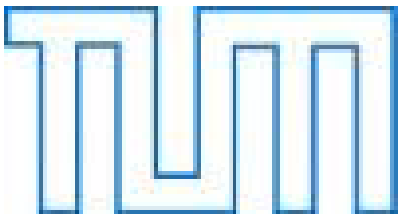


Technische Universität München

Department für Chemie

Bayerisches NMR-Zentrum

Lehrstuhl für biomolekulare NMR-Spektroskopie

Structural and functional characterization of RBM5/6/10 in alternative splicing regulation

Komal Soni

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Franz Hagn
Prüfer der Dissertation:	1. Prof. Dr. Michael Sattler
	2. Prof. Dr. Dierk Niessing

Die Dissertation wurde am 19.06.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 11.07.2017 angenommen.

DECLARATION

I hereby declare that parts of this thesis have been already published in the following scientific journal-

Mourao A*, Bonnal S*, **Soni K***, Warner LR*, Bordonne R, Valcarcel J and Sattler M. "Structural basis for the recognition of spliceosomal SmN/B/B' proteins by the RBM5 OCRE domain in splicing regulation". eLife 2016;5:e14707.

*Equal contribution

Table of Contents

Structural and functional characterization of RBM5/6/10 in alternative splicing regulation

Abstract

Alternative splicing (AS) expands the protein repertoire encoded by the genome whereby the alternatively spliced isoforms can translate into proteins pertaining to distinct, often opposite functions. The regulation of AS is quite complex and involves recognition of *cis*-regulatoy elements in the pre-mRNA by *trans*-acting splicing factors, which act as guides to direct the splicing machinery to the correct splice sites. Aberrant splicing of pre-mRNAs involved in cell proliferation and signaling pathways has frequently been correlated to several diseases in humans, including cancer. The AS regulation of two such pre-mRNA targets: *Fas* and *Caspase-2*, by a family of RNA-binding proteins (RBM5/6/10) has been studied here.

RBM5/6/10 belong to the family of RNA Binding Motif (RBM) proteins. These are multi-domain proteins, where two RNA Recognition Motifs (RRM1,2) and one zinc finger (Zf1) mediate RNA interactions while an OCRE domain is involved in protein-protein interactions. RBM5 is a putative tumor suppressor gene consistent with the frequent deletion of its gene locus in lung cancer. It has been shown to modulate cell proliferation and apoptosis mediated by: death receptor Fas via interactions between OCRE domain and components of the spliceosomal tri-snRNP U4/U6.U5; initiator *Caspase-2* via interactions between RRMs/Zf and U/C rich intronic element of *Caspase-2* pre-mRNA.

So far it is not known how multiple domains in RBM5 that are involved in protein-RNA and protein-protein interactions contribute to its functional activity in splicing regulation. Here, I provide a structural characterization of the RBM5/6/10 OCRE and RBM5 RNA binding domains and a functional analysis of their interactions in AS regulation using an integrated structural biology approach.

Chapter 1 of the thesis introduces the basics of pre-mRNA splicing. It also describes the role and relevance of RBM5 in AS regulation of *Fas* and *Caspase-2* pre-mRNA targets via its different domains. Chapter 2 details the theory of the different structural biology techniques used in this study. Chapter 3 presents the material and methods section providing experimental details.

Chapter 4 provides structural and functional insights into the involvement of RBM5/6/10 OCRE domains in AS regulation of *Fas* pre-mRNA. RBM5 OCRE domain recruits spliceosomal tri-snRNP U4/U6.U5 complex to distal splice sites via a direct interaction with SmN/B/B' C-terminal proline-rich tails. Using a combination of NMR, ITC and CD spectroscopy, I found that RBM5 OCRE domain specifically recognizes a pre-formed poly-proline type II helix comprising of consecutive proline residues, flanked by arginine residues on either side. Additionally, solution NMR structures of RBM10/6 OCRE domains are presented. While RBM10 OCRE domain is structurally conserved in comparison with its RBM5 counterpart, the RBM6 OCRE domain is truncated with only four β-strands instead of six as in the others. Moreover, RBM10 OCRE domain binds to SmN derived PRM with similar affinity as that of RBM5 OCRE while RBM6 OCRE domain is unable to bind. Consistently, *in vivo* splicing assays of *Fas* minigene show that like RBM5, RBM10 OCRE domain is required for formation of anti-apoptotic form of Fas possibly via the same mechanism. On the other hand, RBM6 promotes formation of pro-apoptotic form of Fas in an OCRE independent manner indicating the possibility of involvement of another domain of RBM6.

Chapter 5 and 6 provide details into molecular recognition of *Caspase-2* pre-mRNA by the RNA binding domains of RBM5 using NMR, X-ray crystallography, SAXS and ITC. The crystal structure of RRM1-Zf1 tandem domains describes a novel interaction interface between the two domains. The significance of residues involved in RNA binding is demonstrated by alanine/charge reversal mutations using ITC. I show that Zf1 specifically recognizes a GG motif while RRM1 and RRM2 readily bind to C/U rich RNA motifs. Protein dynamics studied by NMR relaxation experiments show that all the three domains tumble together in solution in both free and RNA bound forms indicating the presence of inter-domain contacts between them. Moreover, SAXS analysis of the free and RNA bound protein shows that the protein adopts a slightly extended conformation upon RNA binding. Additionally, using metal exchange kinetics I show that the presence of an additional cysteine residue adjacent to the metal coordination site in a zinc finger may provide instability to the protein owing to competition between the neighboring cysteines to successfully coordinate the metal ion.

Chapter 7 describes the effects of certain cancer point mutations on the structure and RNA binding of RBM5 domains. A proline mutation (R115P) in RRM1 leads to disruption of the protein fold, while two other point mutations (R140S and R263H) neither affect the structure nor the RNA binding of the individual domains. Another mutation (R263P) known to cause male sterility in mice leads to complete disruption of the secondary structure of the

protein, thereby leading to mice infertility. Chapter 8 provides the discussion of results presented in thesis in comparison to previously published data.

In summary, I demonstrate that RBM5/10 OCRE domains adopt a novel β-sheet fold recognizing proline-rich motifs in the flexible tails of the core spliceosomal SmN/B/B' proteins thereby recruiting tri-snRNP U4/U6.U5 to *Fas* pre-mRNA. The structural basis of this interaction serves as a novel link between a splicing factor and the core splicing machinery. Moreover, the investigations of RNA binding properties of RBM5 suggest distinct possible roles of the multiple domains in AS regulation of different pre-mRNA targets owing to their differential affinity and specificity. The results described in this thesis illustrate how multiple RNA binding domains of RBM5 could cooperate and coordinate with each other for molecular recognition and AS regulation of a variety of pre-mRNA targets via modulation of their dynamic multi-domain arrangement upon RNA binding. This combinatorial control can be essential for expansion of its functional repertoire and adds another level of complexity to the molecular mechanisms underlying alternative splicing and the splicing code.

Strukturelle und funktionale Charakterisierung von RBM5,6 und 10 in der Regulation von alternativem Spleißen

Zusammenfassung

Alternatives Spleißen (AS) erweitert das Proteinrepertoire, das im Genom kodiert ist. Alternativ gespleißte Proteinisoformen können unterschiedliche, oft gegensätzliche Funktionen erfüllen. Die Regulierung des AS ist recht komplex und beinhaltet die Erkennung von Informationen kodierenden *cis*-regulatorischen Elementen durch in *trans* agierende Spleißfaktoren, welche die Spleißmaschinerie zur korrekten Spleißstelle rekrutieren. Abweichendes Spleißen von Proteinen involviert in Zellproliferierung und Signalwege wurde oft mit humanen Krankheiten wie Krebs assoziiert. Die Regulierung von AS zweier solcher mRNAs, *Fas* und *Caspase-2*, durch die Spleißfaktorfamile RBM 5/6/10 wurde in dieser Arbeit untersucht.

RBM 5/6/10 gehören zur Familie der RNA Bindemotif (RBM) Proteine. Es handelt sich hierbei um Multidomänenproteine bei denen zwei RNA Recognition Motif (RRM1,2) Domänen und einem Zinkfinger (ZF1), die Bindung von RNA vermitteln, während die OCRE Domäne in Protein-Protein Interaktionen involviert ist. RBM5 wurde als Tumorsupressorgen beschrieben. Dies ist konsistent mit der Häufigkeit von Deletionen des RBM5 Gens in Lungenkrebszellen. Es wurde gezeigt, dass RBM5 die Zellproliferierung und Apoptose beeinflusst. Dies benötigt im Fall des AS der *Fas* pre-mRNA die Interaktionen der OCRE Domäne mit Komponenten des tri-snRNP U4/U6.U5 des Spleißosoms, andererseits für AS der *Caspase-2* pre-mRNA die Bindung an cis regulatorische RNA Elemente in einem U/C reichen intronischen Element.

Bisher ist nicht bekannt wie die Domänen in RBM5, welche in Protein-RNA- und Protein-Protein- Interaktionen eingebunden sind, zu der Spleißaktivitätregulierung beitragen. In dieser Arbeit wird die OCRE Domäne von RBM5/6/10 und die RNA bindenden Domänen von RBM5 strukturell und funktional auf ihre Interaktionen in der Regulierung von AS hin mit Hilfe strukturbiologischer Methoden untersucht.

Kapitel 1 dieser Arbeit stellt die Grundlagen des Spleißens von pre-mRNA vor. Es beschreibt auch die Rolle und die Wichtigkeit von RBM5 in der Regulierung des AS von *Fas*

und *Caspase-2* pre-mRNA durch seine unterschiedlichen Domänen. Kapitel 2 beschreibt die Theorie der in dieser Arbeit verwendeten strukturbiologischen Methoden. Kapitel 3 enthält die experimentellen Details der für diese Arbeit durchgeführten Experimente.

Kapitel 4 enthält strukturelle und funktionale Einblicke in die Rolle der OCRE Domänen von RBM5/6/10 in der Regulation des AS von *Fas* per-mRNA. Die RBM5 OCRE Domäne rekrutiert den tri-snRNP U4/U6.U5 Komplex des Spleißosoms zur distalen Spleißstelle durch eine direkte Interaktion mit C-terminalen prolinreichen Schwänzen des SmN/B/B'. Durch eine Kombination von NMR, ITC und CD-Spektroskopie habe ich herausgefunden, dass die OCRE Domäne von RBM5 spezifisch eine vorgeformte Polyprolin Typ II Helix erkennt, die aus aufeinanderfolgenden Prolinen flankiert von Argininen besteht. Zusätzlich werden NMR Strukturen der OCRE Domänen von RBM10/6 präsentiert. Während die OCRE Domäne von RBM10 strukturell mit der von RBM5 konserviert ist, ist die OCRE Domäne von RBM6 mit nur 4 von 6 präsenten β-Strängen verkürzt. Darüber hinaus bindet die OCRE Domäne von RBM10 von SMN hergeleitete PRM mit ähnlicher Aktivität wie die OCRE Domäne von RBM5, während die OCRE Domäne von RBM6 diese nicht bindet. Damit konsistent zeigen in vivo Spleißassays mit einem *Fas* Minigen, das RBM5,10, wahrscheinlich nach dem gleichen Mechanismus, die Bildung der anti-apoptotischen Form von FAS befördern. In Kontrast dazu befördert RBM5 die Bildung der pro-apoptotischen Form von *Fas* durch einen OCRE unabhängigen Mechanismus, was auf die Beteiligung einer anderen Domäne in RBM6 hindeutet.

Kapitel 5 und 6 beschreiben mit Hilfe von NMR, Röntgenkristallographie, SAXS und ITC detailliert die molekulare Erkennung der *Caspase-2* pre-mRNA durch die RNA Bindedomänen von RBM5. Die Kristallstruktur der RRM1-ZF1 Tandemdomänen zeigt ein neues Interaktionsinterface zwischen den beiden Domänen auf. Die Wichtigkeit von Aminosäuren, die an der Erkennung er RNA beteiligt sind wird durch Alanin- und Ladungsumkehrmutationen mit Hilfe von ITC gezeigt. Ich zeige, dass ZF1 spezifisch GG Motive erkennt und dass RRM1 und RRM2 C/U reiche RNA Motive binden. Studien der Proteindynamik durch NMR-Relaxationsmessungen zeigen, dass alle drei Domänen sich gemeinsam durch die Lösung bewegen, sowohl in der freien, als auch in der RNA gebunden Form. Dies deutet auf Domänen-Domänen-Kontakte hin. Darüber hinaus zeigt eine SAXS Analyse des freien und RNA gebundenen Proteins, dass das Protein in der RNA-gebundenen Form ein bisschen weniger kompakt ist. Zusätzlich zeige ich mit Hilfe von Metallaustauschkinetiken, dass die Anwesenheit eines zusätzlichen Cysteins benachbart zu der

Metallkoordinationsstelle des Zinkfingers, vielleicht durch kompetitives Binden des Metalls, zu Instabilität des Proteins führt.

Kapitel 7 beschreibt den Effekt von Krebspunktmutationen auf die Struktur und die RNA Bindung der RBM5 Domänen. Eine Prolinmutation (R115P) in RRM1 führt zur Zerstörung der Proteinfaltung. Die anderen Punktmutationen (R140S und R263H) hingegen beinträchtigen weder die Struktur noch die RNA-Bindung der einzelnen Domänen. Eine andere Punktmutation (R263P), welche bekanntermaßen zu Sterilität bei männlichen Mäusen führt, bewirkt eine komplette Entfaltung der Sekundärstruktur des Proteins. Kapitel 8 bietet eine Diskussion der Ergebnisse dieser Arbeit im Vergleich zu bereits veröffentlichen Daten.

Zusammenfassend zeige ich, dass die RBM5/10 OCRE Domänen einen neue β-Faltblattfaltung haben, die prolinreiche Motive in flexiblen Enden des spleißesomalen SmN/B/B' Proteins erkennt und so das tri-snRNP U4/U6.U5 zur *Fas* pre-mRNA rekrutiert. Die strukturelle Basis dieser Interaktion liefert eine der ersten Verbindungen des Spleißfaktors zur Spleißmaschinerie. Darüber hinaus zeigt die Untersuchung der RNA-Bindeeigenschaften von RBM5 die unterschiedlichen möglichen Rollen der Domänen in der Regulierung von AS verschiedener mRNAs dank ihrer unterschiedlichen Affinität und Spezifität. Die Ergebnisse der vorliegenden Arbeit illustrieren auch wie mehrere RNA Bindedomänen von RBM5 durch Modulation ihrer dynamischen Domänenanordnung kooperieren und sich miteinander koordinieren um RNA zu erkennen und AS einer Reihe von pre-mRNAs zu regulieren. Dieses dynamische Verhalten erweitert das funktionelle Repertoire von RNA-bindenden Multidomänenproteinen erheblich und fügt eine weitere Ebene an Komplexität der AS zugrunde liegenden molekularen Mechanismen hinzu.

-

Chapter 1: Introduction I: Biological background

1.1. Splicing and spliceosome assembly

1.1.1. Pre-mRNA splicing

The process of gene expression can be regulated at every level, including transcription, post-transcription and post-translation. This regulation is important for the cell to modify the levels of the specific gene products according to specific requirement. In eukaryotes, a variety of mechanisms involving post-transcriptional gene regulation occur, including splicing, editing and polyadenylation. Here, I only discuss the process of pre-mRNA splicing as it is the main focus of this study. The mechanism of pre-mRNA splicing refers to removal of non-coding introns and ligation of coding exons especially in metazoan organisms. A majority of eukaryotic genes are transcribed into pre-mRNAs that are converted into processed mRNAs via this mechanism.

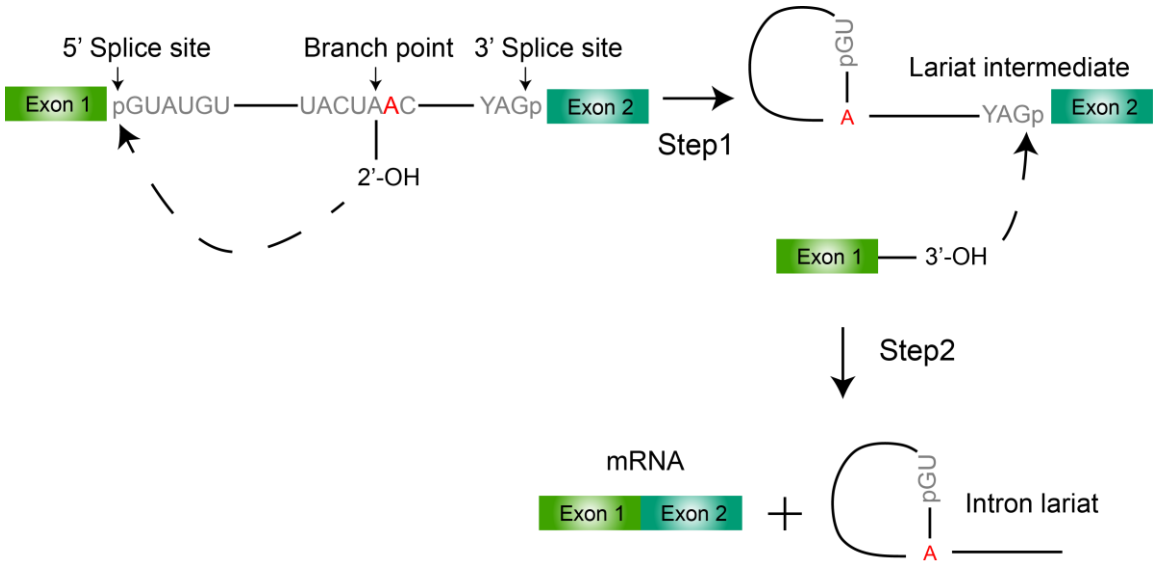


Figure 1 Schematic overview of pre-mRNA splicing reaction

The splicing reaction consisting of two transesterification steps creating a spliced mRNA and intron lariat from the pre-mRNA is shown. (Adapted from (Will and Luhrmann 2011))

The process of pre-mRNA splicing can be easily explained by considering a pre-mRNA consisting of exon1 and exon2 separated by an intron containing the branch point adenosine (**Figure 1**). It consists of two consecutive transesterification reactions whereby first the 2'-OH of the branch point adenosine attacks the 5' splice site (5' ss) leading to displacement of exon1. Next, ligation between 5' ss and branch point leads to formation of a lariat between them. In the second step, the 3'-OH of exon1 carries out a nucleophilic attack on the 3' splice site (3'

ss) of exon2 ligating the two exons and releasing the intron (Moore and Sharp 1993, Black 2003, Wahl, Will et al. 2009).

The process of pre-mRNA splicing is carried out by five different small-nuclear ribonuclear particles (snRNPs) and associated *trans*-acting factors. In these snRNPs seven distinct Sm proteins (SmB/B', SmD1, SmD2, SmD3, SmE, SmF and SmG) assemble as a heptameric ring around a conserved RNA binding motif in the UsnRNAs along with several particle specific proteins. The interactions between UsnRNPs and *trans*-acting factors with the pre-mRNA substrate lead to formation of a highly dynamic spliceosome complex (Lerner, Boyle et al. 1980). This compensates for the minimal information encoded in the splicing substrate, while providing specificity and flexibility. Two different types of spliceosomes exist in eukaryotes: the highly abundant U2-dependent spliceosome, and the minor U12-dependent spliceosome. They particularly differ in the splicing subunits, branch point and splice site sequences amongst other differences (Burge CB 1999). Here I will only discuss the most abundant U2-dependent splicing mechanism.

Spliceosome assembly is an elaborate process powered by ATP hydrolysis and involving sequential assembly and dis-assembly of different complexes on the pre-mRNA (**Figure 2**). First, the U1 snRNP recognizes the 5' ss leading to the formation of E' complex in an ATP independent manner. Due to the relatively weak intensity of this interaction, it is stabilized by other factors, including serine/arginine-rich SR proteins (reviewed in (Long and Caceres 2009, Shepard and Hertel 2009)). At this step, splicing factor SF1 also recognizes the branch point (Berglund, Chua et al. 1997). Next, the U2 auxiliary factor (U2AF) heterodimer is recruited to the polypyrimidine tract (Py tract) and the 3' AG dinucleotide forming the E complex or the commitment complex (Legrain, Seraphin et al. 1988, Nelson and Green 1989, Zamore and Green 1989) at the 3' ss. Subsequently, U2 snRNP replaces SF1 at the branch point by recognition of sequences around the branch point via U2 snRNA in an ATP dependent manner, thereby forming Complex A. Further recruitment of pre-assembled U4/U6.U5 tri-snRNP complex leads to the formation of Complex B where all the splicing components have assembled onto the pre-mRNA. Finally, after significant amount of rearrangements including the formation of intermediate complexes B^{act} and B^{*} involving displacement of U1 and U4 snRNPs, formation of catalytically active spliceosome complex C takes place. Formation of Complex C marks the first chemical reaction of pre-mRNA splicing. Finally, the post-spliceosomal complex is formed after the second chemical reaction takes place. Furthermore,

the spliceosome dissociates from the mRNA and the snRNPs are recycled for additional rounds of pre-mRNA splicing.

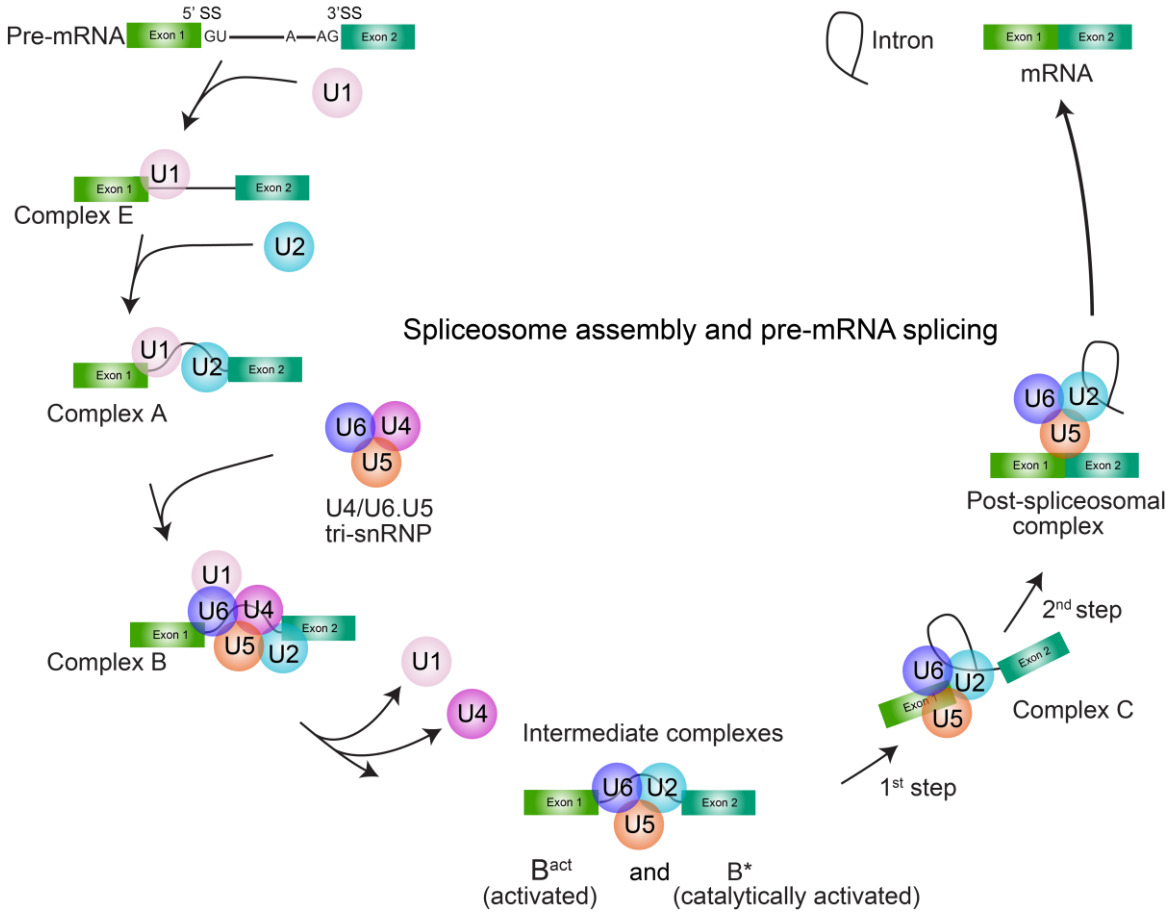


Figure 2 Spliceosome assembly and pre-mRNA splicing

Schematic representation of the process of pre-mRNA spliceosome indicating the detailed steps and intermediate complexes formed (Adapted from (Will and Luhrmann 2011))

1.1.2. Alternative splicing

The process of pre-mRNA splicing can be broadly categorized into constitutive and alternative splicing.. In constitutive splicing, certain exons are always included in the mature mRNA and are referred to as constitutive exons. On the other hand, alternative splicing (AS) is the process by which multiple isoforms of mature mRNA can be produced from a single pre-mRNA by inclusion or exclusion of distinct exons. This process forms the basis of expansion of the protein repertoire encoded by the genome. Interestingly, most eukaryotic genes undergo AS to produce isoforms with distinct and sometimes antagonistic activities.

The process of AS is a tightly regulated and complex process that requires the careful assembly of the spliceosome complexes at the respective splice sites. Since the intronic sequences are generally large, spanning up to hundreds of kilobases in length, they can easily harbor 'decoy' splice sites which might make the process of 'authentic' splice site recognition error prone. Intriguingly, it has been shown that such decoy splice sites marking pseudo-exons are rarely ever spliced (Sun and Chasin 2000), indicating that additional features apart from these core splicing signals must be important in guiding the splicing machinery to the correct positions.

An extensive array of additional signals involved in AS regulation include the *cis*-regulatory RNA elements and the *trans*-acting splicing factors that can either serve as splicing enhancers or repressors. Depending on the location of these *cis*-regulatory elements, they can be classified as exon splicing enhancers (ESE), exon splicing silencers (ESS) or intron splicing enhancers (ISE), intron splicing silencers (ISS). The activities of the *cis*-regulatory elements is context dependent, although they usually function by recruiting the *trans*-acting splicing factors to activate or repress splicing at different stages of the spliceosome assembly at nearby splice sites (Matlin, Clark et al. 2005, Wang and Burge 2008).

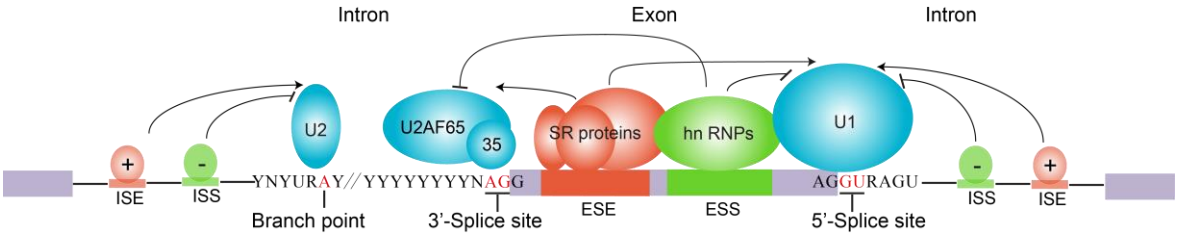


Figure 3 Schematic representation of alternative splicing regulation

The *cis*-acting regulatory elements and *trans*-acting splicing factors involved in alternative splicing regulation are shown. Adapted from (Wang and Cooper 2007, Wang and Burge 2008)

The best characterized ESEs promote splicing by binding to the Serine/Arginine (SR) family of proteins while the best characterized ISSs and ESSs operate by binding to heterogeneous nuclear ribonucleoproteins (hnRNPs) thereby inhibiting splicing (**Figure 3**). Other *trans*-acting splicing factors involved in AS regulation can either be auxiliary factors of the spliceosome or may interact with the core splicing machinery (Zhou, Licklider et al. 2002, Jurica and Moore 2003, Bessonov, Anokhina et al. 2008, Hegele, Kamburov et al. 2012), affecting the splicing decisions in a very diverse set of ways. It becomes therefore, very interesting to study how they regulate the process of alternative splicing.

1.2. Role of RBM5 in alternative splicing regulation

As mentioned previously, the process of pre-mRNA splicing requires the formation of the ribonucleoprotein complex (spliceosome) which needs to be guided to the correct location on the pre-mRNA. The required information as to where splicing should take place is encoded within the pre-mRNA and the 'helper' RNA-binding proteins (RBPs) that are able to recognize and read this information act as guides (Chen and Manley 2009, Nilsen and Graveley 2010). Since many RBPs contain multiple RNA binding domains which may provide differential sequence specificity for RNA recognition, the alterations in their levels and activity serve as the major means of alternative splicing regulation.

In addition, it has been shown that specific point mutations in the *cis*-regulatory elements altering the splice sites thereby affecting splice site selection or in trans-acting factors affecting the protein-RNA recognition, can directly or indirectly lead to disease phenotype (reviewed in (Wang and Cooper 2007, Tazi, Bakkour et al. 2009, Scotti and Swanson 2016). Moreover, the splicing isoforms of proteins generated via the process of AS are frequently reported to have opposite functions. This phenomenon becomes particularly interesting for genes encoding proteins that are involved in cell death pathways whereby alternative splicing gives rise to pro- and anti-apoptotic isoforms of the protein (Schwerk and Schulze-Osthoff 2005). This forms an essential link between cancer and alternative splicing where dysregulation of alternative splicing events can occur, strongly selecting certain variants that would evade cell death.

Here, I will focus on alternative splicing regulation of two genes involved in apoptotic pathways that are regulated by RBM5 protein, among others.

1.2.1. Regulation of *Fas* alternative splicing

Fas protein (also known as CD95) is a widely expressed cell-surface receptor that is member of the tumor necrosis factor and nerve growth factor family of receptors (Nagata and Golstein 1995, Schulze-Osthoff, Ferrari et al. 1998). Interaction of Fas receptor with its ligand (FasL) is critical for shutdown of immune responses (Hughes, Belz et al. 2008, Weant, Michalek et al. 2008) as well as for maintaining immune privileged sites in the body (Krammer 2000, Peter, Budd et al. 2007). Engagement of Fas receptor by Fas ligand on the surface of T-cytotoxic cells can initiate a cascade of reactions mediated via caspase activation leading to

cell death (Bouillet and O'Reilly 2009). Therefore, aberrant splicing of *Fas* pre-mRNA can serve as a potential way for tumor cells to circumvent elimination via the immune system.

The *Fas* pre-mRNA can be alternatively spliced to produce a number of different isoforms. There are eight alternatively spliced variants of *Fas* amongst which two isoforms are of particular importance. Depending upon inclusion or exclusion of exon 6, which encodes the transmembrane domain, either a membrane bound Fas or soluble Fas protein is produced (**Figure 4**). The membrane bound Fas has pro-apoptotic properties as it is able to carry out its normal cell death function via the signaling cascade. Contrastingly, the soluble form of Fas is anti-apoptotic and acts as inhibitor of Fas signaling by binding to the Fas ligand, making it unavailable for binding to membrane bound Fas receptor (Cheng, Zhou et al. 1994, Cascino, Fiucci et al. 1995). Increased concentration of soluble Fas protein is observed in a wide range of tumors.

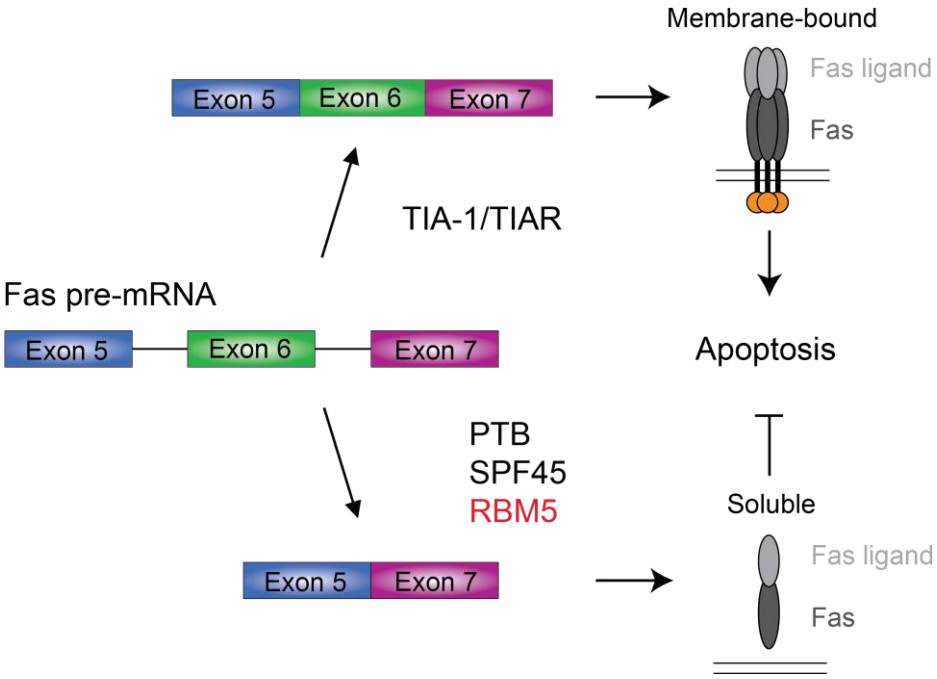


Figure 4 Alternative splicing regulation of *Fas* pre-mRNA

Inclusion or skipping of exon 6 leads to production of either a membrane-bound pro-apoptotic form of Fas or anti-apoptotic soluble form of Fas protein. The proteins that influence alternative splicing of *Fas* are shown.

A number of splicing factors have been implicated in alternative splicing regulation of *Fas* pre-mRNA, as shown in **Figure 4**. T-cell intracellular antigen-1 and TIA-1 related protein (TIA-1/TIAR) bind to U-rich intronic sequences downstream of exon 6, enhance U1snRNP recruitment via direct interactions with U1C protein (Forch, Puig et al. 2002, Izquierdo, Majos

et al. 2005) and promote exon6 inclusion. Contrastingly, a polypyrimidine tract-binding protein (PTB) inhibits inclusion of exon 6 by interfering with U2AF binding upstream of exon6 and via interactions with an exonic silencer (Izquierdo, Majos et al. 2005). Another protein, splicing factor 45 (SPF45) has also been shown to regulate the alternative splicing of *Fas* pre-mRNA (Corsini, Bonnal et al. 2007, Al-Ayoubi, Zheng et al. 2012, Liu, Conaway et al. 2013) by promoting exon 6 exclusion via protein-protein interactions between its UHM domain (U2AF Homology Motif) and ULM (U2AF Ligand Motif) sequences from different splicing factors including SF1, SF3b155 and U2AF65. Finally, it was shown about a decade ago that RBM5 protein is also involved in alternative splicing regulation of *Fas* pre-mRNA where it also required for exon 6 exclusion, thereby promoting the formation of anti-apoptotic form of Fas protein (Bonnal, Martinez et al. 2008).

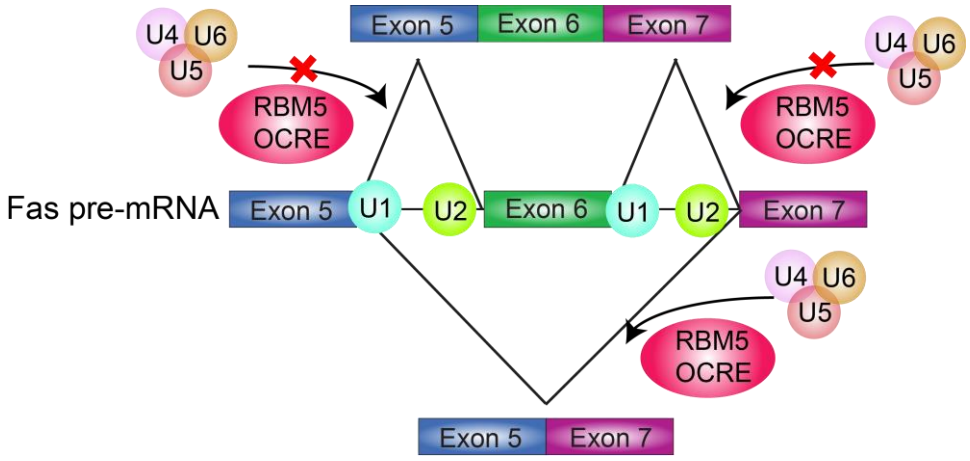


Figure 5 Model depicting role of RBM5 in *Fas* alternative splicing

RBM5 via its OCRE domain inhibits the recruitment of U4/U6.U5tri-snRNP to pre-spliceosomal complexes assembled on introns 5and 6 preventing the formation of mature spliceosomes. It also promotes sequence-dependent distal splice site pairing between 5' splice site of exon 5 and 3' splice site of exon 7. (Adapted from (Bonnal, Martinez et al. 2008))

A detailed mechanism of action of RBM5 to promote exon 6 exclusion in *Fas* mRNA was illustrated by Bonnal *et al.* (**Figure 5**). RBM5 does not affect early splice site recognition processes but influences splice site choice decisions at later steps in the splicing process. After exon 6 definition takes place by recognition of splice sites by U1 and U2snRNP, RBM5 inhibits the splicing of introns 5 and 6 by blocking the recruitment and incorporation of the U4/U6.U5 tri-snRNP complex thereby inhibiting transition of the pre-spliceosomal complexes to mature spliceosomes at these sites. Additionally, it was shown that RBM5 promotes distal splice site pairing between the 5' splice site of exon 5 and 3' splice site of exon 7 in a sequence-specific

manner. Therefore, sequences within exon 6, its flanking sites and distal sites are all required for a complete response from RBM5. This activity of RBM5 was shown to be conferred by one of its domains, known as the OCRE (OCtamer REpeat) domain.

1.2.2. Regulation of *Caspase-2* alternative splicing

Caspase-2 is a highly conserved cysteine-aspartate protease that acts as a tumor suppressor in a variety of cellular processes (Ho, Taylor et al. 2009, Kumar 2009). Alternative splicing of *Caspase-2* pre-mRNA can produce different isoforms depending on the inclusion (*Casp 2S*) or exclusion of exon 9 (*Casp 2L*) as shown in **Figure 6**. The predominant *Casp 2L* isoform produces full-length functional protein with pro-apoptotic properties (Wang, Miura et al. 1994). On the other hand, the *Casp 2S* isoform produces a truncated version of the protein (lacking the enzyme active domain) owing to a frameshift mutation introduced by exon 9 inclusion (Wang, Miura et al. 1994). It is unclear whether the truncated protein simply has anti-apoptotic properties or it is just provided as a substrate for non-sense mediated decay (NMD) thereby reducing *Caspase-2* mRNA levels (David and Manley 2010). Either way, an increase in the *Casp 2S* isoform would be favorable for the cancer cells.

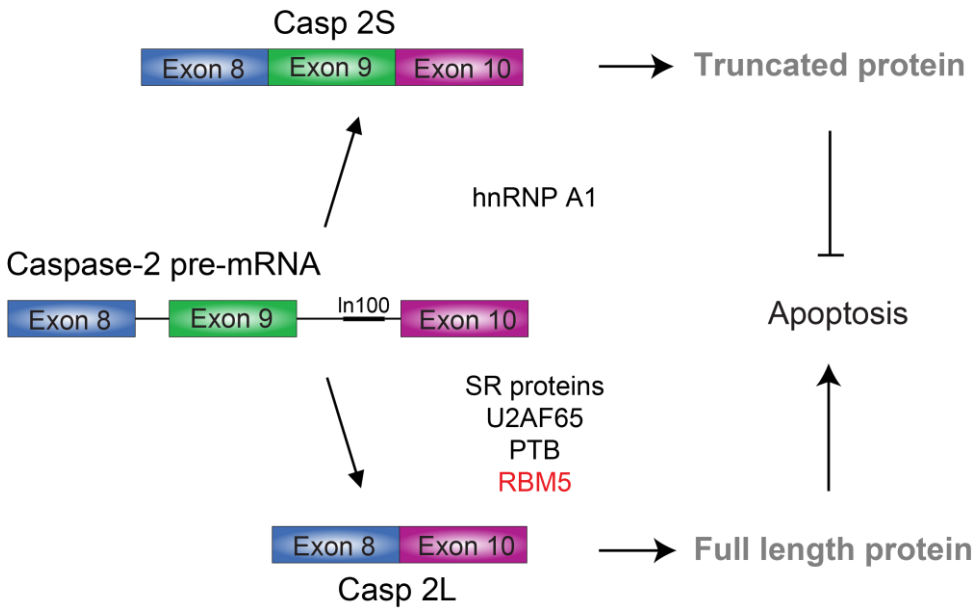


Figure 6 Alternative splicing regulation of *Caspase-2* pre-mRNA

Inclusion or skipping of exon 9 leads to production of either a truncated anti-apoptotic form of Casp2 or a full length protein with pro-apoptotic properties. The proteins that influence alternative splicing of *Caspase-2* pre-mRNA are shown.

Various proteins have been implicated in alternative splicing regulation of *Caspase-2* pre-mRNA (**Figure 6**). It was shown that hnRNP A1 promotes exon 9 inclusion thereby promoting the anti-apoptotic form of the protein, while serine-arginine proteins like SC35/SRSF2 have the opposite effect (Jiang, Zhang et al. 1998). It was later identified that a 100-nucleotide intronic element termed as In100, present in intron 9 inhibits the inclusion of exon 9. A sequence in In100 acts as a decoy 3' splice site forming U2snRNP dependent non-productive spliceosome-like complexes thereby providing a competitive advantage to the exon-skipping splicing event (Cote, Dupuis et al. 2001) (**Figure 7**). It was later shown that the intronic element In100 contains an additional region downstream of the decoy 3' splice site containing several binding sites (U/C-rich repeats) for splicing repressor polypyrimidine tract-binding protein (PTB) (**Figure 7**). Both the regions were shown to be able to repress exon 9 inclusion independently (Cote, Dupuis et al. 2001). Later it was also suggested that In100-like intronic elements might be general splicing repressors of Caspase genes (Havlioglu, Wang et al. 2007).

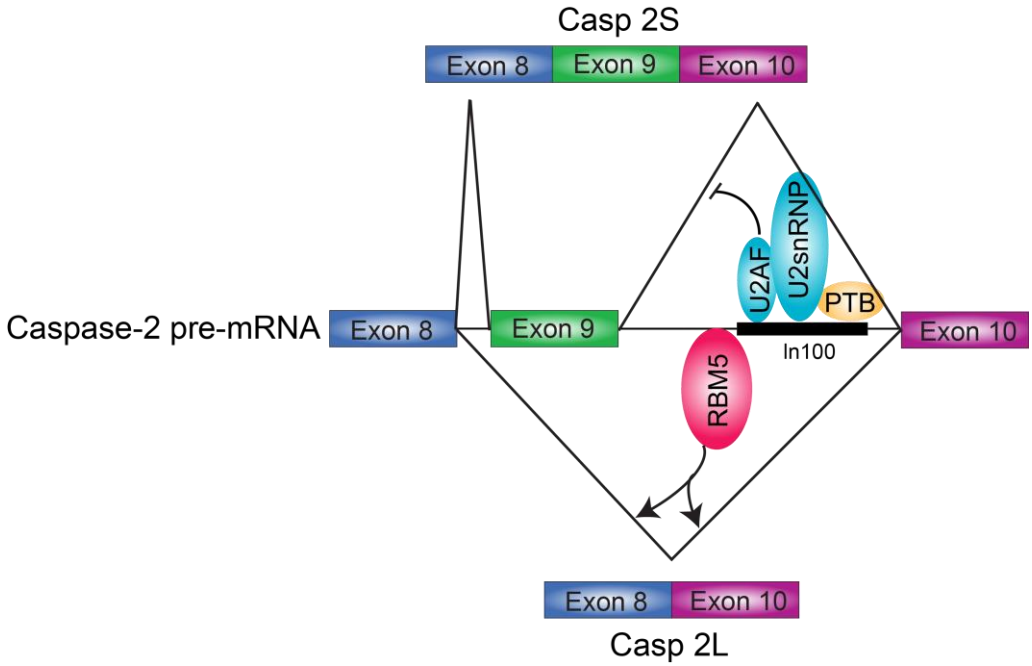


Figure 7 Model depicting role of RBM5 in *Caspase-2* alternative splicing

RBM5 binds to a U/C rich element upstream of In100 in intron 9, promoting splicing between exon 8 and 10 of *Caspase-2* pre-mRNA. Additionally, the intronic element In100 inhibits exon 9 inclusion in a U2snRNP and PTB dependent manner, separately.

Almost a decade ago, it was shown that RBM5 is also involved in alternative splicing of *Caspase-2* pre-mRNA by interacting with U/C-rich elements immediately upstream of the

splicing repressor ln100 (Fushimi, Ray et al. 2008). RBM5 was demonstrated to promote the formation of *Casp 2L* isoform, by activating splice site pairing between 5' splice site of intron 8 and 3' splice site of intron 9 in an ln100-independent manner, thereby promoting apoptosis and acting as a tumor suppressor (**Figure 7**). It was eventually shown that the two RNA recognition motif (RRM) domains of RBM5 confer its *Caspase-2* alternative splicing activity where a mutant lacking the two RRM domains was unable to bind RNA, thereby compromising its effect on *Caspase-2* alternative splicing (Zhang, Zhang et al. 2014).

1.3. RBM5/6/10 family of proteins

RNA-binding proteins (RBPs) are involved in a variety of processes including RNA transport and metabolism, translation, stability and alternative splicing (Glisovic, Bachorik et al. 2008). Generally, these RBPs are multi-domain proteins where individual domains recognize specific targets and therefore carry out a variety of functions. According to a recent study, there are 2,130 known RBPs involved in RNA processing and metabolism in humans (Gerstberger, Hafner et al. 2014). This is not surprising, given the essentially complex nature of the cellular processes RBPs are involved in.

RBM5, RBM6 and RBM10 form a very closely related family of RNA binding motif proteins. The most studied of these RBMs is the putative tumor suppressor protein RBM5. The role of RBM5 came into light with the frequent deletion of a piece of chromosome 3 (3p21.3), encoding the RBM5 gene in heavy smokers, lung cancers and other tissue carcinomas (Angeloni 2007). Loss of heterozygosity at this locus occurs in 95% of small-cell lung cancer (SCLC), as well as in 70% of non-SCLC (Sutherland, Wang et al. 2010). RBM5 has been identified as a molecular signature associated with metastasis, consistent with its down-regulation in a variety of cancers (Edamatsu, Kaziro et al. 2000, Welling, Lasak et al. 2002, Ramaswamy, Ross et al. 2003). In comparison to this, RBM5 is consistently upregulated in breast cancer (Oh, Grosshans et al. 1999, Rintala-Maki, Goard et al. 2007). These observations suggest a strikingly important, albeit complex role of RBM5 in regulating genes important in several cancers. Given the numerous domains in the protein, it can be speculated that the individual domains separately or in conjugation with each other confer complex functionality to the protein making it possible to recognize and regulate a variety of targets.

Additionally, while the effect of RBM5 on *Caspase-2* splicing (promoting the 2L isoform) appears to be in line with its role in promoting apoptosis, the inhibition of *Fas* exon

6 inclusion would protect against Fas mediated apoptosis, which contrasts its role as a putative tumor suppressor. Apart from this, very little is known about its biochemical functions.

RBM6 and RBM10, also have similarity with that of RBM5 in terms of domain organization (**Figure 8**). RBM5 and RBM6 are clustered on the same chromosome, with RBM5 being an immediate telomeric neighbor of RBM6 indicating their origin from gene duplication (Timmer, Terpstra et al. 1999, Lerman and Minna 2000). RBM6 has ~30% sequence similarity with RBM5 protein. Likewise, RBM10 shares ~50% sequence similarity with RBM5 suggesting the proteins to be paralogues, with possibly over-lapping functions (Bonnal, Martinez et al. 2008).

It was recently demonstrated that RBM6 and RBM10 have antagonistic effects on alternative splicing of *NUMB* pre-mRNA which is involved in regulation of NOTCH cell signaling (Bechara, Sebestyen et al. 2013). *NUMB* encodes an inhibitor of NOTCH pathway which is hyper-activated in ~40% of human lung cancers (Dang, Gazdar et al. 2000, Westhoff, Colaluca et al. 2009, Maraver, Fernandez-Marcos et al. 2012), making inhibition of NOTCH pathway a lucrative approach for cancer therapy (Purow 2012). Inclusion or skipping of *NUMB* exon 9 leads to isoforms encoding proteins promoting cell proliferation or cell differentiation, respectively (Verdi, Bashirullah et al. 1999, Toriya, Tokunaga et al. 2006). While RBM10 promotes exon 9 skipping, RBM6 has the opposite effect whereby it promotes exon 9 inclusion and RBM5 has no apparent effect on *NUMB* pre-mRNA splicing (Bechara, Sebestyen et al. 2013). This striking result demonstrates how these very similar RBM proteins have a very diverse set of functions.

Multi-domain RBM5/6/10 proteins each consist of an arginine/serine rich (RS) domain, two RNA recognition motif (RRM) domains, two Zinc finger domains, an OCRE (OCtamer REpeat) domain and a glycine-rich (G-patch) domain at the C-terminus (**Figure 8**). A detailed description of the different RBM domains is presented below.

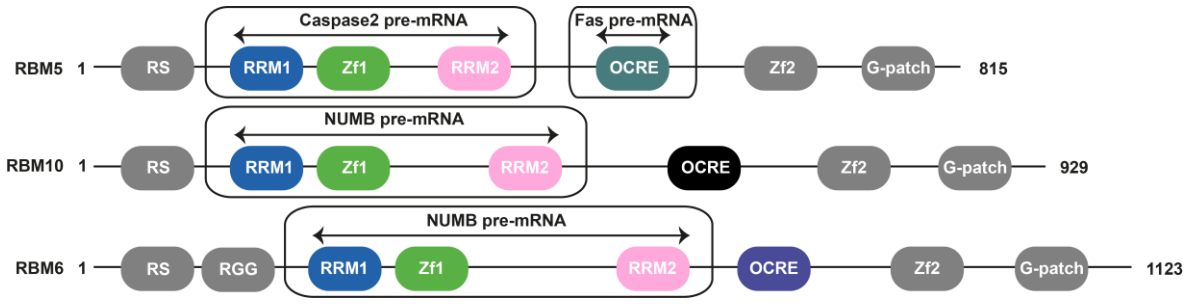


Figure 8 Domain organization of RBM5/6/10

Domain organization of RBM5/6/10 multi-domain proteins is shown. The pre-mRNA targets regulated directly (*Caspase-2/NUMB*) and indirectly (*Fas*) by the different domains are indicated.

The RS domain consists of arginine/serine repeats, which is usually involved in protein-protein interactions with other RS domain containing proteins while being essential for the proper functioning of many of these proteins (Graveley and Maniatis 1998, Philipps, Celotto et al. 2003). Moreover, extensive phosphorylation of RS domains is important for the correct localization of the domain containing proteins (reviewed in (Sanford, Longman et al. 2003)), it can even alter the splicing function of some proteins (Graveley 2000). Additionally, RBM6 also has a low complexity RGG repeat domain containing aromatic residues that are frequently interspersed between the RGG repeats. The RGG domains have multi-functional roles including translational repression, apoptosis, transcription, snRNP biogenesis, and DNA damage signaling, among others (reviewed in (Thandapani, O'Connor et al. 2013)) and often perform their regulatory processes via arginine methylation. They may also be involved directly in RNA binding (Kiledjian and Dreyfuss 1992).

The RRM domain is one of the most abundant type of RNA binding domains found in higher eukaryotes. The RRM domains have a canonical eight-residue motif known as the ribonucleoprotein 1 (RNP1) (Adam, Nakagawa et al. 1986, Sachs, Bond et al. 1986), having the consensus motif [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] and an additional six-residues motif known as ribonucleoprotein 2 (RNP2) (Lahiri and Thomas 1986, Dreyfuss, Swanson et al. 1988), having the consensus motif [ILV]-[FY]-[ILV]-X-N-L, where X can be any amino acid residue. The RRM fold usually consists of βαββαβ topology with the β-sheet interface usually involved in RNA recognition, and two α-helices which pack against the β-sheet interface on both sides (**Figure 9**). Several variants of the RRM domain are known in literature which deviate from the canonical RRM to accommodate for their specific targets.

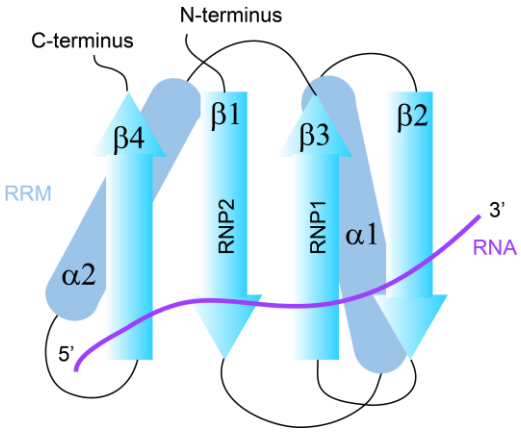


Figure 9 Representative canonical RRM fold

The canonical RRM fold with βαββαβ topology is represented with the RNP1 and RNP2 motifs indicated. The position and orientation of RNA ligand is shown in purple. (Adapted from (Kielkopf, Lucke et al. 2004))

Zinc finger proteins form a diverse family of proteins which are characterized by zinc ion coordination required for stabilization of the protein fold (Laity, Lee et al. 2001).These domains could be involved in protein-protein, protein-RNA or protein-DNA interactions (Laity, Lee et al. 2001) and mediate a range of cellular processes (Krishna, Majumdar et al. 2003). Of the many classes of zinc finger proteins, the most common is the C_2H_2 type having ββα fold where two cysteine residues on the β-strands and two histidine residues on the α-helix coordinate the Zn^{2+} ion (Lee, Gippert et al. 1989). They also form the largest cluster of transcription factors in most species. The second zinc finger domain in RBM5/6/10 (Zf2) is the C_2H_2 type.

Another class of Zinc finger domains is the RanBP2-type zinc finger. As the name suggests, these type of zinc fingers were identified initially in a conserved member of the Ras superfamily, Ran binding protein 2 (RanBP2) which plays a role in nuclear protein import (reviewed in (Steggerda and Paschal 2002)). The multiple zinc fingers in the protein were shown to mediate binding to RanGDP (Yaseen and Blobel 1999). Apart from this, ~30 other proteins contain these domains where they carry out distinct functions (Nguyen, Mansfield et al. 2011). A subset of the RanBP2-type zinc fingers are present in proteins implicated in transcription regulation and RNA processing. They contain the consensus motif sequence W-X-C-X_{2-4}-C-X_3-N-X_6-C-X_2-C, with a single Zn^{2+} ion being coordinated by the four cysteine residues. The RanBP2-type zinc finger fold contains two short β-hairpins sandwiching a conserved tryptophan residue and the Zn^{2+} ion (Hall 2005).

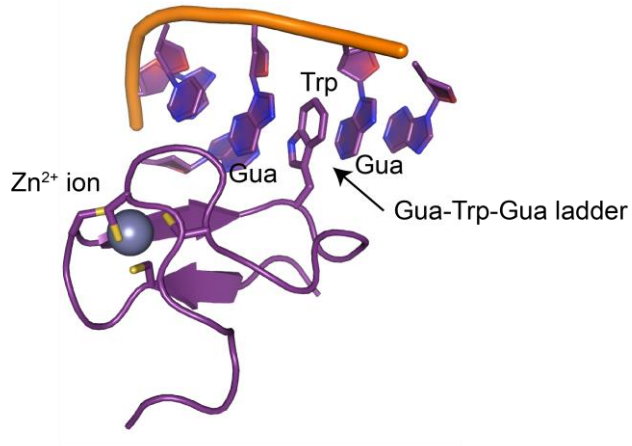


Figure 10 Representative structure of RanBP2-type zinc finger in complex with RNA

Structure of RanBP2-type Zinc finger domain (ZRANB2-F2) in complex with RNA (PDB ID: 3G9Y) is shown. The Gua-Trp-Gua ladder formed between the protein and RNA is indicated. The canonical fold of RanBP2-type zinc fingers where four cysteine residues coordinate the central Zn^{2+} ion is also shown.

Nguyen *et al.* determined the crystal structure of a RanBP2-type zinc finger (ZRANB2-F2) in complex with an RNA sequence (AGGUAA) (Nguyen, Mansfield et al. 2011) which was determined by SELEX (Loughlin, Mansfield et al. 2009) experiments (**Figure 10**). It revealed that amongst other interactions, a surface exposed tryptophan side-chain formed a unique Gua-Trp-Gua ladder, making the sequence specific recognition of Guanine bases possible. Moreover, using single base mutations at 2^{nd} and 4^{th} positions of AGGUAA RNA sequence, it was shown that RBM5 Zf1 displays a 1.5-, 2- and 8- fold preference for guanine at position 4 over uracil, adenine and cytosine, repectively.

About a decade ago, the OCRE (OCtamer REpeat) domain, characterized by an imperfectly repeated octameric sequence was identified in angiogenic factor (VG5Q) and RBM5/6/10 family of RBPs (Callebaut and Mornon 2005). The OCRE domain was shown to be involved in protein-protein interactions of RBM5 with tri-snRNP proteins, mainly U5 snRNP 200 and 220kDa, with SRp20, SmN/B/B', Acinus and U175K (Bonnal, Martinez et al. 2008). Deletion of the OCRE domain (but not the other domains) lead to disruption of RBM5 function in alternative splicing regulation of *Fas* pre-mRNA, thus highlighting the importance of OCRE domain.

The G-patch domain that is characterized by the presence of Gly-rich repeats was predicted to be involved in nucleic acid binding, given its occurrence in a number of RBPs (Aravind and Koonin 1999). It has a consensus sequence of six highly conserved glycine

residues. Recently, it was shown that RBM5 interacts with a protein found in the spliceosome, DEAH-box RNA helicase protein (DHX15) (Fouraux, Kolkman et al. 2002), in a G-patch dependent manner (Niu, Jin et al. 2012).

1.3.1. Structural and functional information available for RBM5 protein

In line with the aim of thesis, it is necessary to understand the wealth of structural and functional information already available on RBM5 protein. As explained above, RBM5 is involved in alternative splicing regulation of *Fas* and *Caspase-2* pre mRNA. While protein-protein interactions mediated via its OCRE domain are necessary in case of *Fas*, protein-RNA interactions via its RNA binding domains are believed to be important in *Caspase-2* alternative splicing regulation. In particular, Zhang and co-workers showed that the RNA binding functionality of RRM1 and RRM2 domains is required for an effect on *Caspase-2* splicing by RBM5 (Zhang, Zhang et al. 2014).

Fushimi *et al.* showed that either the deletion of a 30-nucleotide region upstream of ln100 in intron 9 (5'-CUCUUUUCCUAAGAACUU**GGCUCUUCUCU**-3') or changing U/C-rich sequence to G/A (5'-CUCUUUUCCUAAGAACUU**GGCUCUUCUCU**-3'to 5'-CUCUUUUCCUAAGAACUU**CCGAGAAGAGA**-3') reduced the *Casp 2L/2S* ratio by ~3 fold (Fushimi, Ray et al. 2008), thereby narrowing down the region of interaction for RBM5.

In 2012, *Song et al.* solved the solution NMR structure of the second RRM domain of RBM5 (Song, Wu et al. 2012). It displays the canonical βαββαβ RRM fold with the β-sheet interface involved in RNA binding (**Figure 11A**). Using two RNA sequences: 5'-CUCUUC-3' and 5'-GAGAAG-3', they showed that RBM5 RRM2 domain can preferentially bind to both CU and GA rich sequences with ~57 μM affinity indicating its flexibility in RNA recognition.

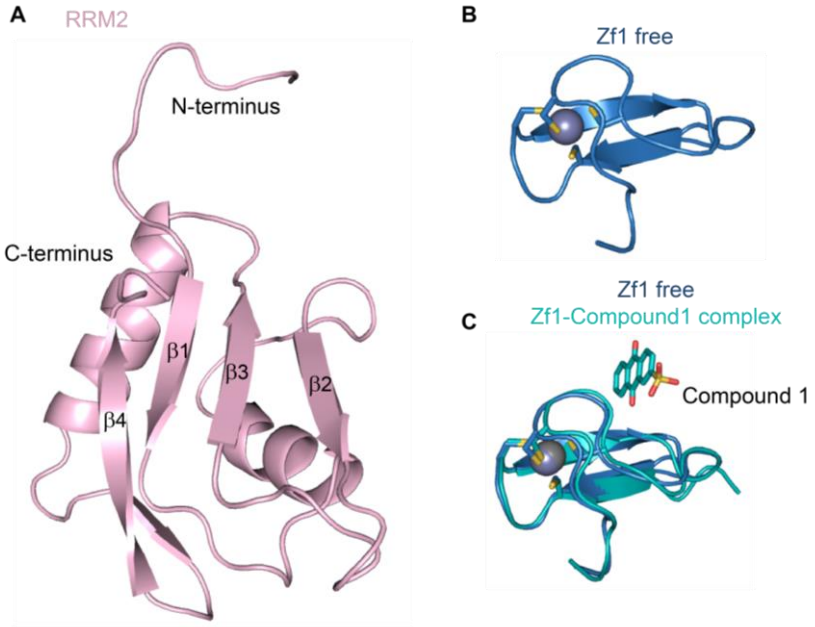


Figure 11 Structural information available for RBM5 protein

(A) Solution NMR structure of RBM5 RRM2 domain (PDB ID: 2LKZ) with the β-sheet interface is shown (B) Solution NMR structure of RBM5 Zf1 domain (PDB ID: 2LK0) in its free form is shown (C) Superposition of solution NMR structures of RBM5 Zf1 domain in its free and compound bound form (PDB ID: 2LK1) is shown in blue and cyan, respectively.

There is also a wealth of information available on RanBP2-type zinc fingers. It was shown that the RanBP2-type zinc fingers preferentially bind to the single stranded RNA sequence 5'-AGGUAA -3' (Nguyen, Mansfield et al. 2011). They also showed that RBM5 Zf1 domain prefers to bind to a guanine at the 4th position with an affinity of ~250-270 nM using fluorescence anisotropy titrations. The solution NMR structure of RBM5 Zf1 domain was further determined by Farina and co-workers (Farina, Fattorusso et al. 2011), using 2D-NMR experiments on the unlabeled protein (**Figure 11B**). The Zf1 domain displays the canonical RanBP2-type zinc finger structure where the Zn^{2+} ion is coordinated by four cysteine residues, stabilizing the protein fold. Due to the given complex nature of RBM5 functionalities, it is tempting to modulate its activity using small molecules. Therefore, Farina *et al.* performed an NMR based fragment-library screen for RBM5 Zf1, which lead to successful identification of a small molecule with binding affinity of ~82 μM. The small molecule-Zf1 complex structure revealed that the compound occupies the RNA binding pocket on Zf1 thereby inhibiting the protein-RNA interaction (**Figure 11C**). A superposition of the free and compound-bound Zf1 structures illustrates that the overall fold of the protein remains unchanged, although a few

changes in the backbone and side-chain conformation of interface residues are observed (**Figure 11C**).

Even though solution NMR structures of Zf1 and RRM2 are available, there is no structural information available for the protein-RNA complex. Moreover, there is also no data available on how the multiple RNA binding domains of RBM5, connected in tandem, might be involved in RNA recognition. The main aims of this thesis are: (1) To understand the structural basis of RBM5/6/10 OCRE –SmN/B/B' interactions in alternative splicing regulation of *Fas* pre-mRNA and (2) To investigate how multiple RNA binding domains in RBM5 cooperate with each other to specifically recognize RNA in alternative splicing regulation of *Caspase-2* pre-mRNA.

Chapter 2: Introduction II: Techniques

Structural biology refers to the structural studies of biological molecules such as proteins and nucleic acids leading to a holistic understanding of complex biological processes. X-ray crystallography has been used as the principal method for determining high resolution structures of biological molecules, but recently the realization of importance of dynamics at every level of study has changed this scenario. For example, the presence of long flexible linkers connecting various domains in a protein reiterates the significance of study of dynamic processes in solution. Therefore, NMR plays a very important and integral part of structural biology. Apart from this, small angle X-ray scattering (SAXS) can additionally be used to obtain low-resolution information on biological complexes in solution, which can also be instrumental in understanding biological processes where high resolution methods fail. Therefore, I adopt an integrated structural biology approach where a combination of methods including NMR, X-ray crystallography and SAXS are used to study protein-RNA complexes.

In this thesis, two basic questions have been addressed using the integrated structural biology approach: how RBM5/6/10 OCRE domains recognize SmN/B/B' polyproline-rich tails to recruit tri-snRNP to *Fas* pre-mRNA; how multiple RNA binding domains in RBM5 cooperate with each other to recognize *Caspase-2* pre-mRNA target for alternative splicing regulation. For this purpose, the techniques used are described in detail.

2.1. NMR spectroscopy

Nuclear Magnetic Resonance spectroscopy is widely accepted as one of the principal techniques in structural biology. It is a powerful method, not only to study high resolution 3D structures of biomolecules but also to probe their dynamic properties in solution. Although the technique is limited by the size of the biomolecules under study, recent advances in hardware and experimental design are pushing this size limit to allow study of much larger biomolecules. In particular, selective labeling schemes and protein deuteration significantly reduce spectral overlap thereby offering the possibility to study huge biomolecular systems. In this work, solution NMR spectroscopy has been used primarily to characterize protein-protein and protein-RNA interactions.

2.1.1. Principles of NMR spectroscopy

The NMR phenomenon is based on the intrinsic property of atomic nuclei to have an overall spin, which is determined by the sum of number of neutrons and protons. The spin can thus take values of zero, fraction or integer and is characterized by a nuclear spin quantum

number I. In the presence of a magnetic field, nuclei with non-zero spin have a magnetic moment associated with them, μ, which is proportional to the spin (Eq. 1)

$$\mu = \gamma I = \gamma \hbar m = \gamma L \qquad \text{Eq. 1}$$

where γ is gyromagnetic ratio, $\hbar$ is Planck's constant, m is magnetic quantum number and L is angular momentum.

The magnetic quantum number can take integer values between -I and I. So, ^{1}H which has spin I = 1/2, upon interaction with an external magnetic field B_0 can exist in two states with energy levels as -1/2 and +1/2 depending upon their alignment with the external magnetic field. This alignment of the nucleus cannot be precisely parallel or anti-parallel, but it precesses at an angle to the magnetic field with an angular velocity known as Larmor frequency (ν_0) or Larmor angular frequency (ω_0) (Eq. 2).

$$\omega_0 = -\gamma B_0 \text{ or } \nu_0 = -\gamma B_0/2\pi \qquad \text{Eq. 2}$$

At equilibrium, the population of spins in the lower and higher energy levels is slightly different and can be described by the Boltzmann distribution (Eq. 3).

$$\frac{N_{upper}}{N_{lower}} = e^{-\Delta E/kT} \qquad \text{Eq. 3}$$

where N_{upper} and N_{lower} represent the number of spins in the upper and lower energy levels, ΔE is the energy difference between the two levels, T is the temperature and k is the Boltzmann constant.

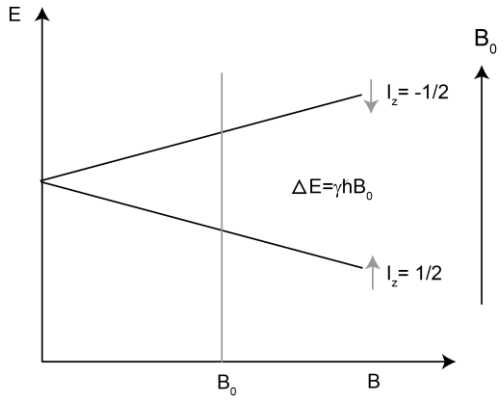


Figure 12 Energy levels of spin half nuclei

Splitting of energy levels of spin half nuclei in the presence of external magnetic field B_0.

In the presence of external magnetic field, this population difference leads to a build-up of net magnetic field in the direction of the external magnetic field (B_0) and gives rise to a macroscopically observable bulk magnetization (M) (**Figure 12**). If a nucleus is irradiated with electromagnetic waves at its Larmor frequency, it will absorb energy and be promoted to the excited, less stable energy level leading to the phenomenon of resonance which is measured in an NMR experiment. The difference in the energy levels of the two spin states of a particular nucleus is dependent upon the gyromagnetic ratio and the external magnetic field (Eq. 4).

$$\Delta E = \hbar\omega_0 = \hbar\gamma B_0 \qquad \text{Eq. 4}$$

Consequently, better sensitivity would be expected for a nucleus with a higher gyromagnetic ratio at a higher field.

2.1.2. Vector formalism

The nuclei precessing at Larmor frequency, in an externally applied magnetic field will also experience a torque which can be expressed as a time derivative of angular momentum as well as a function of the external magnetic field (Eq. 5).

$$T = \frac{\delta L}{\delta t} = \frac{1}{\gamma}\frac{\delta \mu}{\delta t} \text{ and } T = \mu B \qquad \text{Eq. 5}$$

Therefore, using Eq. 5, and representing bulk magnetization (M) as a summation of all nuclear dipoles, the behavior of magnetic moments of spins as a function of time can be described by the Bloch equation-

$$\frac{\delta M(t)}{\delta t} = M(t) * \gamma B(t) \qquad \text{Eq. 6}$$

where *M* represents bulk magnetization vector, γ is the gyromagnetic ratio and B represents the external magnetic field.

In simple terms, in the presence of external magnetic field B_0, which conventionally defines the z-axis of the coordinate system, the bulk magnetization vector (M) also points towards the z-axis. Upon irradiation with a short RF pulse along x-axis, following the right hand rule of electromagnetism, the bulk magnetization will now point towards –y axis, while the angle of rotation (θ) will depend on the length of the RF pulse. The magnetisation will then start precessing in the xy plane, at an angular frequency (ω) generating signal in NMR detection coil. Eventually, the NMR signal will decay due to relaxation effects (transverse relaxation T_2 and longitudinal relaxation T_1).

2.1.3. Product operator formalism

The vector model can only be used to describe basic NMR experiments taking into account only isolated spins. On the other hand, product operator formalism can be used to describe complicated experiments for coupled spin systems. The product operator formalism can be used to describe the states of spin system in density matrix representation. It provides a complete description of complex NMR experiments in quantum mechanical terms where all operators have a clear physical meaning.

An orbiting spin possesses angular momentum, which is a vector quantity pointing in the direction perpendicular to the plane of rotation. The components of this spin angular momentum can be described as operators I_x, I_y, I_z along x, y and z-axis, respectively and the entire spin system can be described by density operator σ(t). Therefore, at any given time point, the state of a single spin-half can be described by the density operator such that it is the sum of different amounts of the operators as x, y and z components (Eq. 7).

$$\sigma(t) = a(t)I_x + b(t)I_y + c(t)I_z \qquad \text{Eq. 7}$$

The values of the operators will vary with time during pulses and delays. At equilibrium, due to the presence of only z-magnetization, the density operator can be equated to the spin angular momentum along z-axis ($\sigma_{eq}=I_z$). During NMR experiments, I_z sequentially transforms and product operators evolve during this time.

The product operator proves to be quite useful as here spin state (eg. I_x along x-axis) and rotational operation (eg. I_x rotation along x-axis) both take the same form. The precession and RF pulses can easily be explained by group theory, where the application of a product operator (rotation) to another product operator (spin state) leads to a changed state.

For example,

$$I_z \xrightarrow{90^\circ I_x} -I_y \qquad I_z \xrightarrow{90^\circ I_y} I_x \qquad I_z \xrightarrow{90^\circ I_z} I_z \qquad \text{Eq. 8}$$

This rotation along different axes can be calculated for any degree of rotation and can be easily compared with the expected results from vector model.

The energy difference between two spin states is not just due to the external magnetic field but also due to local magnetic field experienced by the nuclei. These local magnetic fields shield the individual nuclei from the external magnetic field differently, depending on their chemical environment. Therefore, each nucleus will resonate at a different frequency owing to

its chemical environment and lead to a distinct signal in the NMR spectrum. This is called chemical shift (δ) represented in ppm (parts per million) and is described as-

$$\delta = \frac{\nu_{signal} - \nu_{ref}}{\nu_{ref}} * 10^6 \quad \text{Eq. 9}$$

The chemical shift evolves with an offset (Ω), which is the difference between the signal and reference during the time of precession (t) as follows-

$$I_x \xrightarrow{\Omega t I_z} I_x \cos \Omega t + I_y sin\Omega t \quad \text{Eq. 10}$$

$$I_y \xrightarrow{\Omega t I_z} I_y \cos \Omega t - I_x sin\Omega t$$

$$I_z \xrightarrow{\Omega t I_z} I_z$$

The product operator approach can deal with coupled spin-systems. Since three operators are needed to define each spin, in case of two coupled spin systems, I_{1x}, I_{1y}, I_{1z} define spin 1 and I_{2x}, I_{2y}, I_{2z} define spin 2. Due to J-coupling between the spin systems I_1 and I_2, the states of the spin systems will mix and the result is the product of the two operators ($2I_1I_2$). The operators for two spins evolve under offsets and pulses the same way as those for a single spin. The rotations have to be applied separately for each spin where rotations of one spin do not affect the other. For example, the evolution of I_{1x} under the offset of spin 1 and spin 2 can be represented as below, where spin 2 operators do not have any effect on spin 1 operators-

$$I_{1x} \xrightarrow{\Omega_1 t I_{1z} + \Omega_2 t I_{2z}} I_{1x} \xrightarrow{\Omega_1 t I_{1z}} \xrightarrow{\Omega_2 t I_{2z}} I_{1x} \cos \Omega_1 t + I_{1y} sin\Omega_1 t \quad \text{Eq. 11}$$

2.1.4. NMR experiments for protein sequence assignment

The very first experiment that one should record is a 1D-^{1}H NMR experiment to check for the quality of the protein. This basic experiment presents a fingerprint of all the protons present in the protein. If the protein is well folded, there should be well dispersed signals from the methyl protons present around 0 ppm. Another indication is the extent of spread of the signals in the 1D spectrum. In a folded protein, the chemical shift dispersion is much more than that in an unstructured protein where for example, the proton signals of the backbone amides will be clustered between 6-8 ppm as opposed to 6-10 ppm for a well folded protein. One thing to consider here is that if the protein is majorly alpha helical, then also the chemical shift dispersion could be minimal like that in case of unstructured protein.

With increasing size of the protein, the complexity of the 1D-^{1}H NMR spectrum increases due to extensive signal overlap. Therefore, if the protein looks well folded, one can proceed with 2D-NMR experiments. If the correlation is measured between same nuclei, it is known as homonuclear 2D experiment while if measured between different nuclei, it is known as heteronuclear 2D experiment (for example, ^{1}H-^{15}N or ^{1}H-^{13}C). A 2D ^{1}H-^{15}N HSQC (Heteronuclear Single Quantum Coherence) is a simple experiment to measure the correlation between ^{1}H and ^{15}N nuclei of each amide bond. It represents the fingerprint of the protein as each backbone amide is represented as a single peak in the spectrum (except for proline residues) and is usually the first heteronuclear experiment that is recorded. Additionally, correlations between tryptophan side chain N_{ε}-H_{ε} and asparagine/glutamine side chain N_{δ}-$H_{\delta 2}$/ N_{ε}-$H_{\varepsilon 2}$ are also visible. The arginine N_{ε}-H_{ε} cross-peaks are also visible but as folded signals as the chemical shift of N_{ε} falls outside of the region usually recorded. At low pH, arginine N_{η}-H_{η} and lysine N_{ζ}-H_{ζ} are also visible, but again as folded signals.

For large proteins, the application of 2D-^{1}H-^{15}N HSQC becomes limited due to spectral crowding especially in the central region, and due to transverse relaxation (T_2) effects leading to broad linewidths that decrease the quality of the NMR spectrum. The T_2 relaxation rates for high molecular weight proteins are high, which leads to rapid decay of the NMR signal. Replacing protons (major source of T_2 relaxation) with deuterons can help achieve better signal to noise (Gardner and Kay 1998). About two decades ago, 2D ^{1}H,^{15}N-TROSY experiment (Transverse Relaxation Optimized Spectroscopy) was introduced (Pervushin, Riek et al. 1997, Salzmann, Pervushin et al. 1998). It correlates the same nuclei as ^{1}H,^{15}N-HSQC but decreases relaxation effects to attain better linewidths, spectral resolution and sensitivity by selecting the coherence component where cancellation of relaxation due to dipolar coupling and chemical shift anisotropy occurs. It therefore extends the protein size limitation which could be studied by NMR (Fernandez and Wider 2003).

Next, to assign the correlations observed in the ^{1}H-^{15}N HSQC spectrum, sequential protein assignment is done using triple resonance experiments whereby the backbone resonances of the protein are assigned sequentially. These include HNCA, HNCACB and CBCA(CO)NH (Shan, Gardner et al. 1996, Sattler, Schleucher et al. 1999) triple resonance experiments. The HNCA experiment is the most sensitive followed by CBCA(CO)NH and lastly HNCACB. In these experiments, correlation between backbone amide with C_{α} and C_{β} resonances is achieved. In the HNCACB for every amide visible in the ^{1}H-^{15}N HSQC, two

sets of C_α and C_β resonances are observed, one belonging to the current residue (i) while the other to the previous residue (i-1). Therefore, a sequential walk of the protein sequence is possible by connecting and matching the peak positions of residue i and i-1 between the different experiments (for example, HNCACB and CBCA(CO)NH) via backbone amides as shown in **Figure 13**.

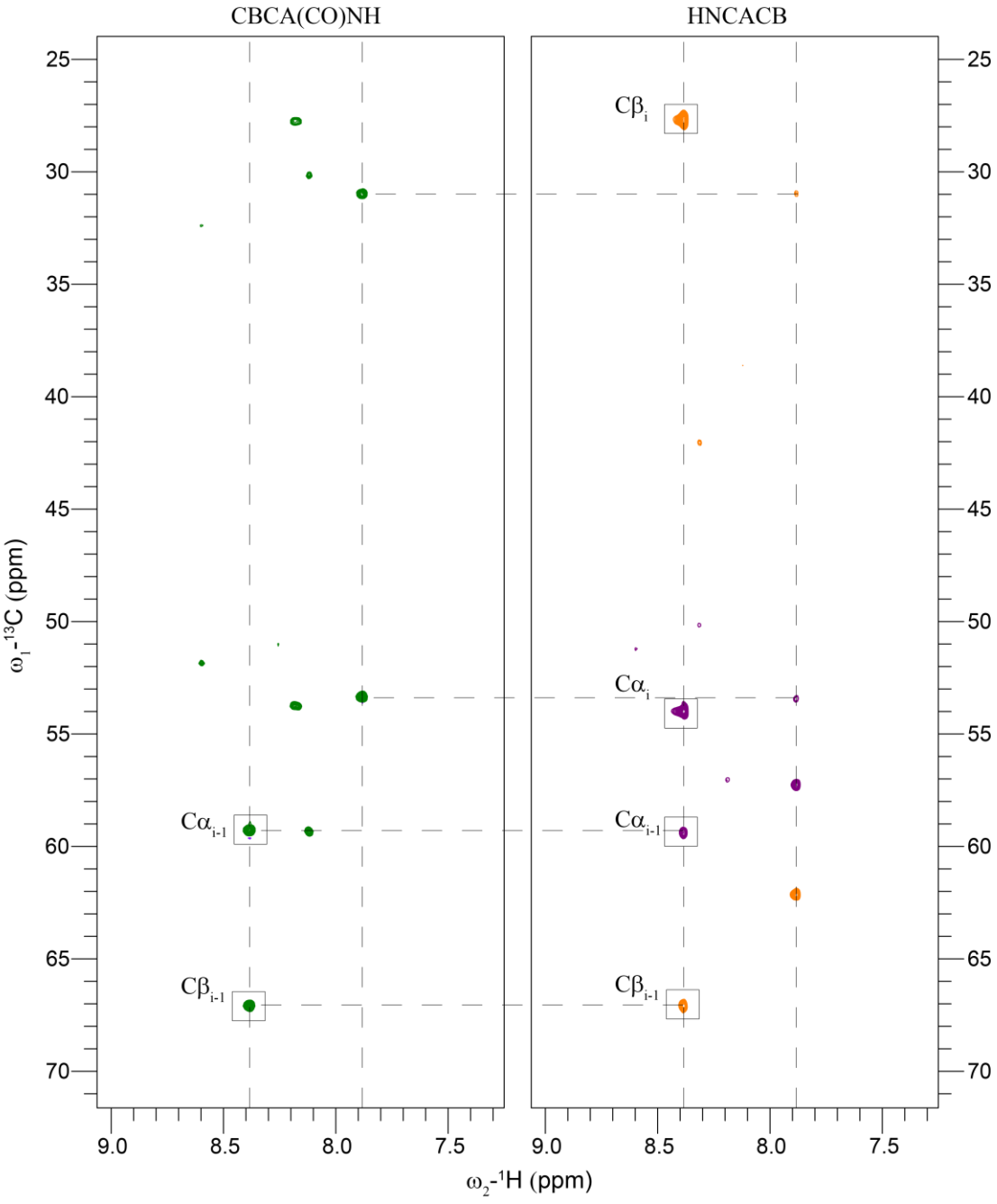


Figure 13 Representation of protein backbone assignment

Sections of CBCA(CO)NH and HNCACB experiments used for backbone assignments for RBM5 RRM1-Zf1-RRM2. Sequential walk linking (i-1) and (i) residues between the two experiments is shown.

After the sequential connectivity is achieved, it is necessary to identify which resonance corresponds to which residue in the protein. For this, certain residues which have very typical C_α or C_β chemical shifts can be used as starting points. For example, a Glycine residue is easily identifiable as it contains only C_α resonance which typically appears ~ 45 ppm or Alanine residue which has a characteristic C_β chemical shift ~15-20 ppm. Similarly, unambiguous assignments can be achieved by looking at neighbouring residues as well which might possibly yield a single solution.

Preliminary information reagrding the secondary structure elements of the protein can be already extracted from the backbone assignments. The Cα and Cβ chemical shifts are sensitive indicators of the secondary structure elements in the protein including α-helix, β-sheet and loops (Spera and Bax 1991). For this purpose, the random coil shifts for each amino acid are subtracted from the actual chemical shift observed for the protein. The random coil chemical shifts can be extracted from previously published databases (Wishart, Sykes et al. 1992). The secondary chemical shifts in the structured parts of the protein differ significantly from random coil chemical shifts with positive deviations for α-helical regions and negative deviations for β-strands.

Next, side chain assignment experiments like HccH-TOCSY and hCCH-TOCSY are recorded to assign all side chain carbon and hydrogen atoms for each residue. Additionally, CC(CO)NH and H(CCO)NH side chain assignment experiments can be recorded which help connect carbon and hydrogen atoms of residue i-1 to the backbone amide of residue i. It is a very helpful experiment as the correlations are directly made to the backbone amides present in ^{1}H-^{15}N HSQC, the only negative point being that Proline residues which do not have an amide resonance in the ^{1}H-^{15}N HSQC would not be visible in these experiments.

After achieving complete or near complete assignment of the protein, one can proceed to structure determination using a variety of experiments including the traditional NOE (Nuclear Overhauser Effect) based experiments as well as experiments used to obtain long-range distance restraints like PRE (Paramagnetic Relaxation Enhancement) or orientation restratints like RDC (Residual Dipolar Coupling) measurements.

2.1.5. Nuclear Overhauser Effect (NOE)

The Nuclear Overhauser Effect arises from the fact that spins which are close in space will have dipole-dipole interaction whereby the spins do not relax independently and have

cross-relaxation effects on each other. Both the spins are therefore connected through space by a mechanism known as cross-relaxation. If two spins (I and S) are coupled to each other, not by J-coupling but due to spatial proximity, and spin I is irradiated, the magnetisation of spin S will be affected as well which will be manifested in the change in signal intensity in the corresponding spectrum.

NOE is characterized by a cross-relaxation rate (σ) and has a strong dependence on the type of nuclei and on the distance between them. It is inverse proportional to the sixth power of the distance between the two spins coupled via dipole-dipole interaction and is only observed over relatively short distances, i.e. <5 Å. It is also dependent on the size of the molecules and the NOESY mixing time. In the initial rate approximation NOE induced peak intensities are proportional to the relaxation rate constants.

For the purpose of structure calculation, NOE cross-peaks in NOESY spectra are manually picked and peak volumes integrated and used as distance restraints. Since the number of such cross-peaks is tremendous, we employ automatic peak assignment using CYANA (Guntert 2004) which uses complex procedures to analyze and assign the complex NOE patterns, iteratively during the process of structure calculation.

2.1.6. Residual Dipolar Coupling (RDC)

The magnetic field experienced by a nucleus depends upon its spatially neighboring nuclei due to dipole-dipole interaction. This phenomenon is known as dipolar coupling.

The magnitude of dipolar interaction depends upon the relative orientation of the nuclei with respect to the external magnetic field and distance between the two nuclei.

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r_{ij})3} \cdot \frac{3cos^2\theta_{ij} - 1}{2} \qquad \text{Eq. 12}$$

where r_{ij} is the inter-nuclear distance between spins i and j, γ_i and γ_j are gyromagnetic ratios of spins i and j, h is the Planck's constant, θ_{ij} is the angle between the inter-nuclear vector and external magnetic field.

In solid state NMR, the dipolar couplings are huge increasing the spectral complexity. On the other hand, in solution NMR, due to Brownian motion of the molecules in solution, the dipolar couplings average out. It has been shown that dilute alignment media can be used to create partial alignment of the biomolecules in solution such that the molecules adopt a

preferred alignment direction in the medium without compromising the simplicity of the spectra (Sanders and Schwonek 1992, Tjandra and Bax 1997). Several alignment media are available including bacteriophage (Pf1) (Zweckstetter and Bax 2001), liquid crystalline media (Rückert and Otting 2000, Lorieau, Yao et al. 2008) and stretched or compressed gels (Ishii, Markus et al. 2001).

Since dipolar coupling is defined in terms of a molecular frame, the measurement of two dipolar couplings can provide orientational information with respect to the molecular coordinate frame and in turn towards each other. It is therefore a useful phenomenon which can be exploited to gain orientational information, especially in case of multi-domain proteins where it can be readily used to define domain orientations.

In the case where anisotropy is introduced due to partial alignment of the protein, not all orientations of the protein can be sampled with equal probability. The alignment of the protein can then be described by an alignment tensor A expressed as a traceless matrix with its principal components A_{xx}, A_{yy}, A_{zz} and the magnitude of dipolar coupling (D_{ij}) can be measured as (Eq. 13).

$$D_{ij}(\theta,\varphi) = D\left\{\left(3cos^2\theta_{ij} - 1\right) + \frac{3}{2}R(sin^2\theta_{ij}cos2\varphi_{ij})\right\} \quad \text{Eq. 13}$$

$$D = \frac{3}{4}.D_{ij}.A_{zz} \quad \text{Eq. 14}$$

$$R = \frac{2}{3}.\frac{A_{xx} - A_{yy}}{A_{zz}} \quad \text{Eq. 15}$$

where D and R are the axial and rhombic components of the molecular alignment tensor A in the principal coordinate frame, θ_{ij} is the angle between the inter-nuclear vector and the z-axis of the alignment tensor and ϕ_{ij} is the angle between the projection of the inter-nuclear vector in the x-y plane and the x-axis.

RDCs are then measured as the difference in coupling in the spectra of the protein in presence (J + Dipolar coupling) or absence (J-coupling) of alignment medium. They are then analyzed using software to couple them to the molecular frame (Dosset, Hus et al. 2001, Zweckstetter 2008). In addition to being used as independent restraints in NMR structure calculations, a great application of RDCs is structure validation, where the experimental RDC data is fitted to back-calculated data from the available solution or crystal structure using PALES (Zweckstetter 2008). The program outputs a data correlation *R* factor and a Cornilescu

Q factor (Cornilescu, Marquardt et al. 1998) which are indicative of the quality of the fit. Since RDCs are quite sensitive, even slight changes in the solution and crystal conformations of the protein can be detected. It therefore becomes necessary to refine the structure using RDCs using programs like Aria (Linge, Habeck et al. 2003) and XPLOR-NIH (Schwieters, Kuszewski et al. 2003).

2.1.7. Paramagnetic Relaxation Enhancement (PRE)

The presence of unpaired electrons at specific sites on the protein alters the relaxation properties of the NMR active nuclei in their close vicinity and enhances nuclear spin relaxation due to hyperfine electron-nucleus coupling. This is called the PRE effect, which can provide useful structural information owing to the distance dependence of the PRE between the paramagnetic center and the nucleus of interest. It is sensitive to minor changes in the environment of the paramagnetic center and can be used to report on dynamic processes containing lowly populated states as well. A big advantage of PRE over NOE is that while NOE can only report on short-range interactions (<5 Å), the PRE effect is very large permitting long-range interactions (<35 Å) to be detected.

For a long time after the introduction of PRE (Solomon 1955, Bloembergen and Morgan 1961), its applicability was only limited to metal-binding proteins (Bertini, Luchinat et al. 2001). In 1980s, for the first time spin-labeling was employed to obtain PRE derived distance restraints on lysozyme and bovine pancreatic trypsin inhibitor (Schmidt and Kuntz 1984, Kosen, Scheek et al. 1986). Briefly, a spin-label with an unpaired electron is chemically coupled to the protein of interest traditionally by introducing surface exposed cysteine mutations to which the spin-label can bind. The spin-label is a small compound containing a nitroxide which specifically binds to the exposed cysteines in the protein. The spin-label is in active state in an oxidizing environment while inactive in reducing environment. Therefore, the PRE effect is detected as the decrease in peak intensity of the oxidized state versus reduced state of the protein as well as their altered R_1 and R_2 relaxation rates.

Typically, ^{1}H-^{15}N HSQC spectra are measured in the oxidized and reduced states whereby peak intensities are extracted. Peak intensity reduction is considered mainly to occur via R_2 relaxation, while the relaxation occurring via R_1 relaxation is considered to be insignificant. Also, relaxation effects on ^{15}N nuclei from the spin-label are considered to negligible compared to ^{1}H due to much lower gyromagnetic ratio of ^{15}N nuclei. The peak

intensity ratio of oxidized and reduced states are related to R_2 relaxation via the following equation (Battiste and Wagner 2000)-

$$\frac{I_{ox}}{I_{red}} = \frac{R_2 exp\ (-R_2^{sp} t)}{R_2 + R_2^{sp}} \quad \text{Eq. 16}$$

where I_{ox} and I_{red} are peak intensities of the spin-labeled protein in oxidized and reduced states, R_2 is the intrinsic relaxation of the amide proton, R_2^{sp} is the spin-contribution to the relaxation rate and t is the total INEPT evolution time of the HSQC.

The paramagnetic rate enhancement (R_2^{sp}) can be converted into distance by accounting for the effect of paramagnetic spins on nuclear magnetic relaxation using the following equation-

$$r = \left[\frac{K}{R_2^{sp}}\left(4\tau_c + \frac{3\tau_c}{1 + \omega_h^2 \tau_c^2}\right)\right]^{1/6} \quad \text{Eq. 17}$$

where r is the distance between the electron and nuclear spins, K is a constant dependent on the type of nucleus, τ_c is the total correlation time for the electron-nuclear interaction, and ω_h is the Larmor frequency of the nuclear spin (proton).

The PRE data can be qualitatively used to validate NMR structure ensemble or crystal structure. It can also be used quantitatively as restraints during structure calculations where r is the distance between an amide proton observed in the ^{1}H-^{15}N HSQC to the spin label attached to the protein. As with NOEs, the PRE distance restraints are also grouped into categories depending on the distance ranges. The PREs might be particularly useful in case of multi-domain proteins where transient interactions between the domains can be easily detected by introducing spin-labels which would otherwise be missed in traditional NOE based experiments.

2.1.8. Structure calculation in solution

NMR solution structure calculation utilizes simulated annealing protocol whereby the system is virtually heated and then slowly cooled down. The software then tries to find the atom coordinates that best fit to the NMR restraints provided. Usually the first step is backbone and side chain assignment, as explained above, followed by collecting and assigning NOE spectra and using NOE peak volumes as distance restraints which are fed into the structure calculation program.

The first structure using NOE-derived distance restraints was calculated three decades ago (Williamson, Havel et al. 1985). Currently additional distance and orientation restraints are obtained from PRE and RDC measurements which help position the different structural elements of the protein (for example, individual domains in a multi-domain protein) as illustrated in **Figure 14**. Dihedral angle restraints from the backbone (ϕ and Ψ) and sometimes from side chains (χ1 and χ2) are also used for the structure calculation. For backbone dihedral angle restraints, a bioinformatics program TALOS+ (Shen, Delaglio et al. 2009) is used.

Apart from the experimental distance and angle restraints, restraints derived from the proper geometry of the molecule, like bond length, chirality or planarity of the aromatic rings and peptide bonds are used during structure calculation. The structure calculation protocol is an iterative process whereby at each step, automated assignment of NOE cross-peaks is done. It consists of a number of cycles whereby an ensemble of lowest energy structures consistent with input restraints is given as the output. With each cycle, the convergence of the NMR ensemble increases with increase in the number of NOE cross-peaks being assigned. The quality of the ensemble is scored by the agreement between the calculated structure and input restraints, as well as the number of restraint violations it has. The stereo-chemical quality of the structure is judged by quantifying the distributions of backbone and side chain dihedral angles, the number of van der Waals steric clashes using NMR software programs like iCING (Doreleijers, Sousa da Silva et al. 2012).

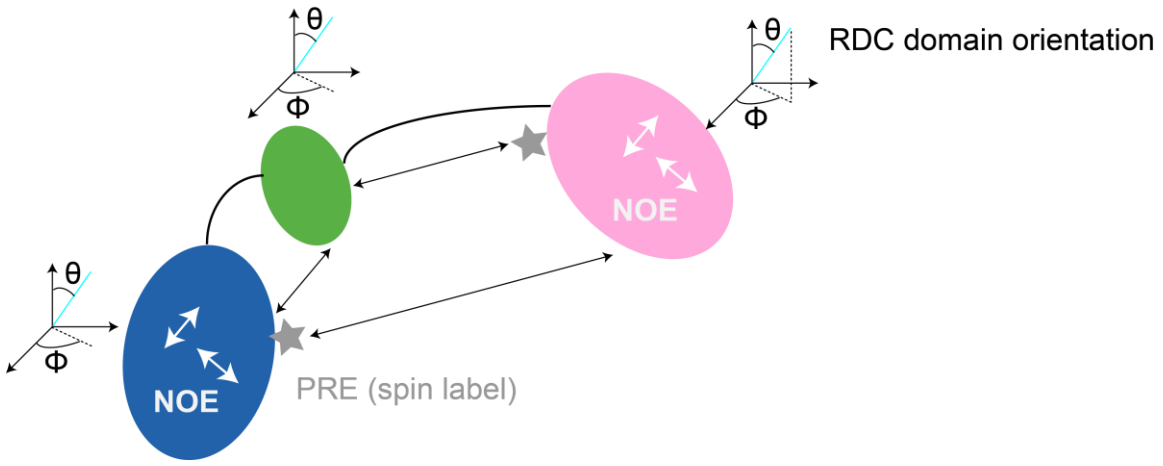


Figure 14 Depiction of different NMR experiments yielding different information

NOE, PRE and RDC experiments can be used as restraints in NMR structure calculation. Different domains of a multi-domain protein are shown in blue, green and pink.

2.1.9. Protein dynamics by NMR

Relaxation is the process by which nuclear spins return to equilibrium state where the population in the different energy levels is described by Boltzmann distribution. In solution NMR, protein dynamics can be studied by exploiting the relaxation properties of the systems under consideration. There are different experiments to study dynamics in the picosecond-nanosecond timescale providing information on backbone and sidechain dynamics to microsecond-millisecond timescale providing information on conformational exchange. To study dynamics at the level of seconds, fast data acquisition techniques allow events to be measured in real-time, for example with the use of SOFAST-HMQC experiments (Schanda, Kupce et al. 2005). Here experiments used to study picosecond-nanosecond dynamics are discussed, since they were majorly used in this thesis. Therefore, one can study and monitor the rate constants for two relaxation pathways: Spin-Lattice relaxation (or longitudinal relaxation, relaxation along z-axis T_1) and Spin-Spin relaxation (or transverse relaxation, relaxation in x-y plane T_2).

The spin-lattice relaxation is induced by the interaction of protein nuclear spins with the surrounding lattice. The lattice is assumed to be in thermal equilibrium and have infinite heat capacity perpetually. After an RF pulse is applied, the rate of spontaneous relaxation of the spins is almost zero, and most of the T_1 relaxation is caused by transient magnetic fields created by random Brownian motion of the spins. The local fluctuations in the magnetic fields create the transition between spin states which in turn leads to recovery of the z-component of the magnetization towards its thermal equilibrium. This recovery is described by the time constant T_1 or the relaxation rate $R_1=1/T_1$. T_1 relaxation occurs most efficiently at the Larmor precession frequency.

The spin-spin relaxation, as suggested by its name, is caused by the interaction between nuclear spins leading to the loss of the coherence between them which leads to loss of magnetisation in the x -y plane. It follows an exponential decay and is described by time constant T_2 or the relaxation rate $R_2=1/T_2$. T_2 relaxation is caused by molecular fluctuations at any frequency. Transverse relaxation is also caused by chemical exchange which might lead to line broadening opening the possibility to study exchanging residues.

The rotational correlation time $\tau_{c.,}$ represents the average time needed by a molecule to rotate by an angle of ~1 radian. It gives information about the molecular size and the flexibility of each amino acid in the protein sequence (Kay, Torchia et al. 1989). The rotational diffusion

or molecular motion can occur at a range of frequencies, and the probability function of finding motions at a given angular frequency (ω) is described by the spectral density function-

$$J(\omega) = \frac{2\tau_c}{1 + (\omega\tau_c)^2} \qquad \text{Eq. 18}$$

It can be shown that for dipolar relaxation, the T_1 relaxation rate is proportional to the square of the dipole filed strength times the spectral density of the filed fluctuation at frequency ω, the spectral density has the following appearance-

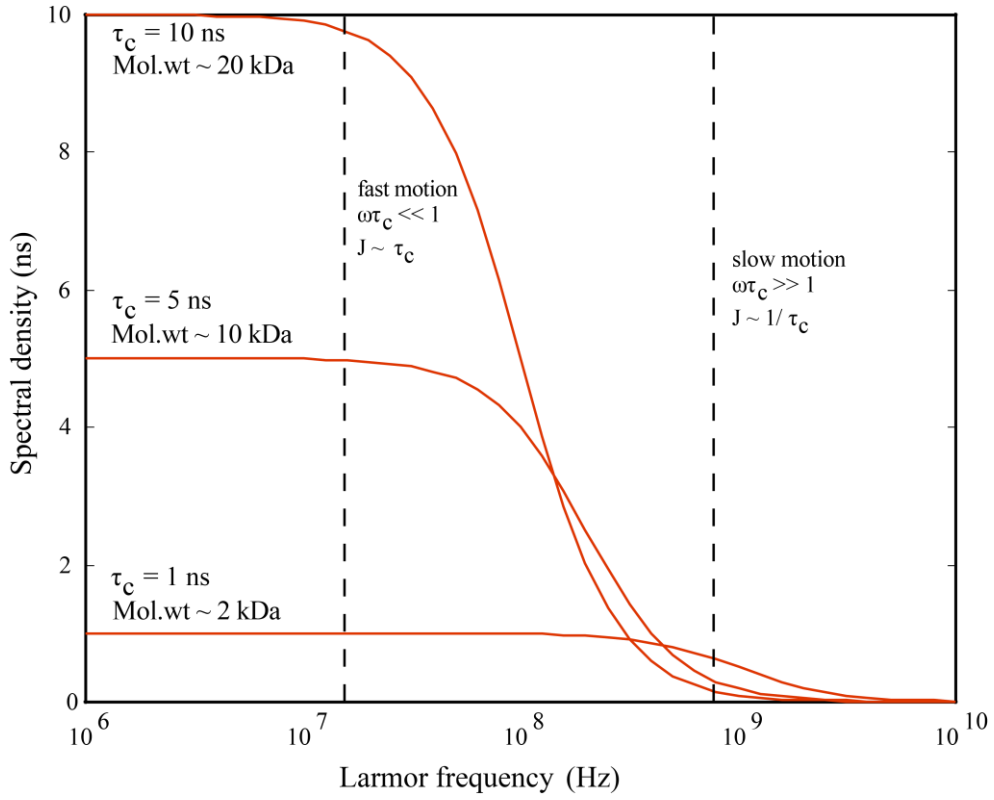


Figure 15 Spectral density function

Efficiency of relaxation for different Larmor frequencies and rotational correlation times τ_c is presented (adapted from Understanding NMR spectroscopy,(Keeler 2002))

Figure 15 shows that for small molecules (with low τ_c values, $\omega\tau_c << 1$), the spectral density function is spread out and there is not much density at the Larmor frequency. Similarly, for large molecules (with high τ_c values, $\omega\tau_c >> 1$), the spectral density is small at the Larmor frequency thereby making T_1 relaxation for both small and large molecules inefficient. However, between the two limits there is efficient T_1 relaxation.

The behavior of *T*1 and *T*2 relaxation as a function of the correlation time is shown in the following figure-

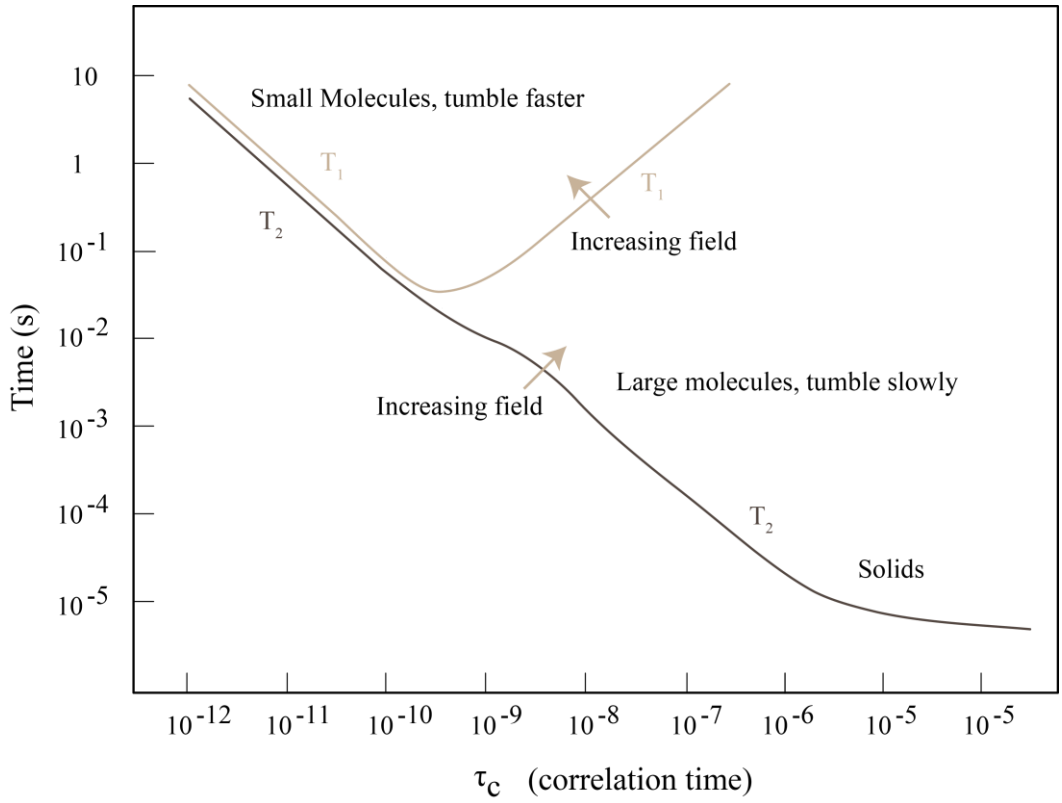


Figure 15 Behavior of T_1 and T_2 as a function of correlation time τ_c

(adapted from (Bloembergen, Purcell et al. 1948))

There are several mechanisms by which the molecular motion of the spins can affect their relaxation including- dipole-dipole relaxation (due to interaction of neighboring spins with each other), chemical shift anisotropy (differential shielding of spins due on differential orientations leading to differential magnetic field at the nucleus).

Another informative experiment is the ^{1}H-^{15}N heteronuclear NOE experiment, which reports on the internal motion of individual amide bond vectors. Those amides which undergo motion faster than the overall tumbling of the molecule show a decreased NOE peak intensity relative to the average of the residues and those having a value lower than 0.77 are considered to be flexible regions of the protein (Kay, Torchia et al. 1989).

The heteronuclear cross relaxation rate (R_s) is measured, which occurs by saturating the proton spin (I) and observing changes in heteronuclear spin (^{15}N, spin S). The steady state NOE enhancement is calculated as follows-

$$NOE = \frac{I_{sat}}{I_{eq}} \qquad \text{Eq. 19}$$

where I_{sat} and I_{eq} are the intensities of a peak in the spectra collected with and without proton saturation.

Usually duplicates of the experiment are collected and analysed in the same manner to calculate the uncertainty of the measurements.

2.2. X-ray Crystallography

Ever since the first structures of myoglobin (Kendrew, Bodo et al. 1958, Kendrew, Dickerson et al. 1960) and hemoglobin (Perutz, Rossmann et al. 1960, Perutz, Miurhead et al. 1968, Perutz, Muirhead et al. 1968) were solved using X-ray crystallography, it has been routinely used to obtain atomic resolution structures of biological macromolecules. Currently, the Protein Data Bank (PDB) has 112,775 structures solved using X-ray crystallography as opposed to only 11,720 solved using NMR and a meagre 1,370 using electron microscopy. These statistics clearly show how instrumental the technique has been in providing structural insights into biological molecules. The ease of use of the technique and the absence of size limitations make it the primary method to study biological macromolecules.

2.2.1. Protein crystallization

To obtain the atomic resolution 3D structure of a protein or protein-ligand complex, a diffraction quality crystal is required. For this, the protein in question must be available in a highly pure and homogenous form. Here Dynamic Light Scattering (DLS) could be used to characterize the polydispersity of the sample, to check for presence of oligomeric states or even aggregates. Next, the suitable conditions required for protein crystallization are usually obtained by using a crystallization screening method like sparse-matrix approach. The quality of the crystals might be improved by fine-tuning the conditions around the parent conditions using a grid-screen method.

For crystallization of the protein to occur, the solution needs to be brought to a supersaturated state via a gradual decrease in the solubility of the protein which is achieved by the addition of precipitants. Of the various methods for protein crystallization developed, the vapor diffusion method (with sitting or hanging drop) remains the most widely used. Here a drop containing protein mixed with the crystallization buffer is setup and is equilibrated against the reservoir solution containing the crystallization buffer. Since the concentration of the precipitant is higher in the reservoir than in the drop, a concentration gradient is developed. Due to this, the water from the drop evaporates towards the reservoir, decreasing the volume of drop while increasing the protein concentration. At some point during this process, the protein will reach supersaturation and hopefully crystallize.

2.2.2. Principles of X-ray crystallography

X-ray diffraction is the phenomenon of the slight bending of X-rays as they passes around the edge of an object. The amount of bending is dependent on the relative size of the wavelength of the incident beam to the size of the opening. When X-rays pass through the crystal, they are diffracted due to their interaction with the electron cloud surrounding the atoms of the crystals. Each diffracted X-ray beam creates a spot on the X-ray detector. The diffraction pattern observed when the X-rays pass through a slit shows constructive and destructive interference arising due to in phase and out of phase interaction of light waves, respectively. The intensities of the spots are subsequently used to calculate the electron density of the molecules within the crystal.

In a protein crystal, the protein molecules are arranged in an ordered manner. It can be considered as a three dimensional crystal lattice having a regular arrangement of repeating elements called unit cells. The unit cell may possess internal symmetry whereby two or more structures in the unit cell are related to each other by a symmetry element and are called symmetry mates. A unit cell can be subdivided into asymmetric units which may further contain more than one protein molecules.

The unit cell is defined by three axes and the angles between them, denoted as a, b, c and α, β, γ, respectively. Each atom in the crystal could be represented by a point to obtain a crystal lattice. Within this crystal lattice, infinite numbers of planes could be drawn through the lattice points and the lattice could be represented by Millers indices (hkl).

The set of equally spaced parallel planes in a single crystal that can be drawn through the lattice points serve as the source of diffraction. These set of planes (with hkl indices and inter-planar spacing d_{hkl}) can produce a strong diffraction pattern only if the Bragg's law is satisfied (Eq. 20).

$$2d_{hkl}sin\theta = n\lambda \qquad \text{Eq. 20}$$

where d_{hkl} is the inter-planar spacing, θ is the angle of diffraction, λ is the incident X-ray wavelength and n is an integer.

The Ewald sphere construction is a geometric construction that helps in visualization of the properties of Bragg's law as shown in **Figure 16**.

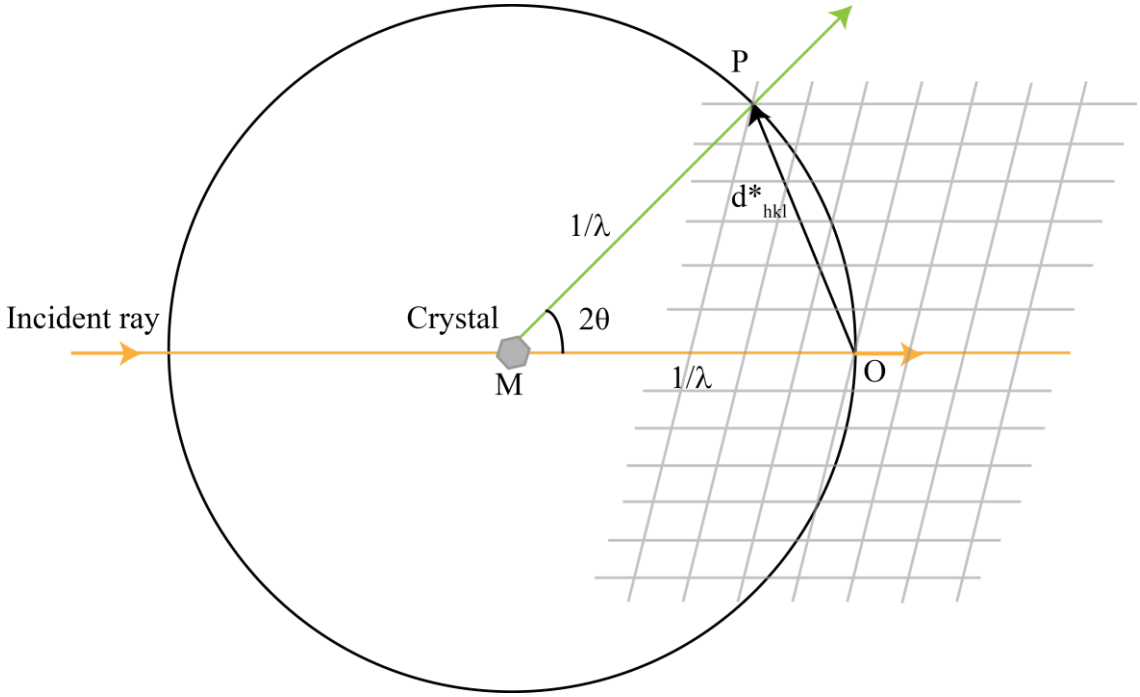


Figure 16 Ewald sphere

The Ewald sphere serves as a useful means to understand the occurrence of diffraction spots from a crystal.

Diffraction from protein crystals can be interpreted by the Ewald's sphere which is a sphere of reflection with radius 1/λ passing through the origin of the reciprocal lattice (O), having the crystal at its center (M). Diffraction occurs whenever a reciprocal lattice point comes in contact with the Ewald sphere. As the incident beam is scattered by the crystal, a reflection occurs in the direction of MP, where P is the operative reciprocal lattice point. The vector joining the origin to the operative reciprocal lattice point is denoted as d^*_{hkl}. Its value is equal to $1/d_{hkl}$ and its direction is perpendicular to the real-space hkl planes. As the crystals rotates, other lattice points come into the contact with this sphere thus producing new reflections.

Every atom in the unit cell contributes to every reflection owing to its chemical nature and relative position. Depending upon this shift in position of one atom relative to the others, the contribution from each atom has a phase shift relative to the others. The intensity of each reflection with Miller indices (hkl) is proportional to the square of the structure factor, which is given by-

$$F_{hkl} = \sum_J f_j exp\left[2\pi i(hx_j + ky_j + lz_j)\right] \qquad \text{Eq. 21}$$

where every atom j in the unit cell contributes to every structure factor (F_{hkl}) owing to its position (x_j, y_j, z_j)and chemical nature (f_j) in the unit cell.

The electron density (ρ_{xyz}) at every given position in the unit cell can be calculated as the Fourier transform of the structure factor-

$$\rho_{xyz} = \left(\frac{1}{V}\right)\sum_{hkl} F_{hkl}\, exp[-i\alpha_{hkl}]\, exp[-2\pi i(hx + ky + lz)] \qquad \text{Eq. 22}$$

where x,y,z in the equation of electron density refer to arbitrary places in the unit cell in contrast to x_j, y_j, z_j in the structure factor equation referring to atomic coordinates and α_{hkl} is the phase. The electron density equation can thus be solved if the structure factor amplitude and phases for all hkl planes are known. In the diffraction experiment, the amplitude of the structure factor is measured while the phases are lost. This is known as the 'phase problem' in crystallography. There are three methods for determination of the phases, namely Molecular Replacement (MR), Multiple Isomorphous Replacement (MIR) and Multi-wavelength Anomalous Dispersion (MAD).

2.2.3. Molecular Replacement (MR)

Molecular replacement is an approach for solving the phase problem of a protein when the structure of a very similar molecule is already known (Rossmann 1972). It is usually successful in cases with high sequence identity (>40 %) between the target protein and its homologue. It involves the solution of rotation and translation functions where the known molecule is rotated in three dimensions such that there is maximum agreement between the calculated structure factors of the model and the actual structure factors from the diffraction. Next, to identify the correct translation, the oriented model is placed at every position in the unit cell to obtain maximum agreement. Once the correct orientation and translation are identified, phases for all structure factors and subsequently the electron density can be calculated.

Since there is an ever increasing number of structures deposited in the PDB, the applicability of MR as the first method to solve the phase problem is also increasing. But it might not always be straightforward as the flexible regions of homologous protein with known structure may not always superimpose with the target protein. In such cases extensive model building may be required after the initial model is obtained, for example in the flexible regions and side-chains.

2.2.4. Multiple Isomorphous Replacement (MIR)

Multiple isomorphous replacement is another approach to solving the phase problem whereby the unknown phases of the target structure are calculated by making known changes to the contents of the crystal without disturbing the structure of the protein (Green, Ingram et al. 1954). This involves the introduction of heavy metal atoms to the protein crystal and detecting differences in the diffraction pattern. Since the heavy metal atoms diffract stronger than the rest of the atoms, their positions and therefore the phases can be estimated.

Next, the diffraction pattern and structure factors of the native versus heavy metal atom crystals are compared. For example, if a structure factor derived from native crystal is significantly stronger than that from the heavy atom crystal, there must be destructive interference from diffraction from the heavy atoms and thus the phases must be ~180° apart. On the other hand, if a structure factor derived from native crystal is significantly weaker than that from the heavy atom crystal, there must be constructive interference from diffraction from the heavy atoms and thus their phases must be fairly close. In this manner the relative phases may be calculated. To resolve the ambiguity that the native crystal phase leads or lags the heavy atom phase, multiple crystals with different heavy metal atoms, which hopefully occupy different positions in the crystals are used.

2.2.5. Multi-wavelength Anomalous Dispersion (MAD)

Multi-wavelength anomalous dispersion is yet another method for solving the phase problem (Hendrickson and Ogata 1997). It comes as an alternative to MIR method due to the possibility of using tunable X-ray beamlines. Since the diffraction pattern is largely dependent on the wavelength of the incident X-rays, the properties of anomalously scattering atoms inherently present in the protein can be utilized. For example, Zn^{2+} in case of Zn-finger proteins or using seleno-methionine labeling methods where methionine residues are replaced by seleno-methionine residues thereby exchanging the Sulphur atom by Selenium. Such anomalous scattering atoms have ‘absorption edges’, around which the scattering in terms of amplitude and phase varies. Next, the phase problem can be solved as in case of MIR. However, there are certain advantages of using MAD over MIR.

The anomalous scattering from an atom near an absorption edge is shifted in phase. Therefore, if scattering from a single crystal is measured at an absorption edge and at another wavelength distant from it, unambiguous phase information can be obtained. Since all the

required information can be obtained from a single crystal, the use of multiple crystals which might be a bit non-isomorphous adding to the background noise, as in case of MIR, is negated.

2.3. Small Angle X-ray Scattering (SAXS)

SAXS is a very useful technique for studying biological molecules in solution. It provides low-resolution information on the overall shape, size and conformational polydispersity of the macromolecules in solution. It becomes particularly valuable in studying biological complexes whereby changes in the shape in the free and bound form of the protein are apparent. In case of multi-domain proteins, it can serve as a powerful complementary technique to reflect the domain orientations in solution. It also comes in handy for validation of high resolution structures obtained using NMR or X-ray crystallography. The technique is gaining popularity due to the ease of availability of high intensity X-ray beams (synchrotron access) and rapid data collection (few seconds at synchrotron). This offers the possibility to even study time-resolved experiments involving kinetics.

For recording SAXS data, usually the capillary is filled with 50-70 µl sample at 1-10 mg/ml concentration. A concentration series can be recorded to test if there is a concentration dependent behavior of the sample. In such a case, the lowest concentration data are used for subsequent analysis.

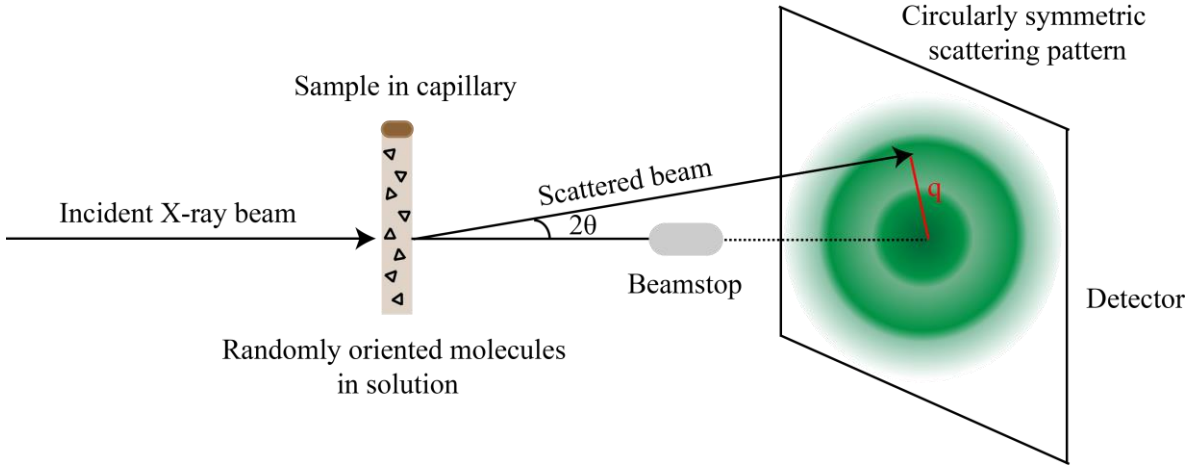


Figure 17 Schematic of SAXS experimental setup

The setup of SAXS experiment is rather simple, the sample is placed in a capillary tube which is exposed to X-ray beam and the intensity of the scattered beam is recorded by an X-ray detector as seen in **Figure 17**. During the experiment, the sample molecules move freely in solution having random orientations, unlike in crystallography where molecules are regularly positioned yielding typical diffraction patterns due to interference. This information is lost in the SAXS measurement which also leads to the low resolution of the technique. Nevertheless,

information on inter-atomic distances is still retained which makes it possible to study the shape and overall structural parameters. In a SAXS experiment, the scattering pattern is described by intensity (I) expressed as a function of the scattering vector q-

$$q = \frac{4\pi sin\theta}{\lambda} \quad \text{Eq. 23}$$

$$I(q) = < \int \left|(\rho(\vec{r}) - \bar{\rho_s})e^{i\vec{q}\vec{r}}d\vec{r}\right|^2 > \quad \text{Eq. 24}$$

where λ is the wavelength of the incident radiation and θ is half of the scattering angle and <> refers to rotational average, $\rho(r)-\rho_s$ is the difference in scattering density between sample molecule at position r and solvent.

Since the buffer also diffracts substantially, the intensity distribution of the macromolecule is obtained by subtracting the SAXS 1D of the buffer from that of the sample.

2.3.1. Structural information in the SAXS curve

There are three basic parts in the SAXS 1D curve from which different data can be extracted- Guinier, Fourier and Porod as illustrated in **Figure 18**.

The experimental radius of gyration (R_g) representing the root mean square of all electrons from the center of mass, can be determined from the Guinier region using Guinier approximation. It was shown by Guinier that for sufficiently small values of q, according to Eq. 25, the plot of ln I(q) vs q^2 should be linear if the condition of $q*R_g < 1.3$ is satisfied for globular proteins. In case of elongated structures, the upper limit for this condition is smaller. Therefore, the Guinier plot provides a good method to check the sample quality as it is greatly affected by aggregation state, polydispersity and improper buffer subtraction.

$$I(q) = I(0)e^{\frac{-q^2R_g^2}{3}} \quad \text{Eq. 25}$$

where I(0) is the intensity of radiation scattered through zero angle.

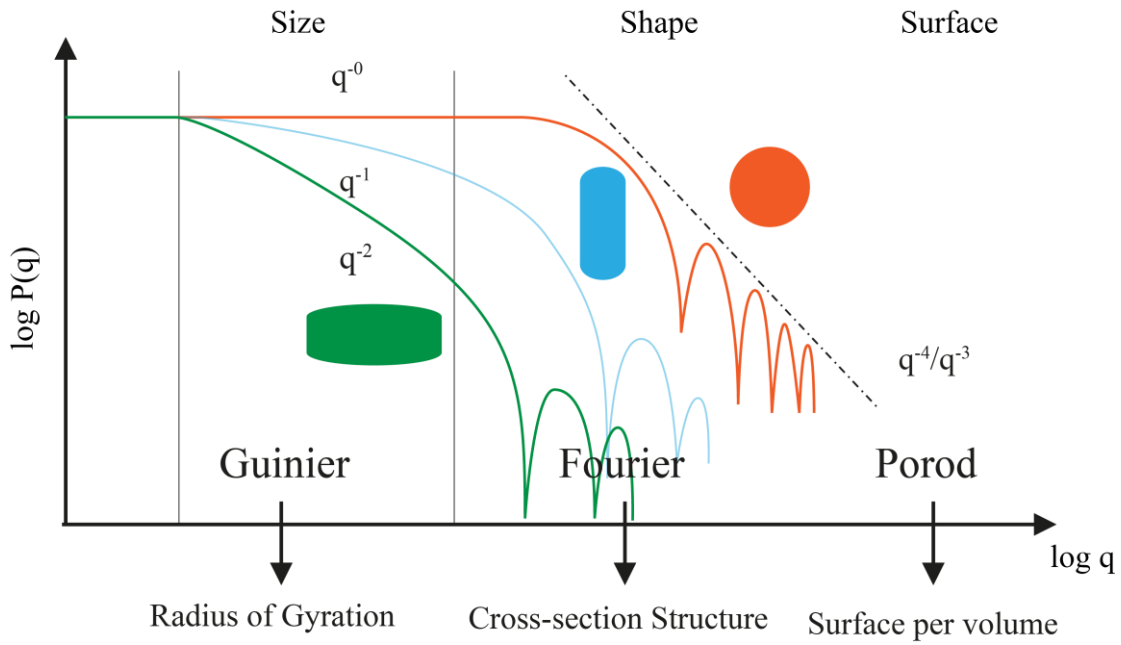


Figure 18 Different regions of SAXS 1D profile

The different regions in a SAXS curve provide information related to size, shape and surface of the biomolecule. Rough SAXS profiles for globular, cylindrical and lamellar molecules are shown in orange, blue and green, respectively. (adapted from The SAXS guide: getting acquainted with the principles (Schnablegger and Singh 2013))

Since the plot of intensity I(q) vs q representing the shape of the molecule is not entirely intuitive, Fourier transform of the scattering profile is used to obtain the pair-wise distribution function p(*r*) which gives the distribution of the electrons that are within distance r of each other. This yields maximum linear dimension D_{max} (p(*r*) at r=0) and R_g. The p(*r*) curve provides information on the overall shape of the molecule and is particularly useful if binding to a ligand induces huge changes in protein conformation for example, formation of closed and extended states.

Lastly, from the Porod region information such as surface-volume ratio can be obtained. Additionally, the Porod plot of q^4 I(q) vs q provides valuable information on molecular weight and Porod volume.

As mentioned previously, the SAXS curve can also be used to validate high resolution three dimensional structures obtained by NMR or crystallography. Here, the theoretical SAXS curve is calculated from the structure and compared to the experimental SAXS curve. The deviation, represented as χ^2 shows the agreement between the curves.

Scope of the thesis

Splice site recognition with the help of *trans*-acting splicing factors plays a key role in alternative splicing regulation. It is intriguing how these splicing factors can specifically yet differentially regulate a repertoire of pre-mRNA targets. Since many of the known splicing factors are multi-domain proteins, it is possible that the individual domains are responsible for molecular recognition of distinct pre-mRNA targets thereby expanding their functional capacity. Additionally, the individual domains with relatively weak RNA binding affinity may cooperate with each other to recognize RNA ligands with high affinity providing a further degree of possible manipulation. The focus of this thesis is to study one such multi-domain splicing factor, RBM5, which regulates alternative splicing of its targets in a diverse set of ways.

RBM5 is a putative tumor suppressor protein that is frequently deleted in lung cancer while it is consistently up regulated in breast cancer, thereby indicating its complex role in tumor progression. It is also known to regulate alternative splicing of death receptor *Fas*, where it promotes the formation of its anti-apoptotic form while in case of initiator *Caspase-2*, it promotes its pro-apoptotic isoform, making the role of RBM5 context dependent. Consequently, it becomes highly interesting to study the involvement of the different domains of the multi-domain protein RBM5.

It was shown previously that the RBM5 OCRE domain regulates *Fas* pre-mRNA splicing via direct interactions with SmN/B/B' proteins part of the core spliceosomal assembly, thereby recruiting the tri-snRNP to distal splice sites. The structural basis of these interactions were unraveled using a combination of NMR spectroscopy, CD and ITC. It was also found that the closely related RBM10 OCRE domain performs a similar function while RBM6 OCRE domain is not able to regulate alternative splicing of *Fas* pre-mRNA owing to its truncated structure.

Additionally, the role of RNA binding domains of RBM5 was investigated with respect to alternative splicing regulation of *Caspase-2* pre-mRNA using an integrated structural biology approach with a combination of NMR, SAXS and X-ray crystallography. In this thesis, it was important to use such complementary methods to not only obtain high resolution structural information, but also to study the dynamics of the individual domains involved in protein-RNA recognition.

Chapter 3: Materials and Methods

3.1. Materials

3.1.1. Buffers

BUFFER	COMPONENTS
Lysis buffer	20mM Tris pH 7.5, 500mM NaCl, 10mM Imidazol, 0.002% NaN_3, 2 mM β-Mercaptoethanol
Elution buffer	20mM Tris pH 7.5, 500mM NaCl, 500mM Imidazol, 0.002% NaN_3, 2 mM β -Mercaptoethanol
TEV cleavage buffer	20 mM Tris pH 7.5, 200mM NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1 dilution buffer	20 mM Tris pH 7.0, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1 ResS-A buffer	20 mM Tris pH 7.0, 50 mM NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1 ResS-B buffer	20 mM Tris pH 7.0, 1 M NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1 lysis buffer	20 mM Hepes-Na, pH 7.5, 500 mM NaCl ,1M Urea , 5 mM β -mercaptoethanol
SP dilution buffer	20 mM Hepes-Na, pH 7.5, 1M Urea ,1mM PMSF, 5 mM β -mercaptoethanol
SP-A buffer	20 mM Hepes-Na, pH 7.5, 100 mM NaCl, 1M Urea ,1mM PMSF, 5 mM β -mercaptoethanol
SP-B buffer	20 mM Hepes-Na, pH 7.5, 2 M NaCl, 1M Urea ,1mM PMSF, 5 mM β -mercaptoethanol
HA-dilution buffer	10 mM K.phosphate, pH 7.4, 5 mM β -mercaptoethanol,1 mM PMSF
HA-A buffer	10 mM K.phosphate, pH 7.4, 75 mM NaCl, 5 mM β -mercaptoethanol
HA-B buffer	10 mM K.phosphate, pH 7.4, 75 mM NaCl, 5 mM β -mercaptoethanol, 12% w/v $(NH_4)_2SO_4$
RRM1-Zf1-RRM2 lysis buffer	20 mM Tris pH 7.0, 500 mM NaCl, 1 M Urea, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1-RRM2 wash buffer	20 mM Na.phosphate pH 7.0, 500 mM NaCl, 1 M Urea, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1-RRM2 elution buffer	20 mM Na.phosphate pH 6.0, 500 mM NaCl, 1 M Urea, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1-RRM2 TEV cleavage buffer	10 mM Na.phosphate pH 7.0, 400 mM NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1-RRM2 dilution buffer	10 mM Na.phosphate pH 7.0, 0.002% NaN_3, 2 mM β -Mercaptoethanol
RRM1-Zf1-RRM2 ResS-A buffer	10 mM Na.phosphate pH 7.0, 50 mM NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol

RRM1-Zf1-RRM2 ResS-B buffer	10 mM Na.phosphate pH 7.0, 1 M NaCl, 0.002% NaN_3, 2 mM β -Mercaptoethanol
NMR/ITC/Crystallisation buffer	
OCRE SEC buffer	20 mM Na.phosphate pH 6.5, 100 mM NaCl, 1 mM DTT
SEC buffer 1	20 mM MES pH 6.5, 400 mM NaCl, 1 mM DTT
SEC buffer 2	20 mM MES pH 6.5, 100 mM NaCl, 1 mM DTT

3.1.2. ^{15}N labelled M9 salts

MEDIUM	COMPONENTS/LITRE
Lysogeny broth (LB) medium	1% tryptone, 0.5% yeast extract, 0.5% NaCl
^{15}N Labelled M9 minimal medium	100 ml M9 salt solution (10X), 20 ml 20% (w/v) glucose, 1 ml 1 M $MgSO_4$, 0.3 ml 1 M $CaCl_2$, 1 ml biotin (1 mg/ml), 1 ml Thiamin (1 mg/ml), 10 ml trace elements solution (100X)
^{15}N, ^{13}C Labelled M9 minimal medium	100 ml M9 salt solution (10X), 2g ^{13}C labelled glucose, 1 ml 1M $MgSO_4$, 0.3 ml 1M $CaCl_2$, 1 ml biotin (1 mg/ml), 1 ml Thiamin (1mg/ml), 10 ml trace elements solution (100X)

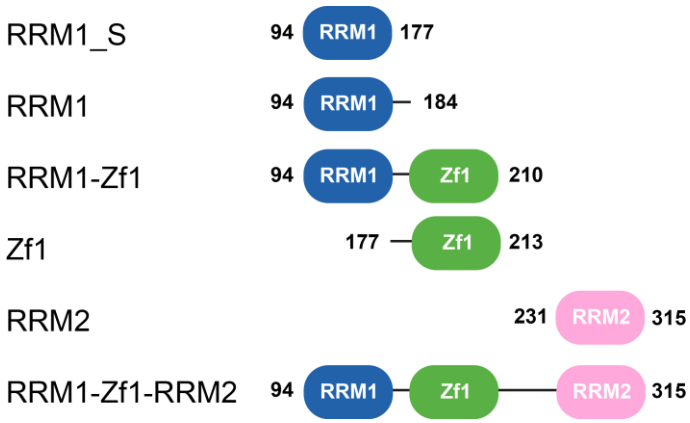
3.1.3. Trace elements solution

TRACE ELEMENTS SOLUTION (100X)	MASS/LITRE
EDTA	5 g/L
$FeCl_3.6H_2O$	0.83 g/l
$ZnCl_2$	84 mg/L
$CuCl_2.2H_2O$	13 mg/L
$CoCl_2.2H_2O$	10 mg/L
H_3BO_3	10 mg/L
$MnCl_2.4H_2O$	1.6 mg/L

3.1.4. List of single-stranded RNA sequences

RNA	SEQUENCE
CU_9	5’- UCUCUUCUC -3’
GGCU_7	5’- CUUGGCU -3’
GGCU_10	5’- UGGCUCUUCU -3’
GGCU_12	5’- UGGCUCUUCUCU -3’
ne_GGCU_13	5’- GAACUUGGCUCUU -3’

3.1.5. Constructs



Note: RRM1_S (residues 94-177) refers to a shorter version of RRM1 domain where the linker connecting RRM1 and Zf1 is deleted.

3.2. Methods

3.2.1. Protein expression and purification

All proteins were expressed in BL21 (DE3) *Escherichia coli* strain. The respective plasmids were then transformed into chemically competent *E.coli* cells and grown overnight at 37 °C on agar plates containing kanamycin resistance (50 µg/ml). 20 ml starter cultures in LB media were inoculated using single colonies from the plates and grown overnight. Next day, scale up of the cultures was done, where the pre-cultures were used to inoculate 1 L LB media containing 50 µg/ml Kanamycin resistance. For constructs containing the zinc finger (Zf1) domain, the cultures were supplemented with 100 µM $ZnCl_2$ solution for proper folding of the domain. The cells were grown up to an OD of about 0.6 at 37 °C after which they were cooled down at 18 °C (RRM1-Zf1-RRM2) or 20 °C (RRM1, RRM1_S, RRM2). Then they were induced with a final concentration of 0.5 mM IPTG solution and grown overnight. Next day, the cultures were centrifuged at 5000 g for 20 min to pellet the cells. In case of RRM1-Zf1 (residues 94-210), the cultures were grown only for 3 h at 37 °C after induction with a final concentration of 0.5mM IPTG solution. The cell pellets were then collected in 50 ml falcon tubes and stored at -20 °C until further use.

For ^{15}N labelling or ^{13}C-^{15}N double labelling of the protein, the 20 ml starter culture was also made in M9 minimal medium instead of LB medium.

For purification of the proteins, the cell pellets were resuspended in 25 ml lysis buffer, supplemented with 0.1 mg/ml lysozyme and 1 mM AEBSF protease inhibitor, followed by incubation with lysozyme on ice for 20 min to weaken the cell walls, before proceeding with sonication. After sonication on ice, the cell lysates were centrifuged at 35000 g for 45 min. The cell debris goes to the pellet while the soluble protein is in the supernatant. The supernatant for the respective protein was then filtered and loaded onto the respective columns.

For OCRE domains (RBM5/6/10) and RRM2, the supernatant was loaded onto a 3 ml bench top Ni^{2+} affinity column equilibrated with lysis buffer. Since the protein of interest has a His-tag, it binds to the column while all other proteins go into the flow through. The column was then washed with 10 CV of lysis buffer after which it was eluted with 20 ml elution buffer. The protein was then mixed with 1 mg/ml TEV protease and cleaved overnight in TEV cleavage buffer at 4°C. Next day, the protein was loaded again onto the Ni^{2+} affinity column where the uncleaved protein, TEV protease and cleaved tag bind to the column while the cleaved protein goes in the flow through. The flow through was then concentrated to a volume of 1 ml in an Amicon® 15 ml concentrator with 3.5 kDa cutoff. It was degassed and loaded onto size exclusion chromatography column (Hiload 16/60 Superdex75 column,GE Healthcare) equilibrated with the respective SEC buffer (OCRE SEC buffer in case of RBM5/6/10 OCRE and SEC buffer 2 in case of RRM2) for final polishing of the protein. Finally, the protein was concentrated to approximately 1 mM concentration and flash frozen in liquid nitrogen in 50 µl aliquots and stored at -80°C until further required.

For RRM1 (RRM1, residues 94-177; RRM1_S, residues 94-184), a cation exchange step was introduced between 2^{nd} Ni^{2+} affinity column and size exclusion chromatography. After 2^{nd} Ni^{2+} column, the protein was diluted 4-fold with RRM1 dilution buffer, filtered and loaded on 1 ml Resource S column (GE Healthcare) equilibrated with RRM1 ResS-A buffer. The protein was eluted with a linear gradient of RRM1 ResS-B buffer from 50 mM NaCl to 1 M NaCl. As a final polishing step, the protein was purified over a size exclusion column pre-equilibrated with SEC buffer 2.

For RRM1-Zf1 (residues 94-210), the protein pellet was lysed in 20 ml RRM1-Zf1 lysis buffer with sonication on ice. After centrifugation of the lysate at 35000 g for 45 min, the supernatant was filtered and diluted 5-fold with SP dilution buffer and loaded on a 20 ml SP-Sepharose column equilibrated with SP-A buffer. The protein was eluted with a linear gradient of SP-B buffer from 100 mM NaCl to 2 M NaCl. The eluted peak fractions were checked on

gel, pooled, diluted 3-fold with HA dilution buffer and loaded on a 15 ml hydroxyapatite (HA) column equilibrated with HA-A buffer. The protein was then eluted with a 2-step gradient of $(NH_4)_2SO_4$ using HA-B buffer. Again, the eluted peak fractions were checked on gel before being pooled and loaded onto size exclusion chromatography column equilibrated with SEC buffer 1. The eluted protein peak fractions were pooled and concentrated to about 1 mM protein concentration in Amicon® 15 ml concentrator, aliquoted in 50 µl fractions, flash frozen in liquid nitrogen and stored at -80°C until further use.

For RRM1-Zf1-RRM2 (residues 94-315), the cell pellet was resuspended in 25 ml lysis buffer, supplemented with 0.1 mg/ml lysozyme and 1 mM AEBSF protease inhibitor. It was then incubated on ice for 20 min before proceeding with sonication. After centrifugation of the lysate at 35000 g for 45 min, the supernatant was filtered and loaded on a 3 ml Zn^{2+} affinity bench top column, equilibrated with RRM1-Zf1-RRM2 lysis buffer. The column was washed with 5 CV RRM1-Zf1-RRM2 lysis buffer and subsequently with 5 CV RRM1-Zf1-RRM2 wash buffer. It was then washed with 5CV RRM1-Zf1-RRM2 wash buffer but with the salt adjusted to 1 M NaCl for removing non-specifically bound nucleic acids. The protein was sequentially washed and eluted with 20 ml each of RRM1-Zf1-RRM2 elution buffer with pH adjusted to 6.0, 5.5, 5.0, 4.5. The eluted fractions were checked on gel and the protein appeared to be mostly pure in fractions with pH 5.5-4.5. For TEV cleavage, 2 mg TEV protease was added to each of the fractions and dialyzed overnight at 4°C in RRM1-Zf1-RRM2 TEV cleavage buffer. After TEV cleavage, the protein was concentrated to 12.5 ml in a 10 kDa cutoff Amicon® concentrator. It was then diluted 8-fold with RRM1-Zf1-RRM2 dilution buffer and loaded on a 1 ml/6 ml Resource S column, equilibrated with RRM1-Zf1-RRM2 ResS-A buffer. The protein was eluted with a linear gradient of RRM1-Zf1-RRM2 ResS-B buffer from 50 mM NaCl to 1 M NaCl. The eluted peak fractions were checked on the gel. The first few fractions from the peak contained TEV protease contamination. These fractions were discarded and the rest were pooled and concentrated again to 1 ml protein solution, after adjusting the final salt concentration to 400 mM NaCl. It was then loaded on a size exclusion column equilibrated with SEC buffer 1. The eluted protein was concentrated, aliquoted in 50 µl fractions, flash frozen in liquid nitrogen and stored at -80°C until further use.

3.2.2. NMR titration experiments

All ^{1}H, ^{15}N Heteronuclear single quantum correlation (HSQC) NMR spectra were recorded at 298K on AVIII600 and AVIII 800 Bruker spectrometers equipped with cryogenic triple resonance probes.

For OCRE domain-SmN peptide titrations, 100 μM of ^{15}N-labeled RBM5 OCRE domain was titrated with 10-fold excess of the respective SmN derived peptide in OCRE SEC buffer, additionally containing 10% D_2O. A semi-quantitative approach was then used to assess the relative binding affinities of wild-type vs mutant peptides to RBM5 OCRE domain. Chemical shift perturbations from 7 RBM5 OCRE residues (Y470, Y471, Y479, D481, N483, S490, Y495) were added for each of the peptide titration and normalized with that of the wild-type peptide to obtain the CSP score.

For RBM5 RRM1 and RRM2, protein-RNA titrations were done with ^{15}N-labeled protein in SEC buffer 2, supplemented with 10% D_2O. The CU_9 RNA was titrated in a step-wise manner to a final ratio of 1:1.

For RBM5 RRM1-Zf1 and RRM1-Zf1-RRM2, since the proteins were not stable at low salt (SEC buffer 2) in the free form, CU_9 or GGCU_12 RNA or ne_GGCU_13 were added to the protein at 1:1 ratio in SEC buffer1. The respective protein-RNA complex was then diluted such that the salt concentration becomes equivalent to that in SEC buffer 2. The sample was then concentrated in a 0.5 ml Amicon centrifugal filter concentrator with 3.5 kDa cut-off. To obtain a comparable spectrum of the free protein in SEC buffer2, the protein sample was diluted to ~50 μM and the ^{1}H-^{15}N-HSQC was measured for longer durations by increasing the number of scans.

3.2.3. NMR backbone and side-chain assignment experiments

All spectra were recorded at 298K on AVIII500, AVIII 600, AVIII800, AVIII950 Bruker spectrometers. For RBM6 OCRE domain, a ^{15}N, ^{13}C labeled 500 μM protein sample was prepared in OCRE SEC buffer with additional 10% D_2O. For backbone resonance assignments, standard experiments including 3D HNCA, HNCACB, CBCA(CO)NH and HNCO were recorded. For side-chain resonance assignments, 3D HCCH-TOCSY with ^{13}C evolution and H(CCO)NH were recorded and used for connecting backbone amide resonances to side-chain resonances. Aromatic resonances were assigned using 2-D ^{1}H-^{13}C HSQC, HBCBCGCDHD, HBCBCGCDCEHE (Yamazaki, Forman-Kay et al. 1993). Additionally, a

^{15}N-edited NOESY-HSQC experiment in 90%H_2O/10%D_2O; aromatic and aliphatic ^{13}C-edited NOESY-HSQC experiments in 100% D_2O were recorded on RBM6 OCRE domain each with 120 ms mixing time.

For RRM1, a ^{15}N, ^{13}C labeled 500 μM protein sample was prepared in SEC buffer 2 with additional 10% D_2O. Standard experiments were used for backbone assignment of the protein (see above) but since the sample was not stable over longer durations, side-chain resonance assignment experiments were recorded only on RNA bound complex. For this, a RRM1:CU_9 RNA complex was made at 1:1.2 ratio and 3D HCCH-TOCSY with ^{13}C and ^{1}H evolution were recorded to connect backbone amide resonances to side-chain resonances . Additionally, ^{15}N-edited NOESY-HSQC experiment in 90%H_2O/10%D_2O; aromatic and aliphatic ^{13}C-edited NOESY-HSQC experiments in 100% D_2O were recorded on RRM1-CU_9 RNA complex each with 120 ms mixing time.

Additionally, to check if inter-molecular NOEs between the protein-RNA complex are observed, a 2D ω_1-filtered NOESY experiment was recorded on a sample where RNA is completely saturated (at protein: RNA ratio of 0.8:1) in 100% D_2O. After confirming the presence of inter-molecular NOEs, aliphatic and aromatic 3D ω_1-filtered edited ^{13}C NOESY experiments in 100% D_2O were also recorded. Furthermore, to see the dispersion of RNA signals in free versus bound form, 2D ^{1}H-^{1}H TOCSY spectra were recorded.

For RRM1-Zf1 and RRM1-Zf1-RRM2, standard backbone resonance assignment experiments (see above) were collected on a ^{15}N, ^{13}C labeled 500 μM protein sample in SEC buffer 1, supplemented with 10% D_2O. Additionally, the chemical shift assignments for Zf1 (ID:17387) and RRM2 (ID:18017) from the BMRB repository were used to assist in the assignment process, wherever necessary.

All spectra were processed in NMRPipe/Draw (Delaglio, Grzesiek et al. 1995) and sequential resonance assignment was done manually in CCPN analysis (Vranken, Boucher et al. 2005).

3.2.4. NMR structure calculation and validation

All NOESY cross-peaks were manually picked in ^{15}N- and ^{13}C-edited NOESY-HSQC experiments in CCPN analysis (Vranken, Boucher et al. 2005). The peak assignment and volume integration was done in an automated manner in CYANA 3.0 (Guntert 2004, Guntert and Buchner 2015). The dihedral angle restraints were predicted using TALOS+ (Shen,

Delaglio et al. 2009) and additionally given as input for the CYANA structure calculation. At this step, 20 structures were generated from CYANA, which were further subjected to water-refinement in ARIA1.2 (Linge, Habeck et al. 2003, Linge, Williams et al. 2003). An ensemble of 40 lowest energy structures were then generated out of which a bundle of 10 representative structures were selected based on Molprobity scores (Davis, Leaver-Fay et al. 2007, Chen, Arendall et al. 2010). The structures were further validated using iCing (Doreleijers, Vranken et al. 2012).

3.2.5. NMR relaxation experiments

To study the molecular tumbling of the RNA binding domains of RBM5 protein in free form, NMR data were recorded at 298 K for 240 μM wild-type RRM1-Zf1, 568 μM RRM1-Zf1 C191G mutant and 300 μM RRM1-Zf1-RRM2 C191G mutant on AVIII600 or AVIII800 Bruker NMR spectrometers in SEC buffer 1. The protein-RNA complexes for both RRM1-Zf1 and RRM1-Zf1-RRM2 were prepared as described in **section 3.2.2**. ^{15}N relaxation data of R_1 and $R_{1\rho}$ experiments were performed as described (Tjandra, Kuboniwa et al. 1995, Massi, Johnson et al. 2004). For relaxation data recorded on wild-type RRM1-Zf1, C191G mutant, RRM1-GGS-Zf1 mutant and RRM1-Zf1 C191G-RNA complex, R_1 data were measured with 10 different relaxation delays and two duplicate delays, 21.6/21.6, 86.4, 162, 248.4, 345.6, 518.4, 669.6, 885.6/885.6, 1144.8, 1382.4 ms and $R_{1\rho}$ data were determined by using 10 different delay points with two duplicate delays, 5/5, 10, 15, 20, 40, 80, 100/100, 130, 160, 180 ms. For relaxation data recorded on RRM1-Zf1-RRM2 C191G mutant in free form, R_1 data were measured with 10 different relaxation delays and two duplicate delays, 21.6/21.6, 86.4, 162/162, 432, 540, 675, 810, 1080, 1350, 1620 ms and $R_{1\rho}$ data were determined by using 12 different delay points with two duplicate delays, 5/5, 10, 15, 20, 30, 50, 75, 80, 100/100, 115, 130, 160 ms. For relaxation data recorded on RRM1-Zf1-RRM2 C191G mutant-GGCU_12/ne_GGCU_13 RNA complexes, R_1 data were measured with 11 different relaxation delays and one duplicate delay, 0, 80, 160, 240/240, 400, 560, 800, 960, 1200, 1440, 1600 ms and $R_{1\rho}$ data were determined by using 11 different delay points with one duplicate delay, 5, 10, 15, 20/20, 30, 40, 50, 60, 80, 100, 120 ms. Duplicate time points were used for error estimation. The transverse relaxation rate R_2 for each residue was estimated by correction of the observed relaxation rate $R_{1\rho}$ with the offset $\Delta\nu$ of the rf field to the resonance using the relation $R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta$, where $\theta = \tan^{-1}(\nu_1/\Delta\nu)$. The correlation time ($\tau_c$) of the protein molecule was then estimated using the ratio of averaged R_2/R_1 values (Daragan, Ilyina et al.

1997). All relaxation experiments were acquired as pseudo-3D experiments and converted to 2D data sets during processing in NMRPipe (Delaglio, Grzesiek et al. 1995). The relaxation rates and error determination were performed by using PINT (Ahlner, Carlsson et al. 2013). Cross-peaks with low intensity or extensive overlaps were removed from the data analysis.

3.2.6. Residual Dipolar Couplings (RDC)

For all RDC measurements, Otting medium containing C12E6-poly (ethylene glycol) and hexanol mixture at a molar ratio of 0.64 having a stability range from approximately 22 °C -32 °C (Rückert and Otting 2000) was used. A 6% PEG-hexanol alignment medium stock was prepared in a solution containing 450 µl SEC buffer 1 (for free protein) or SEC buffer 2 (for protein-RNA complex) and 50 µl D_2O by adding hexanol in steps of 1 µl to a final volume of ~12 µl (for free protein) and ~13 µl (for protein-RNA complex).The addition of hexanol was accompanied with continuous vortexing. Additional 0.3-0.4 µl of hexanol were added after mixing 80 µl of sample with 80 µl of PEG-hexanol alignment medium to yield a final concentration of 3 % PEG-hexanol alignment medium. To prevent the alignment medium from collapsing, all steps were carried out on a thermal-block maintained at 25 °C. Deuterium splitting was measured to check if alignment was achieved and stable alignment with ~12-13 Hz splitting was observed. The dipolar couplings were extracted from 2D in-phase–anti-phase (IPAP) HSQC experiments (Ottiger, Delaglio et al. 1998, Cordier, Rogowski et al. 1999) recorded under both isotropic and anisotropic conditions. The spectra were processed in NMRPipe (Delaglio, Grzesiek et al. 1995) and the splitting was extracted from peak positions in CCPN Analysis (Vranken, Boucher et al. 2005). Only residues forming secondary structure or involved in Zn^{2+} coordination (in case of Zf1) were used for further analysis. PALES software (Zweckstetter 2008) was used for the analysis of RDCs whereby the magnitude of alignment tensor (D_a) and rhombicity (R) were calculated using the principal components of traceless matrix (A_{xx}, A_{yy}, A_{zz})-given by PALES and the absolute value of RDC. In case of RRM1-Zf1 C191G mutant, the ^{1}H-^{15}N RDCs measured from PEG-hexanol alignment medium were used to validate the crystal structure of the wild-type RRM1-Zf1 protein. The Cornilescu *Q* factor (Cornilescu, Marquardt et al. 1998) was used to determine the quality of the fit of experimental versus back-calculated RDCs.

3.2.7. Small angle X-ray scattering (SAXS) experiments

All measurements for RRM1-Zf1 C191G with and without RNA were performed at 25 °C using the BioSAXS beamline BM29, using a 2D Pilatus detector, at the European

Synchrotron Radiation Facility (ESRF) in Grenoble. Fifteen frames with 1s exposure time per frame were recorded for each free protein and buffer sample, using an X-ray wavelength of λ= 0.9919 Å. Measurements were performed in flow mode where samples were pushed through the capillary at a constant flow rate to minimize radiation damage. Frames showing radiation damage were removed prior to data analysis. For protein-RNA complex, an HPLC column was coupled to the SAXS measurement, whereby the sample is injected on the column and the elution peak is automatically used for SAXS measurement.

For data collection and processing, dedicated beamline software BsxCuBE was used in an automated fashion. The one-dimensional scattering intensities of samples and buffers were expressed as a function of the modulus of the scattering vector $Q = (4\pi/\lambda)\sin\theta$ with 2θ being the scattering angle and λ the X-ray wavelength. After buffer subtraction, all the downstream processing was done with PRIMUS (Konarev, Volkov et al. 2003). R_g of all the samples were determined using the same program using Guinier approximation and from p(r) curves. For validation of the crystal structure, CRYSOL (Svergun, Barberato et al. 1995) was employed to fit the back-calculated scattering curves with the experimental SAXS curves.

All measurements for RRM1-Zf1-RRM2 C191G with and without RNA were performed at 5 °C using Rigaku BIOSAXS 1000 and primary data processing was done with Rigaku SAXSLab v 3.0.1r1. Eight frames with 900 s exposure time per frame were collected for free protein and protein-RNA complex. The protein-RNA complex was prepared using a size-exclusion column whereby the excess RNA eluted as a separate peak and the protein-RNA complex peak was pooled and concentrated. Data treatment was done as before, with PRIMUS software (Konarev, Volkov et al. 2003).

3.2.8. Crystallization of R1Zf1 protein

For crystallization of R1Zf1 protein, the protein was concentrated to 10 mg/ml in SEC buffer 1 and sparse matrix crystallization screens were set up at 25 °C and 4 °C. Crystals appeared within 3 days in a drop containing 0.1 M BICINE pH 9.0, 20 % PEG 6000 as very thin joint needles. The condition was optimized by screening various pH and PEG 6000 concentrations to obtain separate but thin needles in solution containing 0.1 M BICINE pH 8.5, 10 % PEG 6000. An additive screen was performed thereafter where needles were optimized to obtain thin plates in a variety of conditions. Finally, crystals with 10 % of 1 M Cesium chloride as additive were pursued further. Crystals were cryo-protected in a solution containing 0.1 M BICINE pH 8.5, 12% PEG 6000, 20% ethylene glycol and flash frozen in liquid nitrogen.

Several datasets for the crystals were collected at ID23-1 beamline capable of MAD measurements from 5 keV to 20 keV energy where anomalous diffraction on Zn^{2+} ion was employed as well as on ID23-2 which is fixed energy (14.20 keV, 0.873 Å) and suitable for data collection on small crystals at ESRF, Grenoble. Datasets from best diffracting crystals were then processed with XDS (Kabsch 2010) software package and the structure was solved by Auto-Rickshaw platform (Panjikar, Parthasarathy et al. 2005, Panjikar, Parthasarathy et al. 2009). The missing residues were built using Coot model building software (Emsley and Cowtan 2004) with multiple rounds of model building and refinement with Refmac software (Murshudov, Vagin et al. 1997) from CCP4 suite (Winn, Ballard et al. 2011).

3.2.9. Static light scattering

All measurements were made with a Malvern Viscotek instrument (TDA 305) connected to an Äkta purifier equipped with an analytical size-exclusion column (Superdex 75 10/300 GL, GE Healthcare). A sample volume of 100 µl containing about 2-4 mg/ml of protein/protein-RNA complex was injected for each run. The SEC buffer 1 was used for free protein runs while SEC buffer 2 was used for protein-RNA complex measurements. Elution profiles were collected for 30 min with a flow rate of 0.5 ml/min and data were collected using absorbance UV detection at 280 nm, right-angle light scattering (RALS) and refractive index (RI). The molecular weights of separated elution peaks were calculated using OmniSEC software (Malvern). As a calibration standard, 4 mg/ml bovine serum albumin was used before all experiments.

3.2.10. Thermofluor assay

The thermofluor assay was performed to assess the stability of the proteins using an Mx2005p qPCR (Agilent) machine. The assay was performed in a high throughout fashion, where 5 µl protein-dye (SYPRO orange) master mix was added to each well in the 96-well plate containing different buffers to be tested for stability of the protein. The master mix was prepared such that the final concentration of the protein and dye in the well was 0.1 mg/ml and 20x, respectively. The melting temperature of the protein in these different conditions was then measured, to report stability of the protein. The SYPRO orange fluorescence was measured as a function of temperature gradient from 25 °C-96 °C. The data was analyzed using the standard pre-installed qPCR software MxPro.

3.2.11. Circular Dichroism (CD) spectroscopy

All CD spectra were recorded on a JASCO-J715 spectropolarimeter and analyzed with Spectramanager version 1.53.00 (Jasco Corp.) with temperature regulation using a Peltier type control system (PTC-348WI). The spectra were recorded at 0.3 mM concentration in OCRE NMR buffer, from 190–260 nm wavelength with a 1.0 nm bandwidth, 0.5 nm pitch at a scan speed of 50 nm/min (20 scans), in cuvettes with 0.1 cm path length, at 5 °C. The spectra were plotted as mean residue ellipticity (deg cm^2/dmol) vs wavelength (nm) after buffer subtraction.

3.2.12. Isothemal Titration Calorimetry (ITC)

All ITC experiments were performed either with MicroCal ITC200 or PEAQ-ITC calorimeters from Malvern. Prior to recording data, the protein samples were dialyzed overnight in the OCRE SEC buffer in case of RBM5 OCRE domain and SEC buffer 2 in all other cases. The 1 mM DTT in buffer was replaced with 2 mM BME for all measurements. The cell was filled completely with 10-30 µM protein and depending on the affinity, the syringe was filled with different concentrations of the respective ligand. A series of 26 injections of 1.5 µl titrant or 39 injections of 1 µl were made into the protein. The data were processed with either the Origin software provided with ITC200 or with PEAQ-ITC Analysis software in case of PEAQ-ITC calorimeter. The data were fit to a one-binding site model.

Chapter 4: Structural and functional insights into RBM5/6/10 OCRE domains

4.1. Characterization of RBM5 OCRE-SmN/B/B' complex

RBM5 OCRE domain recognizes poly-proline rich sequences in the C-terminal tails of SmN/B/B' proteins which contain not just one but a number of poly-proline rich motifs (PRM) arranged tandemly. Each PRM contains three or four consecutive proline residues, flanked by arginine residues at ±3 residues (**Figure 19A**). The solution NMR structure of RBM5 OCRE domain in complex with a single PRM (GMRPPPPGIRG-residue s219-229) from SmN C-terminal tail was solved by a previous doctoral student in the lab (PDB ID: 5MF9) (**Figure 19B**).

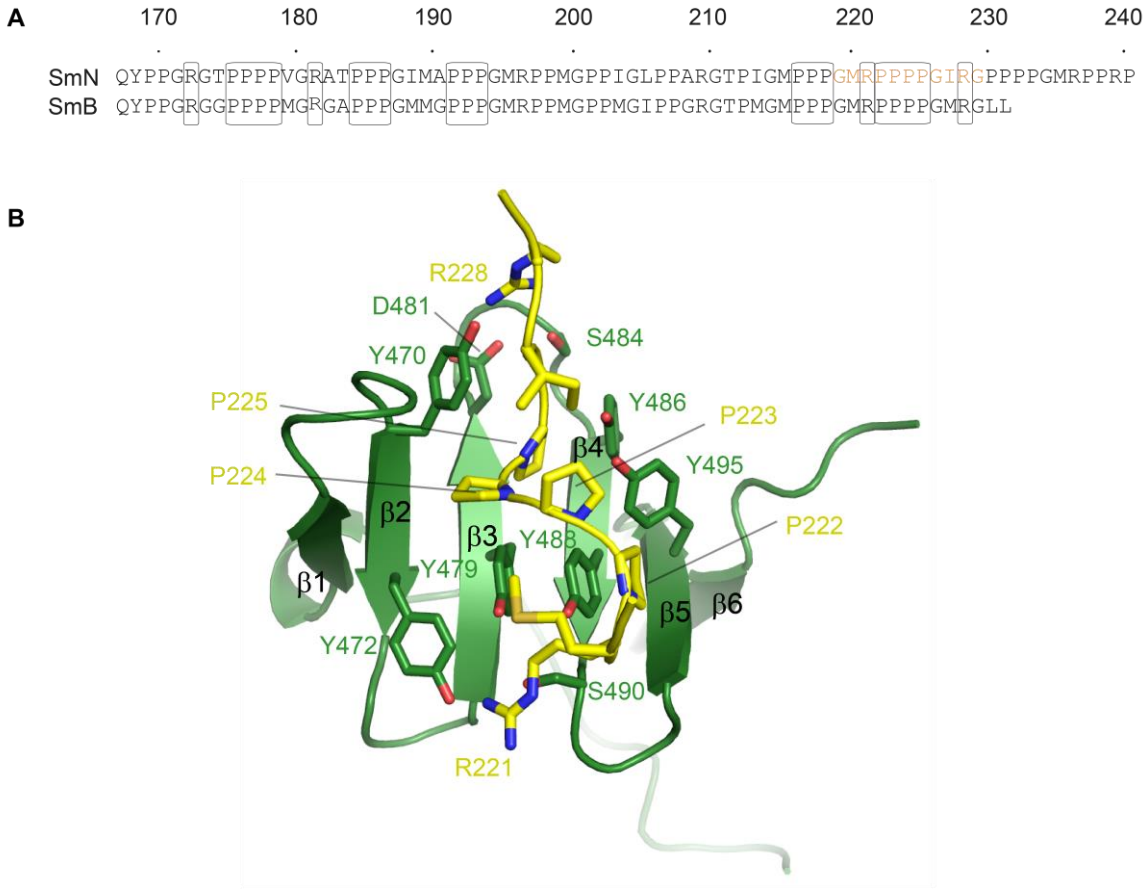


Figure 19 Structure of RBM5 OCRE domain-PRM complex

(A) Sequence alignment of SmN/B poly-proline rich C-terminal tails is shown, with the proline rich motifs and arginine residues at ±3 position highlighted with boxes. The PRM used for structure calculation of RBM5 OCRE domain-PRM complex is marked in orange (residue s219-229). (B) The NMR solution structure of the complex (PDB ID: 5MF9) is shown with RBM5 OCRE domain in green and the PRM in yellow. Important residues involved in intermolecular interactions are labeled.

The structure of the complex clearly shows how the highly aromatic interface of the RBM5 OCRE domain recognizes the PRM. A detailed description of the specific contacts

between the OCRE domain and SmN ligand can be found in the following publication (Mourao, Bonnal et al. 2016).To gain insights into the sequence specific requirements of PRMs for complex formation with RBM5 OCRE domain, I used a combination of ITC and CD and NMR spectroscopy.

4.1.1. Sequence specific requirements of PRMs for RBM5 OCRE binding

I carefully designed several constructs of the SmN/B/B' proteins to test the effect of point mutations in PRMs on binding affinity. A comparison of ITC data between the wild-type and SmN/B sequence variants is shown in **Figure 20**.

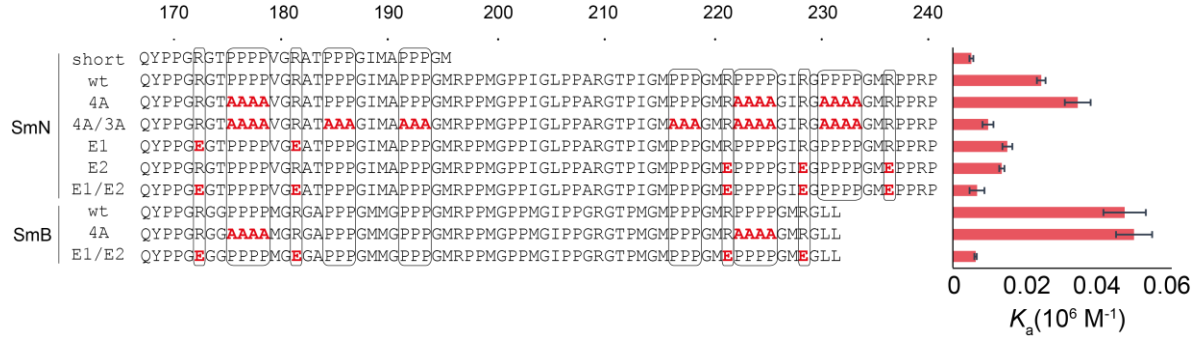


Figure 20 ITC data to probe sequence specific requirements of PRMs for RBM5 OCRE binding

A comparison of affinities of wild-type and mutant PRMs are presented. The point mutations are shown in red and the conserved poly-proline stretches and arginine residues are highlighted in boxes. Note that association constants (K_a) are shown.

The short SmN tail (residues 167-196, denoted as short) containing one four proline PRM and two three proline PRMs has ~195 µM binding affinity as compared to ~41 µM of the longer SmN tail (residues 167-240, denoted as wt) for RBM5 OCRE domain. This clearly indicates that multiple PRMs contribute to the binding affinity as opposed to a single PRM and the overall binding affinity increases with increasing number of PRMs, consistent with avidity effects. Next, the flanking arginine residues on either side of the PRMs were mutated sequentially to include mutations from multiple PRMs (R1->E1, R2->E2 and R1/R2->E1/E2). With increasing severity of the mutations, the binding affinity decreased successively (67 µM, 74 µM and 150 µM, respectively) indicating that the conserved flanking arginine residues significantly contribute to binding.

Finally, to determine if the proline residues are an absolute requirement for binding of PRMs to RBM5 OCRE domain, a set of SmN/B sequence variants were created where either all four-proline PRMs or all four- and three-proline PRMs were mutated to alanine residues

(4P->4A and 4P/3P->4A/3A, respectively). Surprisingly, the affinity of 4P->4A mutants did not decrease while that of the more stringent 4P/3P->4A/3A decreased by ~3 fold. To understand how RBM5 OCRE domain could still bind to the 4A/3A mutant, I used CD spectroscopy. Since RBM5 OCRE domain recognizes poly-proline type II helix, I used CD spectroscopy (**Figure 21**) to check if the conformation is maintained in the 4P->4A mutant. For this purpose, I designed short peptides (10-12 residues) containing 4P and 3P PRMs with and without alanine mutations.

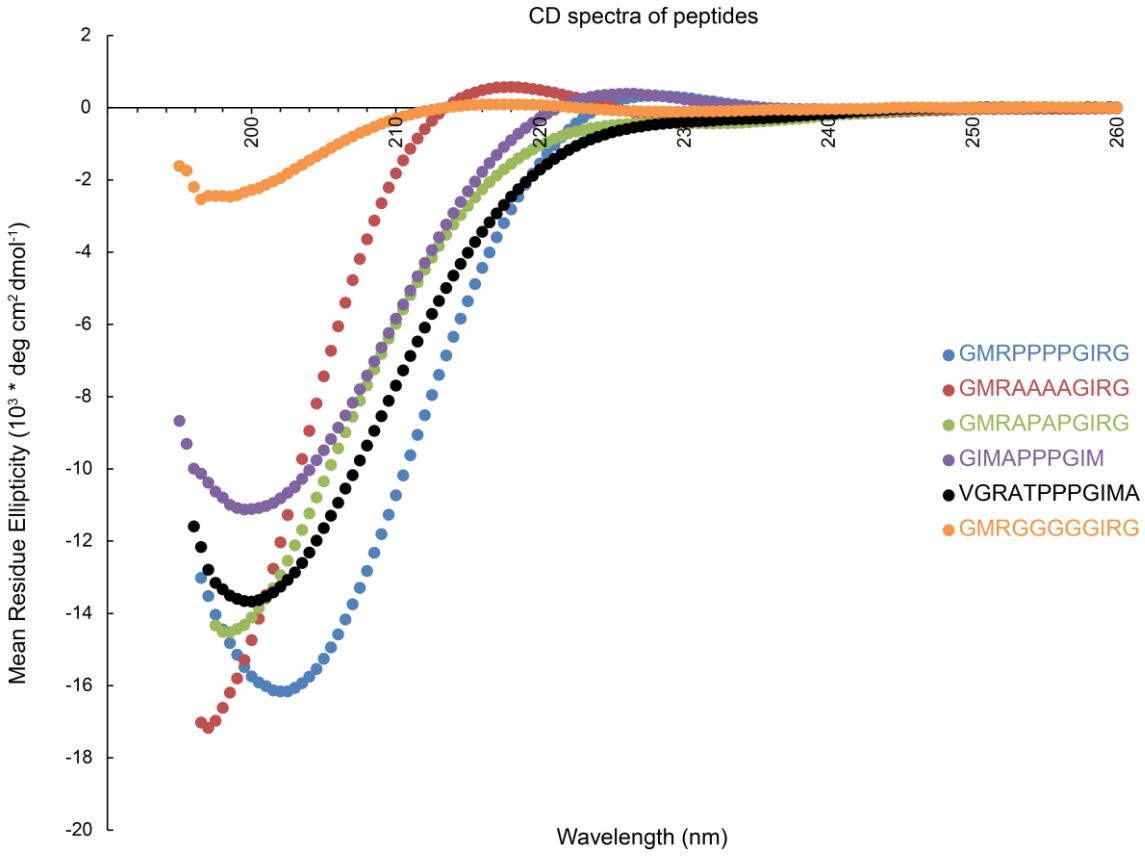


Figure 21 CD spectra of short SmN/B derived peptides

CD spectra of wild-type (blue), 4A (red) and GIMA (purple) peptides show the presence of poly-PPII helical content, while APAP (green), VGRA (black) and 4G (orange) peptides show residual PPII helical content, if any.

In CD spectra, the presence of a strong negative band at 200 nm and a weak positive band at 217 nm are characteristic features of a PPII helix (Drake, Siligardi et al. 1988). With increasing number of proline residues, the positive band in the CD spectrum shifts towards 229 nm (Petrella, Machesky et al. 1996). Therefore, wildtype (blue), 4A mutant peptide (red) clearly show PPII helical conformation. This explains the ITC data where 4P->4A mutation in the C-terminal SmN/B tails did not affect the binding affinity, due to the overall structure of

the peptide still being maintained. Next, I tested the effect of making 4P->APAP mutation (green) on the secondary structure of the peptide. The CD spectrum lacks the characteristic positive band at 217/229 nm indicating that APAP mutation is not tolerated and it breaks the PPII helix. Another mutant 4P->4G (orange) was tested, which showed only residual PPII helical content consistent with previous reports (Kelly, Chellgren et al. 2001, Brown and Zondlo 2012). Two 3P PRM peptides were also tested for their PPII helical propensity-VGRA peptide (black) lacks PPII helical content indicated by absence of a positive band at around 229 nm, while GIMA (purple) peptide still retains its secondary structure, although the negative band at 199.5 nm is less intense as compared to the wildtype peptide . This could be attributed to the presence of an alanine residue before the 3P PRM (GIM**APPP**GIM).

4.1.2. NMR investigations of relative binding affinities of SmN variant peptides

To learn about the contribution of individual residues in the PRM motifs towards affinity, a set of 11-mer peptides with point mutations were used. Single PRMs have very low binding affinity (in the mM range), beyond the detection limit of ITC. Therefore, a semi-quantitative NMR chemical shift perturbation based scoring scheme was designed. For this, single-point NMR titrations of the wild-type and mutant peptides into RBM5 OCRE domain were made and the CSPs from seven most shifting residues were added for each of the peptides and normalized with that of the wild-type peptide. This score was then used as an indirect measure of the binding affinity of the various peptides. ^{15}N-HSQC NMR spectra for three of the peptide titrations into OCRE domain with zoom-ins of the specific residues used for the CSP score calculation are shown in **Figure 22**.

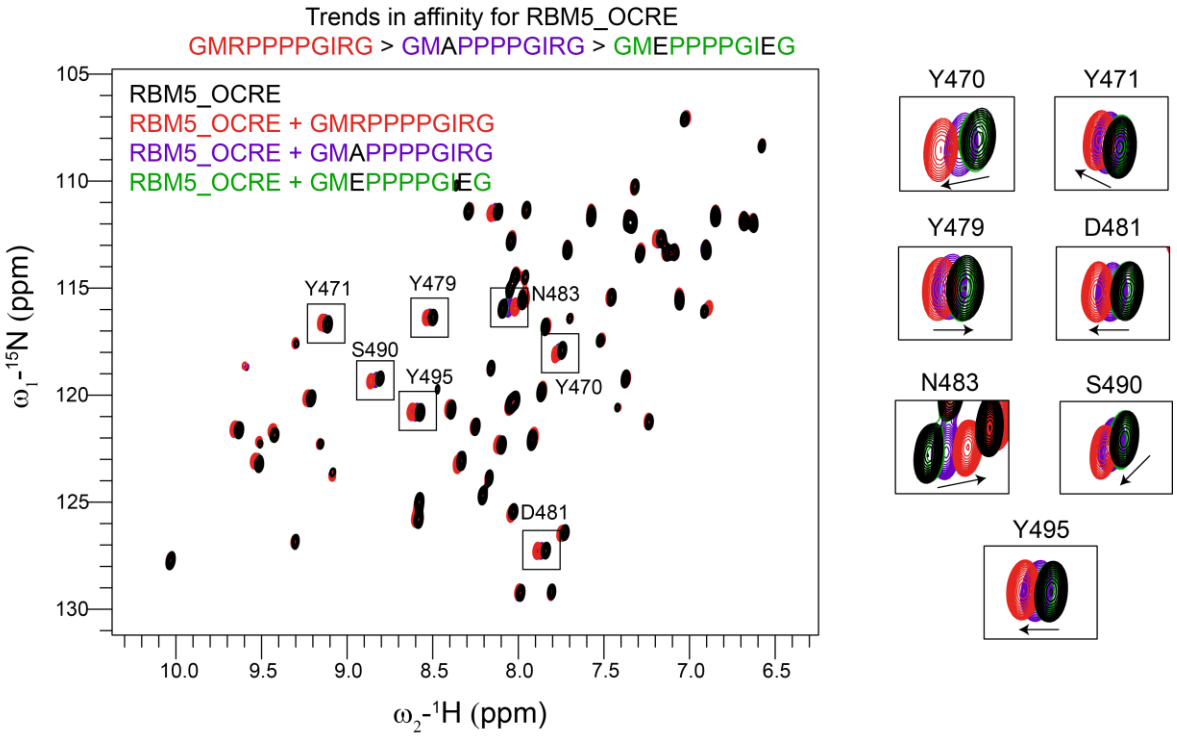


Figure 22 NMR spectra showing residues used for NMR based CSP score calculation

Overlay of 15N-HSQC spectra of free RBM5 OCRE domain (black), and bound to wild-type (GMRPPPPGIRG) and two mutant peptides (GMAPPPPGIRG and GMEPPPPGIEG) in red, purple and green, respectively. Residues used for CSP score calculation are shown as zoom-ins on the right.

The normalized CSP score is shown in **Figure 23**, where higher affinity is represented by a higher score. The comparison suggests that RBM5 OCRE domain has a clear preference for its binding motif, although it can tolerate certain variations. Firstly, a 4P PRM is preferred over 3P PRM peptide as the two 3P PRM peptides (GIMA and VGRA) have a 2-3 fold reduced affinity compared to wild-type peptide. Secondly, the APAP peptide which breaks the PPII helical conformation also leads to a significant decrease in affinity. Thirdly, both flanking arginine residues are important for binding which can be seen by the strong decrease in affinity with charge reversal mutations (compare CSP score of wild-type peptide GMRPPPPGIRG with GMEPPPPGIEG, GMEPPPPGIRG and GMRPPPPGIEG). On the other hand, mutations of these arginine residues to alanine residues have smaller effects possibly as no charge clashes were introduced on the highly negatively charged surface of RBM5 OCRE domain. It is also clear that the first arginine is more important for binding than the second.

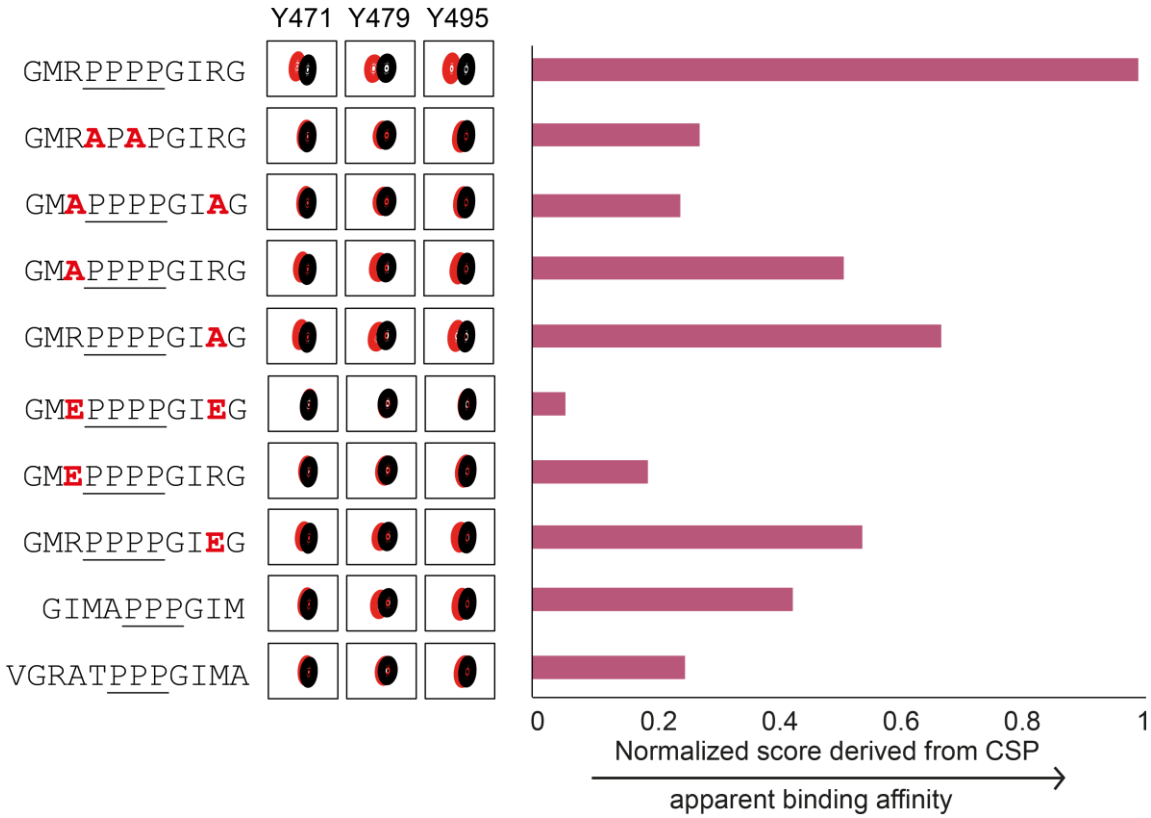


Figure 23 NMR based Normalized CSP score

The CSP score derived from 7 representative residues is shown on the right. The score of the wild-type peptide is taken to be highest and score of the rest of the peptides is derived by normalization to that of the wild-type peptide. The higher the bar, the higher is the binding affinity. The point mutations are shown in red, and the PRM is underlined.

To verify if the specific requirements of arginine residues at ±3 position of the proline-rich repeats can be explained structurally, I went back to look at the solution structure of RBM5 OCRE-PRM complex. Indeed, both the arginine residues (Arg 221 and Arg 228) are involved in specific inter-molecular interactions. The side-chain of Arg 221 makes hydrogen bonds with hydroxyl groups of Tyr 472, Tyr 479 and Ser 490 as shown in **Figure 24**. On the other hand, the side-chain of Arg 228 can also potentially form hydrogen bonds with side-chains of Ser 484 and Asp 481, providing an additional layer of specificity to the PRM recognition by RBM5 OCRE.

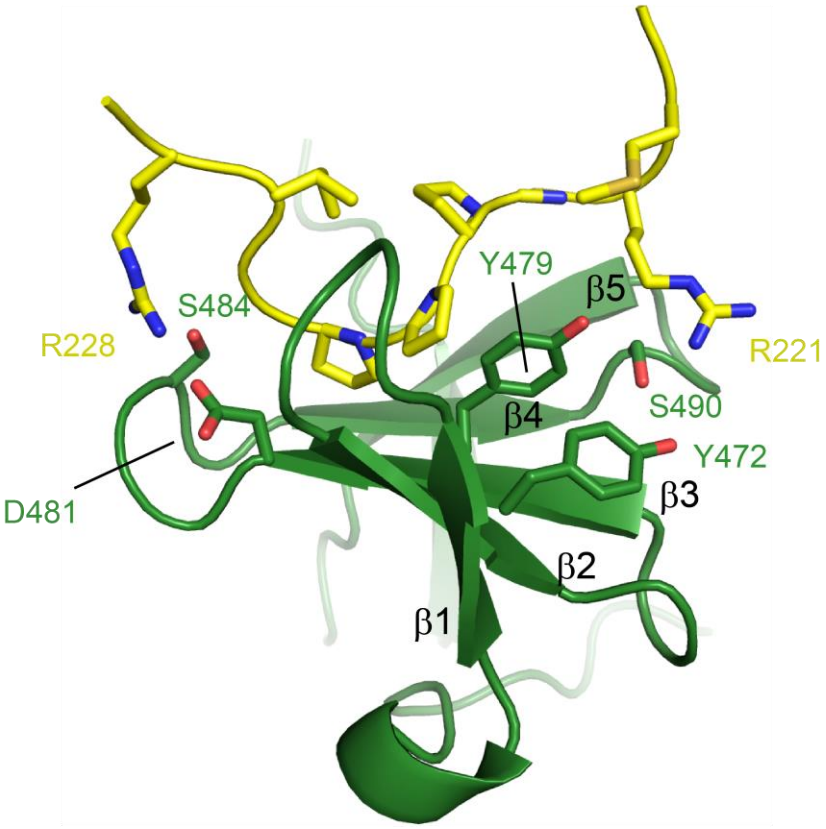


Figure 24 Structure of RBM5 OCRE-PRM complex:Importance of flanking arginine residues

The solution NMR structure of RBM5 OCRE domain-PRM complex clearly demonstrates the importance of the arginine residues flanking the poly-proline stretch. RBM5 OCRE domain residues forming specific contacts with the arginine residues of PRM are labeled.

It was therefore concluded that RBM5 OCRE domain recognizes a PPII helical conformation formed by four consecutive prolines and flanked by positively charged residues on either side of the PRM.

4.2. Characterization of RBM10/6 OCRE domains

The sequence alignment of RBM5/6/10 OCRE domains suggests that RBM10 OCRE domain has all the tyrosine residues required to form the 6-β strands as observed in the RBM5 OCRE domain. On the other hand, RBM6 OCRE domain seems to be truncated and containing the tyrosine repeats enough for formation of only 4-β strands (**Figure 25**).

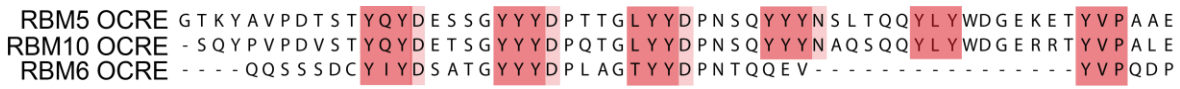


Figure 25 Sequence alignment of RBM5/6/10 OCRE domains

The tyrosine repeat regions possibly forming the β-strands are highlighted in red and the conserved negatively charged residue in pink.

The ^{1}H-^{15}N HSQC spectra of RBM10/6 OCRE domains show good dispersion as would be expected for an all β-strand protein (**Figure 26**).

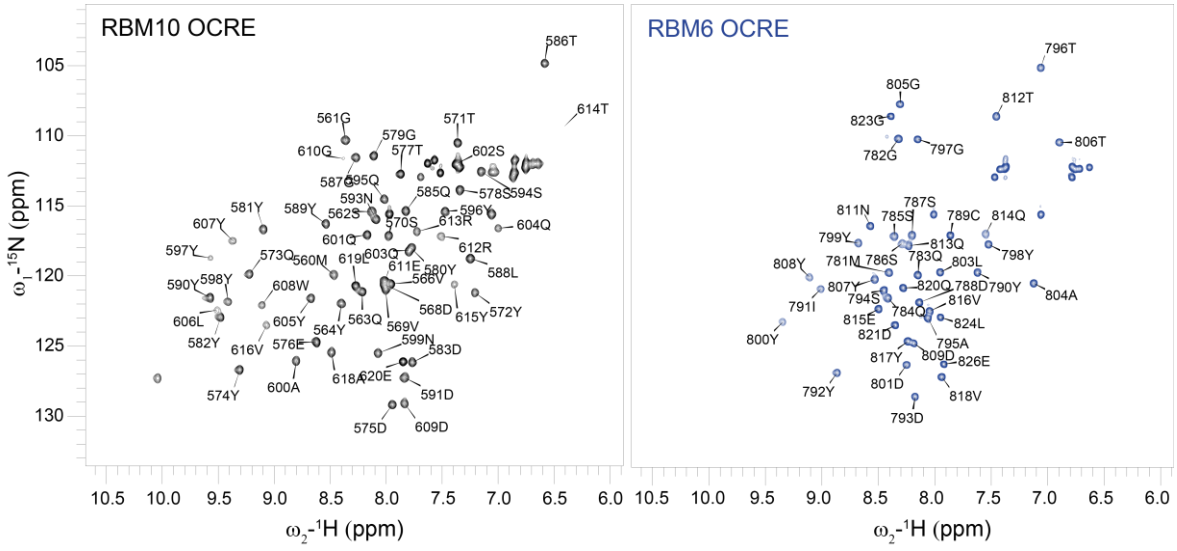


Figure 26 ^{1}H,^{15}N HSQC spectra of RBM10/6 OCRE domains

To understand if the differences in the sequence translate into structural differences, I calculated the solution NMR structures of RBM10/6 OCRE domains.

4.2.1. Solution NMR structures of RBM10/6 OCRE domains

For determining the solution NMR structures of RBM10/6 OCRE domains, the proteins were expressed in ^{13}C, ^{15}N labelled M9 minimal medium with cleavable His-tags in *E.coli* BL21 (DE3) cells. The proteins were purified as described in the Methods section. Next,

various NMR backbone and side-chain assignment experiments were recorded to achieve 97.2% and 96.8% assignment completeness for RBM10 and RBM6 OCRE domains, respectively. Finally, a set of 3D-NOESY experiments (^{15}N-edited NOESY, ^{13}C-edited NOESY for aliphatic and aromatic regions) were recorded for both proteins, to provide information on the short and long range NOEs, which were finally used as input for automatic assignment and structure calculation using CYANA3.0. In case of RBM10 OCRE domain, most of the experiments required for structure calculation were collected by Dr. André Mourão.

Additionally, TALOS+ derived torsion angle restraints were provided as input for structure calculation. The ensemble of 20 structures obtained converged well with a low RMSD between them. The overall quality of the structures improved with each cycle in the CYANA run, which was judged by the decrease in the value of the target function. A final water refinement was done in ARIA after which 10 lowest energy structures were selected. The structure statistics are shown in **Table 1** and **Table 2** for RBM10 and RBM6 OCRE domains, respectively.

Table 1 Structure statistics for RBM10 OCRE domain

Structure calculation restraints	
Distance restraints	
Total NOEs	1468
Sequential (|i- j| <= 1)	671
Medium-range (|i- j| < 5)	255
Long-range (|i- j| >= 5)	542
Dihedral restraints (ϕ+ψ)	99
Quality analysis	
Restraints violations (mean ± s.d)	
Distance restraints (Å)	0.013 ± 0.007
Dihedral angle restraints (º)	0.139 ± 0.06
Deviation from idealized geometry	
Bond length (Å)	0.002 ± 0.000
Bond angles (º)	0.284 ± 0.046
Improper dihedral distribution (º)	0.176 ± 0.049
Average pairwise r.m.s. deviation (Å)[a]	
Heavy	0.39 ± 0.05
Backbone	0.06 ± 0.03
Ramachandran values (%)[a,b,c]	
Most favored regions	90.4
Allowed regions	9.6
Generously allowed regions	0
Disallowed regions	0
WhatIf analysis[a,c]	
First generation packing	2.618 ± 1.442
Second generation packing	5.336 ± 2.777
Ramachandran plot appearance	-2.501 ± 0.388
Chi-1/Chi-2 rotamer normality	-1.666 ± 0.786
Backbone conformation	0.178 ± 0.279

[a] For residues 563-619, [b] With Procheck., [c] Analyzed by iCING. Structure Z-scores, a positive number is better than average.

Table 2 Structure statistics for RBM6 OCRE domain

Structure calculation restraints	
Distance restraints	
Total NOEs	568
Sequential (|i- j| <= 1)	318
Medium-range (|i- j| < 5)	77
Long-range (|i- j| >= 5)	173
Dihedral restraints (ϕ+ψ)	60
Quality analysis	
Restraints violations (mean ± s.d.)	
Distance restraints (Å)	0.031 ± 0.01
Dihedral angle restraints (º)	0.074 ± 0.04
Deviation from idealized geometry	
Bond length (Å)	0.001± 0.000
Bond angles (º)	0.245 ± 0.009
Improper dihedral distribution (º)	0.125 ± 0.01
Average pairwise r.m.s. deviation (Å)[a]	
Heavy	0.52 ± 0.06
Backbone	0.25 ± 0.09
Ramachandran values (%)[a,b,c]	
Most favored regions	82.4
Allowed regions	17.6
Generously allowed regions	0
Disallowed regions	0
WhatIf analysis[a,c]	
First generation packing	0.247 ± 1.653
Second generation packing	4.120 ± 2.295
Ramachandran plot appearance	-4.123 ± 0.428
Chi-1/Chi-2 rotamer normality	-0.767 ± 0.763
Backbone conformation	-0.231 ± 0.594

[a] For residues 788-820, [b] With Procheck., [c] Analyzed by iCING. Structure Z-scores, a positive number is better than average.

The final ensemble of 10 lowest energy structures of RBM10 OCRE domain is shown in **Figure 27** and that of RBM6 OCRE domain are shown in **Figure 28.** The N- and C- termini of both RBM10 and 6 OCRE domains are flexible while the rest of the protein is quite rigid as indicated by a backbone RMSD of 0.06 Å in case of RBM10 OCRE domain and 0.25 Å in case of RBM6 OCRE domain.

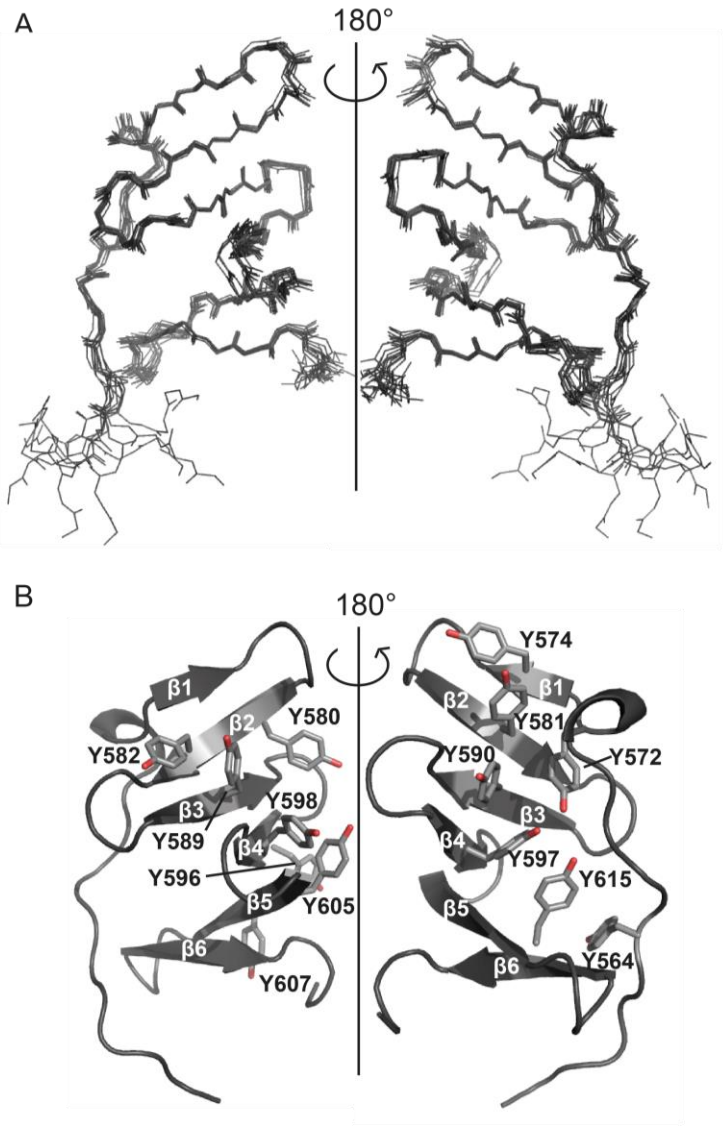


Figure 27 Solution NMR structure of RBM10 OCRE domain

An ensemble of 10 lowest energy structures of RBM10 OCRE domain are shown in panel (A). A single representative structure is shown in panel (B) with the exposed tyrosine residues on each either side of the protein marked.

As expected, the RBM10 OCRE domain is structurally quite similar to that of RBM5 OCRE domain, consisting of 6 anti-parallel β-strands with the N-terminal loop packing against one side of the domain, partially shielding it from the solvent. On the other hand, RBM6 OCRE domain is a truncated OCRE consisting of only 4 anti-parallel β-strands. Both proteins have a series of surface exposed tyrosine residues on either side of the proteins.

In RBM10 OCRE domain, Tyr 580 (β2), Tyr 582 (β2), Tyr 589 (β3), Tyr 598 (β4) and Tyr 605 (β5) form an extended aromatic interface with surface exposed tyrosine hydroxyl

groups (**Figure 27**). On the other side, Tyr 574 (β1), Tyr 581 (β2), Tyr 590 (β3) and Tyr 57 (β4) also form an extensive aromatic interface with residues Tyr 572 and Pro 567 from the N-terminal loop forming an additional hydrophobic cluster along with Trp 608 (β5), Tyr 615 (β6) and Tyr 597 (β4). This hydrophobic cluster also brings the N-terminal loop close to the C-terminus of the protein.

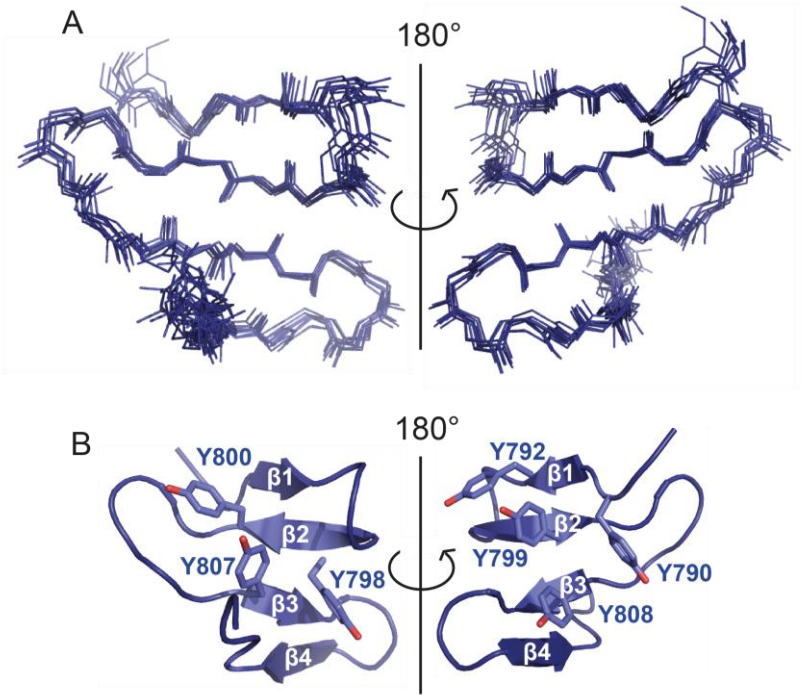


Figure 28 Solution NMR structure of RBM6 OCRE domain

An ensemble of 10 lowest energy structures of RBM6 OCRE domain are shown in panel (A). A single representative structure is shown in panel (B) with the exposed tyrosine residues on each either side of the protein marked.

In RBM6 OCRE domain, Tyr 800 (β2), Tyr 798 (β2) and Tyr 807 (β3) are solvent exposed on one side of the domain, forming a short aromatic interface while Tyr 792 (β1), Tyr 799 (β2), Tyr 807 (β3) along with Tyr 790 from the N-terminus form an extensive network of aromatic interactions (**Figure 28**).

The twisted β-sheet appearance appears to be unique and conserved among the OCRE domains although it is much less apparent in RBM6 OCRE domain due to lack of two β-strands. The electrostatic surface of the proteins is predominantly negatively charged making it a good candidate for protein-protein interactions rather than nucleic acid binding.

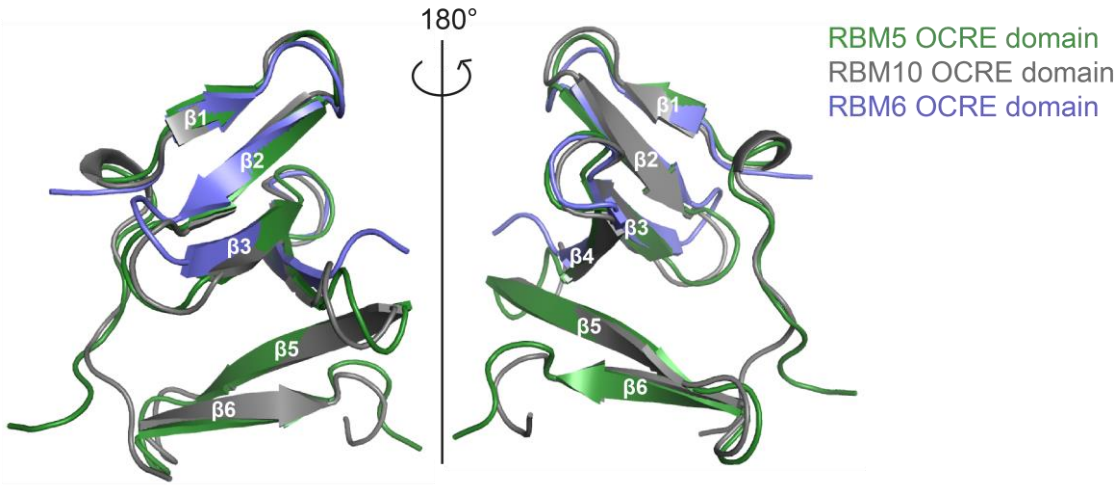


Figure 29 Superposition of RBM5/6/10 OCRE domains

Superposition of RBM5/6/10 OCRE domains are shown in green, grey and purple, respectively. RBM5/10 OCRE domains have β1-6 while RBM6 OCRE domain only has β1-4.

A structural superposition of the three RBM5/6/10 OCRE domains indicates a good agreement between the three domains with a backbone RMSD of 0.527 Å between RBM5/10 OCRE domains and 0.983 Å between RBM5/6 OCRE domains considering only the structured regions (residues 454-509, 563-619, 788-820 for RBM5/6/10 OCRE domains, respectively). The RMSD values are calculated using SuperPose v1.0 webserver (Maiti, Van Domselaar et al. 2004). It becomes clear that RBM6 OCRE domain has the first four β-strands as its counterparts but the last two β-strands, β5 and β6 are absent (**Figure 29**). Interestingly, the hydrophobic cluster bringing the N- terminal loop close to the C-terminus involving residues from β5 and β6 (Trp 608 and Tyr 615 in RBM10 OCRE domain and the corresponding Trp 498 and Tyr 505 in RBM5 OCRE domain) is absent in RBM6 OCRE domain. This could be the reason why the N-terminal loop in this case does not wrap around the domain, but is pushed away towards the solvent.

4.2.2. Binding studies of RBM10/6 OCRE domains

Next, I wanted to examine if RBM10/6 OCRE domains have similar binding characteristics as that between RBM5 OCRE and SmN/B/B' C-terminal tails. Using ITC, SmN C-terminal tails were titrated into RBM10/6 OCRE domains in separate experiments. Since RBM10 OCRE domain is structurally quite similar to that of RBM5, I expected that SmN ligand binding should be conserved between the domains as well. Consistently, RBM10 OCRE domain has ~24 μM binding affinity for SmN C-terminal tail while RBM6 OCRE domain

showed no binding at all to SmN C-terminal tail (performed by Dr. André Mourão). Since ITC has a detection limit and it is possible that RBM6 OCRE domain still binds to SmN tail, just with very low affinity, I used NMR to provide further insights into this.

SmN C-terminal tail derived peptide GMRPPPPGIRG containing 4P PRM was titrated at 1:10 ratio into RBM10/6 OCRE domains. As expected, RBM10 OCRE domain showed chemical shift perturbations while RBM6 OCRE domain did not show any binding (**Figure 30**).

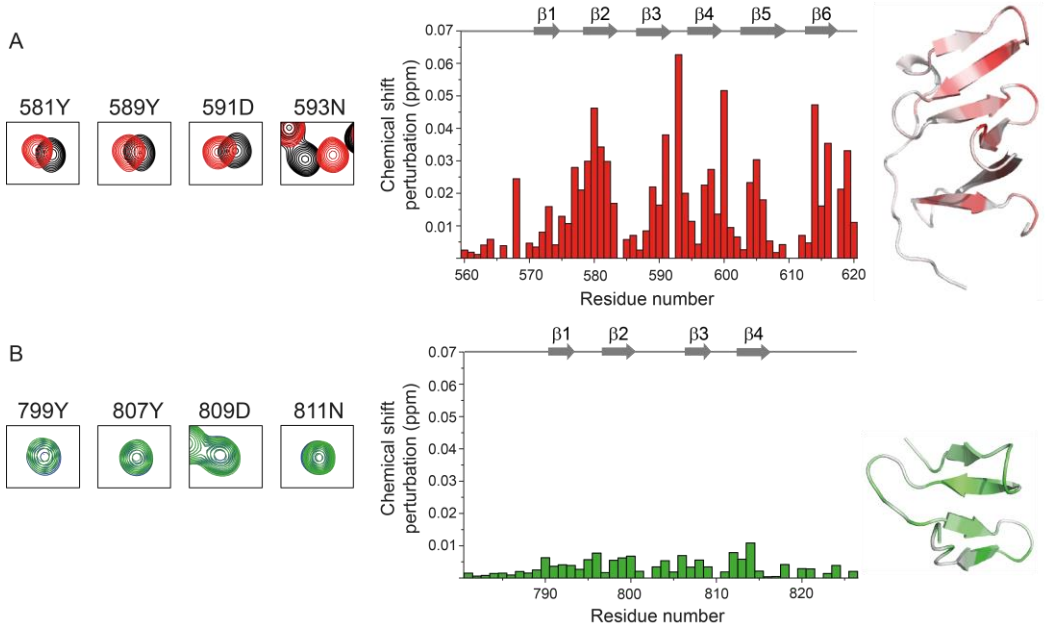


Figure 30 NMR binding characterization of SmN ligand with RBM10/6 OCRE domains

NMR titration analysis of SmN derived 4P PRM peptide GMRPPPPGIRG with RBM10/6 OCRE domains in panel (A) and (B), respectively. Zoom-ins of ^{15}N-HSQC spectra to show four most shifting residues for RBM10 OCRE domain upon PRM binding, and the corresponding residues in RBM6 OCRE domain are presented on the left. The chemical shift perturbation is plotted onto the respective structures, as shown on the right.

It is clear that RBM10 OCRE domain can recognize SmN C-terminal tail in a similar manner as that of RBM5, due to structural conservation between the domains. On the other hand, RBM6 OCRE domain lacking the last two β-strands (β5 and β6) cannot bind to SmN derived ligands (4P PRM or C-terminal tail).

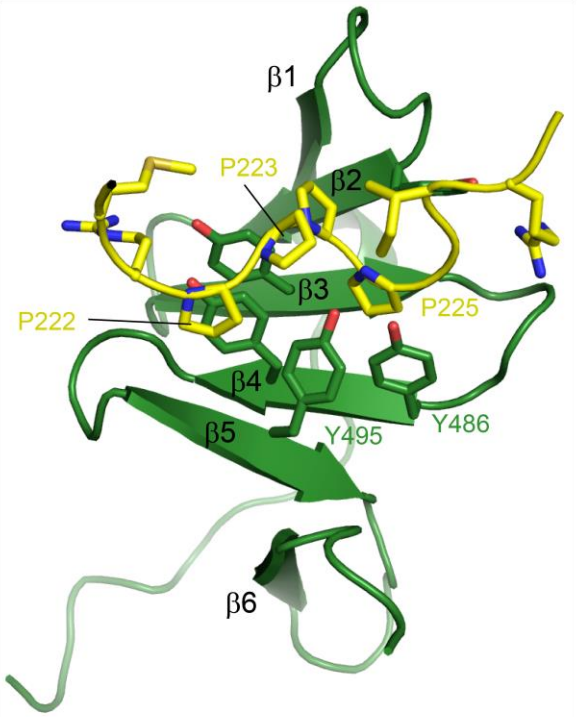


Figure 31 Contribution of β5 strand in PRM recognition by RBM5 OCRE domain

Importance of β5 strand of RBM5 OCRE domain in PRM recognition is illustrated by the specific interactions between Tyr 495, Tyr 486 of the OCRE domain and Pro 223 and Pro 225 of the SmN PRM ligand.

To rationalize this, I looked at the solution NMR structure of RBM5 OCRE domain in complex with the 4P PRM. As illustrated in **Figure 31**, side-chain of Tyr 495 belonging to β5 strand of the OCRE domain makes contacts with Pro 223 of the SmN PRM while additionally stacking with Tyr 486 side-chain (β4), which further stacks with Pro 225 of the SmN PRM. Therefore, Tyr 495 (β5) is involved in a network of stacking interactions which might be de-stabilized in its absence as in the case of RBM6 OCRE domain. Additionally, β4 strand in RBM6 OCRE domain is also shorter and would be unable to provide stacking interactions with Pro 222.

4.2.3. Alternative splicing regulation of *Fas* pre-mRNA by RBM10/6

Previously, it has been shown that RBM5 OCRE domain is important for alternative splicing regulation of *Fas* pre-mRNA (Bonnal, Martinez et al. 2008). Due to structural conservation between RBM5/10 OCRE domains, we wanted to test if RBM10 OCRE domain can also regulate *Fas* pre-mRNA alternative splicing as its RBM5 counterpart. Our

collaborators (Dr. Sophie Bonnal in Dr. Juan Valcárcel's group in Barcelona, Spain) therefore performed *Fas* pre-mRNA *in vivo* splicing assays (**Figure 32**). As expected, RBM10 OCRE domain is also required for exon 6 skipping in the *Fas* minigene reporter, re-iterating the fact that RBM5/10 OCRE domains are structurally as well as functionally conserved (**Figure 32A**).

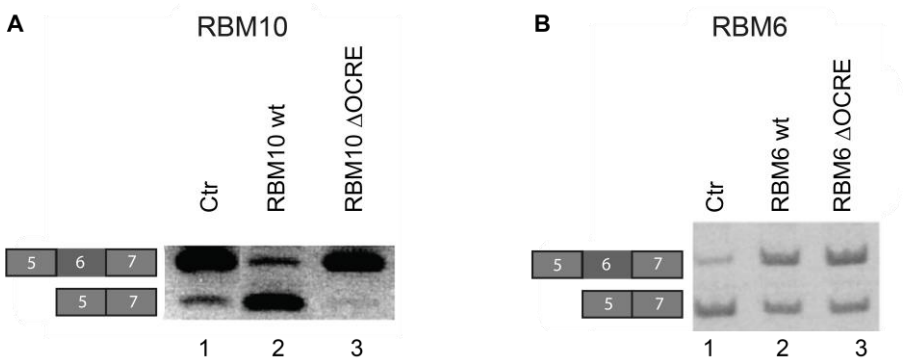


Figure 32 *Fas* pre-mRNA *in vivo* splicing assays

(A) The OCRE domain of RBM10 is required for *Fas* exon 6 skipping from a minigene reporter (compare lanes 2 and 3). (B) The OCRE domain of RBM6 is not required for the function of the protein in the regulation of *Fas* exon 6 inclusion from a minigene reporter (compare lanes 2 and 3). *Fas* genomic sequences between the 5′ end of exon 5 and 47 nucleotides downstream of the 5′ splice site of exon 7 were cloned in an expression vector and transfected into HeLa cells together with either T7-RBM10 wt, T7-RBM10 with OCRE deletion, T7-RBM6 wt, T7-RBM6 with OCRE deletion expression plasmids or T7-ADAR as control. RNA was isolated 24 hours after transfection and analyzed by RT-PCR using vector-specific sequences.

Next, we wanted to probe the requirement of RBM6 OCRE domain on *Fas* pre-mRNA splicing. Over-expression of RBM6 protein promotes *Fas* exon 6 inclusion in the *in vivo* splicing assays (**Figure 32B**). This is contrasting to that of RBM5 and 10 where the over-expression of the respective proteins promotes exon 6 skipping. Further, upon deletion of RBM6 OCRE domain from the full-length protein, the effect of exon 6 inclusion is still maintained. This clearly indicates that another domain, apart from OCRE domain, promotes *Fas* exon 6 inclusion. Since the C-terminus of RBM6 protein contains various other domains which might be involved in protein-protein interactions, including Zf2 and G-patch, it is quite plausible that one of these domains mediates the alternative splicing regulation of *Fas* pre-mRNA. Preliminary data indicate the requirement of Zf2 domain of RBM6 in *Fas* exon 6 inclusion (Dr. Sophie Bonnal, personal communication).

Although RBM5/6/10 have a high degree of sequence similarity and conservation in domain organization, there are indications that they have distinct functions. It has been shown that these RBM proteins have little overlap between the target genes regulated by them (Bonnal,

Martinez et al. 2008, Bechara, Sebestyen et al. 2013). Even in our *Fas in vivo* splicing assays, RBM5/10 seem to have an overlapping function while RBM6 has an opposite effect. This provides a clear indication of distinct functionalities of the RBM5, 10 and 6 proteins.

Chapter 5: Structural and functional investigations of protein-RNA interactions of RBM5 RRM1-Zf1 tandem domains

5.1. Biophysical characterization of RRM1-Zf1: Thermofluor assay

X-ray crystallography and NMR spectroscopy both require highly pure and stable biomolecular samples. Ascertaining the quality of the protein is of utmost importance as all the further experiments depend on the quality of the protein used. Therefore, biophysical characterization of the protein is necessary before any further experimentation is done.

RRM1-Zf1 is a rather difficult protein to work with. After the purification of the protein, the protein could not be concentrated and it blocked the concentrator membrane due to precipitation and possible aggregation. I therefore wanted to make a buffer screen to test the effects of different buffers, salt concentrations and pH on the stability of the protein. A thermofluor buffer screen was made in a 96-well plate where 5 µl of a master mix of RRM1-Zf1 protein (0.1 mg/ml concentration) with SYPRO Orange dye in 1:20 ratio was added to each well. The dye binds to the hydrophobic patches in the protein. As the protein unfolds with increase in temperature, the fluorescence of the dye increases due to its interaction with the hydrophobic residues of the protein. The resulting curve could be used to determine the melting temperature of the protein. If the buffer alters the stability of the protein, the melting temperature of the protein shifts accordingly. The temperature increase and fluorescence monitoring are carried out using a qPCR machine.

The thermal melting curves of the protein in each well with the different buffer conditions were recorded (**Figure 33**). Clear improvements in melting temperature of the protein (which is a direct indication of stability) were observed in three buffers, 40 mM MES pH 6.5, 40 mM Phosphate pH 6.5 and 40 mM Bis-Tris Propane pH 7.0 containing either 200 mM NaCl or 400 mM NaCl. The improvement in melting temperature in the presence of 400 mM NaCl versus that of 200 mM NaCL was 4 °C.

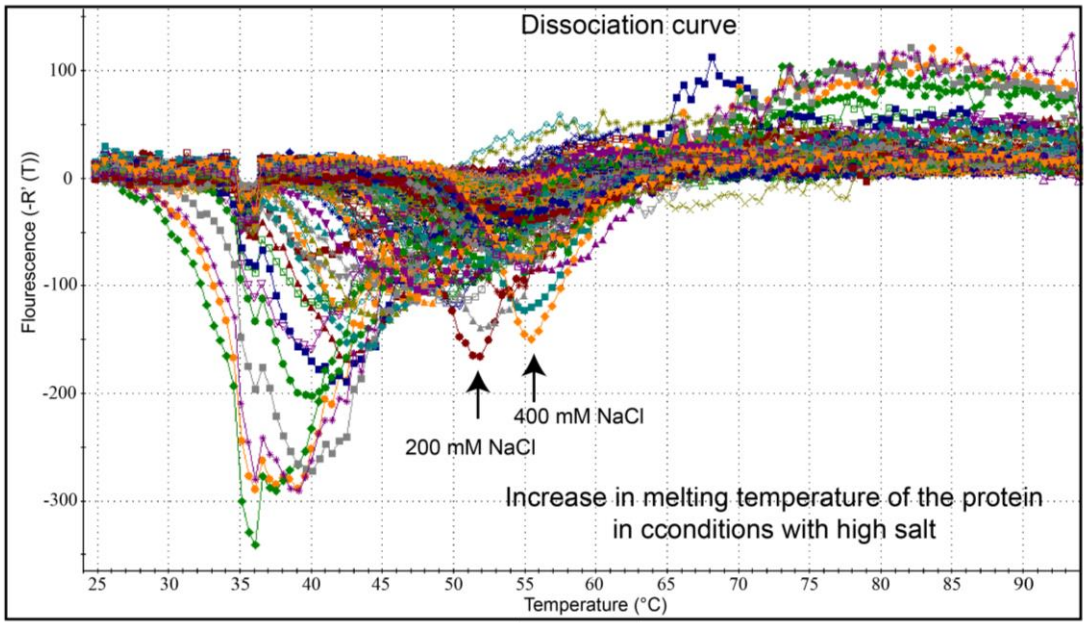


Figure 33 Thermofluor assay buffer screen for RRM1-Zf1 protein

Thermofluor assay of RRM1-Zf1 under different buffer conditions is shown. The dissociation curves representing the thermal melting curves of the protein are plotted. Arrows indicate certain buffers conditions under which the melting of the protein increases.

I finally chose 20 mM MES pH 6.5, 400 mM NaCl buffer for our further experiments with the RRM1-Zf1 protein. Bis-Tris Propane pH 7.0 buffer was discarded as lower pH of sample is better for NMR experiments (to minimize the rate of exchange of amide proton with solvent). Phosphate pH 6.5 buffer was also not pursued further as I wanted to perform Zn^{2+} titrations to see effects on the protein and precipitation of zinc phosphate due to its low solublity would not allow this.

Increasing the salt concentration to 400 mM NaCl solved the protein precipitation problem and it could be easily concentrated to >10 mg/ml without any problems.

5.2. Interaction of RRM1 and Zf1

5.2.1. Backbone assignment of RRM1, Zf1 and RRM1-Zf1 tandem construct

^{13}C, ^{15}N-labeled RRM1 domain (residues 94-184) was expressed in M9 minimal medium and purified at 4°C. It was soon realized that the protein is not stable at room temperature after a few hours and precipitates in the NMR tube. Nevertheless, backbone assignment experiments (see Methods sections for details) were recorded on the protein by

changing the sample after every assignment experiment and finally a 97.8% assignment completeness was achieved.

Zf1 domain was also expressed in M9 minimal medium and purified over the GST column with a final size exclusion chromatography step to produce a ^{15}N-labelled protein sample. A ^{15}N-HSQC spectrum of the protein was recorded (**Figure 34**). The spectrum showed a good overall peak dispersion but also contained some non-uniform NMR signals and some additional amide signals in the central area of the spectrum, possibly due to being part of unstructured regions. The peak assignments were obtained from the chemical shifts deposited in Biological Magnetic Resonance Bank (BMRB entry no.17387). Most of the assignments from BMRB matched well with the ^{1}H,^{15}N-HSQC amide signals and near complete assignment was possible (**Figure 34**).

For assignment of the tandem domain RRM1-Zf1 domain, it was ^{13}C, ^{15}N-labeled by expression in M9 minimal medium and purified. An initial ^{1}H,^{15}N-HSQC spectrum revealed the presence of an overall good spectral dispersion but non-uniformity of certain amide signals (**Figure 34**). Backbone assignment experiments were recorded and 97.4% assignment completeness was achieved.

5.2.2. Initial insights into RRM1 and Zf1 interaction

^{1}H,^{15}N-HSQC spectra report on the chemical environment of the biomolecules in question. Since chemical shifts of the backbone amides represent the basic fingerprint of the protein, changes in the chemical environment either due to ligand binding in case of ligand titrations or due to inter-domain contacts in case of multi-domain proteins can be easily seen in the 2D-^{1}H,^{15}N-HSQC spectra. Likewise, the very first indications of inter-domain interactions between RRM1 and Zf1 were obtained by a direct comparison of the ^{1}H,^{15}N-HSQC spectra of the single and tandem domain constructs (**Figure 34**). An overlay of the ^{1}H,^{15}N-HSQC spectra of RRM1 and Zf1 onto that of the tandem RRM1-Zf1 domains shows significant chemical shift perturbations (CSPs) in both the domains, although more so in the Zf1. This is indicative of a major structural change or domain organization in the Zf1 domain.

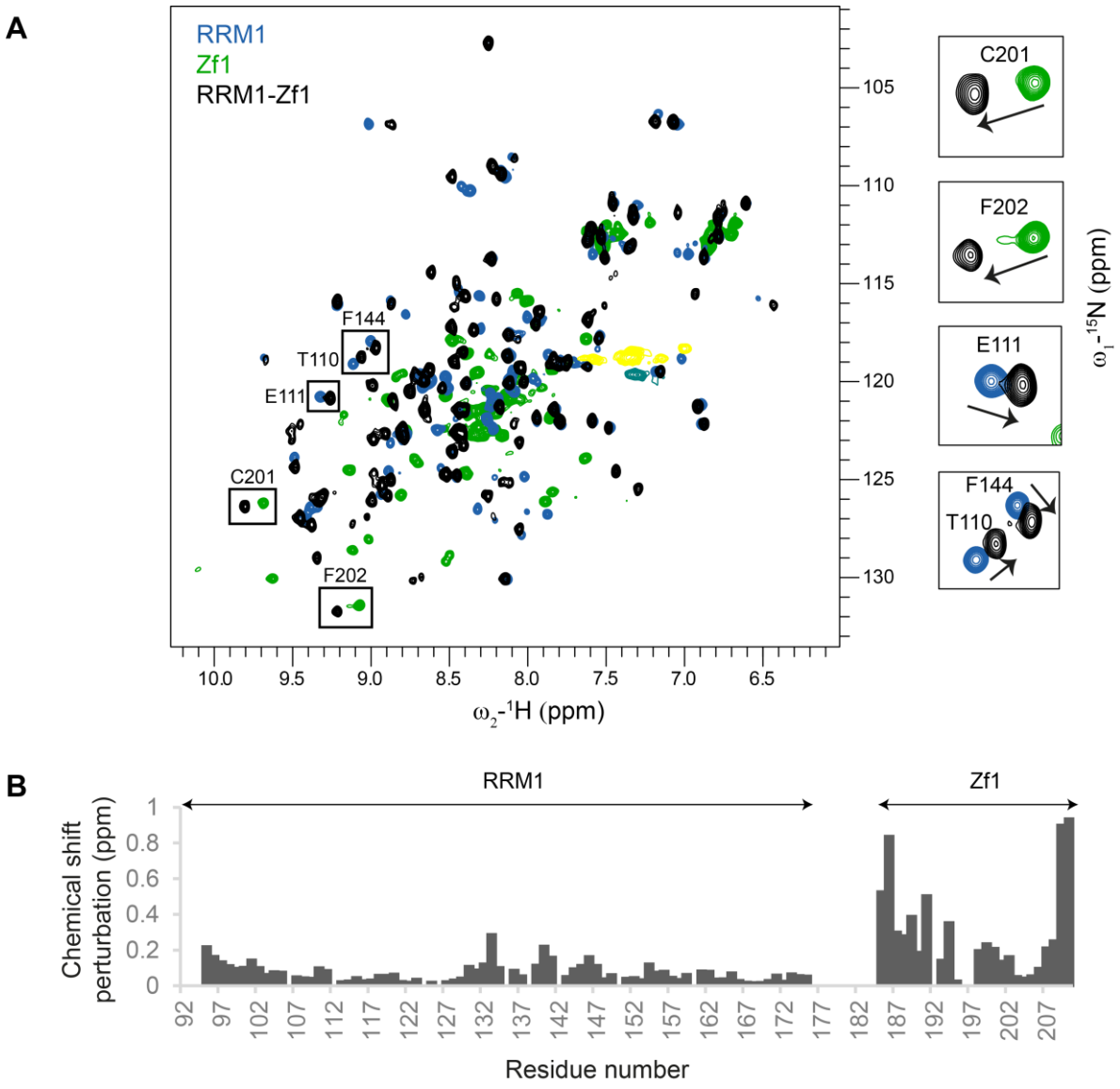


Figure 34 Interaction between RRM1 and Zf1 domains

(A) An overlay of ^{1}H,^{15}N-HSQC spectra of single domains RRM1 and Zf1 with that of tandem domain RRM1-Zf1. Zoom-ins of several residues belonging to each of the domains are shown on the right, to provide a clearer view. (B) The chemical shift perturbations in the RRM1-Zf1 tandem domain compared to individual RRM1 and ZF1 domains are plotted against the residue number in the lower panel. The changes in the Zf1 domain are much more pronounced than in the RRM1 domain.

5.2.3. Relaxation analysis of RBM5 RRM1-Zf1

To understand the system further, ^{15}N-relaxation data on the tandem domains RRM1-Zf1 were recorded. T_1 and $T_{1\rho}$ experiments were collected to estimate the total rotational correlation time. The value of τ_c depends on the overall size of the protein and as a general rule of thumb, the theoretical value is approximately ~0.6 times the size of protein. Since the RRM1

and Zf1 domains have a big difference in their molecular weight (RRM1 ~9 kDa and Zf1 ~3.5 kDa), if the domains do not interact with each other and thus tumble separately in solution, the τ_c values for the two domains should correspond to their individual molecular weight and thus be significantly different.

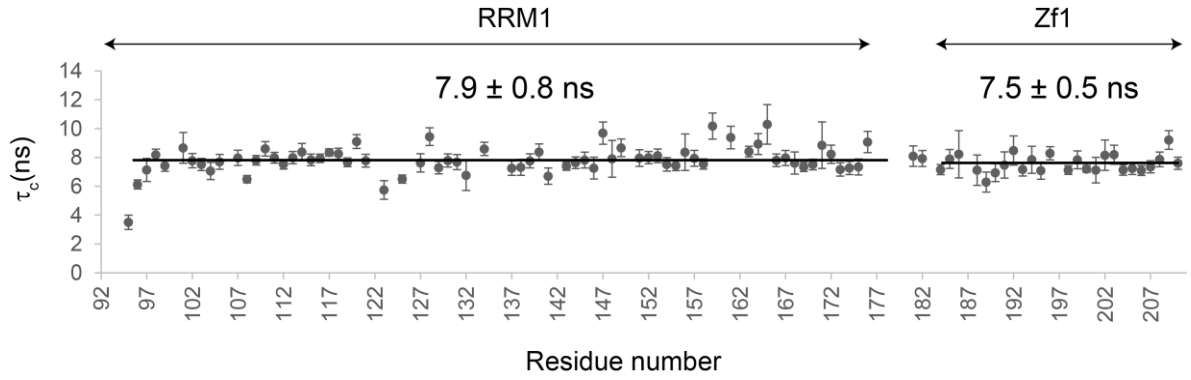


Figure 35 ^{15}N-relaxation data for RRM1-Zf1 tandem domains

The total rotational correlation time (τ_c) calculated from the R_1 and R_2 rates is plotted against residue numbers, the average ± standard deviation values are listed for each domain. Since the difference between the τ_c values for RRM1 and Zf1 is within the error, it is concluded that the two domains tumble together in solution.

Contrastingly, the total rotational correlation time obtained is approximately same for the two domains, τ_c ~8 ns for RRM1 and ~7.5 ns for Zf1; fitting the theoretical value of τ_c = 8.4 ns for RRM1-Zf1 (14 kDa) as seen in **Figure 35**. This suggests that the two domains tumble together in solution and behave as a single entity, possibly due to inter-molecular contacts between them. Since the linker between RRM1 and Zf1 is considerably short (7 residues), in principle it is also possible that the coupling between the two domains is only due to the very short distance between them. In such a case, the correlation time would be artificially higher for each of the individual domains owing to drag or motional coupling. But in this case, the total correlation time is exactly the same for the two domains suggesting that they tumble together in solution.

5.2.4. Crystal structure of RBM5 RRM1-Zf1

After initial stability tests using thermofluor assay, it became clear that the RRM1-Zf1 protein is stable only at high salt (400 mM NaCl). Also, DLS showed that the protein forms dimers at higher concentration, which could increase the possibility of crystallization. Sparse matrix crystallization screens at 10 mg/ml protein concentration and at room temperature and 4 °C were setup. Very thin needle clusters were observed within 3 days in a solution containing

0.1 M BICINE, pH 9.0, 20 % PEG 6000. After further optimization using grid search and additive screening, diffraction quality crystals were obtained in a buffer containing 0.1 M BICINE, pH 8.5, 10 % PEG 6000 with 10 % of 1 M Cesium chloride as additive. The crystals that appeared as thin needle clusters were optimized into thin plates (**Figure 36**).

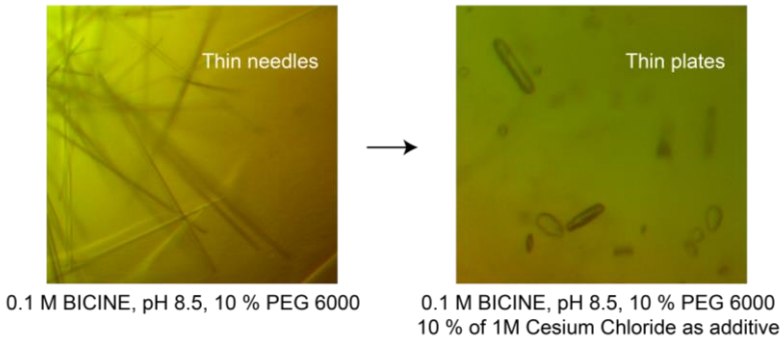


Figure 36 Optimization of RRM1-Zf1 crystals

Optimization of RRM1-Zf1 crystals from thin needles to separate, thin plates using additive screen.

The crystals were set up by hanging drop method containing 1 µl protein (10 mg/ml) and 1 µl crystallization buffer. Crystals were transferred to a solution containing the crystallization buffer with 30% ethylene glycol before being flash frozen in liquid nitrogen. The crystals diffracted to 2.9 Å resolution. The best diffracting crystals were used for further data processing. Even though RBM10 and RBM5 share a high degree of similarity, the homology model obtained for RBM5 RRM1 based on that of NMR structure of RBM10 (PDB ID: 2LXI) as template and RBM5 Zf1 NMR structure (PDB ID 2LK0) were unsuccessful in obtaining a reliable solution using molecular replacement. Since the Zf1 has a Zn^{2+} ion bound, I collected anomalous diffraction data on Zn^{2+} ion on ID23-1 tunable beamline at ESRF Grenoble. The structure was solved by SAD phasing using anomalous signal from zinc.

Clear and unambiguous electron density could be observed for RRM1 and thus it could be used using automated methods. On the other hand, the Zf1 has a number of residues in loops, with only 6 residues forming secondary structure elements and it was thus built manually in Coot. Two copies of the protein was found in the unit cell which are related to each other by a two fold non-crystallographic symmetry. The statistics for data collection and refinement are shown in **Table 3**.

Table 3 Data collection and refinement statistics for RRM1-Zf1 crystal

Parameter	RRM1-Zf1
Wavelength	1.282 Å
Resolution range	33.46 – 2.873 (2.975 – 2.873)
Space group	C 1 2 1
Unit cell	61.1 40.07 96.59 90 95.382 90
Total reflections	34641 (3080)
Unique reflections	5445 (512)
Multiplicity	6.4 (6.0)
Completeness (%)	99.03 (96.60)
Mean I/sigma(I)	13.65 (3.55)
Wilson B-factor	51.49
R-merge	0.1194 (0.5077)
R-meas	0.1302 (0.5566)
Reflections used in refinement	5438 (512)
Reflections used for R-free	541 (51)
R_{work}	0.2036 (0.3128)
R_{free}	0.2727 (0.4389)
Number of non-hydrogen atoms	1912
macromolecules	1877
ligands	2
solvent	33
Protein residues	232
RMS(bonds)	0.010
RMS(angles)	1.85
Ramachandran favored (%)	91.23
Ramachandran allowed (%)	5.26
Ramachandran outliers (%)	3.51
Rotamer outliers (%)	8.04
Average B-factor (macromolecules)	52.85
Statistics for the highest-resolution shell are shown in parentheses.	

RRM1 has the canonical RRM fold with βαββαβ topology, where the four anti-parallel β-strands form the RNA binding interface on one side and two α-helices pack against them, facing the other side of the domain. Similarly, the Zf1 also has the canonical RanBP2 zinc finger structure, where the two β-strands sandwich the conserved tryptophan residue and the Zn^{2+} ion, which is coordinated by four cysteine residues.

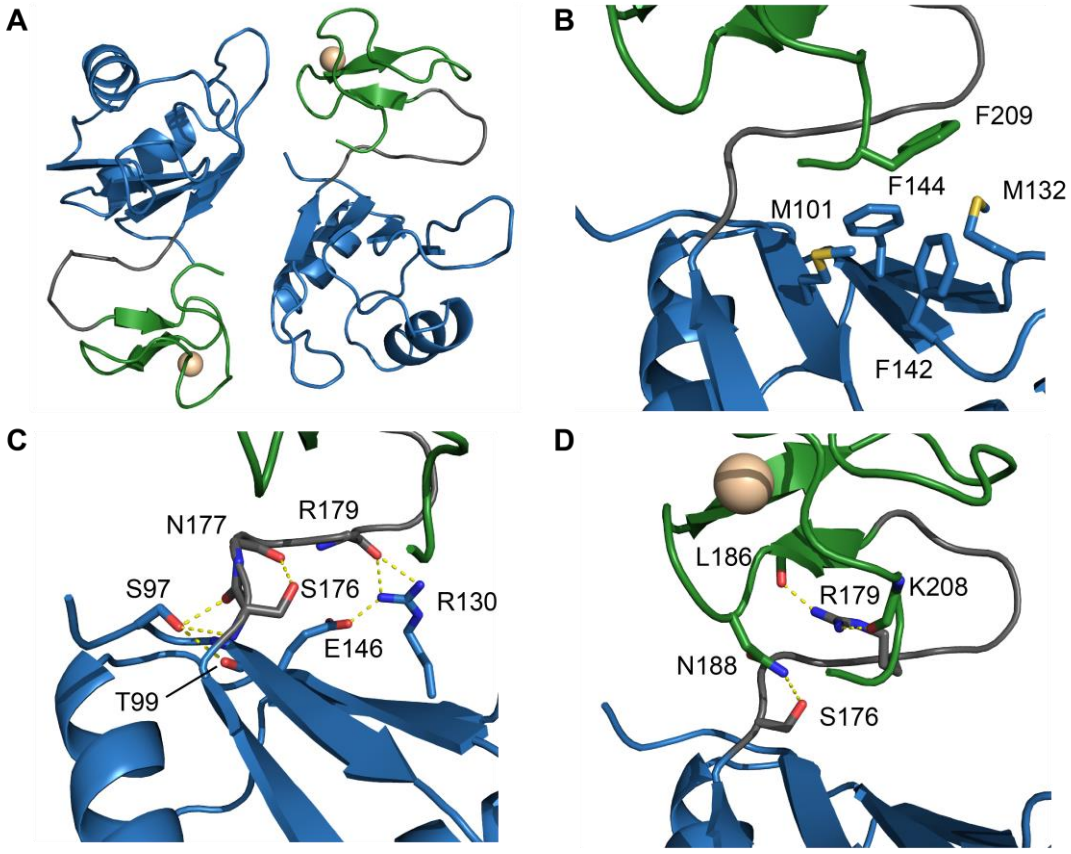


Figure 37 Crystal structure of RRM1-Zf1

The crystal structure of RRM1-Zf1 is shown with RRM1 and Zf1 domains colored in blue and green, respectively. The Zn^{2+} ion coordinated by cysteine residues of Zf1 is colored in golden. (A) Two molecules of RRM1-Zf1 present in the unit cell are shown. (B) The hydrophobic core formed between RRM1 and Zf1 residues is shown. Hydrogen bonds between the linker and RRM1 residues are shown as yellow dotted lines in panel (C), while those between linker and Zf1 are shown in panel (D).

The crystal structure clearly shows an intricate network of molecular interactions not only between RRM1 and Zf1 but also with the linker between them, holding the two domains in close proximity to each other (**Figure 37**). In particular, hydrogen bonds are formed between side-chain hydroxyl groups of Ser 97, backbone carbonyl of Thr 99 and backbone amide and carbonyl of Ser 176. Next, hydrogen bonding takes place between backbone carbonyl of Asn 177 and side-chain hydroxyl group of S176. Another set of hydrogen bonds are formed between backbone carbonyl of Arg 179 from the linker and Glu 146 and Arg 130 from RRM1. Taken together, these interactions form the hinge or anchor region to define the orientation of the linker with respect to RRM1 (**Figure 37C**). Hydrogen bonds are also formed between side chains of Asn 188 from Zf1 and Ser 176, backbone amide of Asn 177 from the linker; between backbone carbonyls of residues Leu 186, Lys 208 from Zf1 and side-chain of Arg 179 from

the linker (**Figure 37D**). Apart from these interactions of the linker with either of the domains, a hydrophobic cluster is formed involving two Methionine residues, Met 101 and M 132, and two Phenylalanine residues, Phe 142 and Phe 144, from RRM1 domain which is further stabilized by T-stacking with Phe 209 from Zf1 domain (**Figure 37B**). This is the only interaction observed between the two domains. It is plausible that formation of this inter-domain hydrophobic cluster is stabilized or promoted in the presence of high salt (400 mM NaCl), which could explain the increase in stability of the protein at high salt as seen in thermofluor assay.

5.2.5. Validation of RRM1-Zf1 crystal structure

Validation of structures determined using X-ray crystallography using other solution-based methods is important, considering the fact that artificial contacts in the crystal structure may be present as a result of crystal packing. I used a combined approach of NMR based Residual Dipolar Couplings (RDCs) and SAXS to determine if the conformation of RRM1-Zf1 present in the crystal structure is maintained in solution or not. For this, a protein stabilizing cysteine mutant referred to as RRM1-Zf1 C191G was used (see **section 5.4** for details). RDC and SAXS experiments were initially performed on the wild-type RRM1-Zf1 protein, but since it was not stable during the course of the experiments, I switched to the mutant protein.

HN-RDCs for RRM1-Zf1 C191G were obtained using 2D-IPAP-HSQC experiments. ^{15}N-lablled protein was purified and IPAP-HSQC was measured on the isotropic sample. Next, sample from the same tube was mixed with 6% PEG-hexanol alignment medium (Rückert and Otting 2000) in a 1:1 ratio diluting the protein and PEG-hexanol alignment medium concentration to half. 3% PEG-hexanol alignment medium was therefore used to attain partial alignment and thereby anisotropy in the protein sample. The absolute values of HN-RDCs were obtained by subtracting the coupling for each backbone amide in isotropic sample (J-coupling) and anisotropic sample (J-coupling + Dipolar-coupling). It is important to maintain the spectral quality in the anisotropic sample to accurately determine the respective peak positions in the IPAP-HSQCs, usually achieved by increasing the experimental acquisition time (**Figure 38A**).

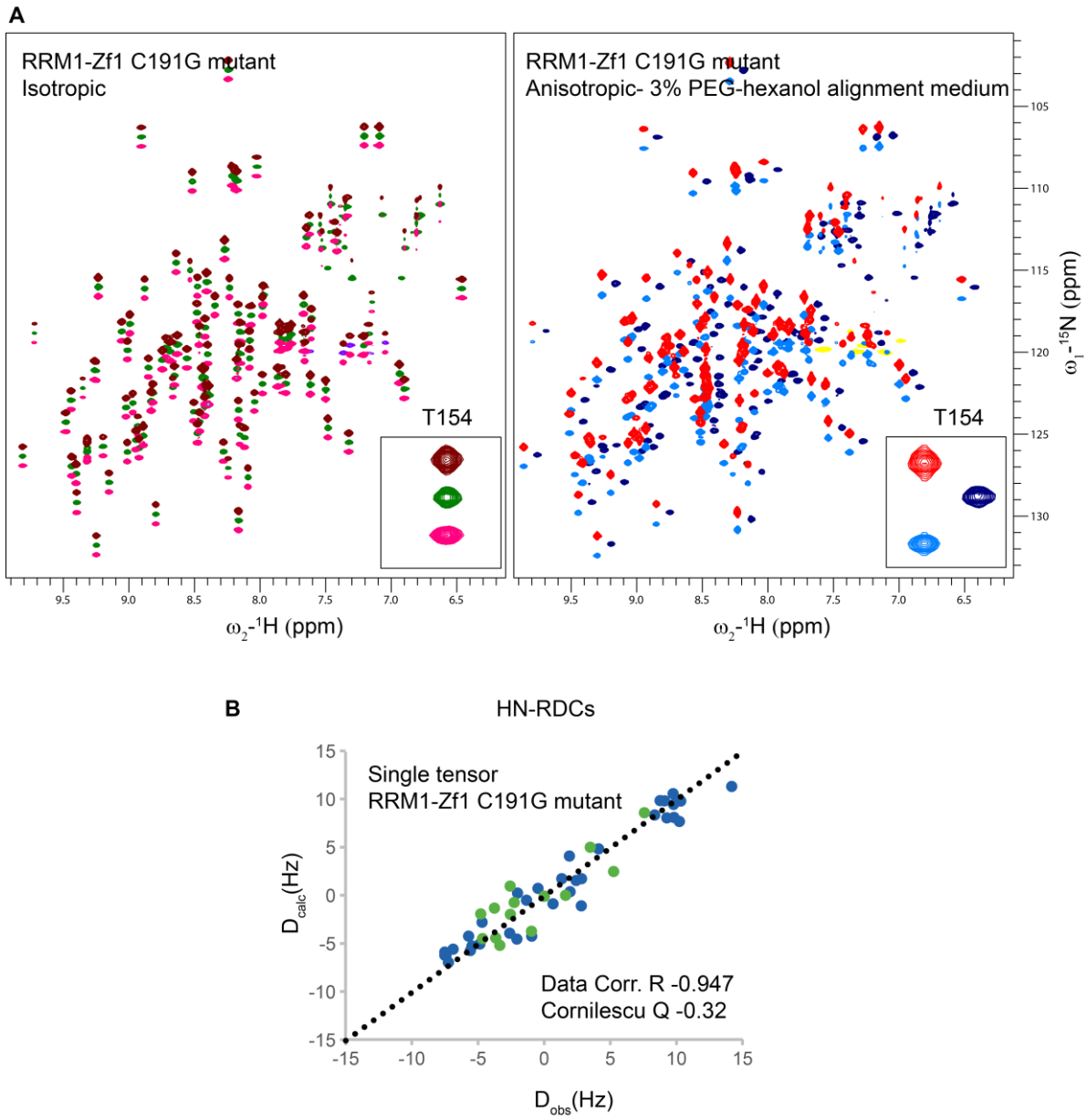


Figure 38 HN-RDCs measured for RRM1-Zf1 C191G mutant

(A) To demonstrate the spectral quality, a superposition of the IPAP-HSQCs which are split and $^{1}H,^{15}N$ -HSQC in the middle is shown for isotropic and anisotropic sample in brown, pink, green and red, cyan and blue respectively.(B) Analysis of HN-RDCs using the software PALES, where the experimental or observed RDCs (D_{obs}) are plotted against the back-calculated RDCs (D_{calc}) from the crystal structure. The data correlation *R* factor and Cornilescu *Q* factor are indicated

To verify the agreement between the crystal structure and the RDCs, PALES (Zweckstetter 2008) program was used. It back-calculates the RDCs from the crystal structure (D_{calc}) and compares them to the experimental RDCs measured (D_{obs}). The Cornilescu quality factor (*Q*) is used to judge the overall quality of the fit between them, along with the data correlation R factor. The value of *Q*-factor ranges between 0.1-1.414 and a lower value is

indicative of a better fit. It should be noted that only residues that are rigid (or part of secondary structural elements) should be used and flexible parts of the protein including loops should be discarded. In our case, since Zf1 has only a handful residues forming secondary structure, other residues in the Zn^{2+} ion coordination region are also used in the calculation. The quality of agreement between the experimental and back-calculated RDCs is quite high as indicated by a *Q*-factor of 0.32 and data correlation R factor of 0.947 (**Figure 38B**).

For measurement of SAXS data, unlabeled RRM1-Zf1 C191G protein was purified and concentrated to 8 mg/ml. SAXS data for a dilution series was measured at 8 mg/ml, 4 mg/ml, 2 mg/ml and 1 mg/ml. Data were recorded on BioSAXS beamline BM29 at ESRF Grenoble. The protein does not show concentration dependent aggregation behavior but does show concentration dependent increase in I_0, possibly due to dimerization. Therefore, it was unsuitable to use the higher concentration data for further analysis. The pairwise distance distribution function (p(*r*)) in SAXS describes the paired set of distances between all the electrons in the macromolecular structure. The p(*r*) function plotted for the lowest concentration (1 mg/ml) showed a D_{max} of 58.2 Å (**Figure 39A**). Furthermore, Crysol program (Svergun, Barberato et al. 1995) was used to fit the crystal structure to the SAXS curve for the lowest concentration (1 mg/ml). The quality of the fit is judged by the Chi-square value (χ^2 = 0.863), which indicated good agreement between the crystal structure and the SAXS data (**Figure 39B**).

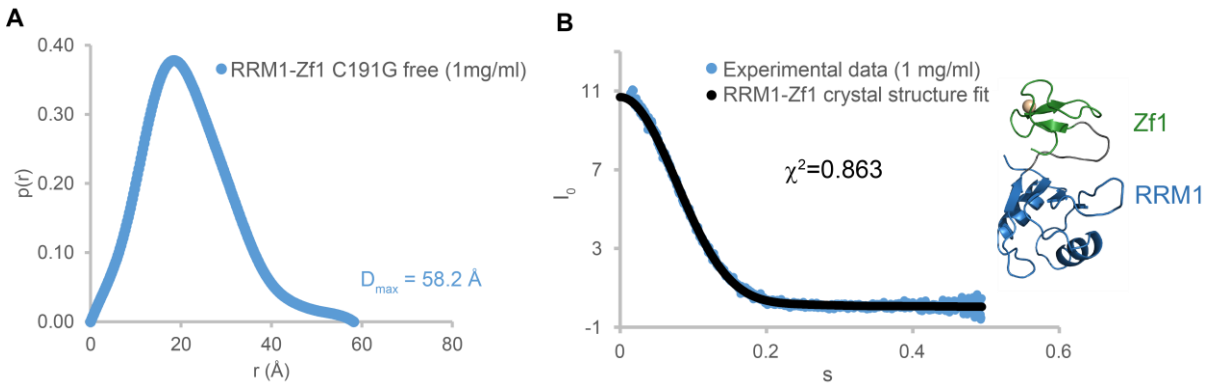


Figure 39 SAXS data for validation of RRM1-Zf1 crystal structure
(A)The p(*r*) curve showing the maximum pairwise distribution for RRM1-Zf1 C191G mutant in its free form is plotted at the lowest concentration (1 mg/ml). (B)The fit between experimental SAXS data for RRM1-Zf1 C191G protein at 1 mg/ml against that of the simulated data from RRM1-Zf1 crystal structure is plotted, as obtained from Crysol software. The Chi-square value is indicated.

Table 4 SAXS data collection and data processing statistics for RRM1-Zf1 C191G

Parameters	RRM1-Zf1 C191G, 1 mg/ml
Data-collection	
Instrument	BioSAXS BM29 ESRF
Beam geometry	10 mm slit
Wavelength (Å)	0.9919
q range ($Å^{-1}$)	0.0029-0.494
Exposure time per frame (s)[a]	1
Concentration (mg ml^{-1})	1
Temperature (°C)	20
Structural parameters	
$I_{(0)}$ (cm^{-1}) [from p(r)]	10.89 ± 00
R_g (Å) [from p(r)]	10.68 ± 0.00
$I_{(0)}$ (cm^{-1}) [from Guinier]	10.89 ± 0.022
R_g (Å) [from Guinier]	10.67 ± 0.002
D_{max} (Å)	58.2
Porod volume estimate ($Å^3$)	24270
Software employed	
Primary data reduction	BsxCuBE
Data processing	PRIMUS

[a] 15 frames were recorded for each sample

Taken together, RRM1-Zf1 crystal structure is validated using RDCs as well as SAXS proving that the conformation of the tandem domain RRM1-Zf1 remains the same in solution.

Further, I also wanted to investigate the consequences of randomizing the linker connecting the two domains. For this, seven residues of the nine residue linker (residue number 178-184) were mutated to a potentially unstructured Gly-Ser repeats (**Figure 40**). This mutant protein is referred to as RRM1-GGS-Zf1 mutant. Since many of the contacts of the linker with either of the domains are mediated by the side-chains of the residues involved, this mutant should disrupt the inter-domain interface between RRM1 and Zf1. ^{15}N-labeled protein was purified from M9 minimal medium and a ^{1}H,^{15}N-HSQC spectrum was recorded. An overlay of the wild-type RRM1-Zf1 spectrum over that of the RRM1-GGS-Zf1 mutant protein shows that a lot of chemical shifts occur in the mutant protein indicating a major structural change (**Figure 40A**, left panel). A superposition with the individual single domain spectra with that of the wild-type and mutant proteins shows that the amide signals in the mutant proteins shift towards those in the single domains. This becomes more clear by looking at zoom-ins of two residues one belonging to each of the domains (Gly 141 from RRM1 and Cys 201 from Zf1), which are shown in **Figure 40A**, right panel.

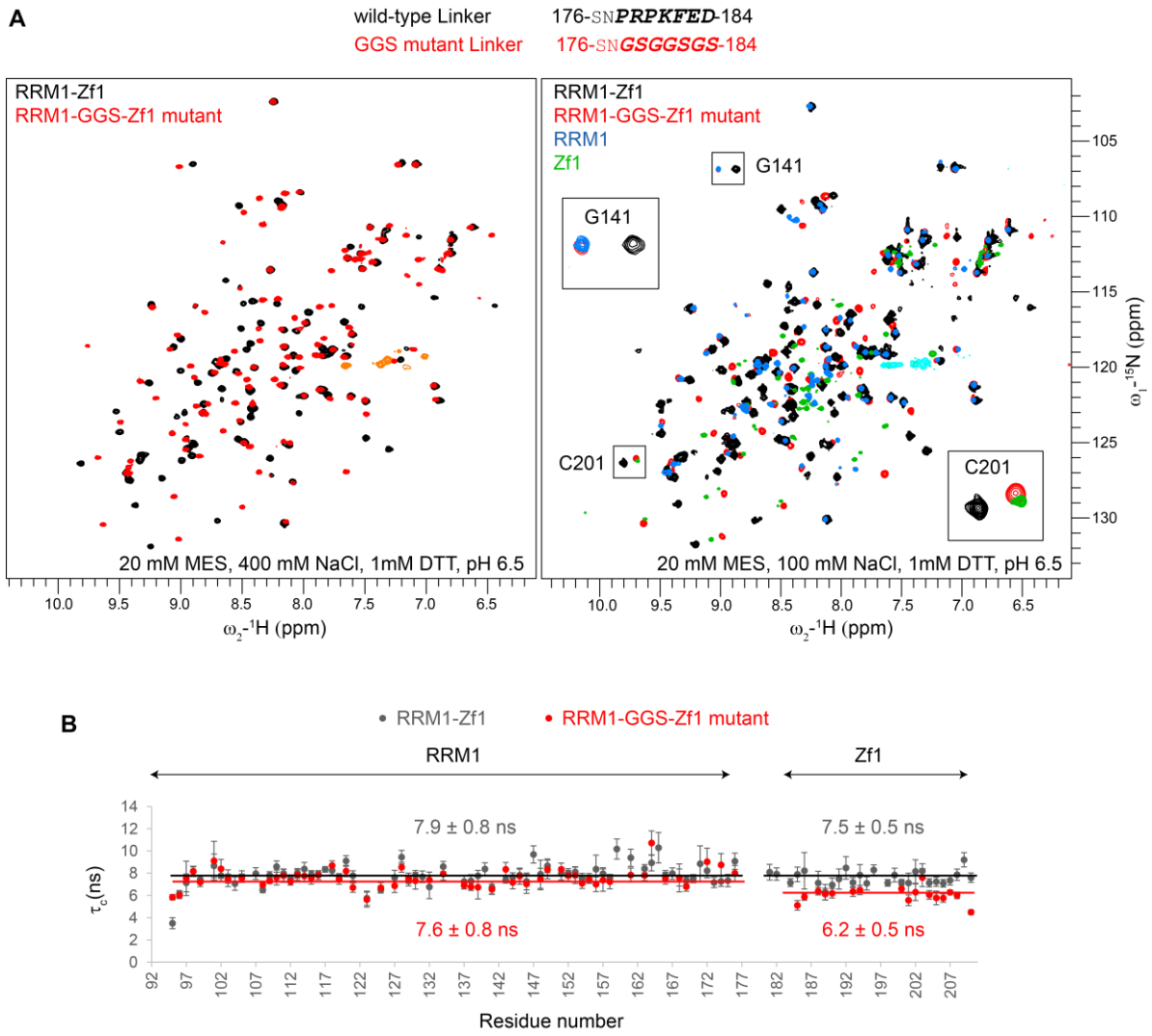


Figure 40 RRM1-GGS-Zf1 mutant disrupts inter-domain contacts

(A) Superposition of wild-type RRM1-Zf1 and RRM1-GGS-Zf1 mutant $^{1}H,^{15}N$ -HSQC spectra is shown in black and red, respectively. The large differences in the spectra indicate a major structural reorganization. The data were collected in high salt buffer (400 mM NaCl) which is needed for stability of the tandem domain constructs at high concentration. On the right, a superposition of wild-type RRM1-Zf1, RRM1-GGS-Zf1 linker mutant and RRM1, Zf1 single domain $^{1}H,^{15}N$ -HSQC spectra is shown in black, red, blue and green, respectively. All these data are collected in low salt buffer (100 mM NaCl) and at low concentrations for tandem domain constructs. Zoom-ins of two residues (Gly 141, Cys 201), each belonging to either of the domains demonstrate that the chemical shifts of residues in single domain are comparable to that in the RRM1-GGS-Zf1 linker mutant (B) The total rotational correlation time (τ_c) calculated from R_1 and R_2 rates is plotted against residue numbers for wild-type RRM1-Zf1, RRM1-GGS-Zf1 linker mutant in grey and red, respectively. The average ± standard deviation values are listed for each domain. Since the difference between the τ_c values of wild-type RRM1-Zf1 and RRM1-GGS-Zf1 linker mutant, especially for Zf1 indicates partial flexibility of the domains with respect to each other introduced in the linker mutant.

To unveil if the linker mutation creates differences in total rotational correlation time (τ_c) in comparison with that of the wild-type protein, ^{15}N relaxation experiments were recorded

with RRM1-GGS-Zf1 mutant (**Figure 40B**). As expected, a considerable drop in the τ_c of the Zf1 domain from ~7.5 ns to ~6.2 ns is observed . It therefore shows that the linker mutation alters the dynamics of the two domains where due to partial disruption of interactions of the linker and the domains, the two domains become partially flexible to each other.

5.3. Investigations of RRM1-RNA interactions

It is known that N-and C-terminal extensions of RNA-recognition motif (RRM) domains often contribute to RNA binding by increasing the protein-RNA interaction network, in addition to the core RNP motifs (Maris, Dominguez et al. 2005). I therefore created a shorter version of RRM1 where the C-terminal linker between RRM1 and Zf1 is chopped off with domain boundary limited to residue 177 and carried out further characterization of the two different versions of the protein.

5.3.1. C-terminal linker of RRM1 makes contacts with the core of the domain

The crystal structure of RRM1-Zf1 tandem domains illustrates that there are a number of contacts between the linker to RRM1 as well as to Zf1 (**Figure 37C**). Next, I wanted to learn if the contacts between the linker and RRM1 are still maintained in the absence of Zf1 domain. This would essentially indicate if the residues involved in RNA binding, on the β-sheet interface, would be potentially made inaccessible due to partial hindrance of the RNA binding surface of RRM1 by the linker.

The idea of this potential interaction between RRM1 and linker in the free form of the protein stems firstly from the crystal structure of the tandem domain. Secondly, a similar stacking between the C-terminal extension of RRM2 and the core RNP motifs was previously suggested (Song, Wu et al. 2012). It was shown that upon RNA titration, dramatic CSPs are observed in this C-terminal extension which becomes more flexible in the RNA bound form. These data indicate that the C-terminal extension is forced to stay away from the core of the domain in the RNA bound form. Still, the C-terminal extended version of RRM2 demonstrated a two-fold increase in RNA binding affinity.

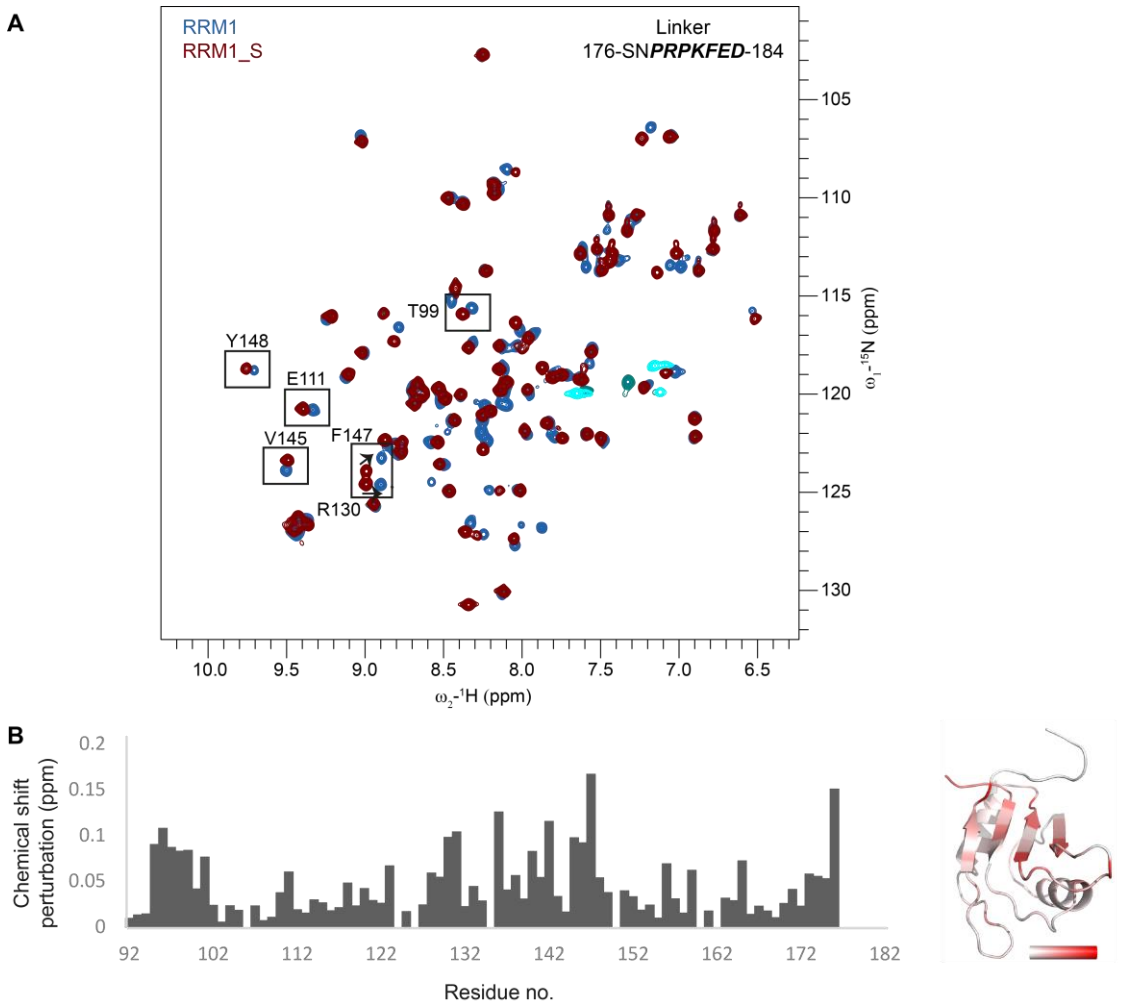


Figure 41 C-terminal extension possibly makes contacts to the core of RRM1

(A) Overlay of ^{15}N-HSQC spectra of RRM1 (containing the C-terminal extension) and RRM1_S (truncated version) is shown in blue and brown, respectively. (B) The chemical shift perturbation plot highlighting the differences between them is shown. These chemical shift changes are plotted onto the structure of RRM1 from white to red with increasing severity of the changes.

To understand if a similar mechanism might exist in the case of RRM1, I used two versions of the protein, one in which the C-terminal linker is present (RRM1) and the other in which the linker is truncated and residues 178-184 are not present (RRM1_S). Both the versions of RRM1 were expressed and purified from ^{15}N-labeled M9 minimal medium. After purification, it was observed that the truncated version of RRM1 (RRM1_S) was also unstable at room temperature and precipitated within a few hours, similar to RRM1. Nevertheless, ^{1}H,^{15}N-HSQC spectra of both the proteins were recorded to point towards any differences that might exist between them (**Figure 41**).

Clear chemical shift perturbations are observed in several residues when the C-terminal extension is deleted in RRM1 (RRM1_S). If this extension would be completely flexible, without making any contacts to the core of the domain, only chemical shift changes would be observed at the end of the protein indicating local effects due to truncation. Upon plotting the chemical shift differences between the truncated and C-terminal extended forms of RRM1 onto the structure in **Figure 41**, it becomes clear that CSPs are observed on the β-sheet interface of the protein. Therefore, it can be concluded that the C-terminal linker of RRM1 makes possible contacts with the core of the domain and perhaps also hindering the RNA binding interface.

5.3.2. RRM1 recognizes a pyrimidine rich RNA ligand

To test the RNA binding capability of RRM1, an RNA sequence motif derived from the intronic region immediately upstream of ln100 element of *Caspase-2* pre-mRNA which has been previously shown to be important for alternative splicing regulation of *Caspase-2* pre-mRNA via RBM5 was used. The RNA sequence is rich pyrimidine (C/U) rich with the sequence- 5'-UCUCUUCUC-3', named as CU_9. NMR titrations of the RNA with RRM1 were made and ^{15}N-HSQC spectra were recorded at each titration step.

An overlay of ^{1}H,^{15}N-HSQC spectra of free and RNA bound RRM1 is shown in **Figure 42**. Chemical shift perturbations are observed on the β-sheet RNA binding interface of the protein (as shown in the CSP on structure plot). This is not surprising as the canonical RNA binding residues are located on RNP2 and RNP1 forming two of the four β-strands. Strikingly, dramatic chemical shifts are observed in the residues forming the linker between RRM1 and Zf1 (or C-terminal extension). These chemical shifts are an order of two greater in magnitude to the strongest shifts in the core domain which are observed in RNP1 residues (residues 140-147). Zoom-into regions of the ^{1}H,^{15}N-HSQC spectra showing residues Asn 177 and Arg 179 of the linker highlight how tremendous the shifts in the linker residues are.

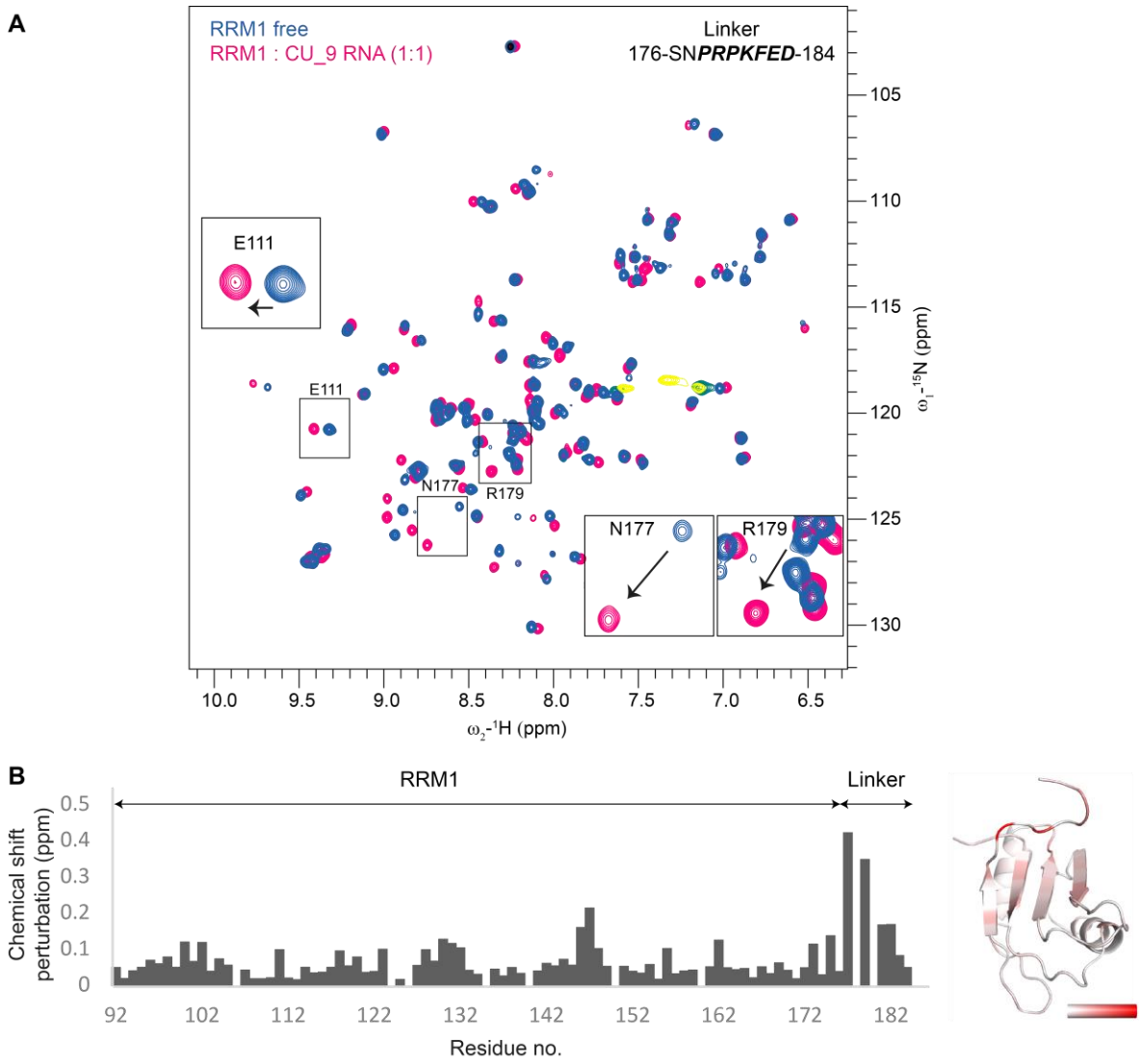


Figure 42 C-terminal linker of RRM1 gets displaced upon RNA binding

(A) Overlay of ^{1}H,^{15}N -HSQC spectra of free RRM1 (containing the C-terminal extension) and C/U rich RNA bound form is shown in blue and pink, respectively. (B) The chemical shift perturbation plot is shown. Highest chemical shifts are observed in the linker residues. These chemical shift changes are plotted onto the structure of RRM1 from white to red with increasing severity of the changes.

Two possibilities arise in this scenario. Either the linker gets directly displaced from the β-sheet interface of the protein or the linker gest displaced and makes additional contacts with the RNA. To further investigate which of the two scenarios exist, I performed ITC experiments where both, the C-terminal extended (RRM1) and truncated (RRM1_S) versions of the protein were tested for binding to the same C/U rich RNA (**Figure 43**).

The ITC binding isotherms show that the affinity of RRM1 and RRM1_S for C/U rich RNA differs by a factor of two ($K_{D\ RRM1}$ ~20±4 μM, $K_{D\ RRM1\_S}$ ~11±5 μM). This is a small difference especially considering the error values which are calculated from two replicates.

Moreover, the dissociation constant of the C-terminal extended version of RRM1 for binding to CU_9 RNA is higher indicating an even lower affinity than the truncated version of the protein. It is therefore concluded that the linker does not contribute towards RNA binding.

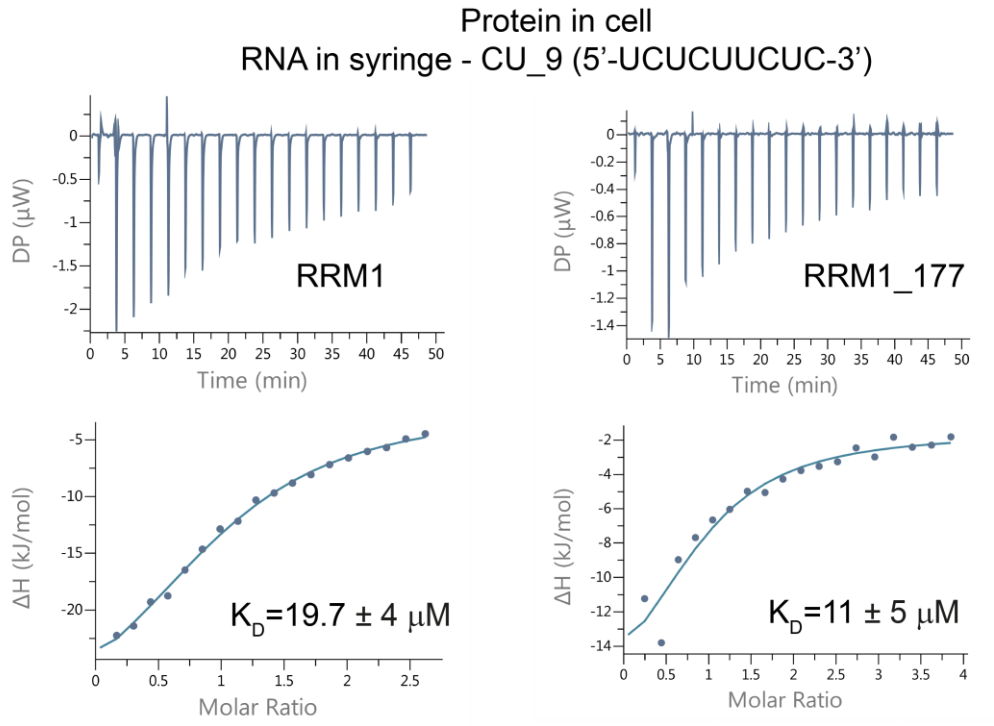


Figure 43 C-terminal linker does not have an effect on RNA binding

ITC binding isotherms of RRM1 (with C-terminal extension) and RRM1_S (truncated version) are shown and the dissociation constants are indicated. In all these experiments, the respective protein is present in the cell and the CU_9 RNA ligand (5'-UCUCUUCUC-3') is titrated from the syringe into the cell

Taken together, it becomes likely that the contacts between the RRM1 core and C-terminal extended linker which are present in the free form of the protein are absent in the RNA bound form of the protein and the linker possibly gets directly displaced upon RNA binding without making any additional contacts to the RNA.

Next, I wanted to investigate the protein-RNA complex but this time by recording RNA based NMR experiments. A 2D ^{1}H-^{1}H TOCSY spectrum provides a fingerprint of the RNA which can be used to study the linewidths of the RNA signals upon complex formation. This is an important step when preparing for NMR-based structure determination of a protein-RNA complex, where essentially sharp NMR signals are preferred not only for the protein but also for the RNA. The H5-H6 correlations in the 2D ^{1}H-^{1}H TOCSY experiment are very useful for this purpose. 2D ^{1}H-^{1}H TOCSY spectra of the free and protein-bound to CU_9 RNA were

recorded (**Figure 44A**). All the RNA bases display chemical shifts with some showing significant changes. Moreover, all the bases display sharp signals which is extremely important.

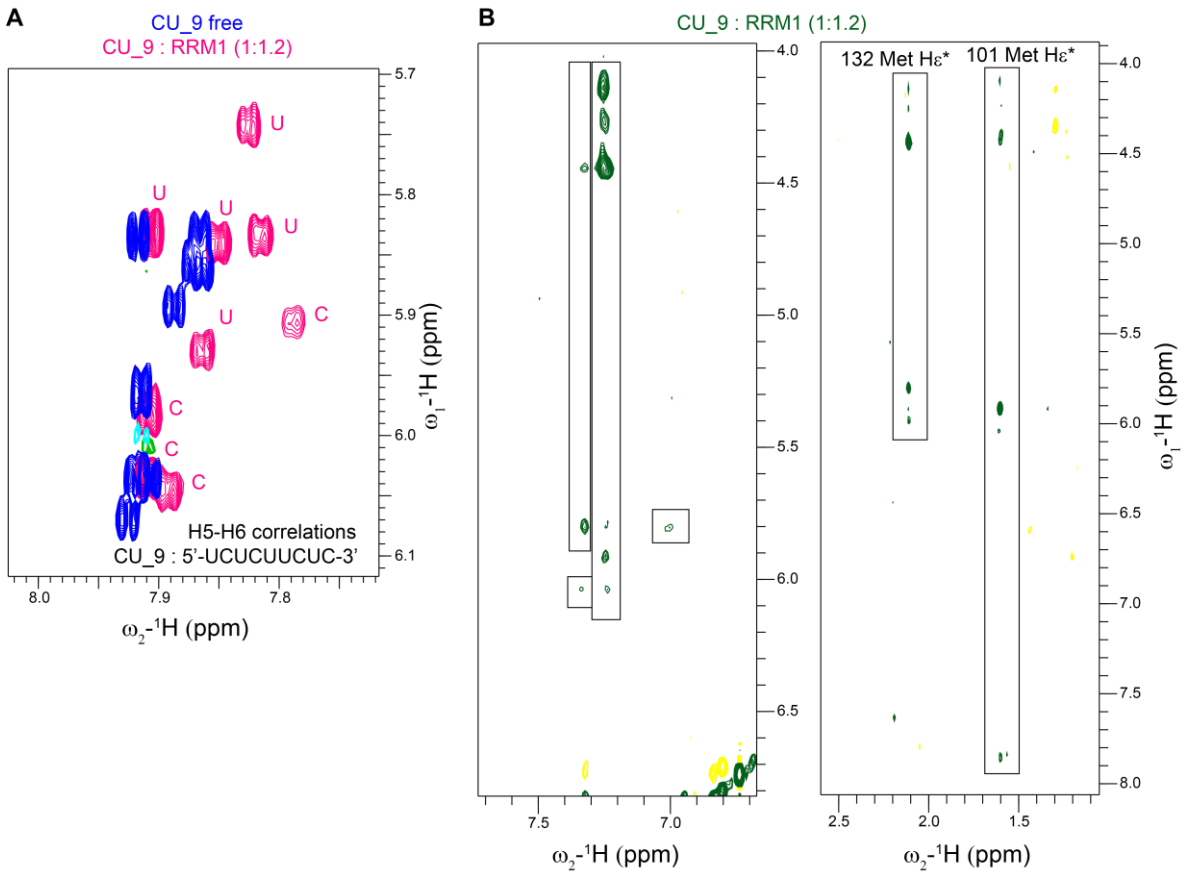


Figure 44 Initial experiments to obtain insights into RRM1-C/U rich RNA complex

(A) An overlay of H5-H6 correlations in 2D TOCSY spectra for free and RRM1 bound is shown in blue and pink, respectively. (B) Specific regions of 2D-ω_1 filtered NOESY spectrum are shown with the intermolecular NOEs between protein and RNA molecules highlighted with boxes.

A natural abundance $^{1}H,^{13}C$-HSQC experiment was recorded on the same sample to allow for assignment of the RNA bases in the bound form. Since C5 carbons of cytosine and uracil have peculiar chemical shifts (average shifts for C5 carbon of cytosine- 96.96 ppm, uracil-103.04 ppm), it was possible to assign the RNA bases in the bound form to either a cytosine or uracil, using C5-H5 correlations from the $^{1}H,^{13}C$-HSQC (**Figure 44A)**.

Next, I recorded a 2D-ω_1 filtered NOESY experiment where only intermolecular NOEs between the protein and RNA should be observed (**Figure 44B**). In total, 24 intermolecular NOEs could be observed between the protein-RNA components. In the first panel, intermolecular NOEs between the side chains of aromatic residues and the H5 of C/U, while the strongest cross-peaks are observed to the sugar protons (between 4-4.5 ppm) of the RNA.

In the second panel, intermolecular NOEs between the methyl protons of the protein and the sugar protons, H5 and H6 of the RNA are observed.

This served as a good starting point for recording the full set of NMR experiments required for structure calculation. Unfortunately, the protein-RNA complex was not stable over long periods of time and severe degradation and precipitation of the complex was observed in the NMR tube after 3-4 days. Therefore, a full NMR based structure calculation was not possible. Additionally, there were problems in assignment of the ^{13}C-aromatic HSQC due to possible exchange broadening processes which made the further analysis even more difficult. Nevertheless, some useful information could still be derived from the data. Using 3D-ω_1 filtered, edited aliphatic ^{13}C-NOESY complemented with 3D-ω_1 edited aliphatic ^{13}C-NOESY, I was able to unambiguously assign the methyl protons as 132 Met Hε* and 101 Met Hε* that show intermolecular NOE cross-peaks to the sugar protons as well as to H5 and H6 of the RNA bases (**Figure 44B, right panel**).

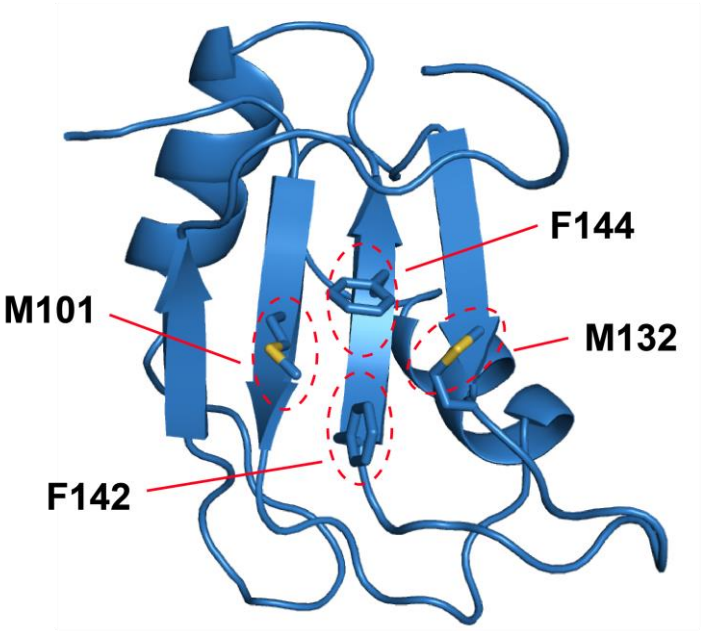


Figure 45 RNA binding residues obtained from filtered NMR experiments

Met 101 and Met 132 in addition to Phe 142 and Phe 144 are shown on the RRM1 structure, highlighted with red dashed circles. The Hε* protons from both the methionine residues generate inter-molecular NOEs to the RNA.

The positions of Met 101 and Met 132 residues in the RRM1 structure are shown in **Figure 45** and they lie on the β-sheet RNA binding interface. Phe 142 and Phe 144 are also indicated as they correspond to the possible RNA binding aromatic residues in the canonical RNP1 motif. The intermolecular NOEs from the aromatic side-chains of the protein to the RNA which could not be assigned due to exchange-broadening could possibly belong to these phenylalanine residues.

Therefore, even though a full NMR based structure calculation of RRM1-RNA complex could not be achieved, it still possible to obtain some useful information on the protein-RNA complex.

5.4. Ambiguous zinc coordination by an additional cysteine in Zf1 renders protein unstable

Even after several attempts to stabilize the protein by changing buffer conditions, pH and temperature, a good quality $^{1}H,^{15}N$-HSQC spectrum of the tandem domain (wild-type RRM1-Zf1) construct could not be recorded. Also, certain residues surrounding the Zn^{2+} ion coordination site showed two sets of amide signals which would then converge to one signal after a few days at room temperature or at 4 °C. This did not present a major problem as the additional amide signals would disappear with time. Additionally, to validate the crystal structure, HN-RDCs were recorded by creating partial alignment in the sample using 3% PEG-hexanol medium. Two sets of amide signals were observed for a number of residues in the protein indicating that two species or conformations of the protein exist in the sample which are in slow exchange (**Figure 46**). Interestingly, this behavior of the protein is not observed under isotropic conditions as evidenced by the presence of singly cross-peak belonging to each amide signal (**Figure 46**). It is therefore possible that either the protein interacts with the alignment medium leading to doubling of the amide signals or the protein indeed has two conformations which are in fast exchange with each other under isotropic conditions, but enters slow exchange regime upon addition of the alignment medium which might interfere with the exchange process between the two conformations. Therefore, due to the aforementioned complications in the analysis of the HN_RDCs, structure validation of the wild-type RRM1-Zf1 using NMR based RDCs was not further pursued.

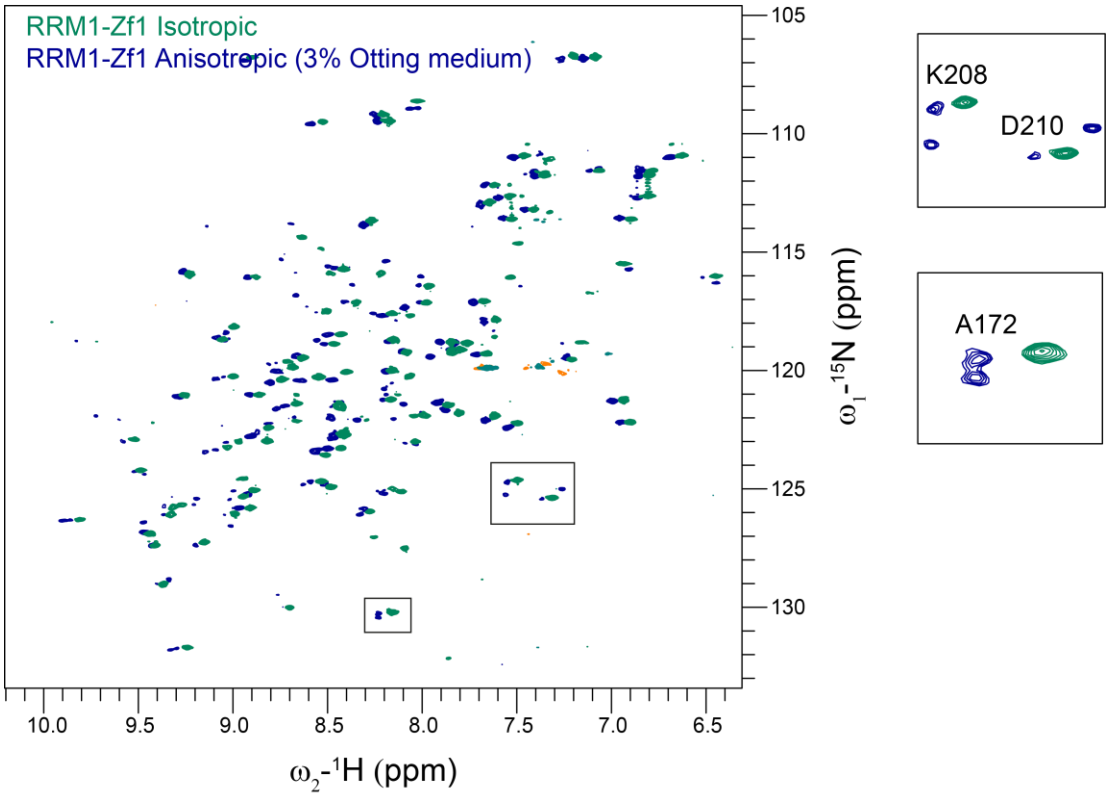


Figure 46 Multiple conformations exist in the Anisotropic wild-type RRM1-Zf1 sample

A superposition of $^{1}H,^{15}N$-HSQC spectra of wild-type RRM1-Zf1 protein under isotropic and anisotropic conditions is shown is green and blue, respectively. Zoom-ins of residues Lys 208, Asp 210 and Ala 172 are shown on the right, clearly illustrating the two sets of amide signals for these residues under anisotropic conditions.

5.4.1. Sequence alignment of RRM1-Zf1 helps to understand the underlying problem

RBM5 Zf1 is a RanBP2-type zinc finger which is defined by the consensus sequence Trp-X-Cys-X_{2-4}-Cys-X_3-Asn-X_6-Cys-X_2-Cys, where the four cysteine residues coordinate the Zn^{2+} ion. Apart from the conserved Zn^{2+} coordinating cysteine residues, there is an additional cysteine residue, Cys 191 which is highlighted in **Figure 47.** This additional cysteine residue is present adjacent to the cysteine residue involved in Zn^{2+} ion coordination. It is also not entirely conserved in different species and this position has a glycine residue in *Danio rerio* (zebrafish) and *Xenopus tropicalis* (frog) while in higher organisms, it has evolved into a cysteine residue. Interestingly, RBM10 also has a glycine residue at this position. Therefore, to test if this extra cysteine residue is actually creating the stability issues, a C191G point mutation was made in the protein.

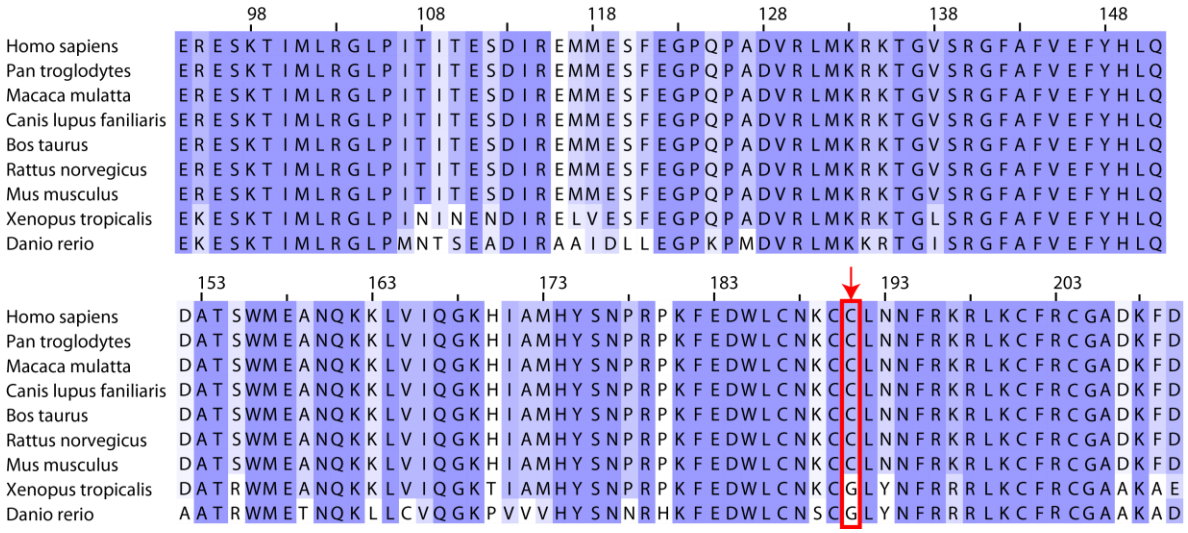


Figure 47 Sequence alignment of RBM5 RRM1-Zf1 domains from different organisms

Sequence alignment of RBM5 RRM1-Zf1 from different organisms is shown. The residues are colored according to sequence conservation. Residue position 191 is highlighted, which is a cysteine residue in higher organisms and while it is glycine in lower organisms.

As expected, upon mutation of the extra cysteine to glycine, the spectral quality of the protein improved. I was also able to record HN-RDCs and SAXS data with the mutant protein without any problems (see **section 5.2.5**). To inspect if the C191G mutant affects the integrity of the protein, ^{15}N-relaxation data was recorded on the mutant and compared to that of the wild-type protein (**Figure 48**).

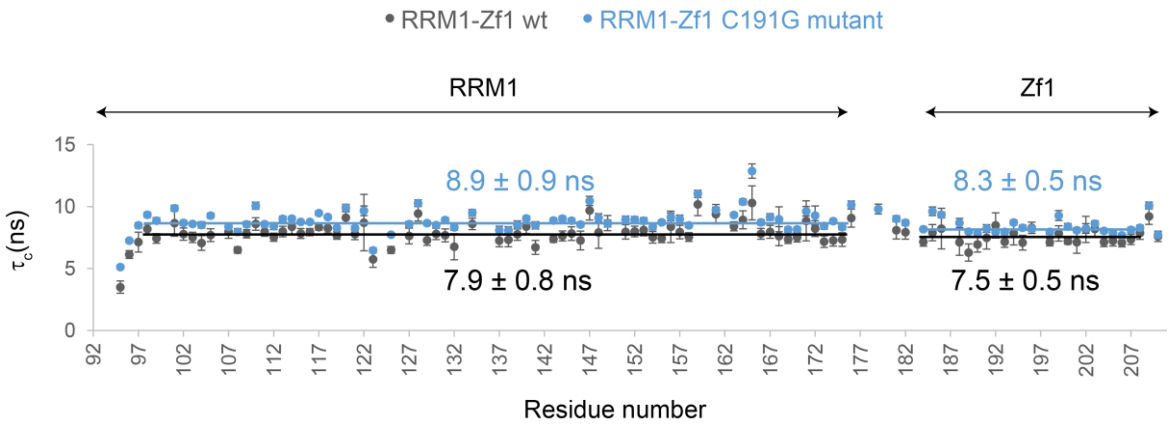


Figure 48 Comparison of ^{15}N-relaxation data for wild-type RRM1-Zf1 and C191G mutant

The total rotational correlation time (τ_c) calculated from R_1 and R_2 rates is plotted against residue numbers for wild-type RRM1-Zf1 and RRM1-Zf1 C191G mutant in grey and blue, respectively. The average ± standard deviation values are listed for each domain. The τ_c value for RRM1-Zf1 C191G mutant is ~1 ns higher than that of the wild-type protein.

The total rotational correlation time of RRM1-Zf1 C191G mutant protein is ~1 ns higher than that of the wild-type protein. This could be attributed to the propensity of the protein to dimerize at higher concentrations (SAXS data, see **section 5.2.5**). Since the C191G mutant data were measured at ~570 μM compared to ~240 μM for that of wild-type protein, the increase in τ_c could be due to concentration dependent dimerization effects. Nonetheless, it still indicates that the integrity of the protein is not compromised.

5.4.2. Zn^{2+}-Cd^{2+} exchange kinetics

Next, I wanted to understand if/how the C191G mutation contributes to the stability of the protein. To this end, I designed a very simple experiment which is based on the strength of coordination of the Zn^{2+} ion by the cysteine residues. It was postulated that the additional cysteine residue competes with its neighboring cysteine for coordinating the Zn^{2+} ion, which leads to the existence of two species or conformations of the protein (as explained above). In the absence of this additional cysteine, the competition between neighboring cysteines does not exist anymore and the protein is more stable. If this theory would hold true, then the Zn^{2+} ion would be loosely bound in the wild-type protein owing to the exchange in coordinating cysteine as opposed to the C191G mutant protein where only a single stable state would exist.

To test this hypothesis, I used a 1:1 Cd^{2+}-EDTA solution added in 5-fold excess to each wild-type RRM1-Zf1 and RRM1-Zf1 C191G mutant proteins. Since the affinities of Zn^{2+} and Cd^{2+} for EDTA are nearly identical (Martell and Smith 1974, Nowack, Kari et al. 2001, Patton, Thompson et al. 2004), an exchange of Zn^{2+} ion by Cd^{2+} ion would occur owing to the inherent higher affinity of thiolate metal ligands for Cd^{2+} than Zn^{2+} (Summers 1988). Such experiments have also been previously done to study the dynamics and metal-exchange properties of different Zinc finger proteins (Houben, Wasielewski et al. 2005). After addition of Cd^{2+}-EDTA solution to the protein sample which took a dead time of 3-4 minutes, the sample was put in the NMR tube and a series of $^{1}H,^{15}N$ -SOFAST-HMQC (Schanda, Kupce et al. 2005) spectra were recorded. Precisely, every data point was recorded after 8 min 29 sec.

As seen from **Figure 49A,** in wild-type RRM1-Zf1 protein the Zn^{2+}-Cd^{2+} exchange is quite fast, and a complete set of new amide signals representing the Cd^{2+}-bound form appear within a few minutes and in the second $^{1}H,^{15}N$-SOFAST-HMQC recorded, the exchange is already complete. This clearly shows that in the wild-type protein, the Zn^{2+} ion is quite loosely bound due to possible existence of the protein in two states where the neighboring cysteines compete with each other for Zn^{2+} coordination. On the other hand, the exchange is quite slow

in the C191G mutant protein indicating that the Zn^{2+} ion is relatively tightly bound. Of course, eventually within 2-2.5 hours, the exchange is also complete in the C191G mutant protein due to firstly, the 5-fold excess of Cd^{2+}-EDTA solution and secondly the inherently high affinity of the Zf1 for Cd^{2+} than Zn^{2+} as mentioned before. The chemical shift perturbation plot of Zn^{2+}/Cd^{2+} bound RRM1-Zf1 C191G mutant protein shows that nearly no changes occur in the RRM1 region, which is expected as only the metal ion coordination site should be effected by the Zn^{2+}-Cd^{2+} exchange (**Figure 49B)**. In the Zf1 domain, a number of residues, including those involved in Zn^{2+} coordination as well as those in close proximity to the coordination site show large CSPs (**Figure 49B**). The residues where amide signals shift to completely new positions, without any signal overlap compared to the Zn^{2+} bound form, are marked with circles and labeled in **Figure 49A**.

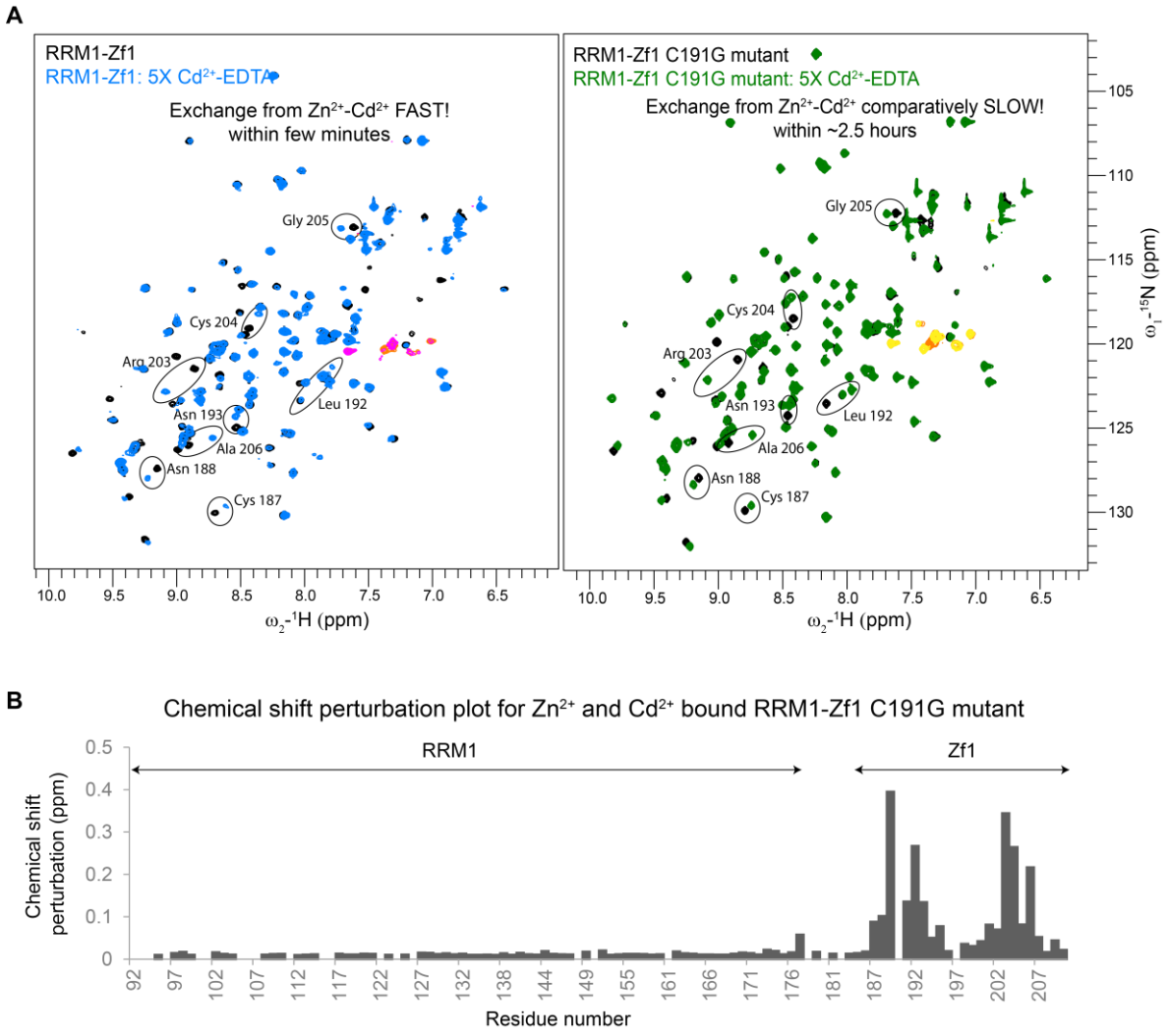


Figure 49 Zn^{2+}-Cd^{2+} exchange occurs faster in wild-type RRM1-Zf1 than C191G mutant

(A) ^{1}H,^{15}N SOFAST-HMQC spectra of wild-type RRM1-Zf1 and C191G mutant proteins bound to Zn^{2+} ion or Cd^{2+} ion is shown in black, blue in the left panel and black, green in the right panel,

respectively. Residues which show highest chemical shift perturbations and therefore, non-overlapping amide signals between the Zn^{2+} and Cd^{2+} bound forms are marked. (B) Chemical shift perturbation plot of RRM1-Zf1 C191G mutant protein when bound to Zn^{2+} ion or Cd^{2+} ion.

Since the Zn^{2+}-Cd^{2+} exchange in the RRM1-Zf1 C191G mutant protein is relatively slow, it is also possible to track the exchange kinetics using real time NMR measurements. In **Figure 50**, the Zn^{2+}-Cd^{2+} exchange curves are depicted for nine residues in Zf1 that show completely isolated sets of amide signals between the two ion bound forms. This is done only the reduce the complexity in data analysis and peak volume integration that may arise in case of overlapping amide signals in Zn^{2+} and Cd^{2+} bound forms. It is clearly demonstrates that the peak intensity of the residues in Zn^{2+} bound form drop over time, while that of the Cd^{2+} bound form increase over time.

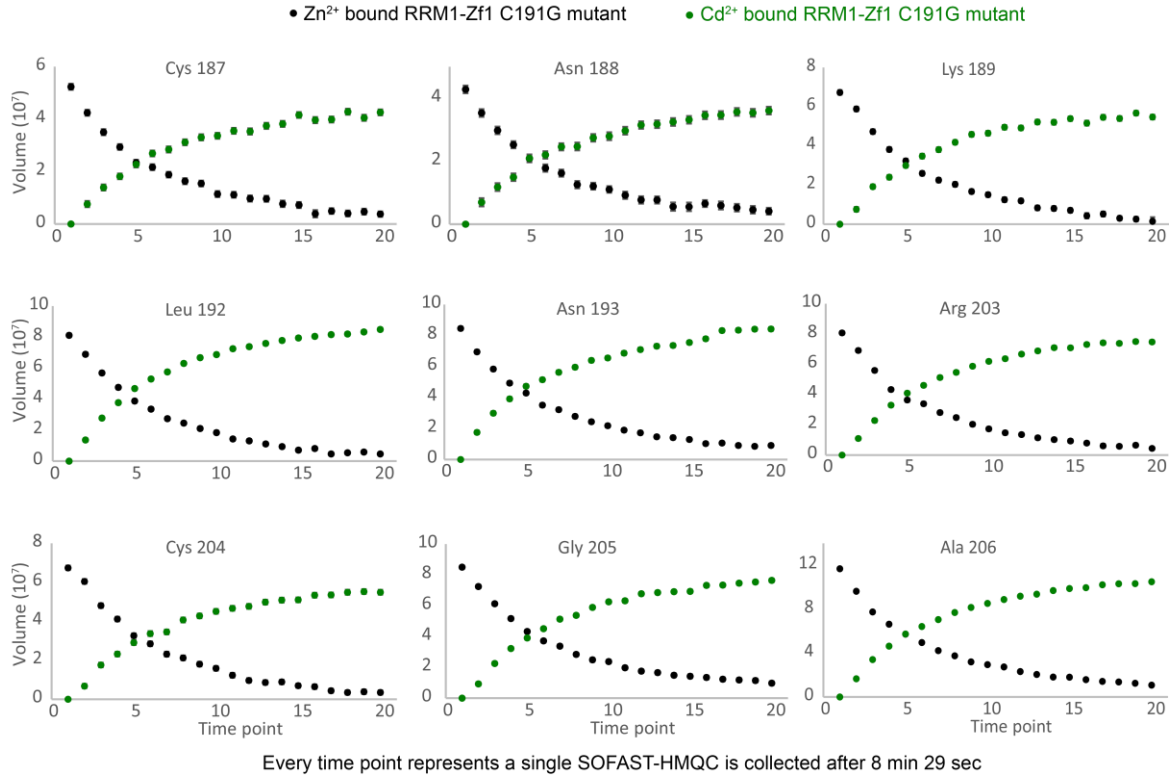


Figure 50 Zn^{2+}-Cd^{2+} exchange kinetics

NMR amide signal intensities of nine individual residues showing the highest chemical shift perturbations were integrated and plotted against time points. Every time point represents a ^{1}H,^{15}N-SOFAST-HMQC that was recorded, and is spaced 8 min 29 sec apart. The change in signal intensity of the amide signals in Zn^{2+} and Cd^{2+} ion bound protein is shown in black and green, respectively.

The differences in chemical shifts and the relatively slow metal ion exchange process allows the metal exchange kinetics to be extracted for the residues with isolated, non-

overlapping amide signals in both metal ion bound spectra (**Table 5**). Similar exchange rates are observed for Cys 187, Asn 188, Lys 189, Leu 192, Asn 193, Arg 203, Cys 204, Gly 205 and Ala 206 in each of the metal ion bound state varying from ~3.4 – 3.9 x10^{-4} s^{-1}.

Table 5 Metal exchange rates for RRM1-Zf1 C191G mutant protein derived from decreasing and increasing amide signal intensities upon exchange of Zn^{2+}-Cd^{2+}

Residue	Ion	Rate (10^{-4} s^{-1})
Cys 187	Zn^{2+}	3.72 ± 0.19
	Cd^{2+}	3.53 ± 0.19
Asn 188	Zn^{2+}	3.92 ± 0.13
	Cd^{2+}	3.35 ± 0.19
Lys 189	Zn^{2+}	3.57 ± 0.15
	Cd^{2+}	3.78 ± 0.15
Leu 192	Zn^{2+}	3.64 ± 0.09
	Cd^{2+}	3.71 ± 0.08
Asn 193	Zn^{2+}	3.70 ± 0.09
	Cd^{2+}	3.36 ± 0.23
Arg 203	Zn^{2+}	3.86 ± 0.14
	Cd^{2+}	3.52 ± 0.08
Cys 204	Zn^{2+}	3.50 ± 0.13
	Cd^{2+}	3.37 ± 0.13
Gly 205	Zn^{2+}	3.78 ± 0.07
	Cd^{2+}	3.40 ± 0.11
Ala 206	Zn^{2+}	3.89 ± 0.12
	Cd^{2+}	3.56 ± 0.08

Residues listed showing Zn^{2+}-Cd^{2+} exchange are either the metal ion coordination residues or are in close proximity to it. The exchange rates are derived from the decreasing and increasing peak intensities of residues affected due to the metal ion exchange using the Nonlinear Curve Fit (Exponential) routine in OriginPro 9.

Taken together, these metal ion exchange kinetic studies clearly demonstrate that wild-type RRM1-Zf1 protein containing the extra cysteine coordinates the Zn^{2+} ion weakly, as evidenced by the fast metal ion exchange upon addition of Cd^{2+}-EDTA solution. The possible reason for this could be due to a direct competition between the neighboring cysteines for metal ion coordination, making the metal ion readily exchangeable. This theory is further supported by the bad quality of ^{1}H,^{15}N-SOFAST-HMQC spectrum of the wild-type protein which could be a result of this exchange between the two conformations where either of the neighboring

cysteine residues coordinate the metal ion (**Figure 49A**, left panel). In contrast, in the absence of the extra cysteine residue, the C191G mutant protein coordinates the metal ion relatively tightly, with slow on/off kinetics as demonstrated by the slow metal ion exchange kinetics. Also, the spectral quality of the protein improves tremendously (**Figure 49**, right panel). Notably, the amide signal doubling upon addition of 3% PEG-hexanol alignment medium to the wild-type RRM1-Zf1 protein, which was done to record HN-RDCs is resolved in the C191G mutant protein (comparing **Figure 38** and **Figure 46**).

5.5. RNA sequence specificities for RRM1-Zf1 binding

RBM5 regulates *Caspase-2* pre-mRNA splicing via direct protein-RNA interactions to an intronic sequence upstream of ln100 element in the pre-mRNA. This intronic region consists of a C/U rich element. Therefore, in this study, all single stranded RNA sequences for RBM5 binding are derived from the C/U rich intronic element. To this end, two different RNA sequences were tested for binding to RRM1-Zf1 protein using NMR and ITC. Additionally, several point mutations in the protein were tested for effects on RNA binding affinity.

5.5.1. Zf1 specifically recognizes a GG motif

To study the RNA sequence specificities of Zf1, two different RNA sequences were titrated to the protein and the $^{1}H,^{15}N$-HSQC spectra were recorded subsequently (**Figure 51**). Negligible chemical shift perturbations were observed upon titration of a C/U rich RNA (**Figure 51**, spectrum in pink) into the protein. Contrastingly, huge chemical shift changes were observed upon titration of an RNA containing an additional GG motif (as seen in the CSP plot). The titration indicated that the formation of the complex was in intermediate-fast exchange on the chemical shift timescale, whereby some signals shifted in position with each titration point while some others simply disappeared. This clearly indicates that the Zf1 protein specifically recognizes a GG motif and does not bind to a C/U rich RNA.

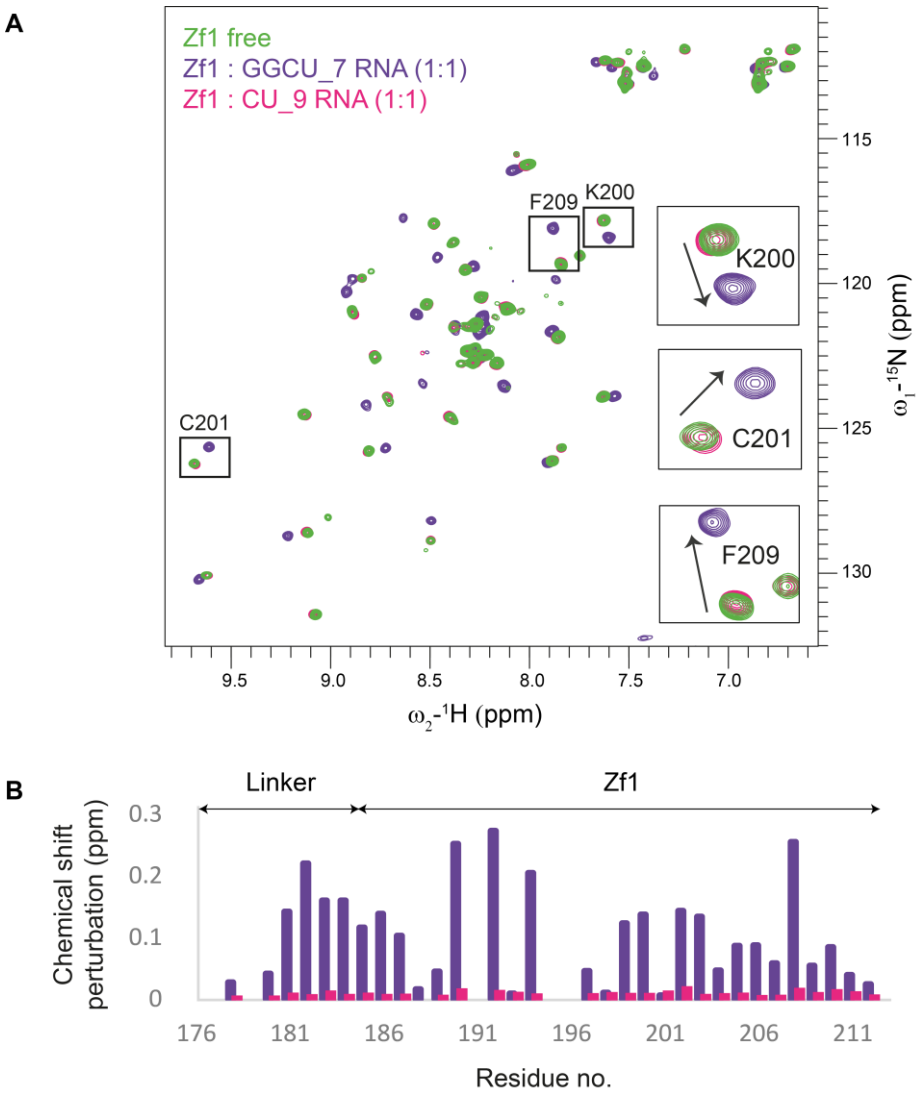


Figure 51 Zf1 requires a GG dinucleotide motif for RNA binding

(A) Superposition of ^{1}H,^{15}N-HSQC spectra of the free Zf1 domain, bound to GGCU_7 (5'-CUUGGCU-3') and CU_9 (5'-UCUCUUCUC-3') RNA is shown in green, purple and pink, respectively. Zoom-ins of three residues are shown on the right. (B) The chemical shift perturbation plot clearly demonstrates that the chemical shifts between the free and CU rich RNA bound spectra are minimal while the Zf1 domain readily recognizes the GG motif containing RNA.

This is not surprising as it was previously suggested that RanBP2-type Zinc fingers preferentially bind to a consensus sequence AGGUAA (Nguyen, Mansfield et al. 2011). By varying bases at the 2nd and 4th position in the RNA sequence and measuring the binding affinity using fluorescence anisotropy titrations, the authors showed that RBM5 Zf1has the highest affinity (~250 nM) for A**GGG**AA and it has a 1.5-, 2- and 8-fold preference for guanine over uracil, adenine and cytosine at the 4th position, respectively.

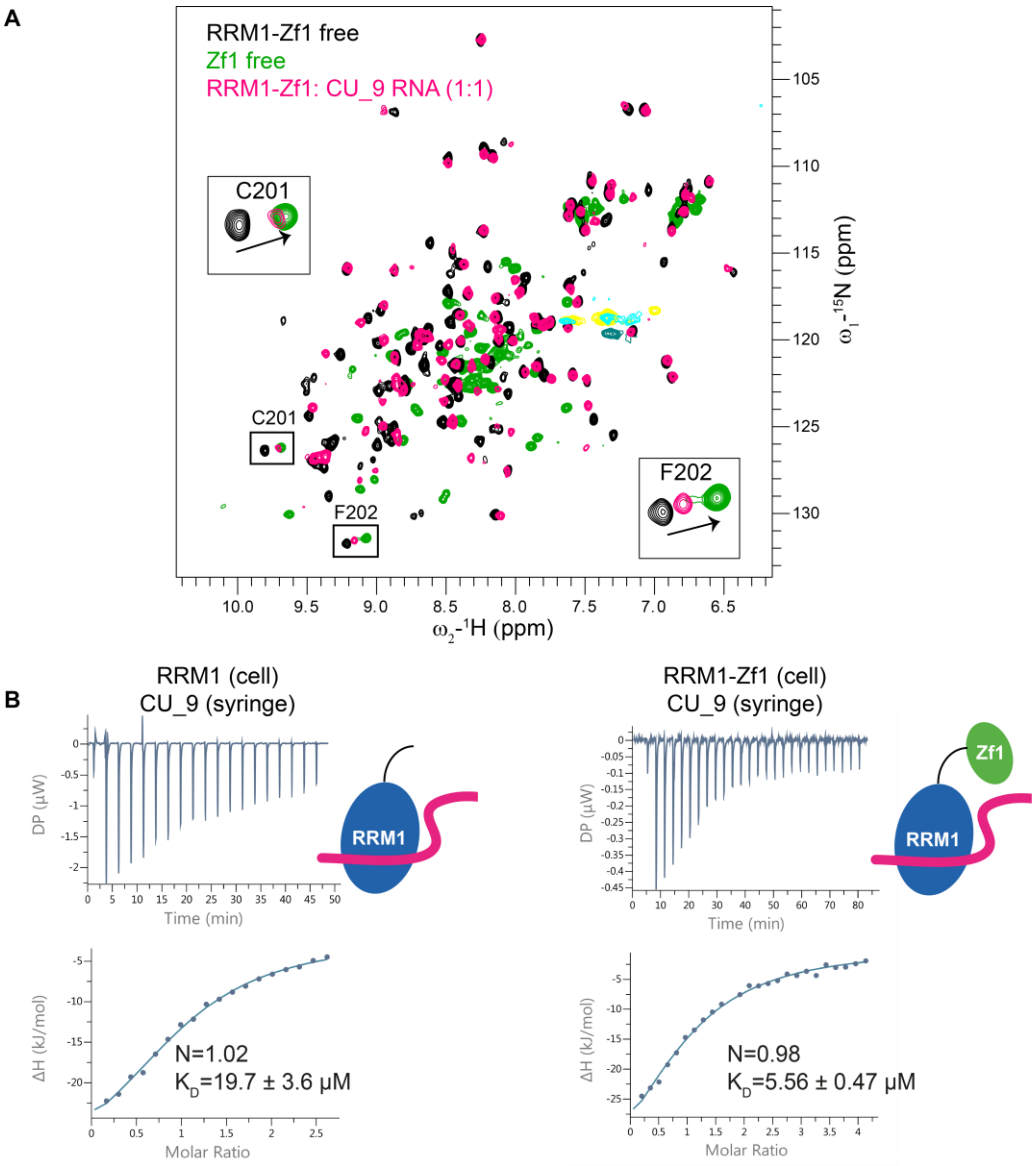


Figure 52 Zf1 does not contribute towards binding to C/U rich RNA

(A) A superposition of $^{1}H,^{15}N$-HSQC spectra of RRM1-Zf1 free, Zf1 free and RRM1-Zf1 bound to C/U rich is shown in black, green and pink, respectively. Zoom-ins of two Zf1 residues show that their peak positions in C/U rich RNA-bound RRM1-Zf1 shift towards their respective positions in Zf1 alone. (B) ITC binding isotherms of RRM1 and RRM1-Zf1 with C/U rich RNA are shown. Upon addition of Zf1 to RRM1 in the tandem construct, there is not much gain in affinity as would be expected if both the domains contribute to RNA binding, comparing left and right panels.

In order to understand what happens in the two-domain context (RRM1-Zf1), NMR titrations of the C/U rich RNA into the protein were performed. An overlay of the ^{15}N-HSQC spectra of free Zf1 and free and RNA bound RRM1-Zf1 shows that there are chemical shift perturbations observed in Zf1 domain in the RNA-bound form of RRM1-Zf1 (**Figure 52A**). Zoomed-in views for two distinct amide signals in the Zf1 domain show that the NMR signals in the RNA-bound form of Zf1 shift towards that of the free Zf1 protein. Since we already

know that Zf1 does not bind to C/U rich RNA, the shifts in Zf1 must occur due to partial displacement of the Zf1 rather than direct RNA binding.

To further confirm this, ITC binding isotherms of RRM1 and RRM1-Zf1 for binding to C/U rich RNA (**Figure 52B**) were recorded. RRM1 has a binding affinity of ~20 μM while that of RRM1-Zf1 is ~6 μM. The gain in affinity by adding an additional domain (Zf1) should be much greater than a factor of 3. It is possible that some residueal binding of Zf1 to the RNA occurs which is reflected in the 3-fold gain in affinity, only due to restraints in the conformational space. Since, Zf1 is still atached to RRM1, it could be conformationally restricted possibly leading to some interactions. Therefore, this provides further proof that Zf1 domain does not bind to a C/U rich RNA.

Now with the knowledge of binding specificities of Zf1, I tested another RNA oligo which contains both C/U rich motif for RRM1 binding and GG motif for Zf1 binding named as GGCU_12 (5'-U**GG**CUCUUCUCU-3'). This RNA oligo is also derived from the *Caspase-2* pre-mRNA intronic sequence, upstream of ln100 element. An NMR titration of the RNA into RRM1-Zf1 protein shows significant chemicals shift perturbations (**Figure 53A**). The complex formation takes place on intermediate-fast exchange timescale, whereby some amide signals shift in position while some others disappear. It is noteworthy, that almost all the RRM1 amide signals can be traced and the NMR signals broadened beyond detection mostly belong to the Zf1 domain. I also recorded an ITC binding isotherm of this RNA for binding to RRM1-Zf1 and obtained a binding affinity of ~84 nM which compared to RRM1 binding to C/U rich RNA with ~20 μM affinity, is a gain of 24-fold in affinity.

Taken together, we now know that an RNA containing C/U rich element and a GG motif can bind to both RRM1 and Zf1 domains with binding affinity in the nanomolar range.

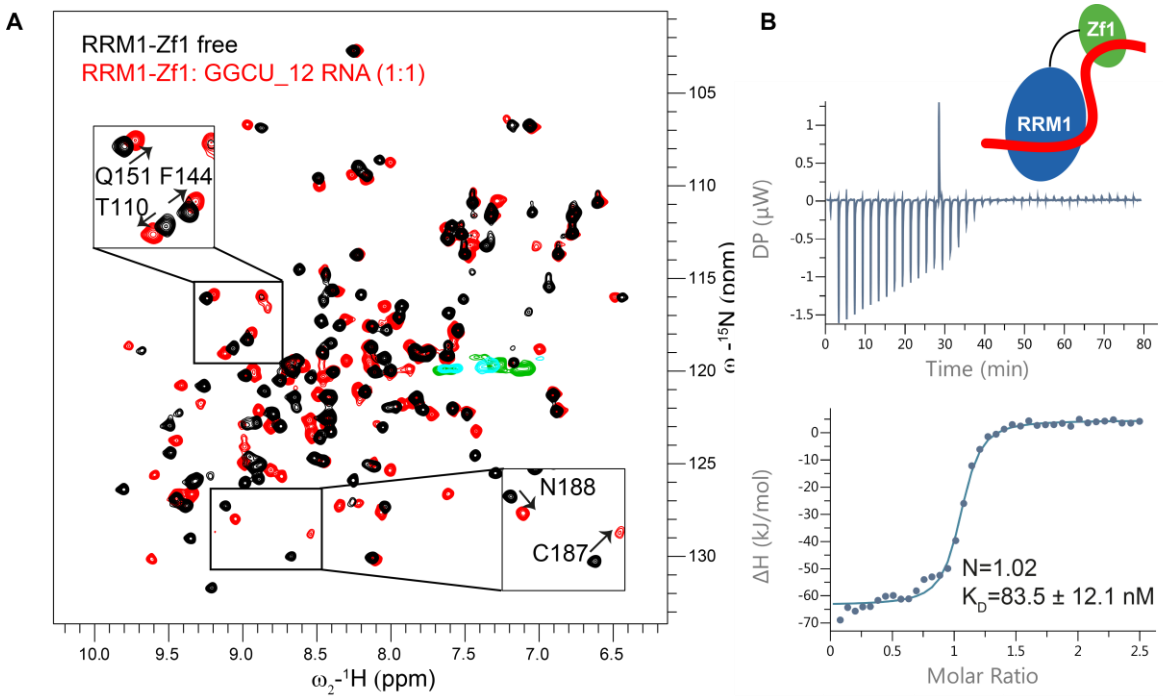


Figure 53 RRM1 binds to C/U rich RNA while Zf1 specifically recognizes 'GG motif

(A) A superposition of $^{1}H,^{15}N$-HSQC spectra of RRM1-Zf1 in free and RNA bound form (GGCU_12) is shown in black and red, respectively. It is clear from the chemical shifts that both the domains bind to RNA (top zoom-in depicts RRM1 residues while the bottom one depicts Zf1 residues). (B) ITC binding isotherm of RRM1-Zf1 with RNA oligo containing both C/U rich and GG rich motifs (GGCU_12). The binding affinity of the tandem domain increases by an order of magnitude upon comparison with that of RRM1 alone (**Figure 52B**, left panel).

5.5.2. Probing residues important for RNA binding in RRM1-Zf1 using point mutations

RRM1 is a canonical RRM domain containing a conserved RNP1 motif with two aromatic residues (Phe 142 and Phe 144) in proper conformation for RNA binding, while the RNP2 motif lacks the canonical RNA binding aromatic residue which is replaced with a methionine instead (Met 101) as shown in **Figure 54**.To test if Phe 142 and Phe 144 are important for RNA binding, a double alanine mutant (F142A/F144A) was made. Upon titration of the C/U rich RNA which binds to wild-type RRM1 with a binding affinity of ~20 µM (as determined using ITC, **Figure 52B**), no chemical shifts were observed in majority of the residues (**Figure 56A**). This clearly indicates that these two phenylalanine residues in RNP1 are extremely important for binding of RRM1 to RNA.

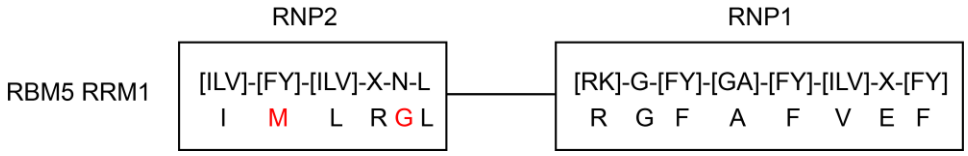


Figure 54 Comparison of canonical RNP motifs with that of RBM5 RRM1

Sequence alignment of canonical RNP motif residues with that of RBM5 RRM1 domain is shown. While RNP1 is conserved, RNP2 lacks the canonical residue while is replaced by Met 101. Additionally, at the 5th position in RNP2, the canonical Asn residue is replaced by Gly 104.

The Zf1 domain also has the canonical RNA binding aromatic residues, Phe 202 and Phe 195 and additional positively charged residues Lys 197 and Arg 198 that are exposed on one side of Zf1, supposedly forming the RNA binding interface (**Figure 55**).

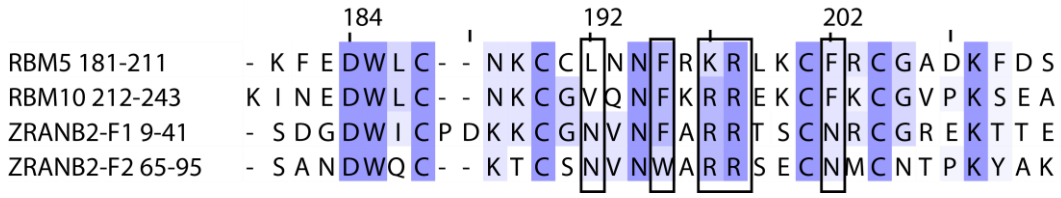


Figure 55 Sequence alignment of RanBP2-type Zinc fingers

Sequence alignment of RanBP2-type Zinc fingers is shown. Boxes highlight the residues which were demonstrated to be involved in RNA binding for ZRANB2-F2.

To gain further insights into the importance of these residues in RNA binding of Zf1, I made several point mutations in the tandem domain. The two phenylalanine residues were mutated to alanine residues (F195A and F202A) and two charge reversal mutations were made (K197E and R198E). Additionally, the phenylalanine double mutant from RRM1 was also cloned in the tandem domain construct (F142A/F144A). All the mutant proteins were expressed and purified as the wild-type protein and using $^{1}H,^{15}N$-HSQC spectra, it was ascertained that the mutants were properly folded.

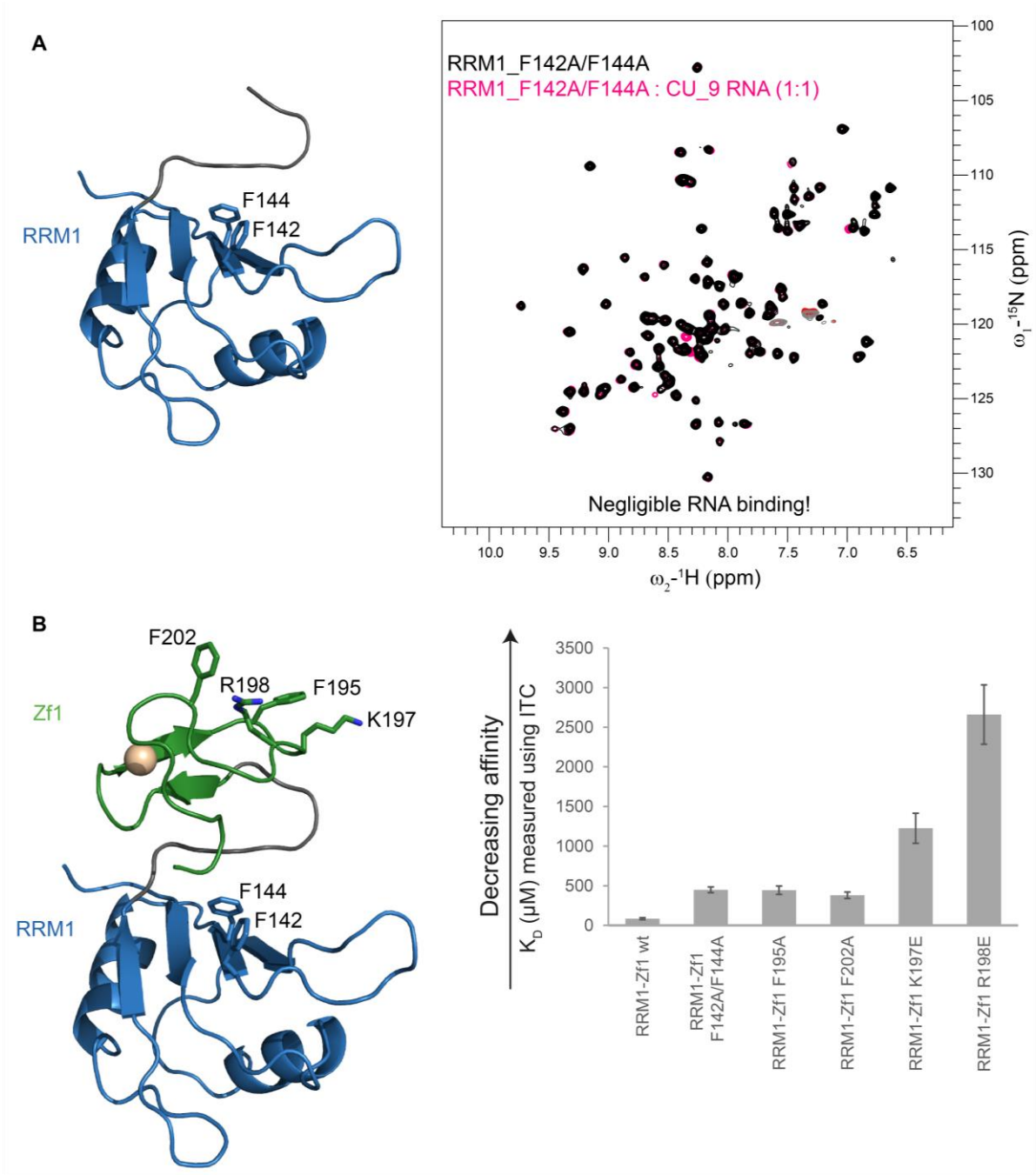


Figure 56 RRM1-Zf1 residues involved in RNA binding

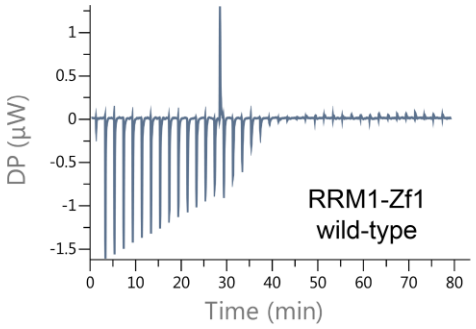
(A) A double mutant RRM1 F142A/F144A does not bind RNA as seen in the superposition of ^{1}H,^{15}N-HSQC spectra of free and RNA (CU_9) bound protein in black and pink, respectively. F142 and F144 are the two canonical RNA binding residues present in RNP1 motif as shown on the left panel. (B) Several point mutations are made in the RRM1-Zf1 tandem domain construct including F142A/F144A in RRM and F195A, F202A, K197E and R198E in Zf1. The position of the mutated residues are shown on the left panel. Binding affinities of wild-type RRM1-Zf1 and the mutants for RNA (GGCU_12) obtained from ITC are plotted in the right panel.

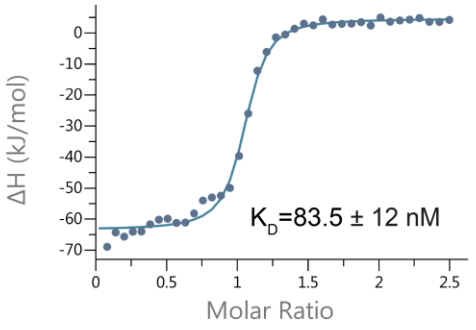
Next, ITC binding isotherms were recorded for all the mutant proteins for binding to RNA containing both C/U and GG motifs (GGCU_12) and compared with that of RRM1-Zf1

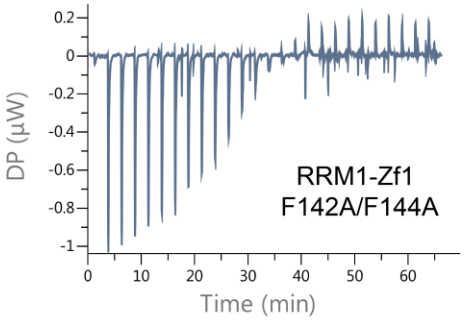
wild-type protein (**Figure 56B**). A ~5-fold drop in affinity is observed upon point mutation of the aromatic residues (F142A/F144A and F195A) while a ~4-fold decrease occurs due to the point mutation F202A. On the other hand, a much greater effect is seen in the charge reversal mutants where a ~15-fold decrease in affinity is observed upon mutation of K197E and a ~32-fold decrease is observed in case of R198E. This is not surprising as the charge reversal mutations may not just cause loss of hydrogen bonds but also charge repulsions with the negatively charges RNA.

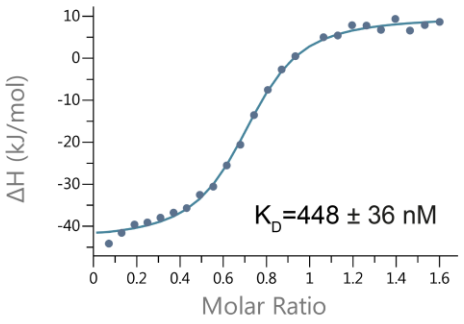
The raw ITC binding isotherms are shown in **Figure 57**. For all the measurements, the respective protein (wild-type or mutant RRM1-Zf1) is loaded in the cell while GGCU_12 RNA (5'-U**GG**CUCUUCUCU-3') is loaded in the syringe and sequentially titrated into the cell.

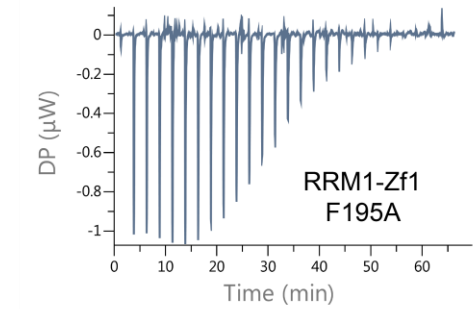
Protein in cell
RNA in syringe - GGCU_12
(5'-U**GG**CUCUUCUCU-3')

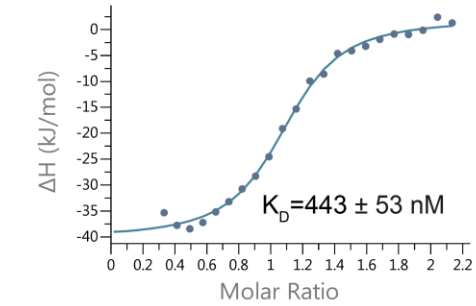


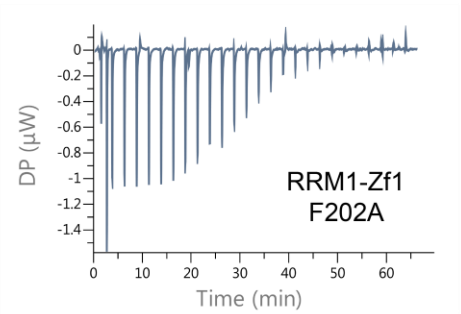


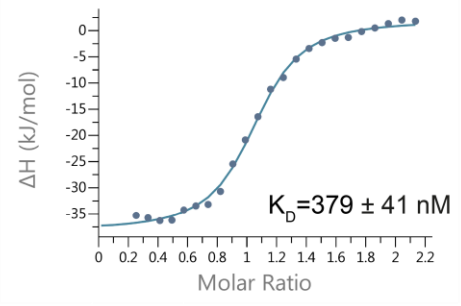












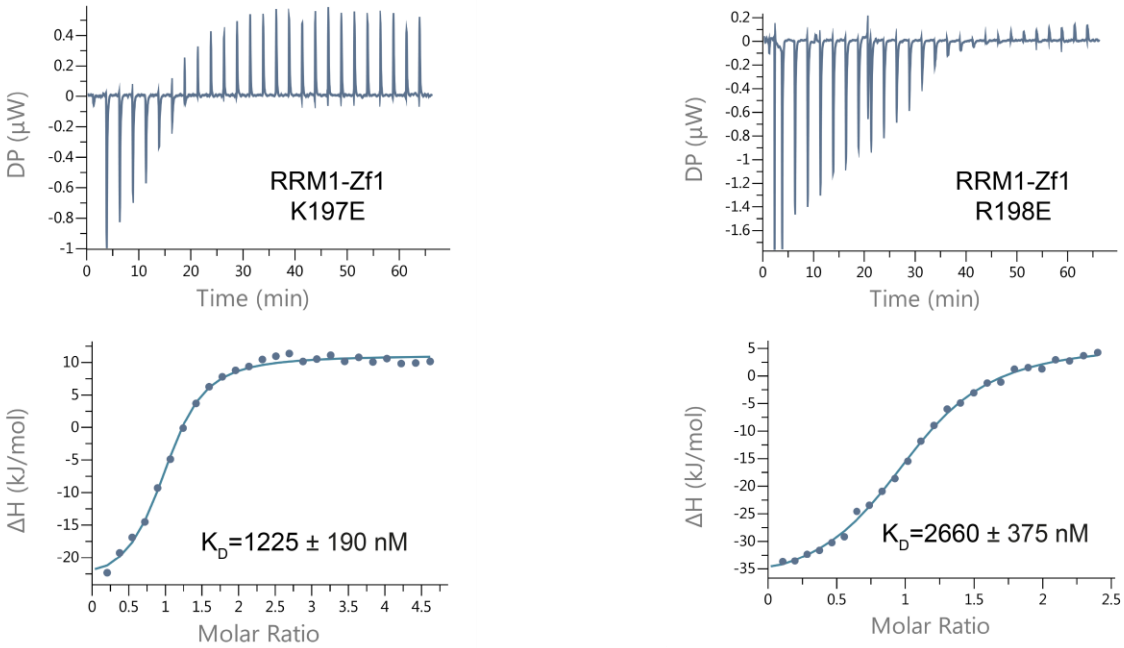


Figure 57 ITC binding isotherms for wild-type RRM1-Zf1 and mutants

ITC binding isotherms of wild-type RRM1-Zf1 and mutants (F142A/F144A, F195A, F202A, K197E and R198E) are shown and the respective dissociation constants are indicated. In all these experiments, the respective protein is present in the cell and the RNA ligand (GGCU_12) is titrated from the syringe into the cell.

5.6. Structural changes in RRM1-Zf1 upon RNA binding

Initial NMR based RNA titrations of the GGCU_12 RNA into RRM1-Zf1 C191G tandem domains showed considerable chemical shift perturbations in both RRM1 and Zf1 domains, although larger chemical shifts were observed in the Zf1 (**Figure 58A, B**). Minor structural re-organisation in the protein can also contribute to the magnitude of CSPs in the Zf1 domain, in addition to direct binding effects. It is noteworthy that the stacking interactions between the RRM1 RNA binding residues and Phe 209 from Zf1 present in the free form of the tandem domain would be destabilized in the presence of RNA (**Figure 58C**). This would be necessary as Phe 209 directly blocks the RNA binding interface of RRM1 domain. Such changes in the Zf1 domain upon RNA binding would then be reflected in the CSP data where chemical shifts would arise not only due to RNA binding but also due to strcutural re-organisation.

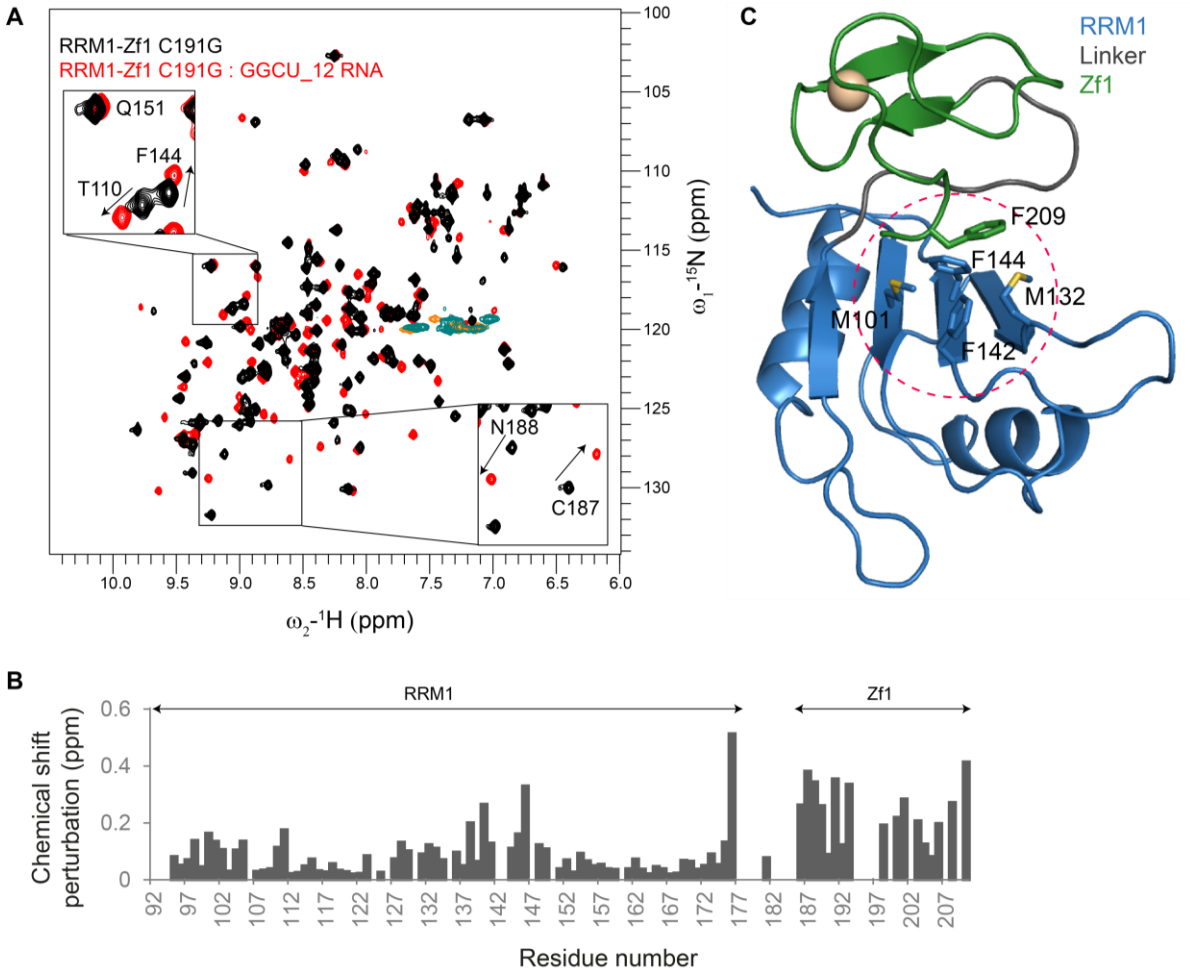


Figure 58 Chemical shift perturbations in RRM1-Zf1 C191G upon RNA binding

(A) Overlay of 1H,15N-HSQC spectra of RRM1-Zf1 C191G mutant protein in its free and RNA bound form is shown in black and red, respectively. (B) Chemical shift perturbation plot between the free and RNA bound RRM1-Zf1 C191G protein is shown. Larger CSPs are observed in the Zf1 domain. (C) Stacking interactions between the RRM1 core RNA binding residues and Phe 209 residue from the Zf1 that block the RNA binding interface of RRM1 are highlighted with a red dotted circle.

Next, I wanted to acquire ^{15}N-relaxation data as well as SAXS curves of the protein in its free and ligand bound form, which could provide substantial information depending upon the degree of structural alterations caused due to ligand binding. It is often seen in case of multi-domain proteins that if the individual domains tumble independently in solution in the free form, their motion is partially restricted in the ligand bound form. Similarly, in case of SAXS, it is often observed that multi-domain proteins where the individual domains are connected via long flexible linkers adopt an extended conformation in the free form of the protein, due to lack of enough inter-domain contacts. On the other hand, a more compact shape is achieved if all the domains bind to a small RNA ligand, decreasing the maximum dimension of the protein. I

therefore measured ^{15}N-relaxation data as well as SAXS on the RNA (GGCU_12) bound form of RRM1-Zf1 C191G mutant protein and compared it with its free form (**Figure 59**).

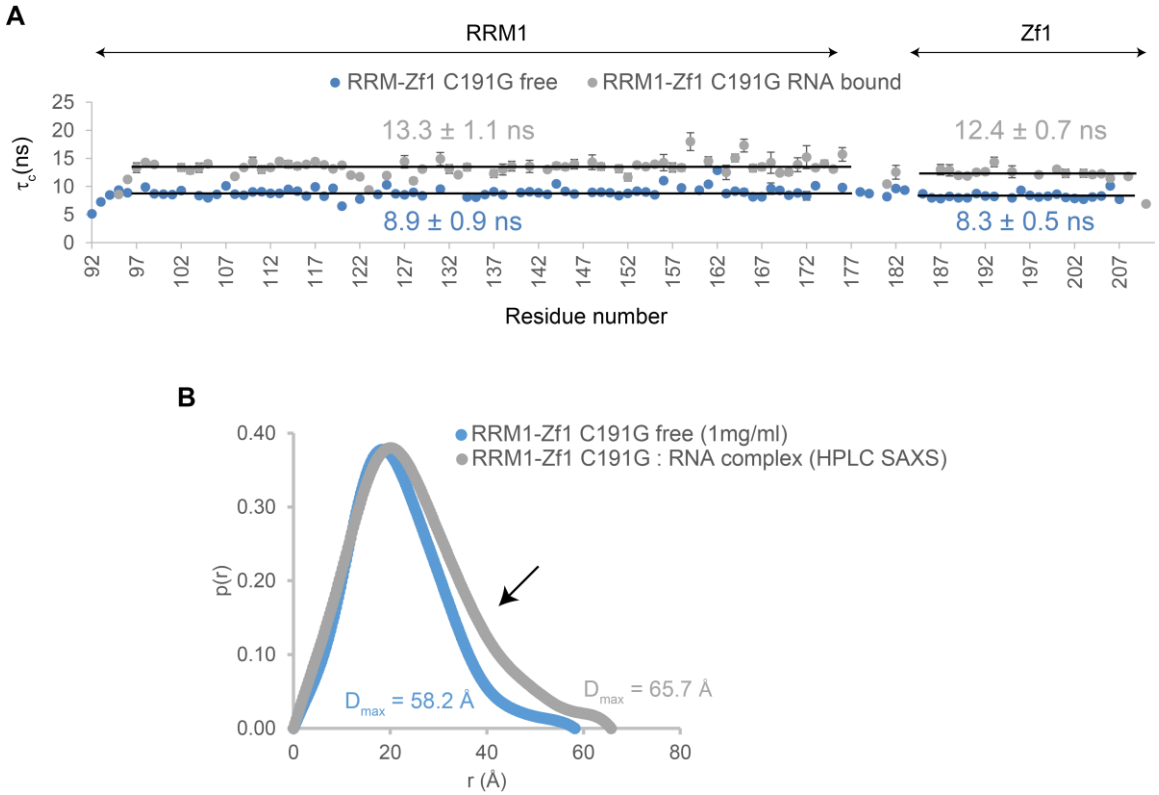


Figure 59 Relaxation and SAXS analysis of RRM1-Zf1 C191G mutant : RNA complex

(A) ^{15}N-relaxation data of free RRM1-Zf1 C191G tandem domains and bound to GGCU_12 RNA are shown in blue and grey, respectively. The average ± standard deviation values for total correlation time are listed for each domain As expected, an overall increase in the τ_c is observed for the protein-RNA complex. A slight increase in the difference between τ_c values of individual domains between free/protein-RNA complex is observed, which might point towards some degree of flexibility of the Zf1 in the presence of RNA. (B) p(*r*) curve showing maximum pairwise distribution of RRM1-Zf1 C191G mutant in complex with GGCU_12 RNA is compared with that of the free form, in grey and blue, respectively. D_{max} is indicated for the respective SAXS curves. The arrow indicates the presence of an extended conformation of the protein-RNA complex as opposed to a comparatively compact shape of the free protein.

R_1 and $R_{1\rho}$ (used for calculation of R_2) experiments were recorded on ^{15}N-labeled RRM1-Zf1 C191G protein-RNA complex and the R_2/R_1 ratio was used for calculation of the total correlation time (τ_c). Since the total correlation time of a biomolecule depends on its size, it is easy to determine if the domains tumble together in free/RNA-bound form behaving as a single moiety regardless of the differences in molecular weight of the individual components or independently. In case of free RRM1-Zf1 C191G protein, τ_c is approximately the same for the individual domains, ~8.9 ns for RRM1 (9 kDa) and ~8.3 ns for Zf1 (3.5 kDa) as shown in

Figure 59A. τ_c should be approximately 0.6 times the molecular weight of the protein, and τ_c for both RRM1 and Zf1 are much higher than the expected value from the molecular weight. Therefore, it can be safely concluded that like wild-type RRM1-Zf1 protein, in the C191G mutant protein the RRM1 and Zf1 domains tumble together in solution in the free form.

Strikingly, even for the tandem domain construct (~14 kDa), a τ_c value of 8.4 ns would be expected, which is lower than the observed values. As already mentioned (**section 5.2.5**), the SAXS data for RRM1-Zf1 C191G free protein shows a concentration dependent increase in I_0 indicative of possible dimerization. This might be a possible explanation for a slightly higher value of τ_c than expected.

Upon addition of RNA, the overall correlation times of the protein increase from ~8.9 ns to ~13.3 ns for RRM1 and ~8.3 ns to ~12.4 ns for Zf1. It indicates that the two domains tumble together also in the RNA bound form. It is noteworthy that the difference in the absolute values of τ_c for RRM1 and Zf1 in the free and RNA bound forms increases from 0.6 ns to 0.9 ns. Although this is a very slight increase which is within the error of measurement, it seems as though the coupling between the two domains in the RNA bound form slightly loosens thereby increasing the gap between their τ_c values. This would fit well with the NMR titration data where the possibility of minor structural re-orientation in the Zf1 domain is observed.

Next, measurement of the SAXS data for the RNA bound form of RRM1-Zf1 C191G protein was made on BioSAXS BM29 at ESRF, Grenoble. A protein-RNA sample with 1:1.2 excess of RNA was loaded onto a size exclusion column directly coupled to the sample inlet for data collection. A single measurement was made via the HPLC-SAXS method. The p(*r*) curve of the protein-RNA complex is plotted to indicate maximum pairwise distance distribution D_{max}=65.7 Å (**Figure 59B**) and compared with that of the free protein at lowest concentration (1 mg/ml).

Table 6 SAXS data collection and data processing statistics for RRM1-Zf1 C191G: RNA complex

Parameters	RRM1-Zf1 C191G: RNA complex HPLC-SAXS
Data-collection	
Instrument	BioSAXS BM29 ESRF
Beam geometry	10 mm slit
Wavelength (Å)	0.9919
q range ($Å^{-1}$)	0.0166-0.494
Exposure time per frame (s)[a]	1
Concentration (mg ml^{-1})	Unknown, due to use of HPLC-SAXS
Temperature (°C)	20
Structural parameters	
$I_{(0)}$ (cm^{-1}) [from p(r)]	5.95 ± 00
R_g (Å) [from p(r)]	10.93 ± 0.00
$I_{(0)}$ (cm^{-1}) [from Guinier]	5.97 ± 0.015
R_g (Å) [from Guinier]	10.93 ± 0.03
D_{max} (Å)	65.7
Porod volume estimate ($Å^3$)	25720
Software employed	
Primary data reduction	BsxCuBE
Data processing	PRIMUS

It is clear that there is no significant change in the D_{max} of the free versus RNA bound forms of the protein, although a marginal increase of 7.5 Å is observed. This is somewhat unexpected given the fact that often multi-domain proteins become more compact upon RNA binding. Nevertheless, since we know that the protein already exists in its compact state in the free form (due to intermolecular contacts between RRM1, linker and Zf1), it is possible that the protein adopts a slightly extended or open conformation. The formation of an extended conformation is indicated by the differences between free and RNA-bound pairwise distance distribution, p(r), curves at higher r values, as indicated by the arrow in **Figure 59B**.

The possibility of existence of an extended conformation as seen from the SAXS data would be complementary to the ^{15}N-relaxation data where an increase in the difference between τ_c values of the two domains is observed in the RNA bound form.

Chapter 6: Structural and functional analysis of RBM5 RNA binding triple domains

6.1. Preliminary analyses of RRM1-Zf1-RRM2 C191G protein

Several attempts to purify the triple domain (RRM1-Zf1-RRM2) protein using different tags and purification protocols failed. Nevertheless, after gaining knowledge that the extra cysteine residue in the Zf1 domain makes the protein unstable using Zn^{2+}-Cd^{2+} exchange kinetic studies, the same point mutation (C191G) was introduced in the context of the triple domain. The protein was cloned and expressed with Thioredoxin TEV cleavable fusion tag in *E.coli* BL21 cells and the cells were grown at 37 °C before induction and at 18 °C after IPTG induction for 16-20 hours (as described in Methods section). During purification over Zn^{2+} affinity column, it was realized that the protein is unstable after the removal of imidazole which is usually used for elution of the protein from the column. To circumvent this problem, a step-wise pH elution of the protein was carried out (see Methods section for details). After further purification using ion exchange and size exclusion chromatography, stable and pure fractions of the protein were obtained. It was noticed that like the tandem domain RRM1-Zf1 (wild-type /C191G mutant), the triple domain C191G mutant protein in its free form could also only be stabilized at high salt concentration (400 mM NaCl). The following sub-sections describe the initial characterization of the protein in detail.

6.1.1. Initial insights from NMR spectra of the free protein

^{15}N, ^{13}C labeled protein was expressed and purified from M9 minimal medium. The first ^{1}H,^{15}N-HSQC that was recorded indicated a good spectral dispersion as is expected for a protein containing a significant amount of β-strand elements (**Figure 60, left panel**). Next, the standard 3D-backbone assignment experiments including HNCACB, CBCA(CO)NH and HNCO were recorded. The spectral quality was not high owing to the size of the protein and the fact that the protein precipitated during the course of the assignment experiments. Still, using this data in addition to the assignments of the individual domains it was possible to assign the protein. It became immediately clear that the signals belonging to the linker have the highest intensity compared to the rest of the protein followed by those of RRM2 domain. On the other hand, some signals from RRM1and Zf1 are quite weak possibly owing to exchange broadening processes. Nevertheless, ~87% backbone assignment completeness was achieved whereby ~85% residues in RRM1, ~65% residues in Zf1 and ~99% residues in RRM2 were assigned.

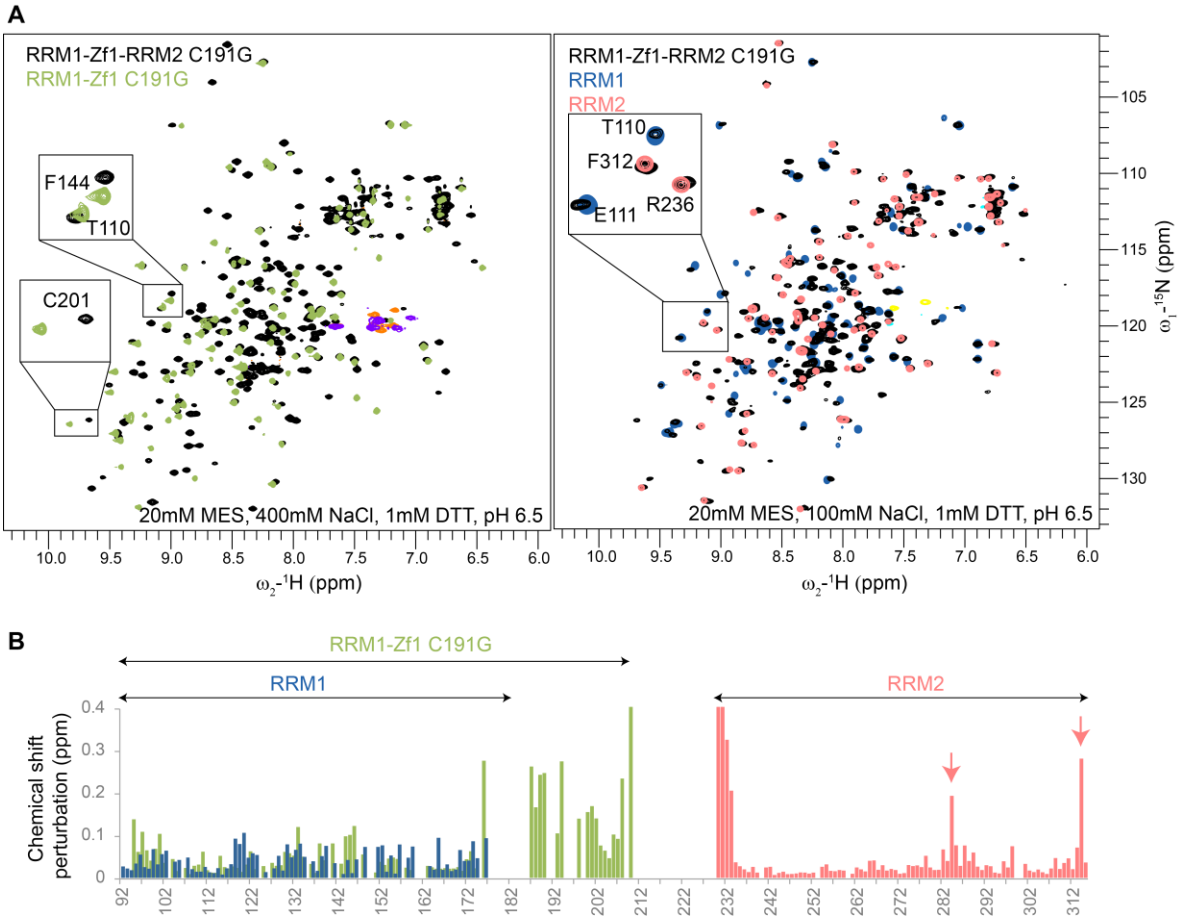


Figure 60 Overlay of ^{1}H, ^{15}N-HSQC spectra of single, tandem and triple domain constructs

(A) Overlay of ^{1}H,^{15}N-HSQC spectra of triple domain RRM1-Zf1-RRM2 C191G and tandem domain RRM1-Zf1 C191G mutant protein is shown on the left in black and green, respectively. On the right side, an overlay of ^{1}H,^{15}N-HSQC spectra of the triple domain protein with that of single domains RRM1 and RRM2 is shown in black, blue and pink, respectively. The buffers in which the spectra were recorded are denoted. (B) The chemical shift perturbation plots indicating the differences between the single and tandem domains compared to the triple domain are shown, colored according to the respective ^{1}H,^{15}N-HSQC spectra. Arrows indicate residues showing significant CSPs in RRM2 which lie in spatially surrounding region of the N-terminus of RRM2.

To gain further insights into differences between single/tandem domains versus triple domain construct, an overlay of the respective ^{1}H,^{15}N-HSQC spectra was made (**Figure 60**). Residues at the N-terminus of RRM2 such as Cys 230 and Asp 231 and the C-terminus of RRM1-Zf1 such as Asp 210 show huge CSPs due to the construct boundaries and therefore, go above the scale in the CSP plot in **Figure 60**. RRM2 shows very little CSPs in general, although certain residues (Ala 284 and Lys 314) spatially surrounding the N-terminus of the protein show significant CSPs as indicated by arrows in **Figure 60**. On the other hand, residues in RRM1 show CSPs throughout the protein (as shown in blue). This could indicate that the changes in RRM1 comparing single vs triple domain constructs could be due to additional

contacts between RRM1 and other components in the context of the triple domain. Additionally, upon comparison of the tandem domain construct (RRM1-Zf1 C191G) with that of the triple domain (RRM1-Zf1-RRM2 C191G) it becomes apparent that there are significant changes especially in the Zf1 domain. This could either be an indication of disturbance of RRM1-Zf1 interface leading to large CSPs in the Zf1 or additional contacts made between the linker and Zf1 which could also lead to substantial CSPs.

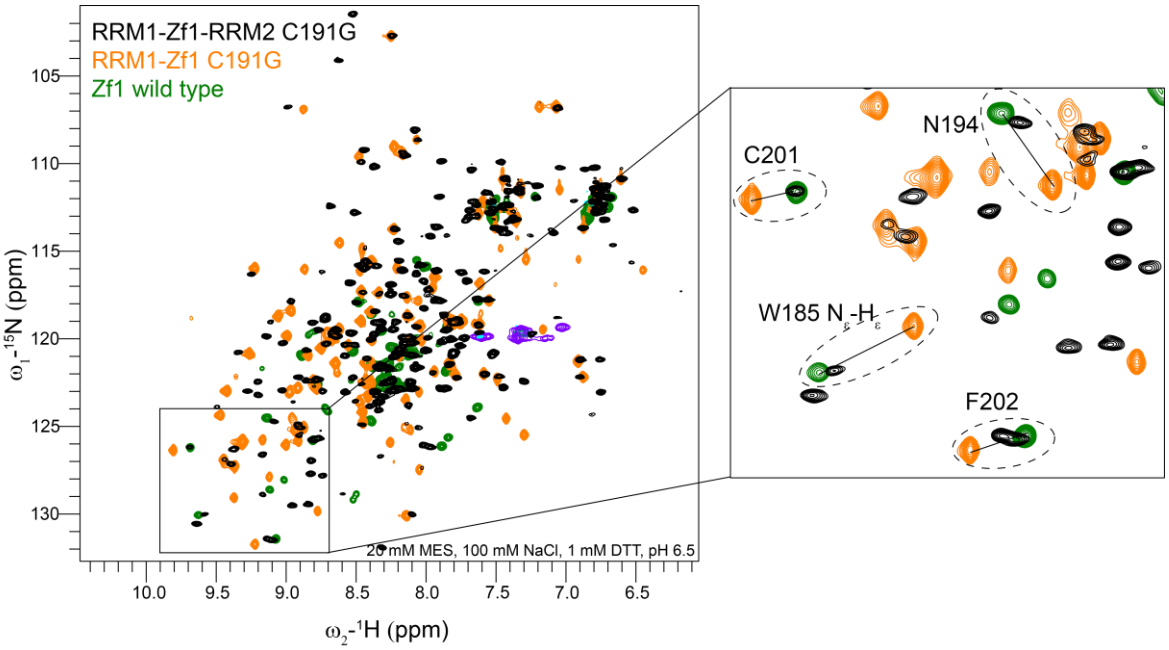


Figure 61 Overlay of ^{1}H, ^{15}N-HSQC spectra of single domain Zf1, tandem and triple domains

Overlay of ^{1}H,^{15}N-HSQC spectra of single domain Zf1 (wild type) with tandem domain RRM1-Zf1 C191G and triple domain RRM1-Zf1-RRM2 C191G mutant proteins is shown in green, orange and black, respectively. A zoomed-in region of the spectra is shown on the right with residues demonstrating clear changes between the different constructs highlighted.

A direct comparison between ^{1}H,^{15}N-HSQC spectra of the single domain Zf1 (wild type) with that in tandem domain and triple domain C191G mutant proteins shows that the residues of the Zf1 domain have similar amide signal positions between the single domain and triple domain constructs while being quite different in the tandem domain construct (**Figure 61**). This would be indicative of the validity of the first hypothesis made above, pointing towards the disturbance of the RRM1-Zf1 inter-domain interface in the triple domain where the Zf1 domain clearly shows shifts similar to that in the single domain. However, the possibility of additional contacts being made between the linker connecting Zf1, RRM2 and Zf1 which might interfere with the interaction between RRM1 and Zf1 cannot be ruled out. It would be interesting to

compare $^{1}H,^{15}N$-HSQC spectra of constructs containing linker extensions on both sides, i.e. RRM1-Zf1-linker and linker-RRM2, in the future.

6.1.2. Characterization of RNA binding properties of RBM5 triple domains

Next, I wanted to study the RNA binding characteristics of RRM1-Zf1-RRM2 C191G protein. I used the prior knowledge about the RNA sequence specificities for recognition via different domains. I know that the Zf1 domain requires a GG motif for RNA binding while RRM1 can bind to a C/U rich RNA sequence. On the other hand, Song et al. (Song, Wu et al. 2012) showed that RRM2 can bind to a C/U or A/G rich RNA sequence with similar binding affinity. They had also used the intronic region upstream of ln100 element in *Caspase-2* pre-mRNA to derive the RNA sequences (5'-CUCUUC-3' and 5'-GAGAAG-3').

The RNA sequence used for characterization of RNA binding properties of RRM1-Zf1 is GGCU_12 (5'-U**GGCU**CCUUCUCU-3'), consisting of 12 bases which should be long enough for binding to all three domain as well. I designed another 13 bases long RNA sequence by extending the 5' site and shortening the 3' end- ne_GGCU_13 (5'-GAACUU**GGCU**CUU-3'). The idea behind testing these two different RNA sequences for binding to the triple domain construct was to achieve domain reorientation in case of linear recognition of the RNA. This simply means that due to the specificity in RNA recognition provided by Zf1, the other two domains have to re-organize with respect to each other depending on the availability of RNA bases as illustrated in **Figure 62**.

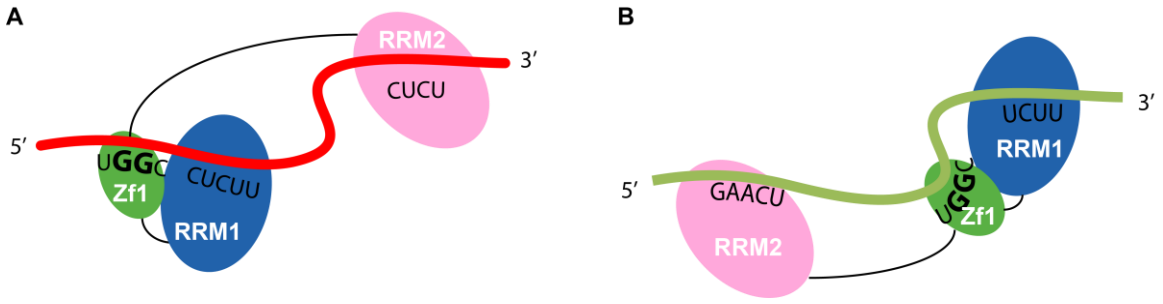


Figure 62 Hypothetical model of RNA recognition

A hypothetical model of RNA recognition by the triple domain construct whereby base-pairing specificity is provided by the Zf1 domain which recognizes the GG motif is for GGCU_12 RNA (A) and ne_GGCU_13 (B). RRM1, Zf1 and RRM2 are color coded in blue, green and pink while GGCU_12 and ne_GGCU_13 RNAs are denoted with red and light green colors, respectively.

To learn if the two different RNA sequences are recognized differently by the triple domain (RRM1-Zf1-RRM2 C119G) protein, titrations were made into the protein and $^{1}H,^{15}N$-

HSQC spectra were subsequently recorded (**Figure 63**). The complex formation between the protein and RNA occurs in fast to intermediate regime on the NMR chemical exchange timescale whereby some signals shift in position with the titration point while some others disappeared during the titration. There are also a handful residues which show two amide signals with different relative intensities. The overall peak intensities for the residues becomes weak owing to exchange broadening processes.

A comparison of the CSP plots of the RRM1-Zf1-RRM2 C191G protein bound to GGCU_12 RNA and ne_GGCU_13 RNA indicates that apart from a few local differences in residues Met 132, Val 138 and Phe 144 of RRM1, both the RNA sequences show similar overall chemical shift perturbation plots. It is noteworthy that if the different RNA sequences would have caused a domain re-organization as I expected, it would be clearly visible in the linker between Zf1 and RRM2 (residues 211-230).

It is therefore unlikely that the RNA is recognized in a linear fashion by the protein. Additionally, we also learn that the presence of purines instead of pyrimidines at the 5'-end of the RNA sequence is also tolerated and the bases are recognized in a similar manner. From the CSP plots, it also becomes clear that there are chemical shift changes observed in the linker between Zf1 and RRM2 domains, indicating either a direct interaction of the linker with RNA or due to allosteric changes.

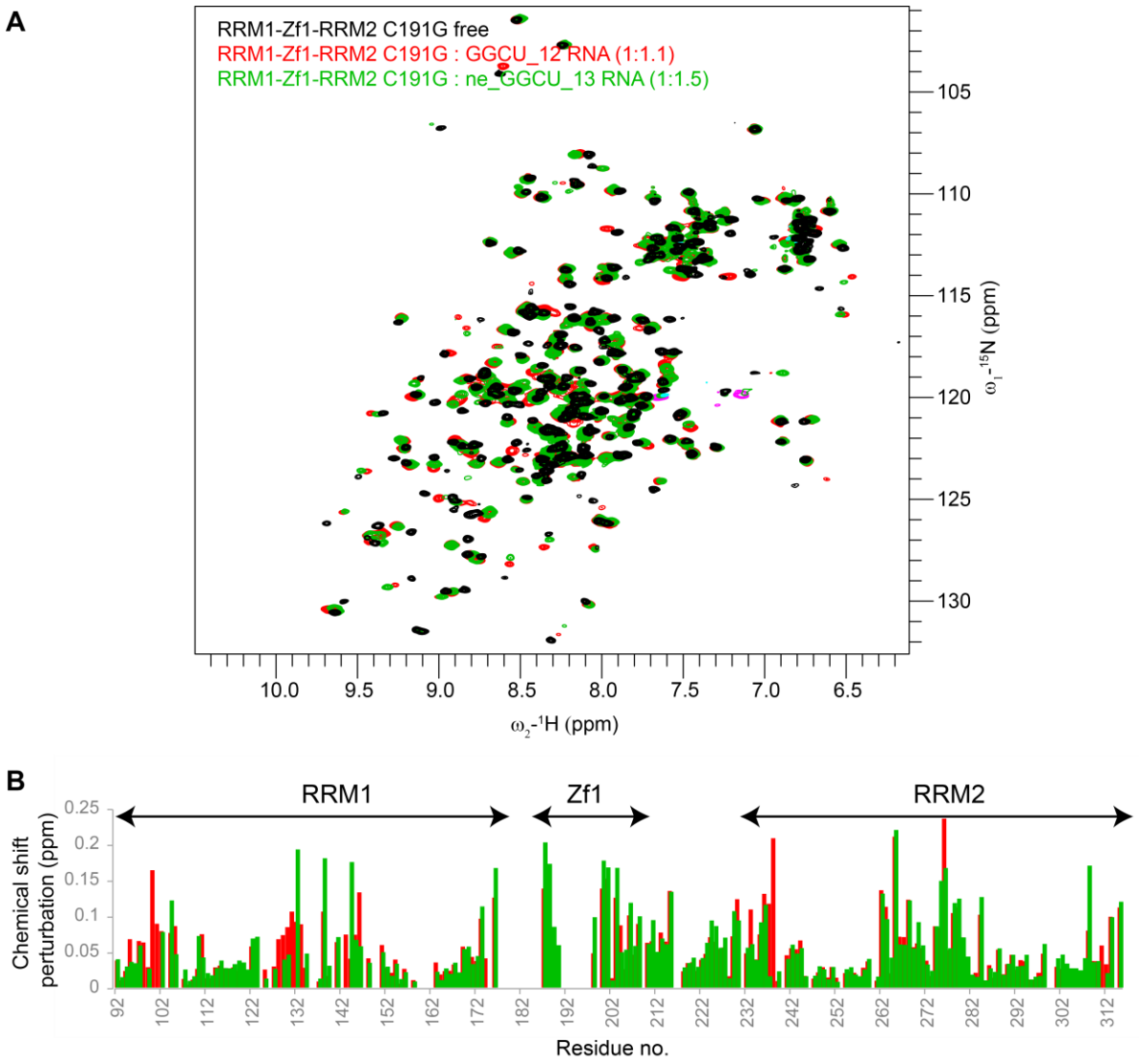


Figure 63 Overlay of free and RNA bound ^{1}H,^{15}N-HSQC spectra of RRM1-Zf1-RRM2 C191G

(B) An overlay of ^{1}H,^{15}N-HSQC spectra of the triple domain construct in free form, in complex with GGCU_12 and ne_GGCU_13 RNAs is shown in black, red and green, respectively. (B) The chemical shift perturbation plots between the free and different RNA bound forms of triple domain construct are shown, color coded according to the ^{1}H,^{15}N-HSQC spectra in red and green , respectively.

Next, I investigated if the C/U rich RNA recognition is conserved in the single domains (RRM1 and RRM2), in the context of the triple domain RRM1-Zf1-RRM2 C191G. To this end, an overlay of CSP plots of RRM1/RRM2 bound to C/U rich RNA and RRM2-Zf1-RRM2 C191G bound to an RNA containing C/U rich and GG motifs (GGCU_12) was made (**Figure 64A**). The overall pattern of CSPs is conserved between the single versus triple domain constructs.

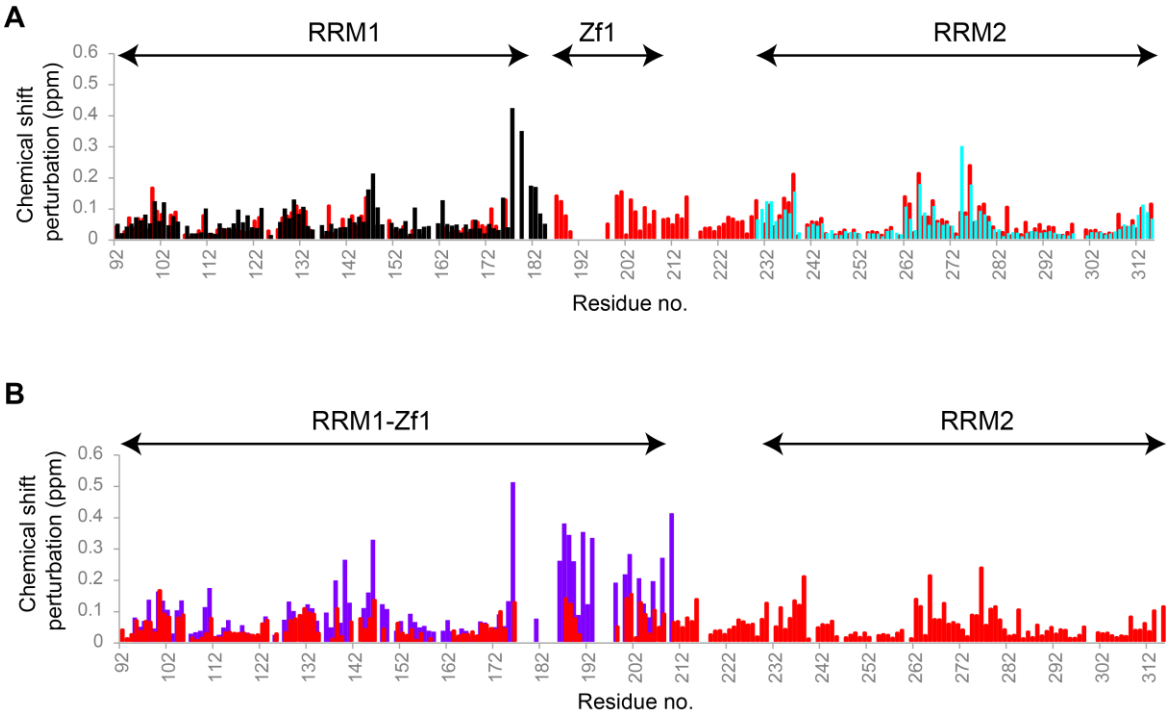


Figure 64 CSP plots comparing free and RNA bound forms of RBM5 protein constructs

(A) The chemical shift perturbation plots of free and RNA bound forms of RRM1 (bound to C/U rich RNA), RRM2 (bound to C/U rich RNA) and RRM1-Zf1-RRM2 C191G (bound to GGCU_12 RNA) are shown in black, cyan and red, respectively. (B) The chemical shift perturbation plots of free/RNA bound forms of RRM1-Zf1 C191G and RRM1-Zf1-RRM2 C191G bound to GGCU_12 RNA are shown in purple and red, respectively.

A similar comparison of the tandem domains RRM1-Zf1 C191G free and RNA bound versus triple domains RRM1-Zf1-RRM2 free and RNA bound proteins is made in the presence of the same RNA (GGCU_12) in **Figure 64B**. It is apparent that the chemical shifts occurring upon RNA binding are different in the tandem and triple domain constructs, with greater changes in the Zf1 than in RRM1. Since we have already seen that there are large changes between the tandem and triple domain constructs already in the free form, it could well be true that these differences are also translated into the RNA bound forms. As stated previously, these changes could either be a result of additional inter-domain (or between with the linker between Zf1 and RRM2) contacts in the triple domain construct which are absent in the tandem domain and/or due to disturbance of RRM1-Zf1 interface in the triple domain construct. In **section 5.6**, the possibility of partial destabilization of the RRM1-Zf1 interface in the RNA bound form of the tandem domains was discussed. In light of the differences between NMR titration data of the triple domains versus tandem domains with RNA, it becomes tempting to speculate that the RRM1-Zf1 interface exists in the free form of tandem domain RRM1-Zf1 while being at least

partly disturbed in the presence of RNA or in the context of the triple domain construct (in both free and RNA bound forms). Validation of this hypothesis by obtaining either high resolution structure of the tandem domains-RNA complex or structural models of the triple domains (free/RNA bound) using PREs, RDCs and SAXS would be essential for conclusion of this study.

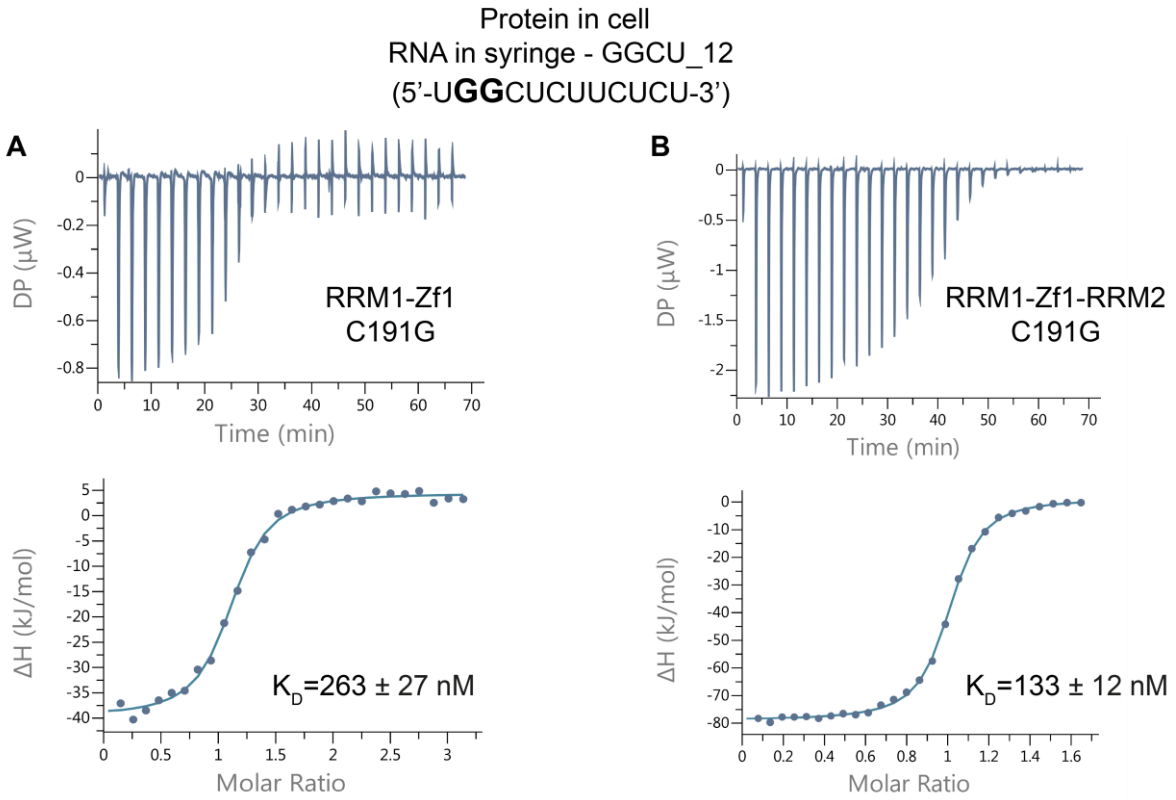


Figure 65 ITC binding isotherms of RRM1-Zf1 and RRM1-Zf1-RRM2 C191G mutants

ITC binding isotherms of tandem domains RRM1-Zf1 and triple domains RRM1-Zf1-RRM2 C191G mutants for binding to GGCU_12 RNA are shown in (A) and (B), respectively. The binding dissociation constants are indicated.

Furthermore, ITC binding isotherms were acquired to obtain the binding affinities of tandem and triple domain C191G mutants for GGCU_12 RNA (**Figure 65**). A 2-fold gain in affinity for GGCU_12 RNA is observed in the presence of the additional RRM2 domain in triple domain construct compared to the tandem domain construct. The gain in affinity due to RRM2 indicates that all three domains are able to bind the RNA, consistent with NMR titration data. In theory, it is expected that the addition of individual domains should have a multiplicative effect on the gain in affinity, which is clearly not the case here. This could be possibly attributed to either the use of a sub-optimal RNA for binding studies or the partial occlusion of one of the domains in the triple domain construct which could hinder RNA binding.

6.2. Multidomain dynamics of RRM1-Zf1-RRM2 C191G

Protein flexibility can play a crucial role in the functioning of the protein. In case of multi-domain protein whereby individual domains are connected by long linkers, it becomes extremely important to study the relaxation properties of the protein to gain initial insights into possible inter-domain contacts. This is easily reflected in the total correlation time (τ_c) which is calculated from R_1 and $R_{1\rho}$ experiments.

The value of the τ_c depends on the overall size of the protein and as a general rule of thumb, the theoretical value should be ~0.6 times the size of the protein. Due to the differences in molecular weight of the individual domains (RRM1 ~9 kDa, Zf1 ~3.5 kDa and RRM2 ~9 kDa), if the domains tumble independently without any inter-domain interactions, the τ_c values should be different for the domains. Contrastingly, if there are inter-domain interactions, the τ_c value of the individual domains would be much higher than the expected value of the single domains and would reflect a value expected for the entire protein (~25 kDa).

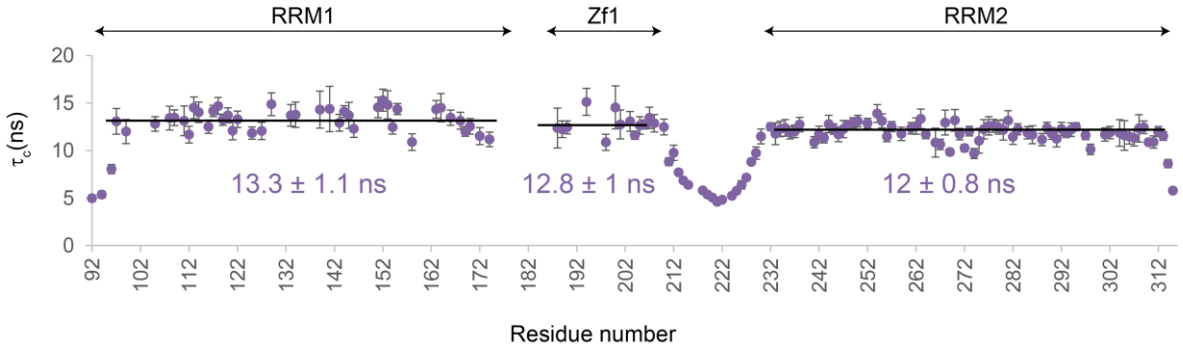


Figure 66 ^{15}N-relaxation data for RRM1-Zf1-RRM2 C191G mutant protein

The total rotational correlation time (τ_c) calculated from the R_1 and R_2 rates is plotted against residue number, the average ± standard deviation values are listed for each domain. Since the difference between the τ_c values for RRM1, Zf1 and RRM2 is within the error, it is concluded that all three domains tumble together in solution.

The total rotational correlation time obtained for the three domains is, τ_c ~13.3 ns for RRM1, ~12.8 ns for Zf1; and ~12 ns for RRM2 (**Figure 66**). Since the molecular weight of the protein is ~25 kDa, a total correlation time of 15 ns would be expeted for the entire protein. The values of τ_c for the individual domains are arguably less than 15 ns but one should consider that τ_c=0.6 times molecular weight of the protein is only an approximation based on a number of assumptions. Additionally, the flexible linker between Zf1 and RRM2 might not contribute

to the overall correlation time owing to its flexibility.Although the τ_c values are a bit different for the domains, they all lie within the standard error demonstrating that all the three domains tumble together in solution. It is noteworthy that the NMR signals of the 18 residue long flexible linker between Zf1 and RRM2 (residues 211-230) also have much high signal intensity in the $^{1}H,^{15}N$-HSQC spectrum, compared to the rest of the protein.. Moreover, the relatively sharp linewidths of RRM2 signals in the $^{1}H,^{15}N$-HSQC spectrum of the triple domain could be an effect of some degree of independent tumbling of RRM2 as it is present just after a flexible linker.

Flexibility of protein may vary greatly in the presence and absence of the ligand which may directly or indirectly be related to the protein function. It is quite possible that upon ligand binding flexible domains or even linkers might become rigid or vice-versa, depending on the mode of recognition of the RNA by protein. It therefore becomes important to study and compare the relaxation properties of the protein in its free versus RNA-bound form.

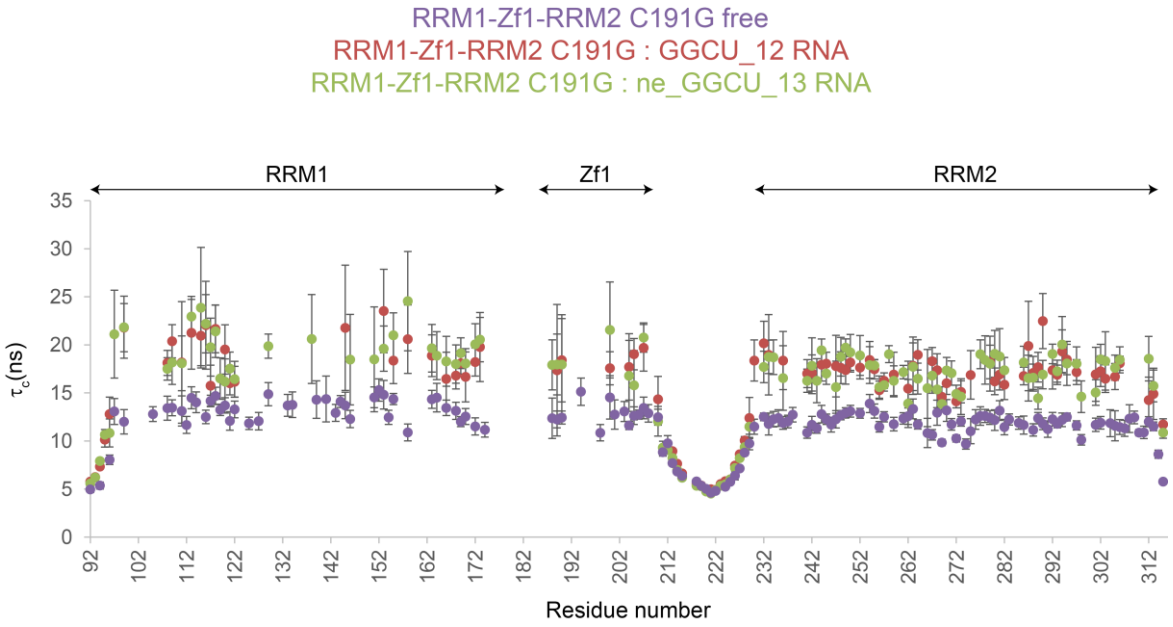


Figure 67 ^{15}N-relaxation data of free/RNA bound RRM1-Zf1-RRM2 C191G mutant protein

The total rotational correlation time (τ_c) calculated from the R_1 and R_2 rates is plotted against residue number for RRM1-Zf1-RRM2 C191G mutant protein in free form, bound to GGCU_12 RNA and ne_GGCU_13 RNA in purple, red and green, respectively.

As mentioned previously, the NMR spectra of the RRM1-Zf1-RRM2 C191G protein-RNA complex become complicated to analyze due to exchange broadening processes causing a major decrease in signal intensity, more so in the RRM1 and Zf1 domains than RRM2 domain.

This is clearly reflected in significant amount of error in estimation of the correlation times (**Figure 67**). Nevertheless, some useful information can still be extracted from the data. As can be seen from **Figure 67**, the overall correlation time (τ_c) of the protein in either of the RNA bound forms (red/green) increases from that of the free protein (purple). This is expected as the protein-RNA complex becomes higher in molecular weight.

As the error bars are quite high, estimation of the total correlation time for individual domains in the RNA bound forms is quite difficult. Still, it is safe to say that the three domains tumble together in solution in both the free and RNA bound forms. It therefore becomes tempting to speculate that the RNA bound conformation of the protein maybe already pre-formed in the free state of the protein.

It is also noteworthy that the linker between Zf1 and RRM2 remains as flexible in both the RNA bound forms as in the free form. This is surprising as considerable amount of chemical shift changes in the linker are observed upon RNA binding (**Figure 63**). As stated before, these changes could be a result of either a direct protein-RNA interaction or may arise from allosteric effects. With the relaxation data at hand, it can be hypothesized that the chemical shifts in the linker maybe an allosteric effect.

SAXS serves as a convenient tool to provide insights into the existence of multiple conformation of a protein in solution. In case of multi-domain proteins, it is quite possible that the protein samples multiple conformations due to the degree of flexibility provided to individual domains by the presence of long, usually unstructured linkers. It becomes even more interesting to see how the SAXS curve changes upon ligand binding and if there are any indications for enrichment of a subset of conformations in the ligand bound versus the free form.

In RBM5, the three RNA binding domains (RRM1-Zf1-RRM2) are connected together by a very short linker (7 residues) between RRM1 and Zf1, and by a relatively long (20 residues) linker between Zf1 and RRM2. Since we already know from the ^{15}N-relaxation data that RRM1 and Zf1 tumble together in solution as a single entity in the tandem domain construct, it could be assumed that these domains remain rigid in the context of triple domain. On the other hand, ^{15}N-relaxation data of the triple domain construct clearly shows that the linker connecting Zf1 and RRM2 is flexible and might also confer some degree of flexibility to the domains even though the protein behaves as a single moiety as indicated by the τ_c values.

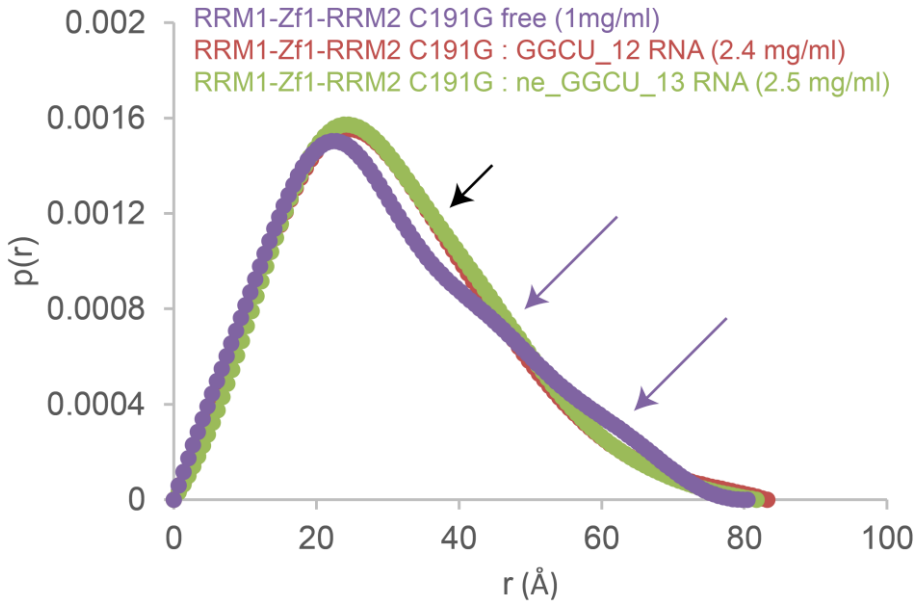


Figure 68 SAXS analysis of RRM1-Zf1-RRM2 C191G mutant protein

p(*r*) curves showing maximum pairwise distribution for RRM1-Zf1-RRM2 triple domain C191G mutant in free and two RNA bound forms (U**GG**CUCUUCUCU, GAACUU**GG**CUCU) are shown in purple, red and green, respectively. Purple arrows indicate two shoulders observed in the p(*r*) curve for the free protein. Black arrow shows that the protein exists in an extended conformation in the both RNA bound forms, compared to the free form.

In line with this, SAXS data for the free and RNA-bound RRM1-Zf1-RRM2 C191G mutant were recorded on a Rigaku BIOSAXS 1000. The protein was concentrated to 4 mg/ml and data for a concentration series: 4 mg/ml, 3 mg/ml, 2 mg/ml, 1 mg/ml and 0.5 mg/ml, was measured. The protein does not show concentration dependent aggregation behavior but does show concentration dependent increase in I_0, possibly due to oligomerization. No visible differences between 0.5 mg/ml and 1 mg/ml concentration data were observed and therefore, the 1 mg/ml data was used for further analysis. The p(*r*) curve describing the pairwise distance distribution is plotted for the lowest concentration (1 mg/ml), and it shows a maximum dimension D_{max} of 80.4 Å (**Figure 68**). Purple arrows marked in **Figure 68** point to the existence of two shoulders in the SAXS curve, which indicate the presence of multiple conformations.

Table 7 SAXS data collection and processing statistics for RRM1-Zf1-RRM2 C191G mutant free and RNA-bound forms

	RRM1-Zf1-RRM2 C191G free (1 mg/ml)	RRM1-Zf1-RRM2 C191G + GGCU_12	RRM1-Zf1-RRM2 C191G + ne_GGCU_13
Data-collection			
Instrument	Rigaku BIOSAXS1000	Rigaku BIOSAXS1000	Rigaku BIOSAXS1000
Beam geometry	10 mm slit	10 mm slit	10 mm slit
Wavelength (Å)	1.5	1.5	1.5
q range ($Å^{-1}$)	0.004-0.65	0.004-0.65	0.004-0.65
Exposure time (s)[a]	900	900	900
Concentration range (mg ml^{-1})	1	2.4	2.45
Temperature (°C)	5	5	5
Structural parameters			
$I_{(0)}$ (cm^{-1}) [from p(r)]	0.7 ± 0.0	0.35 ± 0.0	0.37 ± 000
R_g (Å) [from p(r)]	24.29 ± 0.00	24.08 ± 0.00	24.07 ± 0.00
$I_{(0)}$ (cm^{-1}) [from Guinier]	0.69 ± 0.0054	0.35 ± 0.003	2.37 ± 0.0024
R_g (Å) [from Guinier]	23.33 ± 1.93	24.02 ± 1.71	23.51 ± 1.03
D_{max} (Å)	80.42	83.21	81.7
Porod volume estimate ($Å^3$)	35421.50	43336.1	45519.6
Software employed			
Primary data reduction	Rigaku SAXSLab v 3.0.1r1	Rigaku SAXSLab v 3.0.1r1	Rigaku SAXSLab v 3.0.1r1
Data processing	PRIMUS	PRIMUS	PRIMUS

[a] 8 frames were recorded for each sample

Next, to measure the SAXS data on the protein-RNA complex, the respective RNA was added to the protein in 1:1.1 ratio and loaded onto an analytical size exclusion column, after incubation of the complex at room temperature for 1 hour. This step ensures removal of any excess RNA from the sample which is necessary to avoid scattering from the free RNA. The protein-RNA complex was then concentrated and a dilution series was measured: 4.8 mg/ml, 2.4 mg/ml, 1.2 mg/ml, 0.6 mg/ml, 0.3 mg/ml for GGCU_12 RNA and 4.9 mg/ml, 2.45 mg/ml, 1.22 mg/ml, 0.6 mg/ml, 0.3 mg/ml for ne_GGCU_13 RNA. It should be noted here that the concentration of the protein-RNA complex is not accurate as the absorbance by RNA at 280 nm is not taken into account while during measurement on the Nanodrop. It is however not an issue as I am only interested in the relative values of concentration and not the absolute values.

As for the free protein, a concentration dependent increase in I_0 was observed for both the protein-RNA complexes. Still, it was not so prominent in the lower concentrations making the 2.4 mg/ml and 2.45 mg/ml data points for each of the RNAs useable. The other data points could not be used due to high noise levels. The p(r) curves show maximum dimensions (D_{max})

of 83.2 Å and 81.7 Å for GGCU_12 RNA and ne_GGCU_13 RNA, respectively (**Figure 68**). The shape of the SAXS curve remains exactly the same for both the protein-RNA complexes clearly indicating that no domain rearrangement is observed upon extension of the RNA on either side of the GG motif that confers Zf1 specific binding. There is also no significant difference in maximum dimensions of either the RNA-bound forms or the free form of the protein. The black arrow in **Figure 68** points to the existence of an extended conformation of the protein in the RNA bound form compared to that of the free form. This pattern is similar to that observed for the free/RNA-bound forms of the tandem domain (RRM1-Zf1 C191G mutant), as shown in **Figure 59B**.

Interestingly, the two humps observed in the case of the free protein have disappeared in the protein-RNA complex. The uniformity of the SAXS shape of the protein-RNA complex could indicate that it exists as a 'relatively' homogenous population. It remains to be seen if the RNA bound conformation of the protein is pre-formed in its free state or is just enriched or selected for in the presence of RNA; or an entirely new conformation, distinct from its conformation in the free form is observed.

6.3. *Caspase-2* pre-mRNA *in vivo* splicing assays

To test the effect of RBM5 RNA binding domains on *Caspase-2* pre-mRNA in *in vivo* splicing assays, *Caspase-2* mouse minigene was used. Alternative splicing of *Caspase-2* can either the pro-apoptotic isoform *Casp-2L* or anti-apoptotic isoform *Casp-2S*, depending on exclusion or inclusion of a 61-bp exon 9. It has been shown previously that RBM5 promotes the formation of *Casp-2L* isoform by directly binding to a C/U rich intronic region upstream of the ln100 intronic element (Fushimi, Ray et al. 2008). In order to investigate the degree of involvement of the individual RNA binding domains of RBM5 in alternative splicing regulation of *Caspase-2* pre-mRNA, the changes in the alternatively spliced products of *Caspase-2* minigene (*Casp-2L/2S* ratio) were detected using RT-PCR with primers specific for the minigene isoforms, upon over-expression of RBM5 wild-type or mutant proteins. For this purpose, over-expression of single and tandem RNA binding domain deletion mutants were used in comparison with that of the wild-type protein. It is expected that the AS regulation of *Caspase-2* minigene by RBM5 would be considerably compromised upon over-expression of the mutant protein, where the specific domain/s responsible for its AS regulation function of would be deleted, thereby altering the *Casp-2L/2S* ratio in comparison to the wild-type. The *in*

vivo splicing assays were carried out by Dr. Sophie Bonnal in Dr. Juan Valcárcel's group in Barcelona, Spain.

Single and tandem domain deletion mutants were tested for AS regulation of *Caspase-2* pre-mRNA by observing the change in *Casp 2L/2S* ratio, as shown in **Figure 69**. A more than 2-fold increase in *Casp 2L/2S* ratio was observed upon over-expression of wild-type RBM5 protein, in comparison to that of the control (ADAR). Upon over-expression of deletion mutants of single domains RRM1, Zf1 and RRM2 or tandem domains RRM1-Zf1, negligible changes in *Casp 2L/2S* ratios were observed, when compared to that of wild-type protein. A 2-fold decrease in the *Casp 2L/2S* ratio was observed upon deletion of RRM1-RRM2 domains, indicating possible involvement of these domains in AS regulation of *Caspase-2* pre-mRNA. Upon mutation of the Zf1 binding site on the RNA (GG->UU), regulation of *Caspase-2* pre-mRNA still occurs, although to a lesser extent. Although it is clear that over-expression of RBM5 leads to alterations in *Casp 2L/2S* ratio, the meagre 2-fold change makes it difficult to observe small differences conferred upon over-expression of deletion mutants and/or upon mutation of Zf1 binding site.

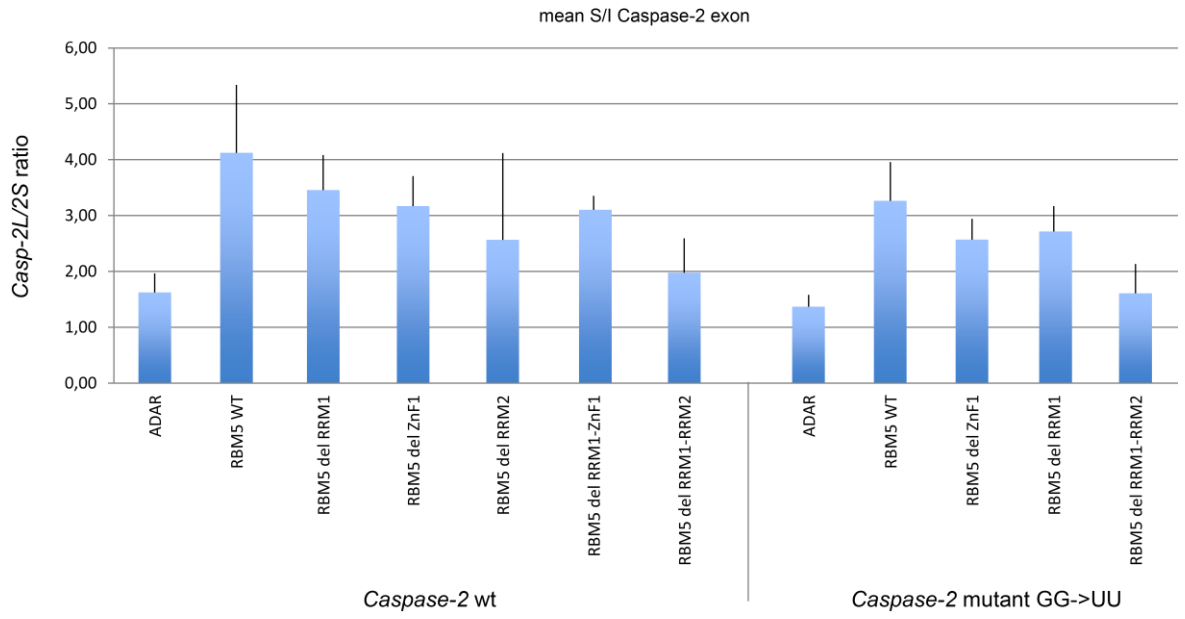


Figure 69 Caspase-2 in vivo splicing assays using Ich2 minigene

Caspase-2 pre-mRNA in vivo splicing assays with changes in Casp 2L/2S ratios upon over-expression of different RBM5 mutants are presented. Data for wild type (wt) and GG->UU mutant Ich2 are shown.

Since it is difficult to derive meaningful information from the Caspase-2 splicing assays, in the future efforts will be employed to search for alternative pre-mRNA targets of RBM5

where a more clear readout of the effects of RBM5 domain deletions or point mutations can be studied.

Chapter 7: Disease linked mutations in RBM5 RNA binding domains

RBM5 was originally identified as a putative tumor suppressor gene consistent with the frequent deletion of its gene locus in lung cancer; while it is over-expressed in breast cancer, suggesting a putative role in promoting tumorigenesis. Therefore, it is very interesting to study certain RBM5 point mutations which are reported in certain cancer patients. To this end, I used cBioPortal (http://www.cbioportal.org/), a portal for cancer genomics that provides visualization, analysis and downloading of large-scale cancer genomics data. It has a nice mutation assessor feature, which can be used to assess the functional impact of the mutation based on conservation. For RBM5, 132 missense mutations are listed. I selected three point mutations based on the functional impact score and the position of the mutation (in the RNA binding domains). For RRM1, I chose R115P and R140S point mutations which are scored to have a 'high' functional impact on the protein via the mutation assessor. For RRM2, I chose R263H point mutation which was ranked to have a 'medium' functional impact. The reason for choosing this particular mutation was that another point mutation in the same residue but to a proline (R263P) was reported previously to cause male sterility in mice (O'Bryan, Clark et al. 2013).

It was suggested that R263P mutation causes pre-mRNA splicing defects in a number of its targets in mice testis, thereby leading to sterility. The authors postulated that since R263 is present on the β-sheet surface and possibly involved in RNA binding, its mutation to a proline might have two-fold effects- disturbance of local secondary structure and abrogation of RNA interaction as a direct effect of substitution of the RNA binding residue R263 to a proline residue. I was therefore interested in studying which of the aforementioned effects is a direct reason of pre-mRNA splicing defects of the mutant protein.

Firstly, the two proline mutations (R115P in RRM1 and R263P in RRM2) were cloned in the single domains, expressed in M9 minimal medium and purified as before. Surprisingly, the mutant proteins did not purify as well as the wild-type proteins, which was already an indication that the mutants had major effects on the proteins. Nevertheless, the purification was still successful and it was possible to record $^{1}H,^{15}N$-HSQC spectra of the mutant proteins. In **Figure 70**, a superposition of $^{1}H,^{15}N$-HSQC spectra of wild-type RRM1 (blue) and R115P mutant (orange) proteins is shown in panel A while that of wild-type RRM2 (pink) and R263P mutant (grey) proteins is shown in panel B. Comparison of the wild-type and mutant spectra clearly shows that in both the cases, the proline mutation disrupts the structure of the protein. Although RRM1 R115P mutant protein is partially unstructured, RRM2 R263P mutant protein is completely unstructured.

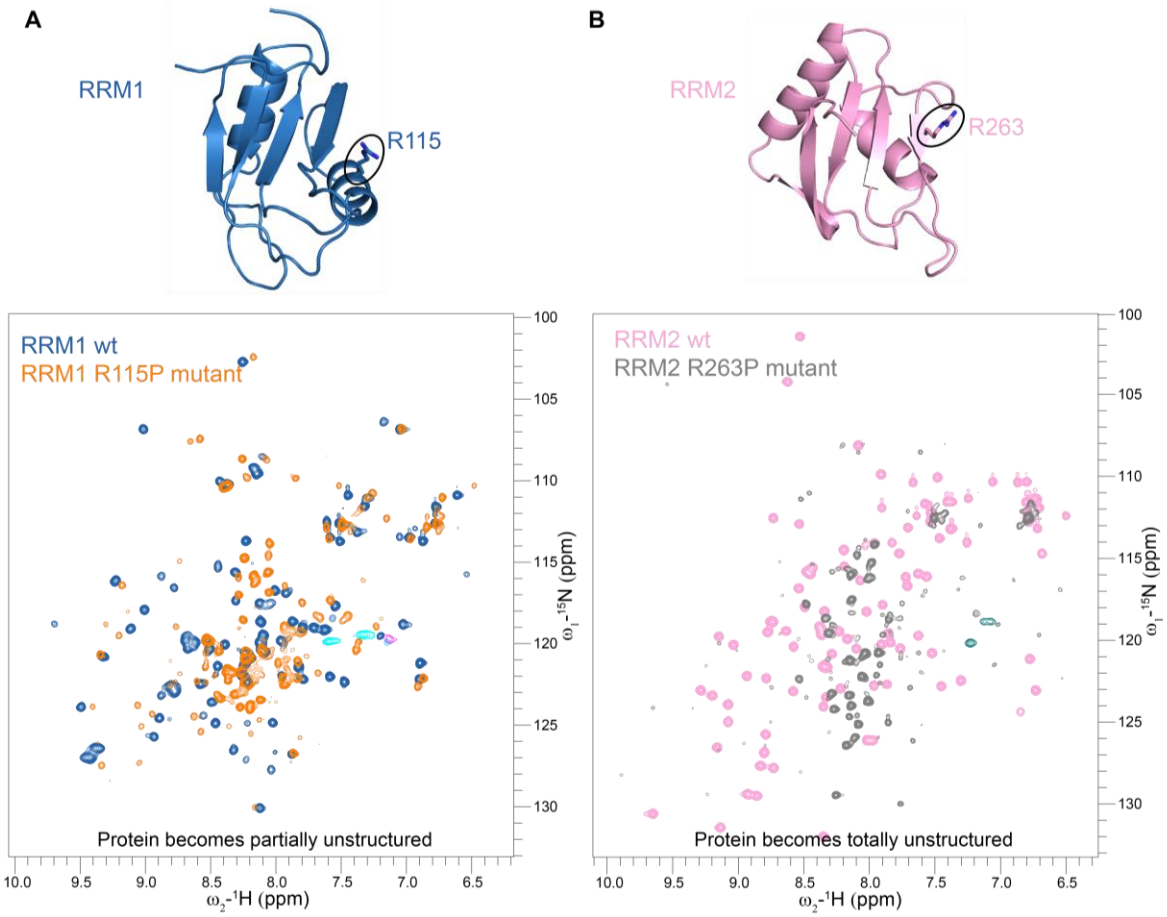


Figure 70 Disease linked mutations affecting the secondary structure of the domains

Superposition of $^{1}H,^{15}N$ -HQSC spectra of wild-type RRM1, RRM1 R115P mutant protein and wild-type RRM2, RRM1 R263P mutant proteins in blue, orange, pink and grey, respectively is shown in panels (A) and (B). The respective position of mutations are shown on the structure of RRM1 and RRM2 (PDB ID: 2LKZ). Both the mutations compromise the structural integrity of the individual domains

.

This suggests that the folding defect in RRM2 R263P mutant is directly translated into functional defects whereby pre-mRNA splicing of its target proteins is affected. Similarly, it is quite possible that the R115P mutation in RRM1 also leads to functional defects due to partially unstructured domain. It is noteworthy that this is just one of the many point mutations in just one protein in the cancer patient and therefore might just be a small contributing factor in causing the disease.

Next, the other two point mutations (R140S in RRM1 and R263P in RRM2) were also cloned and expressed as before. Purification of these two mutants is straightforward, unlike the proline mutants. The proteins also seemed folded upon recording their $^{1}H,^{15}N$-HSQC spectra (right panels in **Figure 71A, Figure 72A**).

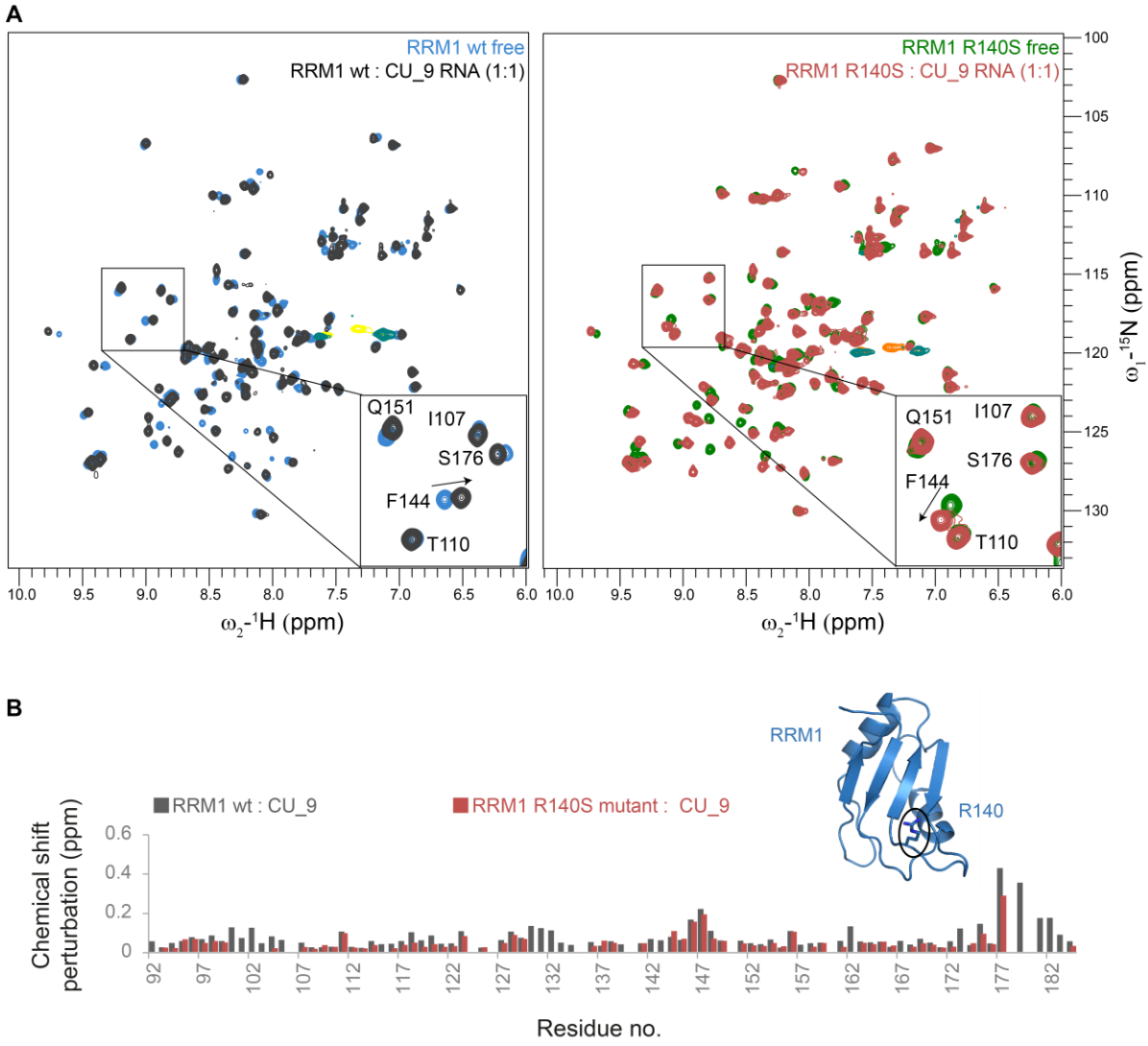


Figure 71 RRM1 R140S cancer mutation does not affect the structure or RNA binding

(A) Overlays of $^{1}H,^{15}N$-HSQC spectra of wild type RRM1 in its free and RNA bound form and RRM1 R140S cancer mutant in its free and RNA bound form are shown in the left and right panels in blue, dark grey, green and maroon, respectively. (B) A comparison of the chemical shift perturbation plots of the wild type RRM1 (dark grey) and R140S mutant (maroon) when bound to the same RNA oligo –CU_9 (5'-UCUCUUCUC-3') in 1:1 ratio is shown. The position of the mutated residue is shown on the structure of RRM1.

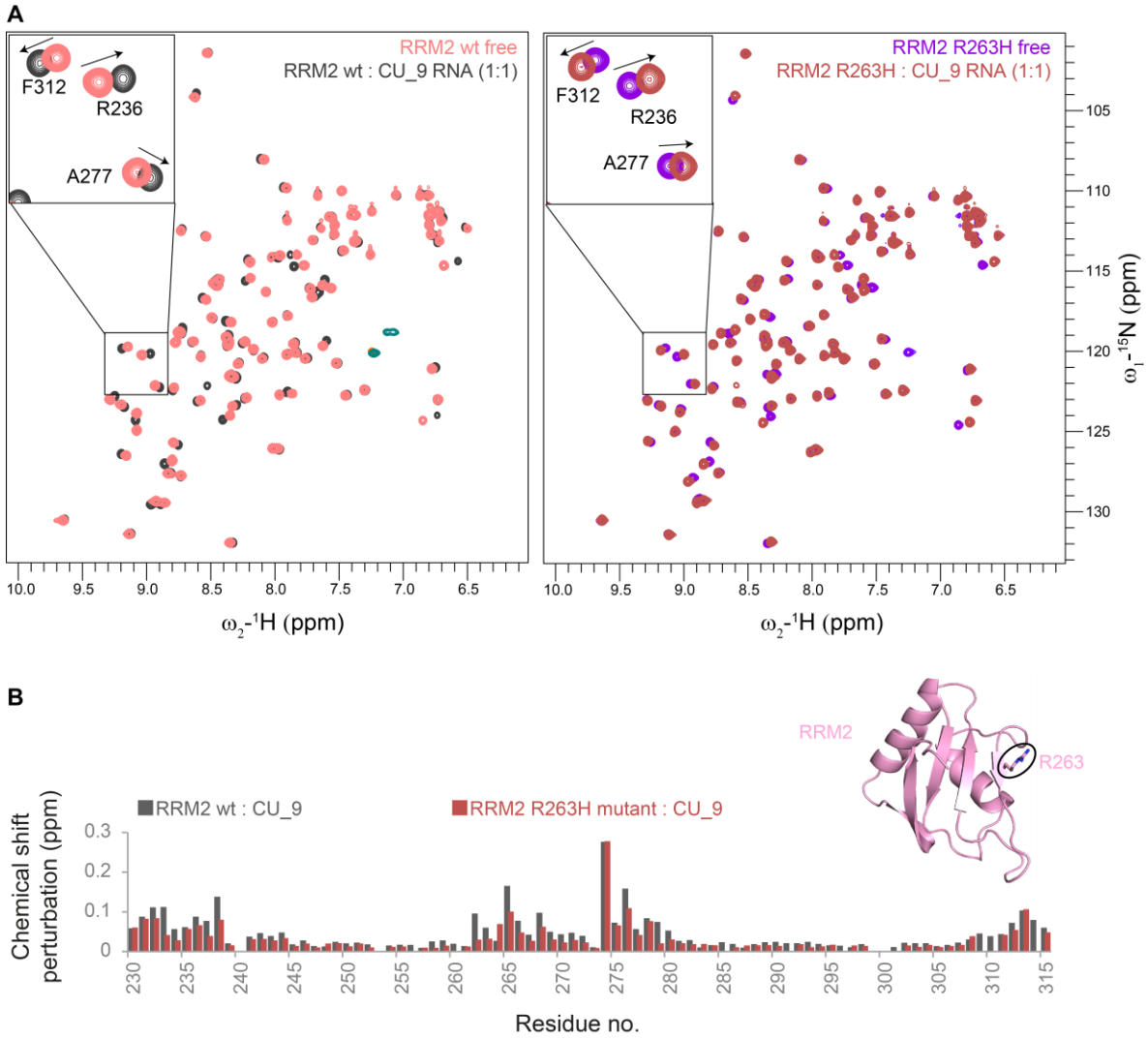


Figure 72 RRM2 R263H cancer mutation does not affect the structure or RNA binding

(A) Overlays of $^{1}H,^{15}N$-HSQC spectra of wild type RRM2 in its free and RNA bound form and RRM2 R263H cancer mutant in its free and RNA bound form are shown in the left and right panels in pink, dark grey, purple and maroon, respectively. (B) A comparison of the chemical shift perturbation plots of the wild type RRM2 (dark grey) and R263H mutant (maroon) when bound to the same RNA oligo –CU_9 (5'-UCUCUUCUC-3') in 1:1 ratio is shown. The position of the mutated residue is shown on the structure of RRM2 (PDB ID: 2LKZ).

Since the mutations did not affect the fold, I titrated the *Caspase-2* derived RNA oligo CU_9 (5'-UCUCUUCUC-3') to check if the respective mutations affect RNA binding. No significant changes in the RNA binding pattern is observed in the mutants upon comparison of the chemical shift perturbation plots of the mutant and wild-type proteins titrated with RNA (**Figure 71B, Figure 72B**). This is not surprising as these are just single point mutations that we are looking at in isolation. It is possible that either these mutations are present in the cancer patient, either by chance or have some other affects (for example on protein-protein interactions etc.), study of which is beyond the scope of this thesis.

Chapter 8: Discussion

8.1. Diverse functionalities of RBM5/6/10 proteins

The conservation of domain organization as well as a high degree of sequence similarity are indicative of overlapping functions between RBM5/6/10 proteins. To ascertain the effects of RBMs in alternative splicing regulation, a large-scale splicing-sensitive microarray analysis was performed whereby each of the individual RBMs was silenced and the effects on splicing events were quantified (Bechara, Sebestyen et al. 2013). Interestingly, depletion of RBM5 affected expression of only 281 transcripts while RBM10 and RBM6 had a more wide-spread effect on 1294 and 1202 transcripts, respectively. Moreover, only 20% overlap between alternative splicing events regulated by RBM5/6/10 was observed suggesting more distinct than overlapping functions of these proteins.

RBM10 OCRE domain is required for AS regulation of *Fas* pre-mRNA

In this thesis, the solution NMR structure of RBM10 OCRE domain is presented. It is a globular domain consisting of six antiparallel β-strands which tyrosine residues exposed on either surface of the domain. Recently, Martin *et al.* also published the solution NMR structure of the RBM10 OCRE domain (Martin, Serrano et al. 2016) (PDB ID: 2MXW). The domain boundaries used in their study extend from residues 558-646 as opposed to those presented in this thesis (residues 562-621). The C-terminal extension in their construct is completely disordered, as seen in **Figure 73A**. Consistently, most of the NOE-based distance restraints were obtained between the residues in the central core of RBM10 OCRE domain.

The superposition of a single representative structure from NMR ensembles of RBM10 OCRE domain determined in this thesis and that from the previously published structure (PDB ID: 2MXW) clearly indicates that the two structures are essentially the same (**Figure 73B**). The backbone RMSD between the two NMR ensembles is 0.813 Å, calculated using residues 563-619 for RBM10 OCRE and residues 564-618 for 2MXW with SuperPose v1.0 webserver (Maiti, Van Domselaar et al. 2004). The arrangement of the surface exposed side-chains of tyrosine residues is also conserved between the two structures, as shown in **Figure 73B.**

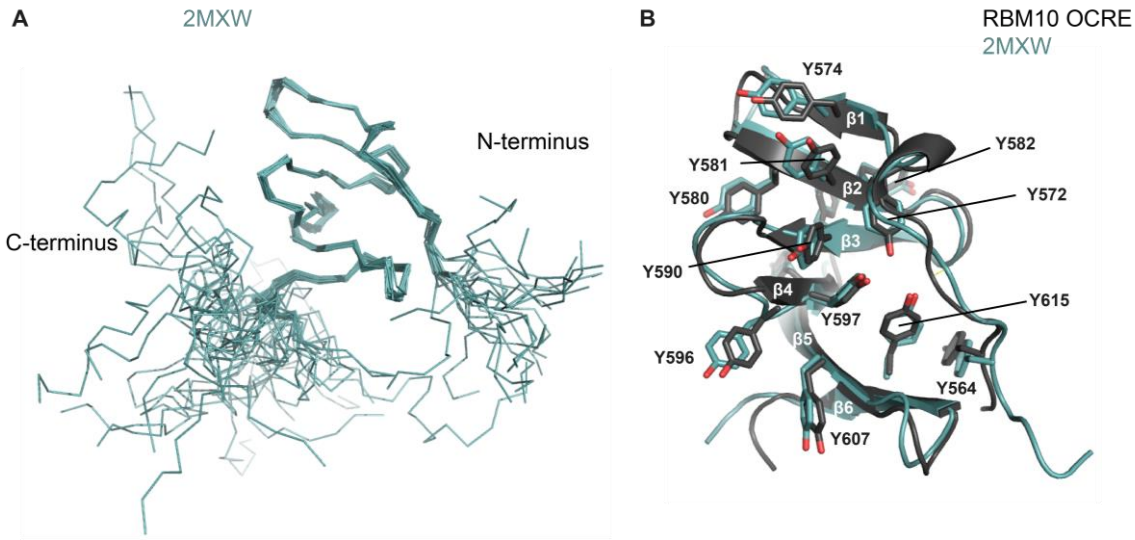


Figure 73 Comparison of RBM10 OCRE domain structures

(A) Solution NMR ensemble of RBM10 OCRE domain (PDB ID: 2MXW). The N- and extended C-termini are unstructured and do not converge in the NMR ensemble upon superposition for the best fit of the central core. (B) Superposition of single representative structures from the NMR ensembles of RBM10 OCRE domain determined in this thesis (black) with that of 2MXW (teal) is presented and the side chains of surface exposed tyrosine residues are denoted.

OCRE domains of RBM5/10 are structurally conserved and both can possibly bind to the core spliceosomal machinery in a similar manner as well, as indicated by similar binding affinities of RBM5/10 OCRE domains for SmN/B/B' C-terminal poly-proline rich tails. Furthermore, *in vivo* splicing assays were done on *Fas* minigene reporter to ascertain similarity between effects of RBM5/10 OCRE domain deletions on AS regulation of *Fas*.

As expected, RBM10 OCRE domain is required for *Fas* exon 6 skipping, consistent with effects of RBM5 OCRE domain. Together with structural and binding data, there are strong indications of the existence of a similar mechanism of *Fas* pre-mRNA AS regulation by RBM10.

A few years ago, Inoue *et al.* also studied the effects of RBM10 on AS regulation of *Fas* pre-mRNA (Inoue, Yamamoto et al. 2014). Interestingly, they found out that depletion of RBM10 alone was sufficient to observe a significant decrease in *Fas* exon 6 skipping (using RNAi mediated knockdown in HeLa and HLE cells), in contrast to the previous observation by our collaborators (Bonnal, Martinez et al. 2008) where simultaneous depletion of RBM5/6/10 was necessary, possibly due to partially redundant functional activities (using RNAi mediated knockdown in HeLa cells). Additionally, they suggested a likely mechanism by which RBM10 regulates AS of *Fas* pre-mRNA. Briefly, RBM10 binds to the 5' splice site

of *Fas* exon 6 that is rich in U- and G- nucleotides (uuguuuggG|GUaaguucuu) possibly via its Zf1 domain which specifically recognizes AGGUAA RNA motifs (Nguyen, Mansfield et al. 2011) with high affinity. Due to this direct protein-RNA interaction, the 5' splice site of exon 6 would be blocked and therefore unavailable for splicing, leading to exon 6 skipping. Furthermore, it was suggested that in case of *Bcl-x* pre-mRNA, RBM10 promotes internal 5' splice site selection by possibly binding to a similar G/U rich sequence (GG|GUAAG) on the 5' splice site on exon 2. It is noteworthy that their model is primarily based on similarity of sequences at the blocked splice sites between *Fas* and *Bcl-x* pre-mRNAs and they did not present any experimental data to support their model.

The mechanism of action of RBM10 in *Fas* AS regulation suggested by Inoue *et al.* is contrasting to that suggested in this thesis whereby the OCRE domain is shown to be responsible.

RBM6 modulates AS of *Fas* pre-mRNA

In this thesis I revealed that RBM6 OCRE domain is a truncated OCRE domain that lacks the ability to recognize the SmN/B/B' poly-proline rich tails. It is noteworthy that RBM6 is still able to regulate the alternative splicing of *Fas* pre-mRNA, although with opposite outcome compared to that of RBM5/10, by promoting exon 6 inclusion. Interestingly, this function of RBM6 is OCRE independent further supporting the data that RBM6 OCRE domain is unable to bind SmN/B/B' tails. It is therefore possible that the AS regulation of *Fas* pre-mRNA by RBM6 protein is conferred via a different mechanism which could perhaps involve another domain in the multi-domain RBM6.

The inability of RBM6 OCRE domain to bind SmN/B/B' C-terminal tails implies that either the domain is just non-functional or it has another function which has not yet been discovered. In addition, the RNA binding domains (RRM1 and Zf1) of RBM6 also lack most of the canonical RNA binding residues and RRM1 also has an unusually low pI of 4.32. Initial NMR experiments involving RNA titrations derived from the *NUMB* RNA (Bechara, Sebestyen et al. 2013) into RRM1 and RRM1-Zf1 domains of RBM6 showed minimal chemical shifts although RRM2 can still bind RNA. Contrastingly, RBM5 and RBM10 (data from another doctoral student in the lab) RNA binding domains can bind their respective target RNA motifs with high affinity.

Another hint for distinct possible functionalities of RBMs came a few years ago, when Heath *et al.* (Heath, Sablitzky et al. 2010) suggested the involvement of RBM6 in co-

transciptional packaging. They also demonstrated the ability of RBM6 to be targeted to splicing speckles, a function which was attributed to its N-terminal multimerization RGG domain. This repeat region is quite short or even absent in both RBM5 and RBM10 while it spans ~22 kDa in the case of RBM6.

In light of these data, it is possible that RBM6 has evolved to perform a diverse set of functions and is possibly involved in protein-protein and protein-RNA interactions via distinct mechanisms from RBM5/10.

8.2. Multipartite RNA recognition

RNA-binding proteins (RBPs) are involved in a diverse set of functions carried out only by a handful of RNA binding domains. Since RBPs usually contain multiple RNA binding modules, it is not surprising that they cater to this large functional diversity by employing a combination of modules. This way the RBPs achieve high affinity and specificity of target recognition. The relatively weak interactions of the individual domains make it easier for their regulation especially in cases where assembly and dis-assembly of complexes is required.

Tandem domains connected together via short linkers act as a classic example of formation of an extended RNA binding interface. Since the individual domains only provide base specific recognition of about 2-3 nucleotides, such a combination of closely positioned RNA binding domains may help in specific recognition of a longer RNA sequence (Shamoo, Abdul-Manan et al. 1995, Mackereth, Madl et al. 2011, Hennig, Militti et al. 2014). The presence of a long flexible linker between the RNA binding modules could also be advantageous in certain situations where the module specific nucleotides are positioned far apart from each other on the same or different RNAs (Braddock, Louis et al. 2002, Oberstrass, Auweter et al. 2005, Stefl, Xu et al. 2006). Another unique way of RNA recognition via dimerization of individual domains leading to a cooperative mode of RNA recognition has also been observed in certain proteins (Liu, Luyten et al. 2001, Ryder, Frater et al. 2004, Beuck, Szymczyna et al. 2010, Meyer, Tripsianes et al. 2010, Feracci, Foot et al. 2016). Additionally, dimerization of RBPs caused by structural rearrangements upon RNA binding can also play an important role in RNA binding (Varani, Gunderson et al. 2000, Chao, Lee et al. 2005, Lingel, Simon et al. 2005).

The inter-domain arrangement during RNA recognition is also quite important. It is frequently observed that domains acting as independent moieties undergo major structural re-

organization upon RNA binding to behave as a single compact molecule as observed in the case of Sxl (Handa, Nureki et al. 1999), PABP (Deo, Bonanno et al. 1999), Hrp1 (Perez-Canadillas 2006) and TIA-1 (Wang, Hennig et al. 2014) proteins. Contrastingly, there exist other examples where a pre-formed arrangement of RNA binding domains is required for target recognition (Hudson, Martinez-Yamout et al. 2004). Another form of modulation of RNA recognition occurs via conformational selection. In the tandem RRM domains of the essential splicing factor U2AF65, an ensemble of active and inactive conformations of the tandem domains exist in the free form and a dynamic population shift from inactive to active conformation occurs in the presence of strong polypyrimidine tract (Mackereth, Madl et al. 2011, Huang, Warner et al. 2014, Voith von Voithenberg, Sanchez-Rico et al. 2016)

The linker length is also important. Theoretically, the affinity of a tandem domain protein for its RNA target could be obtained by multiplying that of the individual domains. But with increasing linker length (>50-60 residues), the domains behave as independent modules without affecting the overall affinity while in case of a short linker, the combined affinity can increase by 10-1000 fold (Shamoo, Abdul-Manan et al. 1995).

RBM5 involves a novel mode of RNA recognition where a stable, compact arrangement between the RRM1 and Zf1 domains is required. The crystal structure of RRM1-Zf1 reveals the tethering of the domains via a tri-partite mechanism where specific contacts between all three parts, i.e. RRM1, linker and Zf1 are essential. Consistently, ^{15}N-relaxation data of the free and protein-RNA complex show that the two domains tumble together in solution in both free and RNA-bound forms , although the existence of a slightly extended conformation in the RNA bound form is indicated by both ^{15}N-relaxation as well as SAXS data. Contrastingly, in case of RBM10, the RRM1 and Zf1 domains tumble independently of one another in the free form (Collins, Kainov et al. 2017), whereas they act as a single moiety upon RNA binding (Martin Ruebbelke, personal communication).

It is noteworthy that the length of the linker is quite short, spanning only 7 residues, possibly providing a pre-formed extended RNA-binding interface in RBM5. Upon RNA titration into the tandem domain construct, large chemical shift perturbations (CSP) are observed in Zf1 domain. This could either be only due to RNA binding as we know that Zf1 is the highest affinity domain or it could be indicative of a cumulative effect of RNA binding and possible domain re-orientation upon RNA binding. Consistent with the latter, SAXS analysis of the free and protein-RNA complex of the tandem domain show a slight increase in the

maximum dimension of the protein upon RNA-binding, which would not be observed in case no domain re-organization occurs.

Additionally, the relatively longer linker between Zf1 and RRM2 compared to that between RRM1 and Zf1 could also be a suggestion of the presence of two separate RNA recognition entities- RRM1-Zf1 and RRM2 in RBM5. The ^{15}N-relaxation data on the triple domain RRM1-Zf1-RRM2 C191G mutant indicate that the three domains tumble together in solution in both free and RNA bound forms. Intriguingly, the SAXS curve for the free protein is indicative of presence of multiple conformations in the free form while that of the RNA bound form seems to be a uniform curve. It is noteworthy that the ^{1}H,^{15}N-HSQC spectra of both the free and RNA-bound RRM1-Zf1-RRM2 C191G mutant show considerable line-broadening in RRM1-Zf1 while the RRM2 signals are relatively sharp, indicative of additional exchange processes on the RRM1-Zf1 side.

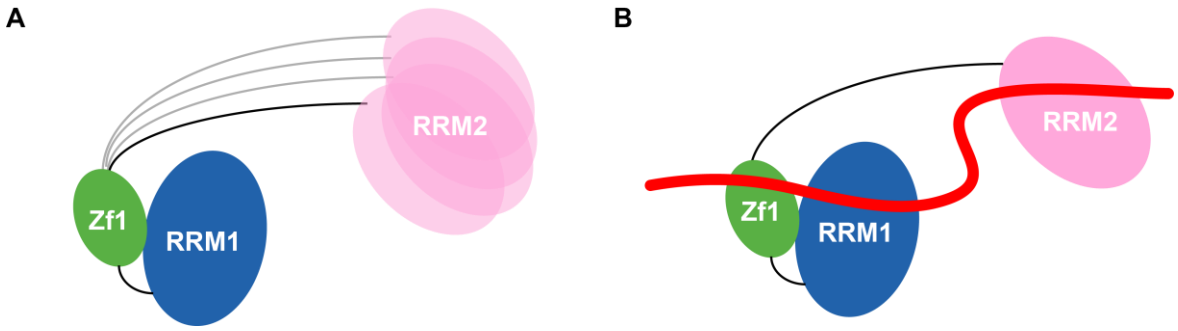


Figure 74 Hypothetical model of RBM5 RNA binding domains

A hypothetical arrangement of the RNA binding domains of RBM5 in the free form (A) and in the presence of RNA (B). RRM1, Zf1 and RRM2 domains are color coded in blue, green and pink, while the RNA is shown in red. In the free form, RRM2 is able to sample multiple conformations while in complex with RNA, it stably achieves a single conformation.

Combining all these data, it is tempting to speculate that RRM1-Zf1 and RRM2 are partially uncoupled in the free form but they behave as a single entity in the RNA-bound form (**Figure 74**). PRE experiments planned in the future could provide more conclusive evidence of the conformation of the triple-domain RBM5 RNA binding region in the presence and absence of RNA. Interestingly, it was recently proposed that in case of RBM10, the RRM2 domain re-orients quasi-independently from RRM1-Zf1 in solution (Collins, Kainov et al. 2017). Such conformational dynamics in multi-domain proteins could be highly relevant in the biological context where the splicing factors need to scan multiple pre-mRNA targets to finally bind and regulate only a handful of them. This would point towards the possibility of

recognition of distinct pre-mRNA targets by either individual domains or a specific combination of domains which may exist either in a pre-formed RNA binding conformation or may undergo structural re-arrangement upon RNA binding. In this way, the functional capacity of the splicing factors can greatly expand.

8.3. Implications of variations in canonical RRM domains

RRM domains usually contain the βαββαβ topology wherein the RNA recognition is conferred via the β-sheet interface. Many variations of this canonical RRM topology are known, which extend the RNA recognition region. One such variation is observed in the RRM domains of PTB whereby the presence of an additional β-strand allows the recognition of one or two extra nucleotides (Oberstrass, Auweter et al. 2005). Another variation of the canonical RNA binding interface is illustrated in case of the RRM domain in Fox-1 where certain loops connecting structured elements are involved in RNA binding (Auweter, Fasan et al. 2006). In yet another variation, these loops are involved in recognition of RNA shape rather than sequence (Skrisovska, Bourgeois et al. 2007).

It has also been observed in certain cases that the N- and C-terminal extensions of the RRM domains (outside the core) are critical for RNA recognition and significantly enhance the RNA binding affinity of the protein, for example in the case of CUG-BP1 (Tsuda, Kuwasako et al. 2009), CB20 (Mazza, Segref et al. 2002), PABP (Deo, Bonanno et al. 1999) and PTB (Oberstrass, Auweter et al. 2005) proteins. It was also shown that both unstructured N- and C-terminal extensions of Tra2-β1 RRM become structured upon RNA binding, while being involved direct protein-RNA interactions (Clery, Jayne et al. 2011, Tsuda, Someya et al. 2011).

Interestingly, another important function of these terminal extensions is to inhibit RNA binding by partially blocking the RNA binding β-sheet interface. In the case of U1A RRM domain, it was demonstrated that the C-terminal helix makes contacts with the RRM core on the upper side of the β-sheet interface to shield the RNA binding hydrophobic surface and provide stability to the protein. Upon binding to RNA, the helix is displaced to make way for the RNA(Avis, Allain et al. 1996). Later it was discovered that displacement of the C-terminal helix upon RNA binding helps in formation of homodimers via association of the C-terminal helices of two monomers (Varani, Gunderson et al. 2000). A similar observation was made in the case of La protein, although the implications of the hindrance in accessibility of RNA binding interface by the C-terminal extension were not entirely clear (Jacks, Babon et al. 2003).

Recently, Song *et al.* (Song, Wu et al. 2012) showed that the N-and C-terminal extensions of RBM5 RRM2 domain are involved in RNA binding. The amide resonances of the N-terminal extension only appeared upon RNA titration and moved in a fast exchange manner during the RNA titration eperiments, indicating its involvement in RNA recognition. Addionally, the amide resonances of the C-terminal extension moved dramatically during RNA titrations, increasing the RNA binding affinity by 2-fold. Interestingly, intermolecular NOEs were observed between the C-terminal extension and the RRM core. ^{15}N-relaxation data of the protein-RNA complex showed that this extension became more flexible upon RNA binding. Therefore, it was concluded that the C-terminal extension packs against the RRM core in the free form while in the RNA bound form, it is forced to stay away from the RRM core.

During NMR-based RNA titrations into RRM1 domain, the amide resonances of the C-terminal linker shift dramatically which could be a result of direct RNA contacts of the linker or due to a direct competition between the RNA and linker to bind to the RRM1 core. Since the binding affinities of RRM1 domain with (RRM1) and without linker (RRM1_S) for CU_9 RNA are almost similar (~20 μM and ~11 μM, respectively), it would suggest a minimal involvement of the linker in RNA binding. Due to the presence of contacts between the RRM1 core and linker residues in the crystal structure of RBM5 RRM1-Zf1 (**Figure 37C**) where the linker blocks the top part of the RRM1 β-sheet interface, it is plausible that such a packing also exists in the absence of Zf1 domain (**Figure 75**). The large shifts in the amide resonances of the linker upon NMR RNA titration experiments could then be attributed to a direct competition between the linker and RNA for binding to the RRM1 core (**Figure 75**). Here, the possibility of indirect effects cannot be ruled out.

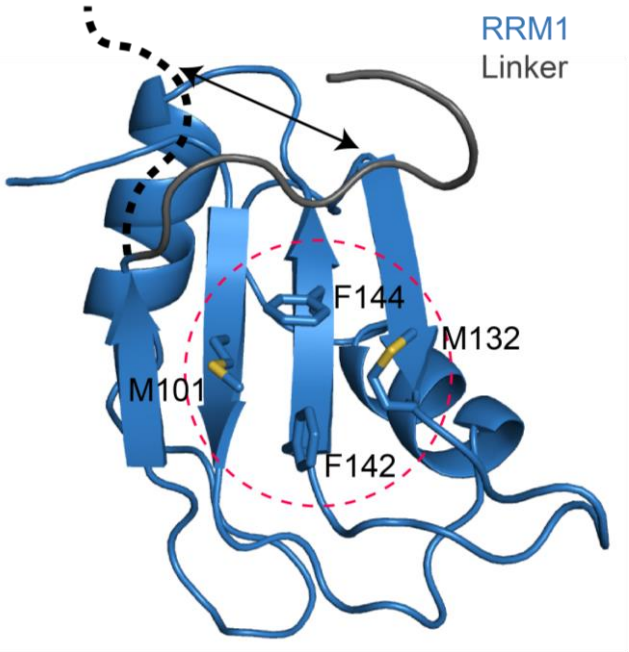


Figure 75 Model indicating possible structural changes upon RNA binding

Crystal structure of RRM1 domain with the linker are shown in blue and dark grey, respectively (derived from RRM1-Zf1 crystal structure). The RRM1 core residues possibly in RNA binding are indicated with a dotted red circle. Structural re-arrangement of the linker that might occur upon RNA binding is shown with an arrow and black dotted line.

This mechanism could also differ in the context of single domain versus that of the tandem or triple domains due to additional contacts between the linker and other components, which might stabilize its conformation. For instance, the crystal structure of RRM1-Zf1 clearly shows that the linker also interacts with the Zf1 domain (**Figure 37D**).

8.4. Sequence specificities of RBM5-RNA interaction

Recent developments in large-scale quantitative data analysis has rendered high-throughput mapping of protein-RNA interactions possible. Traditionally, CLIP based techniques (Cross-Linking and ImmunoPrecipitation coupled with high-throughput sequencing) which involve UV cross-linking of RNA-binding proteins (RBPs) to their respective RNA molecules that are further isolated and sequenced, have provided RNA binding profiles for a variety of RBPs (Licatalosi, Mele et al. 2008, Hafner, Landthaler et al. 2010, Konig, Zarnack et al. 2010, Van Nostrand, Pratt et al. 2016). Although this technique captures biologically relevant *in vivo* interactions, it suffers from a high number of false negatives due to low cross-linking efficiency of several RBPs (Darnell 2010) and is sensitive to differential tissue and time specific expression levels (Blencowe, Ahmad et al. 2009). Other widely used *in vitro* protein-

RNA interaction screening methods include SELEX (Systematic Evolution of Ligands by Exponential Enrichment) combined with high-throughput sequencing (Campbell, Bhimsaria et al. 2012, Campbell, Valley et al. 2014, Ozer, Pagano et al. 2014), which has also been used to decode the RNA binding preferences in the cell (Lorenz, Gesell et al. 2010). However, this technique is greatly biased towards capturing highest-affinity interactions. More recently, another *in vitro* selection method RNACompete has been developed which is able to identify medium-range affinity interactions where the protein in question is incubated with a large molar excess of a diverse RNA pool, after which the protein is isolated by affinity selection and the bound RNAs are subjected to microarray analysis (Ray, Kazan et al. 2013). The main disadvantage of using *in vitro* techniques is that the experiments are performed under non-physiological conditions (Marchese, de Groot et al. 2016). In addition, such data should always be used cautiously keeping in mind that the results might be dominated by the highest affinity binding domain, in case of multi-domain proteins.

Various studies have identified a diverse set of enriched motifs that are potentially recognized by RBM5 using different methods. Initially it was suggested that RBM5 binds polyG homopolymers *in vitro* (Edamatsu, Kaziro et al. 2000). More than a decade later, RBM5 CLIP-seq data (Bechara, Sebestyen et al. 2013) were used to identify consensus motifs (UCAUCGA and AGUAACG) using HOMER software (Heinz, Benner et al. 2010). A closer look into all the top scoring motifs suggests very little overlap between the motifs for example, the motifs AAGGAAAG, CAAGAGUU, AUCUUUGU and CCGGGACA are quite different from each other. One could attribute the emergence of such a diverse set of motifs to the RNA sequence specificities conferred by the different domains. The other explanation would be that the protein-RNA recognition is very unspecific. However, it is reasonable to assume the validity of the first scenario as usually RRMs and Zf domains have some degree of specificity. Interestingly, using RNACompete approach (Ray, Kazan et al. 2013), a consensus motif - GAAGGAA was derived, which is quite similar to one of the CLIP consensus motif (AAGGAAAG).

To further understand the differences in RNA binding specificities of the different RNA binding domains of RBM5, we carried out RNACompete in collaboration with Dr. Debashish Ray in Dr. Tim Hughes's group at University of Toronto. The experiments were performed with different constructs containing either the triple domain (RRM1-Zf1-RRM2), tandem domain (RRM1-Zf1) or the single domains. The rationale behind this was to see if any differences between the enriched motifs from different constructs could be observed.

Interestingly, the experiments on the single domains RRM1 and RRM2 failed possibly due to low affinity of these single domains. On the other hand, results could still be obtained for Zf1 which has been previously suggested to bind its preferred RNA motif (AGGGAA) with affinity in high nano-molar range, which is relatively high compared to that of RRM1: ~20 μM for CU_9 RNA) and RRM2: ~60 μM for CUCUUC/GAGAAG RNA (Song, Wu et al. 2012).

As expected, we did observe minor differences between the enriched motifs (**Figure 76**). We already know that RRM1 can bind a C/U rich RNA and RRM2 also binds CU/AG rich RNA sequences with similar affinity, and Zf1 specifically requires a GG motif for binding. Therefore, it seems likely that the Zf1 domain dominates the results and the enriched RNACompete motifs would not provide a 'true' representation of the RNA binding preferences of RBM5.

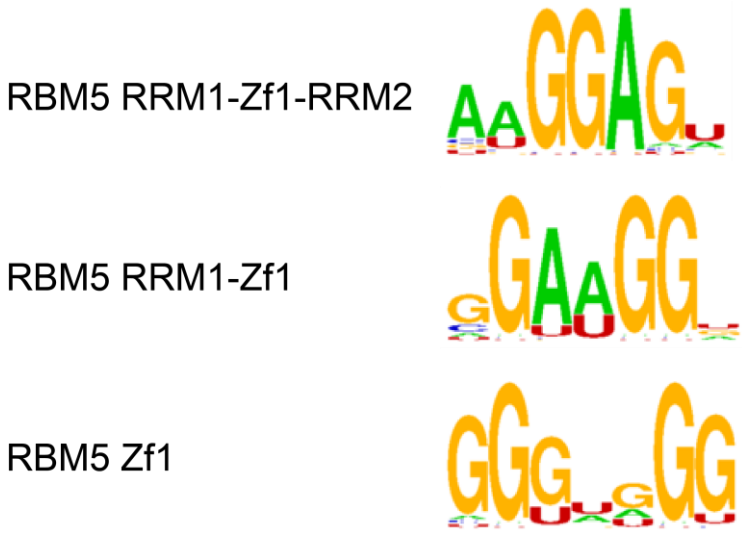


Figure 76 RNACompete top consensus motifs

RNACompete consensus motifs obtained for different RBM5 constructs: RRM1-Zf1-RRM2, RRM1-Zf1 and Zf1 are shown.

Nevertheless, there are differences observed in the consensus motifs for the different constructs which in the simplest interpretation of the data would mean that there are contributions made by the distinct domains. The *Caspase-2* derived RNA sequences used in this thesis for studying protein-RNA interactions- GGCU_12 (5'-UGGCUCUUCUCU-3') and ne_GGCU_13 (5'-GAACUUGGCUCUU-3'), both contain the conserved GG element and a 'GAA' element (ne_GGCU_13), are at least partly representative of the RNACompete motifs.

Recently, Collins and co-workers have shown that RBM10 known to be involved in alternative splicing regulation of *NUMB* pre-mRNA (Bechara, Sebestyen et al. 2013) has two RNA binding modules (Collins, Kainov et al. 2017). The RRM2 domain acts as a separate entity recognizing a C-rich motif while the RRM1-Zf1 recognizes a 'GGA' motif. This would explain the recognition of C-rich intronic sequence upstream of exon 9 in *NUMB* pre-mRNA, which is specifically recognized by RBM10. They additionally found another RBM10 target, TNRC6A, containing the GGA motif important for recognition by RRM1-Zf1 domains. RNAi-mediated knockdown of RBM10 confirmed the specific regulation of *TNRC6A* by RBM10 owing to changes in the levels of the alternatively spliced exons.

Similarly, it has been shown before that RBM5 regulates AS of *AID* pre-mRNA by promoting exon 4 skipping where deletion of the entire N-terminus harboring RRM1,2 and Zf1 severely compromised the function of RBM5 (Jin, Niu et al. 2012). Upon point mutation of conserved phenylalanine to alanine residues in RRM2, a significant decrease in the activity of RBM5 to modulate AS of *AID* pre-mRNA was observed while similar mutations in RRM1 did not have such a pronounced effect. This indicates a higher degree of involvement of RRM2 in AS regulation of *AID* pre-mRNA. Additionally, O′Bryan *et al.* identified a set of 11 putative pre-mRNA targets of RBM5 in round spermatids using microarray analysis out of which the three most enriched targets (St5, Asb1 and Pla2g10) were shown to be directly related to AS regulation by RBM5 RRM2 domain (O'Bryan, Clark et al. 2013). Taken together, RBM5 RRM2 domain is known to regulate a variety of pre-mRNA targets although information on the specific importance of RRM1 and Zf1 domains is still missing.

These data further confirm our hypothesis that multi-domain proteins achieve specific recognition of a diverse set of targets by employing different domains. The fact that RBM5 OCRE domain is involved in alternative splicing regulation of *Fas* pre-mRNA, while the RNA binding domains are involved AS regulation of a variety of pre-mRNA targets (as mentioned above) is already an indication that such a mechanism also exists in RBM5.

Conclusions & Outlook

One of the major goals of this thesis was to understand the structural basis of interaction between the RBM/5/10/6 OCRE domains and the core spliceosomal machinery. To this end, I successfully established the molecular basis of RBM5 OCRE domain-PRM recognition using NMR, ITC and CD spectroscopy. I also calculated the solution NMR structures of RBM10/6 OCRE domains. The structural conservation between RBM5/10 OCRE domains explains a similar mode of recognition of SmN/B/B' derived PRM peptide. Contrastingly, the truncated RBM6 OCRE domain cannot bind to the PRM motif, consistent with the *in vivo* splicing data where no effect of RBM6 OCRE domain was observed on alternative splicing (AS) of *Fas* pre-mRNA. Our collaborators additionally found out that RBM6 has opposite effects on *Fas* AS compared to RBM5/10. In the future, it would be interesting to understand which domain/s of RBM6 confer the AS regulation of *Fas* pre-mRNA.

Another key aim of this thesis was to understand how the multiple RNA-binding domains of RBM5 cooperate with each other to carry out AS regulation of *Caspase-2* pre-mRNA. For this purpose, I used a divide and conquer approach where I started studying protein-RNA interactions of single domains, then tandem and triple domains. We solved the crystal structure of RRM1-Zf1 tandem domains which showed that the two domains are coupled together forming a compact structure. Additional NMR and SAXS data point towards the existence of an extended conformation of the tandem domains in the presence of RNA. I also found out that Zf1 domain preferentially binds to a GG motif, providing a high degree of sequence specificity to the protein-RNA recognition. Furthermore, I used metal exchange kinetics to probe the effects of an extra cysteine residue on the stability of Zf1 domain. Finally, NMR and SAXS were used to provide initial insights into the dynamics of the triple-domain RNA binding region. In the future I plan to record additional PRE experiments which would provide us with long range distance restraints to aid in calculation of a low resolution model of protein-RNA recognition, along with restraints obtained from SAXS and RDCs.

This thesis which aims at obtaining a structural model of RBM5/6/10 protein and RBM5 RNA binding domains along with study of the conformational dynamics of RBM5-RNA recognition is essential in understanding the molecular and structural basis of recognition and differentiation of distinct pre-mRNA targets by RBM5. Studying such a modulation of the dynamics and interplay between the different domains of multi-domain splicing factor RBM5

brings us a step closer in understanding the its role in RNA recognition and alternative splicing regulation.

Appendix

Protein sequences

Protein sequences of different constructs of RBM5 used are shown below. Extra residues after TEV cleavage are shown in red. A non-native point mutation present in all constructs is shown in green (I107T). The position of protein stabilizing C191G mutation is highlighted in grey.

RRM1 (94-184):

GAMGERESKTIMLRGLPTTITESDIREMMESFEGPQPADVRLMKRKTGVSRGFAFVE
FYHLQDATSWMEANQKKLVIQGKHIAMHYSNPRPKFED

RRM1 (94-177):

GAMGERESKTIMLRGLPTTITESDIREMMESFEGPQPADVRLMKRKTGVSRGFAFVE
FYHLQDATSWMEANQKKLVIQGKHIAMHYSN

RRM2 (231-315):

GAMDTIILRNIAPHTVVDSIMTALSPYASLAVNNIRLIKDKQTQQNRGFAFVQLSSAM
DASQLLQILQSLHPPLKIDGKT IGVDFAKS

RRM1-Zf1 (94-210):

MGERESKTIMLRGLPTTITESDIREMMESFEGPQPADVRLMKRKTGVSRGFAFVEFY
HLQDATSWMEANQKKLVIQGKHIAMHYSNPRPKFEDWLCNKCCLNNFRKRLKCFR
CGADKFD

RRM1-GGS-Zf1 mutant (94-210):

MGERESKTIMLRGLPTTITESDIREMMESFEGPQPADVRLMKRKTGVSRGFAFVEFY
HLQDATSWMEANQKKLVIQGKHIAMHYSNGSGGSGSWLCNKCCLNNFRKRLKCFR
CGADKFD

RRM1-Zf1-RRM2 (94-315):

GAMGERESKTIMLRGLPTTITESDIREMMESFEGPQPADVRLMKRKTGVSRGFAFVE
FYHLQDATSWMEANQKKLVIQGKHIAMHYSNPRPKFEDWLCNKCCLNNFRKRLKC
FRCGADKFDSEQEVPPGTTESVQSVDYYCDTIILRNIAPHTVVDSIMTALSPYASLAV
NNIRLIKDKQTQQNRGFAFVQLSSAMD ASQLLQILQS LHPPLKIDGK TIGVDFAKS

NMR chemical shift assignments of RBM6 OCRE domain

Res no.	Chemical shift (ppm)	Atom name
2	49.763	CA
2	4.293	HA
2	1.288	QB
2	16.808	CB
3	119.617	N
3	8.402	H
3	52.959	CA
3	4.306	HA
3	29.975	CB
3	1.857	HB2
3	1.942	HB3
3	29.302	CG
3	2.361	HG2
3	2.419	HG3
4	110.103	N
4	8.318	H
4	42.546	CA
4	3.86	QA
5	119.789	N
5	8.144	H
5	53.219	CA
5	4.248	HA
5	26.833	CB
5	1.881	HB2
5	2.019	HB3
5	31.056	CG
5	2.244	QG
6	121.424	N
6	8.411	H
6	53.247	CA
6	4.258	HA
6	26.827	CB
6	2.017	HB2
6	1.885	HB3
6	31.124	CG
6	2.247	QG
7	117.034	N
7	8.356	H
7	55.747	CA
7	4.386	HA
7	61.122	CB
7	3.764	QB
8	117.529	N
8	8.286	H
8	55.774	CA
8	4.372	HA
8	60.964	CB
8	3.753	HB2
8	3.821	HB3
9	116.966	N
9	8.199	H
9	55.965	CA
9	4.094	HA
9	60.896	CB
9	3.706	HB2
9	3.597	HB3
10	121.721	N
10	8.132	H
10	52.178	CA
10	4.462	HA
10	38.086	CB
10	2.556	QB
11	116.968	N
11	7.859	H
11	55.719	CA
11	4.362	HA
11	25.5	CB
11	2.699	QB
12	119.599	N
12	7.622	H
12	53.754	CA
12	4.894	HA
12	37.022	CB
12	2.932	HB2
12	2.772	HB3
12	6.734	QD
12	6.423	QE
12	129.411	CD1
12	115.602	CE1
13	120.775	N
13	9.006	H
13	57.346	CA
13	4.466	HA
13	37.769	CB
13	1.89	HB
13	0.926	QG2
13	15.038	CG2
13	24.431	CG1
13	1.185	HG12
13	1.527	HG13
13	0.89	QD1
13	10.558	CD1
14	126.744	N
14	8.859	H
14	55.894	CA
14	4.175	HA
14	36.056	CB
14	2.756	QB
14	6.419	QD
14	6.47	QE
14	130.047	CD1
14	115.085	CE1
15	128.443	N
15	8.17	H
15	49.979	CA
15	4.65	HA
15	39.5	CB
15	2.294	HB2
15	2.868	HB3
16	120.875	N
16	8.445	H

Res no.	Chemical shift (ppm)	Atom name
16	57.854	CA
16	3.755	HA
16	60.293	CB
16	3.894	HB2
16	3.823	HB3
17	122.818	N
17	8.057	H
17	51.898	CA
17	4.06	HA
17	1.37	QB
17	15.983	CB
18	105.023	N
18	7.058	H
18	58.806	CA
18	3.989	HA
18	68.202	CB
18	3.431	HB
18	0.893	QG2
18	18.428	CG2
19	110.123	N
19	8.147	H
19	42.544	CA
19	3.874	HA2
19	3.523	HA3
20	117.602	N
20	7.52	H
20	51.846	CA
20	5.109	HA
20	36.974	CB
20	3.294	HB2
20	2.886	HB3
20	6.785	QD
20	6.623	QE
20	128.995	CD1
20	115.346	CE2
21	117.496	N
21	8.673	H
21	54.828	CA
21	5.11	HA

Res no.	Chemical shift (ppm)	Atom name
21	37.634	CB
21	2.752	HB2
21	2.685	HB3
21	6.94	QD
21	6.8	QE
21	129.8	CD1
21	115.669	CE1
22	123.114	N
22	9.34	H
22	53.076	CA
22	4.963	HA
22	39.059	CB
22	2.652	QB
22	6.128	QD
22	6.336	QE
22	114.737	CE1
22	130.307	CD2
23	126.166	N
23	8.245	H
23	46.704	CA
23	4.935	HA
23	39.523	CB
23	2.75	HB2
23	2.123	HB3
24	48.436	CD
24	60.798	CA
24	3.938	HA
24	29.53	CB
24	2.226	HB2
24	1.922	HB3
24	24.202	CG
24	1.922	QG
25	119.582	N
25	7.947	H
25	53.692	CA
25	4.177	HA
25	38.406	CB
25	1.744	HB2
25	1.504	HB3

Res no.	Chemical shift (ppm)	Atom name
25	24.577	CG
25	1.508	HG
25	0.841	QD1
25	0.766	QD2
25	21.826	CD1
25	20.517	CD2
26	120.367	N
26	7.121	H
26	49.536	CA
26	4.206	HA
26	1.25	QB
26	17.693	CB
27	107.629	N
27	8.301	H
27	43.434	CA
27	3.9	HA2
27	3.647	HA3
28	110.351	N
28	6.897	H
28	57.116	CA
28	4.465	HA
28	68.493	CB
28	3.916	HB
28	0.925	QG2
28	19.232	CG2
29	120.088	N
29	8.525	H
29	54.43	CA
29	5.328	HA
29	38.552	CB
29	2.535	HB2
29	2.774	HB3
29	7.039	QD
29	6.597	QE
29	130.614	CD1
29	115.579	CE1
30	119.94	N
30	9.106	H
30	54.185	CA

Res no.	Chemical shift (ppm)	Atom name
30	4.66	HA
30	39.27	CB
30	2.465	HB2
30	2.466	HB3
30	999	QB
30	6.553	QD
30	6.459	QE
30	130.112	CD1
30	115.004	CE1
31	124.619	N
31	8.182	H
31	47.736	CA
31	4.945	HA
31	39.823	CB
31	2.225	HB2
31	2.948	HB3
32	48.808	CD
32	61.439	CA
32	4.273	HA
32	29.568	CB
32	2.053	HB2
32	2.227	HB3
32	24.43	CG
32	2.044	QG
32	4.12	HD2
32	4.223	HD3
33	116.294	N
33	8.565	H
33	52.58	CA
33	4.61	HA
33	36.277	CB
33	2.81	HB2
33	2.962	HB3
34	108.488	N
34	7.45	H
34	58.958	CA
34	4.286	HA
34	67.483	CB
34	4.236	HB
34	1.099	QG2
34	18.739	CG2
35	117.685	N
35	8.229	H
35	54.444	CA
35	3.854	HA
35	24.011	CB
35	2.193	QB
35	31.647	CG
35	2.191	QG
36	116.876	N
36	7.547	H
36	51.851	CA
36	4.484	HA
36	28.574	CB
36	2.046	HB2
36	1.776	HB3
36	30.856	CG
36	2.267	QG
37	122.201	N
37	8.492	H
37	53.846	CA
37	4.375	HA
37	27.459	CB
37	1.807	QB
37	33.625	CG
37	1.997	HG2
37	2.24	HG3
38	122.351	N
38	8.044	H
38	57.856	CA
38	4.082	HA
38	31.598	CB
38	1.538	HB
38	0.495	QG1
38	0.252	QG2
38	18.19	CG1
38	16.974	CG2
39	124.507	N
39	8.234	H
39	54.745	CA
39	4.567	HA
39	36.183	CB
39	2.829	QB
39	7.02	QD
39	6.711	QE
39	130.406	CD1
39	115.361	CE1
40	127.048	N
40	7.936	H
40	56.45	CA
40	4.122	HA
40	30.647	CB
40	1.623	HB
40	0.607	QG1
40	0.442	QG2
40	17.825	CG1
40	17.819	CG2
41	48.181	CD
41	60.103	CA
41	4.164	HA
41	29.344	CB
41	2.191	HB2
41	1.798	HB3
41	24.4	CG
41	1.854	HG2
41	1.914	HG3
41	3.499	HD2
41	3.44	HD3
42	120.71	N
42	8.274	H
42	52.601	CA
42	4.203	HA
42	27.279	CB
42	1.861	HB2
42	1.981	HB3
42	31.035	CG
42	2.295	QG

Res no.	Chemical shift (ppm)	Atom name
43	123.35	N
43	8.346	H
43	49.501	CA
43	38.633	CB
43	2.463	HB2
43	2.645	HB3
44	48.101	CD
44	60.819	CA
44	4.354	HA
44	29.432	CB
44	2.194	HB2
44	1.924	HB3
44	24.214	CG
44	1.934	QG
44	3.737	QD
45	108.482	N
45	8.384	H
45	42.577	CA
45	3.855	HA2
45	3.757	HA3
46	122.789	N
46	7.947	H
46	50.164	CA
46	4.546	HA
46	39.127	CB
46	1.498	HB2
46	1.561	HB3
46	999	QB
46	24.221	CG
46	1.58	HG
46	0.84	QD1
46	0.871	QD2
46	20.527	CD1
46	22.336	CD2
47	47.954	CD
47	60.629	CA
47	4.357	HA
47	29.198	CB
47	1.913	HB2
47	2.192	HB3
47	24.346	CG
47	1.936	QG
47	3.75	HD2
47	3.568	HD3
48	126.12	N
48	7.916	H
48	55.293	CA
48	4.008	HA
48	28.523	CB
48	1.948	HB2
48	1.804	HB3
48	33.963	CG
48	2.142	QG

NMR chemical shift assignments of RBM5 RRM1 (94-184)

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
92	8.526	119.750
93	8.367	110.257
94	8.192	120.881
95	8.219	122.419
96	8.238	120.605
97	6.531	115.751
98	8.004	116.714
99	8.312	115.612
100	9.380	126.412
101	9.340	126.399
102	8.934	125.751
103	8.614	119.952
104	8.174	109.217
105	8.094	119.463
107	8.877	115.869
108	7.046	106.861
109	6.894	121.176
110	9.113	119.092
111	9.323	120.774
112	8.229	113.701
113	7.587	122.037
114	7.622	119.158
115	8.124	119.923
116	8.118	118.651
117	7.793	119.007
118	7.920	116.898
119	7.706	119.029
120	7.607	112.537
121	7.475	122.355
122	8.444	121.379

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
123	8.101	108.527
125	8.388	120.053
127	8.490	123.590
128	7.371	113.160
129	7.789	122.202
130	8.887	124.574
131	9.420	126.990
132	8.022	124.848
133	8.576	122.477
134	8.672	120.291
136	7.171	106.365
137	8.143	109.566
138	7.021	118.828
139	8.686	119.818
140	8.788	122.628
141	9.014	106.860
142	7.044	113.456
143	8.507	120.345
144	9.003	117.934
145	9.487	123.876
146	8.320	126.530
147	8.878	123.152
148	9.685	118.778
149	7.305	111.000
151	9.211	116.048
152	7.187	119.482
153	6.895	122.173
154	8.424	110.021
155	7.868	118.629
156	8.211	124.900

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
157	8.686	119.594
158	8.120	117.507
159	7.555	118.350
160	7.485	111.679
161	7.968	119.846
162	8.443	115.313
163	7.542	117.678
164	7.942	121.991
165	8.040	127.819
166	8.456	124.850
167	9.455	126.995
168	8.253	102.712
169	7.827	121.397
170	8.795	122.907
171	8.209	127.090
172	8.130	130.100
173	8.063	117.616
174	8.302	117.300
175	8.806	122.624
176	8.779	116.560
177	8.556	124.410
179	8.237	121.190
181	8.111	120.351
182	8.086	120.512
183	8.262	121.901
184	7.876	126.770

NMR chemical shift assignments of RBM5 RRM1-Zf1

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
95	8.434	123.015
96	8.174	121.617
97	6.441	116.108
98	7.952	116.289
99	8.418	115.698
100	9.308	126.108
101	9.313	125.665
102	8.973	125.603
103	8.684	119.915
104	8.213	109.177
105	8.077	119.193
107	8.885	116.060
108	7.092	106.767
109	6.940	121.233
110	9.060	118.752
111	9.254	121.054
112	8.276	113.677
113	7.622	121.921
114	7.662	119.305
115	8.168	120.049
116	8.156	118.658
117	7.845	119.272
118	7.987	117.131
119	7.761	118.810
120	7.640	112.893
121	7.496	122.226
122	8.440	121.536
123	8.025	108.661
125	8.410	120.129
127	8.508	123.572
128	7.400	113.107
129	7.813	121.739
130	8.886	125.081
131	9.421	127.444
132	8.112	125.161
133	8.422	121.630
134	8.749	120.369
136	7.198	106.683
137	8.174	109.425
138	7.166	118.819
139	8.635	119.452
140	8.672	122.125
141	8.903	106.755
142	7.222	114.182
143	8.554	120.275
144	8.999	118.156
145	9.480	124.189
146	8.270	125.889
147	8.977	122.870
148	9.710	118.651
149	7.371	110.771
151	9.176	115.606
152	7.163	119.321
153	6.883	122.151
154	8.520	109.435
155	7.843	118.849
156	8.148	125.063
157	8.751	119.690
158	8.154	117.623
159	7.583	118.389
160	7.413	111.602
161	8.114	120.476
162	8.566	115.230
163	7.609	117.861
164	7.948	121.801
165	8.116	127.897
166	8.467	124.786
167	9.445	126.948
168	8.244	102.699
169	7.869	121.388
170	8.900	122.692
171	8.191	126.885
172	8.186	130.242
173	8.139	117.776
174	8.346	116.728
175	8.825	122.625
176	8.630	114.174
177	9.296	125.645
179	9.006	124.899
181	8.435	122.848
182	8.537	117.368
183	7.048	116.455
184	8.446	123.270
185	8.207	115.769
186	9.485	122.005
187	8.767	130.277
188	9.070	126.833
189	9.549	122.560
190	9.023	120.166
191	7.587	116.329
192	8.120	122.292
193	8.542	124.785
194	9.000	126.254
195	7.797	119.107
196	8.029	123.053
198	8.472	118.845
199	9.047	123.252
200	7.676	117.061
201	9.809	126.268
202	9.249	131.761
203	8.873	120.846
204	8.440	118.470
205	7.620	112.164
206	8.912	125.789
207	8.677	121.483
208	7.503	124.695
209	8.520	115.809
210	7.300	125.309

NMR chemical shift assignments of RBM5 RRM1-Zf1-RRM2 C191G

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
92	8.546	119.553
93	8.369	110.049
94	8.205	120.754
95	8.264	122.620
96	8.183	120.750
97	6.519	115.723
98	8.035	116.571
99	8.384	115.808
100	9.372	126.245
101	9.392	126.232
102	8.943	125.517
104	8.202	109.433
105	8.092	119.248
107	8.887	116.034
108	7.088	106.801
109	6.934	121.203
110	9.085	118.818
111	9.315	120.976
112	8.268	113.675
113	7.619	121.956
114	7.663	119.349
115	8.161	119.851
116	8.163	118.667
117	7.853	119.258
118	8.000	117.139
119	7.772	118.835
120	7.651	112.936
121	7.499	122.157
122	8.436	121.309
123	8.024	108.639
125	8.404	120.027
127	8.519	123.506
128	7.431	113.197
129	7.793	122.023
130	8.928	124.902
131	9.426	127.198
132	8.093	125.045
133	8.520	122.088
134	8.726	120.288
136	7.218	106.691
137	8.175	109.544
138	7.100	118.825
139	8.660	119.662
141	8.987	106.815
143	8.535	120.342
144	8.992	117.824
145	9.504	123.794
146	8.338	126.578
148	9.746	118.485
151	9.228	116.011
152	7.239	119.582
153	6.911	122.138
154	8.502	109.744
155	7.872	118.651
156	8.156	124.857
158	8.181	117.572
159	7.604	118.536
163	7.607	117.818
164	7.972	121.882
165	8.082	127.594
166	8.476	124.882
167	9.436	126.892
168	8.239	102.693
169	7.865	121.440
170	8.838	122.904
171	8.238	126.985
172	8.142	130.129
173	8.037	117.601
174	8.339	117.213
176	8.735	115.789
187	8.640	128.911
188	9.181	128.739
189	9.243	122.392
190	8.847	119.057
193	8.374	124.068
194	9.092	124.853
198	8.345	119.196
200	7.687	117.831
201	9.669	126.078
202	9.141	131.454
203	8.866	120.602
204	8.383	118.311
205	7.656	112.336
206	8.839	125.691
207	8.608	121.102
208	7.709	124.570
210	7.817	120.481
211	8.124	115.845
212	8.339	121.680
213	8.187	120.293
214	8.330	122.656
215	8.166	122.699
218	8.461	109.199
219	7.956	113.604
220	8.192	116.139
221	8.387	123.049
222	8.278	116.959
223	8.100	121.700
224	8.366	123.659
225	8.275	117.487
226	8.020	120.698
227	8.121	122.841
228	7.733	119.793
229	7.824	120.809
230	7.518	120.932
231	8.100	121.474
232	7.947	116.091
233	9.048	122.912
234	9.226	123.172
235	8.795	125.679
236	9.037	120.204
237	8.216	115.597
238	7.634	116.119
239	9.161	131.462

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
241	7.961	109.837
242	6.807	121.013
243	8.242	119.357
244	8.794	122.446
245	8.749	118.882
246	7.498	113.753
247	7.323	122.436
248	8.045	118.192
249	8.458	115.594
250	7.477	122.778
251	7.595	111.575
252	7.918	117.758
254	7.706	113.008
255	7.848	120.011
256	8.464	115.988
257	8.028	125.956
258	8.256	122.910
259	8.589	120.311
261	8.088	116.201
262	7.652	119.656
263	9.180	126.572
264	8.855	127.792
265	6.814	124.303
266	8.972	129.524
267	8.602	122.941

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
268	8.844	126.964
269	8.573	116.740
270	8.075	107.985
271	8.229	114.531
272	7.713	116.539
273	8.788	119.453
274	8.366	121.677
275	8.661	104.019
276	6.667	114.572
277	8.940	122.105
278	8.747	118.617
279	9.304	122.960
280	8.872	129.354
281	8.763	127.748
282	8.694	112.372
283	7.486	109.867
284	8.912	125.108
285	8.339	119.118
286	7.915	120.085
287	7.922	122.842
288	8.539	112.746
289	8.131	120.132
290	7.875	119.467
291	8.401	119.749
292	7.775	116.125

Res no.	Chemical shift (ppm) -H	Chemical shift (ppm)-N
293	7.768	120.368
294	8.515	117.957
295	8.489	115.723
296	7.396	113.061
297	6.753	122.924
298	8.263	120.639
301	8.338	121.504
302	7.974	126.144
303	8.181	119.809
304	9.646	130.548
305	8.535	101.565
306	7.531	120.691
307	8.669	119.598
308	8.330	131.886
309	7.991	114.140
310	8.364	118.977
311	8.320	123.229
312	9.160	119.770
313	8.427	123.807
314	8.122	121.035
315	7.883	122.876

Note: Backbone assignments of RRM1 (94-184) were done in SEC buffer 2 while that of RRM1-Zf1 and RRM1-Zf1-RRM2 C191G were done in SEC buffer 1.

Abbreviations

1D, 2D, 3D	One-, Two-, Three-Dimensional
AEBSF	4- Benzenesulfonyl fluoride hydrochloride
BME	β-mercaptoethanol
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear single quantum coherence spectroscopy
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Equilibrium Dissociation Constant
kDa	Kilo Dalton
LB	Lysogeny Broth Medium
MWCO	Molecular weight cut-off
$NI^{2+}/Zn^{2+}/Cd^{2+}$	Nickle/Zinc/Cadmium
OD	Optical Density at 600nm Wavelength
PMSF	Phenylmethylsulfonyl fluoride
CV	Column volume
wt	wild-type
PCR	Polymerase Chain Reaction
ADAR	Adenosine Deaminase acting on RNA
pre-mRNA	precursor messenger RNA
SEC	Size Exclusion Chromatography
R_g	Radius of gyration
RMSD	Root mean square deviation
SAXS	Small angle X-ray scattering
CD	Circular Dichroism
TEV	Tobacco Etch Virus
TOCSY	Total correlation spectroscopy
TROSY	Transverse relaxation optimized spectroscopy
UHM/ULM	U2AF homology motif/UHM-ligand motif
PPII helix	Poly-proline type II helix
PRM	Poly-proline rich motif
RRM	RNA recognition motif
RanBP2	Ran Binding Protein 2
SELEX	Systematic Evolution of Ligands by Exponential enrichment

Table of figures

List of Tables

Acknowledgements

I would like to sincerely thank my supervisor Prof. Dr. Michael Sattler for giving me the opportunity to work towards my doctoral degree in his lab. I really appreciate our short meetings where I often got his valuable suggestions that helped push the projects forward. He gave me freedom to work on a number of projects that helped in expanding my practical knowledge. The time spent in his lab has helped me evolve both professionally and personally.

I would like to thank the coordinators of IMPRS-LS Graduate School -Dr. Hans Joerg Schaeffer**,** Dr. Ingrid Wolf and Maximiliane Reif for providing an excellent support system. They made my life much easier when I initially came to Munich and also afterwards. The graduate school helped to fund my major conferences in addition to providing me the opportunity to attend a variety of workshops held by the IMPRS-LS graduate school. I am really happy to have been a part of it.

I express my gratitude towards my Thesis Advisory Committee members Prof. Dr. Andreas Ladurner (LMU) and Prof. Dr. Iris Antes (TUM) for their extremely valuable suggestions. The yearly meetings helped me gain perspective in my projects. I would especially like to thank my collaborators in University of Cologne, Dr. Jay Gopalakrishnan, Anand Ramani, Dr. Arul Mariappan and Arpit Wason; in Centre de Regulació Genòmica, Barcelona, Dr. Sophie Bonnal and Dr. Juan Valcárcel; and in University of Toronto, Dr. Debashish Ray and Dr. Tim Hughes, for very exciting and fruitful collaborations.

I am grateful to Dr. Lisa Warner for mentoring me during my initial years. Even though she was not my official supervisor, she never shied away from helping me. She proved to be a very good teacher and friend to me. It is difficult to imagine how things would have been without her. Also, I inherited her project after she left the group which in the end turned out to be my main project. I also thank Dr. André Mourão who was also always just an email away. The collaboration with him resulted in an important, timely publication.

I sincerely thank Dr. Arie Geerlof, Astrid Lauxen and Dr. Ana Messias who have created and maintained a great atmosphere in the lab. It is only due to them that everything works so seamlessly in HMGU. I am thankful to Sam and Gerd for maintenance of the NMR facility, providing all the technical help and Waltraud for the administrative work. I also thank Ralf for the maintenance of the SAXS facility. I am grateful to Rainer for ensuring the

accessibility of all possible software as the IT administrator. I am also thankful to Dr Grzegorz Popowicz and Dr. Robert Janowski for their occasional help with crystallography.

I express my deepest regards to my friends Carolina, Eleni, Nishtha, Diana, Leo, Ashish, Martin, and Miriam who made this journey extremely memorable. I thank them for all the nice experiences we shared and for always just being there. I am thankful to all members of the Sattler group for the scientific and not so scientific discussions we had.

I am deeply indebted to my parents and little brother who have always believed in me. They always stood by me loving, supporting and encouraging me in every decision I have made. And to Pravin, who has played many roles in my life as my mentor, teacher and friend but most importantly as the love of my life. I cannot even begin to imagine my life without him. I am truly thankful to Michael for placing me in HMGU and in the same office as Pravin!

Bibliography

Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff and G. Dreyfuss (1986). "mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence." Mol Cell Biol **6**(8): 2932-2943.

Ahlner, A., M. Carlsson, B. H. Jonsson and P. Lundstrom (2013). "PINT: a software for integration of peak volumes and extraction of relaxation rates." J Biomol NMR **56**(3): 191-202.

Al-Ayoubi, A. M., H. Zheng, Y. Liu, T. Bai and S. T. Eblen (2012). "Mitogen-activated protein kinase phosphorylation of splicing factor 45 (SPF45) regulates SPF45 alternative splicing site utilization, proliferation, and cell adhesion." Mol Cell Biol **32**(14): 2880-2893.

Angeloni, D. (2007). "Molecular analysis of deletions in human chromosome 3p21 and the role of resident cancer genes in disease." Brief Funct Genomic Proteomic **6**(1): 19-39.

Aravind, L. and E. V. Koonin (1999). "G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins." Trends Biochem Sci **24**(9): 342-344.

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

Avis, J. M., F. H. Allain, P. W. Howe, G. Varani, K. Nagai and D. Neuhaus (1996). "Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding." J Mol Biol **257**(2): 398-411.

Battiste, J. L. and G. Wagner (2000). "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data." Biochemistry **39**(18): 5355-5365.

Bechara, E. G., E. Sebestyen, I. Bernardis, E. Eyras and J. Valcarcel (2013). "RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation." Mol Cell **52**(5): 720-733.

Berglund, J. A., K. Chua, N. Abovich, R. Reed and M. Rosbash (1997). "The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC." Cell **89**(5): 781-787.

Bertini, I., C. Luchinat and M. Piccioli (2001). "Paramagnetic probes in metalloproteins." Methods Enzymol **339**: 314-340.

Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub and R. Luhrmann (2008). "Isolation of an active step I spliceosome and composition of its RNP core." Nature **452**(7189): 846-850.

Beuck, C., B. R. Szymczyna, D. E. Kerkow, A. B. Carmel, L. Columbus, R. L. Stanfield and J. R. Williamson (2010). "Structure of the GLD-1 homodimerization domain: insights into STAR protein-mediated translational regulation." Structure **18**(3): 377-389.

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

Blencowe, B. J., S. Ahmad and L. J. Lee (2009). "Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes." Genes Dev **23**(12): 1379-1386.

Bloembergen, N. and L. O. Morgan (1961). "Proton relaxation times in paramagnetic solutions. Effects of electron spin relaxation." The Journal of Chemical Physics **34**: 842-850.

Bloembergen, N., E. M. Purcell and R. V. Pound (1948). "Relaxation Effects in Nuclear Magnetic Resonance Absorption." Physical Review **73**(7): 679-712.

Bonnal, S., C. Martinez, P. Forch, A. Bachi, M. Wilm and J. Valcarcel (2008). "RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition." Mol Cell **32**(1): 81-95.

Bouillet, P. and L. A. O'Reilly (2009). "CD95, BIM and T cell homeostasis." Nat Rev Immunol **9**(7): 514-519.
Braddock, D. T., J. M. Louis, J. L. Baber, D. Levens and G. M. Clore (2002). "Structure and dynamics of KH domains from FBP bound to single-stranded DNA." Nature **415**(6875): 1051-1056.
Brown, A. M. and N. J. Zondlo (2012). "A propensity scale for type II polyproline helices (PPII): aromatic amino acids in proline-rich sequences strongly disfavor PPII due to proline-aromatic interactions." Biochemistry **51**(25): 5041-5051.
Burge CB, T. T., Sharp PA (1999). Splicing of Precursors to mRNAs by the Spliceosomes., Cold Spring Harbor Laboratoty Press, Cold Spring Harbor, New York. **The RNA world Second edition:** 525-560.
Callebaut, I. and J. P. Mornon (2005). "OCRE: a novel domain made of imperfect, aromatic-rich octamer repeats." Bioinformatics **21**(6): 699-702.
Campbell, Z. T., D. Bhimsaria, C. T. Valley, J. A. Rodriguez-Martinez, E. Menichelli, J. R. Williamson, A. Z. Ansari and M. Wickens (2012). "Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity." Cell Rep **1**(5): 570-581.
Campbell, Z. T., C. T. Valley and M. Wickens (2014). "A protein-RNA specificity code enables targeted activation of an endogenous human transcript." Nat Struct Mol Biol **21**(8): 732-738.
Cascino, I., G. Fiucci, G. Papoff and G. Ruberti (1995). "Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing." J Immunol **154**(6): 2706-2713.
Chao, J. A., J. H. Lee, B. R. Chapados, E. W. Debler, A. Schneemann and J. R. Williamson (2005). "Dual modes of RNA-silencing suppression by Flock House virus protein B2." Nat Struct Mol Biol **12**(11): 952-957.
Chen, M. and J. L. Manley (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches." Nat Rev Mol Cell Biol **10**(11): 741-754.
Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.
Cheng, J., T. Zhou, C. Liu, J. P. Shapiro, M. J. Brauer, M. C. Kiefer, P. J. Barr and J. D. Mountz (1994). "Protection from Fas-mediated apoptosis by a soluble form of the Fas molecule." Science **263**(5154): 1759-1762.
Clery, A., S. Jayne, N. Benderska, C. Dominguez, S. Stamm and F. H. Allain (2011). "Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1." Nat Struct Mol Biol **18**(4): 443-450.
Collins, K. M., Y. A. Kainov, E. Christodolou, D. Ray, Q. Morris, T. Hughes, I. A. Taylor, E. V. Makeyev and A. Ramos (2017). "An RRM-ZnF RNA recognition module targets RBM10 to exonic sequences to promote exon exclusion." Nucleic Acids Res.
Cordier, F., M. Rogowski, S. Grzesiek and A. Bax (1999). "Observation of through-hydrogen-bond 2hJHC' in a perdeuterated protein." J Magn Reson **140**(2): 510-512.
Cornilescu, G., J. L. Marquardt, M. Ottiger and A. Bax (1998). "Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase." Journal of the American Chemical Society **120**(27): 6836-6837.
Corsini, L., S. Bonnal, J. Basquin, M. Hothorn, K. Scheffzek, J. Valcarcel and M. Sattler (2007). "U2AF-homology motif interactions are required for alternative splicing regulation by SPF45." Nat Struct Mol Biol **14**(7): 620-629.

Cote, J., S. Dupuis, Z. Jiang and J. Y. Wu (2001). "Caspase-2 pre-mRNA alternative splicing: Identification of an intronic element containing a decoy 3' acceptor site." Proc Natl Acad Sci U S A **98**(3): 938-943.
Cote, J., S. Dupuis and J. Y. Wu (2001). "Polypyrimidine track-binding protein binding downstream of caspase-2 alternative exon 9 represses its inclusion." J Biol Chem **276**(11): 8535-8543.
Dang, T. P., A. F. Gazdar, A. K. Virmani, T. Sepetavec, K. R. Hande, J. D. Minna, J. R. Roberts and D. P. Carbone (2000). "Chromosome 19 translocation, overexpression of Notch3, and human lung cancer." J Natl Cancer Inst **92**(16): 1355-1357.
Daragan, V. A., E. E. Ilyina, C. G. Fields, G. B. Fields and K. H. Mayo (1997). "Backbone and side-chain dynamics of residues in a partially folded beta-sheet peptide from platelet factor-4." Protein Sci **6**(2): 355-363.
Darnell, R. B. (2010). "HITS-CLIP: panoramic views of protein-RNA regulation in living cells." Wiley Interdiscip Rev RNA **1**(2): 266-286.
David, C. J. and J. L. Manley (2010). "Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged." Genes Dev **24**(21): 2343-2364.
Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson and D. C. Richardson (2007). "MolProbity: all-atom contacts and structure validation for proteins and nucleic acids." Nucleic Acids Res **35**(Web Server issue): W375-383.
Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.
Deo, R. C., J. B. Bonanno, N. Sonenberg and S. K. Burley (1999). "Recognition of polyadenylate RNA by the poly(A)-binding protein." Cell **98**(6): 835-845.
Doreleijers, J. F., A. W. Sousa da Silva, E. Krieger, S. B. Nabuurs, C. A. Spronk, T. J. Stevens, W. F. Vranken, G. Vriend and G. W. Vuister (2012). "CING: an integrated residue-based structure validation program suite." J Biomol NMR **54**(3): 267-283.
Doreleijers, J. F., W. F. Vranken, C. Schulte, J. L. Markley, E. L. Ulrich, G. Vriend and G. W. Vuister (2012). "NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB." Nucleic Acids Res **40**(Database issue): D519-524.
Dosset, P., J. C. Hus, D. Marion and M. Blackledge (2001). "A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings." J Biomol NMR **20**(3): 223-231.
Drake, A. F., G. Siligardi and W. A. Gibbons (1988). "Reassessment of the electronic circular dichroism criteria for random coil conformations of poly(L-lysine) and the implications for protein folding and denaturation studies." Biophys Chem **31**(1-2): 143-146.
Dreyfuss, G., M. S. Swanson and S. Pinol-Roma (1988). "Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation." Trends Biochem Sci **13**(3): 86-91.
Edamatsu, H., Y. Kaziro and H. Itoh (2000). "LUCA15, a putative tumour suppressor gene encoding an RNA-binding nuclear protein, is down-regulated in ras-transformed Rat-1 cells." Genes Cells **5**(10): 849-858.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.
Farina, B., R. Fattorusso and M. Pellecchia (2011). "Targeting zinc finger domains with small molecules: solution structure and binding studies of the RanBP2-type zinc finger of RBM5." Chembiochem **12**(18): 2837-2845.

Feracci, M., J. N. Foot, S. N. Grellscheid, M. Danilenko, R. Stehle, O. Gonchar, H. S. Kang, C. Dalgliesh, N. H. Meyer, Y. Liu, A. Lahat, M. Sattler, I. C. Eperon, D. J. Elliott and C. Dominguez (2016). "Structural basis of RNA recognition and dimerization by the STAR proteins T-STAR and Sam68." Nat Commun **7**: 10355.
Fernandez, C. and G. Wider (2003). "TROSY in NMR studies of the structure and function of large biological macromolecules." Curr Opin Struct Biol **13**(5): 570-580.
Forch, P., O. Puig, C. Martinez, B. Seraphin and J. Valcarcel (2002). "The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites." EMBO J **21**(24): 6882-6892.
Fouraux, M. A., M. J. Kolkman, A. Van der Heijden, A. S. De Jong, W. J. Van Venrooij and G. J. Pruijn (2002). "The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase." RNA **8**(11): 1428-1443.
Fushimi, K., P. Ray, A. Kar, L. Wang, L. C. Sutherland and J. Y. Wu (2008). "Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5." Proc Natl Acad Sci U S A **105**(41): 15708-15713.
Gardner, K. H. and L. E. Kay (1998). "The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct **27**: 357-406.
Gerstberger, S., M. Hafner, M. Ascano and T. Tuschl (2014). "Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease." Adv Exp Med Biol **825**: 1-55.
Glisovic, T., J. L. Bachorik, J. Yong and G. Dreyfuss (2008). "RNA-binding proteins and post-transcriptional gene regulation." FEBS Lett **582**(14): 1977-1986.
Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." RNA **6**(9): 1197-1211.
Graveley, B. R. and T. Maniatis (1998). "Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing." Mol Cell **1**(5): 765-771.
Green, D. W., V. M. Ingram and M. F. Perutz (1954). "The Structure of Haemoglobin. IV. Sign Determination by the Isomorphous Replacement Method." Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **225**: 287-307.
Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol Biol **278**: 353-378.
Guntert, P. and L. Buchner (2015). "Combined automated NOE assignment and structure calculation with CYANA." J Biomol NMR **62**(4): 453-471.
Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, Jr., A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan and T. Tuschl (2010). "Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP." Cell **141**(1): 129-141.
Hall, T. M. (2005). "Multiple modes of RNA recognition by zinc finger proteins." Curr Opin Struct Biol **15**(3): 367-373.
Handa, N., O. Nureki, K. Kurimoto, I. Kim, H. Sakamoto, Y. Shimura, Y. Muto and S. Yokoyama (1999). "Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein." Nature **398**(6728): 579-585.
Havlioglu, N., J. Wang, K. Fushimi, M. D. Vibranovski, Z. Kan, W. Gish, A. Fedorov, M. Long and J. Y. Wu (2007). "An intronic signal for alternative splicing in the human genome." PLoS One **2**(11): e1246.
Heath, E., F. Sablitzky and G. T. Morgan (2010). "Subnuclear targeting of the RNA-binding motif protein RBM6 to splicing speckles and nascent transcripts." Chromosome Res **18**(8): 851-872.

Hegele, A., A. Kamburov, A. Grossmann, C. Sourlis, S. Wowro, M. Weimann, C. L. Will, V. Pena, R. Luhrmann and U. Stelzl (2012). "Dynamic protein-protein interaction wiring of the human spliceosome." Mol Cell **45**(4): 567-580.
Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.
Hendrickson, W. A. and C. M. Ogata (1997). "[28] Phase determination from multiwavelength anomalous diffraction measurements." Methods in Enzymology **276**: 494-523.
Hennig, J., C. Militti, G. M. Popowicz, I. Wang, M. Sonntag, A. Geerlof, F. Gabel, F. Gebauer and M. Sattler (2014). "Structural basis for the assembly of the Sxl-Unr translation regulatory complex." Nature **515**(7526): 287-290.
Ho, L. H., R. Taylor, L. Dorstyn, D. Cakouros, P. Bouillet and S. Kumar (2009). "A tumor suppressor function for caspase-2." Proc Natl Acad Sci U S A **106**(13): 5336-5341.
Houben, K., E. Wasielewski, C. Dominguez, E. Kellenberger, R. A. Atkinson, H. T. Timmers, B. Kieffer and R. Boelens (2005). "Dynamics and metal exchange properties of C4C4 RING domains from CNOT4 and the p44 subunit of TFIIH." J Mol Biol **349**(3): 621-637.
Huang, J. R., L. R. Warner, C. Sanchez, F. Gabel, T. Madl, C. D. Mackereth, M. Sattler and M. Blackledge (2014). "Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study." J Am Chem Soc **136**(19): 7068-7076.
Hudson, B. P., M. A. Martinez-Yamout, H. J. Dyson and P. E. Wright (2004). "Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d." Nat Struct Mol Biol **11**(3): 257-264.
Hughes, P. D., G. T. Belz, K. A. Fortner, R. C. Budd, A. Strasser and P. Bouillet (2008). "Apoptosis regulators Fas and Bim cooperate in shutdown of chronic immune responses and prevention of autoimmunity." Immunity **28**(2): 197-205.
Inoue, A., N. Yamamoto, M. Kimura, K. Nishio, H. Yamane and K. Nakajima (2014). "RBM10 regulates alternative splicing." FEBS Lett **588**(6): 942-947.
Ishii, Y., M. A. Markus and R. Tycko (2001). "Controlling residual dipolar couplings in high-resolution NMR of proteins by strain induced alignment in a gel." J Biomol NMR **21**(2): 141-151.
Izquierdo, J. M., N. Majos, S. Bonnal, C. Martinez, R. Castelo, R. Guigo, D. Bilbao and J. Valcarcel (2005). "Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition." Mol Cell **19**(4): 475-484.
Jacks, A., J. Babon, G. Kelly, I. Manolaridis, P. D. Cary, S. Curry and M. R. Conte (2003). "Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element." Structure **11**(7): 833-843.
Jiang, Z. H., W. J. Zhang, Y. Rao and J. Y. Wu (1998). "Regulation of Ich-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors." Proc Natl Acad Sci U S A **95**(16): 9155-9160.
Jin, W., Z. Niu, D. Xu and X. Li (2012). "RBM5 promotes exon 4 skipping of AID pre-mRNA by competing with the binding of U2AF65 to the polypyrimidine tract." FEBS Lett **586**(21): 3852-3857.
Jurica, M. S. and M. J. Moore (2003). "Pre-mRNA splicing: awash in a sea of proteins." Mol Cell **12**(1): 5-14.
Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.
Kay, L. E., D. A. Torchia and A. Bax (1989). "Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**(23): 8972-8979.

Kay, L. E., D. A. Torchia and A. Bax (1989). "Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**(23): 8972-8979.
Keeler, J. (2002). Understanding NMR Spectroscopy. University of Cambridge, Department of Chemistry, Wiley.
Kelly, M. A., B. W. Chellgren, A. L. Rucker, J. M. Troutman, M. G. Fried, A. F. Miller and T. P. Creamer (2001). "Host-guest study of left-handed polyproline II helix formation." Biochemistry **40**(48): 14376-14383.
Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-666.
Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore (1960). "Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution." Nature **185**(4711): 422-427.
Kielkopf, C. L., S. Lucke and M. R. Green (2004). "U2AF homology motifs: protein recognition in the RRM world." Genes Dev **18**(13): 1513-1526.
Kiledjian, M. and G. Dreyfuss (1992). "Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box." EMBO J **11**(7): 2655-2664.
Konarev, P., V. Volkov, A. Sokolova, M. Koch and D. Svergun (2003). "PRIMUS: a Windows PC-based system for small-angle scattering data analysis." Journal of Applied Crystallography **36**(5): 1277-1282.
Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe and J. Ule (2010). "iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution." Nat Struct Mol Biol **17**(7): 909-915.
Kosen, P. A., R. M. Scheek, H. Naderi, V. J. Basus, S. Manogaran, P. G. Schmidt, N. J. Oppenheimer and I. D. Kuntz (1986). "Two-dimensional 1H NMR of three spin-labeled derivatives of bovine pancreatic trypsin inhibitor." Biochemistry **25**(9): 2356-2364.
Krammer, P. H. (2000). "CD95's deadly mission in the immune system." Nature **407**(6805): 789-795.
Krishna, S. S., I. Majumdar and N. V. Grishin (2003). "Structural classification of zinc fingers: survey and summary." Nucleic Acids Res **31**(2): 532-550.
Kumar, S. (2009). "Caspase 2 in apoptosis, the DNA damage response and tumour suppression: enigma no more?" Nat Rev Cancer **9**(12): 897-903.
Lahiri, D. K. and J. O. Thomas (1986). "A cDNA clone of the hnRNP C proteins and its homology with the single-stranded DNA binding protein UP2." Nucleic Acids Res **14**(10): 4077-4094.
Laity, J. H., B. M. Lee and P. E. Wright (2001). "Zinc finger proteins: new insights into structural and functional diversity." Curr Opin Struct Biol **11**(1): 39-46.
Lee, M. S., G. P. Gippert, K. V. Soman, D. A. Case and P. E. Wright (1989). "Three-dimensional solution structure of a single zinc finger DNA-binding domain." Science **245**(4918): 635-637.
Legrain, P., B. Seraphin and M. Rosbash (1988). "Early commitment of yeast pre-mRNA to the spliceosome pathway." Mol Cell Biol **8**(9): 3755-3760.
Lerman, M. I. and J. D. Minna (2000). "The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium." Cancer Res **60**(21): 6116-6133.
Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin and J. A. Steitz (1980). "Are snRNPs involved in splicing?" Nature **283**(5743): 220-224.

Licatalosi, D. D., A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell and R. B. Darnell (2008). "HITS-CLIP yields genome-wide insights into brain alternative RNA processing." Nature **456**(7221): 464-469.
Linge, J. P., M. Habeck, W. Rieping and M. Nilges (2003). "ARIA: automated NOE assignment and NMR structure calculation." Bioinformatics **19**(2): 315-316.
Linge, J. P., M. A. Williams, C. A. Spronk, A. M. Bonvin and M. Nilges (2003). "Refinement of protein structures in explicit solvent." Proteins **50**(3): 496-506.
Lingel, A., B. Simon, E. Izaurralde and M. Sattler (2005). "The structure of the flock house virus B2 protein, a viral suppressor of RNA interference, shows a novel mode of double-stranded RNA recognition." EMBO Rep **6**(12): 1149-1155.
Liu, Z., I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Kramer and M. Sattler (2001). "Structural basis for recognition of the intron branch site RNA by splicing factor 1." Science **294**(5544): 1098-1102.
Liu, Y., L. Conaway, J. Rutherford Bethard, A. M. Al-Ayoubi, A. Thompson Bradley, H. Zheng, S. A. Weed and S. T. Eblen (2013). "Phosphorylation of the alternative mRNA splicing factor 45 (SPF45) by Clk1 regulates its splice site utilization, cell migration and invasion." Nucleic Acids Res **41**(9): 4949-4962.
Long, J. C. and J. F. Caceres (2009). "The SR protein family of splicing factors: master regulators of gene expression." Biochem J **417**(1): 15-27.
Lorenz, C., T. Gesell, B. Zimmermann, U. Schoeberl, I. Bilusic, L. Rajkowitsch, C. Waldsich, A. von Haeseler and R. Schroeder (2010). "Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts." Nucleic Acids Res **38**(11): 3794-3808.
Lorieau, J., L. Yao and A. Bax (2008). "Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins." J Am Chem Soc **130**(24): 7536-7537.
Loughlin, F. E., R. E. Mansfield, P. M. Vaz, A. P. McGrath, S. Setiyaputra, R. Gamsjaeger, E. S. Chen, B. J. Morris, J. M. Guss and J. P. Mackay (2009). "The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences." Proc Natl Acad Sci U S A **106**(14): 5581-5586.
Mackereth, C. D., T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcarcel and M. Sattler (2011). "Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF." Nature **475**(7356): 408-411.
Maiti, R., G. H. Van Domselaar, H. Zhang and D. S. Wishart (2004). "SuperPose: a simple server for sophisticated structural superposition." Nucleic Acids Res **32**(Web Server issue): W590-594.
Maraver, A., P. J. Fernandez-Marcos, D. Herranz, M. Canamero, M. Munoz-Martin, G. Gomez-Lopez, F. Mulero, D. Megias, M. Sanchez-Carbayo, J. Shen, M. Sanchez-Cespedes, T. Palomero, A. Ferrando and M. Serrano (2012). "Therapeutic effect of gamma-secretase inhibition in KrasG12V-driven non-small cell lung carcinoma by derepression of DUSP1 and inhibition of ERK." Cancer Cell **22**(2): 222-234.
Marchese, D., N. S. de Groot, N. Lorenzo Gotor, C. M. Livi and G. G. Tartaglia (2016). "Advances in the characterization of RNA-binding proteins." Wiley Interdiscip Rev RNA **7**(6): 793-810.
Maris, C., C. Dominguez and F. H. Allain (2005). "The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression." FEBS J **272**(9): 2118-2131.
Martell, A. E. and R. M. Smith (1974). Critical Stability Constants. New York, Plenum Press.
Martin, B. T., P. Serrano, M. Geralt and K. Wuthrich (2016). "Nuclear Magnetic Resonance Structure of a Novel Globular Domain in RBM10 Containing OCRE, the Octamer Repeat Sequence Motif." Structure **24**(1): 158-164.

Massi, F., E. Johnson, C. Wang, M. Rance and A. G. Palmer, 3rd (2004). "NMR R1 rho rotating-frame relaxation with weak radio frequency fields." J Am Chem Soc **126**(7): 2247-2256.
Mazza, C., A. Segref, I. W. Mattaj and S. Cusack (2002). "Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex." EMBO J **21**(20): 5548-5557.
Matlin, A. J., F. Clark and C. W. Smith (2005). "Understanding alternative splicing: towards a cellular code." Nat Rev Mol Cell Biol **6**(5): 386-398.
Meyer, N. H., K. Tripsianes, M. Vincendeau, T. Madl, F. Kateb, R. Brack-Werner and M. Sattler (2010). "Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain." J Biol Chem **285**(37): 28893-28901.
Moore, M. J. and P. A. Sharp (1993). "Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing." Nature **365**(6444): 364-368.
Mourao, A., S. Bonnal, K. Soni, L. Warner, R. Bordonne, J. Valcarcel and M. Sattler (2016). "Structural basis for the recognition of spliceosomal SmN/B/B' proteins by the RBM5 OCRE domain in splicing regulation." Elife **5**.
Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.
Nagata, S. and P. Golstein (1995). "The Fas death factor." Science **267**(5203): 1449-1456.
Nelson, K. K. and M. R. Green (1989). "Mammalian U2 snRNP has a sequence-specific RNA-binding activity." Genes Dev **3**(10): 1562-1571.
Nguyen, C. D., R. E. Mansfield, W. Leung, P. M. Vaz, F. E. Loughlin, R. P. Grant and J. P. Mackay (2011). "Characterization of a family of RanBP2-type zinc fingers that can recognize single-stranded RNA." J Mol Biol **407**(2): 273-283.
Nilsen, T. W. and B. R. Graveley (2010). "Expansion of the eukaryotic proteome by alternative splicing." Nature **463**(7280): 457-463.
Niu, Z., W. Jin, L. Zhang and X. Li (2012). "Tumor suppressor RBM5 directly interacts with the DExD/H-box protein DHX15 and stimulates its helicase activity." FEBS Lett **586**(7): 977-983.
Nowack, B., F. G. Kari and H. G. Krüger (2001). "The Remobilization of Metals from Iron Oxides and Sediments by Metal-EDTA Complexes." Water, Air, and Soil Pollution **125**(1): 243-257.
O'Bryan, M. K., B. J. Clark, E. A. McLaughlin, R. J. D'Sylva, L. O'Donnell, J. A. Wilce, J. Sutherland, A. E. O'Connor, B. Whittle, C. C. Goodnow, C. J. Ormandy and D. Jamsai (2013). "RBM5 is a male germ cell splicing factor and is required for spermatid differentiation and male fertility." PLoS Genet **9**(7): e1003628.
Oberstrass, F. C., S. D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D. L. Black and F. H. Allain (2005). "Structure of PTB bound to RNA: specific binding and implications for splicing regulation." Science **309**(5743): 2054-2057.
Oh, J. J., D. R. Grosshans, S. G. Wong and D. J. Slamon (1999). "Identification of differentially expressed genes associated with HER-2/neu overexpression in human breast cancer cells." Nucleic Acids Res **27**(20): 4008-4017.
Ozer, A., J. M. Pagano and J. T. Lis (2014). "New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization." Mol Ther Nucleic Acids **3**: e183.
Ottiger, M., F. Delaglio and A. Bax (1998). "Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra." J Magn Reson **131**(2): 373-378.

Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2005). "Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment." Acta Crystallogr D Biol Crystallogr **61**(Pt 4): 449-457.
Panjikar, S., V. Parthasarathy, V. S. Lamzin, M. S. Weiss and P. A. Tucker (2009). "On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination." Acta Crystallogr D Biol Crystallogr **65**(Pt 10): 1089-1097.
Patton, C., S. Thompson and D. Epel (2004). "Some precautions in using chelators to buffer metals in biological solutions." Cell Calcium **35**(5): 427-431.
Perez-Canadillas, J. M. (2006). "Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1." EMBO J **25**(13): 3167-3178.
Perutz, M. F., H. Miurhead, J. M. Cox, L. C. Goaman, F. S. Mathews, E. L. McGandy and L. E. Webb (1968). "Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 A resolution: (1) x-ray analysis." Nature **219**(5149): 29-32.
Perutz, M. F., H. Muirhead, J. M. Cox and L. C. Goaman (1968). "Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 A resolution: the atomic model." Nature **219**(5150): 131-139.
Perutz, M. F., M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will and A. C. North (1960). "Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis." Nature **185**(4711): 416-422.
Pervushin, K., R. Riek, G. Wider and K. Wuthrich (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proc Natl Acad Sci U S A **94**(23): 12366-12371.
Peter, M. E., R. C. Budd, J. Desbarats, S. M. Hedrick, A. O. Hueber, M. K. Newell, L. B. Owen, R. M. Pope, J. Tschopp, H. Wajant, D. Wallach, R. H. Wiltrout, M. Zornig and D. H. Lynch (2007). "The CD95 receptor: apoptosis revisited." Cell **129**(3): 447-450.
Petrella, E. C., L. M. Machesky, D. A. Kaiser and T. D. Pollard (1996). "Structural requirements and thermodynamics of the interaction of proline peptides with profilin." Biochemistry **35**(51): 16535-16543.
Philipps, D., A. M. Celotto, Q. Q. Wang, R. S. Tarng and B. R. Graveley (2003). "Arginine/serine repeats are sufficient to constitute a splicing activation domain." Nucleic Acids Res **31**(22): 6502-6508.
Purow, B. (2012). "Notch inhibition as a promising new approach to cancer therapy." Adv Exp Med Biol **727**: 305-319.
Ramaswamy, S., K. N. Ross, E. S. Lander and T. R. Golub (2003). "A molecular signature of metastasis in primary solid tumors." Nat Genet **33**(1): 49-54.
Ray, D., H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L. H. Matzat, R. K. Dale, S. A. Smith, C. A. Yarosh, S. M. Kelly, B. Nabet, D. Mecenas, W. Li, R. S. Laishram, M. Qiao, H. D. Lipshitz, F. Piano, A. H. Corbett, R. P. Carstens, B. J. Frey, R. A. Anderson, K. W. Lynch, L. O. Penalva, E. P. Lei, A. G. Fraser, B. J. Blencowe, Q. D. Morris and T. R. Hughes (2013). "A compendium of RNA-binding motifs for decoding gene regulation." Nature **499**(7457): 172-177.
Rintala-Maki, N. D., C. A. Goard, C. E. Langdon, V. E. Wall, K. E. Traulsen, C. D. Morin, M. Bonin and L. C. Sutherland (2007). "Expression of RBM5-related factors in primary breast tissue." J Cell Biochem **100**(6): 1440-1458.
Rossmann, M. G. (1972). The Molecular Replacement Method. New York, Gordon & Breach.

Rückert, M. and G. Otting (2000). "Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments." Journal of the American Chemical Society **122**(32): 7793-7797.
Ryder, S. P., L. A. Frater, D. L. Abramovitz, E. B. Goodwin and J. R. Williamson (2004). "RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1." Nat Struct Mol Biol **11**(1): 20-28.
Sachs, A. B., M. W. Bond and R. D. Kornberg (1986). "A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression." Cell **45**(6): 827-835.
Salzmann, M., K. Pervushin, G. Wider, H. Senn and K. Wuthrich (1998). "TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins." Proc Natl Acad Sci U S A **95**(23): 13585-13590.
Sanders, C. R., 2nd and J. P. Schwonek (1992). "Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR." Biochemistry **31**(37): 8898-8905.
Sanford, J. R., D. Longman and J. F. Caceres (2003). "Multiple roles of the SR protein family in splicing regulation." Prog Mol Subcell Biol **31**: 33-58.
Sattler, M., J. Schleucher and C. Griesinger (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Progress in Nuclear Magnetic Resonance Spectroscopy **34**(2): 93-158.
Schanda, P., E. Kupce and B. Brutscher (2005). "SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds." J Biomol NMR **33**(4): 199-211.
Schmidt, P. G. and I. D. Kuntz (1984). "Distance measurements in spin-labeled lysozyme." Biochemistry **23**(18): 4261-4266.
Schnablegger, H. and Y. Singh (2013). The SAXS guide: getting acquainted with the principles. Austria, Anton Paar GmbH.
Schulze-Osthoff, K., D. Ferrari, M. Los, S. Wesselborg and M. E. Peter (1998). "Apoptosis signaling by death receptors." Eur J Biochem **254**(3): 439-459.
Schwerk, C. and K. Schulze-Osthoff (2005). "Regulation of apoptosis by alternative pre-mRNA splicing." Mol Cell **19**(1): 1-13.
Schwieters, C. D., J. J. Kuszewski, N. Tjandra and G. M. Clore (2003). "The Xplor-NIH NMR molecular structure determination package." J Magn Reson **160**(1): 65-73.
Scotti, M. M. and M. S. Swanson (2016). "RNA mis-splicing in disease." Nat Rev Genet **17**(1): 19-32.
Shamoo, Y., N. Abdul-Manan and K. R. Williams (1995). "Multiple RNA binding domains (RBDs) just don't add up." Nucleic Acids Res **23**(5): 725-728.
Shan, X., K. H. Gardner, D. R. Muhandiram, N. S. Rao, C. H. Arrowsmith and L. E. Kay (1996). "Assignment of 15N, 13Cα, 13Cβ, and HN Resonances in an 15N,13C,2H Labeled 64 kDa Trp Repressor−Operator Complex Using Triple-Resonance NMR Spectroscopy and 2H-Decoupling." Journal of the American Chemical Society **118**(28): 6570-6579.
Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**(4): 213-223.
Shepard, P. J. and K. J. Hertel (2009). "The SR protein family." Genome Biol **10**(10): 242.
Skrisovska, L., C. F. Bourgeois, R. Stefl, S. N. Grellscheid, L. Kister, P. Wenter, D. J. Elliott, J. Stevenin and F. H. Allain (2007). "The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction." EMBO Rep **8**(4): 372-379.
Solomon, I. (1955). "Relaxation Processes in a System of Two Spins." Physical Review **99**(2): 559-565.

Song, Z., P. Wu, P. Ji, J. Zhang, Q. Gong, J. Wu and Y. Shi (2012). "Solution structure of the second RRM domain of RBM5 and its unusual binding characters for different RNA targets." Biochemistry **51**(33): 6667-6678.
Spera, S. and A. Bax (1991). "Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts." Journal of the American Chemical Society **113**(14): 5490-5492.
Stefl, R., M. Xu, L. Skrisovska, R. B. Emeson and F. H. Allain (2006). "Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs." Structure **14**(2): 345-355.
Steggerda, S. M. and B. M. Paschal (2002). "Regulation of nuclear import and export by the GTPase Ran." Int Rev Cytol **217**: 41-91.
Summers, M. F. (1988). "113Cd NMR spectroscopy of coordination compounds and proteins." Coordination Chemistry Reviews **86**: 43-134.
Sun, H. and L. A. Chasin (2000). "Multiple splicing defects in an intronic false exon." Mol Cell Biol **20**(17): 6414-6425.
Sutherland, L. C., K. Wang and A. G. Robinson (2010). "RBM5 as a putative tumor suppressor gene for lung cancer." J Thorac Oncol **5**(3): 294-298.
Svergun, D. I., C. Barberato and M. H. J. Koch (1995). "CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates." Journal of Applied Crystallography **28**: 768-773.
Zamore, P. D. and M. R. Green (1989). "Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor." Proc Natl Acad Sci U S A **86**(23): 9243-9247.
Zhang, L., Q. Zhang, Y. Yang and C. Wu (2014). "The RNA recognition motif domains of RBM5 are required for RNA binding and cancer cell proliferation inhibition." Biochem Biophys Res Commun **444**(3): 445-450.
Zhou, Z., L. J. Licklider, S. P. Gygi and R. Reed (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**(6903): 182-185.
Zweckstetter, M. (2008). "NMR: prediction of molecular alignment from structure using the PALES software." Nat Protoc **3**(4): 679-690.
Zweckstetter, M. and A. Bax (2001). "Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage." J Biomol NMR **20**(4): 365-377.
Tazi, J., N. Bakkour and S. Stamm (2009). "Alternative splicing and disease." Biochim Biophys Acta **1792**(1): 14-26.
Thandapani, P., T. R. O'Connor, T. L. Bailey and S. Richard (2013). "Defining the RGG/RG motif." Mol Cell **50**(5): 613-623.
Timmer, T., P. Terpstra, A. van den Berg, P. M. Veldhuis, A. Ter Elst, G. Voutsinas, M. M. Hulsbeek, T. G. Draaijers, M. W. Looman, K. Kok, S. L. Naylor and C. H. Buys (1999). "A comparison of genomic structures and expression patterns of two closely related flanking genes in a critical lung cancer region at 3p21.3." Eur J Hum Genet **7**(4): 478-486.
Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science **278**(5340): 1111-1114.
Tjandra, N., H. Kuboniwa, H. Ren and A. Bax (1995). "Rotational dynamics of calcium-free calmodulin studied by 15N-NMR relaxation measurements." Eur J Biochem **230**(3): 1014-1024.
Toriya, M., A. Tokunaga, K. Sawamoto, K. Nakao and H. Okano (2006). "Distinct functions of human numb isoforms revealed by misexpression in the neural stem cell lineage in the Drosophila larval brain." Dev Neurosci **28**(1-2): 142-155.
Tsuda, K., K. Kuwasako, M. Takahashi, T. Someya, M. Inoue, T. Terada, N. Kobayashi, M. Shirouzu, T. Kigawa, A. Tanaka, S. Sugano, P. Guntert, Y. Muto and S. Yokoyama (2009).

"Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3." Nucleic Acids Res **37**(15): 5151-5166.
Tsuda, K., T. Someya, K. Kuwasako, M. Takahashi, F. He, S. Unzai, M. Inoue, T. Harada, S. Watanabe, T. Terada, N. Kobayashi, M. Shirouzu, T. Kigawa, A. Tanaka, S. Sugano, P. Guntert, S. Yokoyama and Y. Muto (2011). "Structural basis for the dual RNA-recognition modes of human Tra2-beta RRM." Nucleic Acids Res **39**(4): 1538-1553.
Wahl, M. C., C. L. Will and R. Luhrmann (2009). "The spliceosome: design principles of a dynamic RNP machine." Cell **136**(4): 701-718.
Van Nostrand, E. L., G. A. Pratt, A. A. Shishkin, C. Gelboin-Burkhart, M. Y. Fang, B. Sundararaman, S. M. Blue, T. B. Nguyen, C. Surka, K. Elkins, R. Stanton, F. Rigo, M. Guttman and G. W. Yeo (2016). "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)." Nat Methods **13**(6): 508-514.
Wang, G. S. and T. A. Cooper (2007). "Splicing in disease: disruption of the splicing code and the decoding machinery." Nat Rev Genet **8**(10): 749-761.
Wang, I., J. Hennig, P. K. Jagtap, M. Sonntag, J. Valcarcel and M. Sattler (2014). "Structure, dynamics and RNA binding of the multi-domain splicing factor TIA-1." Nucleic Acids Res **42**(9): 5949-5966.
Wang, L., M. Miura, L. Bergeron, H. Zhu and J. Yuan (1994). "Ich-1, an Ice/ced-3-related gene, encodes both positive and negative regulators of programmed cell death." Cell **78**(5): 739-750.
Wang, Z. and C. B. Burge (2008). "Splicing regulation: from a parts list of regulatory elements to an integrated splicing code." RNA **14**(5): 802-813.
Varani, L., S. I. Gunderson, I. W. Mattaj, L. E. Kay, D. Neuhaus and G. Varani (2000). "The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein." Nat Struct Biol **7**(4): 329-335.
Weant, A. E., R. D. Michalek, I. U. Khan, B. C. Holbrook, M. C. Willingham and J. M. Grayson (2008). "Apoptosis regulators Bim and Fas function concurrently to control autoimmunity and CD8+ T cell contraction." Immunity **28**(2): 218-230.
Welling, D. B., J. M. Lasak, E. Akhmametyeva, B. Ghaheri and L. S. Chang (2002). "cDNA microarray analysis of vestibular schwannomas." Otol Neurotol **23**(5): 736-748.
Verdi, J. M., A. Bashirullah, D. E. Goldhawk, C. J. Kubu, M. Jamali, S. O. Meakin and H. D. Lipshitz (1999). "Distinct human NUMB isoforms regulate differentiation vs. proliferation in the neuronal lineage." Proc Natl Acad Sci U S A **96**(18): 10472-10476.
Westhoff, B., I. N. Colaluca, G. D'Ario, M. Donzelli, D. Tosoni, S. Volorio, G. Pelosi, L. Spaggiari, G. Mazzarol, G. Viale, S. Pece and P. P. Di Fiore (2009). "Alterations of the Notch pathway in lung cancer." Proc Natl Acad Sci U S A **106**(52): 22293-22298.
Will, C. L. and R. Luhrmann (2011). "Spliceosome structure and function." Cold Spring Harb Perspect Biol **3**(7).
Williamson, M. P., T. F. Havel and K. Wuthrich (1985). "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry." J Mol Biol **182**(2): 295-315.
Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.
Wishart, D. S., B. D. Sykes and F. M. Richards (1992). "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy." Biochemistry **31**(6): 1647-1651.
Voith von Voithenberg, L., C. Sanchez-Rico, H. S. Kang, T. Madl, K. Zanier, A. Barth, L. R. Warner, M. Sattler and D. C. Lamb (2016). "Recognition of the 3' splice site RNA by the U2AF

heterodimer involves a dynamic population shift." Proc Natl Acad Sci U S A **113**(46): E7169-E7175.
Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue (2005). "The CCPN data model for NMR spectroscopy: development of a software pipeline." Proteins **59**(4): 687-696.
Yamazaki, T., J. D. Forman-Kay and L. E. Kay (1993). "Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings." Journal of the American Chemical Society **115**(23): 11054-11055.
Yaseen, N. R. and G. Blobel (1999). "Two distinct classes of Ran-binding sites on the nucleoporin Nup-358." Proc Natl Acad Sci U S A **96**(10): 5516-5521.