity interaction for the top two-thirds ($K_d = 17.8 \pm 4.3$ nM) as compared to the bottom two-thirds ($K_d = 31.0 \pm 0.1$ nM) of the ARF1 SBS dsRNA (Figure 1C, Supplementary Figure S1E and Table 1). SAXS measurements of STAU1 dsRBD3/4-bound to the top two-thirds of ARF1 SBS dsRNA reveal a 1:1 stoichiometry of the complex as determined by a Porod volume of 38900 Å³ (Supplementary Table S1). These data suggest that both dsRBDs bind in the same 15 bp region formed between nt 81–93 and 194–206 of the ARF1 3'UTR (Figure 1B).

To assess the contributions of the individual STAU1 dsRBDs to the binding affinity, we designed two smaller protein constructs comprising dsRBD3 with 11 amino acids of the linker sequence (102–181) and dsRBD4 (205–274) and compared their binding affinity (Supplementary Figure S2A). FA measurements show that dsRBD3 binds with 2-fold higher affinity ($K_d = 41.4 \pm 1.5$ nM) to top two-thirds ARF1 SBS dsRNA as compared to dsRBD4 ($K_d = 74.8 \pm 17.6$ nM). In both cases, the affinity is lower than observed for the STAU1 dsRBD3/4 construct indicating weak cooperative binding of the tandem dsRBDs driven by the higher affinity of dsRBD3.

Design of STAU1 dsRBD3/4–ARF1 SBS dsRNA complexes for structure determination

The complexity of the ^1H - ^{15}N HSQC spectrum with highly overlapping resonances from the unstructured linker prompted us to simplify the spectra by employing a 'divide-and-conquer' approach working with the individual dsRBDs bound to the top two-thirds of ARF1 SBS dsRNA, termed short ARF1 (sARF1) SBS dsRNA hereafter, separately. We assessed complex formation of the individual STAU1 dsRBDs with sARF1 SBS dsRNA by recording ^1H - ^{15}N HSQC spectra of the individual subcomplexes and compared them to the ^1H - ^{15}N HSQC spectrum obtained for the STAU1 dsRBD3/4 complex. While the ^1H - ^{15}N HSQC spectrum of dsRBD4 displays an almost identical fingerprint of the protein amide resonances as compared to the tandem dsRBD3/4, dsRBD3 shows deviations and broadening of resonances indicating that the latter has different binding characteristics when taken out of the context of the tandem dsRBD3/4 protein construct (Supplementary Figure S2B). The difference is also apparent from ^1H , ^1H -

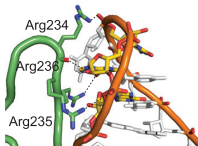
TOCSY spectra recorded on sARF1 SBS dsRNA bound to the tandem or individual dsRBDs. The spectra show that sARF1 SBS dsRNA adopts the same conformation when bound to dsRBD3/4 and dsRBD4 while peak doubling for several pyrimidines upon dsRBD3 binding indicates multiple binding registers or alternate binding orientations along the dsRNA or both (Supplementary Figure S2C).

To reveal the reason for this difference, we recorded 3D ^{13}C F1-filtered F2-edited HSQC-NOESY(42) spectra on the dsRBD3- and dsRBD4-ARF1 SBS dsRNA complexes separately and assigned intermolecular nuclear Overhauser effect (NOE)-derived contacts. While dsRBD4 binds sARF1 SBS dsRNA in one preferred orientation placing the helix $\alpha 1$ at the lower part of the sARF1 SBS dsRNA, dsRBD3 showed multiple binding registers with helix $\alpha 1$ binding the 5' end of the sARF1 SBS dsRNA and the middle part of the stem (Supplementary Figure S3A and B). This promiscuous, non-specific binding mode has also been observed for the isolated *Drosophila* dsRBD3 bound to an artificial 12 bp stem-loop (9). The data suggest that the dsRBD4-sARF1 SBS dsRNA can be determined in isolation while structure determination of the dsRBD3 interaction with sARF1 SBS dsRNA requires the context of the tandem dsRBD3/4 protein. To alleviate spectral overlap, we prepared segmentally labeled STAU1 dsRBD3/4 protein, where only dsRBD3 and part of the linker is ^{13}C , ^{15}N -labeled while the rest of the linker and dsRBD4 is unlabeled and therefore does not contribute to NMR signals in heteronuclear NMR experiments (Supplementary Figure S3C). When this segmentally labeled protein was used for complex formation with sARF1 SBS dsRNA, we observed no strong deviations and broadening of resonances in the ^1H - ^{15}N HSQC spectrum (Supplementary Figure S3D). Likewise, the intermolecular NOEs indicated one preferred orientation also placing the helix $\alpha 1$ of dsRBD3 at the middle part of the sARF1 SBS dsRNA (Supplementary Figure S3E). With these optimized protein-dsRNA complexes, we determined the NMR solution structure of the STAU1 dsRBD4-sARF1 SBS dsRNA subcomplex and of the STAU1 dsRBD3/4-sARF1 SBS dsRNA complex using standard NMR techniques (53,54).

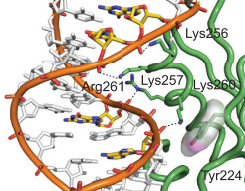
Structure of STAU1 dsRBD4-sARF1 SBS dsRNA complex reveals sequence readout from the minor groove

The final ensemble of the STAU1 dsRBD4-sARF1 SBS dsRNA complex is well-defined and is represented by twenty conformers with a root-mean-square (r.m.s.) deviation of 0.91 Å for all heavy atoms (Table 2 and Figure 2A). Resonance assignment of the sARF1 SBS dsRNA as well as unambiguous assignment of intermolecular contacts was aided by using ^{13}C , ^{15}N -labeled sARF1 SBS dsRNA in complex with ^{15}N -labeled STAU1 dsRBD4 in addition to ^{13}C , ^{15}N -labeled STAU1 dsRBD4-unlabeled sARF1 SBS dsRNA complex commonly used for extracting intermolecular NOEs and RNA assignments (53,54). This combined approach yielded a sufficient number of intermolecular (178) and intra-RNA (717) NOEs for high-resolution structure determination (Table 2).

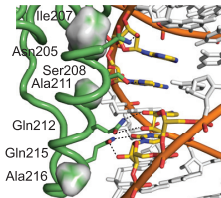
The STAU1 dsRBD4 adopts the canonical α - β - β - β - α dsRBD fold and interacts with the sARF1



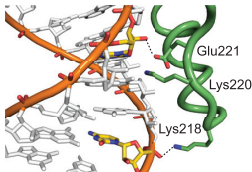
$\beta 1$ - $\beta 2$ loop



helix $\alpha 2$



helix $\alpha 1$



$\alpha 1$ - $\beta 1$ loop

Table 1. Binding affinities of wild-type and mutant STAU1 dsRBDs and ARF1 SBS dsRNA

STAU1 protein ^a	dsRNA ^b	K_d (nM) ^c	Affinity factor ^d
dsRBD3/4	ARF1 SBS (5'-GGCGAGUGCCAGAAGCUGUCUC CUUCGGGAGGCAGUUUCUGGUACUCGCC-3')	26.5 ± 2.9	1.5
dsRBD3/4	top 2/3 ARF1 SBS (5'-GGCAGAAGCUGCCUCUUC GGAGGCAGUUUCUGCC-3') = sARF1 SBS	17.8 ± 4.3	1.0
dsRBD3/4	bottom 2/3 ARF1 SBS (5'-GGAGUGCCAGAAGCUA CGGUUUCUGGUACUCC-3')	31.0 ± 0.1	1.7
dsRBD3	sARF1 SBS	41.4 ± 1.5	2.3
dsRBD4	sARF1 SBS	74.8 ± 17.6	4.3
dsRBD3 I105A S106A	sARF1 SBS	147.1 ± 35.6	3.6
dsRBD3 E110A L113A	sARF1 SBS	31.5 ± 1.7	0.8
dsRBD3 F109A	sARF1 SBS	98.2 ± 8.7	2.4
dsRBD4 Q212A Q215A	sARF1 SBS	121.2 ± 15.0	1.6
dsRBD4 S208A Q212A Q215A	sARF1 SBS	389.6 ± 38.4	5.2
dsRBD3/4 I105A S106A	sARF1 SBS	28.8 ± 2.7	1.6
dsRBD3/4 E110A L113A	sARF1 SBS	18.5 ± 4.3	1.0
dsRBD3/4 I105A S106A E110A L113A	sARF1 SBS	21.8 ± 6.4	1.2
dsRBD3/4 Q212A Q215A	sARF1 SBS	15.3 ± 0.7	0.9
dsRBD3/4 E110A L113A Q212A Q215A	sARF1 SBS	18.5 ± 1.1	1.0
dsRBD3/4 S208A	sARF1 SBS	28.0 ± 4.4	1.6
dsRBD3/4 S208A Q212A Q215A	sARF1 SBS	87.2 ± 3.0	4.9
dsRBD3/4 I105A S106A S208A Q212A Q215A	sARF1 SBS	43.7 ± 5.2	2.5
dsRBD3/4 K133A	sARF1 SBS	23.2 ± 9.7	1.3
dsRBD3/4 R234A R236A	sARF1 SBS	26.0 ± 9.1	1.5
dsRBD3/4 F109A	sARF1 SBS	29.7 ± 5.8	1.7
dsRBD3/4 R209A	sARF1 SBS	17.7 ± 4.1	1.0
dsRBD3/4 F109A R209A	sARF1 SBS	27.7 ± 5.5	1.6

^aSTAU1 protein construct and mutations are indicated.^bARF1 SBS dsRNA used for FA measurements are indicated.^cAll FA experiments were performed in technical triplicates and the reported K_d values are the average of three experiments with error bars. ^dAffinity factor for variant RNA or protein relative to top 2/3 ARF1 (sARF1) SBS dsRNA bound to wild-type STAU1 dsRBD3, dsRBD4 or dsRBD3/4, respectively.

SBS dsRNA in the expected three regions (Figure 2A). Region 3, the N-terminal end of helix $\alpha 2$ of STAU1 dsRBD4 contains the conserved KKNK-motif which makes contacts with the phosphate oxygens from the major groove side via its lysine side chains (Figure 2B and C) (51). In the case of STAU1 dsRBD4, these electrostatic interactions with the phosphodiester backbone occur in pairs with Lys256 hydrogen bonding the phosphate oxygens of G197 and C198 via the backbone amide and side chain amino groups, respectively, and Lys257 hydrogen bonding the phosphate oxygens of A199 via the side chain amino group. On the other side of the major groove of the sARF1 SBS dsRNA, a second pair with Lys260 and Arg261 hydrogen bonds the phosphate oxygens of A84 and G86 via the side chain amino and guanidinium groups, respectively (Figure 2B and C). The interaction of Lys260 is also stabilized by a hydrophobic interaction of its side chain with Tyr224 from the $\beta 1$ strand. These interactions aid to recognize A-form dsRNA and are crucial for dsRNA binding affinity since replacement of basic amino acids in the KKNK-motif by acidic amino acids as found in STAU1 dsRBD5 is incompatible with dsRNA binding (11).

In region 2, interactions of Arg234, Arg235 and Arg236 place the $\beta 1$ - $\beta 2$ loop in the minor groove of the UUCG tetraloop which we used to cap the end of sARF1 SBS dsRNA. Arg234 hydrogen bonds to the phosphodiester backbone via its guanidinium group as well as to a ribose 2' hydroxyl group via its backbone carbonyl oxygen (Fig-

ure 2B and C). The guanidinium group of Arg235 hydrogen bonds to the phosphodiester backbone and Arg236 hydrogen bonds to the O6 carbonyl oxygen of the guanosine of the UUCG tetraloop which adopts a *syn*-conformation. The latter contact is not sequence-specific and could also occur with the O2 carbonyl oxygen of pyrimidines which is consistent with a uracil in this position of the 100 nt loop of ARF1 SBS (23,24). Overall, the $\beta 1$ - $\beta 2$ loop interactions read out minor groove features and shape of an RNA hairpin loop in a non-sequence-specific manner thereby anchoring the STAU1 dsRBD4 at the end of sARF1 SBS dsRNA.

The last canonical interaction point of a dsRBD with dsRNA is region 1 which usually comprises residues 3, 4, 7, 8 and 11 of helix $\alpha 1$ making partially sequence-specific contacts to the minor groove of dsRNA (51). In the case of STAU1 dsRBD4 spacing of residues is maintained but shifted by one amino acid involving Ile207, Ser208, Ala211, Gln212 and Gln215 (Figure 2B). Ile207 makes hydrophobic contacts with the ribose moiety of A85 assisting placement of helix $\alpha 1$ in the minor groove of the sARF1 SBS dsRNA (Figure 2B and C). The side chain hydroxyl of Ser208 forms a hydrogen bond with the N3 of A84 which is considered non-sequence-specific since it could also form with the N3 of guanosine or the O2 carbonyl oxygen of pyrimidines (Figure 2B and C). The methyl groups of Ala211 and also Ala216 are involved in stacking interactions with the ribose moieties of A84 and G206, respectively, further anchoring

Table 2. Structural statistics of Staufen1 dsRBD4 and dsRBD3/4 in complex with sARF1 SBS dsRNA

	dsRBD4–sARF1 SBS dsRNA		dsRBD3/4–sARF1 SBS dsRNA
	dsRBD4	sARF1 SBS dsRNA	dsRBD3/4
NMR distance and dihedral restraints			
Distance restraints			
Total NOE	1359	717	2472
intraresidual	390	358	719
interresidual			
sequential ($ i - j = 1$)	350	258	664
medium range ($1 < i - j < 5$)	272	35	469
long range ($ i - j \geq 5$)	347	66	620
RNA–Protein intermolecular	178		224
RNA–dsRBD3			46
RNA–dsRBD4			178
hydrogen bonds ^a	0	37	0
RNA–Protein hydrogen bonds ^a	13		19
Dihedral restraints ^b	0	168	0
Structure statistics^c			
Mean Violations			
Number of distance constraints $> 0.1\text{--}0.2\text{\AA}$	5.2 ± 1.0		14.8 ± 2.4
Number of distance constraints $> 0.2\text{--}0.3\text{\AA}$	0.0		1.1 ± 0.8
Number of distance constraints $> 0.3\text{--}0.4\text{\AA}$	0.0		0.1 ± 0.3
Number of distance constraints $> 0.4\text{\AA}$	0.0		0.0
Number of dihedral constraints $< 5^\circ$	4.2 ± 0.8		7.1 ± 1.8
Number of dihedral constraints $> 5^\circ$	0.0		0.0
Max. distance constraint violation ($\text{\AA}$)	0.18		0.33
Max. dihedral constraint violation ($^\circ$)	0.43		0.65
Mean Deviation from ideal covalent geometry			
Bond Length ($\text{\AA}$)	0.0109 ± 0.0001		0.0108 ± 0.0001
Bond Angle ($^\circ$)	2.373 ± 0.016		2.369 ± 0.016
Ramachandran plot statistics^{c,d,e}			
Residues in most favored regions (%)	77.7 ± 4.2		71.7 ± 3.5
Residues in additionally allowed regions (%)	21.7 ± 4.2		26.0 ± 3.5
Residues in generously allowed regions (%)	0.2 ± 0.6		1.4 ± 1.3
Residues in disallowed regions (%)	0.3 ± 0.6		0.9 ± 0.9
Average pairwise r.m.s. deviation from mean structure^d			
Protein			
Backbone atoms	0.80 ± 0.13		1.79 ± 0.36
Heavy atoms	1.24 ± 0.11		2.21 ± 0.36
RNA			
Heavy atoms	0.76 ± 0.23		1.23 ± 0.46
RNA–protein complex			
Heavy atoms	0.91 ± 0.15		1.95 ± 0.31

^aHydrogen bond constraints were identified in the course of structure refinement and for imino resonances protected from exchange with H₂O.

^bDihedral angle constraints for sARF1 SBS dsRNA are based on regular A-form geometry and previous structures of the UUCG tetraloop (60).

^cStatistics are computed for the deposited bundle of 20 structures which had lowest violation energy and were selected out of 30 structures with lowest total amber energy from a total of 50 calculated structures. With the exception of maximum constraint violations which give the largest violation among the 20 selected structures, the statistics give the mean and standard deviation.

^dBased on structured residue range as defined: Asn205–Lys 274 (dsRBD4) Lys102–Pro173,Asn205–Lys274 (dsRBD3/4) for the proteins and 1–34 for the sARF1 SBS dsRNA.

^eRamachandran plot, as defined by the program Procheck (46).

helix $\alpha 1$ in the minor groove (Figure 2B and C). This interaction is supported by numerous NOEs between the Ala211 and Ala216 methyl group protons and the riboses of G83, A84 and G206, respectively. The base paired partner of G83, C204 shows canonical interactions with the amide side chain of Gln212 contacting the O2 carbonyl oxygen and the 2' hydroxyl group of the ribose. This type of contact could be maintained by the N3 of a purine and is therefore not considered purely sequence-specific (55). The last position also involves side chain contacts of Gln215 with O2 carbonyl oxygen and the 2' hydroxyl group of U205 (Figure 2B and C). Beside the contacts via helix $\alpha 1$, also one linker residue, namely Asn205, contacts the minor groove side of

the dsRNA by forming a hydrogen bond of the side chain amide to the 2' hydroxyl group of A85 (Figure 2C).

Beside the canonical contacts from regions 1–3 of the STAU1 dsRBD4, we also observed three electrostatic contacts from Lys218, Lys220 and Glu221 of the $\alpha 1$ – $\beta 1$ loop. The side chain amino group of Lys218 hydrogen bonds to the phosphodiester backbone and the side chain carboxyl of Glu221 hydrogen bonds the 2' hydroxyl group of G83. Lys220 also approaches the phosphodiester backbone of A82, suggesting probably a water-mediated interaction (Figure 2C).

Thus STAU1 dsRBD4 interacts with two consecutive pyrimidines on one strand and a purine on the opposite

strand in an A-form dsRNA helix and additional hydrogen bonding and hydrophobic contacts to the minor and major grooves assist these interactions. The conservation of the interacting residues among STAU1 dsRBD4 from different species suggests that this binding mode is conserved (Supplementary Figure S4A).

Structure of STAU1 dsRBD3/4-sARF1 SBS dsRNA complex

The structured domains of the final ensemble of the STAU1 tandem dsRBD3/4-sARF1 SBS dsRNA complex are well-defined and the twenty conformers display a r.m.s. deviation of 1.95 Å for all heavy atoms (Table 2 and Figure 2D). The flexible linker between the dsRBDs could only be partially assigned and is therefore unstructured in the final ensemble of structures (Supplementary Figure S4B). Using the segmentally labeled STAU1 dsRBD3/4 protein with only dsRBD3 and part of the linker being ^{13}C , ^{15}N -labeled allowed us to extract 46 intermolecular NOEs from a 3D ^{13}C F1-filtered F2-edited HSQC-NOESY spectrum and to define the interaction of STAU1 dsRBD3 with sARF1 SBS dsRNA (Supplementary Figure S3C-E). The lower number of intermolecular NOEs for the dsRBD3-sARF1 SBS dsRNA interaction as compared to dsRBD4 is reflected in a lower r.m.s. deviation of 2.68 Å for all protein heavy atoms as compared to 1.37 Å for dsRBD4 (Supplementary Figure S4C). The structure also rationalizes the observed minor differences in the ^1H - ^{15}N HSQC spectra of dsRBD3/4 bound to long ARF1 or sARF1 SBS dsRNA (Figure 1D and Supplementary Figure S1C). Upon shortening of the ARF1 SBS dsRNA signal, differences occur in both helices of dsRBD3 and in the β 1- β 2 loop and β 3 strand of dsRBD4 (Supplementary Figure S5A). All the differences in the amide chemical shifts occur in residues not involved in interaction with the dsRNA and the chemical shift pattern in 3D ^{15}N NOESY-HSQC spectra is preserved (Supplementary Figure S5B and C). Thus, these differences in amide chemical shifts suggest stabilization of secondary structure elements when bound to the sARF1 SBS dsRNA rather than differences in the interaction with the dsRNA. In dsRBD3 which is more dynamic in the free form as compared to dsRBD4, the shift differences might indicate stabilization of the C-terminus of helix α 1 (K114) and the N-terminus of helix α 2 (S156) upon sARF1 SBS dsRNA binding, and in dsRBD4 the differences cluster in one region of the protein indicating a stabilization of the β 1- β 2 loop and adjacent β 3 strand upon binding to the tetraloop of sARF1 SBS dsRNA (Supplementary Figure S5C).

STAU1 dsRBD3 and dsRBD4 bind on the opposite sides of the sARF1 SBS dsRNA and thus do not form a protein-protein interface consistent with the absence of interdomain NOEs in the 3D ^{13}C F1-filtered F2-edited HSQC-NOESY spectrum recorded on the segmentally labeled STAU1 dsRBD3/4-sARF1 SBS dsRNA complex (Figure 2D). As seen for dsRBD4, dsRBD3 also interacts with sARF1 SBS dsRNA in the canonical 3 regions. Helix α 1 is placed in the minor groove through stacking interactions of Ile105, Phe109 and Leu113 with the ribose moieties of U201 and A199 on one strand and C90 on the other strand.

This placement results in an interaction of the side chain carboxyl group of Glu110 with the 2' hydroxyl of the ribose moiety of U88 (Figure 2E). Furthermore, the side chain hydroxyl group of Ser106 interacts with the amino group of G200 which is considered a specific interaction for G-C or C-G Watson-Crick base pairs (Figure 2E) (51). In region 3, Lys157 and Lys158 interact with the phosphodiester backbone of one RNA strand and Lys153 on the opposite strand. The positioning of Lys158 is additionally stabilized by a hydrophobic contact with Ile105 from helix α 1. Finally, in the β 1- β 2 loop, the side chain amino group of Lys133 contacts the ribose O4' or O2' oxygens probably in a water-mediated manner (Figure 2E).

Comparison of the NMR and crystal structures of the STAU1 dsRBD3/4-sARF1 SBS dsRNA complex

The recently published crystal structure of STAU1 dsRBD3/4 bound to the long, 19 bp ARF1 SBS duplex RNA identifies the same set of interacting protein residues with minor deviations but differs both in stoichiometry and arrangement of the dsRBDs along the dsRNA (27). In the crystal structure, long ARF1 SBS dsRNA binds two copies of dsRBD3 on opposite sides of the RNA helix and one copy of dsRBD4 interacting with one copy of dsRBD3 and the dsRNA while the second copy of dsRBD4 is disordered (Supplementary Figure S6A). In our hands, STAU1 dsRBD3/4 did not form a stable 2:1 complex with the long ARF1 SBS dsRNA as shown by our NMR titration experiments which are indicative of stable complex formation only at a protein to RNA ratio of 1:1 (Figure 1D and Supplementary Figure S1A). Similarly, the molecular weight derived from SAXS measurements of STAU1 dsRBD4 and dsRBD3/4-sARF1 SBS dsRNA complexes and overall good fit of solution scattering from atomic models to experimental SAXS data (CRY SOL $\chi^2 = 1.30$ and 1.07, respectively) also supports a 1:1 complex formation in solution (Supplementary Table S1 and Figure S6B-E).

Furthermore, based on our initial NMR experiments, we further shortened the ARF1 SBS dsRNA and showed that complex formation is the same as observed with the longer dsRNA (Supplementary Figure S1B-E), again with a 1:1 stoichiometry. The difference in stoichiometry for the longer ARF1 SBS dsRNA might arise from the fact that our long ARF1 SBS is capped by a stable tetraloop while in the crystal structure a plain dsRNA was used allowing one copy of the two dsRBD3 to interact with a neighboring dsRNA in the crystal lattice *via* the β 1- β 2 loop (27). As a possible consequence, one copy of the dsRBD3 binds in the same region of the ARF1 SBS dsRNA as observed in the NMR structure while the second copy of dsRBD3 partly occupies the dsRBD4 binding side observed in the NMR structure (Supplementary Figure S6A). For one dsRBD3 copy, we detect very similar interactions to the same RNA bases and phosphodiester backbone oxygens as observed in the crystal structure, only the guanine-specific interaction of the β 1- β 2 loop is lost due to our shortened sARF1 SBS dsRNA construct and some interactions of the N-terminal amino acids (K102, S103) are not detected either. For dsRBD4, we obtained 178 intermolecular NOEs and we detect more

contacts than in the crystal structure where one copy of dsRBD4 binds in the opposite direction as compared to the NMR structure and displays high B-factors while the other one is completely disordered. In region 3 of dsRBD4, we observe one additional interaction of an arginine with the phosphodiester backbone and in region 2 we observe contacts from 3 arginines with the phosphodiester backbone and riboses which are not present in the disordered $\beta 1$ – $\beta 2$ loop of the crystal structure. The helix $\alpha 1$ interactions of dsRBD4 occur through the same amino acid side chains (Q212 and Q215) but with other bases (C204 and U205 instead of C90 in the crystal structure) and we observe a minor groove base contact between A84 and Ser208 which is absent in the crystal structure. Moreover, additional contacts with riboses and the phosphodiester backbone from two lysines and a glutamate in the adjacent $\alpha 1$ – $\beta 1$ loop are observed. In summary, one dsRBD3 copy shows the same binding characteristics both in the solution and crystal structure while in solution, dsRBD4 binds in the opposite direction along the dsRNA and partly occupies the position of the second dsRBD3 copy observed in the crystal. Regardless of this difference in positioning and orientation of the dsRBDs, the same amino acid side chains interact with the same type of bases thereby preserving shape and sequence read-out from the minor groove.

Minor groove contacts of STAU1 dsRBD3/4 play little role for binding affinity *in vitro* but are important for target selection *in vivo*

To test the importance of the minor groove base contacts, we mutated key interacting residues of dsRBD3 and dsRBD4 to alanine and tested the effect of these substitutions on sARF1 SBS dsRNA binding affinity by FA measurements. In dsRBD4, mutation of Gln212 and Gln215 to alanine showed only a 1.6-fold reduction in binding affinity, while combination of these mutations with the third minor groove interacting amino acid, Ser208 showed a 5.2-fold drop in affinity (Table 1 and Supplementary Figure S7A). Likewise, mutation of the guanine-interacting Ser106 in dsRBD3 together with the neighboring hydrophobic sugar contact of Ile105 displayed a 3.6-fold drop in affinity (Table 1 and Supplementary Figure S7B). Interestingly, another hydrophobic sugar contact (Phe109) showed a similarly strong 2.4-fold reduction while alanine mutation of the 2' hydroxyl (Glu110) and a hydrophobic sugar contact (Leu113) in helix $\alpha 1$ of dsRBD3 did not affect the affinity (Table 1 and Supplementary Figure S7B).

We next examined the effect of these mutations in the context of the STAU1 dsRBD3/4 protein construct. Alanine mutations in dsRBD3 helix $\alpha 1$ (Ile105/Ser106–1.6-fold, Glu110/Leu113–1.0-fold, and Ile105/Ser106/Glu110/Leu113–1.2-fold) of STAU1 dsRBD3/4 showed only minor or no loss in affinity (Table 1 and Supplementary Figure S7C). Similarly, alanine mutations in dsRBD4 helix $\alpha 1$ (Gln212/Gln215–0.9-fold) or in combination with dsRBD3 helix $\alpha 1$ mutations (Glu110/Leu113/Gln212/Gln215–1.0-fold) also showed wild-type (WT) binding affinities confirming that these residues are not major determinants of binding affinity as already seen with the isolated dsRBDs

(see above, Table 1 and Supplementary Figure S7C). On the other hand, alanine mutations that had larger effects in the isolated dsRBDs also showed consistently stronger loss in the STAU1 dsRBD3/4 construct (Ser208–1.6-fold, Ser208/Gln212/Gln215–4.9-fold, and Ile105/Ser106/Ser208/Gln212/Gln215–2.5-fold) showing that these minor groove interactions are important for dsRNA target binding. In addition, we also tested alanine mutants in $\beta 1$ – $\beta 2$ loop in STAU1 dsRBD3/4 (Lys133–1.3-fold, Arg234/Arg236–1.5-fold) which showed only minor loss in affinity. The alanine mutant of Phe109 which showed a 2.4-fold loss in affinity in the isolated dsRBD3 also showed a 1.7-fold loss in affinity in the STAU1 dsRBD3/4, while the alanine mutant of Arg209 which is not in contact with the dsRNA in showed no loss in affinity. Combination of the latter two mutations in STAU1 dsRBD3/4 showed a 1.6-fold loss in affinity confirming that only Phe109 but not Arg209 contribute to binding affinity (Table 1 and Supplementary Figure S7D). Overall, the binding experiments suggest that minor groove interactions are not the major determinants for binding affinity *in vitro* and that the interaction between STAU1 dsRBD3/4 and sARF1 SBS dsRNA is mainly driven by electrostatic interactions along the phosphodiester backbone as also described for the crystal structure of the complex (27).

To test the relevance of minor groove interactions on STAU1 *in vivo* function, we prepared stable HEK293T-REx and T-REx-HeLa cell lines with inducible expression of FLAG-tagged WT and mutant versions of STAU1, respectively. We aimed to evaluate the effect of these mutation on mRNA binding and steady state levels of ARF1 and X-box binding protein 1 (XBP1), two *bona-fide* targets of SMD(20,23). To assess mRNA binding, we performed RNA-immunoprecipitations (RIP) of STAU1 using the FLAG antibody and quantified the co-precipitated RNAs by RT-qPCR with gene-specific primers directed to either the stem-loop region in the 3'UTR or the coding sequence (Supplementary Figure S8A). Interestingly, the dsRBD4 mutant in Q212A/Q215A that displayed only minor effect on the RNA binding *in vitro* showed decreased RNA binding *in vivo* as evaluated based on the level of co-precipitated RNAs (Figure 3A, Supplementary Figure S8B). The overexpression of this dsRBD4 mutant (Q212A/Q215A) was accompanied by a significant, 2-fold upregulation of XBP1 mRNA and insignificant increase in ARF1 mRNA in HeLa cells (Figure 3B). The overexpression of STAU1 containing mutations in both dsRBD3 and 4 (E110A/L113A/Q212A/Q215A) lead to significant (3- to 4-fold) accumulation of both mRNAs in HeLa cells (Figure 3B) and a similar trend was observed in HEK293T cells (Supplementary Figure S8C). Similar results were obtained when we normalized mRNA levels to their corresponding pre-mRNAs in both cell types upon expression of WT and mutant FLAG-tagged forms of STAU1 (Supplementary Figure S8D). This further supported the conclusion that the dsRBD mutations affect STAU1 function in SMD.

Accordingly, the dsRBD3 and dsRBD4 mutations led to significantly decreased binding to both mRNAs *in vivo* (Figure 3A and Supplementary Figure S8B). In summary,

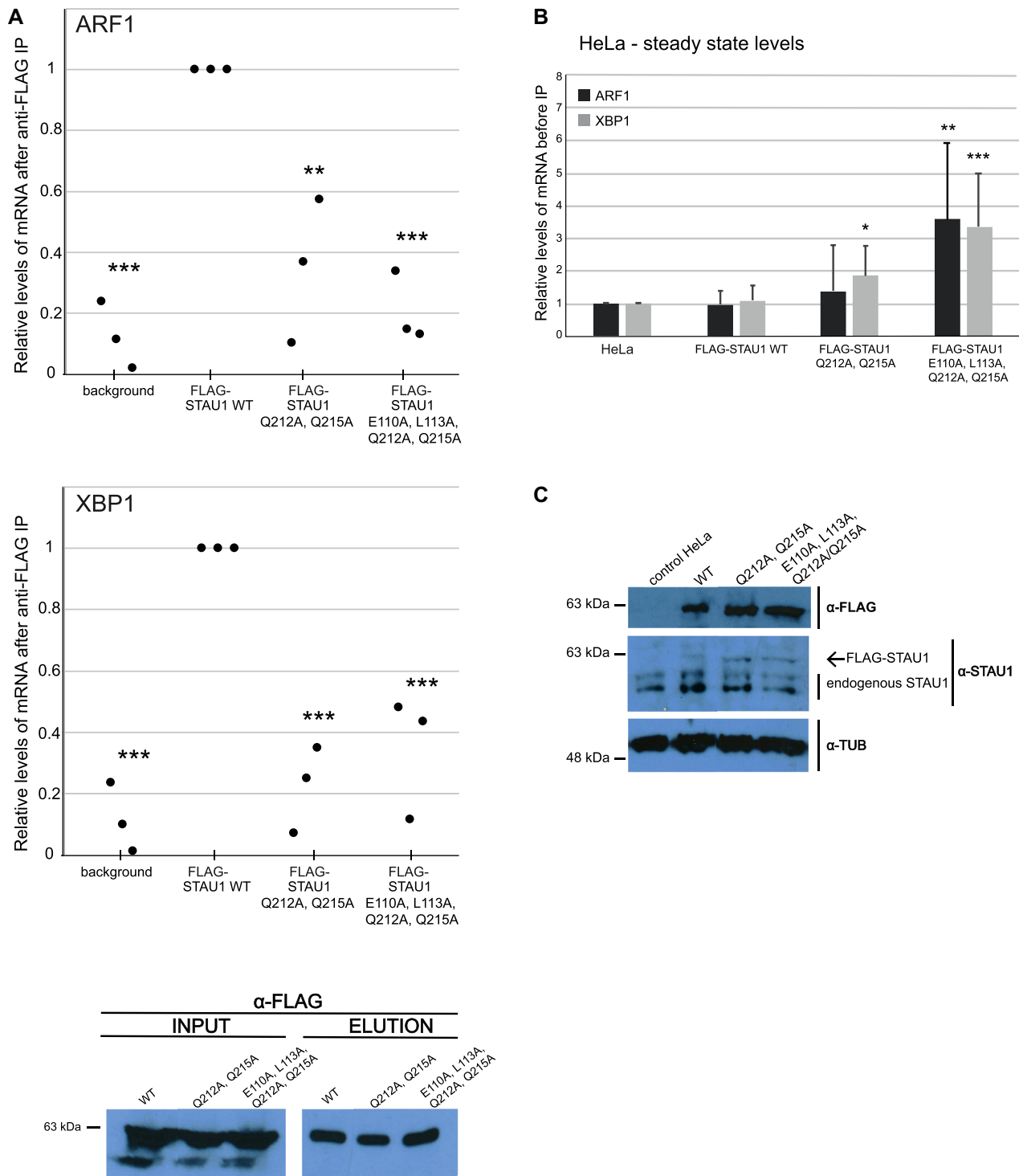


Figure 3. Mutations in dsRBD3 and dsRBD4 impair binding to target mRNAs *in vivo*. (A) RT-qPCR analysis of the amount of ARF1 and XBP1 mRNAs coprecipitated with WT and mutant STAU1 from HEK293T-REx cells. The Y-axis represents the enrichment of mRNAs coprecipitated with STAU1 variants relative to the level in the whole cell lysate (input). Input ARF1 and XBP1 mRNA levels were normalized to GAPDH as an internal control. The qPCR results were analyzed by the $\Delta\Delta C_t$ method. Scatter plot represents relative enrichments of precipitated mRNAs with FLAG-tagged STAU1 mutants relative to the FLAG-WT STAU1. Background is the level of mRNAs unspecifically bound to FLAG-beads from control HEK293T-REx where no FLAG-tagged protein was expressed. Error bars SD ($n = 3-4$ biological replicates), * P -value < 0.1 , ** P -value < 0.01 , *** P -value < 0.001 ; P -values were calculated by two-tailed paired t -test. The western blot below the graphs shows the efficiency of FLAG immunoprecipitation. Input is the whole cell lysate before IP. (B) RT-qPCR analysis of steady-state levels of ARF1 and XBP1 mRNA in T-REx-HeLa cells upon overexpression of stably integrated STAU1 variants. Data were analyzed by $\Delta\Delta C_t$ calculation method and normalized to GAPDH as an internal control. Bar plot represents fold enrichments relatively to control T-REx-HeLa cells. Error bars SD ($n = 3$ biological replicates), * P -value < 0.1 ; P -values were calculated by two-tailed paired t -test. (C) Western blot analysis of STAU1 protein levels upon doxycycline induction using antibodies as indicated on the right. Tubulin was used as a loading control. Control HeLa is a cell line without any integration and expression of a FLAG-tagged protein.

combined mutations in minor groove binding residues in STAU1 dsRBD3/4 show decreased dsRNA binding and mRNA accumulation when expressed in the context of the full-length STAU1 *in vivo*. Thus minor groove contacts are the determinants for target selection *in vivo*.

DISCUSSION

Our solution structure of STAU1 dsRBD3/4 in complex with sARF1 SBS dsRNA reveals how STAU1 selects target mRNAs for SMD. Both dsRBDs show the canonical binding and while dsRBD4 in isolation binds sARF1 SBS dsRNA in one preferred orientation, dsRBD3 shows promiscuous binding in different regions of the dsRNA (Supplementary Figure S3). Only in the context of the tandem dsRBD3/4, dsRBD3 selects a single binding site on the dsRNA. To extract this structural information, we used a segmentally labeled STAU1 protein where only dsRBD3 is isotope labeled and dsRBD4 is not and thus invisible in heteronuclear NMR experiments. With this approach, we could perform isotope-filtered NMR experiments which allowed us to extract two types of information: first, NOE-distance information required for definition of the dsRBD3 binding site on the sARF1 SBS dsRNA on the opposite side of dsRBD4 and secondly that the two dsRBD modules are not involved in protein–protein contacts as seen in the crystal structure (27).

In isolation, dsRBD3 binds with 2-fold higher affinity to sARF1 SBS dsRNA as compared to isolated dsRBD4. This might suggest that the dsRBD3 module binds A-form dsRNA first in a promiscuous manner scanning the A-form dsRNA and this aids in recruiting dsRBD4 for dsRNA binding. The dsRBD4 module then selects a defined binding site on the dsRNA and as a consequence, dsRBD3 also binds in a defined location on the dsRNA. In addition, β 1– β 2 loop interaction of both dsRBDs with opposite ends of the dsRNA might further assist final binding site selection on the dsRNA.

Binding affinity of dsRBDs is mainly driven by electrostatic interactions with the phosphodiester backbone and 2' hydroxyl groups, interactions which select for dsRNA over dsDNA and thus for regular A-form geometry (51). As such, mutations of charged, basic residues in regions 2 and 3 of STAU1 dsRBD3/4 lead to a 4- to 8-fold reduction in binding affinity as compared to WT protein (27). While these interactions constitute the basis for binding affinity for dsRNA, only very few minor groove base contacts in the right register length are providing sequence specificity and thus determine target selection (51). As a result, altering minor groove readout by mutating key protein residues often has only very little effect on affinity (Supplementary Figure S7). This contrasts protein interactions with ssRNA which are often characterized by an intricate network of hydrogen bonds and Van der Waals interactions and thus very sensitive to protein mutations affecting sequence readout.

We observed differential *in vitro* RNA binding affinities of STAU1 dsRBD3/4 mutations when tested in individual dsRBDs or in the context of tandem dsRBD3/4 (Table 1). Interestingly, these minor groove contacting mutations showed much stronger effect in the full-length

protein *in vivo*. The combined mutant in dsRBD3/4 (E110A/L113A/Q212A/Q215A) appeared to significantly affect SMD of ARF1 and XBP1 mRNAs. The differences between the *in vitro* and *in vivo* dsRNA binding can be a consequence of presenting the mutant STAU1 in the context of the full-length protein and testing endogenous full-length mRNA targets. In HeLa cells, the level of expression of FLAG-tagged STAU1 forms was comparable to the level of endogenous STAU1 (Figure 3C). Yet, we observed more than 3-fold upregulation of both mRNA targets in cells expressing mutations in dsRBD3 and 4 (Figure 3B). It is possible that the introduced STAU1 mutations affected yet additional STAU1 activities *in vivo*, such as effects on STAU1 localization or its role in mRNA export or translation.

STAU1 recognizes many different dsRNA targets ranging from mRNA transport and localization elements (1), 5'UTR signals to regulate translation (56) to SMD signals in 3'UTRs to regulate mRNA levels (23) and expansion segments in 18S and 28S rRNA (20). Beside ARF1 SBS, the vast number of STAU1 dsRNA targets, especially transport and localization elements, has not been characterized in more detail (15–20). Many target RNAs form long double-stranded helices, such as duplexes formed by long non-coding RNAs and 3'UTRs which also activate SMD (25) and also transport signals are often complex such as the one of *Drosophila* bicoid mRNA which dimerizes to provide an extended interaction surface for *Drosophila* dm-STAU binding (2,57). The recognition of such large signals by STAU1 might not only require minor groove interactions by dsRBD3/4 but also *in vivo* dimerization (11) and other domains, such as dsRBD2. STAU2, for instance, contains an additional dsRBD1 beside dsRBD2/3/4 and both tandem dsRBD1/2 and dsRBD3/4 modules bind dsRNA with similar affinities and kinetics suggesting cooperative binding of all four domains for stable complex formation and target dsRNA selection (58).

The question remains how STAU1 and STAU2 proteins exhibit their diverse functions in post-transcriptional regulation of gene expression when only a few minor groove interactions determine dsRNA target selection. It is feasible that depending on the complexity of the dsRNA signal, the overall shape of STAU1 or STAU2 complexes with target RNAs could differ. This would then expose different STAU1 or STAU2 protein surfaces for interaction with other proteins mediating RNA transport, translational control, or SMD and thus determine the cellular function of this interaction. Whether this additional layer of complexity confers the diverse functions of STAU1 and STAU2 proteins in post-transcriptional regulation of gene expression needs to be explored with future studies of different full-length STAU1 or STAU2–dsRNA complexes functioning in diverse gene expression pathways.

DATA AVAILABILITY

The coordinates and chemical shifts of the structural ensembles of STAU1 dsRBD4–sARF1 SBS dsRNA and the STAU1 dsRBD3/4–sARF1 SBS dsRNA complexes have been deposited in the OneDep Protein Data Bank under PDB accession number 6SDY and 6SDW, and BMRB accession number 34422 and 34421, respectively. SAXS

datasets, fits and experimental details have been deposited in Small Angle Scattering Biological Data Bank (www.sasbdb.org) (59) as entry SASDF29 (STAU1 dsRBD4-sARF1 SBS dsRNA), SASDF39 (STAU1 dsRBD3/4-sARF1 SBS dsRNA), and SASDF49 (STAU1 dsRBD3/4-long ARF1 SBS dsRNA).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Prof. M. Sattler for providing plasmid of WT and mutant Sortase A, Dr J. Ule for providing the Flag-hSTAU1 plasmid, helpful staff from all CEITEC core facilities and previous as well as current lab members for their useful discussion and technical help during this study.

Authors contributions: P.J.L. and D.K.Y. designed the project. D.K.Y. expressed and purified proteins and performed *in vitro* RNA transcription and purification. D.K.Y. optimized the STAU1-ARF1 complexes and prepared all samples for structural studies. M.Z., M.K., C.B. and D.K.Y. prepared samples for FA measurements. D.K.Y. performed all NMR experiments with help of P.J.L. D.K.Y. and P.J.L. analyzed NMR data and performed structure calculations. M.Z., M.K., C.B. and D.K.Y. performed FA measurements and analysis. T.K., performed and analyzed SAXS measurements. D.Z. performed all the experiments in mammalian cells, constructs, stable cell lines, mRNA quantifications by RT-qPCR, and western blotting. D.Z. and S.V. analyzed the data. P.J.L. wrote the manuscript, D.Z. and S.V. wrote method and results part on the *in vivo* experiments in mammalian cells. All authors discussed the results and approved the manuscript.

FUNDING

EU Seventh Framework Programme under the ‘Capacities’ specific programme [286154-SYLICA]; INBIOR [CZ.1.07/2.3.00/20.0042]; CEITEC 2020 (LQ1601) project with financial contribution made by the Ministry of Education, Youths and Sports of the Czech Republic (MEYS CR) within special support paid from the National Programme for Sustainability II funds; CIISB Research Infrastructure project LM2015043 funded by MEYS CR; Diamond Light Source; B21 Beamline Scientists; Marie Curie Action-Career Integration Grant [PCIG14-GA-2013-630758]; EMBO Installation Grant [3014]; Czech Science Foundation Grant [P305/18/08153S]; Human Frontier Science Program Grant [RGP0024/2008]; Czech Science Foundation Grant [16-21341S to S.V., D.Z.]. Funding for open access charge: Grantová Agentura České Republiky [P305/18/08153S].

Conflict of interest statement. None declared.

REFERENCES

- Heraud-Farlow, J.E. and Kiebler, M.A. (2014) The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. *Trends Neurosci.*, **37**, 470–479.
- St Johnston, D., Beuchle, D. and Nusslein-Volhard, C. (1991) Staufen, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell*, **66**, 51–63.
- Ephrussi, A., Dickinson, L.K. and Lehmann, R. (1991) Oskar organizes the germ plasm and directs localization of the posterior determinant nanos. *Cell*, **66**, 37–50.
- Ferrandon, D., Elphick, L., Nusslein-Volhard, C. and St Johnston, D. (1994) Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell*, **79**, 1221–1232.
- St Johnston, D. (1995) The intracellular localization of messenger RNAs. *Cell*, **81**, 161–170.
- Wickham, L., Duchaine, T., Luo, M., Nabi, I.R. and DesGroseillers, L. (1999) Mammalian staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum. *Mol. Cell Biol.*, **19**, 2220–2230.
- Monshausen, M., Putz, U., Rehbein, M., Schweizer, M., DesGroseillers, L., Kuhl, D., Richter, D. and Kindler, S. (2001) Two rat brain staufen isoforms differentially bind RNA. *J. Neurochem.*, **76**, 155–165.
- Duchaine, T.F., Hemraj, I., Furic, L., Deitinghoff, A., Kiebler, M.A. and DesGroseillers, L. (2002) Staufen2 isoforms localize to the somatodendritic domain of neurons and interact with different organelles. *J. Cell Sci.*, **115**, 3285–3295.
- Ramos, A., Grunert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D. and Varani, G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.*, **19**, 997–1009.
- Martel, C., Dugre-Brisson, S., Boulay, K., Breton, B., Lapointe, G., Armando, S., Trepanier, V., Duchaine, T., Bouvier, M. and DesGroseillers, L. (2010) Multimerization of Staufen1 in live cells. *RNA*, **16**, 585–597.
- Gleghorn, M.L., Gong, C., Kielkopf, C.L. and Maquat, L.E. (2013) Staufen1 dimerizes through a conserved motif and a degenerate dsRNA-binding domain to promote mRNA decay. *Nat. Struct. Mol. Biol.*, **20**, 515–524.
- Luo, M., Duchaine, T.F. and DesGroseillers, L. (2002) Molecular mapping of the determinants involved in human Staufen-ribosome association. *Biochem. J.*, **365**, 817–824.
- Ravel-Chapuis, A., Belanger, G., Yadava, R.S., Mahadevan, M.S., DesGroseillers, L., Cote, J. and Jasmin, B.J. (2012) The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. *J. Cell Biol.*, **196**, 699–712.
- Elbarbary, R.A., Li, W., Tian, B. and Maquat, L.E. (2013) STAU1 binding 3'UTR IRAlus complements nuclear retention to protect cells from PKR-mediated translational shutdown. *Genes Dev.*, **27**, 1495–1510.
- Furic, L., Maher-Laporte, M. and DesGroseillers, L. (2008) A genome-wide approach identifies distinct but overlapping subsets of cellular mRNAs associated with Staufen1- and Staufen2-containing ribonucleoprotein complexes. *RNA*, **14**, 324–335.
- Heraud-Farlow, J.E., Sharangdhar, T., Li, X., Pfeifer, P., Tauber, S., Orozco, D., Hormann, A., Thomas, S., Bakosova, A., Farlow, A.R. et al. (2013) Staufen2 regulates neuronal target RNAs. *Cell Rep.*, **5**, 1511–1518.
- Laver, J.D., Li, X., Ancevicus, K., Westwood, J.T., Smibert, C.A., Morris, Q.D. and Lipshitz, H.D. (2013) Genome-wide analysis of Staufen-associated mRNAs identifies secondary structures that confer target specificity. *Nucleic Acids Res.*, **41**, 9438–9460.
- de Lucas, S., Oliveros, J.C., Chagoyen, M. and Ortin, J. (2014) Functional signature for the recognition of specific target mRNAs by human Staufen1 protein. *Nucleic Acids Res.*, **42**, 4516–4526.
- Ricci, E.P., Kucukural, A., Cenik, C., Mercier, B.C., Singh, G., Heyer, E.E., Ashar-Patel, A., Peng, L. and Moore, M.J. (2014) Staufen1 senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol.*, **21**, 26–35.
- Sugimoto, Y., Vigilante, A., Darbo, E., Zirra, A., Militti, C., D'Ambrogio, A., Luscombe, N.M. and Ule, J. (2015) hiCLIP reveals the *in vivo* atlas of mRNA secondary structures recognized by Staufen 1. *Nature*, **519**, 491–494.
- Popp, M.W. and Maquat, L.E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.*, **47**, 139–165.

22. Park, E. and Maquat, L.E. (2013) Staufen-mediated mRNA decay. *Wiley Interdiscip. Rev. RNA*, **4**, 423–435.
23. Kim, Y.K., Furic, L., Desgroseillers, L. and Maquat, L.E. (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell*, **120**, 195–208.
24. Kim, Y.K., Furic, L., Parisien, M., Major, F., DesGroseillers, L. and Maquat, L.E. (2007) Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J.*, **26**, 2670–2681.
25. Gong, C. and Maquat, L.E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3'UTRs via Alu elements. *Nature*, **470**, 284–288.
26. Gong, C., Tang, Y. and Maquat, L.E. (2013) mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat. Struct. Mol. Biol.*, **20**, 1214–1220.
27. Lazzaretti, D., Bandholz-Cajamarca, L., Emmerich, C., Schaaf, K., Basquin, C., Irion, U. and Bono, F. (2018) The crystal structure of Staufen1 in complex with a physiological RNA sheds light on substrate selectivity. *Life Sci. Alliance*, **1**, e201800187.
28. Elantak, L., Wagner, S., Herrmannova, A., Karaskova, M., Rutkai, E., Lukavsky, P.J. and Valasek, L. (2010) The indispensable N-terminal half of eIF3j/HCR1 cooperates with its structurally conserved binding partner eIF3b/PRT1-RRM and with eIF1A in stringent AUG selection. *J. Mol. Biol.*, **396**, 1097–1116.
29. Liu, H.T. and Naismith, J.H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.*, **8**, 91.
30. Chen, I., Dorr, B.M. and Liu, D.R. (2011) A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 11399–11404.
31. Freiburger, L., Sonntag, M., Hennig, J., Li, J., Zou, P. and Sattler, M. (2015) Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J. Biomol. NMR*, **63**, 1–8.
32. Lukavsky, P.J. and Puglisi, J.D. (2004) Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA*, **10**, 889–893.
33. Easton, L.E., Shibata, Y. and Lukavsky, P.J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA*, **16**, 647–653.
34. Zlobina, M., Sedo, O., Chou, M.Y., Slepankova, L. and Lukavsky, P.J. (2016) Efficient large-scale preparation and purification of short single-stranded RNA oligonucleotides. *Biotechniques*, **60**, 75–83.
35. Ryder, S.P., Recht, M.I. and Williamson, J.R. (2008) Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol. Biol.*, **488**, 99–115.
36. Sattler, M., Schleucher, J. and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Mag. Res. Spectrosc.*, **34**, 93–158.
37. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
38. Kazimierczuk, K., Zawadzka, A. and Kozminski, W. (2008) Optimization of random time domain sampling in multidimensional NMR. *J. Magn. Reson.*, **192**, 123–130.
39. Stanek, J. and Kozminski, W. (2010) Iterative algorithm of discrete Fourier transform for processing randomly sampled NMR data sets. *J. Biomol. NMR*, **47**, 65–77.
40. Lukavsky, P.J. and Puglisi, J.D. (2001) RNAPack: an integrated NMR approach to RNA structure determination. *Methods*, **25**, 316–332.
41. Peterson, R.D., Theimer, C.A., Wu, H. and Feigon, J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J. Biomol. NMR*, **28**, 59–67.
42. Lee, W., Revington, M.J., Arrowsmith, C. and Kay, L.E. (1994) A pulsed field gradient isotope-filtered 3D ¹³C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Lett.*, **350**, 87–90.
43. Herrmann, T., Guntert, P. and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
44. Herrmann, T., Guntert, P. and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, **319**, 209–227.
45. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr, Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
46. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
47. Pollard, T.D. (2010) A guide to simple and informative binding assays. *Mol. Biol. Cell*, **21**, 4061–4067.
48. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
49. Panjkovich, A. and Svergun, D.I. (2018) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics*, **34**, 1944–1946.
50. Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. et al. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.*, **50**, 1212–1225.
51. Masliah, G., Barraud, P. and Allain, F.H. (2013) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol. Life Sci.*, **70**, 1875–1895.
52. Steff, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B. et al. (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, **143**, 225–237.
53. Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. and Allain, F.H. (2011) Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.*, **58**, 1–61.
54. Yadav, D.K. and Lukavsky, P.J. (2016) NMR solution structure determination of large RNA-protein complexes. *Prog. Nucl. Magn. Reson. Spectrosc.*, **97**, 57–81.
55. Gan, J.H., Tropea, J.E., Austin, B.P., Court, D.L., Waugh, D.S. and Ji, X.H. (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, **124**, 355–366.
56. Dugre-Brisson, S., Elvira, G., Boulay, K., Chatel-Chaix, L., Mouland, A.J. and DesGroseillers, L. (2005) Interaction of Staufen1 with the 5' end of mRNA facilitates translation of these RNAs. *Nucleic Acids Res.*, **33**, 4797–4812.
57. Ferrandon, D., Koch, I., Westhof, E. and Nusslein-Volhard, C. (1997) RNA-RNA interaction is required for the formation of specific bicoid mRNA 3'UTR-STAUEN ribonucleoprotein particles. *EMBO J.*, **16**, 1751–1758.
58. Heber, S., Gaspar, I., Tants, J.N., Gunther, J., Moya, S.M.F., Janowski, R., Ephrussi, A., Sattler, M. and Niessing, D. (2019) Staufen2-mediated RNA recognition and localization requires combinatorial action of multiple domains. *Nat. Commun.*, **10**, 1659–1674.
59. Valentini, E., Kikhney, A.G., Previtali, G., Jeffries, C.M. and Svergun, D.I. (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.*, **43**, D357–D363.
60. Lukavsky, P.J., Kim, I., Otto, G.A. and Puglisi, J.D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.*, **10**, 1033–1038.

