Technische Universität München

Fakultät für Chemie
Lehrstuhl für biomolekulare NMR-Spektroskopie

Conformational Dynamics and Mechanisms of RNA Recognition by the Multidomain Splicing Factor U2AF

Carolina Sánchez Rico

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. Carlo Camilloni
Prüfer der Dissertation:	Prof. Dr. Michael Sattler Prof. Dr. Bernd Reif Prof. Dr. Friedrich Förster

Die Dissertation wurde am 15.02.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 21.03.2017 angenommen.

DECLARATION

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

Huang, Jie-rong, Lisa R. Warner, **Carolina Sánchez**, Frank Gabel, Tobias Madl, Cameron D. Mackereth, Michael Sattler, and Martin Blackledge. "Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study." *Journal of the American Chemical Society* 136, no. 19 (May 14, 2014): 7068–76.

Contribution to this work: In this collaborative project I did the validation of the ensemble prediction, by measuring and analyzing PREs at different salt concentrations. Also, I calculated the errors of the PREs for all the mutants used during the ensemble calculation and used later for error crossvalidation.

Voith von Voithenberg Lena, **Carolina Sánchez-Rico**, Hyun-Seo Kang, Tobias Madl, Katia Zanier, Anders Barth, Lisa R. Warner, Michael Sattler, and Don C. Lamb. "Recognition of the 3' Splice Site RNA by the U2AF Heterodimer Involves a Dynamic Population Shift." *Proceedings of the National Academy of Sciences* 113, no. 46 (November 15, 2016): E7169–75.

Contribution to this work: In this collaborative project I prepared all the proteins samples used for FRET studies, did the NMR and ITC experiments and analysis of the disease-related mutations. Discussion of all data during manuscript writing.

Table of Content

Abstract 9

Zusammenfassung 13

Chapter 1: Introduction I: Biological background 17

1.1 Pre-mRNA Splicing 18

1.2 Splicing regulation and Spliceosome Assembly 20

1.2.1 U2AF and its role in 3’ splice site recognition 22

1.3 U2Af65 RNA-binding domains and multidomain dynamics 24

Chapter 2: Introduction II: Solution techniques for integrated structural biology 29

2.1 Nuclear Magnetic Resonance 30

2.1.1 Basic principles of NMR 30

2.2 NMR studies of proteins and protein-ligand interactions 34

2.3 Relaxation in NMR 37

2.3.1 Longitudinal relaxation 38

2.3.2 Transversal relaxation 38

2.3.3 Molecular tumbling and relaxation 39

2.3.4 $\{^1H\}$-^{15}N Heteronuclear NOE 42

2.3.5 Paramagnetic relaxation enhancement (PRE) 43

2.4 Isothermal titration calorimetry (ITC) 45

2.5 Small angle X-ray scattering (SAXS) 47

Scope of the thesis 51

Chapter 3: Materials and Methods 53

3.1 Materials 54

3.1.1 Constructs 54

3.1.2 Solutions and Media 54

3.1.3 Buffers 55

3.2 Methods..55

3.2.1 Protein Expression and Purification 55

3.2.2 NMR titrations 57

3.2.3 NMR dynamic experiments 58

3.2.4 Paramagnetic Relaxation Enhancement, PRE experiments 58

3.2.5 Förster Resonance Energy Transfer, FRET...59

3.2.6 Isothermal Titration Calorimetry, ITC titrations 60

3.2.7 Small Angle Scattering, SAXS 61

3.2.8 Static Light Scattering, SLS...62

3.2.9 Surface Plasmon Resonance, SPR/BIAcore 62

Chapter 4: U2AF65 RRM1,2 free is an ensemble of structures in solution..............65

4.1 Evidence of additional non-compact states of RRM1,2 66

4.2 Generating the pool..67

4.3 Selecting the ensemble for the apoprotein 68

4.4 Error description and crossvalidation 71

4.5 Analysis of the ensemble...74

4.6 Confirmation of ensemble prediction using PREs 77

4.7 RRM1,2 ensemble in solution, not just closed conformation 78

CHAPTER 5: Linker between RRM1 and RRM2 has an autoinhibitory role 81

5.1 Linker between RRM1 and RRM2 interacts with RRM2....................................82

5.1.1 The C-terminal of the linker is involved in the interaction with RRM2...........84

5.2 Effects of the linker/domains interactions on RNA binding..............................89

5.3 Linker provides proofreading against weak Py-tracts 94

5.4 Cooperative effect of the linker and domains for RNA recognition 99

5.5 Chemical Exchange on the free and RNA-bound states..................................108

Chapter 6: Role of the U2AF35 subunit within the U2AF heterodimer.................115

6.1 FRET studies confirm dynamic nature of RRM1,2 116

6.2 Role of U2AF35 during Py-tract recognition by U2AF65 118

6.3 Somatic mutations do not affect the U2AF heterodimer 122

6.4 Effect of the dye-labelling position on RRM1,2 integrity 126

6.5 Fluorophore hydrophobicity affects protein integrity 129

6.6 Fluorophore effect on RNA binding 131

Chapter 7: Discussion 135

Appendix 143

Protein Sequences 144

GS-linker genes 145

U2AF65-RRM12 T_1, T_2, $T_{1\rho}$ relaxation parameters 147

Abbreviations 149

List of Figures 151

Acknowledgements 1555

References 1577

Curriculum Vitae 1633

Conformational Dynamics and Mechanisms of RNA Recognition by the Multidomain Splicing Factor U2AF

Abstract

Protein-RNA interactions are involved in numerous cellular functions, playing central roles in gene regulation and they are involved in every step of RNA metabolism. The RNA binding proteins (RBP) are proteins containing multiple domains connected by flexible linkers, where the individual domains are called RNA binding domains (RBDs) (e.g. the RNA recognition motif (RRM)). Usually the structure and RNA binding modes of a single RRM are well known. However, how multiple RBDs can interact with each other and how multidomain arrangements, dynamics and cooperativity can modulate RNA binding is poorly understood. Interestingly, the multi-RRMs proteins have the ability to regulate their binding specificity and affinity. This regulation is usually done cooperatively between the domains and in some cases the conformational flexibility of the linker of multi-domain RNA binding proteins can further modulate the interactions between the domains and with the RNA.

The U2 auxiliary factor (U2AF) is a multi-domain protein essential for spliceosome assembly. U2AF is a heterodimer composed of a large (65 kDa, U2AF65) and a small (35 kDa, U2AF35) subunit, which recognize the poly-pyrimidine tract (Py-tract) and the AG-dinucleotide, respectively, at the 3’ end of pre-mRNA introns. It was previously reported by our group that the tandem RRM1-RRM2 domains of U2AF65 adopt two distinct conformations, closed (inactive) and open (active), when bound to Py tract RNA.

The goals of this thesis are to characterize the conformational space and dynamics of the tandem domain RRM1-RRM2 of U2AF65 and the effects on RNA binding. In addition, the role of U2AF35 when U2AF65 binds to poly-pyrimidine tract RNAs of

different strength, i.e. binding affinities, is studied. For these studies a number of state-of-the-art solution methods are employed, combining NMR spectroscopy, Small Angle X-ray Scattering and fluorescence spectroscopy. The findings are confirmed by biochemical and biophysical experiments.

Using an ensemble approach, a study of the conformational space sampled by the RRM1-RRM2 tandem-domain is presented. By combining NMR and SAXS data it is found that in the absence of RNA the two domains sample a larger conformational space than previously thought. The ensemble of the tandem domain conformations includes the previously reported closed and open states, but also a substantial fraction of detached conformations. The proximity of the two domains in the closed conformations was found to be of electrostatic origin.

Additionally, the role of the linker connecting the two RRM domains is studied. Based on NMR chemical shift analysis and NOE data it is shown that the C-terminal region of the RRM1-RRM2 linker interacts with the RNA binding surface of RRM2. Mutational analysis combined with NMR and ITC experiments indicate that the linker exerts an autoinhibitory role where the weak intramolecular linker interaction competes with RNA binding. This suggests a novel proofreading mechanism that discriminates against weak (low affinity) Py-tract RNA ligands.

The role of U2AF35 for Py-tract recognition by U2AF65 was investigated by combining NMR and single-pair FRET experiments. FRET experiments show that the presence of the U2AF35 UHM domain (U2AF homology motif) induces a conformational shift of the population towards the open state of U2AF65 RRM1-RRM2, and thereby promotes the binding of weak Py-tracts. NMR paramagnetic relaxation enhancement (PRE) data show that the UHM domain contacts the U2AF65 RRM1 domain, thereby mediating this population shift. These observations reveal an unexpected role for the U2AF35 subunit, which involves a dynamic population shift towards the open conformation to facilitate the recognition of weak Py-tracts at the 3’ splice site. The potential role of disease-linked mutations in U2AF65 recognition of 3’ splice site are investigated, as these have been proposed to disrupt the binding of U2AF35. However, neither of these mutations affect the U2AF heterodimer formation or RNA binding, suggesting that the mutations may impair other interactions.

Finally, the structural effects of the attachment of hydrophobic fluorophores used for FRET experiments were studied using NMR. NMR spectra of U2AF65 conjugated to different commonly used FRET fluorophores reveal distinct effects that can interfere with the conformation and/or dynamics of the protein studied.

In this thesis, an integrated approached has been applied by a combination of solution methods commonly used in structural biology, e.g. NMR, SAXS and FRET. The results obtained provide novel and unexpected insight into the role of conformational dynamics in RNA recognition and splicing activity by U2AF. The results could only be obtained by a combination of complementary solution techniques that are crucial to reveal the interplay between structure, dynamics and function. The combination of these methods is thus strongly advised for studies of protein complexes, where dynamic molecular recognition is expected to play critical roles.

Zusammenfassung

Protein-RNA Interaktionen sind Teil vielfältiger zellulärer Funktionen, spielen eine entscheidende Rolle in der Genregulation und sind in sämtliche RNA Metabolismusschritte eingebunden. RNA-bindende Proteine enthalten oft mehrere RNA bindende Domänen (z.B. das RNA recognition motif (RRM)), die durch flexible Linker miteinander verbunden sind. Normalerweise sind die Struktur und die RNA Interaktion der Einzeldomänen bereits gut verstanden, es ist aber oft unklar wie die einzelnen Domänen miteinander interagieren und wie die Anordnung, Dynamik und Kooperativität der Einzeldomänen die RNA-Bindung beeinflussen. Interessanterweise haben diese Multidomänenproteine die Fähigkeit ihre Bindungsspezifität und -affinität zu modulieren. Diese Regulation geschieht normalerweise kooperativ zwischen den Einzeldomänen. Manchmal kann die konformationelle Flexibilität des Linkers die Interaktion zwischen den Domänen und mit der RNA beeinflussen.

Der U2 auxilary factor (U2AF) ist ein essentielles Multidomänenprotein für die Assemblierung des Spliceosomes. U2AF ist ein heterodimeres Protein, das aus einer großen 65 kDa (U2AF65) und einer kleinen 35 kDa (U2AF35) Untereinheit besteht. U2AF65 erkennt den Polypyrimidinstrang (Py-Strang), während U2AF35 das AG-Dinukleotid am 3‘-Ende von pre-mRNA-Introns erkennt. Es wurde vorher durch unsere Gruppe publiziert, dass die U2AF65 tandem RRM1-RRM2 Domänen zwei unterschiedliche Konformationen annehmen: Die geschlossene, inaktive und die offene, aktive Konformation, die den Py-Strang RNA bindet.

Die Ziele dieser Arbeit sind die Charakterisierung des konformationellen Raumes und der Dynamik der U2AF65 tandem Domänen RRM1-RRM2 und der RNA-Bindung. Darüber hinaus wird der Effekt von U2AF35 auf die Bindung von schwachen und starken Py-Strängen, das heißt mit niedriger und hoher Affinität, untersucht. Für diese Studien wird eine Vielzahl von strukturbiologischen Techniken in Flüssigkeit eingesetzt. NMR-Spektroskopie wird mit Röntgenkleinwinkelstreuung (SAXS) und

Fluoreszenzspektroskopie kombiniert. Die Ergebnisse werden durch biochemische und biophysikalische Experimente erhärtet.

Durch einen Ensembleansatz wird der konformationelle Raum, der durch die U2AF65 RRM1-RRM2 Tandemdomänen eingenommen wird, dargestellt. Durch die Kombination von NMR und SAXS Daten wurde gezeigt, dass die beiden Domänen einen viel größeren konformationellen Raum abdecken als vorher erwartet. Das Ensemble der Tandemdomänen enthält den vorher beschriebenen offenen und geschlossenen Zustand, aber auch eine substantielle Fraktion von distanzierten, nicht kompakten, Konformationen. Die Nähe der beiden Domänen im geschlossen Zustand konnte auf elektrostatische Wechselwirkungen zurückgeführt werden.

Als nächstes wurde die Rolle des Linkers, der die beiden Domänen verbindet, untersucht. Basierend auf der Analyse von chemischen Verschiebungen in NMR Spektren und NOE Daten wurde gezeigt, dass die C-terminale Region des Linkers zwischen RRM1 und RRM2 mit der RNA bindenden Oberfläche der RRM2 Domäne interagiert. Eine Mutationsanalyse kombiniert mit NMR und ITC Experimenten deuten auf eine autoinhibitorische Rolle des Linkers hin, bei der die schwache Interaktion des intramolekulare Linkers mit der RNA Bindung konkurrieren kann. Diese Ergebnisse legen den Schluss auf einen neuen Korrekturlesemechanismus nahe, der zwischen schwachen und starken Py-Strängen unterscheiden kann.

Die Rolle von U2AF35 bei der Erkennung von Py-Strängen durch U2AF65 wurde mit Hilfe von NMR-Spektroskopie und Einzelmolekül-Försterresonanzenergietransfer (FRET)-Spektroskopie untersucht. FRET-Spektroskopie zeigt, dass die Anwesenheit der U2AF35 UHM (U2AF homology motif) Domäne eine konformationelle Verschiebung der Population der RRM1 RRM2 Tandemdomäne von U2AF65 zum offenen Zustand vollzieht und somit die Bindung von schwachen Py-Strängen erleichtert. NMR paramagnetic relaxation enhancement (PRE) Daten zeigen, dass die UHM Domäne die

RRM1 Domäne von U2AF65 berührt und dadurch die Konformationsverschiebung auslöst. Diese Beobachtungen zeigen eine unerwartete Funktion von U2AF35. U2AF35 induziert eine dynamische Populationsverschiebung zur offenen Konformation und erleichtert so die Bindung schwacher Py-Stränge an der 3‘ splice site. Die potentielle Rolle von Mutationen in U2AF65, die eine Rolle bei der Erkennung der 3‘ splice site spielen und mit Krankheiten assoziiert sind, wurden untersucht. Diese Mutationen stehen in Verdacht die Bindung von U2AF35 durch U2AF65 zu unterbrechen. In dieser Arbeit konnte jedoch kein Effekt dieser Mutationen auf die Bildung des U2AF Heterodimers oder die RNA-Bindung beobachtet werden. Dies deutet darauf hin, dass die untersuchten Mutationen andere Interaktionen betreffen.

Abschließend wurden die strukturellen Effekte der benutzen hydrophoben Fluorophore, welche in den FRET-Studien verwendet wurden, auf U2AF65 untersucht. NMR Spektren von U2AF65 makiert mit verschiedenen häufig genutzten Fluorophoren, zeigen unterschiedliche Effekte, die die Konformation und Dynamik des untersuchten Proteins beeinflussen.

In dieser Arbeit wurde ein integrierter strukturbiologischer Ansatz durch die Kombination von strukturbiologischen Methoden in Flüssigkeit benutzt: Z.B. NMR, SAXS und FRET. Die beschriebenen Ergebnisse ermöglichen neue und überraschende Einblicke in die Rolle der konformationellen Dynamik der RNA-Bindung und der Regulation der Splicing Aktivität durch U2AF. Diese Ergebnisse konnten nur durch die Anwendung komplementärer Methoden in Flüssigkeit erreicht werden, die essentiell für die Untersuchung des Zusammenspiels zwischen Struktur, Dynamik und Funktion sind. Daher ist diese Kombination dieser Methoden für die Untersuchung von Proteinkomplexen, bei denen die dynamische molekulare Bindung von Liganden eine entscheidende Rolle spielt, sehr zu empfehlen.

CHAPTER 1:

Introduction I: Biological background

1.1 Pre-mRNA Splicing

How is it explained that hundreds of thousands of different proteins and peptides in humans can be expressed from around 20000 coding genes? (Kim et al., 2014). The more complex the organism, the more complicated regulatory mechanisms are needed to produce specific RNAs and proteins. These regulatory mechanisms are part of the gene regulation carried out by the cell, and specify the fraction of the genetic information that each cell expresses in a particular spatiotemporal context and is vital for its function and adaptation to the changing conditions of its surroundings (Beato, 1993).

After transcription, the resulting pre-mRNA has to undergo global processing events: 5' capping, splicing and 3' polyadenylation. Splicing is one of the main gene regulation mechanisms and it is the process by which internal non-coding sequences of the pre-mRNA are removed through cleavage-ligation reactions (Green, 1986). These non-coding sequences are called introns, while the coding sequences of the pre-mRNA to be ligated are called exons. The removal of introns and the ligation of exons during splicing is a crucial mechanism in eukaryotes to obtain multiple alternative transcripts from a single gene. It is known that alternative splicing plays a crucial role in expanding the functional complexity, protein diversity and in developing of higher eukaryotes. In alternative splicing particular exons can be included or excluded from the final mature mRNA, thus increasing the number of different proteins expressed from a particular gene (Nilsen & Graveley, 2010).

The splicing reaction consist itself of two consecutive trans-esterification reactions (Figure 1). The first step is a hydrophilic attack, where the 2'-OH group of the adenosine in the branch point attacks the phosphodiester bond of the conserved guanine at the 5' splicing site (exon1-intron junction). During the second step another hydrophilic attack takes place. The 3'-OH end of the exon1 attacks the phosphodiester bond of the conserved guanine at the 3' splicing site (intron-exon2 junction). In this final step the 3'-OH of the intron is liberated resulting in a free lariat and the two exons are ligated.

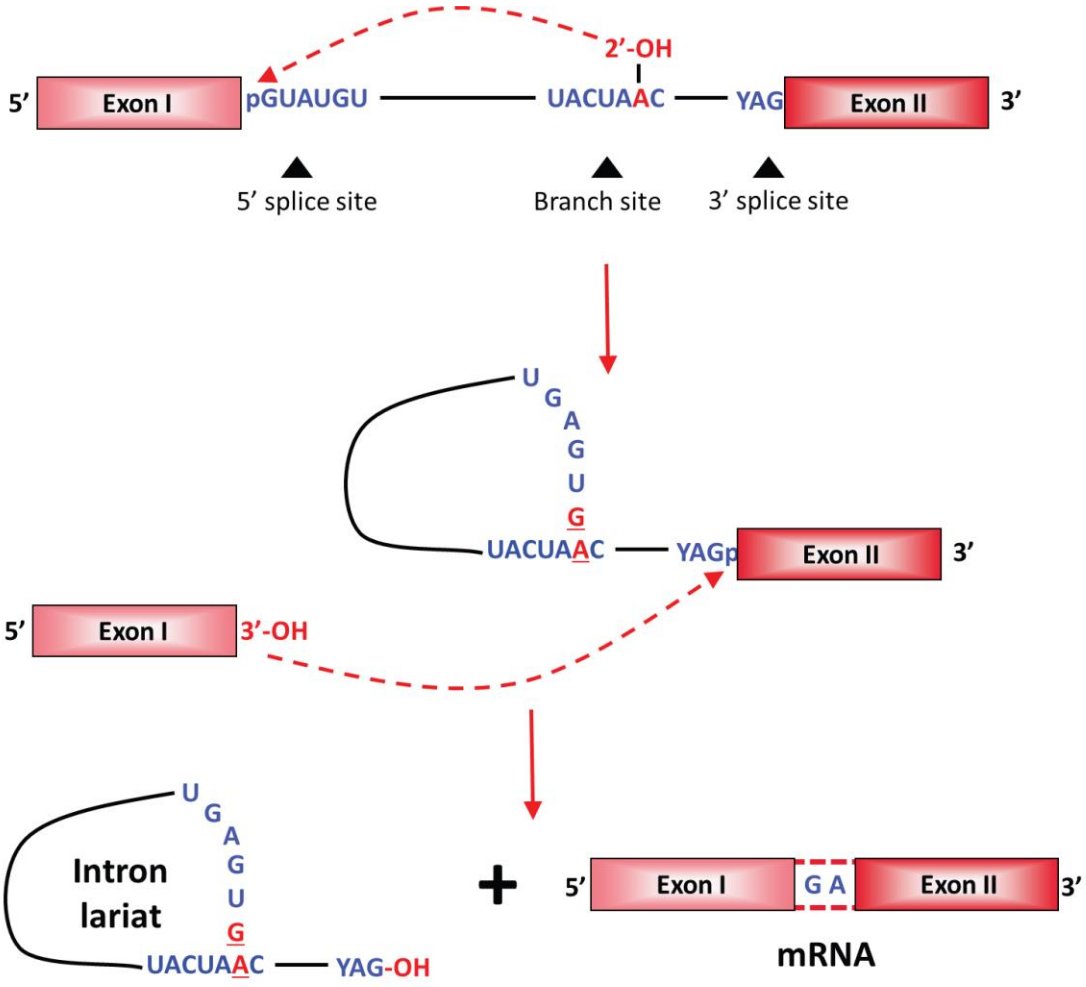
2'-OH
5'
Exon I
pGUAUGU
UACUAAC
YAG
Exon II
3'
5' splice site
Branch site
3' splice site
U G A G U G
UACUAAC
YAGp
Exon II
3'
5'
Exon I
3'-OH
Intron lariat
U G A G U G
UACUAAC
YAG-OH
5'
Exon I
G A
Exon II
3'
mRNA

Figure 1: Schematic overview of pre-mRNA splicing. *The splicing reaction consist of two trans-esterification reactions. In the first reaction, a hydrophilic attack is carried out from the 2'-OH of the branch point adenosine to the conserved guanine at the 5'-SS. The second reaction is another hydrophilic attack, where the 3'-OH end of the released exon I attacks the phosphodiester bond of the conserved guanine at the 3'-SS. Then the mature m-RNA is produced and the intron is released as a lariat.*

There exists group I and group II introns that can execute the splicing reactions by themselves, however the pre-mRNA introns require a complex machinery (spliceosome) of trans-acting protein and RNA factors that carry out the splicing process.

1.2 Splicing regulation and Spliceosome Assembly

The splicing regulation is carried out by ribonucleoproteins (RNPs) which are complexes of RNAs and proteins that work cooperatively in different molecular machines. In the spliceosome machinery there is an intricate engagement of the pre-mRNA during the building of the spliceosome, it doesn’t merely act as substrate, but it is involved in the active sites formation. The active site is built by RNA and protein components which are both involved in splicing catalysis. The spliceosome is a highly dynamic RNP machinery, which is particularly receptive to regulation. The main two aspects susceptible to regulation during splicing are the recognition of functional sites of pre-mRNA by several spliceosome factors to ensure fidelity, and second, numerous weak binary interactions act together to reinforce or reject splice-site choices (Wahl, Will, & Lührmann, 2009).

The main building blocks of the spliceosome are five uridine-rich small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4/U6 and U5. Each snRNP consist of an snRNA and associated proteins that together are responsible for removing most of the pre-mRNA introns (Black, 2003; Will & Lührmann, 2011). During the splicing process there are several assembly and disassembly steps throughout stepwise interactions of the snRNP with the pre-mRNA (Figure 2). During the spliceosome assembly, the U1 binds to the 5’-SS (splice site) of the intron via base paring of the U1-snRNA and the splice site. Furthermore, the earliest assembly stage also involves the binding of a special set of proteins to the 3’-SS: the SF1/BBP (Branch-point Binding Protein) and the U2AF (U2 auxiliary factor) to the branch point site (BPS) and to the polypyrimidine tract (Py-tract) respectively. The early spliceosome assembly complex, called complex E, comprise the binding of the U1 and U2AF to both ends of the intron (Figure 3) and play essential roles in the initial recognition of the 5' and 3' splicing sites (Black, 2003; Wahl et al., 2009). The next step involves the binding of the U2 snRNP to the complex E in an ATP-driven manner, where the snRNA of U2 base-pairs at the BPS and displaces the SF1/BBP.

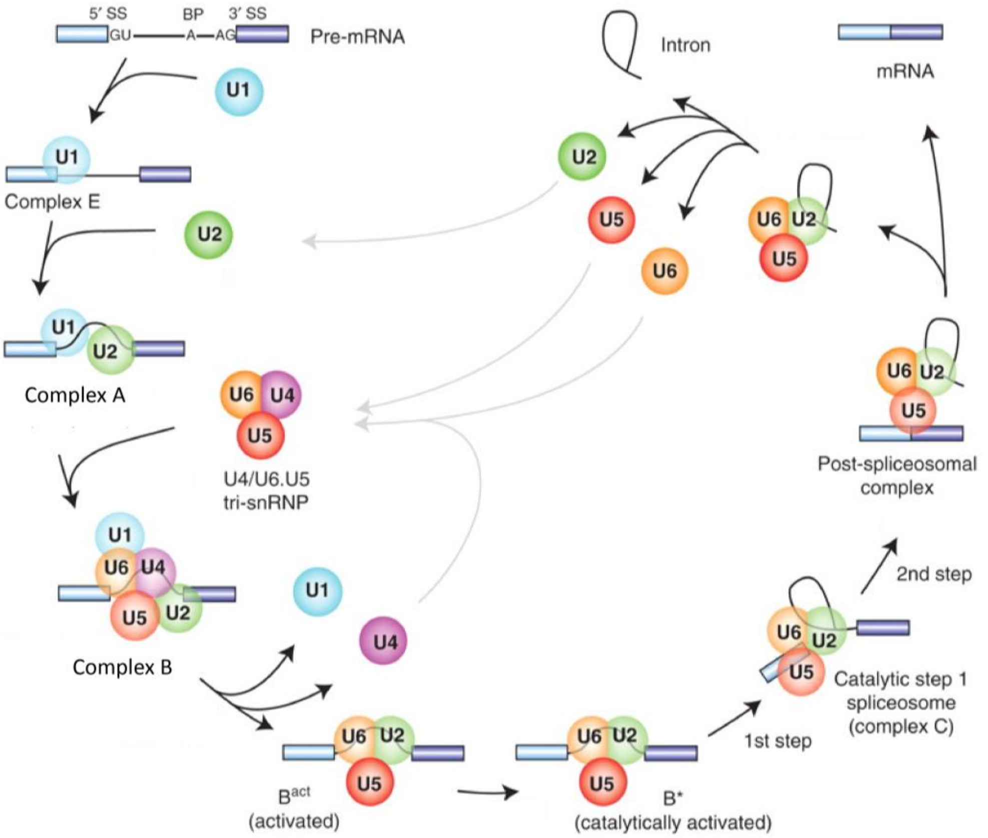
5′ SS
BP
3′ SS
GU
A
AG
Pre-mRNA
U1
Complex E
U2
Complex A
U6
U4
U5
U4/U6.U5
tri-snRNP
Complex B
U1
U4
B^act
(activated)
B*
(catalytically activated)
1st step
Catalytic step 1
spliceosome
(complex C)
2nd step
Post-spliceosomal
complex
Intron
mRNA
U2
U5
U6

Figure 2: Spliceosome Assembly and pre-mRNA splicing. *Schematic representation of the spliceosome assembly stages and pre-mRNA splicing. (Adapted from (Will & Lührmann, 2011))*

This base-pairing interaction is stabilized by the heterodimer protein SF3a and SF3b and also by the arginine-serine rich domain of U2AF65 (large subunit of U2AF) (Valcárcel, Gaur, Singh, & Green, 1996). The new complex formed is called complex A or pre-spliceosomal complex. These RNP rearrangements at this early stages of the spliceosome assembly are reasonably well understood, however less is known for the next steps of the assembly and catalytic activation.

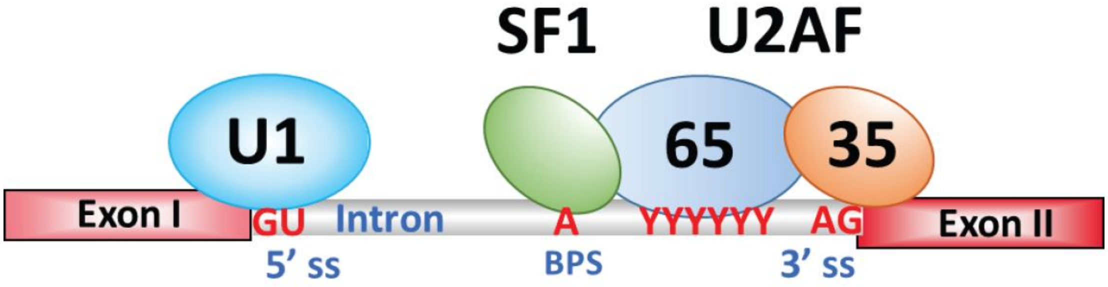

Figure 3: Complex E during spliceosome assembly. *U1 binds to the 5'-SS, the SF1 binds to the branch point at the 3'-SS, U2AF65 binds to the polypyrimidine tract (Py-tract) and U2AF35 recognizes the AG-dinucleotide at the 3'-SS.*

Following in the course of the assembly, the U4/U5/U6 tri-snRNP binds the complex A in order to form the complex B. The complex B has to undergo complicated rearrangements in order to form the complex C, here the U1 snRNP at the 5'-SS is replaced with the U6 snRNP and then the U1 and U4 snRNPs are gone from the complex. This yields to the catalytically activated spliceosome and the first chemical reaction takes place. Additional changes arise in the RNP complex and the second reaction occurs. It is the complex C that catalyzes the two chemical reactions of splicing (Black, 2003; Wahl et al., 2009; Will & Lührmann, 2011).

After that the two exons are ligated and the mature mRNA is released upon dissociation of the spliceosome complex. The intron lariat is degraded and the snRNPs are recycled for new splicing events (Cheng & Menees, 2011; Wahl et al., 2009).

1.2.1 U2AF and its role in 3' splice site recognition

U2AF (U2 snRNP auxiliary factor) is a heterodimeric protein constituted by a large subunit of 65 KDa (U2AF65) and the small subunit of 35 KDa (U2Af35). U2AF is an essential factor in constitutive splicing and its binding is often controlled during alternative splicing (Smith & Valcárcel, 2000). During the primary complex E formation, at the 3'-SS U2AF65 binds the polypyrimidine tract (Py-tract) located downstream of the branch-point site. While U2AF35, which is in contact with U2AF65 forming the U2AF heterodimer, binds the AG-dinucleotide at the 3'-SS (Will & Lührmann, 2011).

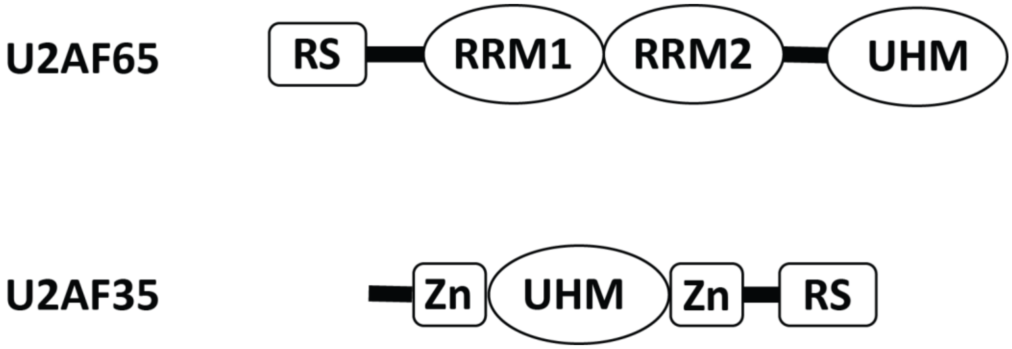

Figure 4: Domain organization U2AF. *The multi-domain organization of U2AF65 and U2AF35 are presented. RS: Arginine-Serine rich motifs, RRMs: RNA binding domains, UHM: U2AF homology motif, Zn: Zinc finger domains.*

U2AF65 is composed of 475 amino acids. It is a multi-domain protein which contains one RS domain (Arg-Ser rich domain), two canonical RNA-recognition motives (RRMs) RRM1 and RRM2, and a third non-canonical RRM called UHM (U2AF homology model) (Figure 4). U2AF35 is composed of 240 amino acids and contains a central UHM domain that is flanked by two Zn-finger domains and a RS domain (Figure 4).

The RS domain of U2AF65 is reported to contribute to the binding of U2 snRNP at the branch point during the complex A formation (Valcárcel et al., 1996). The RRM domains of U2AF65 are in charge of recognizing the Py-tract consensus sequences at the 3'-SS (Ito, Muto, Green, & Yokoyama, 1999), while the UHM domain participates in protein-protein interactions to stabilize the SF1-U2AF65 complex during BPS recognition (Selenko et al., 2003) (Figure 5).

Recently it was shown by Yoshida et. al. (H. Yoshida et al., 2015) that the two zinc finger domains of U2AF35 are directly involved in the binding to the AG-dinucleotide, while the UHM domain of U2AF35 is involved in protein-protein interaction with U2AF65 (Kielkopf, Rodionova, Green, & Burley, 2001) (Figure 5).

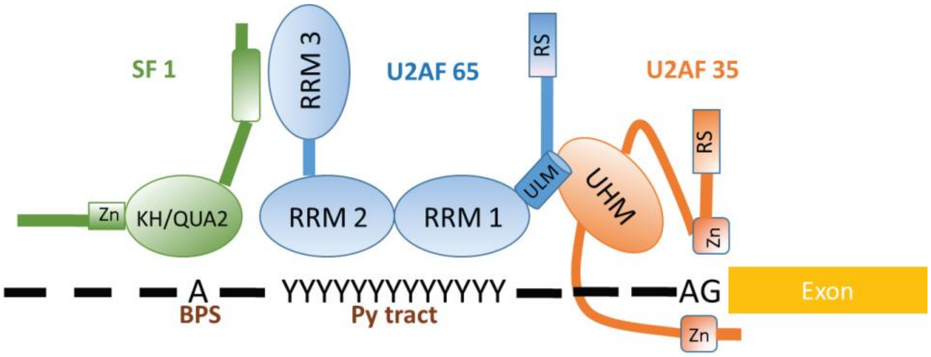

Figure 5: Interactions at the 3' splice site. *Recognition of the 3’-SS involves cooperative binding of SF1, U2AF65 and U2AF35.*

U2AF dissociates at later stages of the splicing, while after the first reaction the U2AF35 interaction at the 3’-SS is replaced by a different group of factors (cited in Wahl et al., 2009).

1.3 U2Af65 RNA-binding domains and multidomain dynamics

One of the biggest challenges of the spliceosome to overcome during splicing is the proper recognition and paring of the splice sites (5’-SS, branch point site and 3’-SS) that are immersed within a numerous of similar sequences. In yeast the 5’-SS and the BPS consensus at the 3’-SS are highly conserved, while they are more degenerated in mammals. On the other hand, the Py-tract sequence (located downstream of the branch point) in mammals is highly conserved. This Py-tract is less conserved in yeast introns, although a U-rich tract enhances yeast splicing (Berglund, Abovich, & Rosbash, 1998; Patterson & Guthrie, 1991).

The two central RNA-recognition motifs (RRMs) of U2AF65 (RRM1 and RRM2) are responsible for the correct recognition of the Py-tract sequence at the 3’-SS. The most conserved signature sequence of RRM is the RNP1 which is an 8-residue motif with the consensus [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] and a second homologous region is the RNP2 which is a 6-residue motif with the consensus: [ILV]-[FY]-[ILV]-X-N-L located approx. 30 residues from the N-terminal of RNP1 (X can be any aminoacid) (Kielkopf, Lücke, & Green, 2004). The tertiary structure of an RRM is comprised of two α-helices that are packed against four antiparallel β-strands, the canonical topology is βαββαβ. The two RNPs establish the two central β-strands and are directly involved in RNA binding (Figure 6) (Mackereth et al., 2011; Sickmier et al., 2006), the RNP1 is located in β-strand 3 and the RNP2 located in β-strand 1. The β-sheets in RRMs are usually found to interact with single-stranded RNA. Generally, an average of two RRMs are observed to act cooperatively within a single polypeptide (Kielkopf et al., 2004).

The proteins containing RRMs are involved in each step of RNA metabolism and they are required to recognize specific RNA sequences. Usually the RNA-RRM interactions are transient interactions and are found in processes such splicing, processing, transport and localization (Cléry & Allain, 2013). RRM domains are found to be involved not only in RNA or DNA recognition but also taking part in protein/protein interactions. RRMs have the ability to regulate their binding specificity and affinity by means of each of their structural elements (α-helices, loops, β-strands). The highly diverse biological functions of proteins containing RRM domains can be explained by the immense versatility of the RRM interactions and their structural characteristics (Cléry, Blatter, & Allain, 2008).

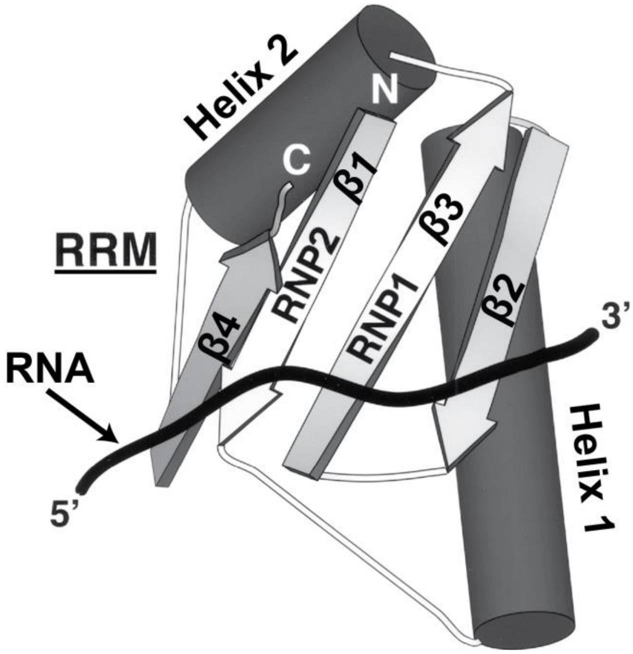

Figure 6: Canonical RRM fold. *The canonical topology of and RRM is βαββαβ, where two α-helices are packed against four antiparallel β-strands. The RNPs involved in RNA-binding laid in the two central β-strands. (Adapted from (Kielkopf et al., 2004), PDB 1URN)*

The great number of multi-domain proteins interacting with RNA or other proteins combining several weak interactions, emphasize the important role of the dynamics throughout the substrate recognition process. These dynamic mechanisms often include cooperation of the different domains and involve different binding modes that can be enabled by flexible linkers. Common mechanisms during RNA recognition by multidomain proteins include fly-casting (induced fit) and conformational selection (Figure 7) (Mackereth & Sattler, 2012).

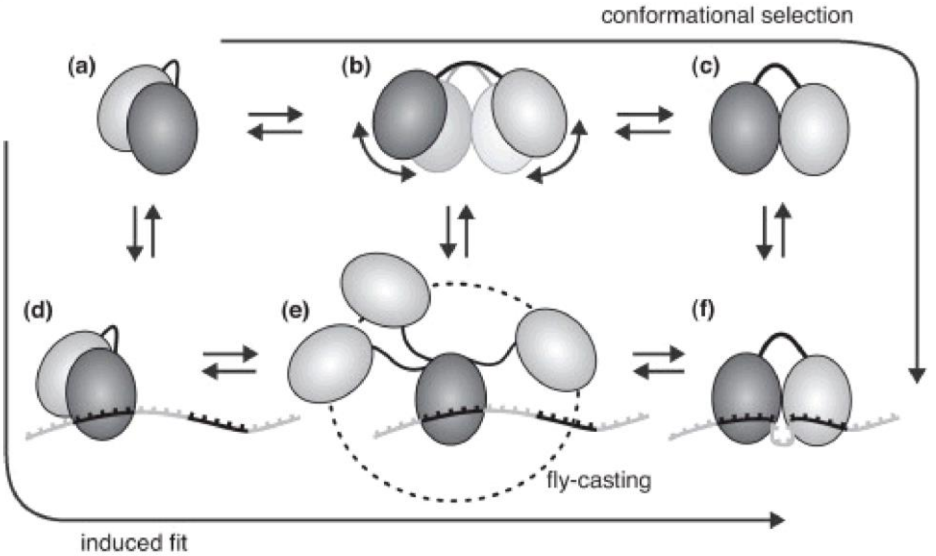
conformational selection
(a)
(b)
(c)
(d)
(e)
(f)
fly-casting
induced fit

Figure 7: RNA binding mechanisms by multidomain proteins. *The conformational selection mechanism requires a pre-existing population of the protein in the active (open) state which is then recognized by the RNA ligand. In the fly-casting or induced fit mechanism an initial RNA binding to one of the domains enable the search by the second domain of specific RNA motifs in a longer RNA ligand, thus inducing the final open conformation. (Adapted from (Mackereth & Sattler, 2012))*

CHAPTER 2:

Introduction II: Solution techniques for integrated structural biology

2.1 Nuclear Magnetic Resonance

The function of all the biomolecules and biological complex is determined by their 3D structure and dynamics which establish the way they interact with each other. The main techniques used to reveal the 3D structure of proteins and nucleic acids are: X-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy (especially Cryo-electron microscopy) and as transversal methods small-angle X-ray and neutron scattering (SAXS and SANS). All of these experimental methods have advantages and limitations. NMR spectroscopy is the only technique that provides atomic resolution structures and in addition allows the study of conformational dynamics of the system over a large range of time-scales (Göbl, Madl, Simon, & Sattler, 2014; Markwick, Malliavin, & Nilges, 2008) (Figure 13). Therefore, NMR is a powerful tool used to investigate structure, dynamics and kinetics of a wide range of biological systems

2.1.1 Basic principles of NMR

The basic phenomenon of NMR is similar to other spectroscopic techniques, where an external radiation causes a transition between the ground state and excited states. For example, in the case of visible light spectroscopy an electron absorbs the energy of light of a certain wave length. In NMR in contrast, the absorbed radiation causes the excitation of the nuclear spin from the ground state to its excited state (Rule & Hitchens, 2005). First of all, for the generation of the ground and excited states in NMR spectroscopy it is required that the nuclei are immersed in a static magnetic field (Zeeman effect). In this case, the energy difference (ΔE) between the two states are of the order of radio frequency (Figure 8).

The energy levels of the two states is given in terms of the dipole moment μ_z along *z*, and the magnitude of the static magnetic field B_0 along *z:*

$$E = -\mu_z \cdot B_0$$

μ_z depends on the nuclear spin quantum number m_z along *z,* and the gyromagnetic ration γ which is constant that depends on the specific nuclei and indicates the sensitivity of the nucleus ($\gamma_{1H}>>\gamma_{15N}>\gamma_{13C}$):

$$\mu_z = m_z \cdot \hbar \cdot \gamma$$

Where ħ is the plank constant divided by 2π.

Then, for spin I=1/2 (as is the case for the nuclear spins of ^{1}H, ^{15}N and ^{13}C), the two energy levels have the following energies:

$$E_\alpha = -\frac{1}{2} \cdot \hbar \cdot \gamma \cdot B_0 \text{ and}$$

$$E_\beta = \frac{1}{2} \cdot \hbar \cdot \gamma \cdot B_0$$

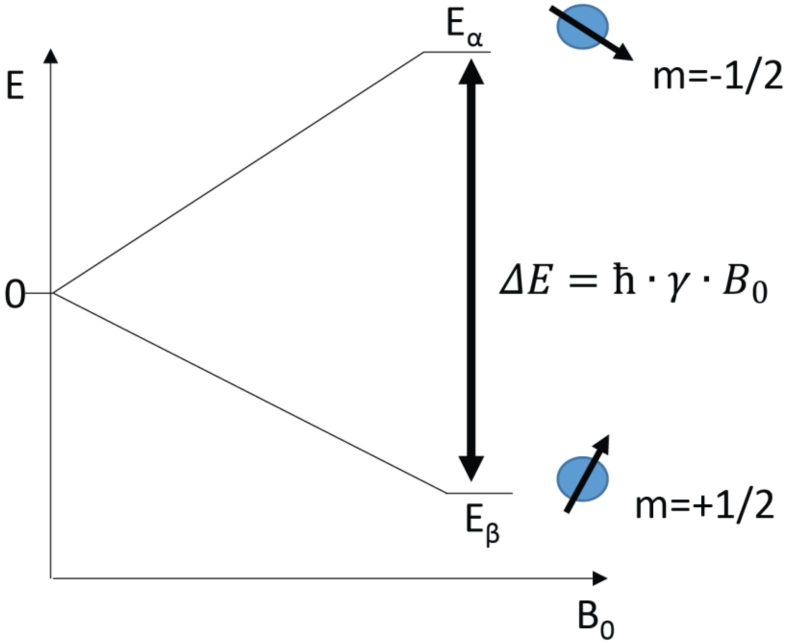

Figure 8: Energy levels of nuclear spin I=1/2. *When atoms of nuclear spin ½ are immerse in a magnetic field, the energy levels are split (Zeeman effect) with an energy difference ΔE. Each level has an associated magnetic quantum number, m.*

The sensitivity in NMR experiments depends among other factors on the gyromagnetic ratio of the specific nucleus, this is the reason why ^{1}H protons are the preferable nuclei to detect during NMR experiments.

For non-continuous-wave NMR spectroscopy, i.e. for Fourier transform (FT) NMR; a series of pulses covering a range of frequencies (“all sequences at once”) are applied to the sample altering its equilibrium and then the system relaxes back to equilibrium. The decay of the signal over time is observed as a free induction decay (FID). The obtained FID is converted by fourier transformation (FT) to a spectrum from the time into the frequency domain.

The resonance condition is fulfilled when the energy of the applied radio waves matches the energy difference between the two spin states ΔE. The resonance frequency is the Larmor frequency (ν):

$$\Delta E_r = h \cdot \nu$$

Then the resonance condition is given by:

$$\hbar \cdot \gamma \cdot B_0 = h \cdot \nu_L$$

$$\therefore\ \nu_L = \frac{\gamma}{2\pi} \cdot B_0\ ,$$

$$\text{with}\ \ \omega_0 = -\gamma \cdot B_0$$

The magnetization along *z* M_0 in the equilibrium can be observed since the spins immerse on the magnetic field along *z* are not equally populated in the two energy levels. A pulse with exactly the Larmor frequency can interact with the spins and transfer the magnetization to the transversal *xy*-plane, where it can be detected and then the signal start to relax back to equilibrium. The relaxation process occurs naturally depending on the size of the molecule and its surroundings (viscosity, temperature etc). The electronic cloud surrounding the nucleus shields it from the external B_0, i.e a nucleus with higher electron density is more shielded from B_0. This shielding occurs because the external magnetic field induces small electronic currents that generate a weak magnetic field $(B_0 \cdot \sigma)$ acting against the external magnetic field, where the shielding constant σ depends on the nucleus and its chemical surrounding and is independent from the strength of B_0.

So we have that the effective magnetic field acting on a particular nucleus is:

$$B_{eff} = B_0 - B_0 \cdot \sigma$$

For example, a methyl carbon has more electron density than a carbonyl carbon and so the methyl carbon 'feels' the static field B_0 less, since it is more 'shielded' and therefore experience a smaller effective field (B_{eff}). Thus, different electron density leads to small changes in energy levels (Figure 9) which give us the chemical shift (δ).

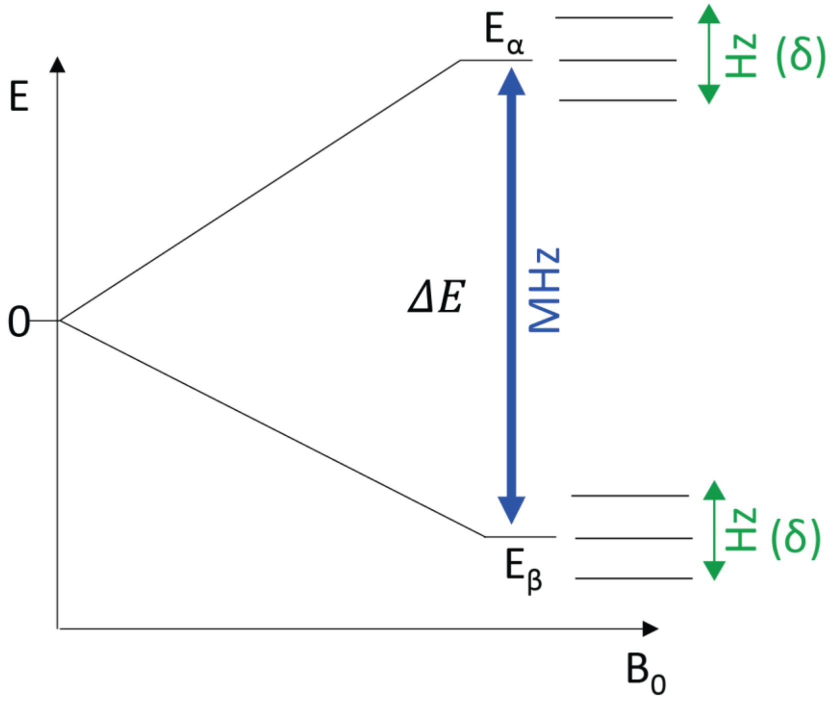

Figure 9: Chemical shift origin. *Different electron densities cause small splitting of the energy levels (hyperfine structure). This leads to a different chemical shift depending on the electron density of each nuclei.*

Therefore, different chemical environments provide different shielding constants, which lead to different Larmor frequencies. Instead of giving the chemical shifts (δ) in frequencies (which is B_0 field dependent), the chemical shifts are normalized to a standard (not field dependent), usually to the frequency of tetramethylsilane for ^{1}H and ^{13}C, and nitric acid for ^{15}N. The chemical shifts are then given in parts per million (ppm) as follow:

$$\delta = \frac{\nu - \nu_{\text{ref}}}{\nu_{ref}} \cdot 10^6 \text{ [ppm]}$$

When the nuclei are connected by covalent bonds, they influence each other by J-couplings. J-couplings are mediated by the electrons of the chemical bond. For more than three bonds, this interaction is very weak. The J-couplings lead to the splitting of the NMR signal (even smaller division within the energy levels in Figure 9). When a spin interacts with other spin, the state of the second spin can either be α or β and then the signal splits up into two peaks of equal intensity. When there are two other spins with equal energy there are four possibilities of states: αα, αβ, βα and ββ. The states αβ and βα are energetically equivalent. Since all four states have the same probability for the spin to be in those states and two states are equivalent, there will be three peaks with an intensity ratio of 1:2:1. For the general case of n (# of neighbors) spins, there will be a multiplicity of n+1 peaks and the intensity will be spread according to the binomial series. The so-called coupling constant is given by the distance between the split peaks. Magnetization can be transfer from one nucleus to another nucleus through covalent bonds over the J-couplings. Higher coupling constants allow more efficient magnetization transfers (Keeler, 2010).

2.2 NMR studies of proteins and protein-ligand interactions

To perform biomolecular NMR spectroscopy it is necessary to have labelling schemes on the sample with NMR-active nuclei. This is done by incorporating ^{15}N, ^{13}C and/or ^{2}H in the sample during the expression of the protein in media containing the corresponding isotope(s). This labelling schemes allow the measurement of multidimensional heteronuclear NMR experiments.

The most basic NMR experiment to be recorded for protein studies is the one-dimensional (1D) ^{1}H- spectrum. This spectrum provides a fast evaluation of the state of the protein, i.e. if it is folded or stable in the buffer. Although the 1D spectrum is unique for each protein, it is too complex to analyze since most of the signals overlap. In the 1D spectrum of folded proteins the proton signals are spread from around -0.5 to 12 ppm.

The fingerprint spectrum of the protein is given by the two dimensional (2D) experiment ^1H,^{15}N-HSQC (heteronuclear single quantum coherence). This spectrum is unique for each protein. During the experiment the only protons visible are the ones bound to the observed heteronuclear atom, in this case ^{15}N. This spectrum has now two dimensions, one presents the frequencies of the ^1H attached to the ^{15}N and the other one shows the frequencies of the ^{15}N attached to the proton respectively as cross-peaks. Since there is one ^{15}N-^1H bond per amino acid in the backbone, each cross-peak in a HSQC experiment represents a specific residue. Proline residues are not visible, since the lack the amide proton. This spectrum contains the signals of the protein backbone amides and usually additional peaks for the side chains of Asn, Gln, Trp, Lys and Arg residues. The sequential assignment of the observed signals is necessary to identify each residue, for this assignment a set of three dimensional (3D) experiments are required (Sattler, Schleucher, & Griesinger, 1999).

The chemical shifts of the nitrogen and proton are very sensitive to changes in their chemical environment. For this reason ^1H,^{15}N-HSQC is largely used in protein-ligand interaction studies, specifically by doing titration series of the protein in presence of increasing ligand concentrations. In this case the protein is ^{15}N labeled and the ligand is unlabeled, the ligand can be another protein, DNA, RNA, compounds or anything else that might interact with the labeled protein. The changes observed in the spectrum (or spectra) depend on the exchange rate (k_{ex}) between free and bound states, k_{ex} is defined as:

$$k_{ex} = k_{on}[L] + k_{off}$$

Where k_{on} and k_{off} are the association and dissociation rates respectively and $[L]$ is the ligand concentration.

If the exchange rate is larger than the difference in resonance frequency ($k_{ex} > \Delta v$), the signal appears at an average, population-weighted chemical shift between free and bound state ($\delta_{observed} = p_{free}\delta_{free} + p_{bound}\delta_{bound}$), then upon addition of more ligand the signal gradually shifts until the final saturated position. This is due to a rapid interconversion between free and bound states which is averaged during the detection time.

When *k_{ex}<Δv*, signals from both states are observed (due to not significant interconversion between free and bound states during the detection time) revealing the different chemical shifts of the free and bound state, their intensities and linewidths (Figure 10). In this limit, the intensity of each peak is directly proportional to the population of each state.

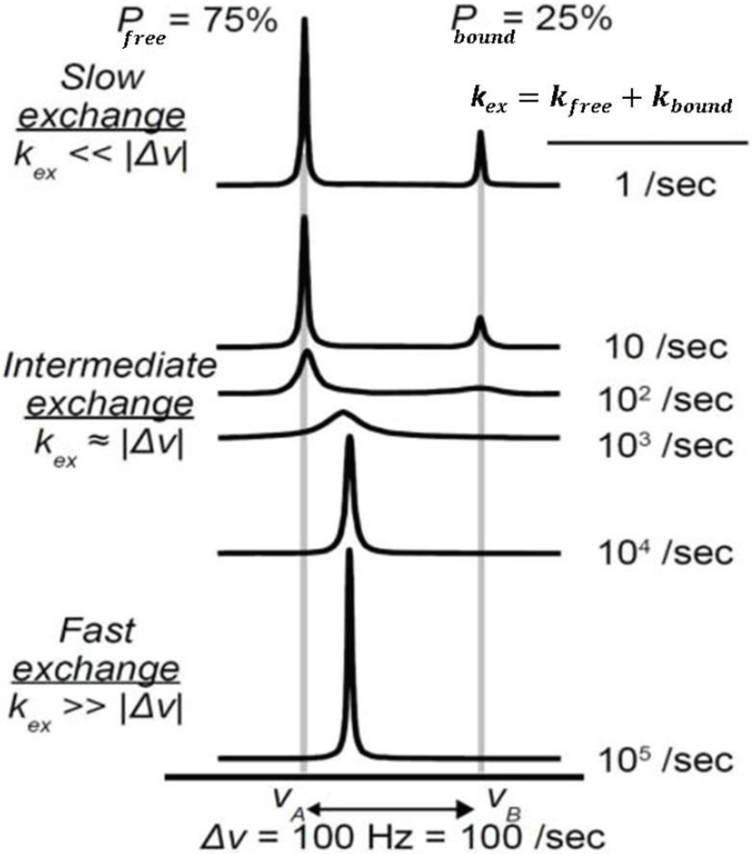

Figure 10: Chemical exchange rates. *When k_{ex}<Δv there is slow exchange and the two signals of the free and bound states can be detected. When k_{ex} ≈ Δv, there is an intermediate regime where one signal is observed at intermediate chemical shift (usually very broad). If k_{ex}>Δv there is fast exchange and the signal appears at an average, population-weighted chemical shift. (Adapted from (Kleckner & Foster, 2011))*

If the $k_{ex} \approx \Delta\nu$, one signal is observed at an intermediate chemical shift (between the chemical shift of the free state (δ_{free}) and the chemical shift of the bound state (δ_{bound})), the linewidth of this peak is very broad (Figure 10) because of interference from $free \leftrightarrow bound$ interconversion during the detection time (Kleckner & Foster, 2011).

Summarizing we have:

$k_{ex} > \Delta\nu$ Fast exchange

$k_{ex} \approx \Delta\nu$ Intermediate exchange

$k_{ex} < \Delta\nu$ Slow exchange

Where $\Delta\nu = \nu_{free} - \nu_{bound}$

Therefore, NMR titrations provide a relatively quick and residue-specific way to determine the protein-ligand interface.

2.3 Relaxation in NMR

After an external RF pulse has been applied to a sample, a natural phenomenon called relaxation takes place. It refers to how the bulk magnetization reaches equilibrium realigning the magnetic moments of individual spins along *z.* In NMR, relaxation is particularly slow in comparison to other molecular excited energy levels, for example in the case of vibrational and rotational energies or in the case of excited electronic states.

The main advantage of this long relaxation lifetime is that it provides us enough time to manipulate the transverse magnetization and to detect it. On the other hand, the main disadvantage of having slow relaxation is that one has to wait enough time for the equilibrium magnetization to be reestablish before repeating the experiment, which is necessary to do multiple times to increase the signal to noise level.

With NMR relaxation experiments is possible to extract information about the environment surrounding the nuclei and the nature of the molecule’s motion, since these two factors directly affect the relaxation rates.

The source of relaxation in a molecule can be seen as a local magnetic field produced by different factors occurring naturally in the surrounding of its spins (for spin-1/2: dipolar coupling > chemical shift anisotropy > spin rotation, and due to paramagnetic species when present). The local B_1 magnetic field is much weaker than the applied external B_0 magnetic field and instead of affecting all the spins in the same way, the effect of the B_1 field is highly localized and it changes as the molecule tumbles due to thermal agitation. There are two forms of relaxation, longitudinal and transverse relaxation. The application of a 90° pulse generates transverse magnetization ($M_{x(y)}$) and alters the population ratios, then the magnetization vector M reaches back its equilibrium M_0 with relaxation time constants T_1 and T_2 (with $R_1=1/T_1$ and $R_2=1/T_2$ as relaxation rates) present in the Bloch equations:

$$\frac{dM_z}{dt} = \frac{M_0 - M_z}{T_1}$$

$$\frac{dM_{x(y)}}{dt} = -\frac{M_{x(y)}}{T_2}$$

2.3.1 Longitudinal relaxation

Longitudinal relaxation, also called spin-lattice relaxation, arises with the movement of spin populations back to their Boltzmann equilibrium distribution, giving away extra energy to the lattice in order to go back to equilibrium. This relaxation is along the *z*-direction (Equation (1)) with a relaxation time T_1 (time for $M_z \rightarrow M_0$). The T_1 values are relatively long because of the deficient transfer of energy from NMR transitions (at Larmor frequency) into thermal energy.

2.3.2 Transversal relaxation

Transversal relaxation, also called spin-spin relaxation, is the process by which the magnetization in the transverse plane decays to zero at equilibrium via the decay of coherences (loss of phase coherence). The time needed for the magnetization in the transverse plane (Equation (2)) to decay to zero is the T2 (time for $M_{xy} \rightarrow 0$).

2.3.3 Molecular tumbling and relaxation

The rotational diffusion (tumbling) of a molecule in solution can be described by a rotational correlation time τ_c. τ_c is the average time required for a molecule to tumble through an angle of ~1 radian. As the molecule rotates the spins remain aligned (bulk magnetization) with the external magnetic field B_0. However, magnetic spin-spin interactions, for example dipole-dipole coupling between neighboring spins, depend on the distance and orientation (relative to the axis of the B_0 field) of the internuclear vector. Therefore, these magnetic interactions that produces the local magnetic fields are modulated by the molecular tumbling. As a result, the spins experience *fluctuating* local magnetic fields that allow the spins return to equilibrium.

To describe the relation between the molecular tumbling and the T1, T2 relaxation, it is necessary to introduce the spectral density function J(ω). J(ω) is the probability of finding a component of the random motion of the molecule at a particular frequency ω. Since the integral of J(ω) over all frequencies is constant, a slowly tumbling molecule (large molecular weight) has more contributions at low frequencies, and a faster tumbling molecule (low molecular weight) has more contributions at higher frequencies (Figure 11). The simplest form of the spectral density is given by:

$$J(\omega) = \frac{2\tau_c}{1+(\omega\tau_c)^2}$$

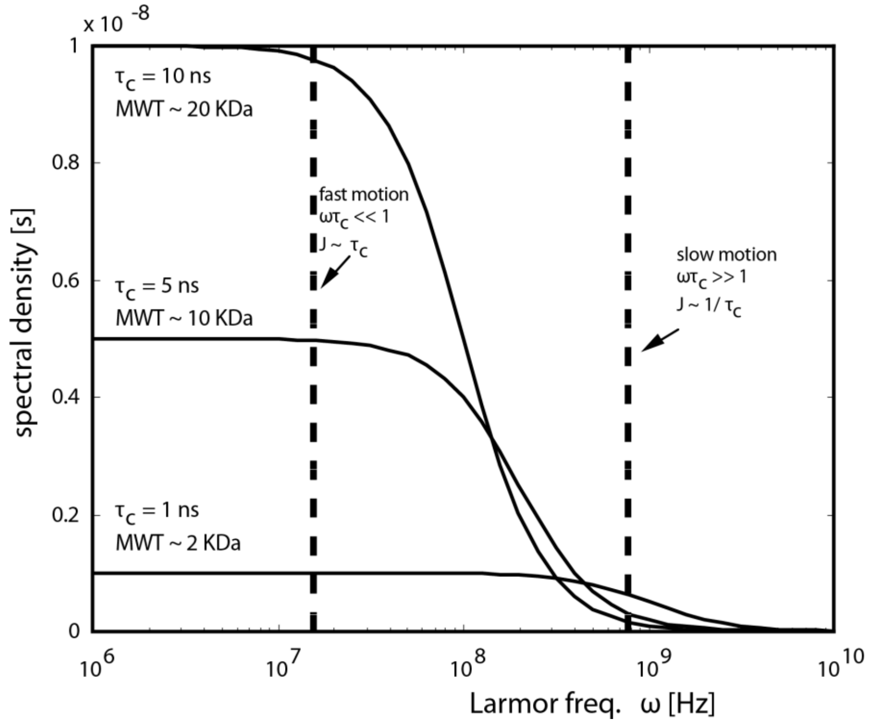

***Figure 11: Spectral Density.** Plot of the spectral density function for molecules of different molecular weight. (Adapted from (Grzesiek, 2003))*

The spin-lattice relaxation rate $1/T_1$ depends on the probability that the local fields have a component oscillating at the Larmor frequency, then we have:

$$1/T_1 = d^2\{ J(\omega_H-\omega_N)+3J(\omega_N)+6J(\omega_H+\omega_N)\}+c^2 J(\omega_N)$$

Here *d* and *c* represent the strengths of the dipolar interaction between the proton and nitrogen, and the chemical shift anisotropy (Kay, Torchia, & Bax, 1989). The T_1 time is relatively long due to deficient mechanisms to transfer energy (from spin population exchange) to the lattice in terms of thermal energy.

The spin-spin relaxation life times of stationary states are decreased, leading to random variations of the precession frequencies and loss of phase coherence. Processes that reduce the life time of spin states cause line-broadening, thus T_2 is directly related to the linewidth: $T_2 = 1/(\pi\Delta\nu)$. T_2 relaxation also depends on the spectral density function (Hore, 2015), however in terms of the spectral density, the component at frequency zero (zero motion) dominates the contribution. This is because motion at any frequencies diminishes the contribution coming from the distribution of the components of the local fields along *z* (Keeler, 2010), then $1/T_2 \sim J(0)$. Finally we have that the relaxation rate $1/T_2$ is given by:

$$1/T_2 = 0.5d^2\{4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\} + 1/6c^2\{3J(\omega_N) + 4J(0)\}$$

The dependence of T_1 and T_2 on τ_c is plotted in Figure 12. Where for slow tumbling molecules the T_2 time is shorter (Figure 12, right) given the large contribution of the spectral density at $\omega=0$ (Figure 12, left when $\tau_c = 4/\omega_0$), while for faster tumbling molecules the contribution at zero frequency is much less (Figure 12, left when $\tau_c = 0.25/\omega_0$), so T_2 is longer (Figure 12, right). In the case of T_1, we have that its main contribution comes when $\omega=\omega_0$. Then in the plot of $J(\omega)$ we focus on the curves when $\omega/\omega_0=1$ (Figure 12, left). For large and small molecules (slow and fast tumbling respectively), the spectral density contribution is low, which means T1 is larger. On the other hand for medium-size molecules the contribution of the spectral density is larger and then T_1 is shorter, thus the minimum T_1 time is for medium size molecules.

Under the following considerations (at least partially fulfilled in biomolecular NMR) of isotropic tumbling of the molecule, no chemical exchange, large molecules (τ_c>0.5 ns) and high magnetic fields (500 MHz or larger), one can approximate a relationship between τ_c and T_1, T_2 relaxation:

$$T1/T2 \approx 1+1/2\ \omega_0^2\ \tau_c^2 \rightarrow \tau_c \approx 1/\ \omega_0 \sqrt{(2T_1/T_2)}$$

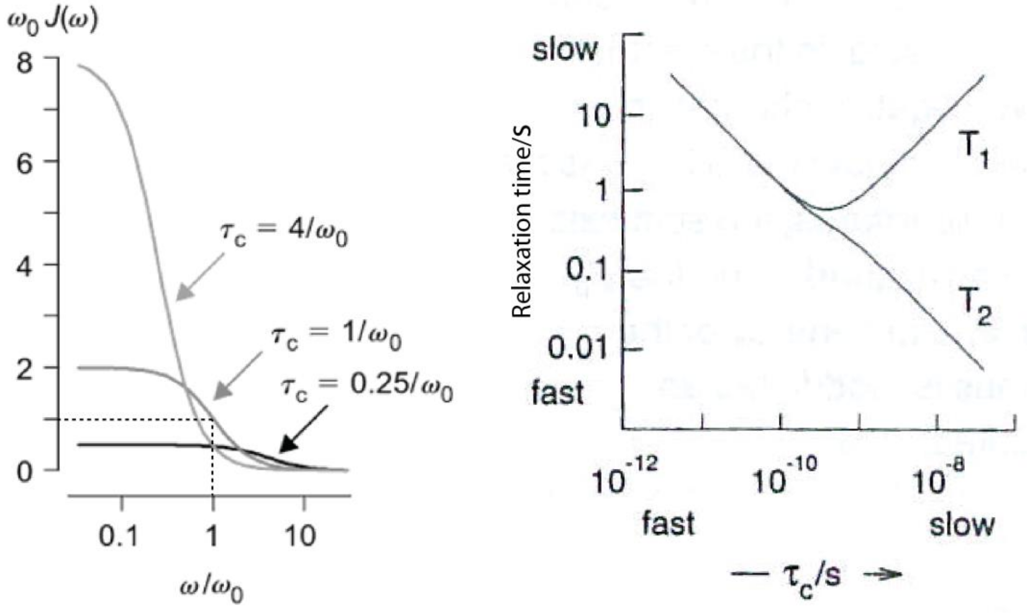

Figure 12: Relation between molecular tumbling and T1, T2 relaxation. *For large molecular weight proteins (which are slowly tumbling) the T2 relaxation times depend mainly on the spectral densities at low frequencies J(0), which causes fast T2. For T1 relaxation, when ω=ω₀, the spectral density contribution is larger for middle size proteins, thus the T1 time is shorter, while is longer for small molecules and big proteins. (Adapted from (Hore, 2015))*

2.3.4 {^{1}H}-^{15}N Heteronuclear NOE

The {^{1}H}-^{15}N heteronuclear NOE (hetNOE) gives information about motion of individual N-H bonds. The basic idea is that by irradiating the protons, the *z*-magnetization of the ^{15}N nuclei are enhanced due to cross-relaxation. Hence, when the magnetization is rotated to the transverse plane, stronger signals of ^{15}N are observed because of the NOE enhancement (Keeler, 2010). The values of steady-state NOEs compare the *z*-magnetization of the S-spin (^{15}N-spin) in the thermal equilibrium to the *z*-magnetization of the S-spin at equilibrium as the I-spin (^{1}H-spin) is saturated:

$$\mathrm{NOE}(I_z \rightarrow S_z) = \frac{\ll S_z \gg_{I_{sat}}}{\ll S_z \gg_{eq}}$$

The H-N bonds that undergo motion faster than the overall molecular tumbling display a decreased NOE intensity. The ^{1}H-^{15}N heteronuclear NOE has an average value of 0.77 and in general, values lower than 0.68 indicate flexible regions, e.g N and C terminal of the protein, linkers and loops. During the hetNOE experiments especial attention is required, due to the long T_1 of water molecules when the recycle time is not long enough, an NOE can be produce as consequence of chemical exchange between the amide protons and the water molecules, artificially decreasing the measured NOEs (Kay et al., 1989).

An overview of the type of protein motions at different NMR time scales is given in Figure 13, where T_1, T_2 and NOE relaxation NMR experiments are used to study dynamics between pico-seconds to nano-seconds.

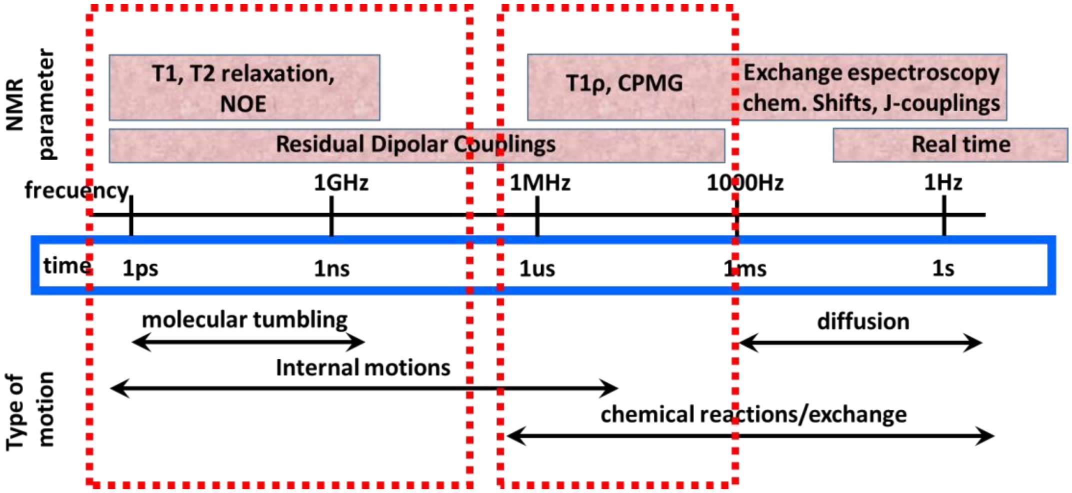
NMR parameter
T1, T2 relaxation, NOE
T1ρ, CPMG
Exchange espectroscopy chem. Shifts, J-couplings
Residual Dipolar Couplings
Real time
frecuency
1GHz
1MHz
1000Hz
1Hz
time
1ps
1ns
1us
1ms
1s
molecular tumbling
diffusion
Internal motions
chemical reactions/exchange
Type of motion

Figure 13: Time scales investigated with NMR. *Several molecular processes can be study by NMR at different time scales. In particular T1, T2 and NOE relaxation are used to study dynamics on the picoseconds to nanoseconds.*

2.3.5 Paramagnetic relaxation enhancement (PRE)

Paramagnetic NMR is a powerful tool used to investigate multiple dynamics processes occurring in macromolecules. Particularly PRE, rises from magnetic dipolar interactions between the nuclear spin and the unpaired electrons of a paramagnetic spin label, and as a result there is an increase in the nuclear relaxation rates. The PRE effect is very large in comparison to NOE, the last one has a short-range effect between proton spins (<6 Å), while the PRE effect has a large-range due to the large magnetic moment of the unpaired electron, allowing the detection of distances up to 35 Å (Clore & Iwahara, 2009). Hence, PRE has become the preferred method to obtain long-rage distance information that can be used for structure calculation. To use of PREs with non-metal binding proteins, the attachment of a paramagnetic center (e.g MTSL, IPSL) to a specific, solvent exposed site is necessary.

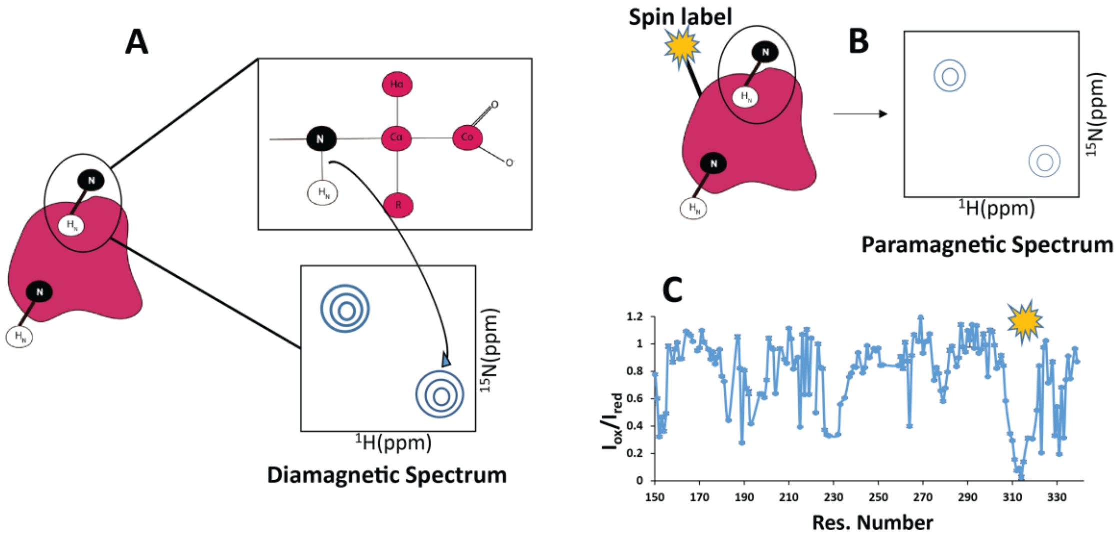
A
Hα
N
Cα
Co
O
O
H_N
R
N
H_N
N
H_N
15N(ppm)
1H(ppm)
Diamagnetic Spectrum
Spin label
N
H_N
N
H_N
B
15N(ppm)
1H(ppm)
Paramagnetic Spectrum
C
1.2
1
0.8
0.6
0.4
0.2
0
I_ox/I_red
150
170
190
210
230
250
270
290
310
330
Res. Number

Figure 14: Schematics of PREs. *In (A) is represented the spectrum without spin-label (diamagnetic spectrum). When a spin-label is attached to a specific residue, the neighbor residues 'feel' the effect of the unpaired electron causing a faster relaxation and therefore their intensities are reduced in the paramagnetic spectrum (B). The PRE signature can be obtained by the intensity ratio between* I_{para} *and* I_{dia} *(C).*

In biomolecular NMR, distances between the paramagnetic center and the nucleus can be determined from the increased in transversal relaxation rates due to the unpaired electron (additional R_2^{sp}=paramagnetic rate enhances). First, we have that the transverse relaxation rates (intrinsic and spin contribution) can be related to the intensity ratio of the oxidized and reduced resonances (Figure 14), I_{ox} and I_{red} respectively (Battiste & Wagner, 2000):

$$\frac{I_{ox}}{I_{red}} = \frac{R_2 \exp(-R_2{}^{sp} t)}{R_2 + R_2{}^{sp}}$$

Where R_2 is the intrinsic relaxation rate, R_2^{sp} is the spin contribution relaxation rate and t is the total evolution time. R_2 can be estimated from the reduced spectrum and together with the intensity ratios, it is possible to fit R_2^{sp}. Finally R_2^{sp}, can be converted into distances (for the effect of paramagnetic spins on nuclear magnetic relaxation)using the following equation described in (Battiste & Wagner, 2000):

$$r^6 = \left[\frac{k}{R_2^{sp}}\left(4\tau_c^* + \frac{3\tau_c^*}{1 + \omega_h^2 \tau_c^{*2}}\right)\right]$$

Where τ_c^* is the correlation time for this electron-nuclear interaction, *r* is the distance between the electron and nuclear spins, ω_h is the Larmor frequency of the proton nuclear spin and *K* is a constant 1.23×10^{-32} cm^6 s^{-2} depending on the gyromagnetic ratio (of protons) and electronic *g* factor.

For calculating distances τ_c^* is assumed to be equal to the global correlation time of the protein. Owing the distance dependence of the PRE as r^{-6}, this distance *r* between the electron and nuclear spin can be used as distance restrain during structure calculation of proteins, having that the nuclei closer in space to the paramagnetic center will relax faster (loss more intensity in the oxidized state) than those farther apart. Additional to structure calculation PREs can be used for detecting transient contacts between domains in biomolecules, to detect encounter-like contacts in protein/protein and protein/nucleic acid, given their high sensitivity detecting lower populated states (Clore & Iwahara, 2009).

2.4 Isothermal titration calorimetry (ITC)

ITC is a physical technique that is broadly used to quantify thermodynamic parameters of interactions in solution. Isothermal titration calorimeters measure the heat change that occurs when two molecules interact. In biological systems, interactions of macromolecules are at the core of their functionality and these interactions are described by their kinetic and thermodynamic properties. The thermodynamics of association can be characterized by the stoichiometry of the interaction (N), the free energy of binding(ΔG), entropy (ΔS), enthalpy (ΔH), heat capacity binding (ΔC) and by the association constant (K_a). ITC measures the binding equilibrium directly from determining the heat exchange involved during the association of a ligand with its binding partner. Directly from a single experiment the following values are determined: the K_a, N and ΔH. While ΔG and ΔS are calculated from the association constant. And the dependence on the temperature of ΔH can be determined by doing a series of titrations at varying temperatures, determines the ΔC. The relation between the Gibbs energy (free energy) and entropy can be determined with the equation:

$$\Delta G= -RTln(K_a)= \Delta H - T\Delta S$$

Where T is the absolute temperature in kelvin and R is the gas constant.

During the titrations of the ligand to its binding partner, heat is released or absorbed depending on the redistribution of non-covalent bonds as the interacting molecules go from free to bound. Using ITC those heat changes are determined by the measure of the differential power applied to the cell heaters necessary to maintain a zero temperature difference with the reference cell while the binding partners are combined. A schematic representation of the basic components of the ITC system is shown in Figure 15, where the sample is in the sample cell and ligand in the syringe, the reference cell normally contains water. The experimental binding isotherm is characterized by the c value (unitless constant), given by the product of K_a , N (number of ligand binding sites on the binding partner) and the concentration of the macromolecule in the sample cell M:

$$c= Ka[M]N$$

c should be between 1 and 1000, preferable between 10 and 500 in order to obtain optimal data.

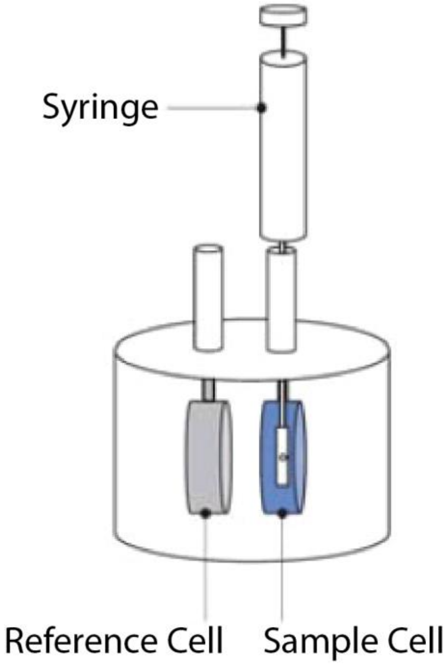

Figure 15: ITC schematics. *The ligand is placed in the syringe to be titrated to its binding partner located in the Sample cell. The reference cell has watter or buffer and allows the measurement of the temperature difference with the sample cell at each titration point. (From ("MicroCal Auto-ITC200 user manual (English)," n.d.)*

2.5 Small angle X-ray scattering (SAXS)

SAXS offers physical information and structural analysis for particles of 1-100 nm, while for bigger molecules it provides information about the average size of the particles and their shapes in solution. The samples in a SAXS experiment are exposed to X-rays of a certain wavelength that scatter elastically in a small angle between 0 and 5 degrees to produce spatially averaged intensity distribution (Boldon, Laliberte, & Liu, 2015). The intensity can be expressed as function of the scattering vector q resulting from a photon of λ wavelength that scatters off the sample at an angle 2θ:

$$q = \frac{4\pi \sin(\theta)}{\lambda}$$

The atoms inside the sample will scatter the incident X-rays in all directions, this gives a background radiation that is nearly constant at small angles. The particles (cluster of atoms) in the sample will generate additional scattering (excess scattering) due to the fact that the particles are in the size-range of the x-ray wavelength and are made of different materials with different densities (contrast). Then, measuring the angle-dependent distribution of the scattered radiation gives the possibility of study the average particle structure and shape (Schnablegger & Singh, 2011). In the case of a macromolecule in solution, its intensity distribution is obtained by subtracting the profile of the buffer from the total profile.

The SAXS profile has three regions that contain different information. (Figure 16).

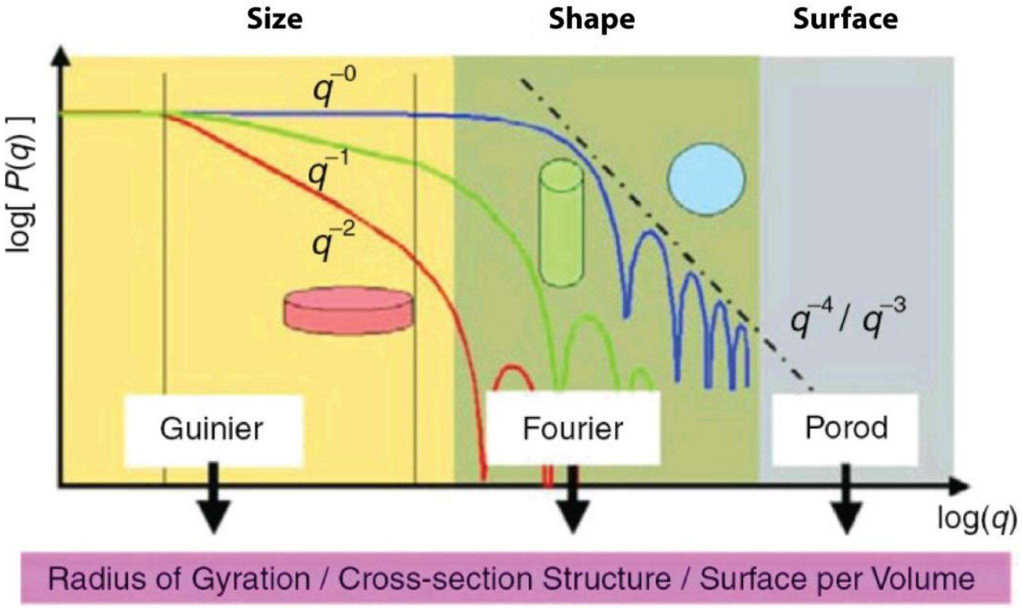

Figure 16: SAXS profile regions. *In the double logarithmic plot an initial slope of 0, -1 or -2 indicates globular, cylindrical or lamellar shape, respectively. (Adapted from (Boldon et al., 2015))*

First, the Guinier region, where the radius of gyration can be obtained. The radius of gyration is largely affected by aggregation, improper subtraction of the buffer and polydispersity. Second, the Fourier region, from where the pair distribution function $\rho(r)$ can be determined:

$$\rho(r) = \frac{1}{2\pi^2 r}\int_0^\infty qP(q)\sin(qr)\,dq$$

Where *P(r)* is the particle form factor. $\rho(r)$ concerns to the electron distribution averaged over a radius *r.* The $\rho(r)$ curves are used to obtain the averaged particle shape.

Third, the Porord region serves to determine the Porod invariant that gives surface information and provides useful information like the Porod volume and molecular weight for compact particles at high *q* values. The Kratky plot ($q^2I(q)$ *vs q*) is a useful tool especially for determining if the protein is unfolded or if disordered conformations are present in the sample (Boldon et al., 2015).

Scope of the thesis

Numerous cellular functions involve multi-domain proteins where their domains are connected by flexible linkers of different lengths. In particular, RNA interactions with these multi-domain proteins play essential roles in gene regulation and are involved in each step of RNA metabolism. RNA binding proteins (RBPs) are mostly composed of multiple domains: RNA binding domains (RBDs) or RNA recognition motifs (RRMs). It is widely studied and usually known the structure and how a single RRM domain recognizes its RNA/DNA or protein binding partners. However, the way multiple domains interact with each other and/or the substrate by modulating their structure and dynamics is not well understood. Typically, these interactions are weak and required to be regulated very specifically since they are involved in numerous machineries that mediate key cellular functions. The regulation of these weak interactions are mostly done cooperatively between the domains and in some cases the conformational flexibility of disordered linker regions can further modulate the multi-domain interactions with their substrates.

The multi-domain protein U2AF is essential during spliceosome assembly and it is composed of a large subunit U2AF65 and a small subunit U2AF35. The RRM1-RRM2 domains of U2AF65 are in charge of the recognition of the polypyrimidine tract (Py-tract) at the 3’-SS, while U2AF35 interacts with the AG dinucleotide at the intron terminal. The Py-tract can vary in length and strength (i.e. different binding affinities). The regulatory mechanisms by which U2AF is able to specifically recognize such diverse sequences still remain unclear.

The first aim of this thesis is to study the conformational space sampled by the RRM1-RRM2 multi-domain in the free form in solution and thus understand the interplay between the domains connected by a flexible linker.

An important goal of this work is to unveil the role of the flexible linker between the two domains as a modulator of the RRM1-RRM2 conformation and during RNA binding of different strength Py-tracts.

A further investigation was performed in order to understand the function (if any) of U2AF35 during Py-tract recognition. A study was carried out on the conformational changes and dynamics of U2AF65 in the presence of U2AF35 UHM domain (U2AF homology motif) before and during RNA recognition.

Finally, using NMR, a methodical study was performed regarding the effect of different FRET fluorophores on the integrity of the U2AF protein.

During the course of this work, we apply an integrated approach combining different solution methods used in structural biology, e.g. NMR, SAXS, and FRET. Solution methods in structural biology are the most suitable to study the interdependence of structure and dynamics of biomolecular systems.

CHAPTER 3:

Materials and Methods

3.1 Materials

3.1.1 Constructs

RRM12 (U2AF65)	148-342 construct cloned in pETM11 vector (6His, TEV cleavage site)
nxRRM12 (U2AF65)	140-342 construct cloned in pETM11 vector (6His, TEV cleavage site)
GS-linker constructs	Residues between 233-254 were replaced totally or partially by $(GGS)_N$ sequence, cloned in pETM11 vector
URRM12 (U2AF65)	88-342 construct cloned in pETM11 vector (6His, TEV cleavage site)
UHM (U2AF35)	39-152 (C67S, C102S1) construct cloned in pET9d-NHis vector (6His)

*****All the constructs and plasmid were provided by the Sattler group and the PEPF (Protein Expression and Purification) facility. The GS-linker constructs were purchased from Eurofins Genomics and further cloned into the pETM11 vector.

3.1.2 Solutions and Media

LB medium (Lysogeny broth)	10g/L Tryptone, 5g/L Yeast Extract, 5g/L NaCl
^{15}N labelled M9 Medium (1 liter)	100 ml M9 salt solution (10X), 10ml trace elements (100X), 20ml 20% glucose, 1mM $MgSO_4$, 0.3 mM $CaCl_2$, 50 mg/ml kanamycin, 1 mM thiamin, 1 mM biotin
^{15}N, ^{13}C labelled M9 medium(1 liter)	100 ml M9 salt solution (10X), 10ml trace elements (100X), 2 g ^{13}C-glucose, 1ml 1M $MgSO_4$, 0.3 mM $CaCl_2$, 50 mg/ml antibiotics (accordingly to plasmids/cells), 1 mM thiamin, 1 mM biotin

M9 salt solution (10X)	75.2 g/L $Na_2HPO_4 \cdot 2H_2O$, 30g/L KH_2PO_4, 5 g/L NaCl, 5 g/L $^{15}NH_4Cl$
Trace elements (100X)	5 g EDTA, 0.83 g/L $FeCl_3 \cdot 6H_2O$, 84 mg/ L $ZnCl_2$, 13mg/L $CuCl_2 \cdot 2H_2O$, 10 mg/L $CoCl_2 \cdot 6H_2O$, 10 mg/L H_3BO_3, 1.6 mg/L $MnCl_2 \cdot 6H_2O$

3.1.3 Buffers

Lysis	50 mM Tris (pH 7.5), 500 mM NaCl, 10 mM imidazole, 5 mM ß- Mercaptoethanol
Wash buffer	50 mM Tris (pH7.5), 1 M NaCl, 10 mM imidazole, 5 mM ß-Mercaptoethanol
Elution buffer	50 mM Tris (pH 7.5), 200 mM NaCl, 300 mM, 300 mM imidazole, 5 mM ß- Mercaptoethanol
Exchange buffer	50 mM Tris (pH 7.5), 50 mM NaCl,5 mM ß- Mercaptoethanol
AX-A buffer	0 M salt 50 tris pH 7.5 5 mM ß- mercaptoethanol
AX-B buffer	1 M salt 50 tris pH 7.5 5 mM ß- mercaptoethanol
NMR buffer	20 mM phosphate (pH 6.5), 50 mM NaCl, 1 mM DTT
ITC buffer	20 mM phosphate (pH 6.5), 50 mM NaCl, 3 mM ß-mercaptoethanol

3.2 Methods

3.2.1 Protein Expression and Purification

The U2AF65 RRM1-RRM2, URRM1-RRM2, and all the single point mutations plasmids were transformed into BL21 (DE3) E-coli cells and U2AF35-UHM into BL21(DE3) pLysS cells and grown overnight on kanamycin (50 µg/ml) agar plates. The cells grew initially in an overnight preculture at 37°C, inoculated with colonies from the fresh agar plate, in 100 ml LB medium with 50 µg/ml kanamycin. The next day, the cells were spun down and resuspended in one liter of the final growing medium (LB or M9) starting with an OD_{600} of ~0.1. The cells were grown at 37°C until an OD_{600} of 0.6-0.8 and then induced with 500 µM IPTG (Isopropyl-β-D-thiogalactopyranosid) and then the protein was expressed at 20°C overnight. Next day, cells were harvested by centrifuging for 25 min at ~5000 rpm. After harvesting, the pellet was washed with 1X PBS (Phosphate-buffered saline) and centrifuged for 15 min (in case of not immediate purification, the pellet was stored at -20°C).

For protein purification the pellet was resuspended in ~30 ml lysis buffer, 200 µl of 40 mM AEBSF protease inhibitor, 50 ul of 1 mg/ml DNAse and a pinch of Lysozyme were added. To lyse the cells, the cell suspension was sonicated on ice with three cycles of 4 min at 52% amplitude (Bandelin Sonaplus sonicator) and 2 min on ice in between cycles. The lysate was centrifuged for 50 min at ~16000 rpm; in the meanwhile the IMAC column for His-tag purification was prepared, using 3ml of Ni-NTA beads per liter of culture. The column was washed with water and equilibrated with lysis buffer. After centrifugation the supernatant was filtered through a 0.45 µm filter and passed through the IMAC column 2 to 3 times, the column was washed 4 times with 30 ml of wash buffer and afterwards the protein was eluted with 15 ml of the elution buffer and collected in a falcon tube. For the TEV tag (Tobacco-Etch-Virus protease) cleavage, the eluted protein plus 1ml of 1 mg/ml TEV protease was dialyzed against one liter of exchange buffer, in order to remove imidazole and to achieve an efficient TEV cleavage. The next day, the cleaved protein was passed again through the with exchange buffer equilibrated IMAC column and the protein was collected without tags, which stick to the column.

To obtain pure protein, Size Exclusion Chromatography (SEC), using a Superdex75 Highload 16/60 GE Healthcare column, was performed. The column was equilibrated with NMR buffer at a flow rate of 1 ml/min, the protein was concentrated to 5 ml and then loaded on the SEC column. Afterwards the protein was collected from the peak fractions. The pureness and yield of the desired protein was checked by SDS-PAGE. Finally the protein was concentrated to 1-2ml using Amicon Ultra centrifugal filter with MWCO 10 KDa and snap frozen in liquid nitrogen and stored at -80°C.

For the purification of the U2AF heterodimer (HD), both proteins U2AF35 and U2AF65 were purified in IMAC column separately. In the case of the purification of the UHM-U2AF35 protein it was required to proceed with inclusion bodies purification, where after the harvesting and sonication it was necessary to use 8 M urea in the buffers and to resuspend the pellet to proceed with the rest of the steps until the dialysis to exchange buffer to remove the urea slowly. In this case the construct has no TEV cleavage site, so it was not necessary to add TEV protease.

For the formation of the U2AF heterodimer (HD), both proteins U2AF35 and U2AF65 being both in the exchange buffer (50 mM NaCl) were incubated together for aprox. 1 h, after that the HD was loaded in an anion exchange column Resource Q GE Healthcare, then the protein was eluted with a linear gradient of NaCl from 0 to 500 mM Nacl (AX-A and AX-B buffers) using a flow rate of 2 ml/min until final volume of 30 ml. The peak fractions were collected and analyzed on SDS-PAGE gel, then they were concentrated and loaded on the SEC column and again the peak fractions were analyzed on SDS-PAGE. Finally the protein was concentrated and left at 4°C until its use, the HD could not be frozen since it presented degradation after thawing. However both proteins (U2AF35 and U2AF65) could be frozen separately until the preparation of the HD was necessary.

3.2.2 NMR titrations

The NMR experiments were performed using AV500 MHz Cryo, AV600 MHz Cryo, AV800 MHz Cryo or AV900 MHz Cryo Bruker NMR spectrometers and TopSpin software. The samples contained ~300 µM protein concentration and adding RNA step by step increasing its concentration until achieve saturation (i.e no chemical shift or intensity changes, for U2AF65:RNA usually 1:1.2-2 ratio). The titration points were followed by recording $^{1}H,^{15}N$-HSQCs (Heteronuclear single quantum correlation) experiments at 298 K.

3.2.3 NMR dynamic experiments

The experiments recorded to study dynamics on the RRM1-RRM2 U2AF65 protein were ^{15}N-T1 (T1 relaxation, to study spin-lattice relaxation), using 12 to 14 relaxation time delays between 0.0216 and 1.6632 s including two duplicates. ^{15}N-T1ρ (T1ρ relaxation, to study spin-lattice relaxation in the rotating frame), using 12 relaxation time point delays between 0.005 s to 0.14 s with at least two duplicates. ^{15}N-T2 (T2 relaxation, to study spin-spin relaxation), using 12 relaxation time point delays between 0.01208 and 0.19328 s (values of the different delays for T1, T2, T1ρ and pulses are shown in the appendix). ^{15}N-HetNOE (Heteronuclear NOE, to study flexibility) and Relaxation Dispersion (CPMG) experiments (to study excited states). All the samples used contained between 300 µM to 1 mM of protein. The data processing and analysis were done using NMRPipe and CCPNMR-Analysis software (Delaglio et al., 1995; Vranken et al., 2005)

3.2.4 Paramagnetic Relaxation Enhancement, PRE experiments

The PRE experiments were performed using a spin-label which is a stabilized nitroxyl radical (with a free electron) called IPSL (3-[2-iodoacetamido]-PROXYL), which was covalently attached to a single cysteine mutant in the protein. The single point mutations were obtained following the protocol from QuikChange II Site-Directed Mutagenesis Kit from Agilent, here are the mutations used: RRM1-RRM2_318C, RRM1-RRM2_155C, RRM1-RRM2_155C_318C, nxRRM1-RRM2_318C, nxRRM1-RRM2_155C, nxRRM1-RRM2_254G_318C, RRM12_GSlinker_318C.

After the protein purification the labelling procedure was performed: First of all a strong reduction of the cysteines was done with 10 mM DTT for several hours or overnight at 4°C. The next day a thoroughly buffer exchange to the labelling buffer (1 M Tris buffer pH 8.0 plus 50 mM NaCl) was done using desalting column or SEC column. Then for the reaction the IPSL was added with a 6 to 10 times excess to the protein. The reaction was performed in the dark, overnight at 4°C. The next day the labeled-protein was washed extensively with NMR buffer using Amicon Ultra centrifugal filter with MWCO 10 KDa in order to remove the excess of IPSL avoiding solvent PRE effects and to completely exchange the buffer to the NMR buffer.

The final samples contained between 200 to 300 µM of protein. First an $^{1}H,^{15}N$-HSQC of the oxidized protein (oxidized spin label attached to the protein) was recorded with an inter-scan delay D1 of 5 s, then in order to reduce the spin label (its unpaired electron) 2-3 mM ascorbic acid was added from a stock of 100 mM. Finally, an $^{1}H,^{15}N$-HSQC of the reduced protein (reduced spin label) was recorded with the same parameters as the HSQC of the oxidized sample.

The evaluation of the PRE effect was done by taking the ratio between the oxidized spectra and the reduced spectra of every residue.

3.2.5 Förster Resonance Energy Transfer, FRET

The single molecule FRET experiments were done in collaboration with the group of Don Lamb from LMU, in particular with the doctoral student Lena Voithenberg. In order to do the experiments it is necessary to attach two fluorophores (donor and acceptor) to the protein, in our case the fluorophores (dyes) were attached covalently via the maleimide group of the fluorophore to the thiol group of a cysteine in the protein (sulfhydryl-maleimide coupling). The two artificial cysteines were designed to be located in each of the two domains, one in RRM1 and the other one in RRM2, optimizing the positions such that in the free form their distance is minimal (i.e high

FRET efficiency) and in the RNA-bound form it is maximal (i.e low FRET efficiency). We used the following pairs of cysteine mutants in the RRM12 and URRM12 constructs: L187C_A318C, L187C_G322C, L187C_G326C.

We used a variety of fluorophores in the green and red wavelengths bought from AttoTec and Life Technologies called Atto532, Alexa647, Atto647N, Cy5.

The general protein labelling procedure was done as follow: The double-cysteine and single-cysteine mutants protein were strongly reduced adding 10 mM DTT overnight, then it was buffer exchanged to tris buffer pH 7-7.5, then an excess of 10 to 20 times of the dye was added and incubated in the dark for 2 hours, finally the protein was extensively washed to remove the excess of free dye.

The solution-based single-pair FRET measurements (spFRET) were done on a confocal microscope with multiparameter fluorescence detection (Kudryavtsev et al., 2012). The free protein was measured at concentrations range of 20-50 pM in potassium phosphate (pH 6.5) buffer containing 50 mM NaCl. the RNA bound form of the protein was measured by adding 5 µM of the respective RNA.

3.2.6 Isothermal Titration Calorimetry, ITC titrations

To perform the ITC experiments, the protein was dialyzed overnight against the ITC buffer and the RNA was dissolve in the same batch of buffer as well to avoid buffer mismatch effects. The protein has to be of high purity to obtain reliable data. The cell and syringe should be thoroughly washed and carefully filled as described in the manual of the machine, special care has to be taken while washing the cell. Usually an incubation with 20% detergent at 60°C for 30 min followed by extensive wash steps with water (around 10 times, preferable autoclaved water) and with the ITC buffer (took from the dialysis) were performed before starting the measurement. The cell cleaning is extremely important since any impurity can give very ugly peaks and/or totally unexpected results. The ITC cell was filled with 200 µl (using 300 µl to ensure complete filling without bubbles) of 20 to 30 µM protein, the reference cell was filled with water and the syringe was filled using 60 µl of 200 to 300 µM of RNA (ligand). The main consideration to define the concentration or the concentration range to be used to obtain good data is the so-called C-value which predicts the shape or sigmoidicity of the curve and it is defined as: $C = K \cdot M_{tot} \cdot n$, where K is the association constant, M_{tot} is the total macromolecule concentration in the cell, and n is the stoichiometry parameter. The optimal values for C are between 5 and 500 ("MicroCal Auto-ITC200 user manual (English)," n.d.). So, for the experiments showed here the concentrations in the cell were taken such that the C-value was within this range to obtain good data.

The experiments were run at 25°C and the experimental set-up was set as follow: Starting with the first injection of 0.4 µl and 0.8 s of duration, and then followed by 20 injections of 2 µl and 4 s of duration. All the injections had 180 s of spacing and a filter period of 5 s, the reference power was set to 6 µcal/sec, the initial delay to 300 s and stirring speed of 750 rpm.

The titrations were done using a Malvern MicroCal iTC200 device and the Origin software included for analysis.

A novel use for the ITC raw data is to study kinetics, which can be done using the KinITC package (Burnouf et al., 2012) from the AFFINImeter software. This was done in the web-based application using the previous acquired ITC data, however the fitting was not accurately enough to do an analysis of the kinetics of the RNA binding to the proteins studied.

3.2.7 Small Angle Scattering, SAXS

All the SAXS experiments were performed by triplicate, ranging concentrations between 10 mg/ml to 2.5 mg/ml, the volume required per measurement was 70 µl in NMR buffer containing 50 mM NaCl to keep good contrast level. If the salt concentration is high the contrast is low. To reduce beam damage it is recommendable to add some additives like glycerine (max. 5%) , thioethanol, TCEP, or DTT, the last one was included in the buffer that contains 1 mM DTT. The exact same buffer is measured twice for referencing which then requires a total volume of 150 ul. The higher protein concentration the better the signal, however the measurement at different protein concentrations is necessary to observe if the protein presents aggregation during the measurements which would make the data not suitable for accurate analysis since in SAXS the average dispersion of the particles present in solution are analysed.

The samples should be of high purity and as well characterized as possible, including SDS-PAGE, gel filtration, SLS, DLS or NMR spectra.

The SAXS experiments were performed in a Rigaku BioSAXS-1000 machine with a collimation using a 2D Kratky system which doesn't require a desmearing process for data analysis.

The analysis of the SAXS data was done using the software Primus.

3.2.8 Static Light Scattering, SLS

The SLS technique measures the intensity of the light scattered to obtain the average molecular weight of a biomolecule, the experiment was done to confirm the size and the complex formation of protein and RNA. Specifically the protein concentration was 300 uM at the injection, however it becomes diluted while it goes through the column, the SLS detector is attached to the size exclusion column (SEC) S75, calibration was done with BSA and it was run at a flow rate of 1 ml/min.

3.2.9 Surface Plasmon Resonance, SPR/BIAcore

Two different approaches were used to study RNA-protein interactions, first by immobilizing protein to the surface of the chip and second by immobilizing the RNA instead. To perform the SPR experiments a NTA (carboxymethylated dextran pre-immobilized with nitrilotriacetic acid) sensor chip was used as well as an SA (carboxymethylated dextran pre-immobilized with streptavidin) sensor chip both from GE-Healthcare, the first one was used to immobilize the histidine-tagged protein to the surface. In this case, it was necessary to use an additional procedure for amine coupling following the protocol from Kimple et.al. in 2010 (Kimple, Muller, Siderovski, & Willard, 2010) aiming for high immobilization levels of the protein. The experiment was run at 25°C in a Biacore 3000 GE-Healthcare machine. The SA sensor chip was used to attach covalently a biotinylated RNA (Biotin-5 PEG-U4A8U4) to the surface, and in this case the RNA was immobilized to a level such the initial RU (response units) were around 100 RU following the wizard provided by the manufacturer for SA sensor chips. The biotinylated RNA and the protein were dissolved in the running buffer (10 mM HEPES pH 7.5, 1 mM DTT, 0.02% NP40, 200 mM NaCl) and the experiment was run at 10°C in a Biacore X-100 GE-Healthcare machine. In all cases, the reference cell was selected such that the ligand flew first through the reference cell and was the flow cell immediately afterwards through the immobilization cell.

CHAPTER 4:

U2AF65 RRM1,2 free is an ensemble of structures in solution

4.1 Evidence of additional non-compact states of RRM1,2

During previous studies in the Sattler lab, it was shown that the tandem RRM1-RRM2 domains in solution adopt a closed conformation or inactive conformation in the free form, and adopt an open conformation upon RNA binding (Mackereth et al., 2011). Additional to this previous NMR studies of RRM1-RRM2 in solution, available SAXS data of the free RRM1-RRM2 suggested that the two domains adopt not only compact conformations, i.e. when the two domains are in proximity in space (closed and open conformations), but that the two domains can adopt more extended or detached conformations. Those extended conformations could be observed from the plot of the pair distance distribution function (Figure 17) that is the average over the entire conformations present in solution. In Figure 17 the black curve corresponds to the RRM1-RRM2 free and the red to the RNA-bound RRM1-RRM2. The RNA bound form (red curve), represents a globular compact conformation of the protein in this case the open state. In contrast, the additional shoulder at larger Dmax in the SAXS curve of the free protein (black curve) represents non compact states in addition to the compact open and closed states.

Given this new evidence of extended conformations of the system, it was obvious to believe that RRM1-RRM2 apoprotein is a highly dynamic system that coexists as multiple conformations in solution, and consequently to study further the greater conformational space of the tandem domains we worked in collaboration with the group of Martin Blackledge at IBS, Grenoble, France, in particular with his student Jie-rong Huang.

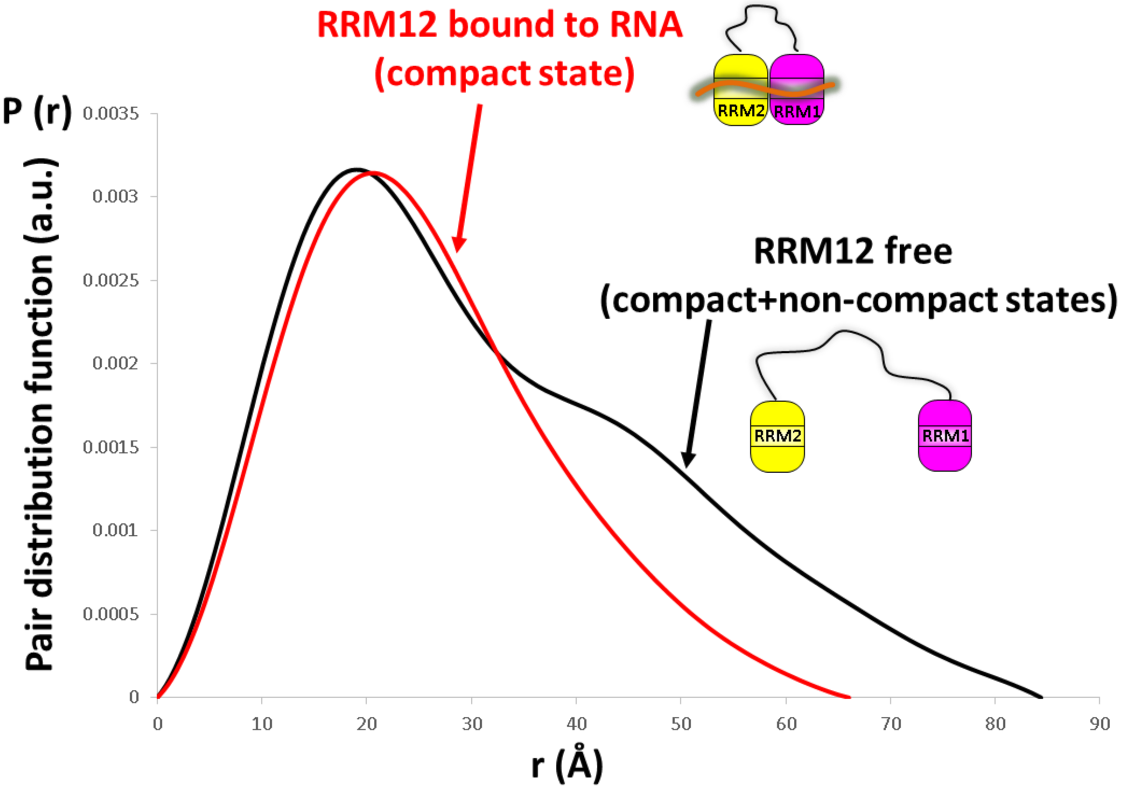

***Figure 17: Pair distribution function of the SAXS data for the RRM1-RRM2** free in solution (black) and for the U9 RNA-bound RRM1-RRM2 (red)*

4.2 Generating the pool

Jie-rong Huang from the Blackledge group started the work by generating a pool of structures for the so-called detached conformation using the statistical coil-model based software flexible-meccano (Ozenne et al., 2012). A randomized linker (i.e. random-coil linker) was built between the end of the C-terminal of RRM1 and N-terminal of RRM2. The RRM1 and RRM2 domains were built by the same software keeping fixed its dihedral angles, i.e. using the known structures of the domains. The tail at the RRM2 C-terminal was also constructed from the statistical coil model as the linker. In total this pool of the detached conformations was composed of 10000 structures.

Next, for generating the pools of the closed and open conformations the XPLOR-NIH software was used, taking as starting structure the NMR-solved structures of each of both conformations (PDB 2YH0 and 2YH1 for the closed and open conformations respectively). The calculations were done starting at 3500 K for a 20 ps dynamics run and then cooling down to 300 K by 50 K decrements of 0.5 ps dynamics runs for equilibration. To maintain the covalent geometry, only the standard potentials were used without additional experimental restrains during the simulated annealing. For the open and closed states the calculations were repeated 5000 times each to ensure that the conformational space of the linker was well sampled.

4.3 Selecting the ensemble for the apoprotein

From the three pools generated earlier (detached conformations and predefined open and closed states) an ASTEROIDS selection described previously by Blackledge group (Nodet et al., 2009; Salmon et al., 2010) was performed from the pool of detached conformations containing the randomized linker. For this selection experimental NMR data (PREs and RDCs) and particularly SAXS data were included. The SAXS data were sensitive to detect the detached conformations and so these data were key during the ensemble selection. Explicit inclusion of the closed and open states (Figure 18B, C) in the prior distribution gives the result that one quarter of conformers adopted these conformations (mainly the closed state). In contrast, while sampling of the pool with only ‘detached’ conformers (Figure 18A) gave as a result a population redistribution that over-samples these closed and open states. The selected ensembles also revealed a small contribution from the open state, although just around 1% were sampled in this state, it was consistently detected. The open and closed conformations appeared within a continuous ensemble envelope (Figure 19) although in very different conformations.

For the optimization of the size of the ensemble different sizes ranging from 10 to 300 structures were used for the selection using ASTEROIDS. 10% of experimental data randomly selected were not used for the selection in order to use it for cross-validation. For the error calculation of the ensemble size, the procedure was repeated 100 times and the X^2 between experimental data and back-calculated for 90% of data used as restrains (active) and the X^2 between experimental data and back-calculated for the 10% left out data (passive) converged when the size of the ensemble is around 200 structures or conformers, the more detailed procedure for the determination of the ensemble can be found in Huang et al., 2014.

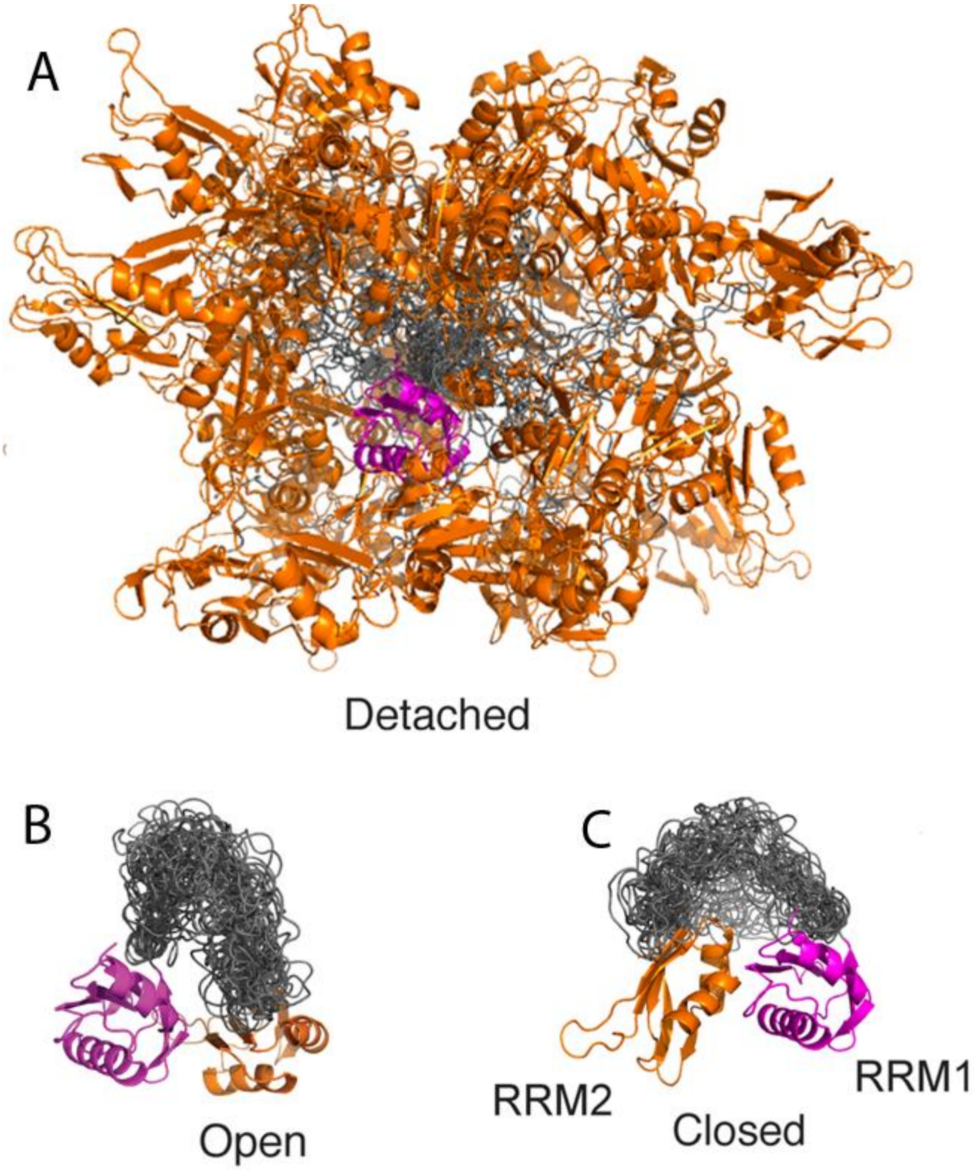
A
Detached
B
Open
C
RRM2
Closed
RRM1

Figure 18: Representative structures of the ensembles, *RRM1 is shown in purple, RRM2 in orange and linker and C-terminus in gray. (A) Representative structures of the detached ensemble, (B) Representative structures of the open ensemble, (C) Representative structures of the closed ensemble.*

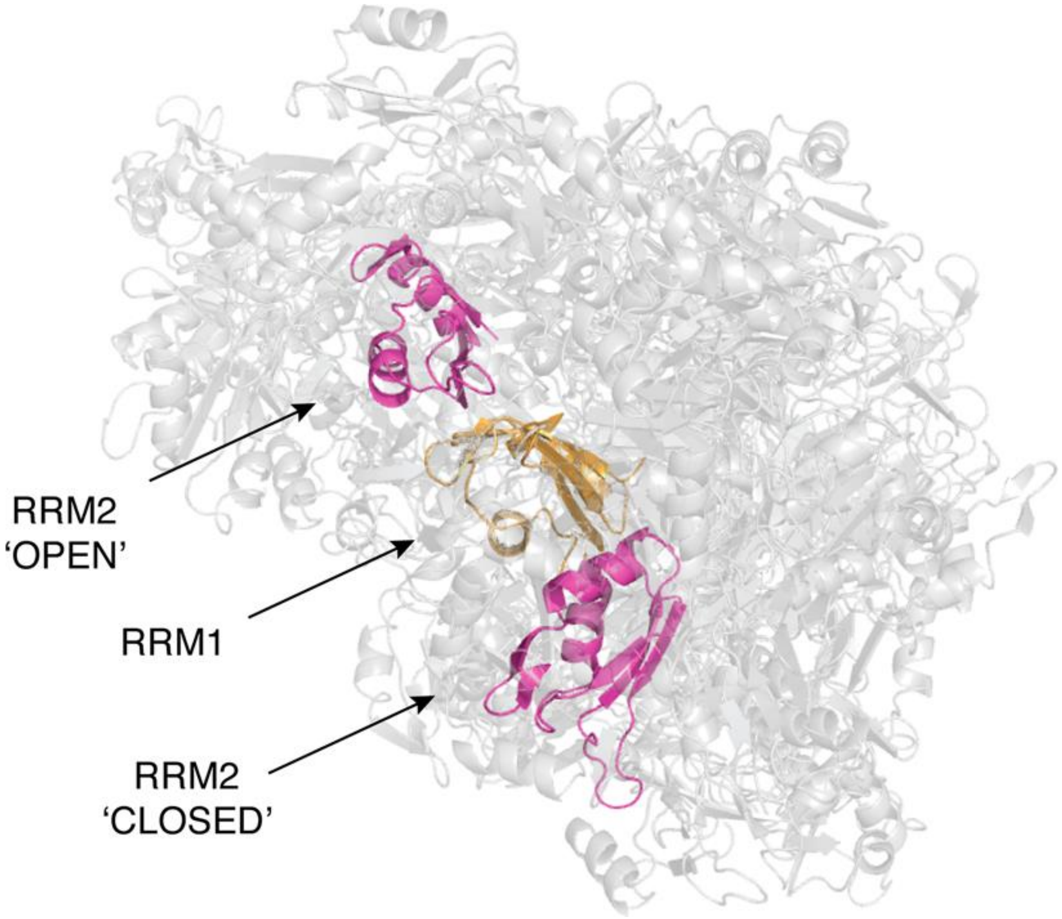

Figure 19: Position of the closed and open states with respect the ensemble. *In yellow is RRM1 fixed and in purple are the open and closed relative positions of RRM2 with respect to the ensemble of all RRM2 structures that were selected from the pool of detached conformations (gray)*

4.4 Error description and crossvalidation

The ensembles were selected based on the agreement with the experimental data (Figure 20 A-C), for the χ^2 calculation and crossvalidation two types of calculation were done, selecting from the detached pool alone or from a combination of open, closed and detached pools. As mentioned before, 10% of all data were left out from the analysis and these were used as "passive" reporters on the predictive nature of ASTEROIDS method. This operation was well suitable to predict independent PREs and

RDCs, however for the SAXS data removing randomly 10% of the data points from a single curve doesn't provide a completely independent R-free measurement, this is due to the low resolution of SAXS data. Moreover, the number of structures necessary to reproduce all independent data was completely dominated by the NMR data in comparison to SAXS data. Hence, this analysis was used to optimize the size of each of the ensembles (Figure 20 D, E), as a result an optimal ensemble size of 200 structures best reproduced experimental data. The resulting χ^2 was approximately one (Table 1).

	PRE+RDC+SAXS			PRE+RDC+SAXS[g]		
	(1184)			No selection (1184)		
Pool[a]	DET,CL,OP	DET	CL,OP	OP	CL	DET
$\chi^{2\ b}$	1502	1577	23321	28130	30341	13376
χ^2/N^c	1.27	1.33	19.70	23.8	25.6	11.3
DET(%)[d]	75.0	100	0	0	0	100
CL(%)[e]	24.3	0	18.3	0	100	0
OP(%)[f]	0.7	0	81.7	100	0	0

Table 1: ***[a] Three different pool used in the selection: detached (DET), open (OP), and closed (CL).*** *[b]* $\boldsymbol{\chi^2 = \sum(param_{calc} - param_{exp})^2 / \sigma^2}$ *is summed over all the experimental data points. [c] Value in b divided by the number of measured points. [d]Percentage of structures derived from the pool detached (DET) in the final selection. [e]Percentage of structures derived from pool closed (CL) in the final selection. [f] Percentage of conformers derived from pool open (OP) in the final selection. [g]Reproduction of the experimental data by the three pools (no selection was done).*

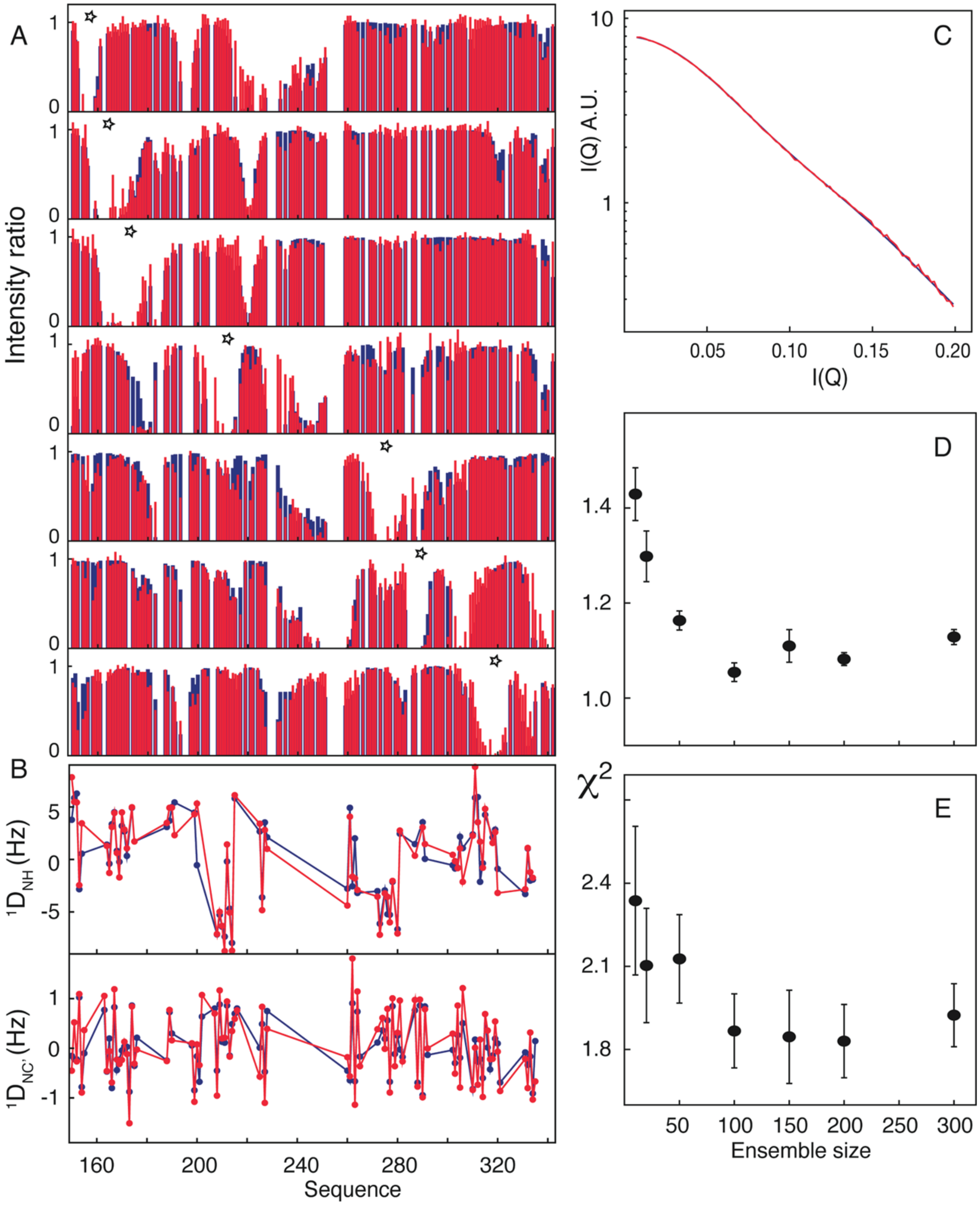

Figure 20: Crossvalidation. Comparison between fitted data (blue) and experimental data (red). *(A) PRE data (intensity ratio between paramagnetic and diamagnetic states (I^{para}/I^{dia}), with the spin labels located at (stars) 155, 164, 171, 209, 273, 287 and 318,*

(B) RDC data and (C) SAXS data. (D) Reduce χ^2 for the active and (E) passive data sets as function of the ensemble size.

4.5 Analysis of the ensemble

After the accurate ensemble selection of the 200 structures that comprises a statistical description of the conformational space sampled by RRM1-RRM2 free in solution, a further analysis of this ASTEROIDS ensemble was done, including free-energy landscape mapping which was fully data-driven (no force field used during this procedure) to select representative ensembles within the distribution of conformers. This analysis provided a description of the conformational space sampled by the system and estimates uncertainties and confidence levels associated with the free-energy description. Application of the free-energy mapping onto the RRM1-RRM2 tandem domains revealed an equilibrium containing highly populated states where RRM1 and RRM2 are in closed proximity, i.e the closed and open states previously observed. Additionally, both conformations remained within a continuous ensemble envelope (Figure 19) that might suggest a pathway of states for interchange between open and closed without provoking large-scale jumps.

Finally, from the ensemble description, the proximity of RRM1 and RRM2 domains in the region resembling the closed state appears to be of electrostatic origin. This was observed after the analysis of conformational properties of the system (Figure 21), where a negatively charged patch of RRM1 (Figure 21 orientation II) which included D206 and E207 (Figure 22), established the main source of electrostatic domain-domain contacts. Additionally, hydrophobic patches in RRM1 (Figure 21 orientations I and III) were found to very infrequently interact with RRM2. To understand in more detail the interaction of electrostatic origin between RRM1 and RMR2, a surface study (Figure 22) was done, and it showed that the RRM2 interfaces that are closest to D206/E207 have a predominantly positive charge distribution (e.g. K286, K328, R334), this finding might explain the bias towards an ensemble with compact conformers and suggested that this ensemble has an electrostatic origin.

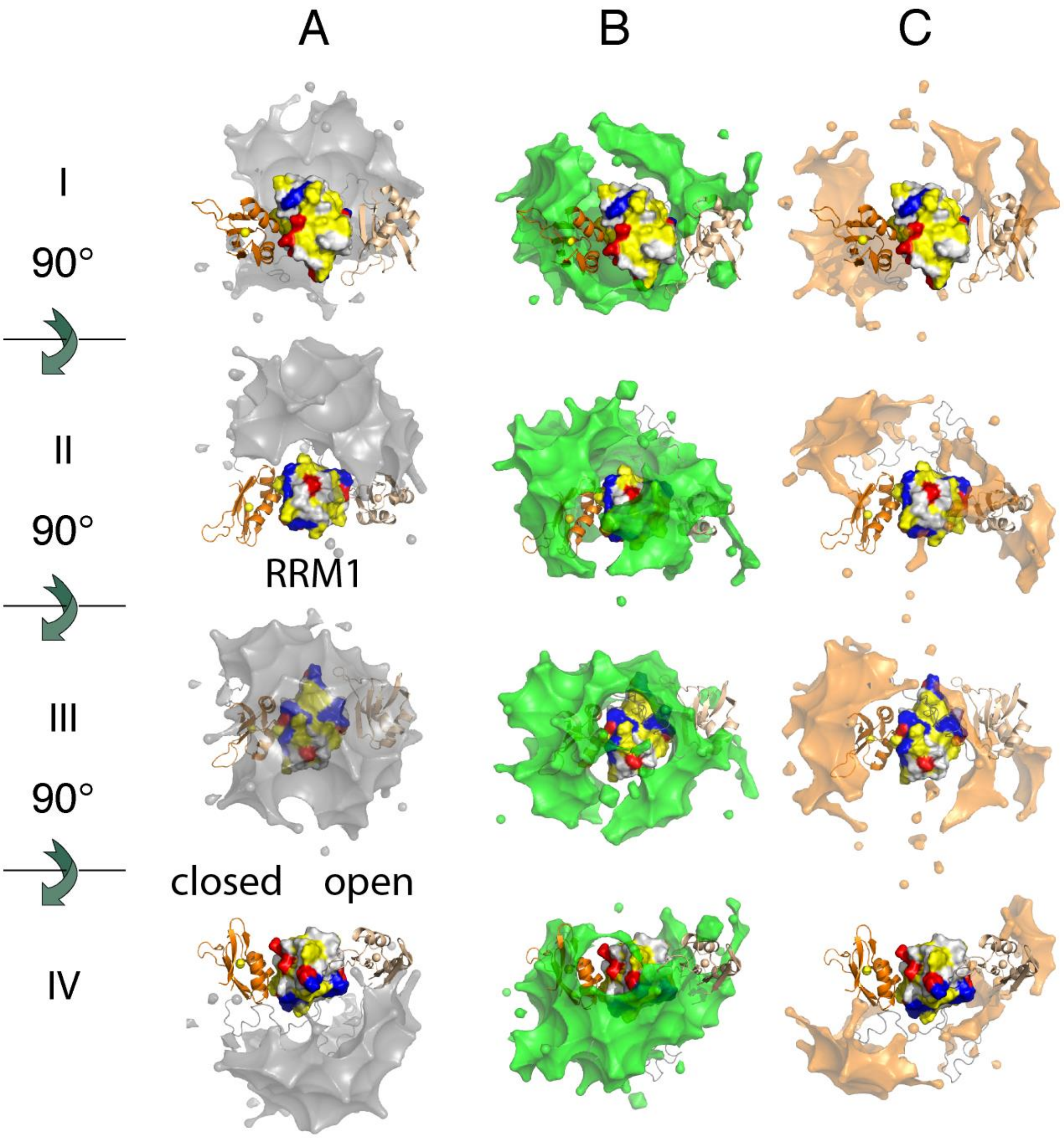

Figure 21: 3D density maps (5% population contour) of RRM2 distribution with respect to RRM1 fixed in the center. *The RRM1 surface is colored according to surface charge, positive-blue and negative-red, RRM2 domain in closed conformation is shown in orange and RRM2 domain in open conformation is shown in magnolia. Hydrophobic*

surfaces are shown in yellow. (A) Prior sampling pool of the detached conformations, (B) Selection based on ASTEROIDS using NMR and SAXS experimental data from detached conformations, and (C) selections based on ASTEROIDS using NMR and SAXS experimental data from a combination of the closed, open and detached conformations.

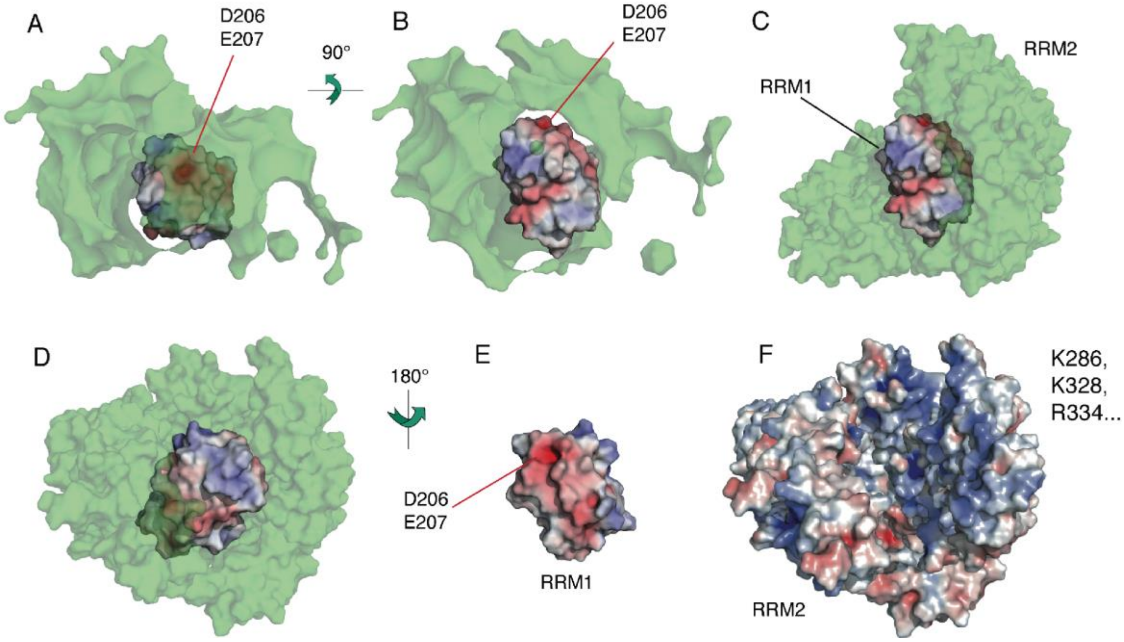

Figure 22: Origin of the recognized encounter complex is electrostatic. *(A,B) Redistribution of RRM2 sampling (green surface distribution) covers acidic patch of RRM1 (surface in Figure 6). (C) Green surface shows RRM2 domains from detached selection (Figure 3A and Figure 6B, whose com values are less than 28 Å from RRM1 (colored surface) com (orientation as in (B)). (D) Alternative orientation of (C). (E) Charge distribution of the face of RRM1 in most frequent contact with RRM2, showing position of acidic patch. (F) Electrostatic distribution on RRM2 surfaces shown in C, D. Most basic region coincides with positions in closest contact with the acidic patch on RRM1 (E).*

4.6 Confirmation of ensemble prediction using PREs

To verify whether or not this ensemble prediction was correct, we performed PRE experiments on RRM1-RRM2 at different salt concentration, using as buffer MES / Bis-tris pH 6.5, since the use of phosphate buffer introduce salt itself. The protein concentration was exactly the same in each sample as well as the DTT concentration and the buffer concentration, the labeling of the protein with the IPSL spin label was done as described previously in the methods section (see section 3.2.4). Regarding the ^{1}H,^{15}N-HSQCs recorded for both, the paramagnetic and diamagnetic states of the spin label attached to RRM1-RRM2, the parameters were kept exactly the same for each of the samples at different salt concentrations, just the pulse calibration and water suppression were adjusted before running each experiment, the inter-scan delay (D1) was set to 3 s.

The reduction of signature PRE effect in the regions of the RRM1-RRM2 interdomain interface upon increasing salt concentration (Figure 23) is consistent with an encounter-like system that is destabilized upon the increase of ionic force between the RRM1-RRM2 domains that are in compact conformations within the ensemble. So, this experimental result verified the prediction from the ensemble that the encounter-like RRM1-RRM2 interdomain contacts are driven by electrostatic interactions.

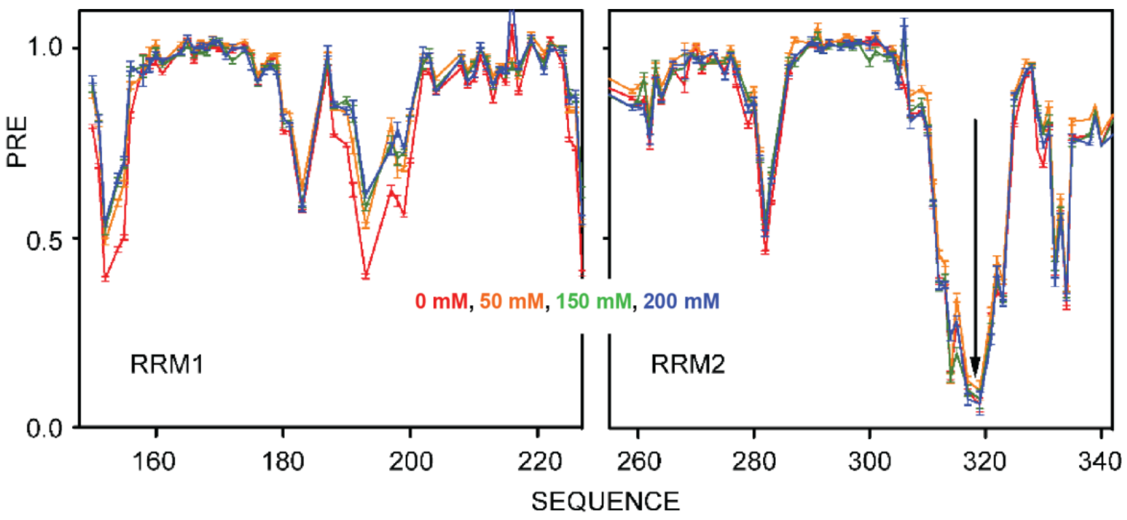

Figure 23: PRE measurements for spin-labeled A318C RRM1-RRM2 at different salt concentrations *NaCl (red – 0mM, orange – 50mM, green – 150mM, blue – 200mM). Negligible salt-dependent effect is seen on RRM2 (containing spin label), while there is clear abrogation of the broadening in RRM1 as a function of salt.*

4.7 RRM1,2 ensemble in solution, not just closed conformation

Using the ASTEROIDS approach to select ensembles from different pools of conformers, the U2AF65 tandem domain RRM1-RRM2 that contain a disordered linker showed large scale conformational dynamics, that include not only the closed (inactive conformation) or open (active, RNA-bound conformation) states that were previously believed to be in equilibrium in solution (Mackereth et al., 2011), but it was found that there is a large population of non-compact or detached conformations (Figure 24). This is a relevant finding since dynamics of multidomain proteins are of key importance in many signaling processes, molecular recognition and specifically in the case of U2AF65. The dynamic interplay between RRM1 and RRM2 serves on their main function during the 3' splice site recognition by identifying the corresponding Py-tract sequences present in human pre-mRNA introns. So, the fact that RRM1-RRM2 exist as an ensemble of closed, open but also detached conformations, help us to understand how dynamic this system is which can be the key for U2AF65 to be so versatile in recognizing a multitude of different Py-tract sequences present during the pre-mRNA processing in eukaryotes.

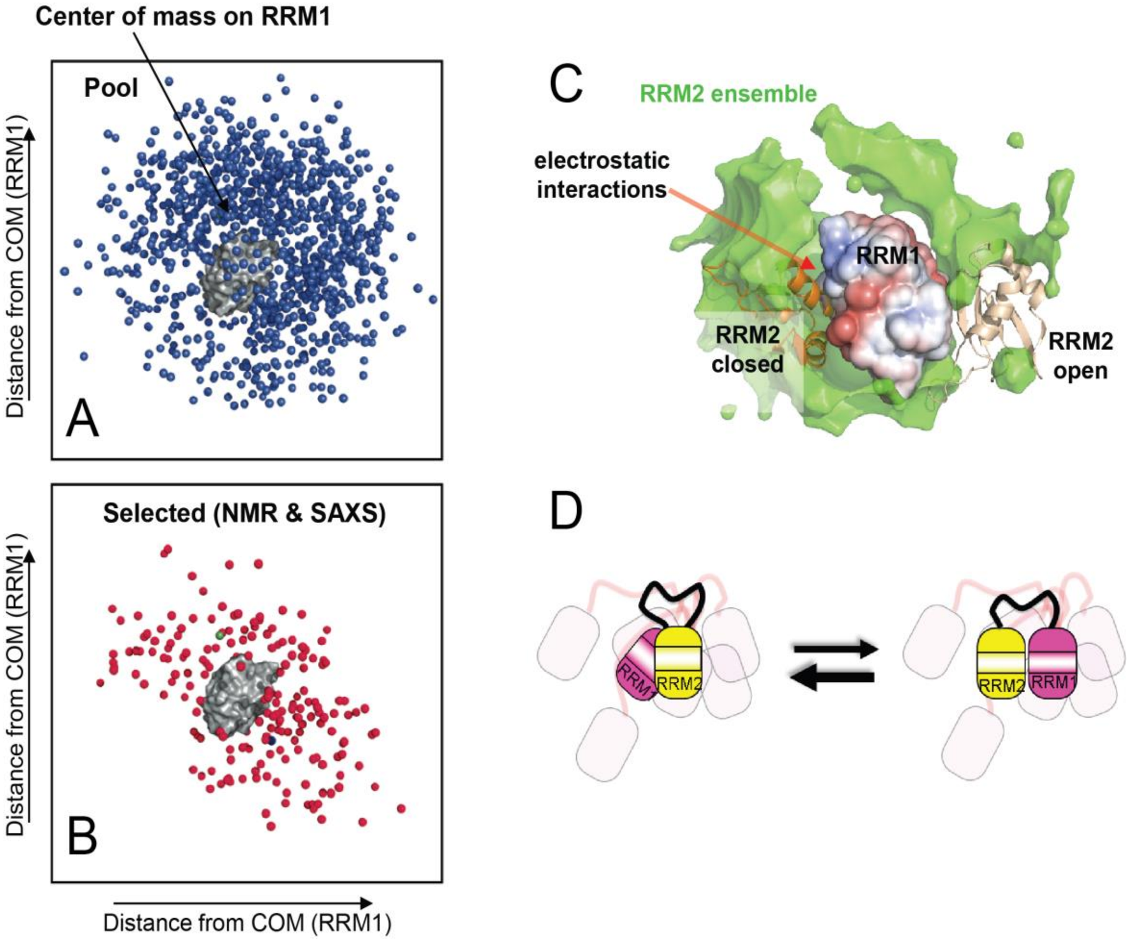
Center of mass on RRM1
Pool
A
Distance from COM (RRM1)
Selected (NMR & SAXS)
B
Distance from COM (RRM1)
Distance from COM (RRM1)
C
RRM2 ensemble
electrostatic interactions
RRM1
RRM2 closed
RRM2 open
D
RRM1
RRM2
RRM2
RRM1

Figure 24: Summary of the ensemble selection of RRM1-RRM2 and conclusion. *(A) Pool of detached conformations, where RRM1 (gray) is fixed in the center of mass and the distribution of positions of RRM2 (blue dots) in respect to RRM1 (fixed position). (B) Selected ensemble using the experimental NMR and SAXS data. (C) Surface representation of the RRM2 distribution (green) in respect to RRM1 (colored) evidencing the interactions of electrostatic origin. RRM2 in the closed conformation is shown in orange and RRM2 in the open conformation is shown in magnolia. (D) Blobs representation of RRM1-RRM2 free in solution, where the closed and open states but also a great deal of detached conformations coexist.*

CHAPTER 5:

Linker between RRM1 and RRM2 has an autoinhibitory role for RNA binding

5.1 Linker between RRM1 and RRM2 interacts with RRM2

In the previous study of the ensemble calculation, pools of structures of RRM1-RRM2 were generated using the computational method flexible-meccano (Ozenne et al., 2012) to produce randomized linkers (random-coil linker) for each structure. However, there were large conformational differences between the RNA-bound RRM1-RRM2 solution structure solved by NMR (Mackereth et al., 2011) and a previous published RNA-bound RRM1-RRM2 X-ray structure (Sickmier et al., 2006) (Figure 25), in which the linker was removed in order to obtain crystals.

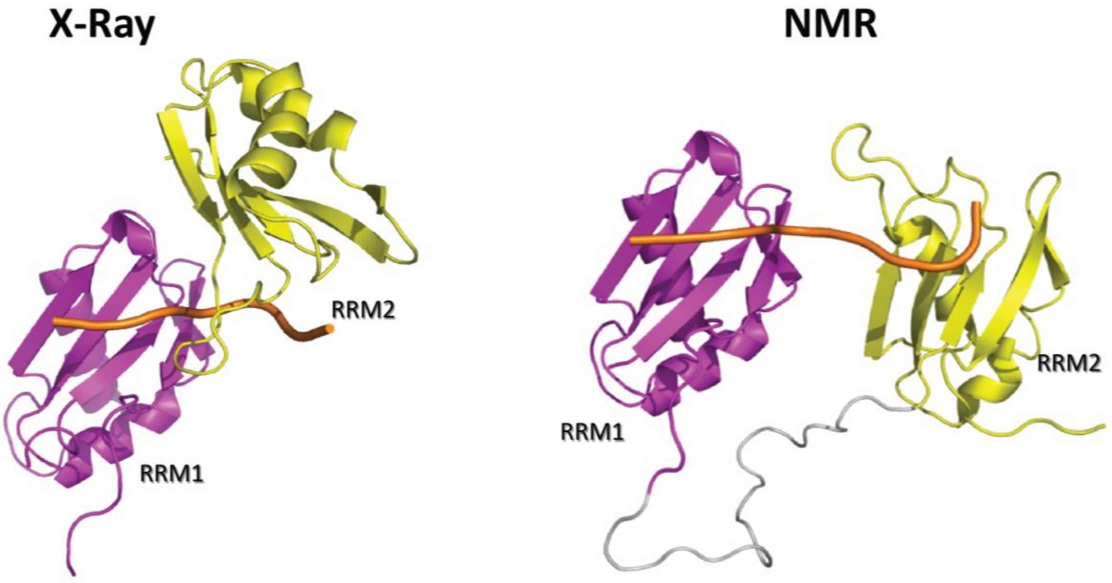

Figure 25: RNA-bound RRM1-RRM2 structures. *On the left is presented the X-Ray structure of RRM1-RRM2 bound to R(P*UP*UP*UP*UP*UP*UP*U) RNA (Sickmier et al., 2006) and on the right is presented the solution structure solved by NMR of the RRM1-RRM2 bound to U9 Py-tract RNA (Mackereth et al., 2011).*

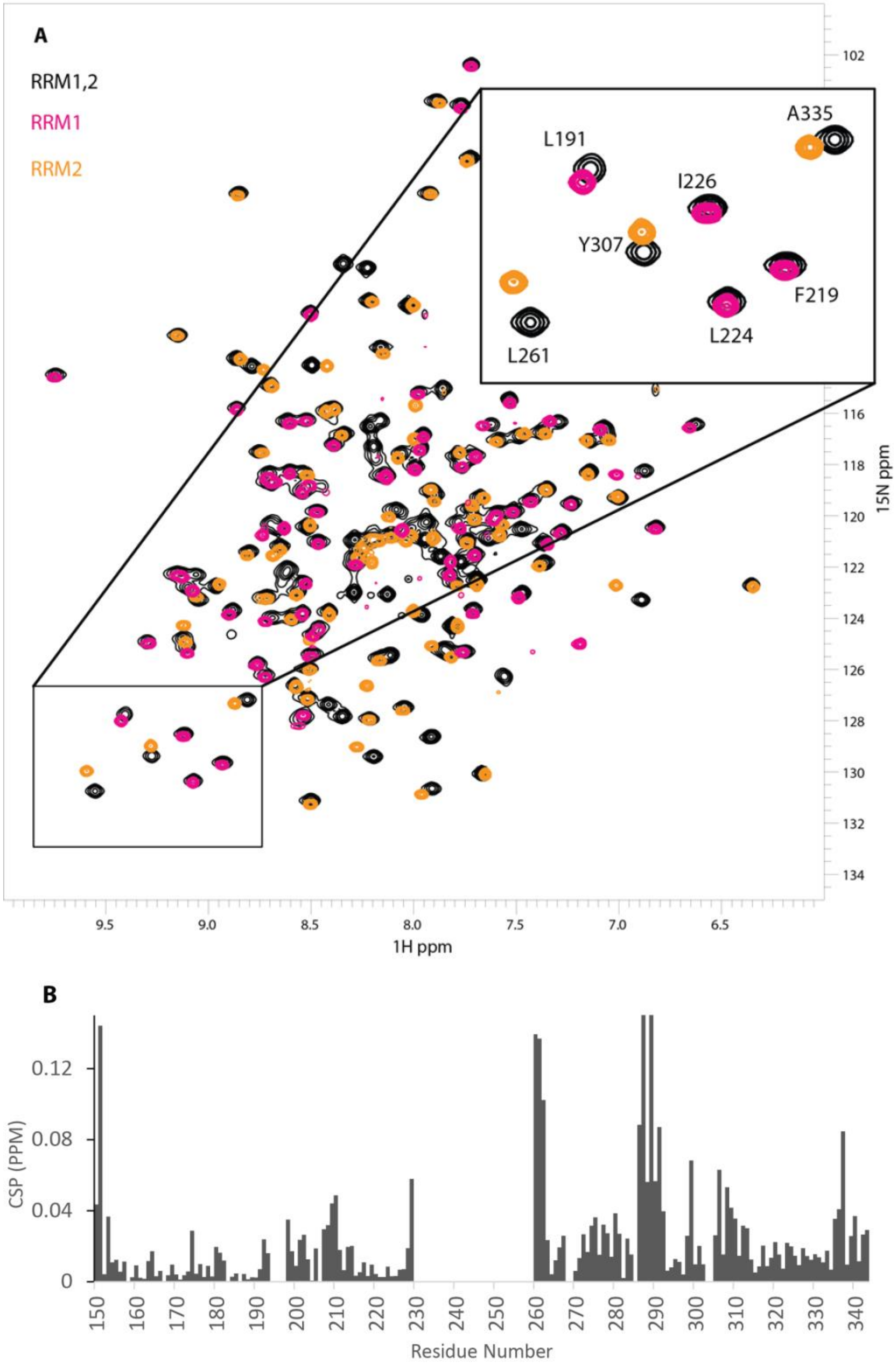

Figure 26: RRM1, RRM2 vs RRM12. *(A) Overlapped spectra of the single domains RRM1 (pink), RRM2 (yellow) and RRM1-RRM2 (black), zoomed in are some of the residues*

belonging either to the RRM1 or the RRM2 domain in the three constructs. (B) Chemical shift perturbations plot of the tandem domain RRM1-RRM2 vs. the single domains RRM1 and RRM2.

Subsequently, given the difference between the solution and crystal structure, and the fact that the pools generated in chapter 4 used randomized linker, it was necessary to further establish if the linker had effects on or interacts with the tandem domains RRM1-RRM2 and or with the RNA in the bound form.

First evidence about the role of the linker could be drawn from PRE experiments, where certain linker residues, particularly residue V254, showed large PRE effect in each of the mutants. Even more striking are chemical shift perturbations when comparing the ^{1}H,^{15}N-HSQC spectrum of the single RRM1 and RRM2 domains versus the spectra of the tandem domain RRM1-RRM2 (Figure 26). It was evident that mainly residues in RRM2 within the tandem RRM1-RRM2 showed significant shifts in comparison to the single RRM2 domain (Figure 26B). These results indicated that the linker is interacting at least partially with the RRM2 domain. The next question was to find out which region of the linker is interacting with RRM2 and which effect this interaction has on RNA binding. To do this, we designed a linker containing a GGS sequence (GS-linker). The results are presented in the next section.

5.1.1 The C-terminal of the RRM1-RRM2 linker is involved in the interaction with RRM2

We endeavored to study the linker interactions with the domains, its dynamics and possible function. For this purpose we designed different GS-linker RRM1-RRM2 constructs (Figure 27), which means that the complete linker or either the N- or C-terminal half of it were replaced by a (gly-gly-ser)$_N$ sequence, these kind of sequences are known to have random coil behavior and lack of complicating ionic interactions (Walsh, Meier, Ishima, & Gronenborn, 2010). Therefore the GS-linker removes all possible interactions which the WT-Linker and the RRM domains have.

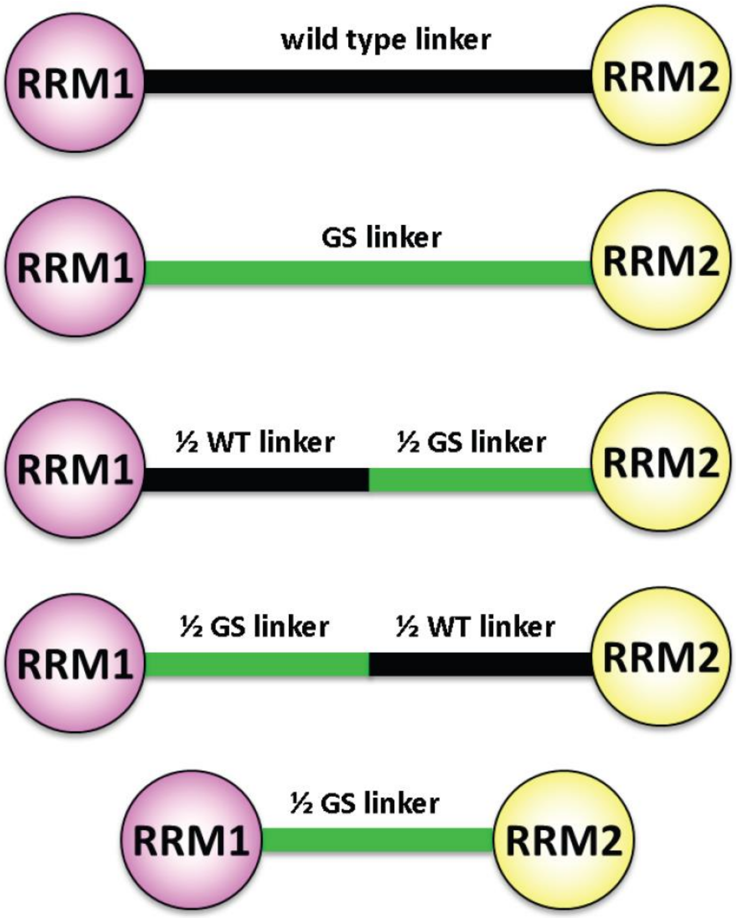
wild type linker
RRM1
RRM2
GS linker
RRM1
RRM2
½ WT linker
½ GS linker
RRM1
RRM2
½ GS linker
½ WT linker
RRM1
RRM2
½ GS linker
RRM1
RRM2

Figure 27: Representation of GS-linker RRM1-RRM2 constructs. *In green are represented the parts of the linker between RRM1 and RRM2 that were replaced by the gly-gly-ser sequence. The full sequence replaced with the GS repetition is: QPLPGMSENPSVYVPGVVST. The sequence replaced in the first half of the linker is: QPLPGMSENP, and the sequence replaced in the second half of the linker is: SVYVPGVVST. One more GS construct was designed to have half of the length of the wild type linker.*

$^{1}H^{15}N$-HSQC spectra for all the constructs were recorded, re-assigned and analyzed. The chemical shift differences of the tandem domains containing the different GS linker constructs in respect to the tandem domain with the wild type linker (WT-RRM1,2) showed that the C-terminal part of the linker is the region that interacts with RRM2 (Figure 28 B). These differences in chemical shift clustered around the central β-sheets of RRM2, suggesting that the linker interacts with these β-strands. Replacing only the C-terminal part of the linker with the GGS motif has the same effect on the RRM2 as the full GS-linker mutant. Hence, it must be the C-terminal part of the linker, which interacts with the RRM2. Additional much weaker chemical shift perturbations were observed in RRM1 due to weak interactions with the N-terminal part of the linker (Figure 28 C). Again this is confirmed by the mutant, where only the N-terminal part is replaced by GGS, which shows the same changes in chemical shifts in RRM1 compared to the WT as the full GS linker mutant.

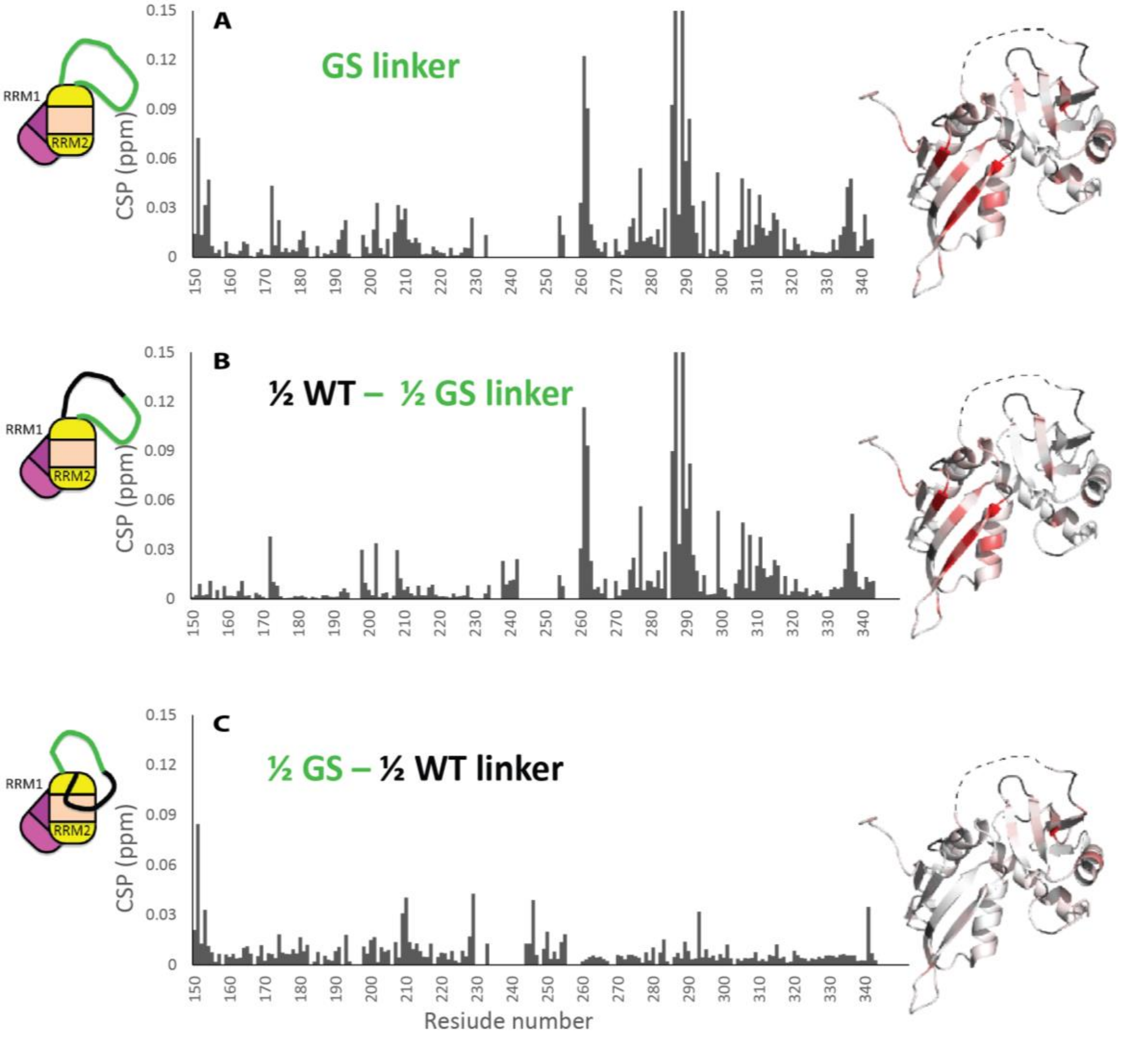

Figure 28: Chemical shift perturbations of the GS-linker constructs. *(A) CSP of the full GS-linker construct vs the wild type linker RRM1-RRM2. (B) CSP of the 1/2WT-1/2GS-linker construct vs the wild type linker RRM1-RRM2. (C) CSP of the 1/2GS-1/2WT-linker vs the wild type linker RRM1-RRM2. On the left are shown blobs representations of the GS-linker constructs. On the right the corresponding chemical shift perturbations are mapped onto the RRM1-RRM2 structure in a red color gradient, such that the more intense red-colored residues are the ones with the largest CSP.*

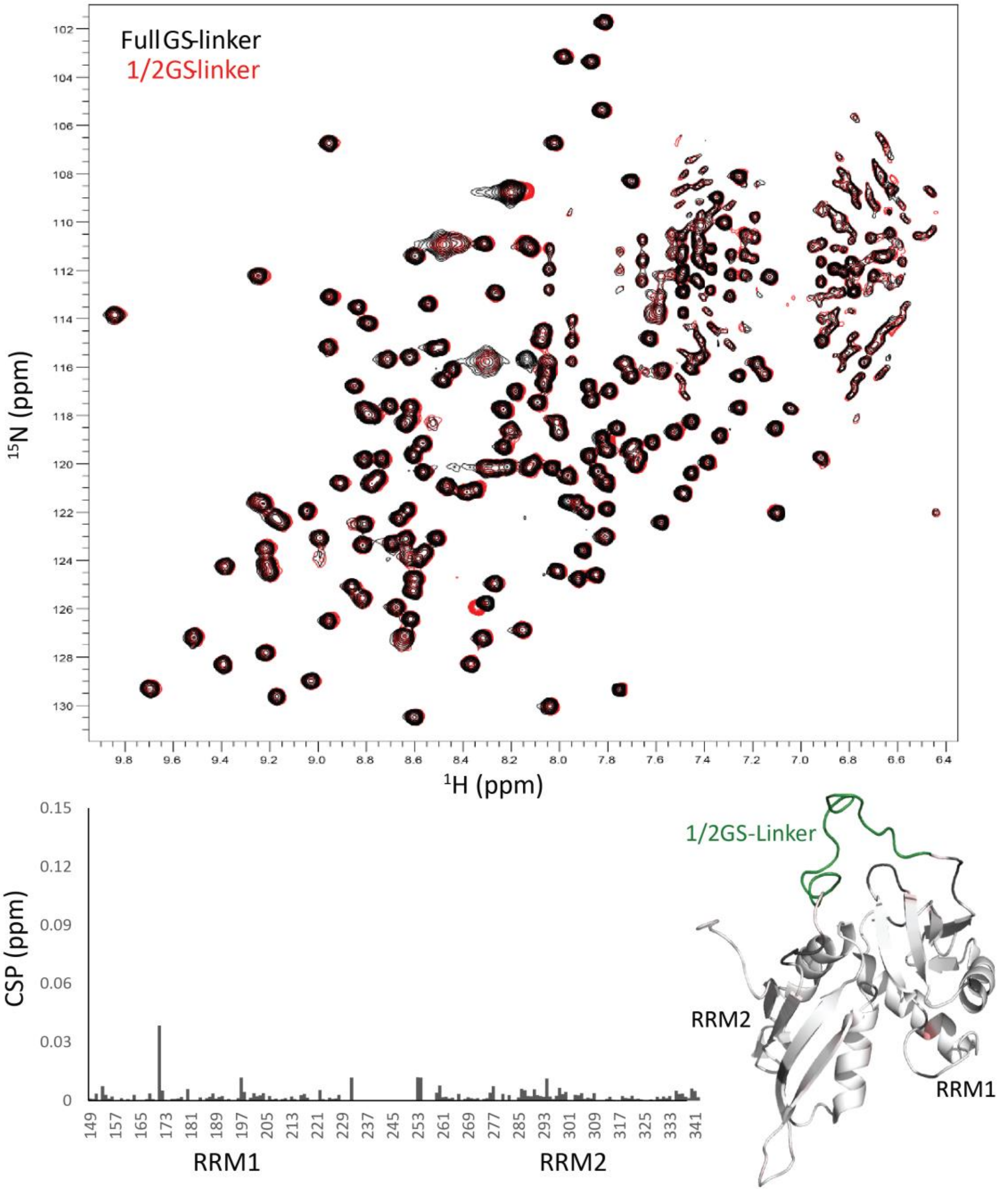

Figure 29: Chemical shift perturbation of different length GS-linker. *Here are presented the chemical shift differences between the full GS-linker and the half-length GS-linker RRM1-RRM2 . On the right the chemical shift perturbations are mapped onto the RRM1-RRM2 structure in red color such that the red color represents CSP values of 0.1 ppm or higher.*

To study if the length of the linker had any effect on RRM1-RRM2 we compared the full GS-linker construct versus the half-length GS-linker (Figure 29) and the chemical shift differences were neglectable, implying that the length of the GS-linker does not have any effect on the domains.

5.2 Effects of the linker/domains interactions on RNA binding

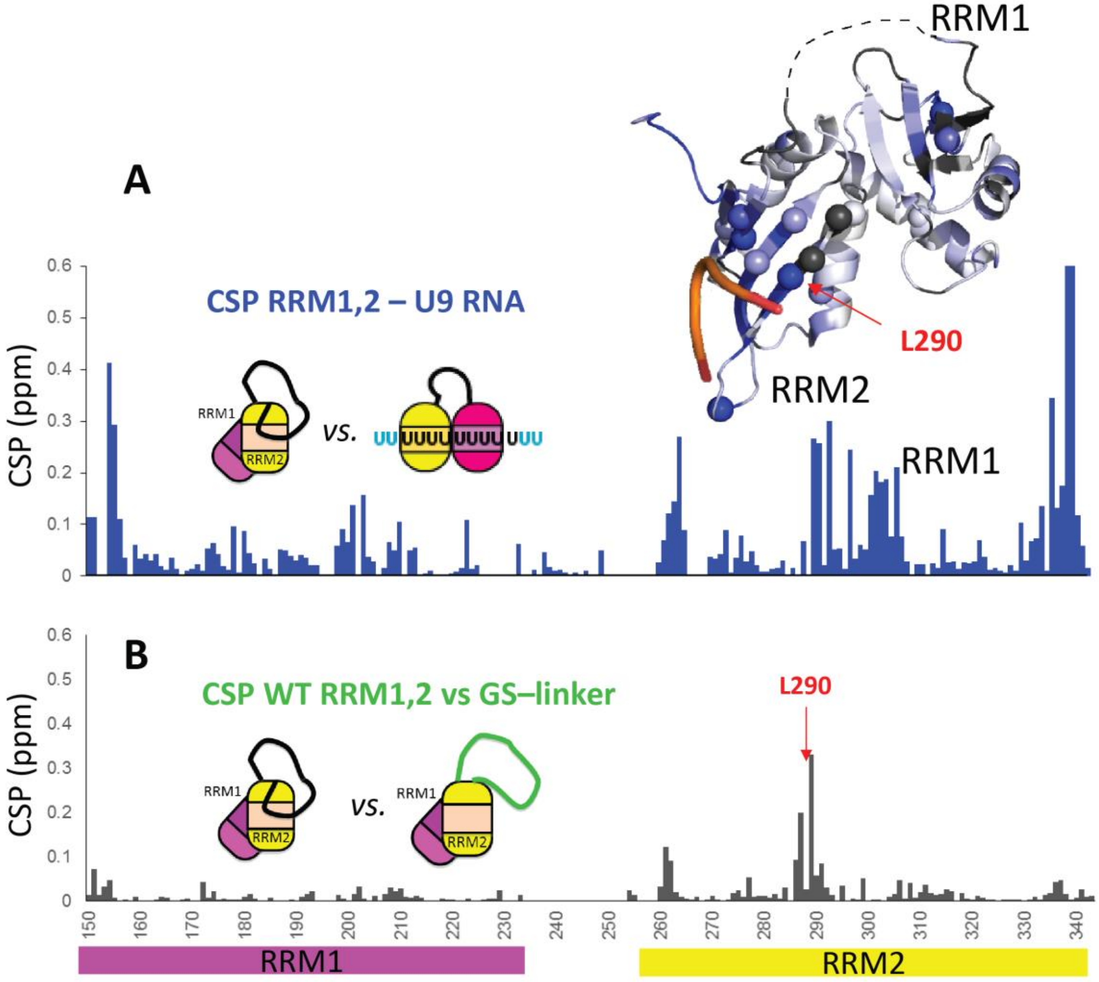

Figure 30: U9-bound RRM1-RRM2 vs. GS-linker chemical shift differences. *(A) Plot of the chemical shift perturbations of the RRM1-RRM2 vs the U9-bound RRM1-RRM2. (B) Plot of the chemical shift perturbations of the RRM1-RRM2 vs GS-linker RRM1-RRM2. On the structure are mapped the CSP of (B) where the blue gradient represents the residues with larger (darker blue) CSP values and lower CSP values (lighter blue). It is*

also shown the RNA (maroon) binding interface of the RRM2 domain. The spheres represent the residues with a CSP larger than 0.2 ppm upon RNA binding.

Besides small interactions with the C-terminal and other minor encounter-like regions in RRM2, as mentioned before the linker residues seem to interact mainly around the central β-sheets of RRM2, which is also the key region for RNA binding. When comparing the region in RRM2 that is implicated in the interactions with the linker versus the regions of the wild type RRM1-RRM2 involved in RNA binding (Figure 30), it appeared that the region in RRM2 interacting with the linker corresponds to the RRM2 regions involved in RNA binding. Also, the regions in RRM1 weakly interacting with the linker are also involved in RNA binding.

This becomes even more evident when looking at the superimposed spectra of the GS-linker RRM1-RRM2 construct with the wild type and with the single domains which (Figure 31) show that residues of RRM2 within the GS-linker construct shift towards the RRM2 peak positions in the single domain. On the other hand, the peaks of RRM1 within the GS-linker construct remain almost unaffected in the same position as they are in the wild type construct. When mapping those shifts on the RRM1-RRM2 structure it can be seen that they overlap with the RNA interface. This result was surprising and allowed us to speculate about a possible task of the linker as inhibitor during RNA binding.

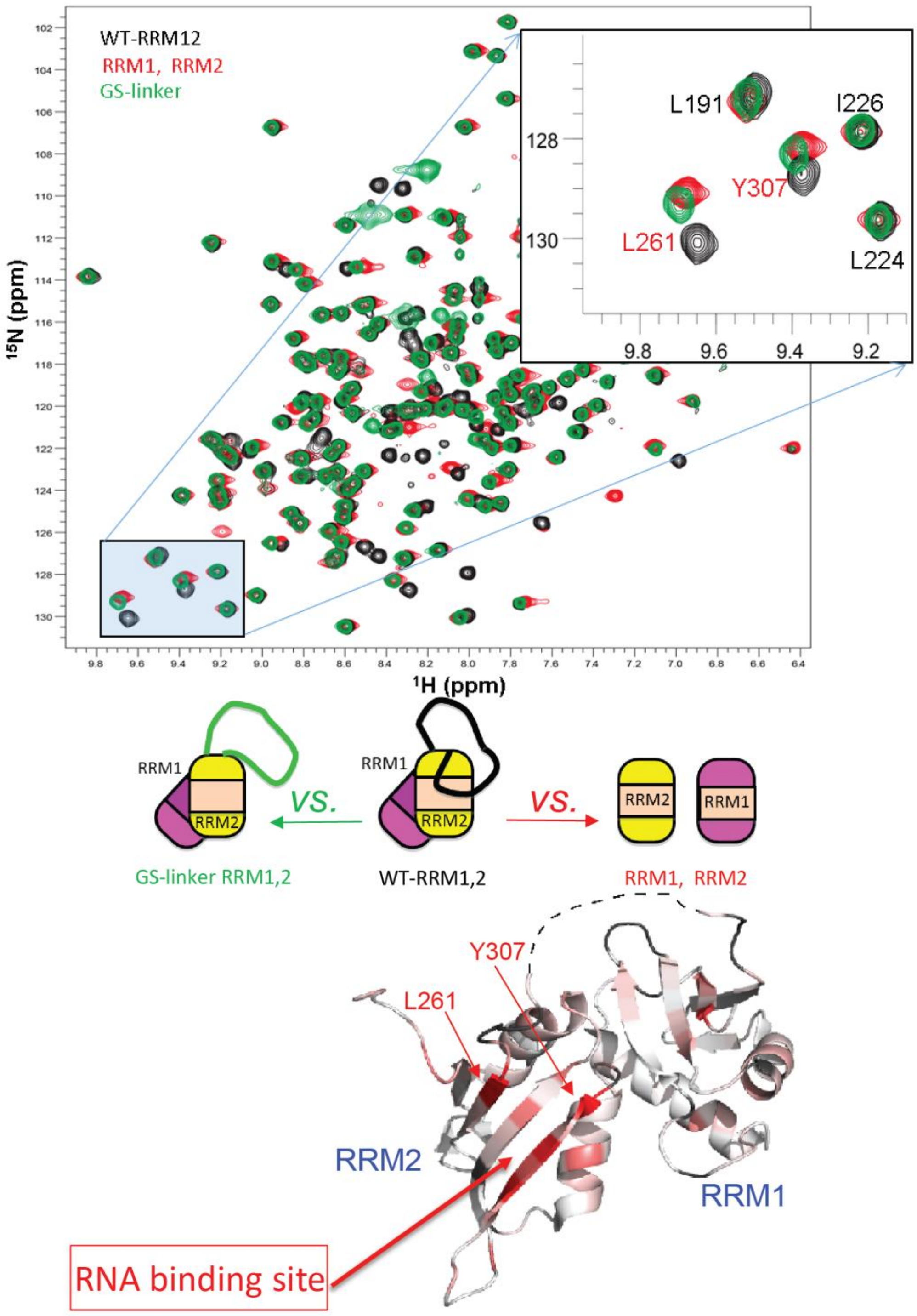

Figure 31: WT-RRM1-RRM2 vs GS-linker vs single domains. *Upper panel: Overlapped spectra of the WT-RRM1-RRM2 (black), GS-linker RRM1-RRM2 (green) and single*

domains RRM1 and RRM2 (red). Lower panel: Structure with the mapped chemical shift perturbations of WT vs GS-linker constructs, and blobs representations of the different constructs.

This new findings of a strong interaction between RRM2 and the C-terminal part of the linker motivated us to look into the sequence conservation of linker residues. The high conservation throughout several species, especially in the C-terminal part of the linker, and the high content of hydrophobic residues show the biological importance of the found interactions between RRM2 and the C-terminal part of the linker (Figure 32).

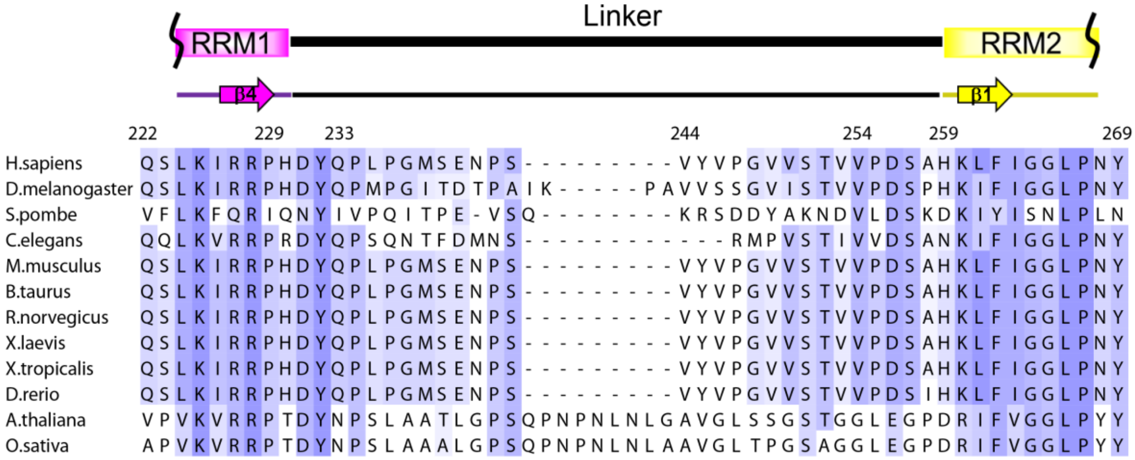
Linker
RRM1
RRM2
β4
β1
222
229
233
244
254
259
269
H.sapiens QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
D.melanogaster QSLKIRRPHDYQPMPGITDTPAIK-----PAVVSSGVISTVVPDSPHKIFIGGLPNY
S.pombe VFLKFQRIQNYIVPQITPE-VSQ--------KRSDDYAKNDVLDSKDKIYISNLPLN
C.elegans QQLKVRRPRDYQPSQNTFDMNS------------RMPVSTIVVDSANKIFIGGLPNY
M.musculus QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
B.taurus QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
R.norvegicus QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
X.laevis QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
X.tropicalis QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSAHKLFIGGLPNY
D.rerio QSLKIRRPHDYQPLPGMSENPS---------VYVPGVVSTVVPDSIHKLFIGGLPNY
A.thaliana VPVKVRRPTDYNPSLAATLGPSQPNPNLNLGAVGLSSGSTGGLEGPDRIFVGGLPYY
O.sativa APVKVRRPTDYNPSLAAALGPSQPNPNLNLAAVGLTPGSAGGLEGPDRIFVGGLPYY

Figure 32: Sequence Alignment. *Alignment of the linker among different species showing its high degree of sequence conservation.*

Thus, we decided to study which is the effect of the linker during binding of two different RNAs, representing a strong (U9-RNA) and a weak Py-tract (U4A8U4-RNA). For this purpose, we did NMR titrations with all of the GS-linker constructs (Figure 33) and we observed differences not only on the binding patterns among the constructs but also differences in the magnitude of the shifts.

When focusing the analysis in two constructs: WT-linker and GS-linker constructs (Figure 34) it was easier to evaluate the effects on RNA binding (e.g. on residues in RRM1: G154, I191, F202, I226, Linker: V254, and RRM2: L261, G264, V291, Y307, S336) since there are clear differences in the titrations patterns (Figure 43), especially on the RRM2 domain. This result then, showed that the linker interactions have an effect on RNA binding.

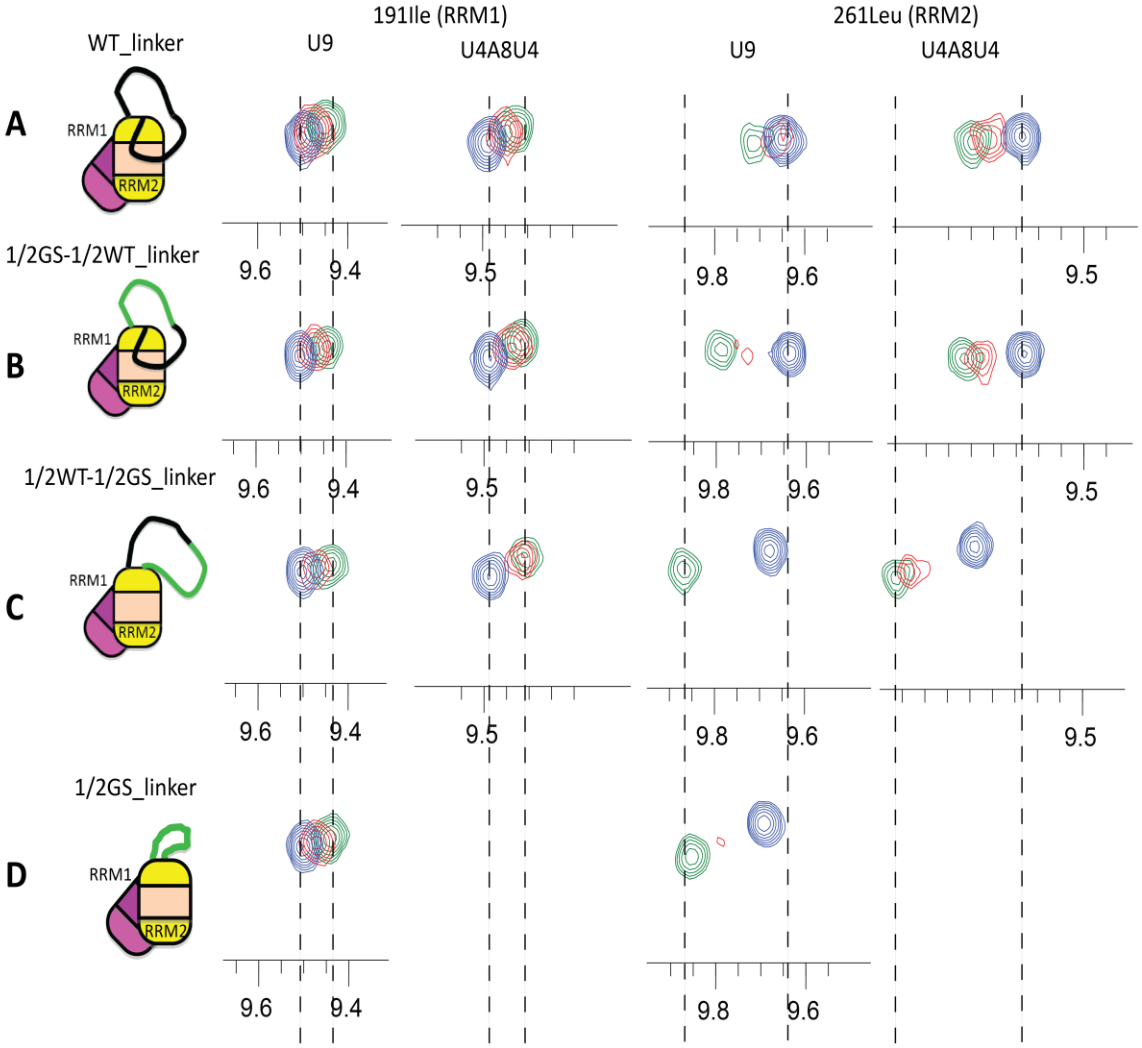

Figure 33: NMR titrations of different GS-linker constructs. *Shown are the peaks of residues I191 and L261 from NMR titrations with U9 and U4A8U4 of (A) the wild type RRM1-RRM2, (B) 1/2GS-1/2WT-linker RRM1-RRM2, (C) 1/2WT-1/2GS-linker RRM1-RRM2, (D) 1/2GS-linker RRM1-RRM2. In blue are the peaks of the reference point, i.e. of the free protein, in red are the peaks of 1:0.5 protein:RNA ratio and in green are the peaks corresponding to the 1:1 protein:RNA ratio.*

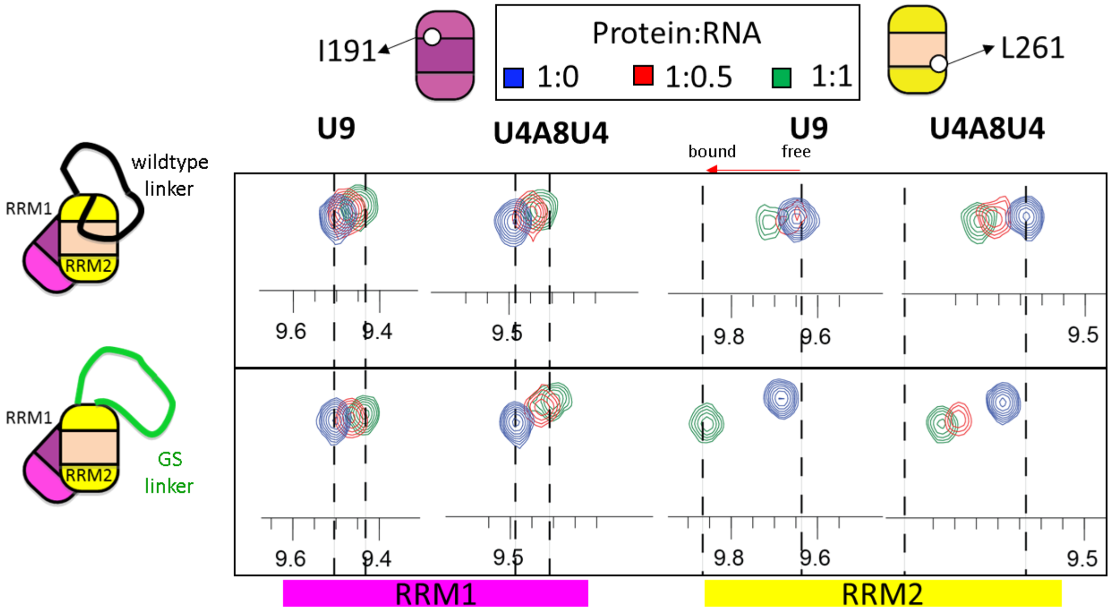

Figure 34: NMR titrations of the WT vs GS-linker constructs. *Differences between a RRM1 residue I191, and a RRM2 residue L261 during the titrations with U9 RNA and U4A8U4 RNA on the wild type RRM1,2 and the GS-linker RRM1,2.*

5.3 Linker provides proofreading against weak Py-tracts

After the results of the NMR titrations, to further understand what the regulatory role of the linker would be, ITC titrations were measured with the GS-Linker-RRM1,2 constructs (**Error! Reference source not found.**). The titrations were done with both, he strong Py-tract U9 and the weak Py-tract U4A8U4. In both cases there were small differences respect to the published values of the wild type RRM1-RRM2 (Mackereth et al., 2011: K_D RRM1,2 + U9:1.32 µM, K_D RRM1,2 + U4A8U4:16.4 µM). Then we did extensively titrations series using the WT-RRM1-RRM2 and the GS-linker RRM1-RRM2 constructs, with both RNAs (Figure 36). Surprisingly the K_D difference between the wild type and the GS-linker constructs when binding the strong Py-tract U9 was very slight (1.28 µM for the WT-RRM1,2, and 1.58 µM for the GS-linker construct), however the K_D when binding the weak Py-tract was significantly different with approximately 2 to 3 times higher affinity of the GS-linker construct versus the wild type RRM1-RRM2

(11.72 µM for the WT-RRM1,2, and 3.35 µM for the GS-linker). While the K_D values for the extended constructs nxRRM1,2 and nx_eGSlinker were 0.22 µM and 0.26 uM for U9 respectively, and 2.58 µM and 1.26 µM for U4A8U4 respectively. This finding set the core of the suggested regulatory role of the linker, since it can act as an autoinhibitor for RNA binding but even more unusual function, it can serve as proof-reading against weak Py-tracts.

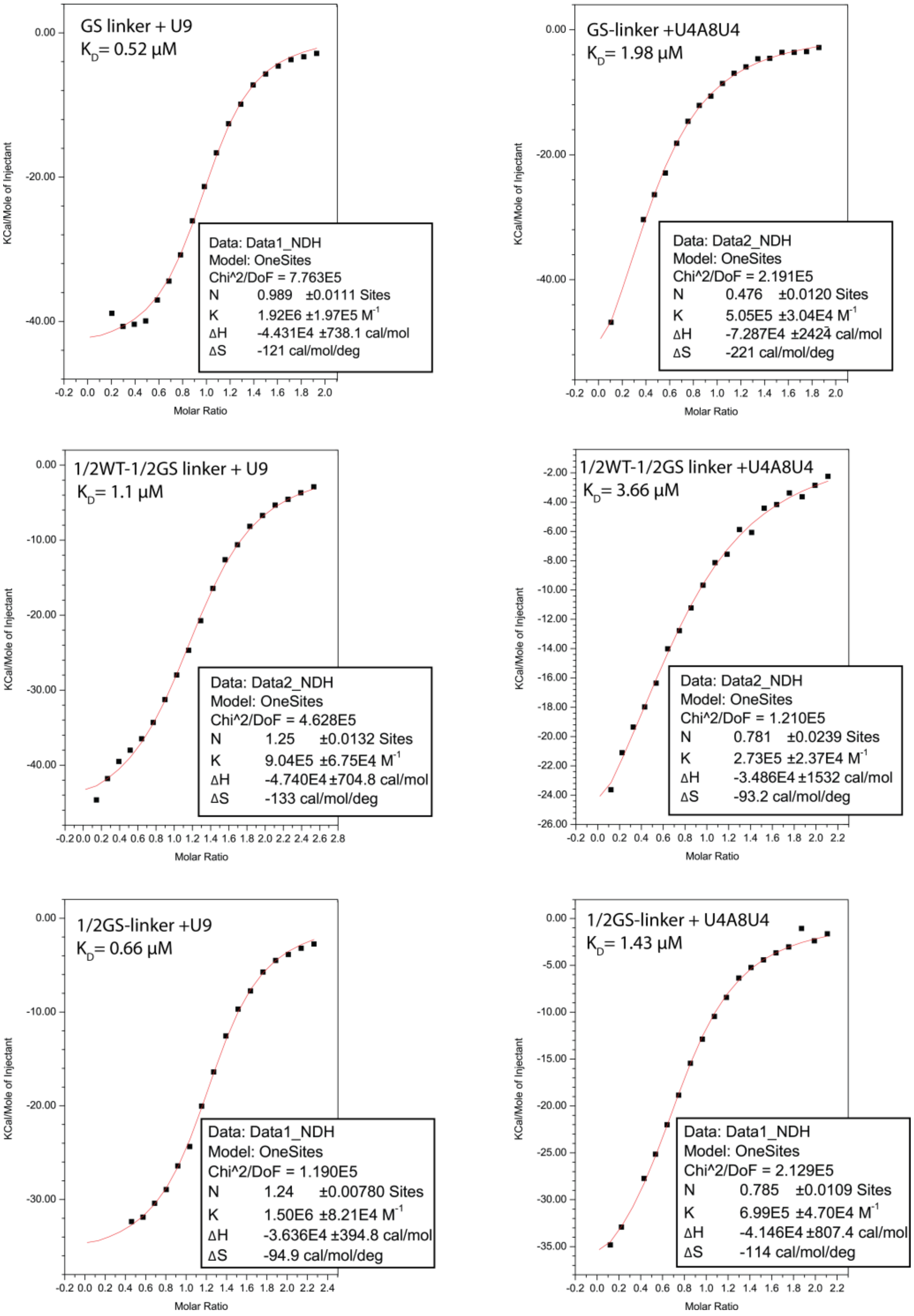

Figure 35: ITC titrations of the GS-linker constructs. *ITC titrations with the strong Py-tract U9 RNA and the weaker Py-tract U4A8U4 RNA at 295 K. Upper panel: GS-linker RRM1,2, middle panel: 1/2WT-1/2GS-linker RRM1,2, lower panel: 1/2GS-linker RRM1,2.*

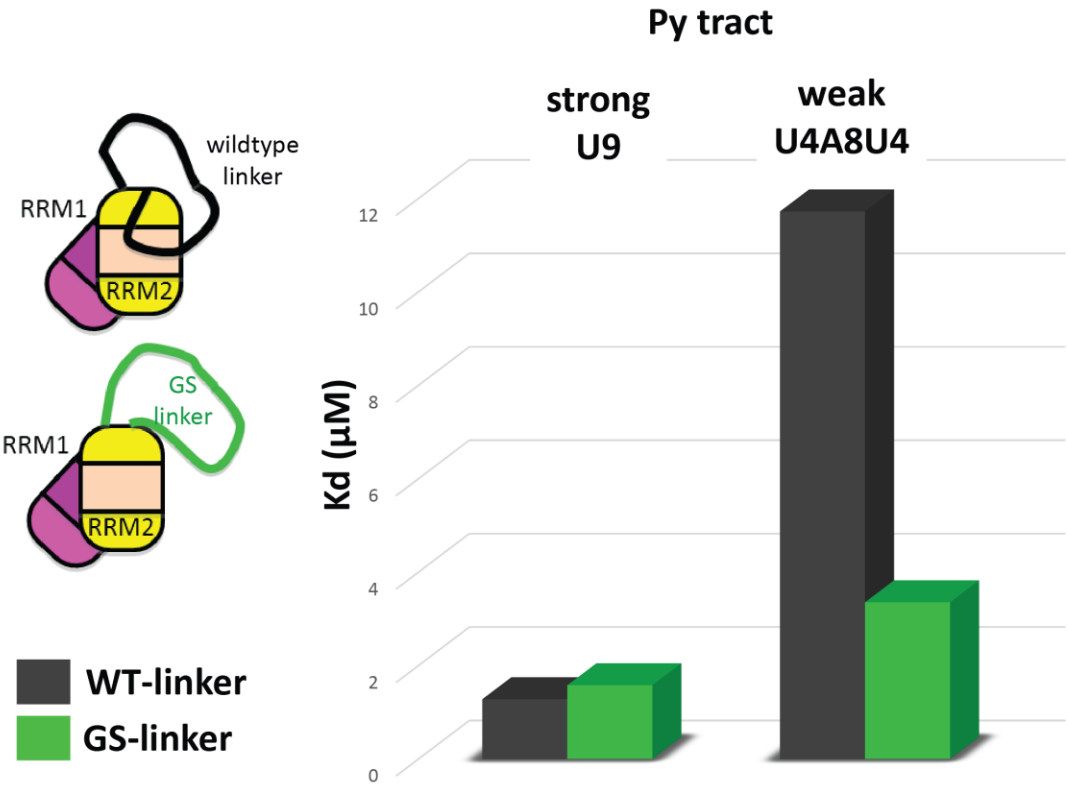

Figure 36: Bar Plot ITC titrations. *In green are plotted the dissociation constant (Kd) for the GS-linker construct from ITC titrations with U9 (left) and U4A8U4 (right), in gray the Kd values of the wild type RRM1,2 with the same RNAs are plotted.*

In order to verify that the results on the different affinities when RRM1-RRM2 is bound to different RNAs are not due to dimerization, (e.g. binding of two protein molecules to one RNA molecule) we performed SAXS experiments (Figure 37). This results demonstrated given the radius of gyration of the scattered particles that there is no dimerization upon RNA binding. It was confirmed, as it was shown previously by Mackereth et. al. that RRM1-RRM2 when bound to a strong Py-tract U9 RNA adopts a compact conformation in comparison to the free form (Figure 37 blue curves), and this result was also observed for the GS-linker construct (Figure 37 green curves) although in this case it seems that the RNA is bound in a slightly more extended conformation. However, in the case when RRM1-RRM2 is bound to a weaker Py-tract U4A8U4 RNA, the SAXS data showed that it exists as a more extended conformation or several extended conformations are present and that it might be a more dynamic complex even in the bound form. This result with the weak Py-tract appeared to be true also for the GS-linker construct, which shows similar SAXS curves compared to the wt-linker (Figure 37 red and orange curves).

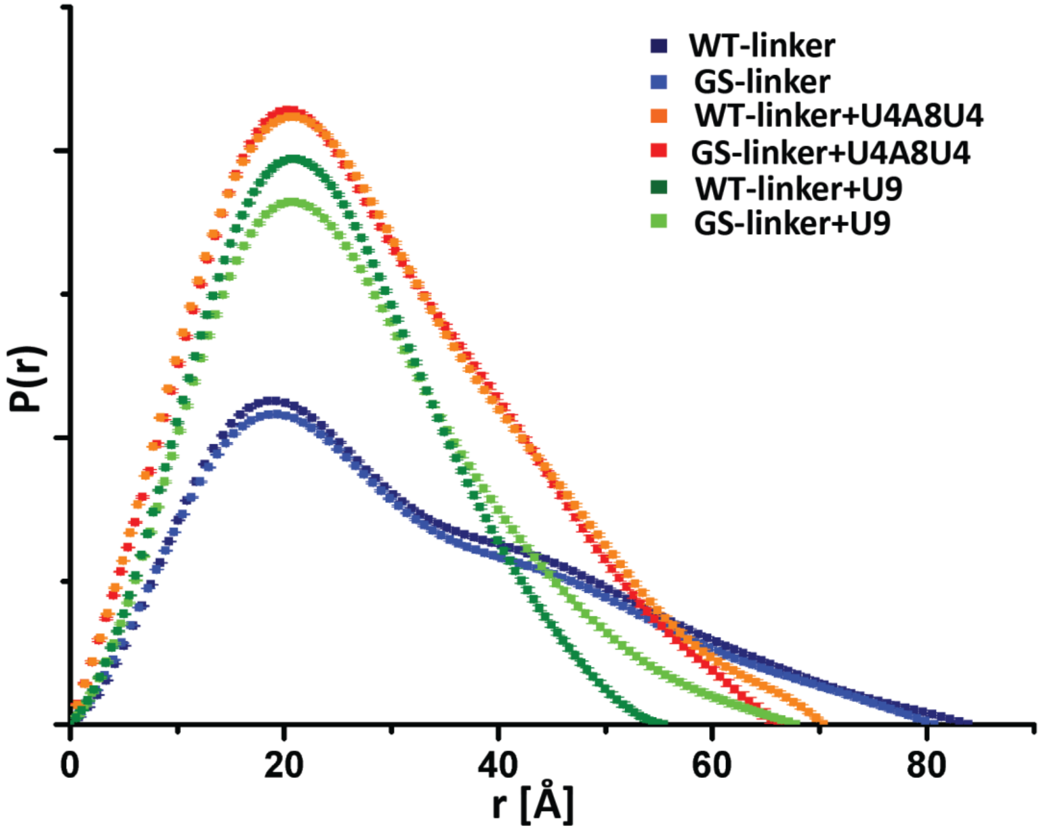

Figure 37: SAXS data of the WT and GS-linker RRM1,2. *Pair distance distribution plot of the free WT-RRM1,2 (dark blue), GS-linker RRM1,2 (light blue), WT-RRM1,2 + U9 (dark green), GS-linker RRM1,2 + U9 (light green), WT-RRM1,2 + U4A8U4 (orange) and GS-linker RRM1,2 + U4A8U4 (red).*

Because of the difference in affinity when the wild type linker RRM1-RRM2 vs. the GS-linker are bound to the weak Py-tract U4A8U4, we decided to investigate if there were kinetic differences between both constructs during binding RNA (Figure 38). This was investigated using the SPR technique using a Biacore X-100 machine. Although there were differences in the kinetic parameters, some of the values obtain for the association/dissociation rates were outside the range of the machine, meaning that the binding events are faster than the detection time. However, one can speculate from the values obtained (Figure 38), that the wild type linker construct has a faster association/dissociation rates in comparison with the GS-linker construct and this would go in agreement with the affinity for the weak Py-tract being higher for the GS-linker construct and lower affinity for the wild type linker construct.

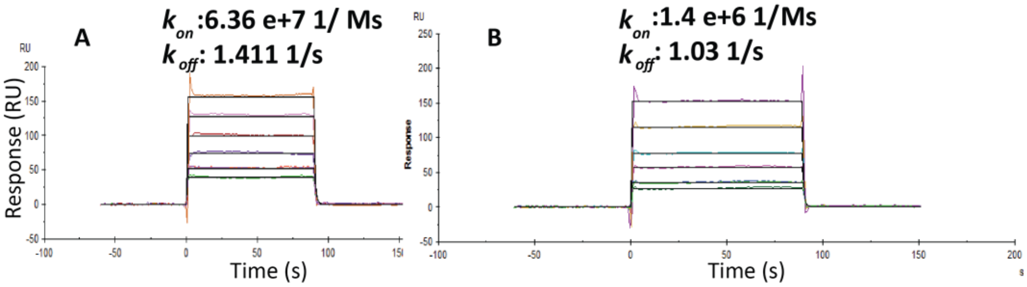

Figure 38: Biacore studies. *Sensogram of the biotinylated U4A8U4 (biotin-5-PEG-U4A8U4) when binding (A) nxRRM1,2 construct and (B) the nxRRM1,2- eGS-linker construct. Kinetic constants are indicated on top of the sensogram.*

5.4 Cooperative effect of the linker and domains for RNA recognition

From the latest findings it became interesting then to try to refine the previous structure of RRM1-RRM2 in order to show the interactions of the linker. This has been done in collaboration by a colleague in the lab Hyun-Seo Kang, who did an NOE-based refinement of the RRM1-RRM2 structure (Figure 39) finding for example multiple NOEs between the V254 at the C-terminal of the linker with A287 in RRM2. The structure shows that the C-terminal end of the linker packs against the central β-strands of RRM2, as it was suggest by the CSP pattern of the GS and WT linker RRM1,2 constructs.

Additionally from the hetNOE experiments one could notice that the residues at the C-terminal of the linker showed less flexibility (Figure 39).

Recently a crystal structure of RRM1-RRM2 bound to poli-U RNA (Agrawal et al., 2016) was published where Agrawal et al. showed that an extension in the N-terminal of RRM1-RRM2 is also involved in RNA recognition, plus some residues in the linker are shown to be in contact with the RNA. This finding motivated us to look into this N-terminal extended construct (nxRRM1-RRM2) by NMR as well. The chemical shift perturbations of nxRRM1-RRM2 versus the shorter one RRM1-RRM2 (Figure 40), presented an interesting pattern suggesting that the N-terminal extension might be interacting with a middle region inside the RRM1 domain and with the N-terminal part of the linker. The hetNOE of the URRM1-RRM2 construct also showed that, as well as the C-terminal part of the linker, the N-terminal part displayed reduced flexibility. To confirm if the N-terminal part of the linker is also interacting with the RRM1 domain, another GS-linker construct was designed on the extended nxRRM1-RRM2 (including two extra residues G253 and G254).

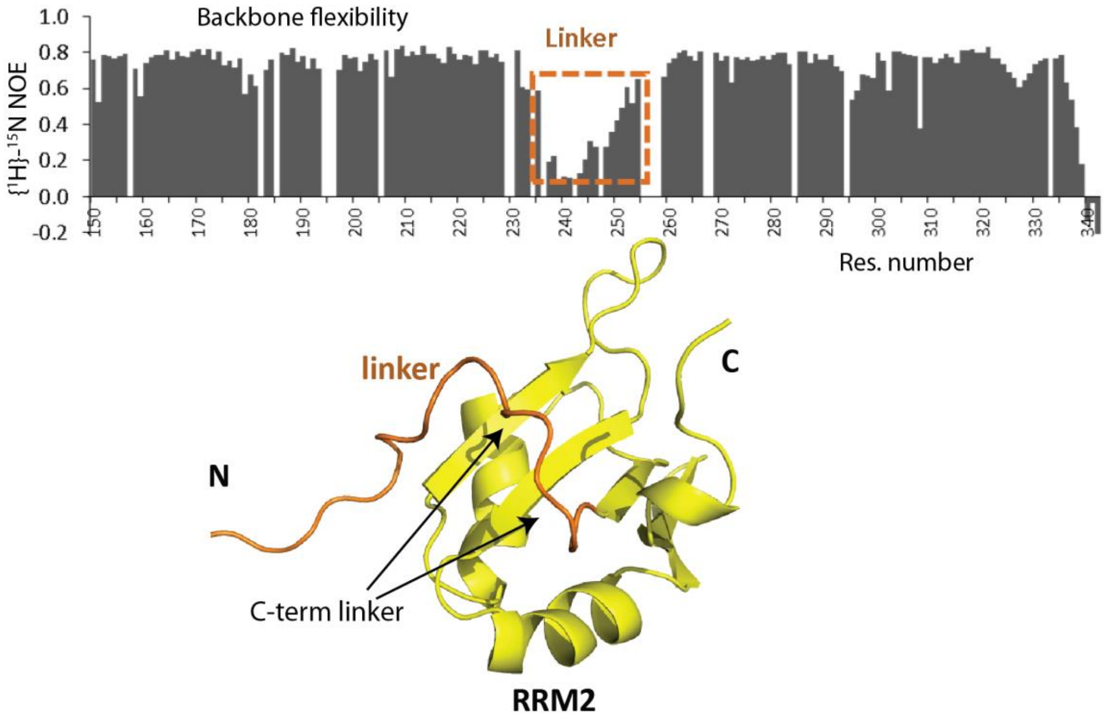

Figure 39: hetNOE and refined structure. *Plot of the heteronuclear NOE values (upper panal) of U2AF65 RRM1,2 and structure (lower panel)of the NOE-based refinement of the RRM1,2 structure (by Hyun-Seo Kang). The C-terminal of the linker presents higher hetNOE values in agreement with the structure showing some sections of the C-terminal of the linker (maroon) in contact with RRM2 (yellow).*

The chemical shift perturbations of this GS-linker nxRRM1-RRM2 vs. the nxRRM1-RRM2 with WT linker (Figure 41) showed how both domains are affected by the linker, especially the N-terminal extension of RRM1 affected the N-terminus of the linker. While as shown before RRM2 is affected by the C-terminus of the linker. The largest chemical shifts are still within the RRM2 domain. Even more interesting is that in all of these regions is at least one residue involved in RNA binding, affirming the theory of a regulatory role of the linker in RNA recognition and possibly RNA proofreading.

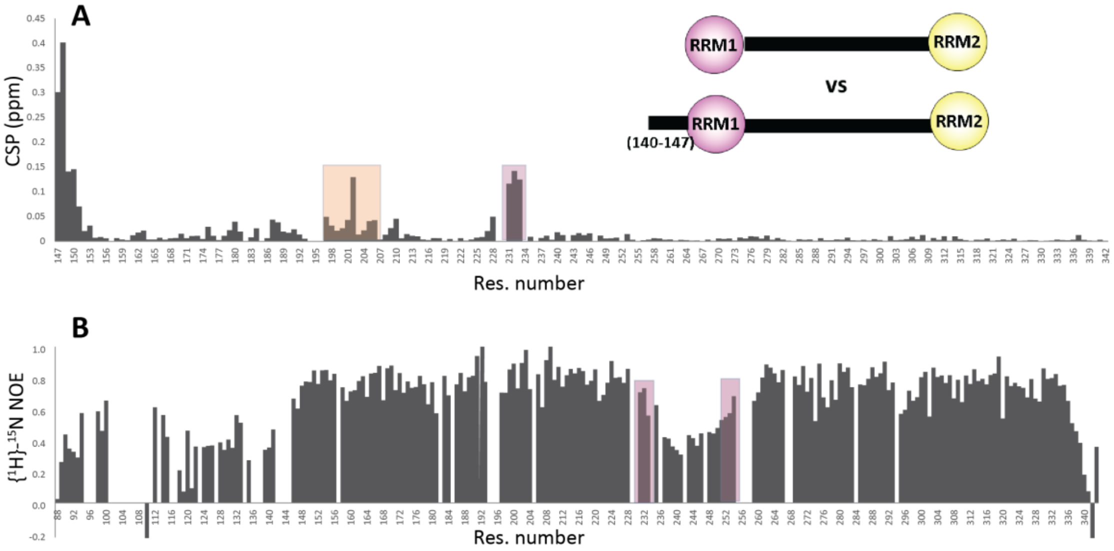

Figure 40: Interactions of the RRM1 N-terminal extension. *(A) Chemical shift perturbations of the RRM1,2 vs the nxRRM1,2 construct, in pink the residues of the N-terminal of the linker interacting with the N-terminal of RRM1 are highlighted. In guava color a region within RRM1 domain affected by its N-terminal extension is highlighted. (B) hetNOE values of the URRM1,2 construct that includes the N-terminal extension of RRM1. In pink the N-terminal and C-terminal regions of the linker show more rigidness in comparison with the rest of the flexible linker.*

These results are remarkable in the sense that they imply that in the free form there is a network between the linker and the two RNA binding domains that might serve to select the appropriate Py-tracts, and furthermore, there is a different conformational rearranging including the linker upon binding of RNAs of different affinity (weak and strong Py-tracts), as it is shown by SAXS data (Figure 42). The SAXS data shows that the extended construct nxRRM1,2 adopts extended conformations when bound to the weak Py-tract U4A8U4, in comparison to the more compact conformation when it is bound to the strong Py-tract U9, this is in agreement with the SAXS data of the RRM1,2 construct (Figure 37). These results might indicate that the protein complex with the weak Py-tract is a dynamic system or even that there is more than just one conformation involved, i.e. might be an ensemble as well as in the free form although probably with less conformations present, this was evident from a slight shoulder at the tale of the SAXS pair distribution function of nxRRM1,2 bound to U4A8U4 (Figure 42).

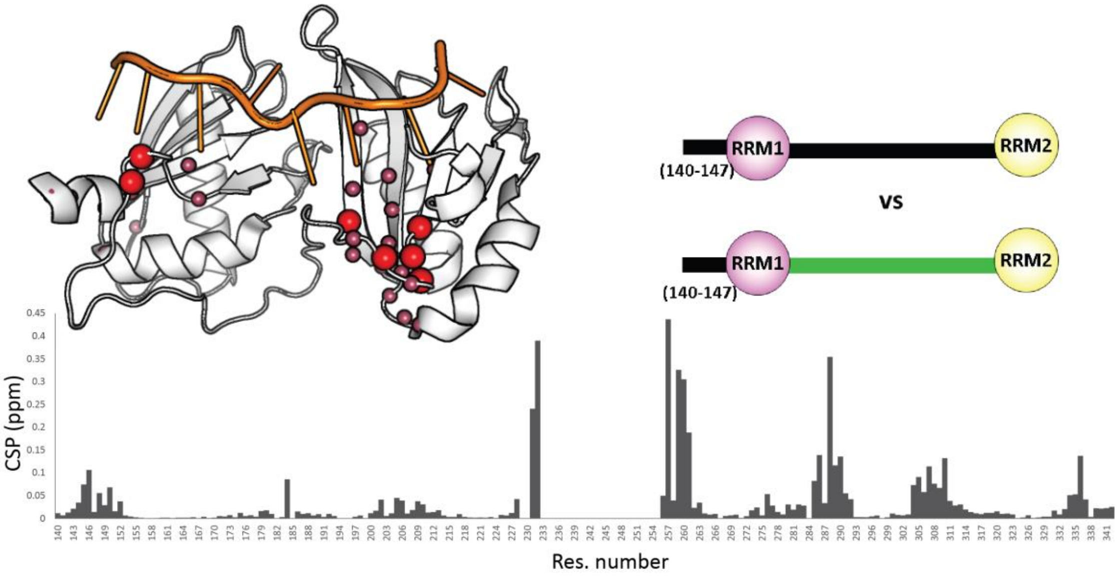

Figure 41: nxRRM1-RRM2 versus eGS-linker. *Chemical shift perturbations of the wild type nxRRM1,2 vs the eGS-linker nxRRM1,2 showed that different regions in both RRM1 and RRM2 domains are interacting with the linker, the largest shifts occurred in the RRM2 domain. The largest chemical shift differences are plot on the structure of the bound form (PDB 5EV1), the bigger red spheres represent the residues with CSP>0.188, the smaller pink spheres represent the residues with 0.033<CSP<0.188.*

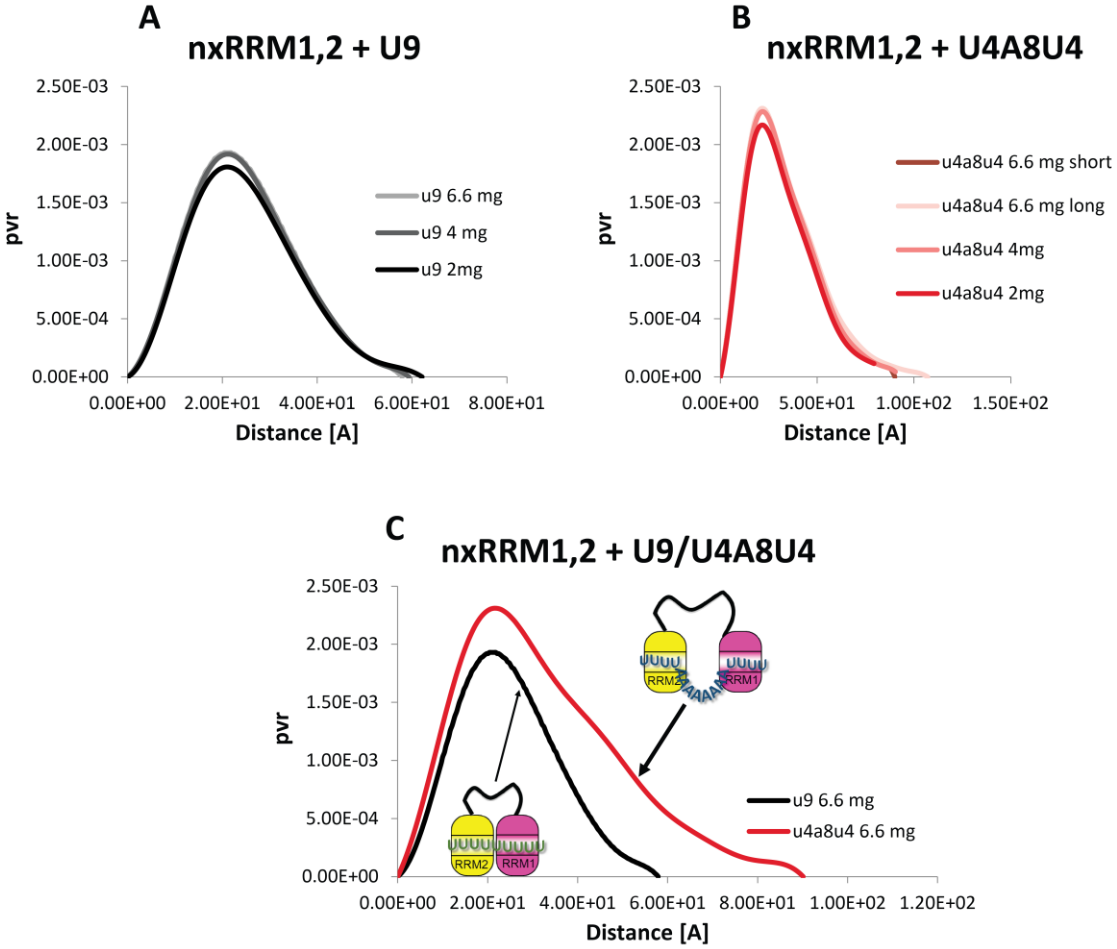

Figure 42: SAXS data of the nxRRM1,2 bound to U9 and U4A8U4. *(A) Pair distance distribution function of different concentrations of the nxRRM1,2 construct bound to U9. (B) Pair distance distribution function of different concentrations of the nxRRM1,2 construct bound to U4A8U4. (C) Comparison between the pair distance distribution function of the nxRRM1,2 construct bound to U9 versus bound to U4A8U4.*

To make evident the fine-tuning role of the linker during RNA binding of different strengths, there were chosen several residues (Figure 43) in both domains showing the different binding patterns of the wild type RRM1,2 and the GS-linker RRM1,2 . Particular differences are observed for the linker residue V254 in the two different protein constructs when titrating with the two different RNAs.

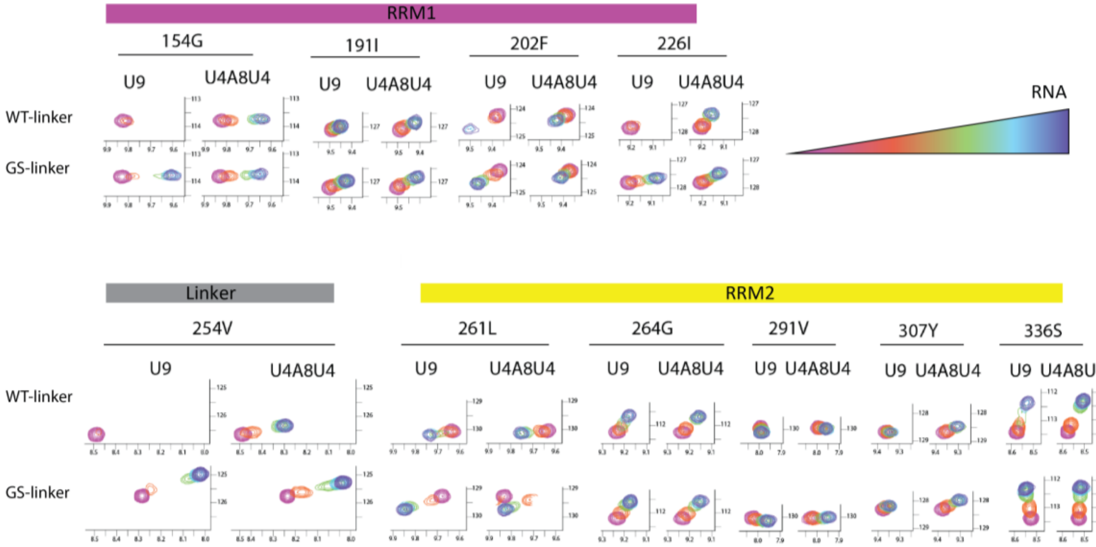

Figure 43: Different patterns on RNA binding. *Above are presented different RRM1 residues during titrations of the wild type RRM1,2 and the GS-linker RRM1,2 constructs with U9 and U4A8U4. Below are shown titrations of residues belonging to the linker and RRM2 domain on the wild type RRM1,2 and the GS-linker RRM1,2 constructs with U9 and U4A8U4.*

An specific difference can be seen in the case of the binding of U9 vs U4A8U4 to the wild type RRM1,2, in the case of the strong Py-tract it was only possible to observe the shift of the first titration point but after adding more U9 RNA there was line broadening of the peak beyond detection, on the other hand when the wild type was titrated with the weak Py-tract it was possible to follow each of the peaks until the final site. In the case of the titrations with the GS-linker construct, the linker residue V254 (which is right after the $(GGS)_N$ mutation) can be followed in both cases, when bound to U9 and U4A8U4 with slightly different patterns. Based on these results for the V254 residue which interacts with RRM2, and given the recent X-tal structure of the RNA-bound form(Agrawal et al., 2016), it is likely that V254 is interacting with the RNA, as well as residues in the N-terminal and C-terminal of the linker, and this interactions are different for the different Py-tracts (different strength Py-tracts). This supports the premise that the linker between RRM1 and RRM2 can act as proof-reading of different strength Py-tracts. Similar results on the titrations were observed for the residues G154 and I226 in RRM1, these residues are in regions that interact with the linker (Figure 41).

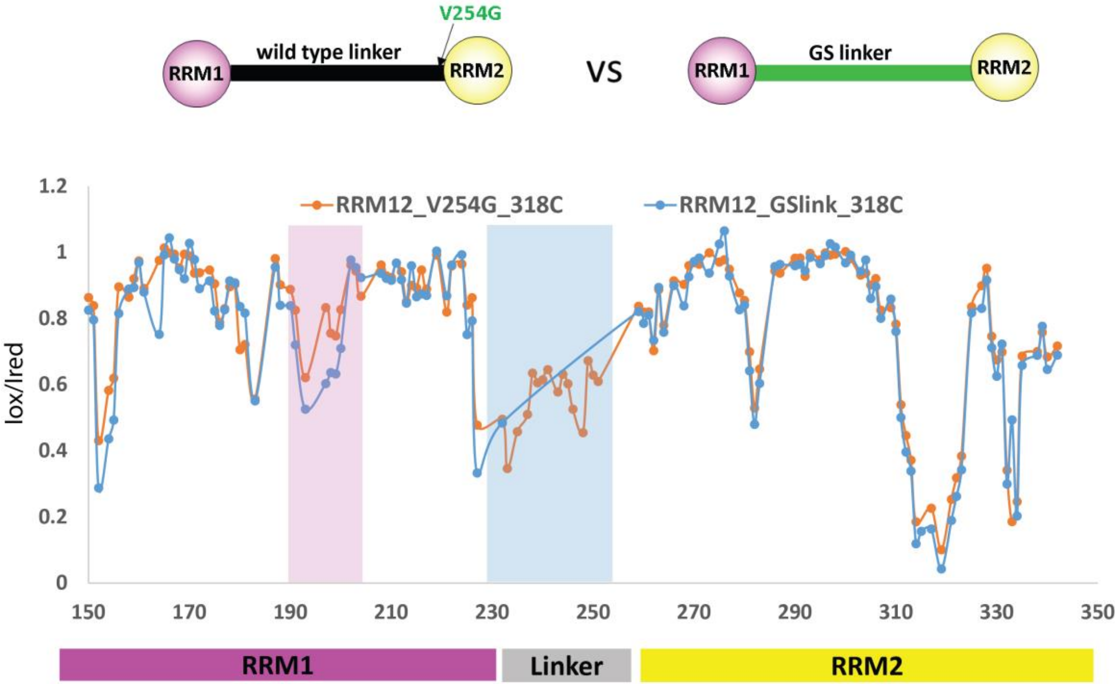

Figure 44: PREs WT-RRM1,2_254G and GS-linker RRM1,2. *The PREs were measured using the IPSL spin label attached to the position C318. Represented in orange and blue are the PRE values (Iox/*I_{red}*) of the V254G RRM1,2 and the GS linker RRM1,2 respectively.*

Additional PRE experiments with the spin label at C318 were done to compare the effect on interdomain contacts due to the linker contacts with RRM2 (Figure 44). It is presented the comparison of the PRE between the GS-linker construct and the RRM1,2_V254G mutant, where the residue V254 was one of the key residues in the linker that folds onto the RRM2 domain. From this comparison it is observed that when mutated the V254 residue to a glycine it disrupt the linker interactions with RRM2, showing more distant interdomain contacts (separation of the domains), even more than the initial GS-linker mutation of the linker (232-252) (Figure 44 pink highlight). This result was also observed for the nxRRM1,2 construct (Figure 45), in this case it was possible to observe the effect of the V254G mutation directly on the linker region (Figure 45 blue highlight), in the N-terminal extension of RRM1 and in the interdomain region (Figure 45 pink highlight). It was shown an increment on the PRE effect in the linker region (particularly in the C-terminal), when the mutation V254G is present, showing that probably the linker becomes more flexible in that region and so it can go closer to the spin label located at C318. The N-terminal of RRM1 showed significant PRE effect in similar magnitude as in the N-terminal of the linker, this goes in agreement with the results showing that the N-terminal of RRM1 bends in contact with the N-terminal of the linker (Figure 40). And finally an important result (Figure 45 second pink highlight) showed that the two domains are getting more separated upon the disruption of the linker-RRM2 interactions.

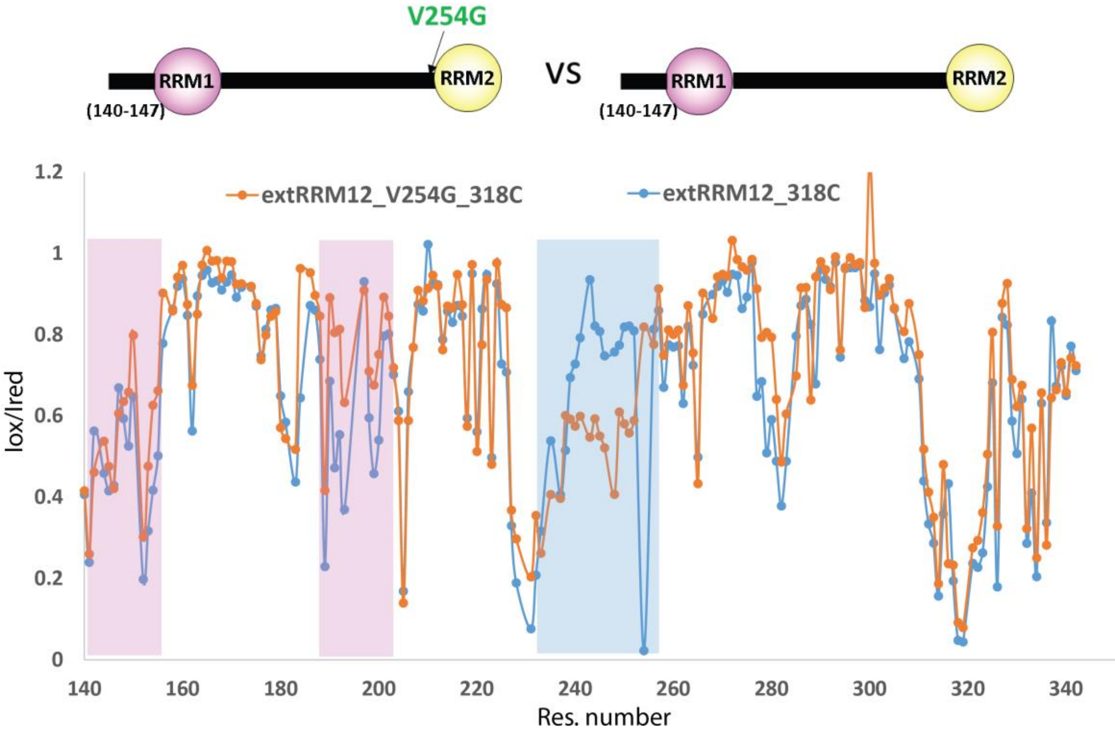

Figure 45: PREs nxRRM12 vs nxRRM12_254G. *Comparison of the PRE values (Iox/Ired) with the spin label at the C318 position in: the nxRRM1,2 construct (blue) and the nxRRM1,2 with the V254G mutation (orange).*

PRE experiments on the RNA-bound to both Py-tracts were recorded when the spin label was attached to the position C155, however these data was noisy and appropriate interpretations could not be properly given.

5.5 Chemical Exchange on the free and RNA-bound states

Preliminary relaxation dispersion studies were performed on the tandem domain RRM1,2 and nxRRM1,2 to study the chemical exchange going on due to the presence of intermediate states.

Initially, we looked into the line width (exchange broadening) differences between residues on the single domains RRM1 and RRM2 versus the line width (due to exchange broadening) of residues in the tandem domain RRM1,2 (Figure 46). The comparison between several residues, showed that some of them present little or no difference in line shape when they belong to either single domains or the tandem domain. However, several other residues presented small and large differences in line width (intensity changes), this finding indicates that there are excited states or intermediate states present in the tandem domain RRM1,2 .

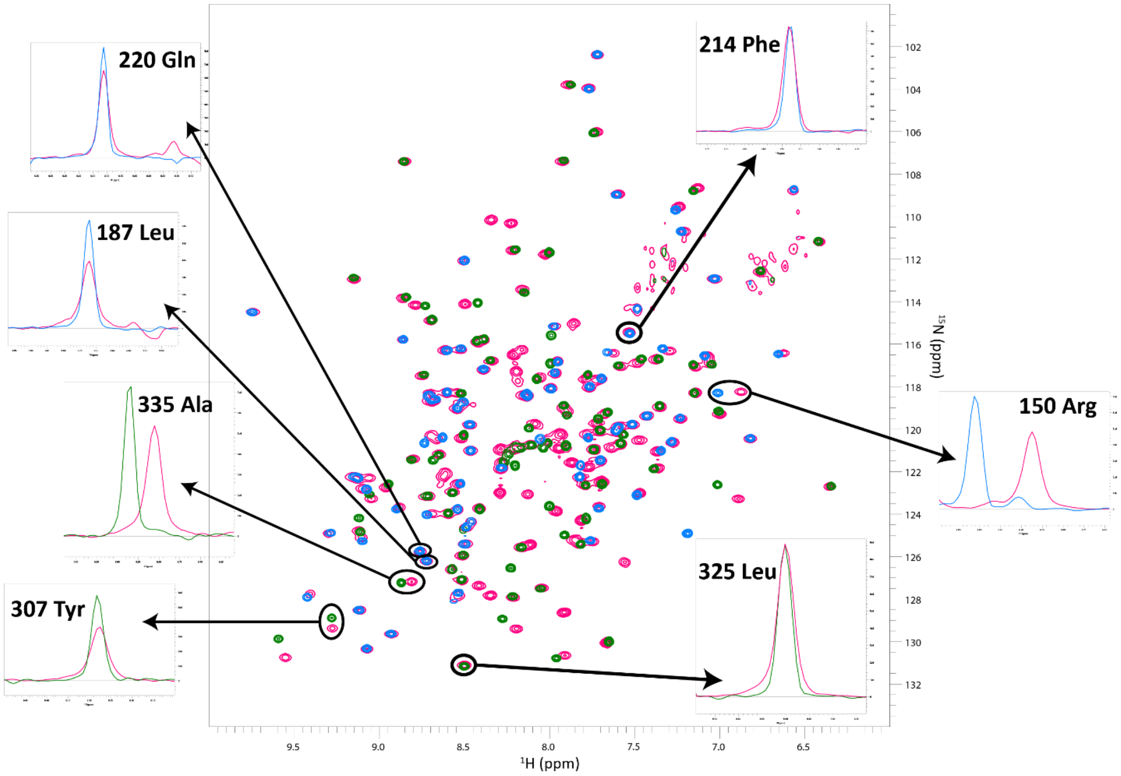

Figure 46: Line width differences between single domains vs tandem RRM1,2 domains. *Overlapped spectrum of the single domains RRM1 (blue) and RRM2 (green) and the tandem RRM1,2 (pink) domains. On the side there are the projections showing the line width of residues in RRM1 or RRM2 when in the single domains versus when belonging to the tandem domain.*

An intensity distribution versus residue number of the single domains and the tandem domain is presented in Figure 47, there are shown residues in the surface representation that display larger line width in both, closed and open conformations. Despite some of the residues are in the interface between the domains, not all the residues with decreased intensity correspond to the closed conformation interface, which is in agreement with the ensemble description (chapter 4).

These data were the initial incentive to study the chemical exchange in the system using relaxation dispersion experiments.

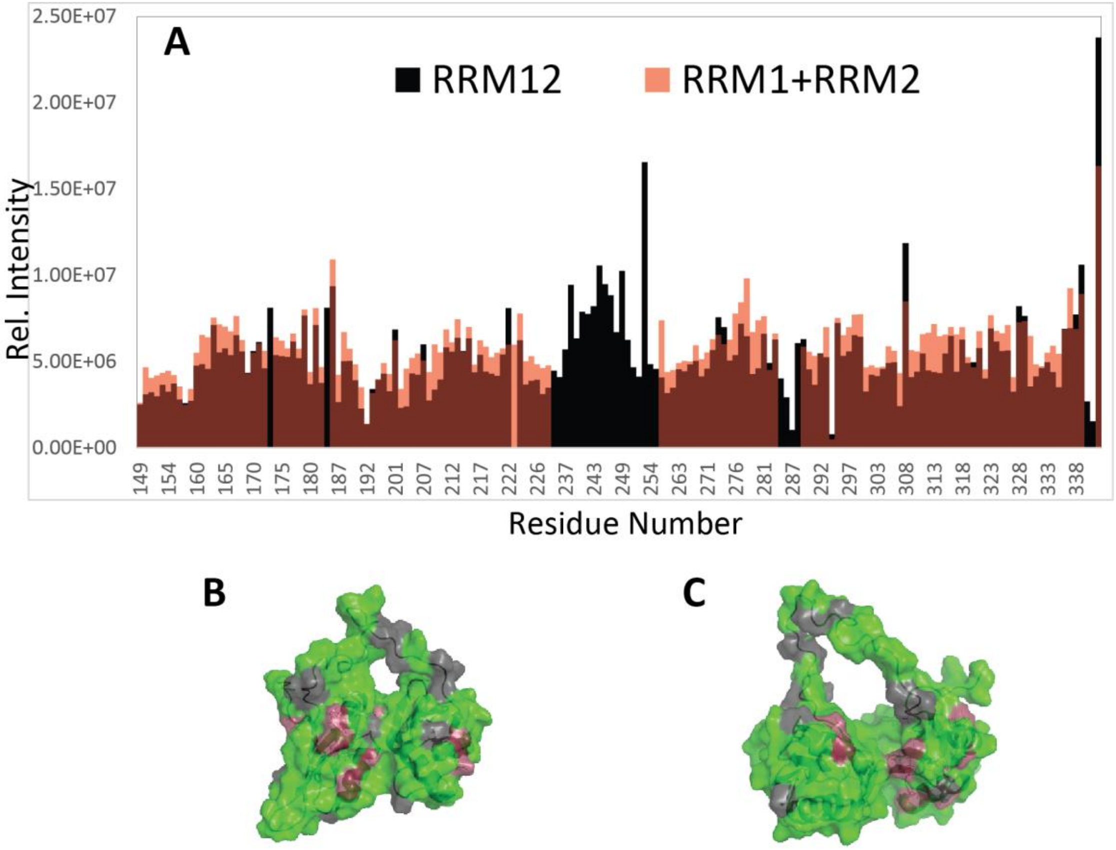

Figure 47: Intensity distribution single domains vs tandem domain. *(A) Intensities distribution on the tandem RRM12 domains (blue) and the single RRM1 and RRM2 (red). (B) Free RRM1,2. In red are colored the residues with decreased intensity (up to 40% decrease) and gray are unassigned residues. K286 and A287 are on the interface in this close conformation. (C) Bound RRM12. In red are colored the residues with decreased intensity and gray are unassigned residues. L279 and T280 are on the interface of this open conformation.*

One important factor to study before performing the relaxation dispersion experiments was to check if there were concentration dependent effects on that can cause chemical exchange. For this we calculated the correlation time at different concentrations (Figure 48), and it was observed that the correlation time increased with concentration. This finding might suggest that there is nonspecific partial aggregation of RRM1,2 and this can lead to chemical exchange, however there were no chemical shifts between the spectrum at different concentrations. Hence, the following relaxation dispersion experiments were recorded using concentrations between 300 µM to 500 µM maximum.

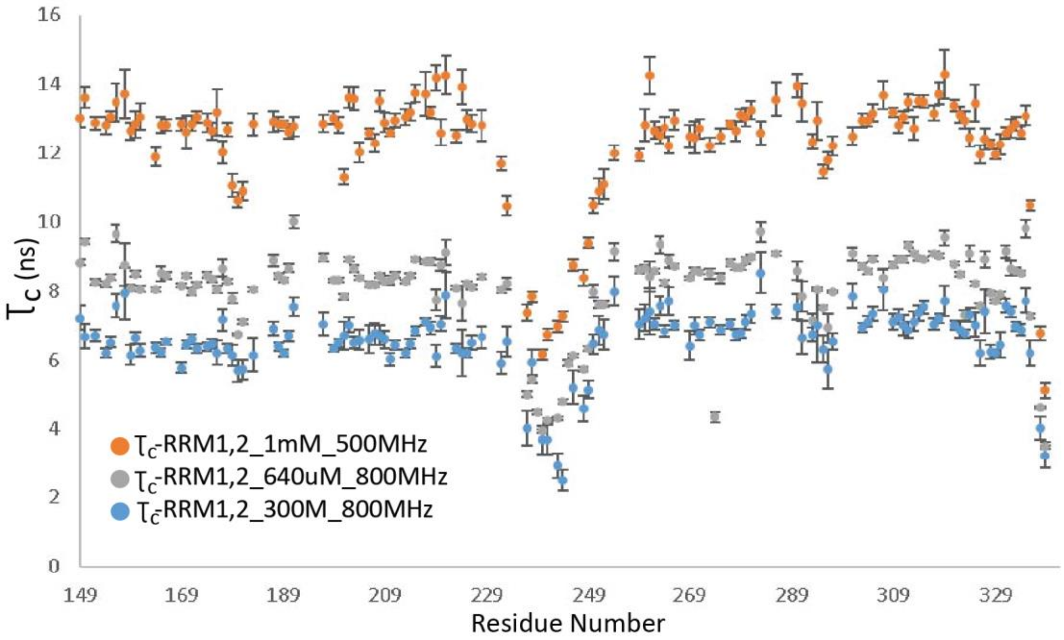

Figure 48: τ_c at different concentrations. *In orange are presented the τ_c values for 1 mM concentration of RRM1,2 protein, in gray the τ_c for 640 µM concentration and in blue the τ_c values for 300 µM protein concentration.*

The relaxation dispersion experiments were done by my colleague Sam Asami. Initially the experiments were done in the RNA-free U2AF65 RRM1,2 at different temperatures (298 K, 288 K, 278 K) and different fields (800 MHz, 500 MHz) (Figure 49). The results of the free RRM1,2 showed dispersion of different residues all over both domains and linker at different temperatures, these differences are expected given the different energy landscape at different temperatures, however there wasn't a clear pattern or region with strong chemical exchange that could provide explanation regarding the previous calculated ensemble or the linker interaction with RRM2.

Then we decided to measure relaxation dispersion on the compact U9-RNA-bound U2AF65 in our attempt of having a system with reduced dispersion i.e. that presents clearer regions undergoing chemical exchange given that this is a more rigid system. A thorough titration of the RNA into the protein was done in order to have completely saturated protein, i.e. RNA was added until no changes in intensities of any residue were observed. However, the results of the relaxation dispersion experiments (Figure 50) showed a much larger chemical exchange all over the two domains and the linker. This result was surprising since the RNA-bound U2AF65 was expected to have less dispersion than the free form of the protein. On the other hand, it has been studied (Schneider et al., 2015) for disordered proteins that it is necessary to add up to more than ten times the ligand concentration in order to fully saturate the excited states as well. This result for RNA-bound RRM1,2 suggests that the excited states or intermediate states of the bound form were not saturated at the ligand concentrations at which the ground state was fully saturated.

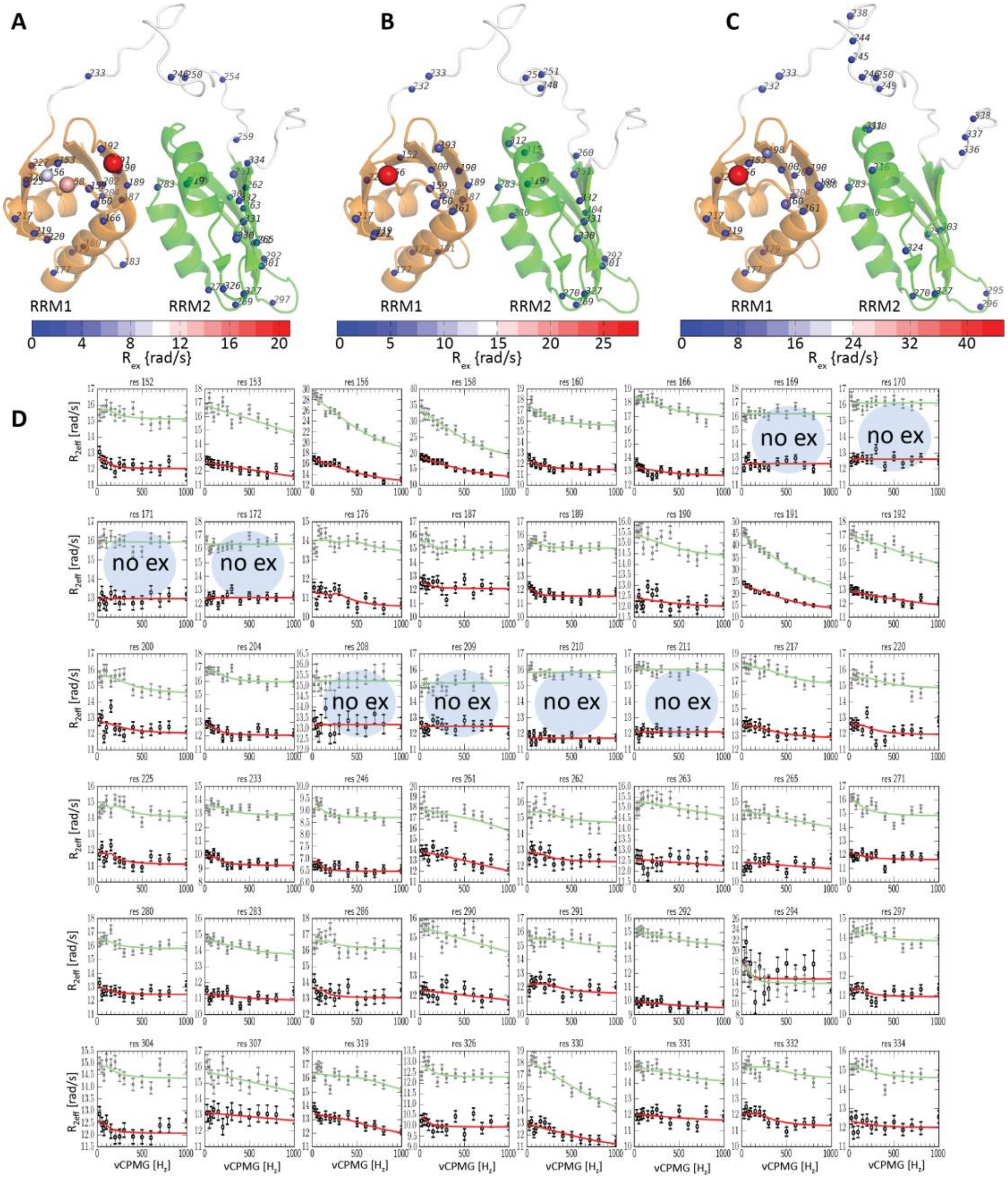

Figure 49: Chemical exchange on the free RRM1,2. *(by Sam Asami)Residues showing different chemical exchange with R_{ex} values varying according to the color code, (A) at 298 K, (B) 288 K and (C) at 278 K. (D) Relaxation curves (R_{ex} vs ν_{cpmg}) at 298 K of several residues at two different fields, red 500 MHz and green 800 MHz. Highlighted are residues that don't present chemical exchange, i.e. flat lines.*

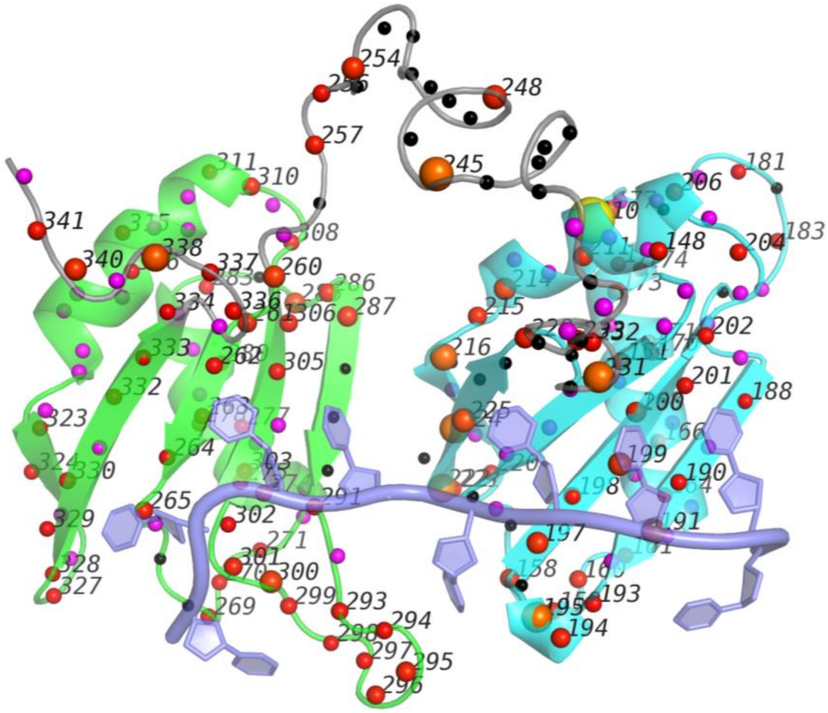
254
255
248
257
245
311
310
181
206
183
341
340
338
337
260
308
148
204
286
287
336
306
334
215
216
232
231
202
333
262
305
201
188
332
200
323
225
199
190
324
330
264
303
291
198
265
302
329
191
197
271
328
327
301
300
299
293
160
193
158
269
294
195
194
298
297
295
296

Figure 50: Chemical exchange on the RNA-bound U2AF65. *(by Sam Asami) nxRRM1,2 bound to U9 RNA, the spheres represent the residues undergoing chemical exchange in the bound state. The ground state was fully saturated with RNA. The observation of exchange all over the protein might indicate not saturation of the invisible (excited) states.*

CHAPTER 6:

Role of the U2AF35 subunit within the U2AF heterodimer

6.1 FRET studies confirm dynamic nature of RRM1,2

In collaboration with the group of Don Lamb particularly with Lena Voithenberg in LMU, spFRET (Single pair Föster resonance energy transfer) experiments were performed on U2AF to further investigate the role of conformational dynamics in the free and bound form(Voithenberg et al., 2016). For these experiments two cysteine mutations were done in the protein, one in RRM1 and other in RRM2 of U2AF65. The mutations were designed such that they are in close proximity when RRM1,2 adopts a closed conformation and the mutations are in maximal distance when the tandem domains adopt an open conformation (Figure 51A). Donor and acceptor fluorophores were stochastically attached to these cysteine mutations.

The spFRET efficiency histograms were calculated for the free RRM1,2, U9-bound and U4A8U4-bound RRM1,2 (Figure 51B, C). The free form of RRM1,2 was found to populate a high FRET state (Figure 51B, gray) with a mean FRET efficiency of around 0.78. From the analysis of the lifetime information of the donor fluorophore it was determined that this FRET efficiency comes from an average value obtained from a highly dynamic population of molecules (Figure 51C, left) where the intensity-determined FRET efficiency was plot versus donor lifetime showing that all molecules deviate from the static FRET line.

When RRM1,2 is bound to U9, the equilibrium between the open and closed conformation shifted substantially such that the complex spends ~90% of the time in the open conformation (the population lies on the static FRET line) (Figure 51C, middle), which resulted in a shift of the apparent FRET efficiency to an average value of ~0.5 (Figure 51B, red). Finally when RRM1,2 is bound to the weaker Py-tract U4A8U4, the apparent FRET efficiency is ~0.63 (Figure 51B, blue) and the population slightly deviates from the static FRET line (Figure 51C, right).

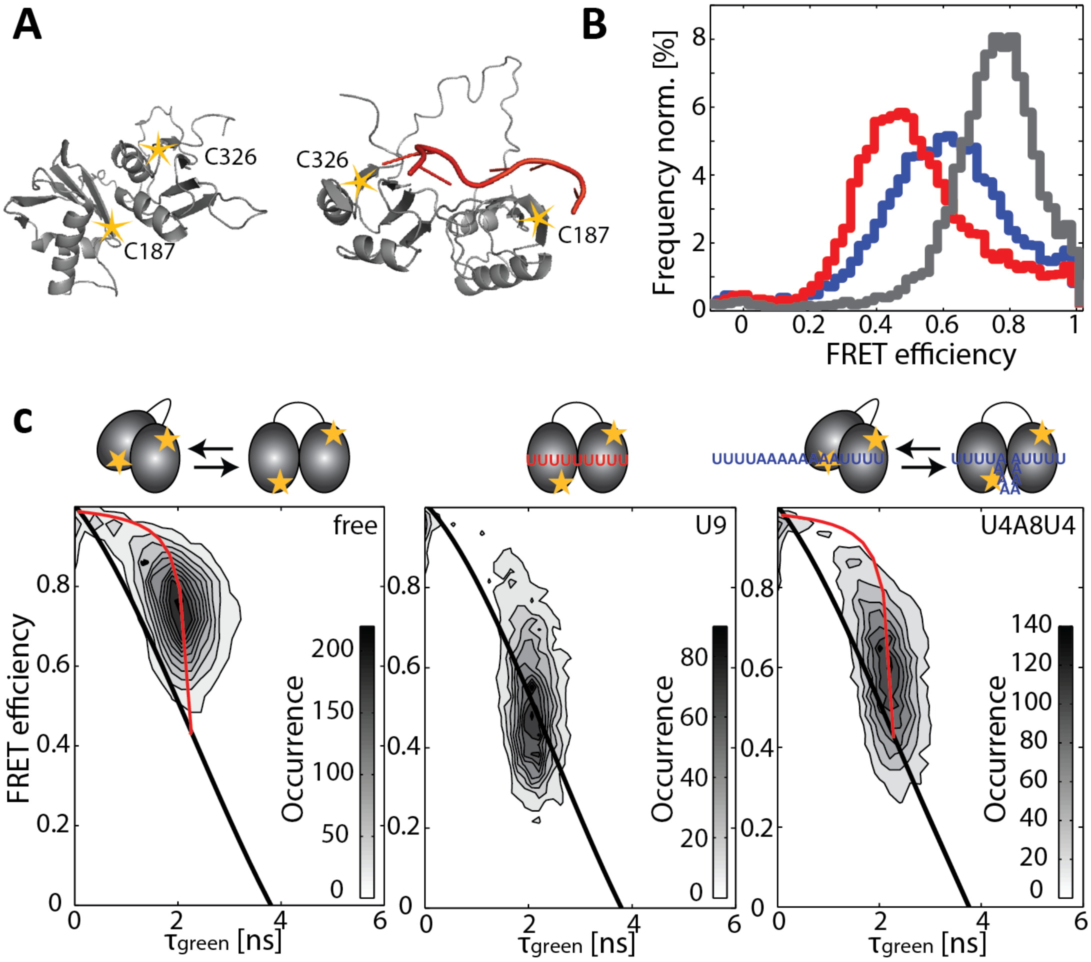

*Figure 51: **Conformation and dynamics of RRM1,2 in solution.** (A) Closed (left) and open, RNA-bound (right) structure of U2AF65(RRM1,2), labeling sites shown as stars. (B) spFRET efficiency distribution histogram of RRM1,2-Atto532-Alexa647 in the absence of RNA (gray), in the presence of U9 (red), or in the presence of U4A8U4-RNA (blue). (C) Histogram of spFRET efficiency as a function of donor lifetime. Populations of static molecules are described by the polynomial static FRET line (black line), while molecules undergoing conformational dynamics on the timescale of µs-ms deviate from this line (dynamic FRET curve: red line). Representations of the open and closed conformations of the molecules are displayed schematically as a simple interpretation of the histograms.*

6.2 Role of U2AF35 during Py-tract recognition by U2AF65

It was shown by Yoshida et al. (H. Yoshida et al., 2015) that the UHM domain of U2AF35 is not directly involved in RNA recognition but instead the two zinc fingers of U2AF35 contact the AG-dinucleotide at the 3' splicing site. Here we investigate the role of the UHM domain of U2AF35 during the Py-tract recognition by the ULM-RRM1,2 (URRM1,2) of U2AF65. Thus, we investigated the minimal U2AF heterodimer (URRM1,2/U2AF35(UHM)) when binding different Py-tracts sequences (Figure 52B). Remarkably, even in the free form, the presence of UHM(U2AF35) shifts the population of RRM1,2 towards an open conformation, which means that in the minimal heterodimer U2AF35 UHM domain enhances the open state of U2AF65 RRM1,2 domains in comparison to the open state of RRM1,2 alone (Figure 52A, B, gray, C). When the minimal heterodimer was bound to the strong Py-tract U9 RNA, there was a shift to the open conformation similar to the shift in FRET efficiency of the RRM1,2 alone (Figure 52A,B, red). However, the most interesting result comes when the minimal heterodimer is bound to a weaker Py-tract U4A8U4, since it was observed a shift of the FRET efficiency to the same value as for the strong Py-tract (Figure 52A,B blue). This means that the presence of UHM(U2AF35) has a large effect in shifting RRM1,2 towards the open state when the weak Py-tract is bound. In all the cases: bound to strong or weak Py-tract or free, the minimal heterodimer (URRM1,2/U2AF35(UHM)) is more likely to be found towards more open state when compared with the RRM1,2. Additionally, contributions of further RNA sequences containing the AG-dinucleotide were studied (Figure 52A, B, right) and the differences between the minimal heterodimer and RRM1,2 were similar to those when studied the U9 or U4A8U4 RNA sequences.

To find out more about the molecular mechanisms possibly involved in the interactions between the two U2AF subunits, PRE experiments were recorded (by Katia Zanier) in the URRM1,2/U2AF35(UHM) heterodimer (Figure 53) when the spin label was attached to the UHM domain, the region in URRM1,2 affected by this spin label, corresponds to its regions in proximity to the U2AF35(UHM) (Figure 53, pink highlight). This regions corresponds to the ULM motive which is required for the binding to UHM(U2AF35) and additional there is an interface with RRM1, which means that the UHM of U2AF35 is also in contact with RRM1 of U2AF65, and this interface mediates the population shift to the open state of RRM1,2.

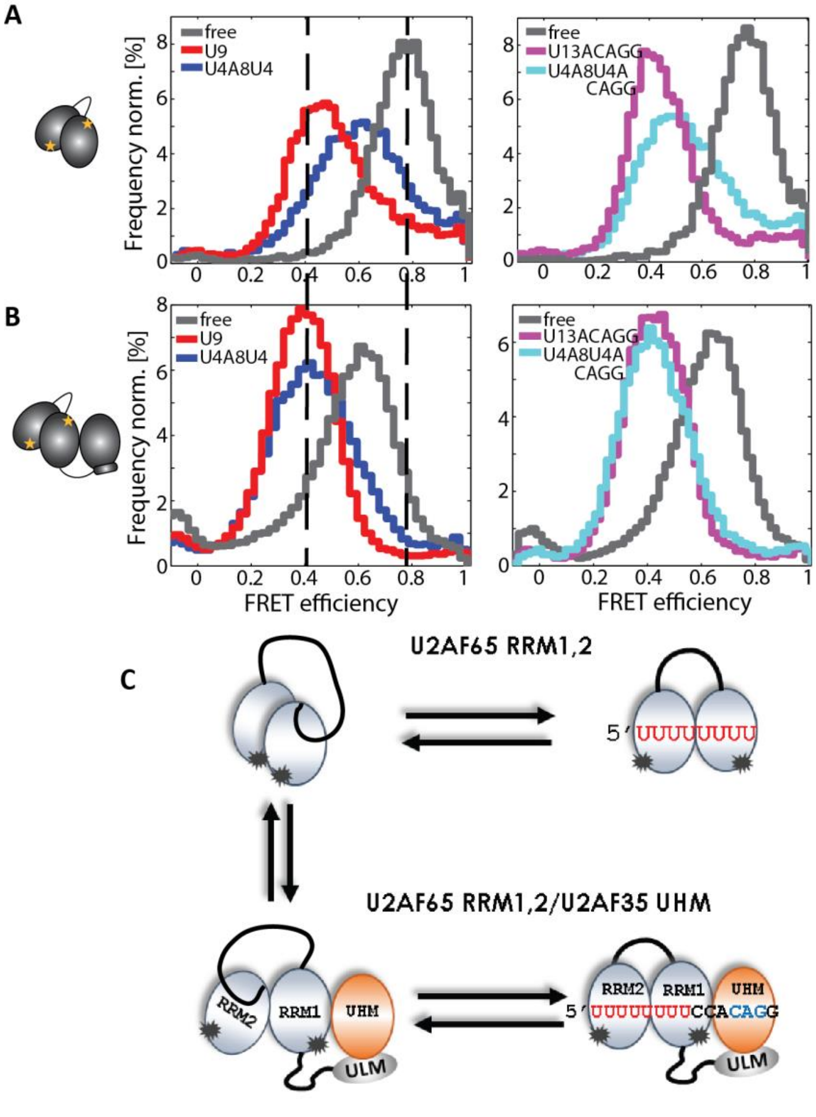

Figure 52: spFRET experiments of RRM1,2 and URRM1,2/U2AF35(UHM). *Comparison between the conformations of the dual-labeled RRM1,2, and*

URRM1,2/U2AF35(UHM).(A) Schematic representation of the minimal RRM1,2. spFRET efficiency histograms of RRM1,2 in the free form (gray) or in the presence of U9 (red), U4A8U4 (blue), U13ACAGG (pink), and U4A8U4ACAGG (cyan). (B) Schematic representation of URRM1,2/U2AF35(UHM) and spFRET efficiency histograms of the heterodimer bound to the same RNA sequences as in (A). The dashed black lines serve as guides to the eye to show the open and closed conformations. (C) Schematic representation of how the presence of the U2AF35 shifts the RRM1,2 equilibrium towards more open conformation.

A model is proposed to explain the ongoing dynamics of U2AF during RNA binding (Figure 54). Gathering the previous results RRM1,2 free in solution are an assemble of compact and non-compact states.

The encounter-like interdomain contacts are driven by linker-RRM2 interactions and electrostatic contacts between the two domains. When binding a strong Py-tract (U9 RNA) RRM1,2 adopts a preferentially open conformation (Figure 54A), however when binding a weak Py-tract (U4A8U4) there is more dynamics going on in the bound state, presenting an equilibrium between closer states (when bound to just one of the domains) and open state (when the RNA is bound to both domains) (Figure 54B, Figure 51C right).

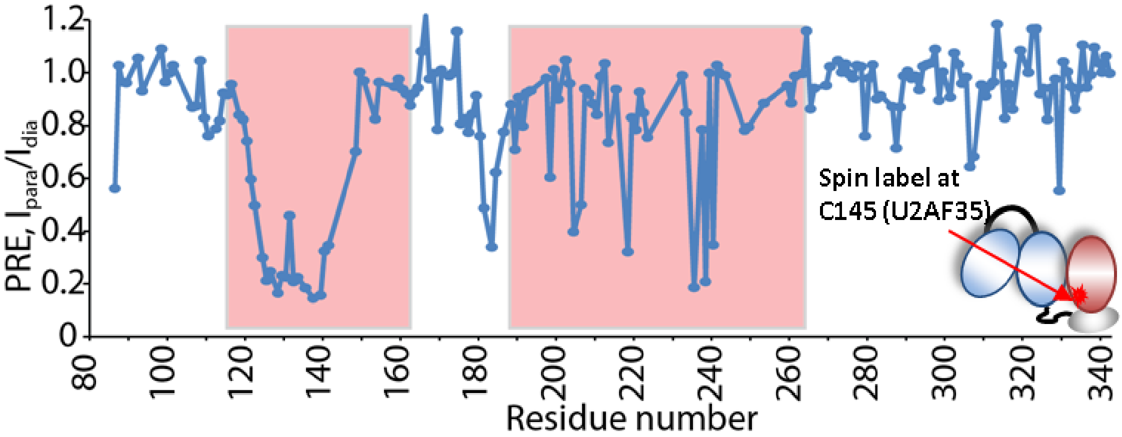

Figure 53: PRE experiment on the U2AF heterodimer. *I_{para}/I_{dia} values versus residue number of the minimal U2AF heterodimer when the spin label is attached to the UHM(U2AF35) at the position 145. Presented is the plot for the URRM1,2(U2AF65) PRE values, showing the residues nearby (pink highlight)) to the attachment site in U2AF35.*

When the UHM (U2AF35) is present to form the minimal U2AF heterodimer (URRM1,2(U2AF65)/UHM(U2AF35)), there is a pre-existing conformations towards the open state, such that when the weak Py-tract is bound the RRM1,2 domains adopt the full open conformation in the similar way as for the strong Py-tract (Figure 52B, Figure 54C). Given that, we can say that the small subunit U2AF35 plays an important role during the binding of weak Py-tracts and might lead to an explanation of how U2AF is capable of recognizing so many different Py-tract sequences present in the 3’ splice site of the eukaryotes pre-mRNA.

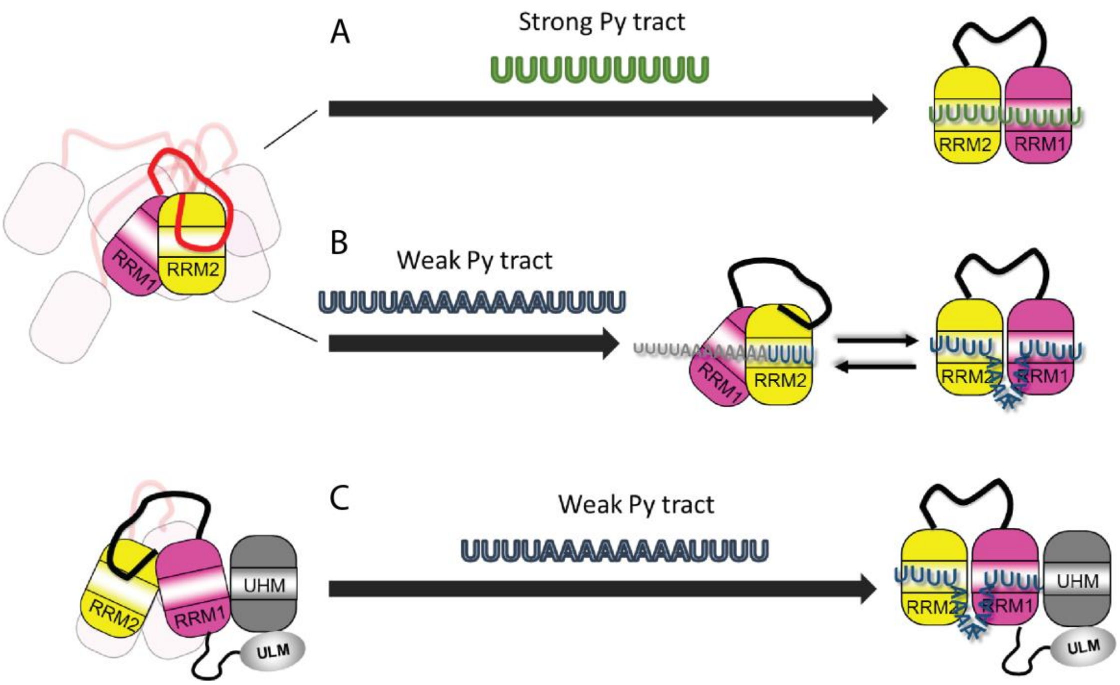
A
Strong Py tract
UUUUUUUUU
UUUUUUUUU
RRM2
RRM1
RRM1
RRM2
B
Weak Py tract
UUUUAAAAAAAAUUUU
UUUUAAAAAAAAUUUU
RRM1
RRM2
UUUU
UUUU
RRM2
RRM1
C
Weak Py tract
UUUUAAAAAAAAUUUU
RRM2
RRM1
UHM
ULM
UUUU
UUUU
RRM2
RRM1
UHM
ULM

Figure 54: Schematic model on the role of U2AF35 during RNA binding. *(A) RRM1,2 binds a strong Py-tract (U9) adopting a full open conformation, while when (B) binding a weak Py-tract (U4A8U4) there is a dynamic equilibrium in the bound state between an open state and a partially closed one. (C) in the presence of the UHM(U2AF35), URRM1,2(U2AF65) is shifted towards a more open (active) conformation and when binding a weak Py-tract the bound form goes to full open conformation in similar way as for the strong Py-tract.*

6.3 Somatic mutations do not affect the structure or RNA binding of the U2AF heterodimer

We investigated the somatic mutations involved in myelodysplastic syndromes (MDS). These mutations were presented by Yoshida et al. (K. Yoshida et al., 2011) which belong to U2AF65: L187V and M144I and are thought to interfere with the U2AF heterodimer formation. First of all to investigate if there were structural differences we compared the NMR ^{1}H,^{15}N-HSQC spectrum of the wild type heterodimer versus the heterodimers containing the M144I or L187V mutations, i.e. URRM1,2-M144I/U2AF35(UHM) and URRM1,2-L187V/U2AF35(UHM) (Figure 55). Apart from local chemical shift differences in the spectrum (Figure 55, left), one can conclude that the heterodimer is formed even on the presence of the mutations.

To study the effect of the MDS mutations on RNA binding we did ITC titrations on the wild type and mutated heterodimers with both, weak and strong Py-tracts (Figure 56).

The K_D values are almost the same for the wild type and mutated heterodimers in the case when is bound to the U13ACAGG strong Py-tract (Figure 56A) and also in the case of the U4A8U4ACAGG weak Py-tract (Figure 56B), additionally the signature plots of the different RNA titrations show that the binding is carried out with the similar thermodynamic characteristics as for the binding to the wild type heterodimer (Figure 56, bar plots). Given these results we can say that the presented mutations don’t interfere with RNA binding of the U2AF heterodimer.

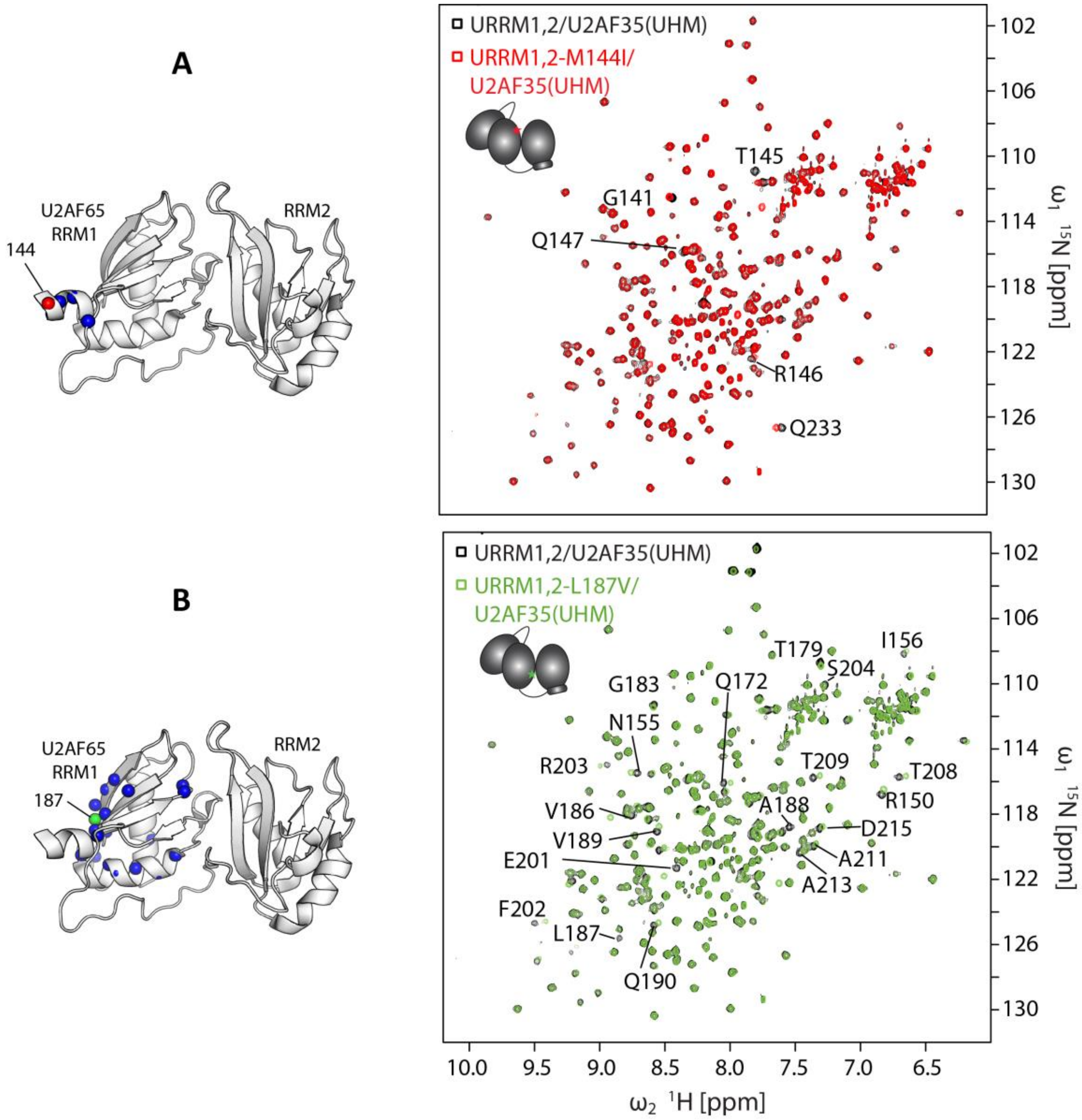

***Figure 55: NMR study on the U2AF heterodimer when containing the MDS somatic mutations**. ^1H,^{15}N-HSQC spectrum of U2AF heterodimers formed with U2AF35(UHM) and URRM1,2 harboring the somatic point mutations M144I and L187V. URRM1,2/U2AF35(UHM) (black) versus (A) URRM1,2-M144I/U2AF35(UHM) (red), and (B) URRM1,2-L187V/U2AF35(UHM) (green). Notable chemical shift changes in U2AF65 observed for the heterodimers harboring the mutations are indicated as blue spheres on the RNA-bound structure (PDB 5EV1).*

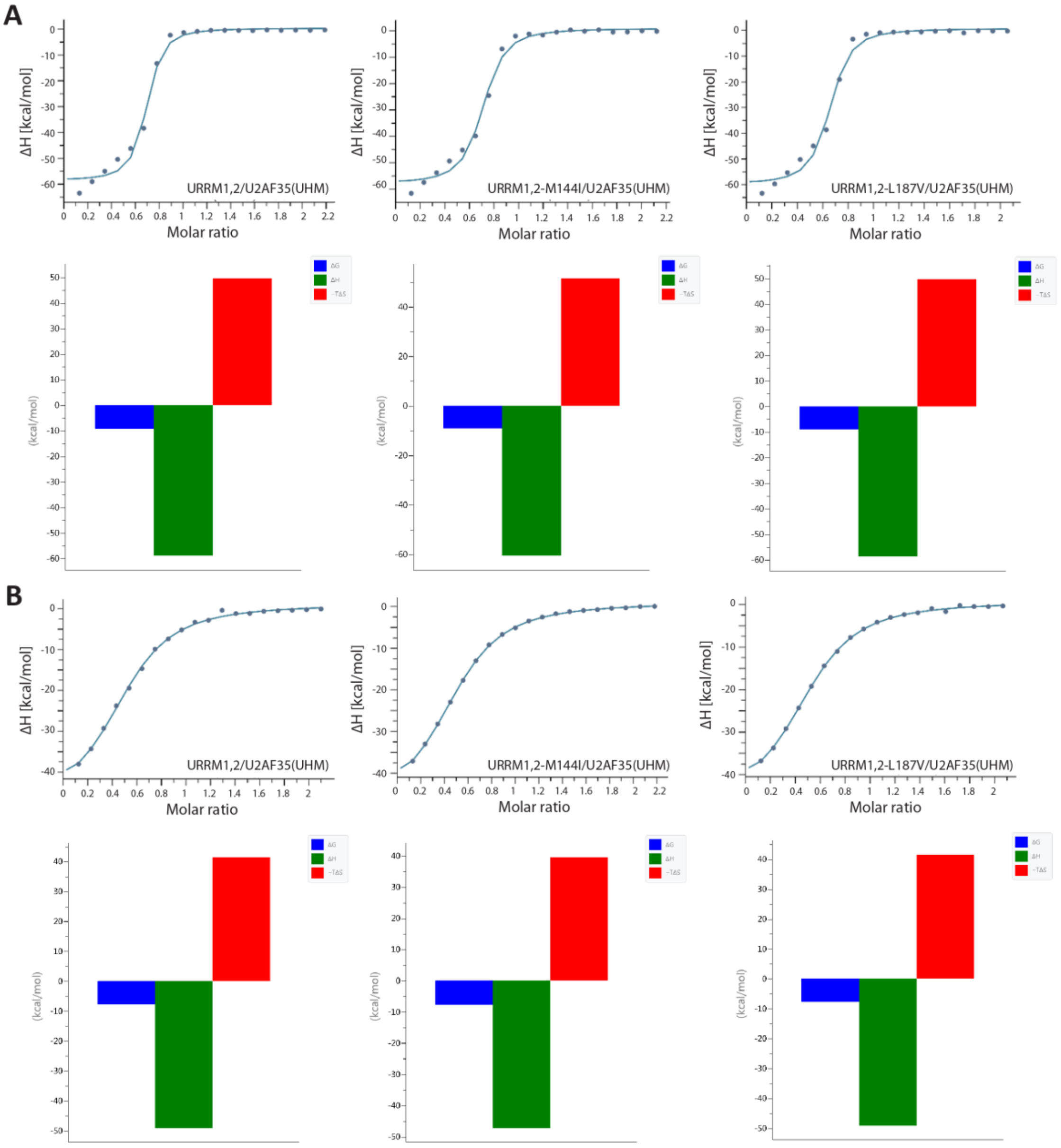

Figure 56: Effects of the MDS mutation on RNA binding. *Influence of the M144I and L187V mutations on the RNA binding of URRM1,2/U2AF35(UHM). Binding affinity curve of (A) U13ACAGG when bound to (Left) URRM1,2/U2AF35(UHM) K_D: 124 nM, (Middle) URRM1,2-M144I/U2AF35(UHM) K_D: 209 nM, and (Right) URRM1,2-L187V/U2AF35(UHM) K_D: 186 nM. (B) U4A8U4ACAGG when bound to (Left) URRM1,2/U2AF35(UHM) K_D: 1.96 uM, (Middle) URRM1,2-M144I/U2AF35(UHM) K_D: 2.23 uM, and (Right) URRM1,2-L187V/U2AF35(UHM) K_D: 1.97 uM. The binding*

affinities were determined by isothermal titration calorimetry and below each curve is presented its corresponding thermodynamic values plot (signature plot).

6.4 Effect of the dye-labelling position on RRM1,2 integrity

We did a rigorous study of the effect of the FRET dyes used during our studies on the protein they are attached to. Primary we analyzed the effect of the dyes at different labeling position within the RRM1,2 U2AF65 protein (Figure 57). The dye Atto647N was attached to three different positions: the 318 residue which is on the surface of the α-helix 2 in RRM2 (Figure 57A, left), the 322 which is at the end of α-helix 2 (Figure 57A, middle) and is a more flexible location, and finally at the position 316 which is in a loop (Figure 57A, right) and so is a flexible position. From the comparison of the ^{1}H,^{15}N-HSQC spectrum and peak intensity plots of the protein with and without the dye attached (Figure 57B, C) we can observe that attaching the fluorophore at position 318 affects a large number of residues in a stronger manner. In most of the cases the peaks lost up to a 90% of its intensity upon the attachment of Atto647N. For the other two positions there was loss of intensity in less extreme manner, however in the case of the dye attached to 322 position still several residues show reduction of intensity between 50-90%. Finally, in the case when the dye is attached to the 316 residue, the intensity signals are comparable to the ones of the spectra of the unlabeled protein, only with few residues exhibiting line-broadening (intensity loss).

The later results means that in the case of fluorophores, in this particular case the Atto647N, is of fundamental importance to determine a correct position for its conjugation. For some studies, for instance in the case of EPR or PRE studies is necessary to make the spin label as rigid as possible (Mackereth et al., 2011; Simon, Madl, Mackereth, Nilges, & Sattler, 2010). However, in the case of FRET using bigger fluorescent molecules it was shown here that the attachment to a secondary structure motive that is relatively rigid affects significantly the integrity of the protein provoking partial unfolding of the region involved. Thus we found that an optimal position for attachment was in a more flexible location at the position 316 which is located in the loop L5, further away from the helix α2.

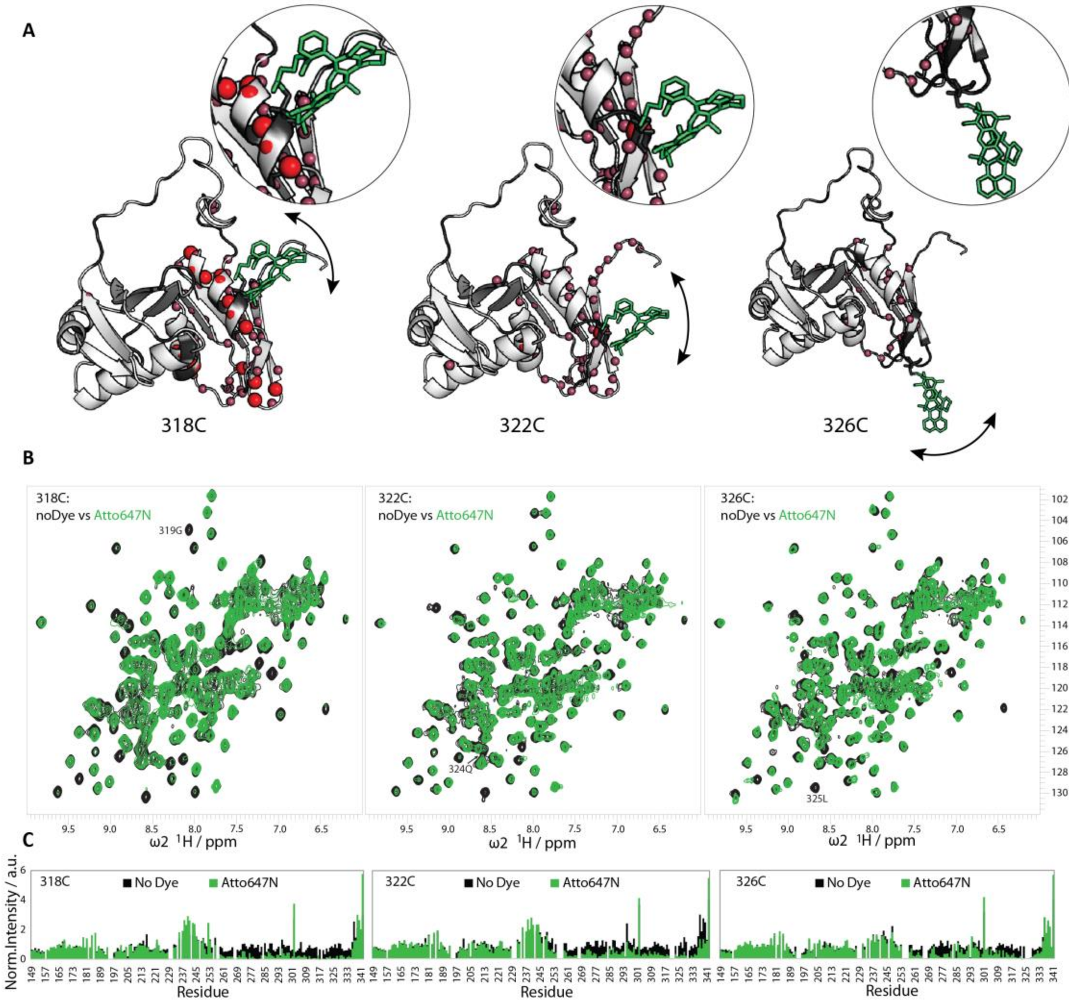

Figure 57: Labeling of the acceptor fluorophore Atto647N to different positions of U2AF65 RRM1,2. *(A) Structural representations of RRM1,2 with the Atto647N fluorophore placed at positions C318 (left), C322 (middle), and C326 (right). Red spheres indicate the residues with loss of up to 90% of peak height upon fluorophore labeling and the pink spheres represent the residues with a loss between 90-50%. (B)* $^{1}H,^{15}N$*-HSQC spectra of unlabeled RRM1,2 (black) compared to RRM1,2 fluorescently labeled with Atto647N (green) on the positions C318 (left), C322 (middle), and C326*

(right). (C) Intensity vs. residue plots of the labeled and unlabeled U2AF65 in the respective positions (318, 322, and 326).

The study of the Atto647N was of particular interest since during the FRET studies using this acceptor fluorophore an artifact at very high FRET efficiency was observed, which was absent when using other acceptor fluorophores such Alexa647 or Cy5 (Figure 58). When FRET experiments were performed with Atto647N in both cases, in the free RRM1,2 protein (Figure 58B, left) or when RNA was bound (Figure 58B, right), a population was found to have very high FRET efficiency that does not correspond to a genuine system. The accessible volume for the donor and acceptor fluorophores are also shown for the free form (closed conformation) and the RNA-bound form (open conformation) (Figure 58A).

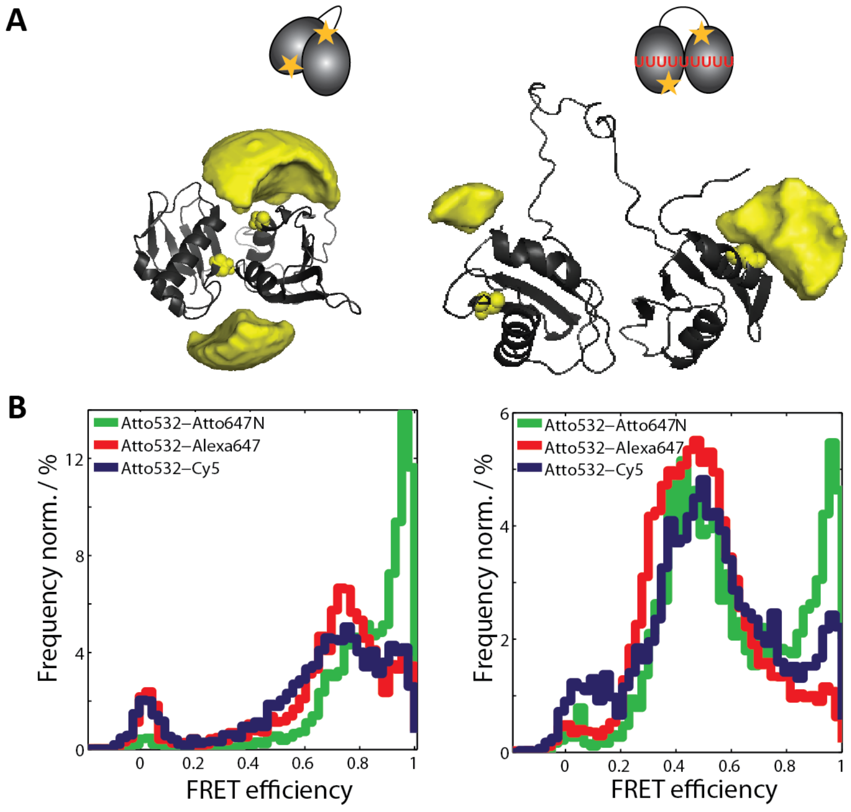

Figure 58: FRET artifacts caused by the acceptor fluorophore. *SpFRET experiments on the U2AF65 RRM1,2-C187-C318. (A) Schematic representations and accessible volume*

calculations of the fluorophores attached to positions C187 and C318 of RRM1,2 in its closed state in the absence of RNA (left) and in the presence of RNA, open state (right). (B) FRET efficiency histograms of RRM1,2 labeled with Atto532 and the acceptor fluorophores Atto647N (green), Alexa647 (red), and Cy5 (blue) in the absence (left) and presence (right) of U9 RNA.

6.5 Fluorophore hydrophobicity affects protein integrity

To further investigate the impact of different dyes on the integrity of the protein we performed NMR experiments when different dyes are attached, Atto532, Atto647N, Alexa647 and Cy5 (Figure 59). In this case we selected the position 316 as it was shown previously to be an optimal position within the protein. Overall the spectrum and the intensity plots when attached any of the dyes have most of the signals and the peak intensities were not extremely reduced when comparing with the spectra of the unlabeled protein. Nevertheless, one can observe that the spectra of Alexa647 and Atto532 labeled protein (Figure 59A, C) have less number of amide signals affected by the conjugation of these dyes in comparison to the Cy5 dye (Figure 59D) whose spectra shows doubling of some peaks reflecting weak interactions with the surface of the protein. The spectra of the Atto647N-labeled RRM1,2 shows larger line-broadening than the rest of the dyes. The Atto647N acceptor fluorophore is highly hydrophobic, which likely explains its strong effects onto the protein conformation.

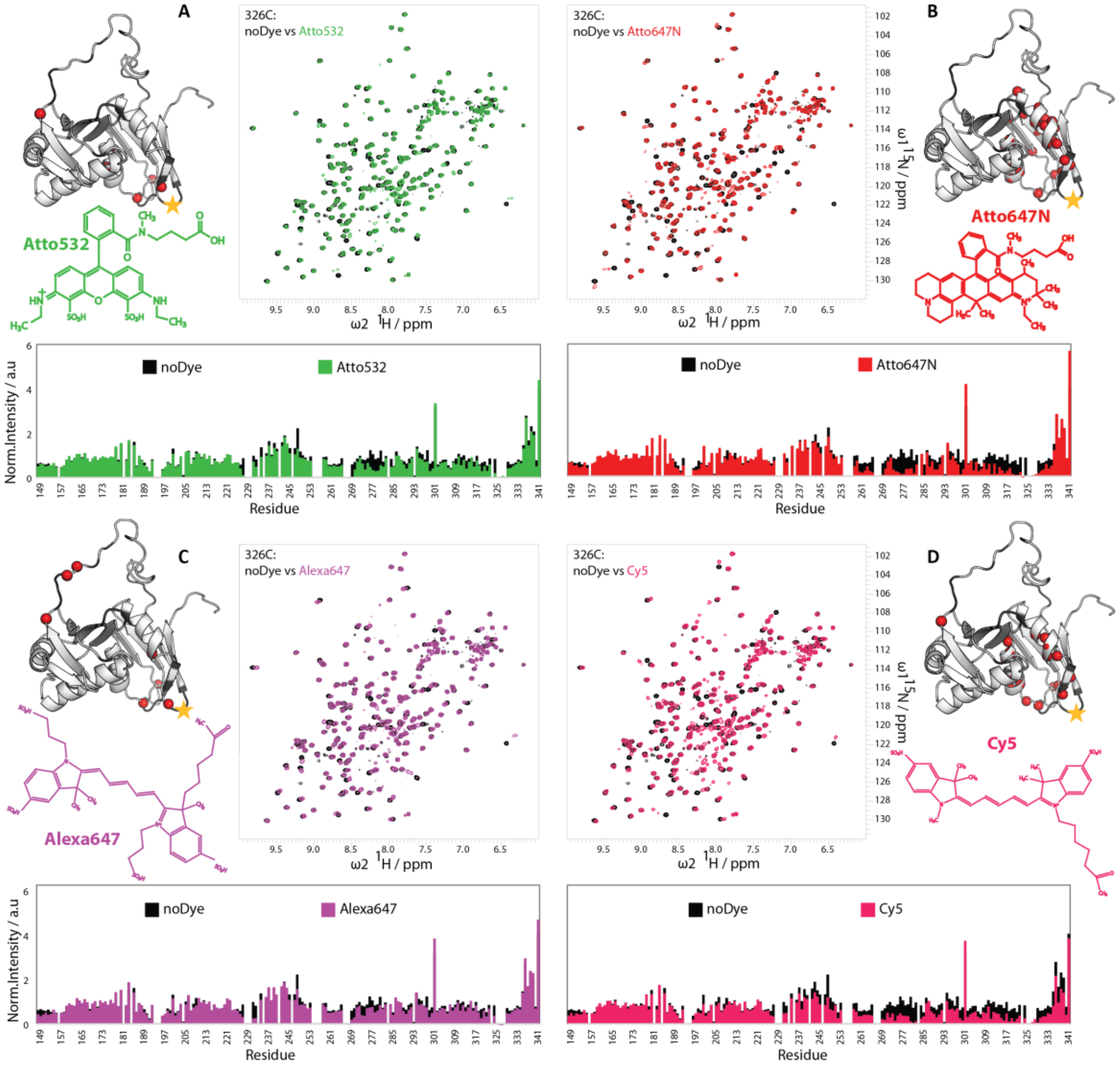

Figure 59: Effect of different fluorophores at position C326 on the conformation of U2AF65 RRM1,2. *^{1}H,^{15}N-HSQC spectra and intensity vs residue plots of unlabeled RRM1,2 (black) compared to (A) RRM1,2-C326 fluorescently labeled with Atto532 (green), (B) Atto647N (red), (C) Alexa647 (magenta), or (D) Cy5 (pink). Next to each spectrum are the structural formulas of the respective dyes and the structural representations of RRM1,2, where the red spheres indicate the residues with loss of more than 60% of peak height upon fluorophore labeling.*

6.6 Fluorophore effect on RNA binding

To evaluate the effect of the fluorophores on RNA-binding, we labeled the protein with acceptor fluorophores at the two positions, 187C in RRM1 and 326C in RRM2. First with Alexa647 only and then with Atto647N (Figure 61). The spectrum of the double-labeled protein presents most of the peaks for both of the dyes (Figure 61C, up). Yet when titrating RRM1,2 with a strong affinity RNA U9 Py-tract, the RRM1,2 spectra labeled with Alexa647 show chemical shift changes and line-broadening for NMR signas of residues involved in RNA binding, which is expected during the titration.

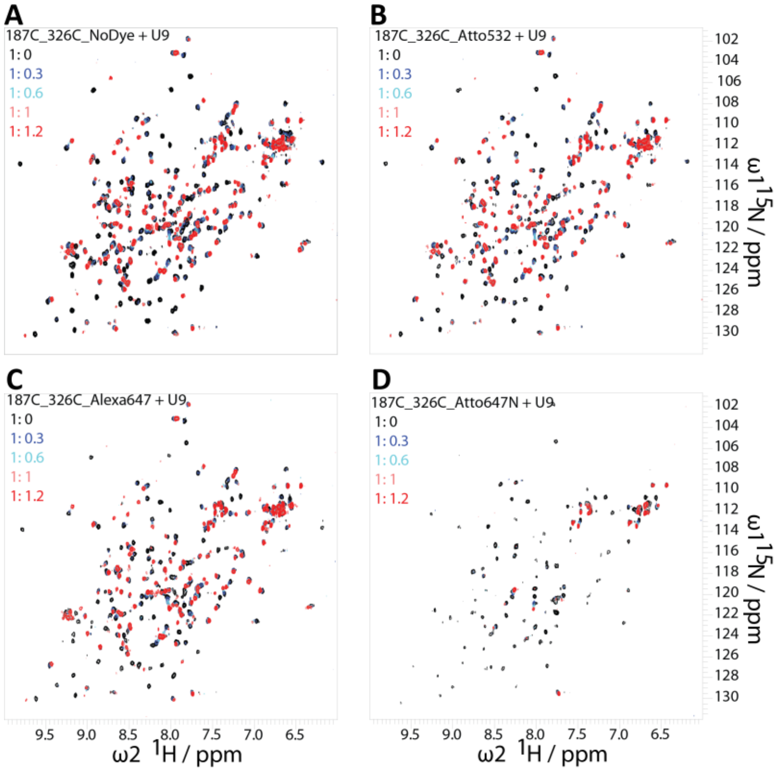

Figure 60: U9 titrations of 187C_326C_RRM1,2 with and without dyes attached. *(A) No dye attached. (B) Atto532 attached. (C) Alexa647 attached. (D) Atto647N attached.*

However, during the titrations of the Atto647N-labeled protein most of the NMR signals are shifted and/or line-broadened beyond detection (Figure 61C, down). A comparison with the RNA titrations of the protein without dye attached is presented in (Figure 60).

SpFRET experiments were performed using the same labeling positions (187 and 326) on RRM1,2 with the acceptor dyes Alexa647 and Atto647N and using Atto532 as donor. In the free RRM1,2, the FRET efficiency histogram with Atto647N shows a high FRET peak (Figure 61B), but to less extend as observed previously for position C318 (Figure 58). Changing the position of the fluorophore from C318 to C326, with a larger distance to the α-helical region in the RRM2 domain, slightly reduces the amplitude of the artificial high FRET peak from the inactive population compared to the active fraction of the molecules. However, there is still a minor (although reduced) population of inactive molecules present, as indicated by the FRET artifact. This is in agreement with the stronger reducution of peak intensities in ^{1}H,^{15}N HSQC NMR spectra, when attaching Atto647N in position C318 compared to position C326, which was described previously. Alexa647 does not show this high FRET artifact when attachted to position C326.

The FRET efficiency histograms of the RNA-bound RRM1,2 using Alexa647 and Atto647N were surprisingly similar (Figure 61B, middle and down) despite the artifact produced by Atto647N. This artificial high FRET peak observed when Atto647N is attached to the C326 position is reduced in comparison to the observed peak in the bound form when Atto647N is attached to the 318C position (Figure 58).

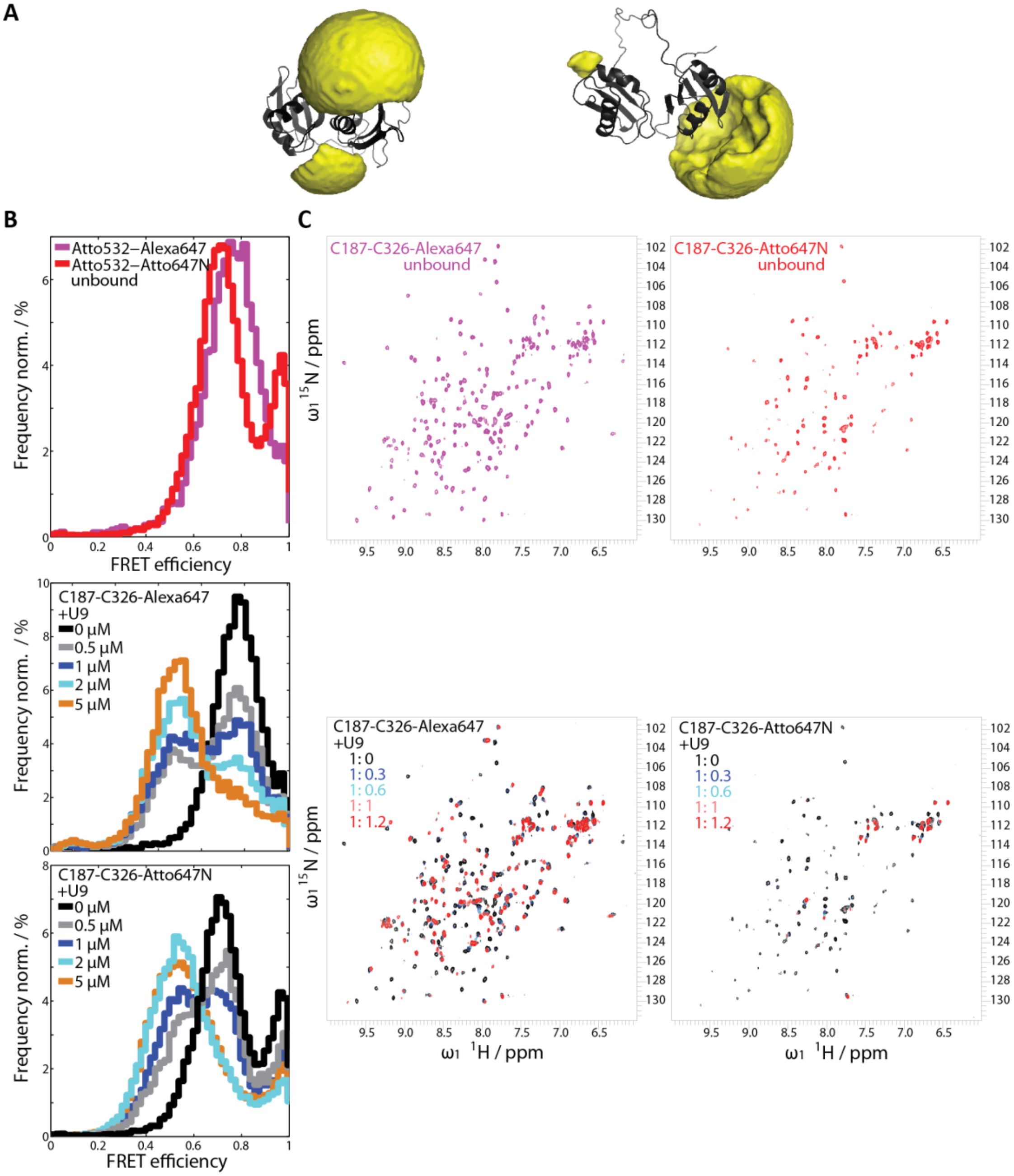

Figure 61: Effect of fluorophores on RNA binding. *(A) Accessible volume calculations of the fluorophores on positions C187 and C326 of RRM1,2 in the closed state in the absence of RNA (left) and in the presence of RNA (right). (B) FRET efficiency histograms of RRM1,2 labeled with Atto532 and Atto647N (red) or Alexa647 (magenta) in its free form (upper panel). SpFRET efficiency histograms of RRM1,2 labeled with Atto532 and*

Alexa647 (middle panel) or Atto532 and Atto647N (lower panel) upon addition of RNA at concentration of 0 µM black, 0.5 µM gray, 1 µM dark blue, 2 µM light blue, 5 µM orange. (C) Upper panels: ^{1}H,^{15}N-HSQC spectra of unlabeled RRM1,2 (black) and RRM1,2 labeled on positions C187 and C326 with Alexa647 (magenta) or with Atto647N (red). Lower panels: 1H,15N HSQC spectra of RRM1,2 fluorescently labeled with Alexa647 (left) or Atto647N (right) during titration series with U9-RNA (black to red).

CHAPTER 7:

Discussion

Dynamics of multi-domain proteins play essential roles during the regulation of numerous biological processes. Particularly, during the spliceosome assembly and during mRNA splicing. RNAs and multi-domain proteins cooperate extensively to perform specific functions, which include the correct recognition and coupling of the different splicing sites. Here a study of the conformation and dynamics of the splicing factor U2AF and its RNA recognition was performed combining solution state methods in structural biology.

In the past years it has been widely studied ensembles and conformations adopted by intrinsically disordered proteins (IDPs) in solution (Fisher & Stultz, 2011; Jensen et al., 2009). However, less is known about domains that are connected by flexible linkers and containing flexible regions. Here was revealed that in absence of RNA, the RRM1-RRM2 tandem domains of U2AF65 exist as an ensemble of conformations sampling a larger conformational space than initially thought. This was shown in this study by an ensemble approach. First of all, to guarantee the unbiased sampling available to the system, a randomized linker between RRM1 and RRM2 was used, with this a pool of structures was generated such that largely covers the RRM1-RRM2 conformational space. From this pool, a selection of ensembles was performed using the ASTEROIDS algorithm (Nodet et al., 2009; Salmon et al., 2010), in which conformers that are in agreement with the NMR and SAXS experimental data were selected. The selected ensemble of the tandem domains includes the 'closed' and 'open' conformations that were previously studied, however a large fraction of the ensemble is comprised by detached (extended) conformations. A prediction of the ensemble established that the proximity of the domains in the region of the 'closed' conformation is of electrostatic origin, since charge complementary patches were identify at the interface. To verify this prediction, PREs at different salt concentrations were measured. The PRE (I_{ox}/ I_{red}) values at the inter-domain interface decreased as the salt concentration increased, this implies that the transient contacts between the domains were weakened due to high salt. This results support the suggested electrostatic counter-like interactions between the two domains. Additionally, the ensemble also confirmed the existence of a significant population in the open state, which is the conformation of the domains when bound to RNA. This result support the role of conformational selection mechanism during RNA recognition that has been proposed before (Jenkins, Agrawal, Gupta, Green, & Clara L. Kielkopf, 2013; Mackereth et al., 2011). These new results on the RRM1,2 domains in solution allow us to understand the interplay between the two domains, and this underline the basis for the recognition of diverse pre-mRNA intron sequences, which is the intriguing role of U2AF65. In order to understand the mechanisms during RNA recognition it would be important to study the energetic landscape changes of U2AF65 in presence of Py-tracts of different strengths.

During the pool generation and subsequent ensemble selection the liker between the two domains was randomized assuming that no contacts between the domains and the linker were present. However, a more attentive examination of the linker was required, given the different configuration of the NMR structure of the RNA-bound RRM1-RRM2 (Mackereth et al., 2011) versus a previously solved X-tal structure (Sickmier et al., 2006) where the linker was deleted. For this purpose several GS-linker constructs ($(GGS)_N$ random coil sequence) were designed, where 20 residues of the wild type linker were replaced completely or partially in the different constructs. By calculating chemical shift perturbations between the spectra of the different GS-linker constructs versus the wild-type RRM1-RRM2, it was revealed that the linker interacts with the two RRM domains. These contacts were observed between the C-terminus of RRM2 and the C-terminal part of the linker; the N-terminus of RRM1 (including the extended N-terminal of RRM1, presented in the X-tal structure from Agrawal et al., 2016) and the N-terminal part of the linker, and particularly large chemical shifts were observed between the C-terminal region of the linker and the β2 strand of RRM2. These contacts were confirmed by the NOE-based refinement of the free nxRRM1-

RRM2 structure (by Hyun-Seo Kang). These linker/domain interactions might drag the domains closer and thus promote the encounter-like contacts by allowing the charged patches to encounter.

Suspiciously, the large interaction of the linker with RRM2 overlaps with regions involved in RNA binding, particularly with the RNPs in the central β-strands of RRM2. For this reason, a study of the role of the linker in RNA binding was performed in order to establish if the linker has an auto-inhibitory function during RNA recognition.

The NMR and ITC titrations of the different GS-linker and wild-type constructs with the strong and weak Py-tracts (U9 and U4A8U4 respectively) showed that the linker competes with RNA binding, particularly with the weak affinity Py-tract. This results were supported by in vitro iCLIP data (not shown) with natural pre-mRNA sequences. Hence, we believe that the linker has an auto-inhibitory function which leads to a fine tuning modulation of RNA binding by U2AF65, providing a proof-reading mechanism to discriminate against weak Py-tracts. When contrasting the interactions of the linker with the N-terminus and C-terminus of RRM1,2, the RRM2 RNPs plus additional encounter-like interactions with the domains (Figure 41), in comparison with the RNA-titrations of nxRRM1-RRM2 and the recent X-tal structure of the RNA-bound RRM1-RRM2 (Agrawal et al., 2016) one can observe that each of those regions are involved in RNA binding. Moreover, SAXS, hetNOE data of the RNA-bound U2AF65 and NMR titrations series of the GS-linker construct vs. the wild type with the different Py-tracts, suggest that linker residues (e.g. V254) as well as other residues in the domains (Figure 43) interact in different way with the weak and strong Py-tracts. Additional FRET studies on U2AF65 showed that in the presence of the strong Py-tract the protein-RNA complex is rigid, while it displays internal motion when it is bound to the weak Py-tract. This goes in agreement with the larger radius of gyration observed in SAXS for the complex when bound to the weak Py-tract versus bound to the strong Py-tract. The previous observations showed how a flexible linker can assist the autoregulation of a multi-domain protein. Additional indications from other systems in the Sattler lab (e.g RBMs, TIA-1. Not published) allow to speculate that the linker in several multi-domain proteins, particularly RNA-binding proteins, have additional roles for conformation and RNA-binding modulation.

While the presence of U2AF65 is sufficient for splicing in vitro of introns containing strong Py-tracts, U2AF35 is needed for splicing in vivo of weak Py-tracts (Guth, Martínez, Gaur, & Valcárcel, 1999; Pacheco, Coelho, Desterro, Mollet, & Carmo-Fonseca, 2006). Recently it was revealed that the two zinc finger domains of U2AF35 directly contact the AG-dinucleotide at the 3'-end of the intron (H. Yoshida et al., 2015). The next goal of this thesis was to study the role of the U2AF35 UHM domain during Py-tract binding of U2AF65. Combining FRET and NMR studies, it was revealed that the small subunit U2AF35(UHM) allosterically strengthens the U2AF65(URRM1,2) binding of weak Py-tracts. In the free form, U2AF35(UHM) induces a population shift towards the open conformation of U2AF65(URRM1,2). This result was supported by NMR data that show an interaction interface between the RRM1 domain of U2AF65 and the UHM domain of U2AF35. This population shift towards the open state induced by the UHM domain of U2AF35, greatly enhances the U2AF65 binding of weak Py-tracts. Hence, this study gives a novel role for the UHM domain of U2AF35 during the 3'-SS recognition, which in combination with the AG-dinucleotide recognition by the zinc finger domains improves the fidelity of this splice site by the U2AF heterodimer. Since the open state of RRM1-RRM2 within the U2AF heterodimer is significantly populated already in absence of RNA, this gives support for the conformational selection mechanism of RNA-binding proposed previously (Mackereth et al., 2011). Moreover, here it was shown that the recognition of Py-tracts of different strength yields to different dynamic behavior in the protein-RNA complex, and this might be correlated to different splicing efficiencies (Mackereth et al., 2011).

Previous studies have shown that somatic mutations in pre-mRNA splicing factors are present in most of the patients with myelodysplastic syndrome (MDS) (Haferlach et al., 2014). These mutations are more common in U2AF35. However, the confirmed mutation L187V and the unconfirmed mutation M144I in U2AF65 (K. Yoshida et al., 2011) were suggested to indirectly contribute to MDS progression by disrupting the interaction between the two U2AF subunits (Agrawal et al., 2016). Here it was proved that these MDS mutations in U2AF65 neither affect the formation of the U2AF heterodimer, nor the RNA binding of different Py-tracts. The overlap of NMR spectra of the wild type heterodimer U2AF65(URRM1,2)/ U2AF35(UHM), versus the ones containing the mutations, showed minimal differences in the conformation of the heterodimer. Even though the L187V and M144I mutations in U2AF65 are located in a solvent exposed region and at the N-terminus of RRM1 respectively, they do not affect the interaction with U2Af35 during the heterodimer formation. The ITC titrations of the mutated heterodimers with weak and strong Py-tracts showed minimal differences in the binding affinities as well as in the thermodynamic signatures in comparison to the wild type. Therefore, these MDS mutations might be involved in a different interaction of U2AF65 with another factor involved in spliceosome assembly.

In fluorescence spectroscopy, particularly in spFRET organic fluorophores with optimized photophysical properties are commonly used (e.g. Atto532, Atto647N, Alexa647). In this thesis a thorough study of the effects of commonly used fluorophores on the integrity of the U2AF65 protein is presented. Using NMR, it was revealed that the attachment of fluorophores to the surface of a rigid secondary element (e.g. α-helices) affects the integrity of the protein, particularly with the attachment of the acceptor fluorophore Atto647N. Unlike spin-labels commonly used in NMR, large and hydrophobic fluorophores should be rather located in flexible regions as loops. Additionally, a comparison among several donor and acceptor fluorophores was done, giving as a result that less hydrophobic and less aromatic fluorophores (e.g Alexa647, Atto532) are better preserving the integrity of the protein in contrast to highly aromatic ones (e.g Atto647N) even when located in a flexible loop. This study allows to explain the observation of additional subpopulations at high FRET efficiency when the acceptor fluorophore Atto647N is used. Subsequently, the RNA-binding to the protein when dyes are attached to two positions simultaneously (in RRM1 and in RRM2), is severely affected when the acceptor Atto647N is attached. Therefore, the specific characteristics of each fluorophore and the position of the conjugation site should be taken into account when performing FRET experiments, both *in vitro* and *in vivo.* Not only the optimal photophysical properties should be the parameter to decide which dyes to use, but their effect on the specific biomolecular system should be considered.

The solution techniques used throughout this thesis have several advantages in comparison to static or solid state methods. In particular, NMR is the only technique that allows to investigate the function and dynamics of biomolecular systems at atomic level. However, there are size limitations on the system to be studied as well as the low sensitivity of NMR in comparison to other techniques.

Combining SAXS measurements with atomic-resolution methods is a powerful way to validate the conformation of the protein (protein/ligand) in solution. It does not required large quantities of sample, however it is very sensitive to aggregation. SAXS provides an average shape of all the particles present, thus is very important to have a sample as homogenous as possible.

FRET is a powerful tool to study conformation and dynamics of biomolecular systems at different time scales. In particular, single molecule FRET offers the advantage that individual molecules can be investigated. The amount of sample required is in the order of picomolar. The fluorophores used in FRET are usually expensive and depending on their physicochemical properties can affect the protein and/or its biomolecular interactions.

Appendix

Protein Sequences

RRM1-RRM2 U2AF65

(M)ARRLYVGNIPFGITEEAMMDFFNAQMRLGGLTQAPGNPVLAVQINQDKNFAFLEFRSVD
ETTQAMAFDGIIFQGQSLKIRRPHDYQPLPGMSENPSVYVPGVVSTVVPDSAHKLFIGGL
PNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSKGYAFCEYVDINVTDQAIAGLNGMQLG
DKKLLVQRASVGAKNA

nxRRM1-RRM2 U2AF65

(M)VGSQMTRQARRLYVGNIPFGITEEAMMDFFNAQMRLGGLTQAPGNPVLAVQINQDKNFA
FLEFRSVDETTQAMAFDGIIFQGQSLKIRRPHDYQPLPGMSENPSVYVPGVVSTVVPDSA
HKLFIGGLPNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSKGYAFCEYVDINVTDQAIA
GLNGMQLGDKKLLVQRASVGAKNA

GS-linker-RRM1,2 U2AF65

(M)ARRLYVGNIPFGITEEAMMDFFNAQMRLGGLTQ
APGNPVLAVQINQDKNFAFLEFRSVDETTQAMAFDGIIFQGQSLKIRRPHDYGSGGSGGS
GGSGGSGGSGGSVVPDSAHKLFIGGLPNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSK
GYAFCEYVDINVTDQAIAGLNGMQLGDKKLLVQRASVGAKNA

nx-eGSlinker-RRM1,2 U2AF65

(M)VGSQMTRQARRLYVGNIPFGITEEAMMDFFNAQ
MRLGGLTQAPGNPVLAVQINQDKNFAFLEFRSVDETTQAMAFDGIIFQGQSLKIRRPHDY
GSGGSGGSGGSGGSGGSGGSGGPDSAHKLFIGGLPNYLNDDQVKELLTSFGPLKAFNLVK
DSATGLSKGYAFCEYVDINVTDQAIAGLNGMQLGDKKLLVQRASVGAKNA

UHM-U2AF35_C67S_C102S

(M)DKPTFSQTIALLNIYRNPQNSSQSADGLRSAVSDVEMQEHYDEFFEEVFTEMEEKYGEV
EEMNVSDNLGDHLVGNVYVKFRREEDAEKAVIDLNNRWFNGQPIHAELSPVTDFRE

ULM-RRM1-RRM2 U2AF65

(M)VRKYWDVPPPGFEHITPMQYKAMQAAGQIPATALLPTMTPDGLAVTPTPVPVVGSQMTR
QARRLYVGNIPFGITEEAMMDFFNAQMRLGGLTQAPGNPVLAVQINQDKNFAFLEFRSVD
ETTQAMAFDGIIFQGQSLKIRRPHDYQPLPGMSENPSVYVPGVVSTVVPDSAHKLFIGGL
PNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSKGYAFCEYVDINVTDQAIAGLNGMQLG
DKKLLVQRASVGAKNA

GS-linker genes

The GS-linker genes were bought from Eurofins Genomics, Germany.
Restriction sites: NcoI, NotI

>RRM12 GS-linker (GSGGSGGSGGSGGSGGS)

CcatgGCCCGGCGCCTCTACGTGGGCAACATCCCCTTTGGCATCACTGAGGAGGCCATGATGG
ATTTCTTCAACGCCCAGATGCGCCTGGGGGGGCTGACCCAGGCCCCTGGCAACCCAGTGTTG
GCTGTGCAGATTAACCAGGACAAGAATTTTGCCTTTTTGGAGTTCCGCTCAGTGGACGAGaCT
ACCCAGGCTATGGCCTTTGATGGCATCATCTTCCAGGGCCAGTCACTAAAGATCCGCAGGCCT
CACGACTACggt agt gga ggc tcc ggt gga tca ggc ggc tcc gga ggc agtggt ggc tcg ggc gga
agcGTGGTCCCCGACTCTGCCCACAAGCTGTTCATCGGGGGCTTACCCAACTACCTGAACGATG
ACCAGGTCAAAGAGCTGCTGACATCCTTTGGGCCCCTCAAGGCCTTCAACCTGGTCAAGGAC
AGTGCCACGGGGCTCTCCAAGGGCTACGCCTTCTGTGAGTACGTGGACATCAACGTCACGGA
TCAGGCCATTGCGGGGCTGAACGGCATGCAGCTGGGGGATAAGAAGCTGCTGGTCCAGAGG
GCGAGTGTGGGAGCCAAGAATGCCtgagcggccgc

>RRM1,2 (1/2wt-1/2GS)-linker (QPLPGMSENP-SGGSGGSGGS)

CcatgGCCCGGCGCCTCTACGTGGGCAACATCCCCTTTGGCATCACTGAGGAGGCCATGATGG
ATTTCTTCAACGCCCAGATGCGCCTGGGGGGGCTGACCCAGGCCCCTGGCAACCCAGTGTTG
GCTGTGCAGATTAACCAGGACAAGAATTTTGCCTTTTTGGAGTTCCGCTCAGTGGACGAGaCT
ACCCAGGCTATGGCCTTTGATGGCATCATCTTCCAGGGCCAGTCACTAAAGATCCGCAGGCCT
CACGACTACcagccgcttcctggcatgtcagagaacccc tcc gga ggc agtggt ggc tcg ggc gga
agcGTGGTCCCCGACTCTGCCCACAAGCTGTTCATCGGGGGCTTACCCAACTACCTGAACGATG
ACCAGGTCAAAGAGCTGCTGACATCCTTTGGGCCCCTCAAGGCCTTCAACCTGGTCAAGGAC
AGTGCCACGGGGCTCTCCAAGGGCTACGCCTTCTGTGAGTACGTGGACATCAACGTCACGGA
TCAGGCCATTGCGGGGCTGAACGGCATGCAGCTGGGGGATAAGAAGCTGCTGGTCCAGAGG
GCGAGTGTGGGAGCCAAGAATGCCtgagcggccgc

>RRM1,2 (1/2GS-1/2WT)-linker (GSGGSGGSGG-SVYVPGVVST)

CcatgGCCCGGCGCCTCTACGTGGGCAACATCCCCTTTGGCATCACTGAGGAGGCCATGATGG
ATTTCTTCAACGCCCAGATGCGCCTGGGGGGGCTGACCCAGGCCCCTGGCAACCCAGTGTTG
GCTGTGCAGATTAACCAGGACAAGAATTTTGCCTTTTTGGAGTTCCGCTCAGTGGACGAGaCT
ACCCAGGCTATGGCCTTTGATGGCATCATCTTCCAGGGCCAGTCACTAAAGATCCGCAGGCCT
CACGACTACggt agt gga ggc tcc ggt gga tca ggc
ggctccgtctatgtgcctggggttgtgtccactGTGGTCCCCGACTCTGCCCACAAGCTGTTCATCGGGGG

CTTACCCAACTACCTGAACGATGACCAGGTCAAAGAGCTGCTGACATCCTTTGGGCCCCTCAA
GGCCTTCAACCTGGTCAAGGACAGTGCCACGGGGCTCTCCAAGGGCTACGCCTTCTGTGAGT
ACGTGGACATCAACGTCACGGATCAGGCCATTGCGGGGCTGAACGGCATGCAGCTGGGGGA
TAAGAAGCTGCTGGTCCAGAGGGCGAGTGTGGGAGCCAAGAATGCCtgagcggccgc

>RRM1,2 1/2GS-linker (GSGGSGGSGG)

CcatgGCCCGGCGCCTCTACGTGGGCAACATCCCCTTTGGCATCACTGAGGAGGCCATGATGG
ATTTCTTCAACGCCCAGATGCGCCTGGGGGGGCTGACCCAGGCCCCTGGCAACCCAGTGTTG
GCTGTGCAGATTAACCAGGACAAGAATTTTGCCTTTTTGGAGTTCCGCTCAGTGGACGAGaCT
ACCCAGGCTATGGCCTTTGATGGCATCATCTTCCAGGGCCAGTCACTAAAGATCCGCAGGCCT
CACGACTACggt agt gga ggc tcc ggt gga tca ggc ggc
GTGGTCCCCGACTCTGCCCACAAGCTGTTCATCGGGGGCTTACCCAACTACCTGAACGATGAC
CAGGTCAAAGAGCTGCTGACATCCTTTGGGCCCCTCAAGGCCTTCAACCTGGTCAAGGACAG
TGCCACGGGGCTCTCCAAGGGCTACGCCTTCTGTGAGTACGTGGACATCAACGTCACGGATCA
GGCCATTGCGGGGCTGAACGGCATGCAGCTGGGGGATAAGAAGCTGCTGGTCCAGAGGGCG
AGTGTGGGAGCCAAGAATGCCtgagcggccgc

U2AF65-RRM12 T_1, T_2, $T_{1\rho}$ relaxation parameters

Experimental parameters for data collected at different protein concentrations. The T1, T2 and T1rho pulse sequences for the 300uM and 1mM samples are the same. While the T1, T2 and T1rho pulse sequences for the 640uM have bipolar gradients in the t1 evolution time.

	300uM, 800MHz			640uM, 800MHz, bipolar gradients.			1mM, 500MHz		
	T1	T2	T1ρ	T1	T2	T1ρ	T1	T2	T1ρ
T1,T2_ Vclist(s), T1ρ_ Vdlist(s)	0.0216 0.0216 0.0864 0.162 0.2484 0.3456 0.432 0.5148 0.6696 0.8856 1.3392 1.6632	0.01208 0.01208 0.02416 0.03624 0.04832 0.07248 0.09664 0.09664 0.1208 0.14496 0.16912 0.19328	0.005 0.005 0.015 0.025 0.04 0.04 0.055 0.065 0.08 0.11 0.11 0.14	0.0216 0.0216 0.0864 0.162 0.162 0.2484 0.3456 0.432 0.432 0.5148 0.6696 0.8856 1.3392 1.6632	0.01208 0.01208 0.02416 0.03624 0.04832 0.07248 0.09664 0.09664 0.1208 0.14496 0.16912 0.19328	0.005 0.005 0.015 0.025 0.04 0.04 0.055 0.065 0.08 0.11 0.11 0.14	0.1188 0.1188 0.3456 0.432 0.5148 0.6696 0.8856 1.3392 1.6632	0.01208 0.01208 0.02416 0.03624 0.04832 0.07248 0.09664 0.09664 0.1208 0.14496 0.16912 0.19328	0.005 0.005 0.015 0.025 0.04 0.04 0.055 0.08 0.11 0.14
TD1	12	12	12	14	12	12	9	12	10
TD2	256	256	256	128	256	150	256	256	256
TD3	1024	1024	1024	1024	2048	1024	1024	1024	1024
P1@pl1	11.01us@-12.3	11.01us @-12.3	10.4 us@ -12.3	10.94us@-12.5	10.94us @-12.5	10.94us @-12.5	8.7us @-8dB	9.65us @-8dB	8.7us @-8dB

P5@pl3	36.2us @-23.71	36.2us @-23.71	36.2 us@ -23.7	37.2us @-23.5	34.57us @-23.5	35.5 us @-23.5	42us @-18dB	42us @-18dB	42us @-18dB
D1	1.5s	1.5s	1.5s	1.5s	1.5s	1.5s	1.5s	1.5s	1.5s
Spin lock			2KHz 125us @ -12.9 dB			2KHz, 125us @-12.93dB (should be -12.6 dB)			2KHz,125 us @-8.53 dB
CPMG		144.8us @-17.7dB			148.44 us@ -17.5dB Should be 137.96 us			167.6us @-12dB	
Sp10 dB (wg power)	22.32dB@1ms	22.32dB@1ms	22.32dB @1 ms	22.23dB@1ms	23.4dB @ 1ms	22.25dB @1 ms	28.16dB@1ms	28.16dB@1ms	29.21dB @1 ms
Sp11dB (fb power)	24.05dB@1ms	24.05dB@1ms	24.05dB @1 ms	24.79dB@1ms	22.43dB@1ms	24.79dB @1 ms	29.21dB@1ms	29.21dB@1ms	28.16dB @1 ms

Abbreviations

1D, 2D, 3D	One-, Two-, Three-Dimensional
AEBSF	4- Benzenesulfonyl fluoride hydrochloride
BME	β-mercaptoethanol
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
FRET	Förster resonance energy transfer
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
Ni2+	Nickle
OD	Optical Density at 600nm Wavelength
pre-mRNA	precursor messenger RNA
Rg	Radius of gyration
RMSD	Root mean square deviation
SAXS	Small angle X-ray scattering
TEV	Tobacco Etch Virus
UHM	U2AF homology motifs
ULM	UHM-ligand motif

List of Figures

Figure 1: Schematic overview of pre-mRNA splicing.. 19
Figure 2: Spliceosome Assembly and pre-mRNA splicing. 21
Figure 3: Complex E during spliceosome assembly.. 22
Figure 4: Domain organization U2AF.. 23
Figure 5: Interactions at the 3' splice site. 24
Figure 6: Canonical RRM fold. 26
Figure 7: RNA binding mechanisms by multidomain proteins. 27
Figure 8: Energy levels of nuclear spin I=1/2.. 31
Figure 9: Chemical shift origin.. 33
Figure 10: Chemical exchange rates. 36
Figure 11: Spectral Density. 40
Figure 12: Relation between molecular tumbling and T1, T2 relaxation. 42
Figure 13: Time scales investigated with NMR. 43
Figure 14: Schematics of PREs. 44
Figure 15: ITC schematics. 47
Figure 16: SAXS profile regions.. 48
Figure 17: Pair distribution function of the SAXS data for the RRM1-RRM2 67
Figure 18: Representative structures of the ensembles 70
Figure 19: Position of the closed and open states with respect the ensemble. 71
Figure 20: Crossvalidation.. 73
Figure 21: 3D density maps 75
Figure 22: Origin of the recognized encounter complex is electrostatic. 76
Figure 23: PRE measurements for spin-labeled A318C RRM1-RRM2 78
Figure 24: Summary of the ensemble selection 79
Figure 25: RNA-bound RRM1-RRM2 structures. 82

Figure 26: RRM1, RRM2 vs RRM12. 83
Figure 27: Representations of GS-linker RRM1,2 constructs 85
Figure 28: Chemical shift perturbations of the GS-linker constructs. 87
Figure 29: Chemical shift perturbation of different length GS-linker. 88
Figure 30: U9-bound RRM1-RRM2 vs. GS-linker chemical shift differences.. 89
Figure 31: WT-RRM1-RRM2 vs GS-linker vs single domains. 91
Figure 32: Sequence Alignment. 92
Figure 33: NMR titrations of different GS-linker constructs. 93
Figure 34: NMR titrations of the WT vs GS-linker constructs... 94
Figure 35: ITC titrations of the GS-linker constructs.. 96
Figure 36: Bar Plot ITC titrations 97
Figure 37: SAXS data of the WT and GS-linker RRM1,2 98
Figure 38: Biacore studies 99
Figure 39: hetNOE and refined structure.. 101
Figure 40: Interactions of the RRM1 N-terminal extension. 102
Figure 41: nxRRM1-RRM2 versus eGS-linker 103
Figure 42: SAXS data of the nxRRM1,2 bound to U9 and U4A8U4. 104
Figure 43: Different patterns on RNA binding. 105
Figure 44: PREs WT-RRM1,2_254G and GS-linker RRM1,2. 106
Figure 45: PREs nxRRM12 vs nxRRM12_254G. 107
Figure 46: Line width differences between single vs tandem RRM1,2 domains.. 109
Figure 47: Intensity distribution single domains vs tandem domain. 110
Figure 48: $Ʈ_c$ at different concentrations. 111
Figure 49: Chemical exchange on the free RRM1,2… 113
Figure 50: Chemical exchange on the RNA-bound U2AF65. 114
Figure 51: Conformation and dynamics of RRM1,2 in solution. 117
Figure 52: spFRET experiments of RRM1,2 and URRM1,2/U2AF35(UHM). 119

Figure 53: PRE experiment on the U2AF heterodimer.. 121
Figure 54: Schematic model on the role of U2AF35 during RNA binding.. 122
Figure 55: NMR study on the U2AF heterodimer-MDS mutations 124
Figure 56: Effects of the MDS mutation on RNA binding. 125
Figure 57: Labeling of fluorophore to different positions... 127
Figure 58: FRET artifacts caused by the acceptor fluorophore.. 128
Figure 59: Effect of different fluorophores at position C326 130
Figure 60: U9 titrations of 187C_326C_RRM1,2. 131
Figure 61: Effect of fluorophores on RNA binding. 133

Acknowledgements

First of all I would like to express my gratitude to my supervisor Prof. Dr. Michael Sattler for giving me the opportunity to work in his lab. When I first came I did not have much experience in biochemistry lab work and nevertheless he allowed me to work in an interesting project. Thanks to that, during these years in the Sattler group I have grown not only professionally but also personally.

I would like to thank my graduate school GRK1721, led by Prof. Dr. Karl-Peter Hopfner and the great job of Dr. Petra Runge-Wollman. It has been an enriching experience to share with students and PIs from different fields of structural biology. Besides the financial support, we have received beneficial training in different techniques and soft skills.

My special regards to the members of my thesis advisory committee, Prof. Dr. Bernd Reif from Technical University, Munich and Prof. Dr. Friedrich Förster from University of Utrecht (previously MPI-Biochemistry) for their valuable suggestions.

I would like to thank my collaborators throughout my doctoral studies. First, Prof. Dr. Martin Blackledge and Dr. Jie-rong Huang from the IBS, Grenoble, France for the valuable collaboration during the ensemble project. Thanks to Prof. Dr. Don Lamb and especially to Dr. Lena Voithenberg at Ludwig-Maximilians-Universität München for the fruitful collaboration combining FRET and NMR techniques. Thanks to Dr. Hyun-Seo Kang for the collaboration and discussions on the linker project. Thanks to Dr. Sam Asami for his work on the relaxation dispersion experiments and his help setting up NMR experiments. Thanks to Stefanie Ebersberger and Prof. Dr. Julian Koenig at IMB, Mainz, Germany for the iClip studies on the linker project.

Particularly I want to thank Dr. Lisa Warner for all her instructions at the beginning of my project and her valuable guidance on NMR. I want to thank also Gülden Demiraslan and Dr. Kostas Tripsianes for their initial training in the lab and their patience with my inexperience.

Special acknowledgements to Ralf for recording all the SAXS data. As well to Rainer, Gerd and Waltraud at TUM for maintaining and managing the NMR facilities and the administrative work. Thanks to Ari and Astrid at Helmholtz Zentrum for their help and for keep running the labs in good conditions. I also thank to the members I have met in the Sattler group as well as in Madl, Niessing and Reif groups.

A big thank you to all the people that have made these years of studies a great experience not only academically but personally, especially to Ashish, Johannes, Komal, Pravin, Eleni, Diana, Leo, Anna, Alex, Hamed and Irina for the nice work atmosphere and great times together. Especial thanks to my friends, nearby or far away, thanks for your friendship and support all along.

Finally, I want to express my deep gratitude to my family, my parents and my two sisters. What I am and what I have achieved is thanks to them, to their love, encouragement and support in every step I take. To Martin, my love, my friend, my partner. I cannot imagine a more wonderful coincidence in life. Thanks for all your support, encouragement and for making me a better person.

References

Agrawal, A. A., Salsi, E., Chatrikhi, R., Henderson, S., Jenkins, J. L., Green, M. R., Kielkopf, C. L. (2016). An extended U2AF65-RNA-binding domain recognizes the 3′ splice site signal. *Nature Communications*, *7*, 10950.

Battiste, J. L., & Wagner, G. (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.

Beato, M. (1993). Gene Regulation by Steroid Hormones. In M. Karin (Ed.), *Gene Expression* (pp. 43–75). Birkhäuser Boston.

Berglund, J. A., Abovich, N., & Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Development*, *12*(6), 858–867.

Black, D. L. (2003). Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry*, *72*(1), 291–336.

Boldon, L., Laliberte, F., & Liu, L. (2015). Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Reviews*, *6*.

Burnouf, D., Ennifar, E., Guedich, S., Puffer, B., Hoffmann, G., Bec, G., … Dumas, P. (2012). kinITC: A New Method for Obtaining Joint Thermodynamic and Kinetic Data by Isothermal Titration Calorimetry. *Journal of the American Chemical Society*, *134*(1), 559–565.

Cheng, Z., & Menees, T. M. (2011). RNA splicing and debranching viewed through analysis of RNA lariats. *Molecular Genetics and Genomics*, *286*(5–6), 395–410.

Cléry, A., & Allain, F. H.-T. (2013). *FROM STRUCTURE TO FUNCTION OF RNA BINDING DOMAINS*. Landes Bioscience.

Cléry, A., Blatter, M., & Allain, F. H.-T. (2008). RNA recognition motifs: boring? Not quite. *Current Opinion in Structural Biology*, *18*(3), 290–298.

Clore, G. M., & Iwahara, J. (2009). Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chemical Reviews*, *109*(9), 4108–4139.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3), 277–293.

Fisher, C. K., & Stultz, C. M. (2011). Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology*, *21*(3), 426–431.

Göbl, C., Madl, T., Simon, B., & Sattler, M. (2014). NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *80*, 26–63.

Green, M. R. (1986). PRE-mRNA Splicing. *Annual Review of Genetics*, *20*(1), 671–708.

Grzesiek, S. (2003). Notes on relaxation and dynamics. *Course Notes, EMBO Practical Course on NMR, Heidelberg*.

Guth, S., Martínez, C., Gaur, R. K., & Valcárcel, J. (1999). Evidence for Substrate-Specific Requirement of the Splicing Factor U2AF35 and for Its Function after Polypyrimidine Tract Recognition by U2AF65. *Molecular and Cellular Biology*, *19*(12), 8263–8271.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., … Ogawa, S. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*, *28*(2), 241–247.

Hore, P. (2015). *Nuclear Magnetic Resonance*. Oxford University Press.

Huang, J., Warner, L. R., Sanchez, C., Gabel, F., Madl, T., Mackereth, C. D., … Blackledge, M. (2014). Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study. *Journal of the American Chemical Society*, *136*(19), 7068–7076.

Ito, T., Muto, Y., Green, M. R., & Yokoyama, S. (1999). Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF65). *The EMBO Journal*, *18*(16), 4523–4534.

Jenkins, J. L., Agrawal, A. A., Gupta, A., Green, M. R., & Clara L. Kielkopf. (2013). U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Research*, *41*(6), 3859–3873.

Jensen, M. R., Markwick, P. R. L., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., … Blackledge, M. (2009). Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure*, *17*(9), 1169–1185.

Kay, L. E., Torchia, D. A., & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, *28*(23), 8972–8979.

Keeler, J. (2010). *Understanding NMR Spectroscopy* (2nd ed.). Chichester, U.K: John Wiley & Sons.

Kielkopf, C. L., Lücke, S., & Green, M. R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes & Development*, *18*(13), 1513–1526.

Kielkopf, C. L., Rodionova, N. A., Green, M. R., & Burley, S. K. (2001). A Novel Peptide Recognition Mode Revealed by the X-Ray Structure of a Core U2AF35/U2AF65 Heterodimer. *Cell*, *106*(5), 595–605.

Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., … Pandey, A. (2014). A draft map of the human proteome. *Nature*, *509*(7502), 575–581.

Kimple, A. J., Muller, R. E., Siderovski, D. P., & Willard, F. S. (2010). A capture coupling method for the covalent immobilization of hexahistidine tagged proteins for surface plasmon resonance. *Methods in Molecular Biology (Clifton, N.J.)*, *627*, 91–100.

Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta*, *1814*(8), 942–968.

Kudryavtsev, V., Sikor, M., Kalinin, S., Mokranjac, D., Seidel, C. A. M., & Lamb, D. C. (2012). Combining MFD and PIE for Accurate Single-Pair Förster Resonance Energy Transfer Measurements. *ChemPhysChem*, *13*(4), 1060–1078.

Mackereth, C. D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., … Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, *475*(7356), 408–411.

Mackereth, C. D., & Sattler, M. (2012). Dynamics in multi-domain protein recognition of RNA. *Current Opinion in Structural Biology*, *22*(3), 287–296.

Markwick, P. R. L., Malliavin, T., & Nilges, M. (2008). Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLOS Computational Biology*, *4*(9), e1000168.

MicroCal Auto-ITC200 user manual (English). (n.d.). Retrieved February 2, 2017, from http://www.malvern.com/de/support/resource-center/user-manuals/MAN0552EN.aspx

Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, *463*(7280), 457–463.

Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., & Blackledge, M. (2009). Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *Journal of the American Chemical Society*, *131*(49), 17908–17918.

Ozenne, V., Bauer, F., Salmon, L., Huang, J., Jensen, M. R., Segard, S., … Blackledge, M. (2012). Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*, *28*(11), 1463–1470.

Pacheco, T. R., Coelho, M. B., Desterro, J. M. P., Mollet, I., & Carmo-Fonseca, M. (2006). In Vivo Requirement of the Small Subunit of U2AF for Recognition of a Weak 3′ Splice Site. *Molecular and Cellular Biology*, *26*(21), 8183–8190.

Patterson, B., & Guthrie, C. (1991). A U-rich tract enhances usage of an alternative 3′ splice site in yeast. *Cell*, *64*(1), 181–187.

Rule, G. S., & Hitchens, T. K. (2005). *Fundamentals of Protein NMR Spectroscopy* (2006th ed.). Dordrecht: Springer.

Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., & Blackledge, M. (2010). NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *Journal of the American Chemical Society*, *132*(24), 8407–8418.

Sattler, M., Schleucher, J., & Griesinger, C. (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.

Schnablegger, H., & Singh, Y. (2011). The SAXS guide: getting acquainted with the principles. *Anton Paar GmbH, Austria*.

Schneider, R., Maurin, D., Communie, G., Kragelj, J., Hansen, D. F., Ruigrok, R. W. H., … Blackledge, M. (2015). Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *Journal of the American Chemical Society*, *137*(3), 1220–1229.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., & Sattler, M. (2003). Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF65 and SF1/mBBP. *Molecular Cell*, *11*(4), 965–976.

Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., & Kielkopf, C. L. (2006). Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Molecular Cell*, *23*(1), 49–59.

Simon, B., Madl, T., Mackereth, C. D., Nilges, M., & Sattler, M. (2010). An Efficient Protocol for NMR-Spectroscopy-Based Structure Determination of Protein Complexes in Solution. *Angewandte Chemie International Edition*, *49*(11), 1967–1970.

Smith, C. W. J., & Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Sciences*, *25*(8), 381–388.

Valcárcel, J., Gaur, R. K., Singh, R., & Green, M. R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science (New York, N.Y.)*, *273*(5282), 1706–1709.

Voithenberg, L. V. von, Sánchez-Rico, C., Kang, H.-S., Madl, T., Zanier, K., Barth, A., … Lamb, D. C. (2016). Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proceedings of the National Academy of Sciences*, *113*(46), E7169–E7175.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., … Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics*, *59*(4), 687–696.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, *136*(4), 701–718.

Walsh, J. D., Meier, K., Ishima, R., & Gronenborn, A. M. (2010). NMR Studies on Domain Diffusion and Alignment in Modular GB1 Repeats. *Biophysical Journal*, *99*(8), 2636–2646.

Will, C. L., & Lührmann, R. (2011). Spliceosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, *3*(7), a003707.

Yoshida, H., Park, S.-Y., Oda, T., Akiyoshi, T., Sato, M., Shirouzu, M., … Obayashi, E. (2015). A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes & Development*, *29*(15), 1649–1660.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., … Ogawa, S. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, *478*(7367), 64–69.

Curriculum Vitae

Carolina Sánchez Rico

From Mistrato, Colombia
Department of Chemistry, Technical University of Munich
Lichtenbergstrasse 4, 80747 Garching
E-mail: carolina.sanchez@tum.de

Education

03/2012 – present	**Ph.D. in Structural Biology** *Technical University of Munich, Chair of Biomolecular NMR Spectroscopy, Munich, Germany* Thesis' title: “Dynamics of RNA recognition by the multidomain splicing factor U2AF”
02/2008-12/2010	**M. Sc. in Physics** *CINVESTAV-IPN (Research and Advanced Studies Center of the National Polytechnic Institute of Mexico), Mexico City, Mexico* Thesis' title: “Study of the production of free radicals from amyloid beta with Cd and Cr by EPR”
10/2001-12/2007	**B. Sc. in Physical Engineering** *National University of Colombia, Medellin, Colombia* Thesis' title: “Optical and Structural properties of ZnO and Zn1-xMnxO thin films synthesized by a Sol-Gel technique”

Awards and Fellowships

08/2016	**Young Investigator Award and Poster Prize** *International Conference on Magnetic Resonance in Biological Systems, Kyoto, Japan*
09/2008-08/2010	**Fellow granted by CONACyT (Mex)** To perform master studies

Publications

- **“Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift”.**
  Von Voithenberg L, **Sanchez C**, Kang HS, Madl T, Zanier K, Barth A, Warner LR, Sattler M and Don C. Lamb. Proc. Natl. Acad. Sci., no. 46 (2016)7169–75

- **“Transient Electrostatic Interactions Dominate the Conformational Equilibrium Sampled by Multidomain Splicing Factor U2AF65: A Combined NMR and SAXS Study”**
  Huang JR, Warner LR, **Sanchez C**, Gabel F, Madl T, Mackereth CD, Sattler M and Blackledge M. *J Am Chem Soc. 136 (2014) 7068-7076*

- **“Comparison of the effect of chronic cadmium exposure on the antioxidant defense systems of kidney and brain in rat”**
  Mendieta JE, Miliar A, Correa J, **Sanchez C**, Ramirez D, Trujillo J and Rosales MC. *Toxicol. Mech. Methods., 23(2013) 329-336*

- **“On the paramagnetic behavior of heavily doped $Zn_{1-x}Mn_xO$ films fabricated by Pechini's method”**
  Sánchez C, Paucar C, A. Mosquera A, Gómez A, Moran O. *Superlatt.&Microstruc. 52(2012) 249-260*

- **“Nanocryszalline ZnO films prepared via polymeric precursor method (Pechini)”**
  Sanchez C, Doria J, Paucar C, Hernandez M, Mosquera A, Rodriguez JE, Gomez A, Baca E and Moran O. *Physica B 405 (2010) 3679-3684*

- **“Applicability of the polymeric precursor method to the synthesis of nanometric single and multi-layers of $Zn_{1-x}Mn_xO$ (x=0-0.3)"**
  Mera SJ, Cordoba C, Benavidez J, Paredes O, Doria J, Gomez A, **Sanchez C**, Paucar C and Moran O. *J Mater Sci 45(2010) 5398-5405*

Manuscripts under Preparation

- **“Effects of fluorophore attachment on protein conformation and dynamics studied by spFRET and NMR”**
  Carolina Sanchez-Rico*, Lena Voith von Voithenberg*, Lisa Warner, Don C. Lamb, Michael Sattler

- **“Autoinhibitory role of the linker between RRM1 and RRM2 domains of the U2AF65 splicing factor during RNA-binding”**
  Carolina Sanchez Rico*, Hyun-Seo Kang*, Stefanie Ebersberger, Ralf Stehle, Sam Asami, Julian Koenig, Michael Sattler

Selected Conferences and Workshops

08/2016	• **International Conference on Magnetic Resonance in Biological Systems**, *Kyoto, Japan*
03/2015	• **Keystone Symposia, Hybrid Methods in Structural Biology**, *Tahoe, California, USA*
06/2014	• **EMBO workshop, Magnetic Resonance for Cellular Structural Biology**, *Principina Terra, Grosseto, Italy*
12/2013	• **Communication and Presentation Workshop**, *LMU, Munich, Germany*
09/2013	• **Scientific Writing Workshop**, *Gene Center, Munich, Germnay*
03/2013-10/2015	• **SAXS, Illustrator and SPR workshops**, *LMU and Helmholtz-Zentrum München*