

Technische Universität München

Lehrstuhl für biomolekulare NMR-Spektroskopie

Department Chemie

Structural studies on cooperative 3' splice site recognition by the SF1-U2AF65 complex

Yun Zhang

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. St. A. Sieber

Prüfer der Dissertation:

1. Univ.-Prof. Dr. M. Sattler
2. Univ.-Prof. Dr. A. Itzen

Die Dissertation wurde am 19.03.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 22.04.2013 angenommen.

Immer nur lernen, ohne dabei nachzudenken, das
führt zu Verwirrung. Immer nur nachdenken, ohne
dabei zu lernen, das führt zu Erschöpfung.

Konfuzius

DECLARATION

I hereby declare that parts of this Thesis are already published in scientific journal:

Zhang, Y., Madl, T. (shared first author), Bagdiul, I., Kern, T., Kang, H.S., Zou, P., Mausbacher, N., Sieber, S.A., Kramer, A., and Sattler, M. (2013). Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3'-splice site recognition. *Nucleic Acids Res* *41*, 1343-1354.

I. ABSTRACT

Recognition of the 3' splice site is a key step in pre-mRNA splicing and accomplished by a dynamic complex comprising splicing factor 1 (SF1) and the U2 snRNP auxiliary factor 65-kDa subunit (U2AF65). Both proteins mediate protein-protein and protein-RNA interactions for cooperative RNA-binding during spliceosome assembly. This work presents structural investigations of cooperative recognition of 3' splice site by SF1-U2AF65 complex. The main focus is on the N-terminal domain of SF1, which is posttranslationally modified by phosphorylation. Additionally, the contribution of the U2AF65 RS (arginine-serine-rich) domain to binding of the intronic branch point sequence (BPS) by SF1 was investigated. Structural analysis by NMR methods combined with SAXS analysis provided novel insight about mechanisms of cooperative recognition of 3' splice site.

Chapter 1.1 introduces the biological background of pre-mRNA alternative splicing and splicing factors involving in recognition of 3' splice site. Chapter 1.2 reviews aspects of NMR methods used for structural analysis and characterization of biomolecules in solution. In Chapter 2, materials and methodology employed for biochemical experiments and structural analysis are documented.

Chapter 3.1 presents NMR and SAXS data that characterize the structural basis for cooperative RNA binding by SF1-U2AF65. The solution structure of a novel helix hairpin domain in the N-terminal region of SF1 (SF1^{NTD}) is described. The structure of a complex of the N-terminal domain of SF1 (SF1^{NTD}) with the C-terminal U2AF homology motif domain of U2AF65 (U2AF65^{UHM}) reveals that, in addition to the known U2AF65^{UHM}-SF1 interaction, the helix-hairpin domain forms a secondary, hydrophobic interface with U2AF65^{UHM}, which locks the orientation of the two subunits. Mutational analysis shows that the helix hairpin is essential for cooperative formation of the ternary SF1-U2AF65-RNA complex. Chapter 3.2 describes NMR and biochemical investigation of SF1 phosphorylation. Structural and dynamic data reveal that tandem serine phosphorylation of a conserved SPSP motif rigidifies a long unstructured linker in the SF1 helix hairpin. Phosphorylation does not significantly alter the overall conformation of SF1, SF1-U2AF65 or the SF1-U2AF65-RNA complexes, but

nevertheless enhances RNA binding by SF1-U2AF65. Chapter 3.3 addresses the studies on U2AF65 RS domain. NMR chemical shifts and paramagnetic relaxation enhancement data indicate that the U2AF65 RS domain exhibits secondary structural elements and tertiary contacts in solution. Chapter 4 reviews the results of this thesis and discusses the role of phosphorylation and U2AF65-binding of SF1 during 3' splice site recognition.

The work presented in this thesis provides a structural investigation of the cooperative recognition of the 3' splice site by SF1-U2AF65, and reports on the effects of SF1 phosphorylation. The results indicate that the helix hairpin domain of SF1 is required for cooperative 3' splice site recognition presumably by stabilizing a unique quaternary arrangement of the SF1-U2AF65-RNA complex. NMR analysis of the U2AF65 RS domain shows that the RS domain comprises two α -helical regions, which could be interaction interfaces with SF1 KH-QUA2. Altogether, this work provides a significant step towards solving the overall structure of the ternary SF1-U2AF65-RNA complex and thus understanding the molecular basis of 3' splice site recognition and regulation.

II. TABLE OF CONTENTS

I. ABSTRACT	I
II. TABLE OF CONTENTS.....	III
1 INTRODUCTION	1
1.1 Biological background introduction.....	1
1.1.1 Alternative pre-mRNA splicing.....	1
1.1.2 Pre-mRNA splicing by the spliceosome	4
1.1.2.1 The spliceosome.....	4
1.1.2.2 Spliceosome assembly.....	5
1.1.3 Structural studies of Complex E.....	7
1.1.3.1 Complex E.....	7
1.1.3.2 5' splicing site.....	8
1.1.3.3 3' splice site.....	10
1.1.4 Splicing factor 1 (SF1)	12
1.1.5 SF1 phosphorylation	13
1.1.6 The U2AF65 arginine-Serine-rich (RS) domain	15
1.2 Nuclear magnetic resonance (NMR) spectroscopy.....	17
1.2.1 The principles of NMR.....	17
1.2.2 NMR for protein analysis.....	21
1.2.3 Protein dynamics by NMR	23
1.2.4 Paramagnetic relaxation enhancement (PRE).....	24
1.2.5 NMR studies for protein-ligand interaction.....	26
1.2.6 Structure calculation	27
1.3 Scope of the Thesis	29
2 MATERIALS AND METHODS.....	31
2.1 Materials	31
2.1.1 Chemicals and enzymes.....	31
2.1.2 Bacterial strains and vectors.....	33
2.1.3 Recipes.....	34
2.1.3.1 Media and buffers for protein expression	34
2.1.3.2 Buffers for protein purification	36
2.2 Biochemical methods	39
2.2.1 Cloning.....	39
2.2.2 Protein expression and NMR sample preparation.....	41
2.2.3 In vitro phosphorylation	42
2.2.4 Isothermal titration calorimetry.....	42
2.2.5 Circular Dichroism	42
2.2.6 Spin labeling.....	43

2.2.7	Electrophoretic mobility shift assays (EMSAs)	43
2.3	NMR and SAXS techniques	45
2.3.1	NMR measurements and assignments	45
2.3.2	Secondary chemical shift analysis and chemical shift perturbations.....	45
2.3.3	Relaxation analysis.....	46
2.3.4	Structure calculations	47
2.3.5	Small-angle X-ray scattering (SAXS)	47
3	RESULTS.....	49
3.1	Structural basis of SF1-U2AF65 cooperative RNA binding.....	49
3.1.1	SF1 constructs	49
3.1.2	NMR analysis and structure of the N-terminal region of SF1	52
3.1.3	SF1 ^{NTD} provides a secondary interface in the SF1 ^{NTD} -U2AF65 ^{UHM} complex	58
3.1.4	Interactions between SF1 and RNA.....	65
3.1.5	Cooperative recognition of RNA by SF1-U2AF65 complex	67
3.1.6	Discussion: Cooperative binding of SF1-U2AF65 to Pre-mRNA.....	69
3.2	Structural analysis of SF1 phosphorylation.....	71
3.2.1	Phosphorylation mimics of SF1 (S80E, S82E)	71
3.2.2	Phosphorylation of SF1 on S80 and S82.....	74
3.2.3	NMR analysis of SF1 ^{NTD} and SF1 phosphorylation	77
3.2.4	Spin labeling.....	79
3.2.5	Effects of phosphorylation on SF1 structure.....	81
3.2.6	Role of phosphorylation of SF1 for cooperative RNA binding.....	84
3.2.7	Discussion: The influence of phosphorylation of SF1 on cooperative assembly of the SF1-U2AF65-RNA complex.....	86
3.3	NMR characterization of the U2AF65 arginine-serine-rich (RS) domains	88
3.3.1	Preparation of U2AF65 RS domain.....	88
3.3.2	NMR assignment and NMR analysis of U2AF65 RS domain	90
3.3.3	U2AF65 RS and SF1-BPS interaction	93
3.3.4	Discussion: RS domain contacts SF1 and RNA at the branch point.....	96
4	DISCUSSION.....	98
4.1	A model for the role of phosphorylation and U2AF65-binding of SF1 during 3' splice site recognition.....	98
4.2	Conclusions	101
5	LITERATURE REFERENCES.....	102
6	LIST OF FIGURES.....	112
III	APPENDIX.....	VI
III.I	DNA sequences.....	VI

III.II Protein sequences	VII
III.III Resonance assignments of SF1^{NTD}.....	VIII
III.IV Abbreviations.....	XXI
ACKNOWLEDGEMENTS.....	XXIII
CURRICULUM VITAE.....	XXIV

1 INTRODUCTION

1.1 Biological background introduction

1.1.1 Alternative pre-mRNA splicing

In February 2001, the first draft of the human genome sequence was successfully revealed (Lander et al., 2001; Venter et al., 2001), it surprisingly contains only approximate 23,000 genes, much less than the originally predicted numbers of genes (Roest Crolius et al., 2000), by contrast, the genome of the nematode *C. elegans* contains about 20,000. These findings indicate the lack of an association between gene number and organismal complexity. The major factor for this phenomenon is proposed to be alternative splicing, which plays essential roles in expanding the functional complexity, protein diversity, and organismal complexity of higher eukaryotes (Graveley, 2001; Kim et al., 2007; Nilsen and Graveley, 2010).

Alternative splicing, the removal of non-encoding sequences (introns) and the joining of coding sequences (exons), is an crucial step in eukaryotic gene expression by which multiple distinct transcripts are generated from a single gene (Graveley, 2001). Up to 92%~94% of human multi-exon genes are estimated to undergo alternative splicing (Modrek and Lee, 2002; Pan et al., 2008; Wang et al., 2008). In addition to the splice-site consensus sequences, alternative splicing is regulated by a number of auxiliary elements known as exonic and intronic splicing enhancers (ESEs and ISEs, respectively) and silencers (ESSs and ISSs, respectively). These auxiliary elements are involved in defining both constitutive and alternative exons, deciding which exon is removed and which exon is included. Enhancers can activate adjacent splice sites or antagonize silencers, whereas silencers can repress splice sites or enhancers. Exon inclusion or skipping is determined by the balance of these competing influences, which in turn might be determined by relative concentrations of the cognate RNA-binding activator and repressor proteins (Matlin et al., 2005) (Figure 1.1A).

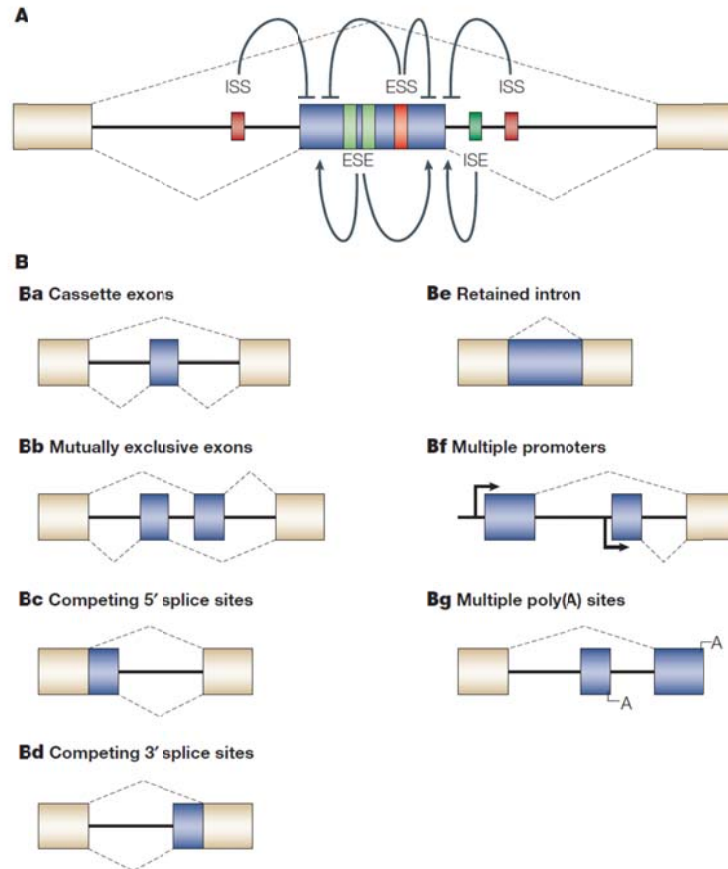


Figure 1.1 Elementary alternative splicing patterns and regulatory elements (adapted from (Matlin et al., 2005)). (A) In addition to the splice-site consensus sequences, ESEs/ISEs and ESSs/ISSs elements can influence alternative splicing. (B) Elementary alternative splicing events represent binary choices. (Ba) Cassette exons are discrete exons that can be independently included or excluded from the mRNA. (Bb) Mutually exclusive splicing involves the selection of only one from an array of two or more exon variants. (Bc,d) Competing 5' and 3' splice sites represent 'exon modification' events. (Be) Retained introns. (Bf) The use of alternative promoters. (Bg) 3' end processing sites.

In a typical eukaryotic multi-exon mRNAs, the splicing pattern can be altered in many different ways (Figure 1.1B). Most of exons are constitutive and are always spliced or kept in the final mRNA. Whereas, another type of exon that sometimes included and sometimes excluded from the mRNA is called a cassette exon (Figure 1.1Ba), they can be further subdivided into 'skipped' or 'cryptic' exons according to whether the main observed variant includes or excludes the exon, respectively. Exons can also be shortened or lengthened by altering the position of their splice sites (Figure 1.1Bb). In addition, competing 5' and 3' splice sites represent 'exon modification' events. Global computational analyses that consider only these first four categories find that 60% involve cassette and mutually

exclusive exons, whereas 40% are exon modifications (Clark and Thanaraj, 2002) (Figure 1.1Bc,d). The remaining categories include retained introns (Figure 1.1Be). Alternative splicing in conjunction with the use of alternative promoters (Figure 1.1Bf) or 3' end processing sites (Figure 1.1Bg). Regulation of the last two categories need not be at the level of splicing.

Canonical mechanism of alternative splicing suggest that serine/arginine (SR) proteins typically bind to ESEs and heterogeneous nuclear ribonucleoproteins (hnRNPs) recognize ESSs or ISSs (Chen and Manley, 2009). SR proteins have a modular structure with one or two N-terminal RNA Recognition motif (RRM)-type domains that bind RNA, and C-terminal domains that are enriched in arginine and serine residues (RS domains). RS domains are also found in other core splicing factors such as U2AF65 (U2 auxiliary factor 65 kDa) and U2AF35. They mediate both protein-protein and protein-RNA contacts (Shen et al., 2004). Detailed descriptions of U2AF65 RS domain can be found in section 1.1.8.

Pre-mRNA splicing is carried out by the spliceosome, a 60S protein-RNA assembly that consists of five uridine-rich small nuclear ribonucleoprotein particles (snRNPs), U1, U2, U4, U5 and U6 and many non-snRNP proteins (Black, 2003; Wahl et al., 2009). Initial recognition of the correct pairs of 5' and 3' splice sites by the spliceosome is a critical step in the processing of both constitutively and alternatively spliced pre-mRNAs. In the early complex of spliceosome, the 3' splice site is recognized exclusively by proteins for constitutive splicing. Furthermore, choices of alternative splicing also have long been thought to be achieved at the stages of splice site recognition and early spliceosome assembly. This raises the importance of understanding the mechanism of splice site recognition during spliceosome assembly, which will be discussed in the following section (s. 1.1.2).

1.1.2 Pre-mRNA splicing by the spliceosome

1.1.2.1 The spliceosome

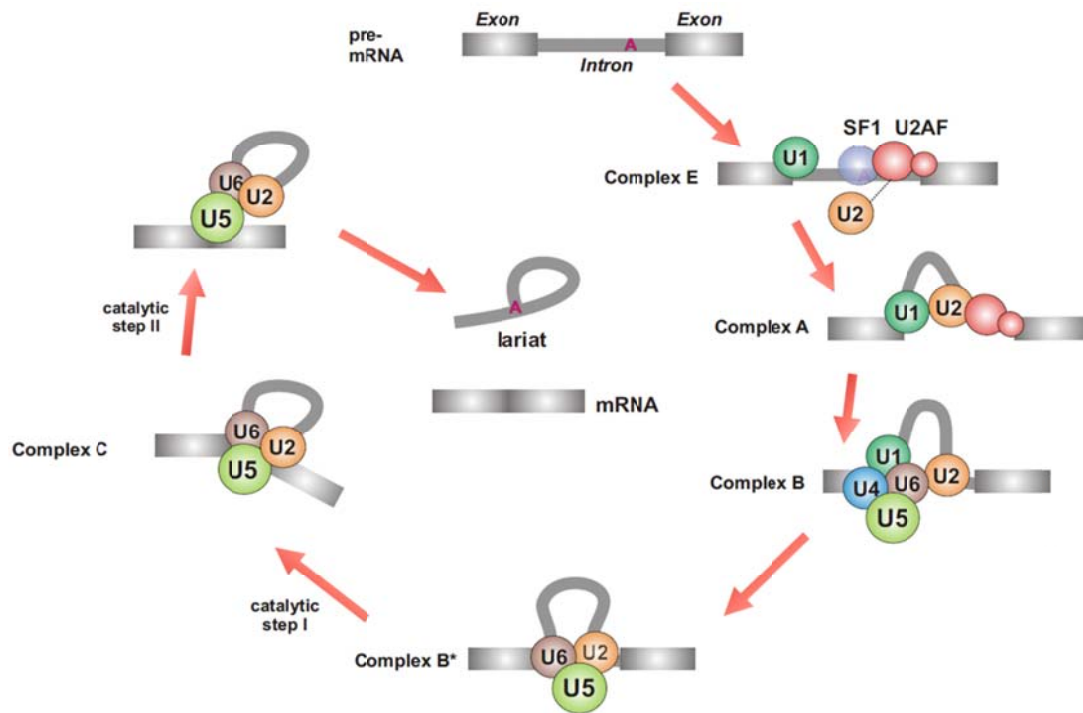


Figure 1.2 Schematic representation of the spliceosome assembly (adapted from P. Selenko & M. Sattler). The assembly of splicing factors and snRNPs onto the pre-mRNA and the dynamic rearrangements of snRNP subunits of the spliceosome are depicted. During the first ATP-dependent step of splicing, complex E is converted into complex A. The association of the U4/U6·U5 tri-snRNP with complex A to form complex B. U6 base pairs with both the 5' splice site and U2 snRNA and this activated complex B* undergoes the first catalytic step, The U2/U6 snRNA in complex C in conjunction with the pre-mRNA promotes the second catalytic step.

Pre-mRNA splicing requires two trans-esterification-reactions steps, each involving a nucleophilic attack on terminal phosphodiester bonds of the intron. In the first step, a conserved 2' hydroxyl of adenosine within the branch site sequence (BPS) of the pre-mRNA intron performs a nucleophilic displacement at the 5' splice site, releasing a free 5' exon and a lariat intron and 3' exon intermediate. In the second step, the attack of the 3' hydroxyl of the upstream 5' exon at the 3' splice site yields ligated exons and the liberated intron lariat (Figure 1.2). Both chemical steps are catalyzed by the spliceosome, a massive structure consisting of five snRNPs and a great number of auxiliary proteins accurately recognize the splice sites (Black, 2003; Wahl et al., 2009). Each snRNP consists of a unique small nuclear

RNA (snRNA) as well as specific proteins. There are several stages of spliceosome formation, in which conserved intron sequences direct the assembly of snRNPs onto the pre-mRNA. Both of the major (or U2-type) spliceosome and minor (or U12-type) spliceosome present many similar conserved features, which mediates splicing of a class of rare metazoan introns (Dietrich et al., 1997; Sharp and Burge, 1997; Tarn and Steitz, 1997; Will et al., 1999).

1.1.2.2 Spliceosome assembly

As a factory for pre-mRNA alternative splicing, spliceosome assembly begins with the base pairing of U1 snRNA to the conserved 5' splice site and the recognition of non-snRNP proteins to the 3' splice site. Splicing Factor 1 or Branchpoint Binding Protein (SF1/BBP) binds to the branch point sequence (Berglund et al., 1997) in an ATP-independent manner to form the E' complex (or commitment complex in *S. cerevisiae*). The E' complex can be converted into the E complex by the recruitment of U2 auxiliary factor (U2AF) heterodimer (comprising U2AF65 and U2AF35) to the polypyrimidine tract and 3' terminal AG158.

The ATP-independent E complex is converted into the ATP-dependent pre-spliceosome complex A. In complex A, the U2 snRNP, which is only loosely associated with the pre-mRNA in complex E (Das et al., 2000; Hong et al., 1997), now becomes tightly bound, inducing the replacement of SF1 by U2 snRNP at the branch point sequence. Several of the U2 snRNP-associated SF3a/b subunits contact the pre-mRNA (Gozani et al., 1998; McPheeters and Muhlenkamp, 2003). Additionally, a cross-intron interaction bridging the U1 and U2 snRNPs is mediated by the RNA-dependent DExD/H helicase Prp5 in yeast and mammals (Xu et al., 2004).

Further recruitment of the U4/U6–U5 tri-snRNP leads to the formation of the B complex, which contains all spliceosomal subunits that carry out pre-mRNA splicing. Subsequently, base pairing interactions involving U1 snRNA/5' splice and U4/U6 snRNA are disrupted, and the U1 and U4 snRNPs leave. U6 base pairs with both the 5' splice site and U2 snRNA and this activated spliceosome (complex B*) undergoes the first trans-esterification step (Figure 1.2) (Makarov et al., 2002). This is followed by extensive conformational changes and remodeling, including the loss of U1 and U4 snRNPs, ultimately resulting in the formation of the C complex, the mature spliceosome (Reed, 2000; Staley and Guthrie, 1998). The U2/U6 snRNA in complex C in conjunction with the pre-mRNA is thought to form the active site of

the spliceosome and promotes the second catalytic step (Collins and Guthrie, 2000; Makarov et al., 2002; Valadkhan and Manley, 2001).

A great numbers of studies have investigated the structural information of some spliceosomal factors, Table 1.1 presents the studies of SF1, U2AF65 and their binding partners.

Table 1.1 Structural studies of SF1, U2AF65 and their binding partners.

Factors	Organism	Reference
SF3b155	human	(Bessonov et al., 2010; Cass and Berglund, 2006)
SF3b155	human	(Spadaccini et al., 2006)
SF1 KH-QUA2	human	(Liu et al., 2001)
U2AF65 UHM	human	(Selenko et al., 2003)
U2AF35	human	(Kielkopf et al., 2001)
SPF45	human	(Corsini et al., 2007)
Prp40p WW-WW	yeast	(Wiesner et al., 2002)
FBP21 WW-WW	human	(Klippel et al., 2011)
SF3b	human	(Golas et al., 2003)
U2AF65 RRM12	human	(Mackereth et al., 2011)
SR45	human	(Day et al., 2012)

1.1.3 Structural studies of Complex E

1.1.3.1 Complex E

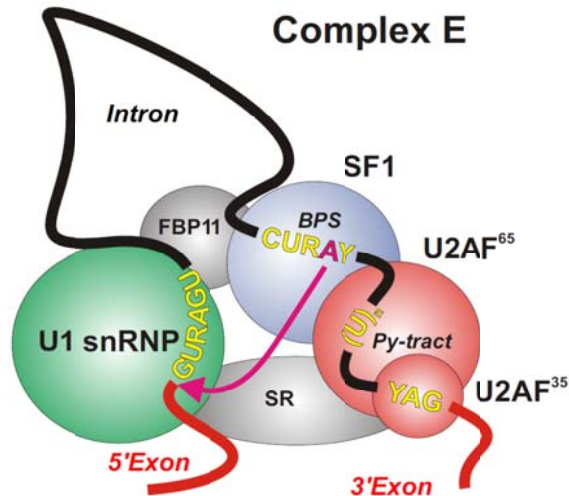


Figure 1.3 Arrangement of proteins and pre-mRNA in complex E (adapted from P. Selenko & M. Sattler). The 5' splice site is specifically recognized by the U1 snRNP (green), while the U2AF heterodimer (U2AF65/U2AF35, red) interacts with the Py tract and the 3' AG, respectively. SF1/BBP (blue) binds to the BPS further upstream. Within in the BPS the BPA (magenta) acts as the nucleophile attacking the phosphate group at the 5' splice site in the first trans-esterification reaction. Protein-protein interactions bridging components at the 5' and 3' splice sites are shown in grey.

The formation of Complex E, as the initial stage in the spliceosome assembly, is an early step in the regulation of splicing. Choices of alternative splicing are initially controlled at the stage of splicing sites definition in the Complex E. In this complex, the 5' splice site is bound with U1 snRNA by base pairing. Non-snRNP proteins simultaneously interact with the 3' splice site. Splicing Factor 1 or Branchpoint Binding Protein (SF1/BBP) recognizes the branch point sequence (BPS), and the two subunits of the heterodimeric U2 Auxiliary Factor (U2AF65 and U2AF35) bind to the polypyrimidine (Py) tract and the 3' splice site, respectively. In mammals, these sequence elements are arranged in a typical way, that the 3' end of the intron was defined by the AG dinucleotide (3' AG), the Py tract located at five nucleotides upstream and extending for ten or more nucleotides, and the BPS generally located upstream to Py tract extending to 30 nucleotides.

The serine-arginine-rich (SR) protein interacts with exon sequences and contributes to complex E formation (Chiara et al., 1996). Exon definition, which involves the SR protein-mediated bridging of 3' and 5' splice sites across the intervening exon (Berget, 1995; Smith and Valcarcel, 2000), may facilitate the recognition of splice sites in long introns (> 1000 nt). Furthermore, cross-intron bridging interactions involves SR protein-mediated contacts between the U1 snRNP at the 5' splice site and U2AF at the 3' splice site (Wu and Maniatis, 1993), and an interaction between the U1 snRNP-associated splicing factors Prp40/FBP11 and SF1 (Abovich and Rosbash, 1997; Reed, 2000). In complex E, all essential pre-mRNA sequences are within close spatial proximity, and the activities of the U1 snRNP, U2AF and SR proteins induce this conformation (Kent and MacMillan, 2002).

1.1.3.2 5' splicing site

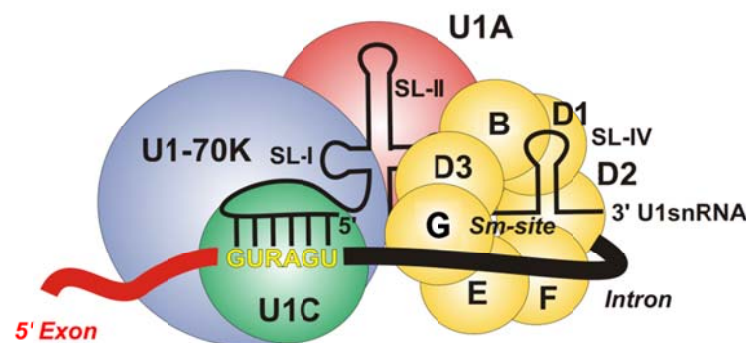


Figure 1.4 Schematic representation of the structural arrangement of protein and RNA components of the U1 snRNP at the 5' splice site (adapted from P. Selenko & M. Sattler). Exon/intron RNA is depicted as a thick red/black line. Base pairing of the 5' end of U1 snRNA (black) with conserved 5' splice site nucleotides is indicated. Protein components of the U1 snRNP and their interactions with each other, the U1 snRNA and the intron are illustrated.

In Complex E, the structure of the U1 snRNP involved in 5' splice site recognition has been reported recently (Pomeranz Krummel et al., 2009). The single-stranded 5' terminal sequence of snRNA in the U1 snRNP's base pairs with intron nucleotides at the 5' splice site in an anti-parallel fashion (Will and Luhrmann, 2001). Structural studies reveal that the U1 snRNA comprises four stem-loops (SL-I, II, III, and IV). SL-I, II, and III are in a four-way helical form and separated from the 3' terminal SL-IV by the canonical, single-stranded Sm-binding site sequence, where is recognized by the seven Sm or core proteins (B/B', D1, D2, D3, E, F,

and G). These Sm proteins are common in all snRNPs and assemble around the Sm-binding site to form the core snRNP. In addition, mammalian U1 snRNP contains four U1-specific proteins (U1A, U1C, and U1-70K) (Figure 1.4), that U1-70K and U1A bind to SL-I and II, respectively, U1C incorporation into the snRNP requires its interaction with U1-70K (Figure 1.4).

Interaction of the U1 snRNP with proteins that associate with the pre-mRNA in the vicinity of the 5' end of the intron additionally contributes to the 5' splice site definition. For instant, binding of U1 snRNP to protein TIA-1 is essential for U1 snRNP recognizes 5' splice sites followed by uridine (U)-rich sequences (Del Gatto-Konczak et al., 2000; Forch et al., 2000). These U-rich sequences are also bound by proteins that would otherwise interact with pyrimidine-rich sequences (i.e. the Py tract) close to the 3' splice site. For example, U2AF65 and the *Drosophila* female specific sex lethal (Sxl) protein, which are involved in 3' splice site selection, are claimed to interact with these U-rich stretches and directly compete with TIA-1 (Forch et al., 2001). In addition, U2AF65 RS domain is proposed to be a critical factor to promote U1 snRNP association with 5' splice sites followed by U-rich sequences (Forch et al., 2003).

Combining biochemical data with known three-dimensional structures of the individual U1 snRNP components, the structural arrangement of proteins and the U1 snRNA was determined. Furthermore, more than a decade ago, electron microscopy was used to determine the fully assembled human U1 snRNP structure at 10 Å resolution (Stark et al., 2001). And crystal structures of U1 snRNP are released recently (Pomeranz Krummel et al., 2009; Weber et al., 2010), showing the core RNP presents multiple attachment sites for the U1-specific 70K protein. Each of the seven Sm proteins recognize one nucleotide of the Sm site RNA. Proteins D1 and D2 move the snRNA through the Sm ring, and proteins F and E mediate a direct interaction between the Sm site termini. Terminal extensions of proteins D1, D2 and B/B', and extended internal loops in D2 and B/B' support a four-way RNA junction and a 3' terminal stem-loop on opposite sides of the Sm core RNP, respectively (Weber et al., 2010).

1.1.3.3 3' splice site

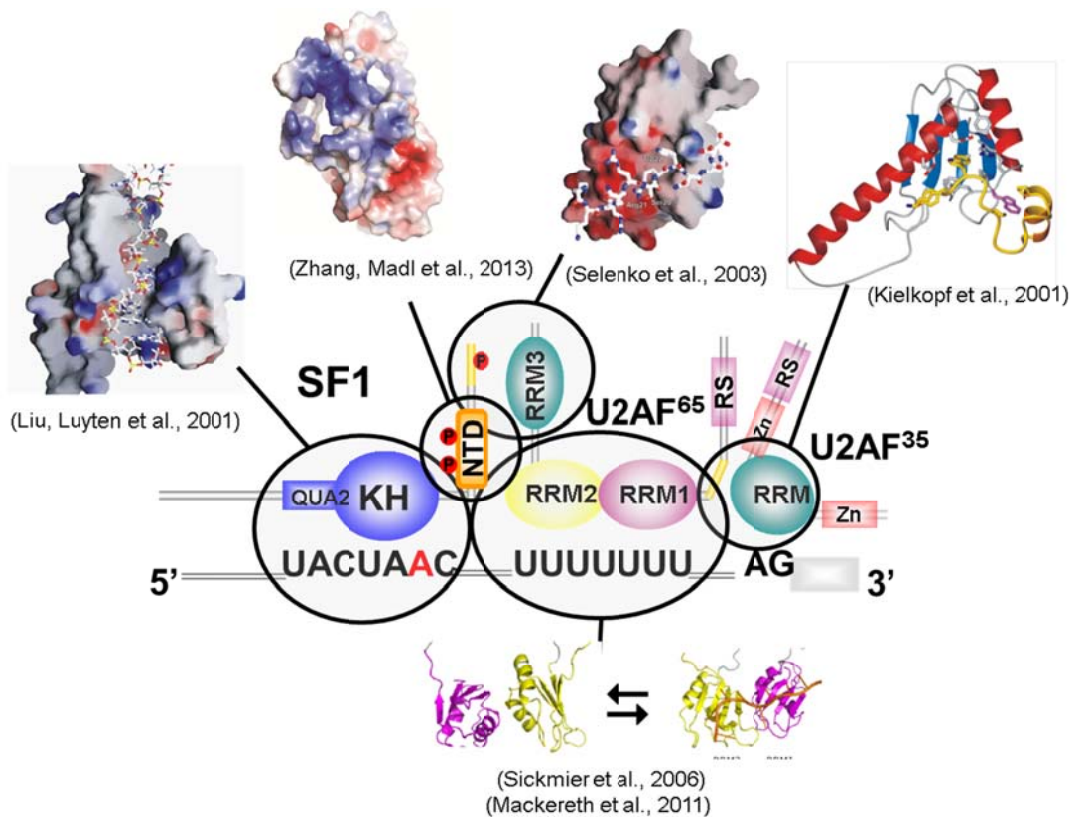


Figure 1.5 Schematic overview and topology of molecular interactions that define the 3' splice site. The BPS is bound by the KH-QUA2 region of SF1 (Liu et al., 2001). The N-terminus of SF1 interacts with U2AF65-RRM3, and phosphorylation of SF1 is involved in the regulation of 3' splicing site recognition (Selenko et al., 2003; Zhang et al., 2013). The RS domain of U2AF65 is located near the BPS and may interact with the RS domain of U2AF35 (not indicated in the figure) (Valcarcel et al., 1996). RRM1 and 2 of U2AF65 bind to the Py tract (Mackereth et al., 2011; Sickmier et al., 2006).

Comparing to the definition of the 5' splice site by binding of the U1 snRNP to pre-mRNA, recognition of the 3' splice site is more complex and involves more splicing factors. In complex E, 3' AG dinucleotide defines the 3' terminus of the intron, and the Py tract locates at five nucleotides upstream extending for ten or more nucleotides to reach the BPS. These three intronic elements, the 3' AG, the Py tract and the BPS, are specifically recognized by the U2AF35, U2AF65 and SF1, respectively (Figure 1.5). Hence at this stage in constitutive splicing, proteins are exclusively involved in recognition of 3' splicing site (Black, 2003; Smith and Valcarcel, 2000). In addition, protein-protein interactions between U2AF65 and U2AF35 and between U2AF65 and SF1 are proposed to promote cooperative 3' splice site

recognition in complex E and ensure high accuracy of 3' splice site definition. The phosphorylation of splicing factors, SF1 and U2AF heterodimer RS domain for instance, provides opportunities of post-translational regulations in pre-mRNA alternative splicing. The protein-protein and protein-RNA interactions in complex E are relative weak, which presumably supports the dynamic assembly of the spliceosome and facilitates the subsequent replacement of SF1 by the U2 snRNP in complex A (Figure 1.2).

A recent study has revealed that the U2AF65 tandem RRM domains 1 and 2 (RNA recognition motif, RRM1-2) recognize the Py tract, induces a population shift between "closed" and "open" conformational arrangement. They have shown that the correlation relationship between the length of the Py tract and the binding affinity of U2AF65 RRM1-2, that further indicates how the 'strength' of a given Py tract is coupled to the efficiency of spliceosome assembly. They adopted a molecular rheostat-like model to describe the equilibrium between the two conformations that quantitatively connects the natural variations in polypyrimidine tract nucleotide composition, length and functional strength to the efficiency to recruit U2 snRNP to the intron during spliceosome assembly (Mackereth et al., 2011).

SF1, one of the major factors involving in 3' splice site recognition, binds to BPS within pre-mRNA and simultaneously interacts with U2AF65 via UHM Ligand Motif (ULM) - U2AF Homology Motif domain (UHM) interaction, it has been proposed that the N-terminus of SF1 (SF1^{NTD}) comprises a helix hairpin fold (SF1^{HH}) providing an additional binding interface to U2AF65^{UHM}. More descriptions of SF1 in detail are presented in the next section.

1.1.4 Splicing factor 1 (SF1)

As described in section 1.1.2, there are two trans-esterification-reactions steps required for the pre-mRNA alternative splicing. In the first step, the 2' hydroxyl group of adenosine within the branch site sequence (BPS) attacks the phosphate at the 5' splice site, which leads to the formation of a 2'-5' phosphodiester bond in the lariat structure (Figure 1.2). Thus, this branch point adenosine (BPA) can be considered as a key nucleotide in the pre-mRNA intron. It is part of the seven-nucleotide BPS and is highly conserved in *S. cerevisiae* (UACUAAC, the underlined adenosine denotes the BPA), but more divergent in other organisms, including *S. pombe* and higher eukaryotes. In mammals, the branch point consensus sequence is YNCURAY (Y = pyrimidine; R = purine; N = any nucleotide). In the *S. cerevisiae* commitment complex, the branch point binding protein (BBP) recognizes the BPS. Whereas, in the mammalian complex E, SF1, ortholog of BBP, was reported to be involved in this interaction (Berglund et al., 1997).

The solution structure of the human SF1 KH-QUA2 domain in complex with a branch point RNA demonstrated that this region is necessary and sufficient for recognition of the BPS (Liu et al., 2001). The KH domain of SF1 consists of a three stranded anti-parallel β -sheet packed against three α -helices in the type of β 1- α 1- α 2- β 2- β 3- α 3 secondary structure. The QUA2 domain adopts an additional amphipathic α -helix (α 4) that packs on top of helices α 1 and α 3. In conjunction with the core KH domain, this entity forms a compact structure which defines an enlarged KH fold. The BPA is deeply buried in a hydrophobic pocket, formed by the β 2, α 1, α 2 and the QUA2 helix α 4 and thus shielded from potential interaction with other molecules. Therefore, the BPA base is positioned in a pre-bulged orientation that may prime it for subsequent base pairing with U2 snRNA. In addition to hydrophobic contacts, electrostatic interactions are another effect between conserved positively charged side chains in SF1 KH-QUA2 domain and the solvent-exposed phosphate backbone of the BPS RNA. Electrostatic contacts further define this adenosine-specific binding pocket.

In addition to the protein-RNA interaction, the N terminus of SF1 binds to U2AF65 UHM (or RRM3) (Rain et al., 1998), the C-terminal non-canonical amino acid sequence RRM. Due to the complex is sensitive to salt concentration, it is reasonable to suggest electrostatic effects contribute to the SF1-U2AF65UHM interaction (Rain et al., 1998). Study of the three-dimensional structure of the SF1-U2AF65UHM protein complex revealed the

molecular details underlying this interaction (Selenko et al., 2003). In the complex, SF1 is bound in an extended conformation at U2AF65 UHM helical α C surface opposite the classical RNA binding region. The conserved Trp22 of SF1 is deeply buried between helices α A and α B of U2AF65-UHM and coordinated by conserved hydrophobic residues, which provide a unique binding pocket for the Trp side chain.

Ser20 within SF1 N-terminal domain can be phosphorylated, which in turn inhibits the SF1-U2AF65UHM interaction and blocks pre-spliceosome assembly (Wang et al., 1999). From the structure, Ser20 of SF1 is solvent-exposed and accessible for potential phosphorylation. Although it cannot be excluded that phosphorylation could alter specific interactions of Ser20 (i.e. intra-molecular salt bridges with preceding positively charged residues in SF1), a phosphorylated Ser20 would not appear sterically restrained, suggesting that unfavorable electrostatic interactions induced by phosphorylation inhibits U2AF65 UHM binding. Recent researches have proposed that the phosphorylation of SF1^{NTD} at residue Ser80 and Ser82 may play regulation roles in 3' splice site recognition (Wang et al., 2013; Zhang et al., 2013). In next section, more information is discussed.

1.1.5 SF1 phosphorylation

It was long thought that protein phosphorylation is required for pre-mRNA alternative splicing and offers a dynamic way to regulate protein activity and stability (Olsen et al., 2006; Ptacek and Snyder, 2006; Stamm, 2008). The serine/arginine-rich (SR) proteins are especially interesting because their phosphorylation-dependent control of spliceosome assembly, in which phosphorylation of arginine-serine domains regulates the localization and interactions between SR proteins and the pre-mRNA (Ghosh and Adams, 2011). In addition to SR proteins, phosphorylation of other classes of splicing factors begins to be understood at the molecular level. For example, the SF3b155 subunit of the U2 snRNP interacts with U2AF65 and it has been suggested that the SF3b155 subunit can be phosphorylated at the region adjacent to the binding interface (Boudrez et al., 2002; Seghezzi et al., 1998). SF3b155 phosphorylation has been proved to promote U2AF65 dissociation from the activated spliceosome (Bessonov et al., 2010; Shi et al., 2006; Wang et al., 1998).

SF1 and U2AF65 cooperatively recognize pre-mRNA at 3' splice site. Additionally, N-terminal domain of SF1 binds to U2AF65 C-terminal region via a ULM-UHM interaction. It has been

discussed in section 1.1.4 that the phosphorylation of Ser20 within SF1 N-terminal domain cGMP-dependent protein kinase-I (PKG-I) inhibits the SF1-U2AF65^{UHM} interaction and blocks pre-spliceosome assembly (Wang et al., 1999). Later on, three dimensional structural information of SF1^{ULM}-U2AF65^{UHM} complex was released, showing that Ser20 of SF1 is solvent-exposed and favorably accessible for potential phosphorylation. The Ser20, phosphorylation site, is not involved in the directly binding to U2AF65^{UHM} (Selenko et al., 2003). Their studies have proposed a combination of steric interference and unfavorable electrostatic contacts of phosphorylated Ser20 with the negatively charged U2AF65^{UHM} helix A induces the inhibition of SF1-U2AF65 interaction and spliceosome assembly.

It has been shown that human SF1 is tandem phosphorylated at the two highly conserved Ser80 and Ser82 residues in proliferating human embryonic kidney cells. These two serine residues are respectively followed by a proline, which forms a SPSP pattern (Ser80-Pro81-Ser82-Pro83). The SPSP pattern is thought to ensure the specific phosphorylation on Ser80 and Ser82 (Manceau et al., 2006). Subsequent phospho-proteome analyses of HeLa (Beausoleil et al., 2004), lymphoma (Shu et al., 2004), and prostate cancer cells (Myung and Sadar, 2012) further confirm the prevalence of SF1 Ser80 and Ser82 phosphorylation and suggest a role in cancer initiation or progression. KIS kinase named from kinase interacting with stathmin (KIS), or UHMK1, was reported to specifically phosphorylate SF1 at positions Ser80 and Ser82 (Manceau et al., 2008; Manceau et al., 2006). Based on their biochemical results, they claimed that phosphorylation by KIS kinase enhances SF1-U2AF65 interactions and facilitates formation of the ternary complex among SF1, U2AF65, and the pre-mRNA (Manceau et al., 2006). Recent studies claimed that phosphorylation induces a disorder-to-order transition within a previously unknown SF1/U2AF65 interface based on crystal structures of the phosphorylated SF1 in complex with the C-terminal domain of U2AF65 and of the unphosphorylated SF1 domain (Wang et al., 2013). Whereas, the NMR solution structure of phosphorylated SF1 shows that phosphorylation rigidifies the linker connecting α -helix1 and α -helix2 within SF1^{NTD}, however it does not noticeably affect the conformation of SF1 or SF1-U2AF65 and does not modulate cooperative RNA binding (Zhang et al., 2013). Descriptions in detail are presented in the discussion section s.4.1.

1.1.6 The U2AF65 arginine-serine-rich (RS) domain

Reams of different proteins are involved in the individual splicing steps and the regulatory processes that lead for example to alternative products (alternative splicing), an essential one of them is SR protein family, which is a class of general factor that in constitutive splicing and can also modulate alternative splicing. All SR proteins share a similar bipartite structure composed of two functional domains: an N-terminal RNA binding domain, consisting of several RNA recognition motifs (RRMs) and a C-terminal arginine- and serine-rich RS domain. The RRM is sufficient for sequence-specific RNA binding, whereas the RS domain is believed to be an essential mediator and regulator in pre-mRNA alternative splicing.

Several key interactions with components of the splicing machinery were already highlighted and indicate that RS domains serve as a dynamic and versatile binding platform for proteins, RNA, and ligands. RS domains promote for example protein-protein interactions that facilitate recruitment of the (pre)spliceosome (Kohtz et al., 1994; Wu and Maniatis, 1993) and contact the pre-mRNA directly via the Branch Point and the 5' splicing site (Hertel and Graveley, 2005; Shen et al., 2004; Shen and Green, 2004). Furthermore, they participate in many other cellular processes such as in mRNA nuclear export, nonsense-mediated mRNA decay and mRNA translation (Caceres et al., 1997; Long and Caceres, 2009).

In vivo, RS domains are found to be extensively phosphorylated on serine residues enabling post-translational modification of the interaction network. It is believed that the subcellular localization as well as the activity is regulated on the level of phosphorylation (Lin and Fu, 2007).

As RS domains are involved in several different tasks, disturbance of the protein integrity and interactions by mutation lead to human disease. Evidence emerged that defects of SR proteins are related with cancer, SMA (spinal muscular atrophy) and HIV (Long and Caceres, 2009).

RS domains are not only detected in SR proteins, but also exist in other splicing factors (i.e. essential splicing factors U2AF65/U2AF35 (Valcarcel et al., 1996)) and play significant roles in constitutive and regulated splicing (Black, 2003; Graveley, 2000). U2AF65 and U2AF35 bind to the polypyrimidine (Py) tract and 3' AG splice site, respectively, and initiate

spliceosome assembly by promoting the interaction between U2 snRNP and the branchpoint. It has been shown that the binding of U2AF65 to the Py tract direct the RS domain to contact BPS and promotes base pairing with U2 snRNP (Nolen et al., 2001; Shen et al., 2004).

Despite all the available functional data, detailed insight into the molecular mechanisms remains elusive. This is probably due to a common feature of all RS domains: although they share a high content of arginine and serine residues, they seem to be intrinsically unstructured and highly dynamic. Furthermore, inclusion of phosphorylation in the structural studies is crucial in order to understand how the binding-strength and -specificity of RS domains for their interaction partners is modulated.

Therefore, in this thesis, the molecular details of the U2AF65 RS domain and its interactions were studied in solution using Nuclear Magnetic Resonance (NMR) Spectroscopy. The main focus will be on the RS domain of the essential splicing factor U2AF65 which acts at a central position of the splicing pathway and interconnects RNA and proteins(Valcarcel et al., 1996), potentially in a phosphorylation-dependent mechanism (Olsen et al., 2006). As described in section 1.1.3.3, U2AF65 RS domain forms a ternary protein-RNA-protein complex with SF1KH-QUA2 domain at the branch point. Therefore, together with our structural studies of SF1, the structural insight for RNA- and protein interaction by RS domains will contribute significantly to understanding the mechanisms of 3' splice site recognition and its regulation.

1.2 Nuclear magnetic resonance (NMR) spectroscopy

1.2.1 The principles of NMR

Nuclear magnetic resonance, or NMR, occurs when the nuclei of certain atoms are immersed in a static magnetic field and affected by an additional oscillating magnetic field. Some nuclei experience this phenomenon and others do not, dependent upon whether they possess an intrinsic property called spin. The spin is defined by I , the fourth quantum number, for any given wave function, which is obtained from the Schrödinger equation:

$$H\Psi = E\Psi$$

where Ψ is the wavefunction of the particle, H the Hamilton operator and E the energy eigenvalue for the system.

I can be zero or can have integral/fractional values. If the spin value is non-zero a nucleus, that represents a spinning charge in terms of classical mechanics, is associated with a magnetic moment. Only the spins of nuclei which have non-zero I -values can be detected by a NMR spectroscopy. The ones that are common used in biological NMR are ^1H , ^2H , ^{13}C , ^{15}N and ^{31}P , because these nuclei are either present in biological molecules in high amounts or can be relatively easily enriched by incorporation through different uniform or selective labeling techniques.

Stable isotopes ^1H , ^{13}C , ^{15}N and ^{31}P are spin $I = \frac{1}{2}$ nuclei. The interaction of these nuclei with an external magnetic field gives rise to two different spin states with a certain energy level separated by a discrete energy difference, characterized by $I = +\frac{1}{2}$ and $I = -\frac{1}{2}$. The magnetic moment of the lower energy state $+\frac{1}{2}$ is aligned with the external magnetic field; the one of the higher energy with $-\frac{1}{2}$ spin state is opposed to the external magnetic field (Figure 1.6). The equation that describes the energy of each state in the conventional laboratory coordinate system is

$$E = -\gamma I B_0 = -\gamma m \hbar B_0$$

where γ is the gyromagnetic ratio; \hbar is the reduced Planck constant; B_0 is the strength of the external magnetic field

The energy difference correlates with a distinct frequency, the Larmor frequency, which depends on γ and B_0 as given by

$$\Delta E = \gamma \hbar B_0 = m \hbar \omega_0$$

where ω_0 is the Larmor frequency

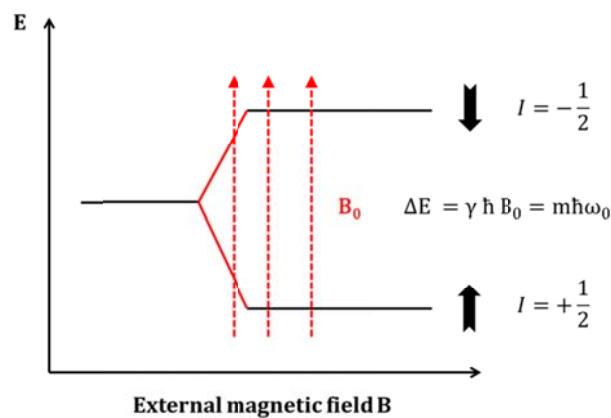


Figure 1.6 Two energy levels for nuclei with $I = \frac{1}{2}$ in an external homogeneous magnetic field. I is the nuclear spin, B is the strength of external magnetic field, γ is the gyromagnetic ratio of the nucleus, h is the Planck's constant and \hbar is the Planck's constant divided by 2π . The energy term is indicated as E . In an external magnetic field, two energy levels exist for I nuclei with a distinct energy difference ΔE . Black arrows indicate the alignment of the spins with respect to the vector of external magnetic field (B_0 , red arrow).

The population of the energy states is given in the Boltzmann equation

$$\frac{N_{\text{higher}}}{N_{\text{lower}}} = e^{-\Delta E/kT}$$

Here N_{higher} and N_{lower} represent the number of spins on the higher and lower energy levels respectively and ΔE is the energy difference between the two levels, T is the temperature in Kelvin and k represents the Boltzmann constant.

The sum of a large ensemble of spins in the external magnetic field can be described as a macroscopically observable equilibrium magnetization M in the vector model. In the presence of an external magnetic field B_0 , the orientation of vector M in respect to a coordinate system in which the orientation of the external magnetic field corresponds to the z axis. Without additional external influences, the magnetic moment of the spins is in equilibrium. They can align with or opposite to the external magnetic field B_0 with the distribution of the two spin states follows Boltzmann equation. Whereas, the numbers of spins on the higher (N_{higher}) and lower energy (N_{lower}) levels are not equal, thus the difference between N_{higher} and N_{lower} creates a net bulk magnetization, represented as a vector M , along the direction of the magnetic field. An additional electromagnetic field B_1 in the transverse plane can affect the orientation of M . In an NMR spectrometer, an radiofrequency (rf) pulse along the x -axis turn M towards the $-y$ -axis. The angle α will depend on the length of the pulse. The direction of the rotation follows the right-hand rule based on physics of electromagnetism. The magnetization will start a precession around the z -axis or external magnetic field B_0 with the Larmor frequency ω , generating the free induction decay (FID) in NMR detection coil. The measured FID is a time-domain signal and contains all the frequencies the sample in a sum of cosine waves. Fourier Transformation (FT) is able to transfer the time-domain FID into the frequency-domain signal, which is the output spectrum in an NMR experiment.

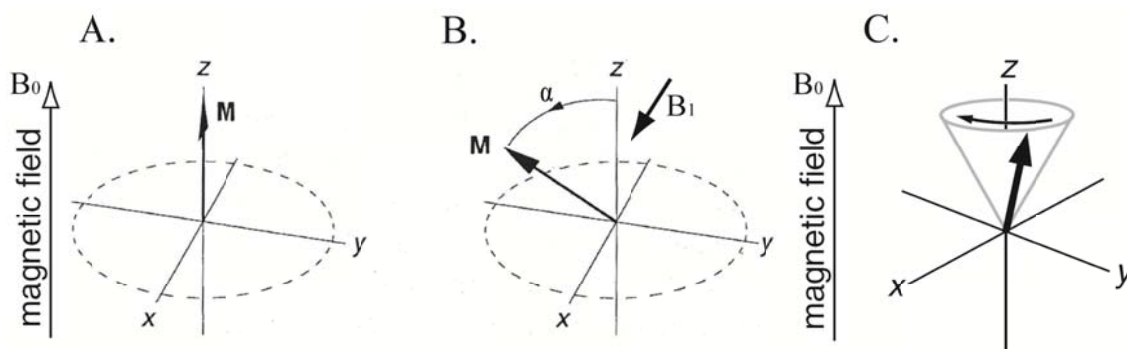


Figure 1.7 The vector model of NMR. (A) The bulk magnetization vector M with an orientation in respect to the orientation of the external magnetic field B_0 . **(B)** Effect of a radiofrequency pulse B_1 along x on the magnetization vector M . α is the flip angle, enforced by the pulse. Adapted from (Hore et al., 2001) **(C)** A precession of the bulk magnetization vector M around the z -axis, collinear with the external magnetic field B_0 , after a radiofrequency pulse. Adapted from (Keeler, 2005).

The gyromagnetic ratio of certain nuclei such as ^1H and ^{15}N decides their Larmor frequencies. In a magnetic field strength of 14.1 Tesla, ^1H precesses at a Larmor frequency of 600 MHz and ^{15}N with 60 MHz. The energy difference between the two spin states is dependent not only on the strength of the external magnetic field but also on the local magnetic field experienced by the spins, therefore, nuclei of the same type within a molecule possess distinct resonance frequencies. The variation of the spin resonance frequencies of the same kind of nucleus due to the variations in the local chemical environment is called the chemical shift (CS). These changes in frequency are relatively small and given in units of ppm (parts per million).

$$\delta_{ppm} = \frac{\nu_{exp} - \nu_{ref}}{\nu_{ref}} \times 10^6$$

where ν_{exp} and ν_{ref} are the resonance frequencies obtained for nucleus in the experiments and for the reference component respectively. In protein NMR, the reference component is usually the tetramethylsilane, TMS.

1.2.2 NMR for protein analysis

The application of NMR to study structure and dynamics of biomolecules (i.e. proteins, nucleic acids) on atomic level is one of the greatest achievements in the last century. On the way towards the successful NMR measurement of proteins, two main limiting factors should be taken in account, one is the signal overlap and the other is the fast transverse relaxation rates. Proteins contains a large number of nuclei of the same atom type, thus, in the early protein NMR experiments, measurements of the naturally abundant NMR-active isotope ^1H in one-dimensional spectra can only provide limit information due to lack of resolution of the signals. In addition, the large size of proteins and nucleic acids causes fast transverse relaxation rates, that broadening the line-width of the signal. Developments in both molecular biology field as well as NMR technology have made a remarkable progress in overcoming these problems. The naturally abundant isotope ^{12}C and ^{14}N can be respectively replaced by the NMR-active nuclei ^{13}C and ^{15}N , by expressing proteins in corresponding isotopic minimal medium. The recombinant expression of proteins in bacteria enables the ^{13}C and ^{15}N labeled protein purified and enriched. These samples provide a way for develop the multidimensional heteronuclear NMR method, which correlates two or more nuclei NMR signals via J-coupling and dipolar-coupling. 3D NMR generates extremely higher signal resolution then 1D NMR. The proton density, a major source of relaxation, can be reduced by expressing protein in a partially deuterated or perdeuterated background and thus large proteins are accessible for NMR analysis. Specific labeling schemes enable specifically isotope labeling on certain amino acids or chemical groups, contributing to the simplification of NMR spectra. At the same time, the progress in hardware of NMR spectrometers increases magnetic field strength, in additional, cryoprobes enhance the sensitivity of the machines. These ongoing advances in NMR extend the research of biomolecules sizes larger than 300 kDa.

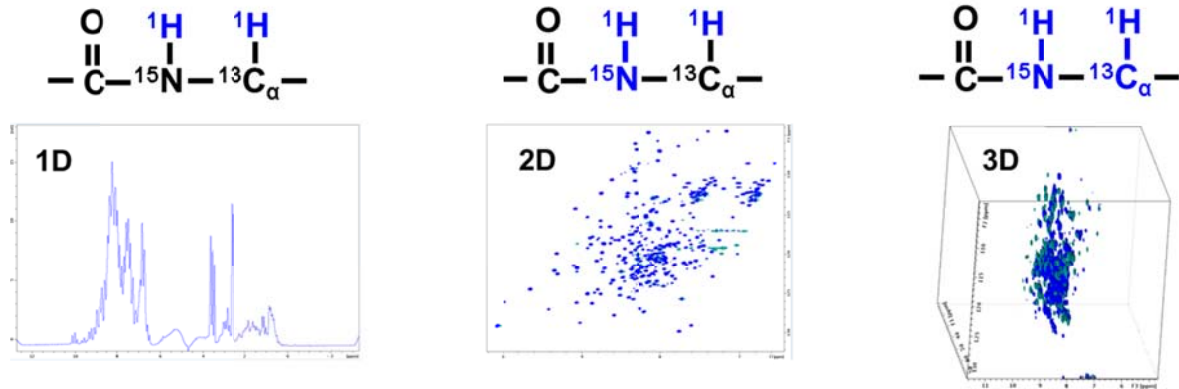


Figure 1.8 Protein NMR spectra. In 1D spectrum (left), only ^1H is detected, which results in signal overlap. The signal overlap can be reduced by heteronuclear multidimensional NMR experiments. The “fingerprint” ^1H , ^{15}N HSQC spectrum (middle) detects amide protons and is characteristic for a protein. ^{15}N , ^{13}C labeled samples are used for triple resonance experiments like an HNCA (right) for backbone assignments. The schematic of the nuclei (blue) used in the experiments are shown above the spectra. The measurement is for phosphorylated SF1.

Although a one-dimensional ^1H spectrum of a protein is far too complex for interpretation due to heavily signals overlap, lineshape and dispersion of the signals still already permit first analysis of the protein structure. For example, the presence of the signals below 0.5 ppm arising from the protons of the aliphatic methyl groups can indicate a folded protein. One of the most important two-dimensional heteronuclear NMR experiments for the proteins is the ^1H , ^{15}N -HSQC (heteronuclear single quantum correlation spectrum) correlating the nitrogen atom of an NH_x group with the directly attached proton. The backbone amide of every amino acid (with the exception of proline) gives rise to one signal in the spectrum. The ^1H , ^{15}N HSQC spectrum is often referred to as the “fingerprint” spectrum, since it is unique for every protein (Figure 1.8).

The sequential assignment of the individual backbone atoms requires a set of triple resonance experiments (Sattler et al., 1999). In these experiments, the frequency of the amide proton is correlated with those of the scalar coupled carbon atoms of the amino acid, and the magnetization of the amide proton can be transferred to the CO, C_α or C_β of the same residue or the previous one in the amino acid sequence. The information obtained in these experiments allows for sequential backbone assignment of a protein. For the assignment of the side chain atoms, NMR TOCSY experiments are applied to correlate the ^1H and ^{13}C frequencies of all atoms within a residue.

The Nuclear Overhauser Effect (NOE), which transfers polarization by dipolar interactions between atoms close in space, is one of the most important effects in NMR. The NMR structure calculation is based on the distance restraints derived from many hundreds of NOE-peaks. In isotope-edited NOESY-HSQC experiments, magnetization is transferred to protons nearby and the HSQC sequence selects protons bound to either ^{15}N or ^{13}C , which can provide sufficient resolution. Information from backbone and side-chain experiments is combined to obtain complete resonance assignments of proteins.

1.2.3 Protein dynamics by NMR

NMR of proteins allows for studying not only the structure but also the dynamic properties of proteins. For investigation of protein dynamics, two relaxation pathways are usually monitored and the rate constants are measured. The longitudinal relaxation (also called the spin-lattice relaxation), characterized by a single time constant T_1 or the corresponding relaxation rate $R_1 = \frac{1}{T_1}$, is induced by the interaction of the spins with their surrounding (the lattice). The transverse relaxation (also called the spin-spin relaxation), described by a single time constant T_2 or the relaxation rate $R_2 = \frac{1}{T_2}$, is caused by the interaction of the nuclear spins.

Measurements of ^{15}N relaxation provide information about backbone dynamics covering a picosecond to second timescale. Dynamics on the pico- to microsecond timescale like molecular tumbling and internal motion are reflected by T_1 . T_2 is additionally influenced by dynamics in the micro- to millisecond timescale and thus sensitive to chemical exchange. The correlation time τ_c describes molecular tumbling and can be calculated from the ratio between both relaxation rates including information about the ^{15}N frequency. In ^1H - ^{15}N heteronuclear NOE experiments, motions of distinct N-H bond vectors are sampled. Those that undergo motion faster than the overall tumbling of the molecules show a decreased NOE intensity relative to the average observed for the majority of the residues. The average value for the ^1H - ^{15}N heteronuclear NOE ratio is approximately 0.77, higher values indicating that the N-H vector is rigid with respect to the rest of the protein and lower values for parts with increased backbone mobility (Kay et al., 1989).

1.2.4 Paramagnetic relaxation enhancement (PRE)

In the 1950s, Solomon and Bloembergen demonstrated that the presence of unpaired electrons in the molecular system enhances the magnetic field modulation at a nucleus and nuclear spin relaxation due to nucleus dipolar couplings. This effect is named paramagnetic relaxation enhancement (PRE) (Solomon, 1955; Solomon and Bloembergen, 1956). In protein NMR experiments, PREs are generated by spin labeling, attaching nitroxide stable radicals such as iodoacetamido proxyl to the protein for example. Compared to NOE (similarly generated by dipolar couplings, the distance between two spins limited up to 6 Å), PREs can extend the effects to a distance of more than 40 Å (Otting, 2008). Therefore, PRE is increasingly becoming a method to provide long-range distance information that can complement NOE restraints. Distances between the spin label and nuclei can be determined from the increased R2 relaxation rates.

In the protein NMR, PREs can be monitored on the HN-resonances in the ^1H , ^{15}N HSQC spectra. The intensity of the peaks is acquired from the spectra in the presence of a paramagnetic spin label (or oxidized state) and diamagnetic analogue (or reduced state). Then, PRE represented in R2 relaxation rates can be calculated using equation (Battiste and Wagner, 2000):

$$\frac{I_{ox}}{I_{red}} = \frac{R_2 e^{-R_2^{PRE}t}}{R_2 + R_2^{PRE}}$$

where I_{ox} and I_{red} are the peak intensities corresponding to amides in the $^1\text{H}^{15}\text{N}$ -HSQC for a sample with a spin label in the oxidized and the reduced state respectively, R_2 is the R_2 relaxation rate for each amide ^1H , t is the total evolution time of the HSQC and R_2^{PRE} the paramagnetic rate enhancement coming from the spin label.

R_2^{PRE} were converted into distances by use of the following equation for the effect of paramagnetic spins on nuclear magnetic relaxation

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

Where r is the distance between the electron and nuclear spins, τ_c is the correlation time for this electron-nuclear interaction, ω_h is the Larmor frequency of the nuclear spin (proton), and K is the physical constant $1.23 \times 10^{-32} \text{cm}^6 \text{s}^{-2}$, dependent on the type of the nucleus. For distance calculations, an approximation can be made that τ_c equals the global correlation time of the protein. Notably, the magnitude of PREs is proportional to the distance r and declines with r^6 . Thus, the r calculated from PREs can be used as distance restraints for the structure calculation, especially for the large proteins.

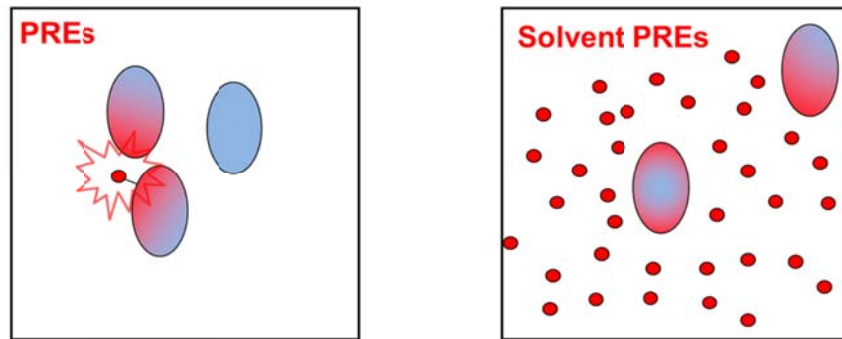


Figure 1.9 Paramagnetic relaxation enhancement. Schematic representation of paramagnetic relaxation enhancement effects (red) on proteins (gray ellipse) with (A) the attachment of a paramagnetic spin label to a fixed position in the protein or (B) a paramagnetic environment (solvent PREs).

The spin label first explored for the protein NMR study was a Fe^{2+} as a natural paramagnetic ligand for metal binding proteins (Weiner, 1986). The site-specific attachment of paramagnetic ions to proteins was developed and allowed common spin labels, such as iodoacetamido proxyl, to attach to the protein via cysteine residues (Su and Otting, 2010). Gd(DTPA-BMA) exerts strong PRE effects and can thus be used for measuring solvent PREs, which can provide structural information of protein surfaces and protein-protein interaction interfaces (Figure 1.9). For this approach, R_1 are measured in the presence of increasing Gd(DTPA-BMA) concentrations. Solvent PREs can be included in structure calculations or used to validate a structure by comparing experimental with back-calculated values (Madl et al., 2009).

1.2.5 NMR studies for protein-ligand interaction.

NMR titration, recording a series of ^1H , ^{15}N HSQC spectrum of the protein in presence of increasing ligand concentrations, is a common tool to investigate protein-ligand interactions such as protein-protein and protein-nucleic acid interactions. Chemical shift perturbations provide information of the amide proton signals of involved residues. The changes observed in the spectrum depend on the exchange rates between free and ligand-bound state. The exchange constant k_{ex} is

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

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

On the NMR timescale,

$k_{ex} \gg \Delta\omega$ fast exchange

$k_{ex} \approx \Delta\omega$ intermediate exchange

$k_{ex} \ll \Delta\omega$ slow exchange

where $\Delta\omega$ is the difference in frequencies of the same resonance in the free and bound state.

When ligand binding in fast exchange, signals possess an average chemical shift between free and bound state depended on the population of each state, thus, the position of the peak is gradually shifted until saturation. Intermediate exchange induces line broadening and thus decreased signal intensity. In a slow exchange regime, the population of the free and bound state can be observed simultaneously, giving rise to two separate signals with intensities corresponding to the percentage of free and bound protein. Thus, NMR titrations allow for a quick and residue-specific determination of the protein-ligand interaction interface.

1.2.6 Structure calculation

Combination of information, such as chemical shift assignments, distance restraints and torsion angles are used for NMR structure calculation. The chemical shift already contains useful structural information. $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ resonances of a specific amino acid have characteristic values given for a random coil environment. The chemical shift of $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ altered in structured regions, the difference from the $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ values of structured region to the random coil values is named secondary chemical shift. In an α -helix, $^{13}\text{C}\alpha$ carbons shift to downfield and $^{13}\text{C}\beta$ values move to upfield, whereas, opposite effects are observed for carbon atoms in a β -sheet. Therefore, the secondary chemical shift for α -helical form is indicated in positive values and for β -sheet conformation is shown in negative values. The distance restraints are derived from the signal intensity of cross-peaks in NOESY spectra by program, such as CYANA3.0 (Guntert, 2009). Torsion angles of each residue in protein can be predicted by the program, TALOS+ (Shen et al., 2009), using the chemical shifts of atoms from the protein backbone. Nowadays, additionally restraints obtained from paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC) measurements are employed to determine the relative position of structural elements within the molecule.

NMR structure calculation is based on a restrained molecular dynamics simulation using a target function which combines different potential energy functions. The target function includes energy terms for known parameters as bond lengths, angles, van der Waals- and electrostatic forces, the so-called force field. Additionally, energy functions for NMR data such as distance restraints from NOEs and torsion angles derived from chemical shift and sequence information are incorporated. The program starts with approximately 100 randomized structures at high temperature. In the following steps of the calculation the temperature is lowered and the different restraints are varied randomly to reduce violations and minimize the target function in a simulated annealing process. Here are some of the most used NMR structure calculation programs. Ambiguous Restraints for Iterative Assignment (ARIA)(Nilges et al., 1997), Combined assignment and dynamics algorithm for NMR applications (CYANA) (Schmucki et al., 2009) and CNS/ARIA (Linge et al., 2003). The root mean square coordinate deviation (RMSD) describes the precision of the output structure between the structures in the final ensemble. Accuracy of the structure is judged

by the absence of substantial violations for distance restraints. Structure validation will be done in the program like the iCING server (Doreleijers et al., 2012).

1.3 Scope of the Thesis

Removal of introns from pre-mRNAs is a key step in mammalian gene expression performed by the spliceosome, a large and dynamic ribonucleoprotein particle. Alternative splicing is essential to generate different proteins from a given primary transcript by differential inclusion or exclusion of coding sequences (exons) (Nilsen and Graveley, 2010). Given the importance of splicing for gene expression, numerous diseases have been linked to aberrant splicing and mutations within the consensus sequences at the 5' and 3' ends of introns interfere with spliceosome assembly (Cooper et al., 2009). Recently, frequent missense mutations in genes of splicing factors that mediate 3' splice site recognition, such as splicing factor 1 (SF1), and the U2 snRNP auxiliary factor (U2AF) large (U2AF65) and small subunits (U2AF35) have been linked to tumorigenesis (Yoshida et al., 2011). The regulation of splicing by extracellular signals and signal transduction cascades is poorly understood. With respect to 3' splice site recognition phosphorylation of conserved serine residues (Ser20, Ser80, and Ser82) in SF1 have been implicated in regulating the stability of the ternary SF1-U2AF65-RNA complex *in vitro* (Manceau et al., 2006; Wang et al., 1999). However, both the structural and functional roles of SF1 phosphorylation remain elusive. In addition, it has been proposed that binding of U2AF65 RRM1-2 to Py trace drives RS domain to contact BPS and SF1^{KH-QUA2}, inducing a compact SF1-U2AF65^{RRM123}-RNA conformation, whereas, none of structural details of U2AF65 RS domain were described.

The aims of this thesis are to study molecular and structural details involving SF1 and U2AF65 as key proteins in the early complex E during spliceosome assembly. When this project was started structural information of SF1 was limited to KH-QUA2 and N-terminal ULM regions, determined by Sattler's group (Liu et al., 2001; Selenko et al., 2003). The structural details of SF1^{HH} domain between KH-QUA2 and ULM is unknown. The molecular weight of SF1^{HH} is 13.7 kDa, which is ideal size for protein NMR. Therefore, solution state NMR was chosen to study SF1^{HH}. Structural roles of tandem phosphorylation of a SPSP (Ser80-Pro81-Ser82-Pro83) motif in SF1^{HH} remain unclear. Mass Spectrometry was applied to characterize the specifically phosphorylated SF1^{HH} and NMR was selected to investigate the conformational changes upon phosphorylation. For structural and biochemical Analyses of novel interactions in the SF1-U2AF65-RNA complex, which molecular size is large for NMR spectroscopy. Therefore, by combining NMR and Small-Angle X-ray Scattering (SAXS),

structural information of the ternary complex was explored. 3D molecular structure of highly dynamic U2AF65 RS domain is hard to be determined by X-ray crystallography. Thereby, NMR and the methods studying molecular dynamics, like relaxation time analysis, were employed to reveal the structural details of U2AF65 RS domain.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals and enzymes

Table 2.1 Chemicals

Chemical	Supplier
¹³ C labeled Glucose, 98% pure	CIL
² H, ¹³ C labeled Glucose, 98% pure	CIL
¹⁵ N labeled Ammonium chloride, 99% pure	euriso-top
3-(2-Iodoacetamido)-PROXYL	Sigma-Aldrich
30% Acrylamide/Bis-acrylamide (37.5:1)	SERVA
40% Acrylamide/Bis-acrylamide (19:1)	SERVA
Adenosine 5'-triphosphate (ATP) disodium salt hydrate	Sigma-Aldrich
Agarose	VWR
Ammonium peroxodisulfate (APS)	Merck
Ampicillin	SERVA
Ascorbic acid	Sigma
Bacto tryptone	Merck
Biotin	AppliChem
Bromophenol blue	Sigma-Aldrich
Cobalt(II) chloride hexahydrate	Roth
Copper(II) chloride dihydrate	Merck
Cytidine 5'-triphosphate (CTP) disodium salt	Sigma-Aldrich
D2O	euriso-top
Desoxyribonucleotides	Promega
Disodium hydrogen phosphate	Merck
Dithiothreitol (DTT)	Roth
DMSO	euriso-top
Ethanol absolute	VWR
Ethylendiamintetra-acetate (EDTA)	Roth
Glycerol	VWR
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	CALBIOCHEM
n-Hexanol	Fluka
Imidazole	Merck
Isopropyl b-D-thiogalactoside (IPTG)	Roth
Kanamycin	SERVA
Magnesium chloride hexahydrate	Merck
Manganese(II) chloride hexahydrate	Merck
β-Mercaptoethanol	Roth

Nickel-Nitrilotriacetate (Ni-NTA)	QIAGEN
Pefabloc SC	SERVA
Polyethylenglycol 8000 (PEG 8000)	Promega
Sodium azide	Merck
Sodium carbonate	Merck
Sodium chloride	Merck
Sodium dihydrogen phosphate	Merck
Sodium dodecyl sulphate (SDS)	Merck
N,N,N',N'-Tetramethylethylen-Diamin (TEMED)	Fluka
Thiamin	AppliChem
TritonX-100	Sigma
Trizma [®] Base (Tris)	Merck
Trizma [®] hydrochloride (Tris-HCl)	Merck
Uridine 5'-triphosphate (UTP) trisodium salt hydrate	Sigma-Aldrich
Urea	VWR
Xylene cyanole	Sigma-Aldrich
Yeast extract	Merck

Table 2.2 Enzymes

Enzyme	Concentration	Supplier
Acc65	10U/ μ l	NEB
NcoI	10U/ μ l	NEB
AleI	20U/ μ l	NEB
EcoRI	20U/ μ l	NEB
DpnI	10U/ μ l	Promega
Pfu DNA polymerase	3U/ μ l	Promega
DNase I	1mg/ml	SERVA
T4 lysozyme	crystalline	SERVA
TEV protease	1mg/ml	In-house production
KIS Kinase	5U/ μ l	In-house production
T4 DNA ligase	3U/ μ l	Promega

2.1.2 Bacterial strains and vectors

Table 2.3 Bacterial strains

<i>E. coli</i> Strain	Genotype
BL21 (DE3)	F ⁻ ompT gal dcm lon hsdS _B (r _B ⁻ m _B ⁻) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
DH5α	F ⁻ , endA1, hsdR17 (rk ⁻ , mk ⁺), supE44, thi-1, recA1, gyrA96, relA1, s80dlacZ M15
XL1-Blue	endA1 gyrA96(nal ^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB ⁺ lacI ^q Δ(lacZ)M15] hsdR17(r _K ⁻ m _K ⁺)

Table 2.4 Plasmids

Plasmid	Vector	Insert	Reference
SF1 ^{ULM}	pETM-30	GST-SF1 1-25	(Selenko et al., 2003)
SF1 ^{NTD}	pETM-11	His ₆ -TEV-SF1 1-145	(Zhang et al., 2013)
SF1 ^{HH}	pETM-11	His ₆ - TEV-SF1 27-145	(Zhang et al., 2013)
SF1 ^{KH-Qua2}	pETM-11	His ₆ - TEV-SF1 133-260	(Liu et al., 2001)
SF1	pETM-11	His ₆ - TEV-SF1 1-260 C171A	(Zhang et al., 2013)
SF1 S20C	pETM-11	His ₆ - TEV-SF1 1-260 C171A S20C	(Zhang et al., 2013)
SF1 T137C	pETM-11	His ₆ - TEV-SF1 1-260 C171A T137C	(Zhang et al., 2013)
SF1 N166C	pETM-11	His ₆ - TEV-SF1 1-260 C171A N166C	(Zhang et al., 2013)
SF1 S80E	pETM-11	His ₆ - TEV-SF1 1-260 C171A S80E	(Zhang et al., 2013)
SF1 S82E	pETM-11	His ₆ - TEV-SF1 1-260 C171A S82E	(Zhang et al., 2013)
SF1 S80ES82E	pETM-11	His ₆ - TEV-SF1 1-260 C171A S80E S82E	(Zhang et al., 2013)
U2AF65 ^{RS}	pETM-11	His ₆ - TEV-U2AF65 1-84	(Zhang et al., 2013)
U2AF65 ^{RRM12}	pETM-11	His ₆ - TEV-U2AF65 147-347	(Mackereth et al., 2011)
U2AF65 ^{UHM}	pETM-11	His ₆ - TEV-U2AF65 342-475	(Selenko et al., 2003)
U2AF65 ^{RRM123}	pETM-11	His ₆ - TEV-U2AF65 147-475	(Zhang et al., 2013)
U2AF65	pETM-11	His ₆ - TEV-U2AF65 1-475	(Zhang et al., 2013)
KIS kinase	pETM-11	His ₆ -TEV-UHMK1 1-419	(Zhang et al., 2013)

2.1.3 Recipes

2.1.3.1 Media and buffers for protein expression

Table 2.5 Media for protein expression

Medium	Component	Amount per liter
lysogeny broth (LB) rich medium pH=7.5	Bacto Tryptone	10 g
	NaCl	10 g
	yeast extract	5 g
	antibiotic (50mg/ml)	1 ml
¹⁵ N labeling medium	M9 medium (10x) (s. below)	100 ml
	TES (100x) (s. below)	10 ml
¹³ C ¹⁵ N labeling medium ^[1]	Glucose 20%(w/v)	20 ml
	or U-[¹³ C]-D-glucose ^[1]	2 g
	or U-[² H]-D-glucose ^[2]	2 g
² H ¹³ C ¹⁵ N labeling medium ^[2]	MgSO ₄ (1M)	1 ml
	CaCl ₂ (1M)	0.3 ml
	Biotin (1mg/ml)	1 ml
	Thiamin (1mg/ml)	1 ml
	antibiotic (50mg/ml)	1 ml

^[1] For ¹³C ¹⁵N labeling medium, glucose is replaced with U-[¹³C]-D-glucose.

^[2] For ²H ¹³C ¹⁵N labeling medium, U-[²H]-D-glucose is used instead of glucose; all components were lyophilized, dissolved in D₂O and sterile filtrated with a 0.22μM filter (Millipore)

M9 medium 10X	Mass per liter
Na ₂ HPO ₄	60g
KH ₂ PO ₄	30g
NaCl	5g
¹⁵ N NH ₄ Cl	5g

Trace element solution (TES) 100X	Mass per liter
EDTA	5g
FeCl ₃ x6H ₂ O	0.83g
ZnCl ₂	84mg
CuCl ₂ x2H ₂ O	13mg
CoCl ₂ x6H ₂ O	10mg
H ₃ BO ₃	10mg
MnCl ₂ x6H ₂ O	1.6mg

2.1.3.2 Buffers for protein purification

Table 2.6 SF1 and U2AF65 purification buffer

Buffer	Component	Concentration
lysis buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	imidazole	5 mM
	TCEP	1 mM
	glycerol	5% (v/v)
	DNase (SERVA)	1.5 µg/ml
	protease inhibitor mix HP (SERVA)	Manufacturer's instructions
	lysozyme (Roth)	spatula-tip
wash buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	imidazole	25 mM
	TCEP	1 mM
	glycerol	5% (v/v)
elution buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	imidazole	250 mM
	glycerol	5% (v/v)
TEV-cleavage buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	glycerol	5% (v/v)
	TEV	0.1 µg TEV/mg protein
	β-mercaptoethanol	40 mM
NMR buffer (Storage buffer)	potassium phosphate buffer, pH 6.5	20 mM
	NaCl	50 mM
	DTT	5 mM

Table 2.7 RS domain purification buffer

Buffer	Component	Concentration
lysis buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	imidazole	5 mM
	TCEP	1 mM
	glycerol	5% (v/v)
	DNase (SERVA)	1.5 µg/ml
	protease inhibitor mix HP (SERVA)	Manufacturer's instructions
	lysozyme (Roth)	spatula-tip
RS wash buffer	NaCl	4M
	Tris/HCl, pH 8.0	50 mM
	imidazole	25 mM
	TCEP	1 mM
	Urea	1 mM
	glycerol	5% (v/v)
elution buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	imidazole	250 mM
	glycerol	5% (v/v)
TEV-cleavage buffer	NaCl	500 mM
	Tris/HCl, pH 8.0	50 mM
	glycerol	5% (v/v)
	TEV	0.1 µg TEV/mg protein
	β-mercaptoethanol	40 mM
NMR buffer (Storage buffer)	potassium phosphate buffer, pH 6.5	20 mM
	NaCl	50 mM
	DTT	5 mM

Table 2.8 SF1 phosphorylation and purification buffer

Buffer	Component	Concentration
phosphorylation	MgCl ₂	15 mM
buffer	MES, pH 8.0	50 mM
	glycerol	25% (v/v)
	EDTA	1 mM
	DTT	5 mM
	ATP	5 mM
MonoQ loading	Tris/HCl, pH 7.2	20 mM
buffer		
MonoQ elution	Tris/HCl, pH 7.2	20 mM
buffer	NaCl	1 M
NMR buffer	potassium phosphate buffer, pH 6.5	20 mM
(Storage buffer)	NaCl	50 mM
	DTT	5 mM

All NMR samples contained 10 % D₂O.

2.2 Biochemical methods

2.2.1 Cloning

Plasmids for expressing human U2AF65^{UHM} (residues 372-475), U2AF65^{RRM123} (residues 147-475), as well as human SF1^{ULM} (residues 1-25), SF1^{HH} (residues 27-145), SF1^{NTD} (residues 1-145) and SF1 (residues 1-260), were prepared in a modified pET-M11 vector containing an N-terminal His₆ tag followed by a tobacco etch virus (TEV) protease cleavage site.

Table 2.9 PCR sample component

Component	Concentration	Volume (μl)
template plasmid	50-100ng/μl	1
primer fwd	10pmol/μl	1
primer rev	10pmol/μl	1
<i>Pfu</i> DNA polymerase (Promega)	5U/μl	0.25
<i>Pfu</i> reaction Buffer(Promega)	10X	5
dNTPs mix	10mM each	1
total volume (add H ₂ O)		50

Table 2.10 Program for QuickChange PCR

Step	Temp (°C)	Duration	Setting
Denaturation	95	2min	
Denaturation	95	30sec]
Annealing	57	30sec	18cycles
Elongation	68	10min]
Final extension	68	20min	
Final hold	10	~	

Plasmids for expressing human U2AF65^{UHM} (residues 372-475), U2AF65^{RRM123} (residues 147-475), as well as human SF1^{ULM} (residues 1-25), SF1^{HH} (residues 27-145), SF1^{NTD} (residues 1-145) and SF1 (residues 1-260), were prepared in a modified pET-M11 vector containing an N-terminal His₆ tag followed by a tobacco etch virus (TEV) protease cleavage site. Point mutations primers were designed using the QuikChange® Primer Design Program by Agilent

Technologies and purchased from Eurofins MWG Operon. Application of that technique allowed conveniently induced the point mutations into the plasmid; the whole mutated plasmid sequence was amplified in a polymerase chain reaction (PCR) and could be transformed directly into the competent cell. *Pfu* polymerase which is capable of proof reading (Lundberg et al., 1991) was used to increase the amplification fidelity. PCR sample composition and PCR program are listed in Table 2.9 and Table 2.10. The PCR products were incubated with 1 μ l *DpnI* for 1h at 37 °C. After heat inactivation at 60 °C for 60 min, 10 μ l of the sample were transformed into *E. coli* XL1-Blue cells. SF1 deletions of helix α 1 (Δ 35-68), α 2 (Δ 96-128), the complete helix hairpin (Δ 35-128), or the α 1- α 2 linker (Δ 75-90) were generated by PCR of SF1²⁻³²⁰ in pTRCHis A (Rain et al., 1998) with reverse and forward primers complementary to sequences upstream and downstream, respectively, of the desired deletion. The primers contained KpnI restriction sites and PCR products were ligated after KpnI digestion. The deleted protein sequences were thus replaced by Gly-Thr. Full-length KIS kinase (comprising the kinase domain and a UHM domain) was cloned by PCR ligation of the DNA encoding the short KIS isoform (residues 1-344) and the KIS UHM (residues 315-419). All plasmids were verified by sequencing at GATC Biotech and were purified using the Wizard® Plus SV Minipreps DNA Purification System (Promega).

2.2.2 Protein expression and NMR sample preparation

Expression plasmids were transformed into *E. coli* BL21(DE3) cells, grown in standard LB medium or in minimal M9T medium supplemented with 2 g/L [U-¹³C]-glucose and/or 1 g/L [¹⁵N]-ammonium chloride as the sole carbon and nitrogen sources (s. 2.1.3.1). SF1 (residues 1-260) and U2AF65^{RRM123} were prepared as [U-²H,¹⁵N]-labelled proteins for NMR titrations and relaxation measurements or as [U-²H,¹³C,¹⁵N] deuterated samples with methyl protonation of isoleucine, leucine and valine residues as described (Tugarinov et al., 2006) for chemical shift assignments. Protein synthesis was induced by the addition of 0.5 mM IPTG at OD₆₀₀ ~0.8. After protein expression for 16 h at 25°C cells were collected by centrifugation, lysed by sonication in the presence of lysozyme and EDTA-free *Complete protease inhibitor* (Roche Applied Science), and then re-suspended in lysis buffer (Table 2.6). The sample was loaded onto Ni-NTA chromatography resin (Qiagen), and washed with 20 column volumes of lysis buffer followed by five column volumes of wash buffer (Table 2.6), and then the sample was eluted with the elution buffer (Table 2.6). Removal of the His₆-tag with 0.1 µg TEV/mg protein protease required 12 h at 4°C. The TEV protease, the histidine-tag and uncleaved protein were removed by a second passage of the sample through Ni-NTA resin. The eluate was further purified by gelfiltration on a Superdex 75 column (GE) using sodium phosphate (pH 6.5), 50 mM NaCl, and 1mM DTT as buffer. NMR samples were concentrated to 100 to 600 µM in NMR buffer (Table 2.6).

KIS kinase was expressed in standard LB medium following the same protocol. The protein was purified at 4°C without cleavage and removal of the His₆-tag. Aliquots of the recombinant KIS kinase were stored at -80°C in buffer consisting of 50 mM MES (pH 8.0), 15 mM MgCl₂, 25% (v/v) glycerol, 2 mM DTT, and 1 mM EDTA.

His₆-tagged SF1²⁻³²⁰ and deletion mutants for gel shift experiments were expressed in *E. coli* and purified as described (Rain et al., 1998). Proteins were dialyzed against 20 mM HEPES-KOH pH 7.9, 100 mM KCl, 20% (v/v) glycerol, 0.2 mM EDTA and 0.5 mM DTT.

A 20-mer RNA containing the BPS and the Py tract (5'-UUAUACUAACAAUUUUUUUUU-3') was purchased from BioSpring GmbH, dissolved in H₂O to a final concentration of 10 mM and added to the protein samples before the NMR and SAXS measurements.

2.2.3 In vitro phosphorylation

Phosphorylation of SF1^{NTD} and SF1 for NMR analysis was performed in 10 ml phosphorylation buffer (Table 2.8) in the presence of 1 ml 10 × phos-stop (Roche), 40 µg of KIS kinase and 2 mg of substrate. 100mM ATP stock solution was prepared by dissolving ATP powder (Sigma) in kinase buffer and adjusting the pH to 8.0. Aliquots were stored at -20°C. The reaction typically required 24-48h at 30°C to fully phosphorylate both Ser80 and Ser82. The phosphorylation products were purified with a MonoQ ion exchange column (GE). The reaction products were buffer exchanged to MonoQ loading buffer (Table 2.8) and loaded onto the column. The NaCl concentration in the elution buffer was gradually increased over 40 column volumes from 0-1 M, and 1-ml fractions were collected. The phosphorylation products were separated depending on the phosphorylation state. The double-phosphorylation of Ser80 and Ser82 in phosphorylated SF1^{NTD} (pSF1^{NTD}) and SF1 (pSF1) was confirmed by SDS-PAGE, mass spectrometry (Figure 3.22) and, for isotope labelled samples, by NMR-spectroscopy (Figure 3.24).

2.2.4 Isothermal titration calorimetry

Calorimetric titrations were performed using an iTC200 microcalorimeter (MicroCal) at 25°C. The buffer used for the protein and ligand samples was 20 mM sodium phosphate (pH 6.5), 50 mM NaCl. The 200-µl sample cell was filled with a 5 or 10 µM solution of protein and the 40-µl injection syringe with 50 or 100 µM of the titrating ligand. Each titration consisted of a preliminary 0.2-µl injection followed by 20 subsequent 2-µl injections. The heat of the injections was corrected for the heat of dilution of every ligand into the buffer. At least two replicas were performed for each experiment. Binding thermodynamic models corresponding to bimolecular complex formation were fitted using routines provided by the manufacturer.

2.2.5 Circular Dichroism

The stability of phosphorylated protein was determined by using Circular Dichroism (CD) spectrum. In this study, 230nm wavelength was chosen to monitor. 300 µL of 10 µM protein in phosphate buffer (NMR) is good for the CD measurements, Tris buffer is not appropriate

for thermal denaturation and it should avoid to use high NaCl concentration. The temperature varied from 10 to 60 °C and the heating rate was 2 °C per minute. The curves were fitted with Boltzmann function.

2.2.6 Spin labeling

A stable nitroxide radical, 3-(2-iodoacetamido)-PROXYL (Sigma-Aldrich), was attached to single cysteine variants of SF1, in which the wildtype cysteine 171 was mutated to alanine, were generated as described in 2.2.1. Single cysteine residues were introduced into variants SF1 at position S20, T137 and N166, respectively (s. Table 2.4). ¹⁵N labeled SF1 were expressed and purified according to the description in 2.2.2. Variants SF1 were phosphorylated as described in 2.2.3. 5 mM DTT was adjusted in the storage buffer to ascertain a complete reduction of the cysteine. For the spin labeling, the buffer was exchanged to 1 M Tris; pH 8.0 using a PD-10 desalting column (GE Healthcare). High concentration of Tris buffers the hydriodic acid, which was produced during the labeling reaction. A five-fold excess of iodoacetamide proxyl was added and incubated in the dark at room temperature overnight. The unbound spin label was removed and the buffer was exchanged to NMR buffer via a PD-10 desalting column (GE Healthcare) and concentrated in an Amicon Ultra-15 Centrifugal Filter Units (Millipore) to 100 μM. After recording a ¹H, ¹⁵N HSQC spectrum in the oxidized state the spin label was reduced by adding a three-fold excess of ascorbate; pH 6.5, and incubating for at least 30 min. Another ¹H, ¹⁵N HSQC spectrum was measured and the PRE effect on backbone amide protons was determined by calculating the signal intensities in the spectra as $\frac{I_{ox}}{I_{red}}$ (I_{ox} the signal intensity in the oxidized state, I_{red} the signal intensity in the reduced state). Spectra were recorded on a Bruker Avance III 800MHz spectrometer at 298 K.

2.2.7 Electrophoretic mobility shift assays (EMSAs)

The 3' splice site RNA (GGUCAUACUAACCCUGUCCCUUUUUUUUCCACAG|C; | denotes the 3' splice site) is derived from the 3' splice site of AdML intron 1 by replacing the original BPS with a consensus BPS (underlined). The RNA was synthesized with the T7-MEGA shortscript kit (Ambion) in the presence of [α -³²P] UTP and gel-purified. RNA binding was performed for 15 min at room temperature in 10-μl reactions containing U2AF65^{RRM123}, SF1 proteins and

50 pmol [α -³²P] UTP-RNA. Assays with SF1²⁻³²⁰ and deletion mutants were performed in the presence of 12 % (v/v) glycerol, 12 mM Hepes-KOH (pH 7.9), 4 mM potassium-phosphate (pH 6.5), 60 mM KCl, 10 mM NaCl, 1.8 mM MgCl₂, 0.5 mM DTT, 0.12 mM EDTA, 5 μ g tRNA and 8U RiboShield™ ribonuclease inhibitor (Dundee Cell Products). Assays with SF1 and pSF1 contained final concentrations of 8 mM potassium-phosphate (pH 6.5), 20 mM NaCl, 0.5 mM DTT, 5 μ g tRNA and 8U RiboShield™ ribonuclease inhibitor. Reaction products were resolved in native 5% polyacrylamide gels (acrylamide:bisacrylamide = 80:1) in 0.5 x TBE at 4°C for 3 h at 150 V.

2.3 NMR and SAXS experiments

2.3.1 NMR measurements and assignments

All samples were measured at 298 K in NMR buffer with 10% $^2\text{H}_2\text{O}$ added for the lock. NMR spectra were recorded on AVIII 500, AVIII 600, AVIII 750, AVIII 800 or AVI 900 Bruker NMR spectrometers, equipped with cryogenic triple resonance gradient probes (with the exception of the AVIII 750MHz). Data were processed in NMRPipe/Draw (Delaglio et al., 1995) and analyzed in Sparky 3 (T. D. Goddard and D. G. Kneller, University of California). Protein backbone assignments were obtained from HNCACB and HNCA spectra, and by comparison of $^1\text{H}, ^{15}\text{N}$ -HSQC and -TROSY spectra with previously reported data (Liu et al., 2001; Selenko et al., 2003). Amino acid side chain resonance assignments were obtained from HCCH-TOCSY, ^{15}N - and ^{13}C -edited NOESY-HSQC/HMQC experiments (Sattler et al., 1999). Intermolecular NOEs between the U2AF65^{UHM} and SF1^{NTD} were identified for well-resolved peaks in the 3D ^{13}C - and ^{15}N -edited NOESY-HSQC experiments. $\text{H}^{\text{N}}\text{-N}$, $\text{N-C}'$, $\text{H}^{\text{N}}\text{-C}'$ residual dipolar couplings for SF1^{HH} were recorded using HNCQ-based NMR experiments (Sattler et al., 1999). Alignment media consisted of 15 mg/ml Pf1 phage (Profos AG, Regensburg, Germany) (Hansen et al., 1998).

2.3.2 Secondary chemical shift analysis and chemical shift perturbations

$^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ secondary chemical shifts were calculated according to

$$\Delta\delta (^{13}\text{C}\alpha, ^{13}\text{C}\beta) = (\delta^{13}\text{C}\alpha(\text{exp}) - \delta^{13}\text{C}\alpha(\text{rc})) - (\delta^{13}\text{C}\beta(\text{exp}) - \delta^{13}\text{C}\beta(\text{rc}))$$

in which $\delta^{13}\text{C}(\alpha,\beta)(\text{exp})$, experimentally obtained chemical shift and $\delta^{13}\text{C}(\alpha,\beta)(\text{rc})$, random coil chemical shift values (Wishart and Sykes, 1994).

Chemical shift perturbations of signals in $^{15}\text{N}, ^1\text{H}$ HSQC spectra were calculated according to

$$\Delta\delta (\text{NH}) = (\Delta\delta^1\text{H})^2 + \left(\frac{\Delta\delta^{15}\text{N}}{6}\right)^2$$

where $\Delta\delta$ = chemical shift difference of the signal in the different spectra.

2.3.3 Relaxation analysis

^{15}N R_1 , $R_{1\rho}$ relaxation rates and ^1H - ^{15}N heteronuclear NOEs of the SF1^{NTD} and U2AF65^{UHM} complexes were recorded at 750 MHz, of SF1^{NTD} at 600 MHz and of SF1 at 800 MHz proton Larmor frequency at 298 K as described (Farrow et al., 1994). Values were determined from the ratio between signal intensity in the experiment with and without saturation with the error calculated from experimental noise. Two delays in each experiment were recorded in duplicates to define the experimental error. Spectra were processed with NMRPipe. Local correlation times were derived from the ^{15}N R_2/R_1 ratio according to

$$\tau_c \cong \frac{1}{2\pi\nu^{15}\text{N}} \sqrt{\left(\frac{3}{2}\right) \left(\frac{R_2}{R_1} - \frac{7}{6}\right)}$$

with $\nu^{15}\text{N}$ = the resonance frequency of ^{15}N in Hz (Farrow et al., 1994).

2.3.4 Structure calculations

The structure of SF1^{HH} was determined using standard approaches. Automated NOESY cross-peak assignments and structure calculations with torsion angle dynamics were performed with CYANA 3.0 (Guntert, 2009). NOE-derived distance restraints derived from CYANA together with ϕ and ψ backbone dihedral angle restraints derived from TALOS+ (Shen et al., 2009) based on secondary chemical shifts and residual dipolar couplings were used during the structure calculation and for final water refinement (Linge et al., 2003). Structures were validated with iCing (<http://nmr.cmbi.ru.nl/icing/>). Molecular images were generated with PyMol (Schrödinger).

The structure of the SF1^{NTD}-U2AF65^{UHM} complex was calculated using our previously reported protocol (Simon et al., 2010). The SF1^{HH} structure determined here and the structure of the U2AF65^{UHM}/SF1^{ULM} complex (Selenko et al., 2003) were used as input structures for semi-rigid calculation of the complex structure. The input structures were maintained using non-crystallographic symmetry (NCS) restraints with a modified version of Aria1.2/CNS (Simon et al., 2010). Distance restraints derived from intermolecular NOEs and backbone torsion angle restraints derived from TALOS+ (Shen et al., 2009) were employed during molecular dynamics and simulated annealing. The final structures were refined in a shell of water molecules (Linge et al., 2003).

2.3.5 Small-angle X-ray scattering (SAXS)

SAXS data for solutions of SF1^{NTD}, pSF1^{NTD}, SF1, pSF1, U2AF65^{UHM}, U2AF65^{RRM123}, SF1^{NTD}-U2AF65^{UHM}, pSF1^{NTD}-U2AF65^{UHM}, SF1-U2AF65^{RRM123}, pSF1-U2AF65^{RRM123}, SF1-U2AF65^{RRM123}-RNA, and pSF1-U2AF65^{RRM123}-RNA were recorded at the X33 beamline of the European Molecular Biology Laboratory (EMBL) at Deutsches Elektronen Synchrotron (DESY) (Hamburg) using a MAR345 image plate detector. The scattering patterns were measured with a 2-min exposure time (8 frames, each 15 seconds) for several solute concentrations in the range from 1 to 10 mg/ml. Radiation damage was excluded based on a comparison of individual frames of the 2-min exposures, where no changes were detected. Using the sample-detector distance of 2.7 m, a range of momentum transfer of $0.01 < s < 0.6 \text{ \AA}^{-1}$ was covered ($s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 1.5 \text{ \AA}$ is the X-ray wavelength).

All SAXS data were analyzed with the package ATSAS. The data were processed using standard procedures and extrapolated to infinite dilution with the program PRIMUS (Konarev et al., 2003). The forward scattering, $I(0)$, and the radius of gyration, R_g , were evaluated using the Guinier approximation. The values of $I(0)$ and R_g , as well as the maximum dimension, D_{max} , and the inter-atomic distance distribution functions, $(P(R))$, were computed with the program GNOM (Svergun, 1992). The scattering from the high resolution models was computed with the program CRY SOL (Svergun et al., 1995). The masses of the solutes were evaluated by comparison of the forward scattering intensity with that of a bovine serum albumin reference solution (molecular mass 66 kDa). For back-calculation of SAXS data from the NMR ensemble of the SF1^{NTD}-U2AF65^{UHM} complex, the flexible regions (SF1^{NTD}: 1-12, 26-38, 70-95, 128-145) were randomized with CORAL (Petoukhov and Svergun, 2005).

3 RESULTS

3.1 Structural basis of SF1-U2AF65 cooperative RNA binding

3.1.1 SF1 constructs

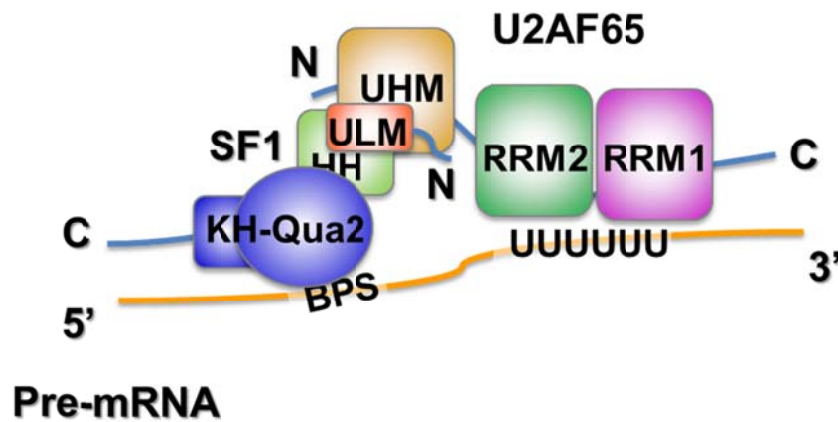


Figure 3.1 Proteins and domains involved in 3' splice site recognition. Diagram of the SF1-U2AF65 3' splice site recognition complex representing domains of SF1 and U2AF65 (HH, helix hairpin; KH, hnRNP K Homology; QUA2, Quaking homology-2; RRM, RNA recognition motif; ULM, U2AF homology domain (UHM) ligand motif). pre-mRNA is described in orange.

SF1-U2AF65 complex recognizes 3' splicing site of pre-mRNA in the complex E (Figure 3.1). A highly conserved tryptophan in the short N-terminal SF1 region named as 'U2AF ligand motif' (ULM) is extended into a hydrophobic pocket formed in a C-terminal region termed 'U2AF homology motif' (UHM) of U2AF65 (Kielkopf et al., 2004; Selenko et al., 2003). SF1 central hnRNP K homology and quaking 2 (KH-QUA2) regions specifically bind to the branch-point-sequence (BPS), in particular mediated by the key branch-site adenosine (Berglund et al., 1997; Liu et al., 2001). A poorly conserved domain composed of zinc knuckles is essential for the RNA recognition of non-vertebrate SF1 homologues, but little contributes to RNA binding affinity of vertebrate SF1 homologues (Berglund et al., 1998; Garrey et al., 2006). At the C-terminus, a proline-rich domain of SF1 interacts with other splicing factors, e.g. Prp40p containing a WW-domains (Abovich and Rosbash, 1997; Bedford et al., 1998; Goldstrohm et al., 2001).

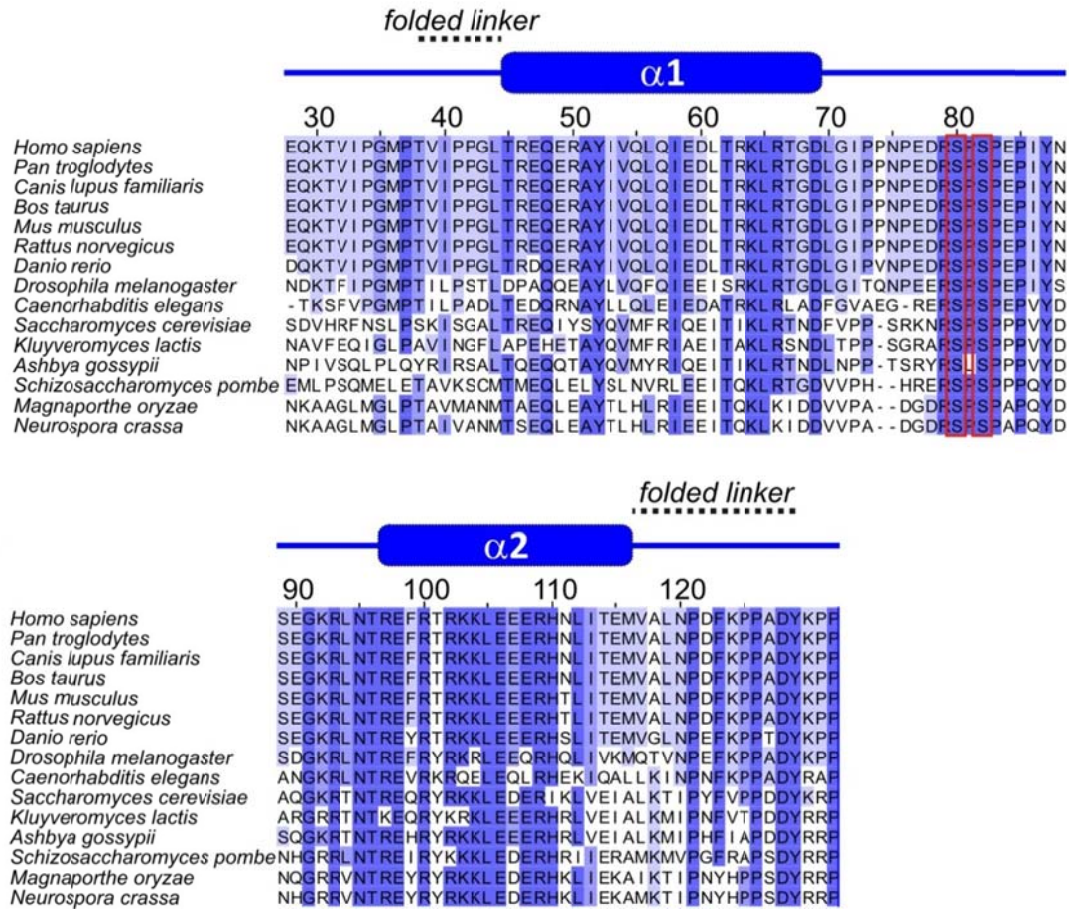


Figure 3.2 Multiple sequence alignment of SF1 N-terminal domains (SF1^{NTD}). Highly conserved residues are shown on a dark blue background, and the darkness of the blue background represents the relatively conservation of SF1^{NTD}. Sequences were taken from the UniProt database (www.uniprot.org) and residue numbers are given for *Homo sapiens* SF1. The secondary structure of SF1^{NTD} is depicted above the alignment. The phosphorylated serine residues (Ser80, Ser82) are indicated in red boxes. Sequences were aligned with ClustalW (Larkin et al., 2007) and analyzed with Jalview 2 (Waterhouse et al., 2009).

In contrast, the structural and functional information of the region of SF1 between UHM-bound ULM and RNA-recognition KH-QUA2 domains (residues 26–132) remain poorly understood. This region harbors the tandem serine motif (Figure 3.2) that is phosphorylated by U2AF homology motif kinase 1 (UHM-K1), and the phosphorylation of SF1 is reported to enhance the formation of SF1-U2AF65-RNA ternary complex (Manceau et al., 2006). A multiple sequence alignment of SF1 shows that this region is highly conserved, 48% sequence identity, e.g. between *Saccharomyces cerevisiae* and human SF1 in this region. (Figure 3.2)

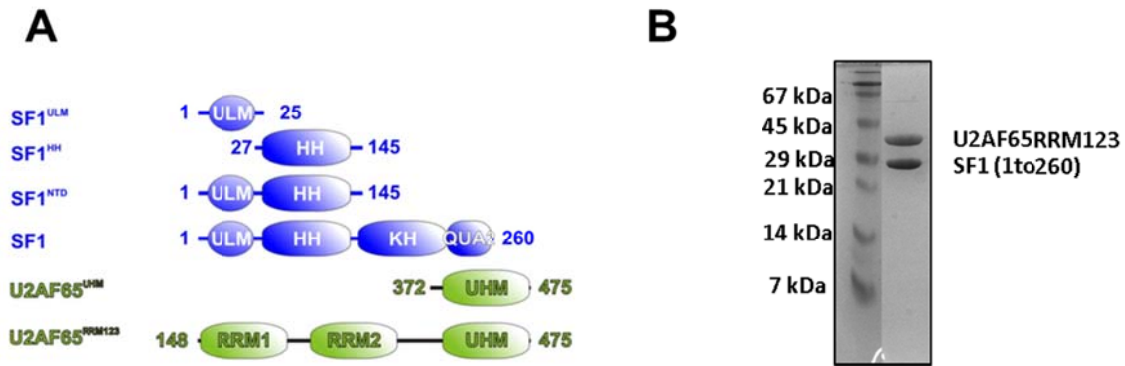


Figure 3.3 Diagram of the protein constructs used in our experiments. (A) SF1^{ULM} is the binding site of U2AF65^{UHM}. SF1^{KH-QUA2} and U2AF65^{RRM123} recognize branch-point-sequence and poly-pyrimidine track of pre-mRNA, respectively. The domains boundary residues are presented with the numbers. SF1^{HH} domain represents the α -helices region in SF1^{NTD} **(B)** SDS-PAGE of SF1-U2AF65^{RRM123} complex. The molecular weight marker is presented on the first lane.

In human cells, the tandem SF1 'SPSP' motif (Ser80-Pro81-Ser82-Pro83) is phosphorylated by UHMK1 kinase. Although few reports referring to the functions of the SF1 tandem phosphorylation and no clear UHMK1 homologues are evident in *Saccharomyces cerevisiae*, the SPSP motif of yeast SF1 is conserved (Figure 3.2) and an adjoining PPxY motif has been claimed to mediate interactions with the WW domain of the splicing factor Prp40p (Abovich & Rosbash, 1997). Despite these functional implications, the structural information of SF1 domain between ULM and KH-QUA2 domains has not been studied. As a step towards determining this potentially novel structure, we therefore cloned and expressed recombinant proteins comprising residues 1-145 (SF1^{NTD}) and 27-145 (SF1^{HH}) for NMR analysis (Figure 3.3A). As a binding partner to SF1 with 1:1 ratio (Figure 3.3B), U2AF65^{RRM123} was cloned with different constructs as well (Figure 3.3A).

3.1.2 NMR analysis and structure of the N-terminal region of SF1

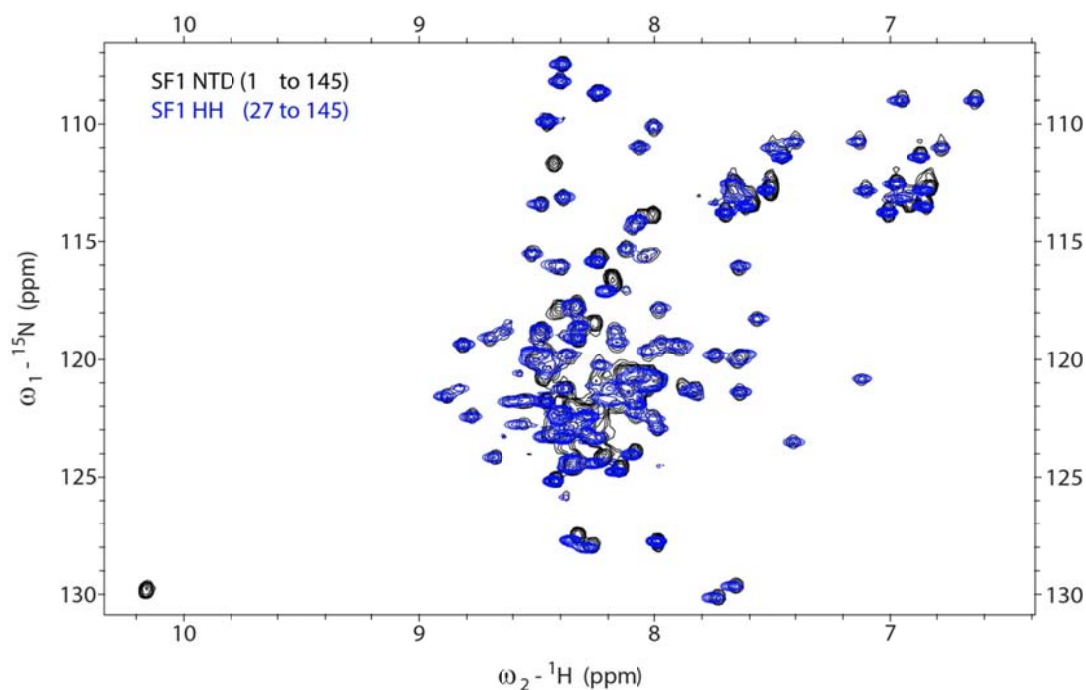


Figure 3.4 Overlay of the ^1H ^{15}N HSQC spectra of SF1^{NTD} (1-145; black) with SF1^{HH} (27-145, blue). Signals from the disordered N-terminus between residues 1 and 26 cluster in the center of the spectrum. The chemical shifts for the dispersed resonances correlate well between the SF1^{NTD} and SF1^{HH} constructs.

^1H , ^{15}N HSQC NMR spectra were recorded on the region of SF1^{NTD} (1-145) as well as SF1^{HH} comprising residues from 27 to 145 (Figure 3.4). The signals in the ^1H , ^{15}N HSQC NMR spectrum of the N-terminal SF1^{ULM} region cluster in the center of the spectrum around 8 to 8.5 ppm as typically seen for unfolded proteins, which is consistent with the previous report (Selenko et al., 2003). The chemical shifts for the dispersed resonances correlate well between the SF1^{NTD} and SF1^{HH} constructs, which indicates the removal of SF1^{ULM} does not alter the structure of SF1^{NTD}.

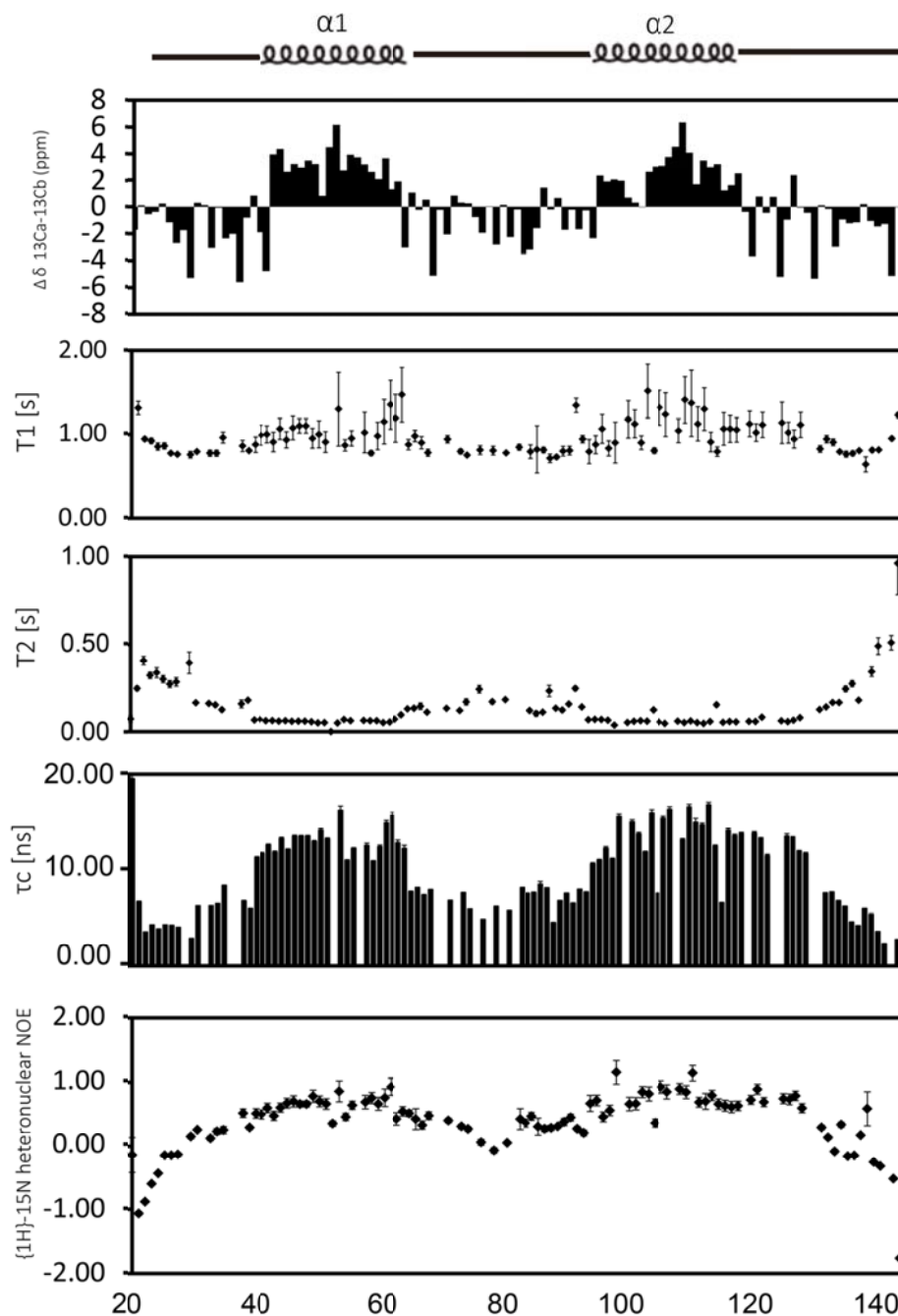


Figure 3.5 Secondary chemical shift and dynamic analysis of SF1^{HH} plotted against the amino acid sequence. The $^{13}\text{C}_\alpha - ^{13}\text{C}_\beta$ secondary chemical shifts of SF1^{HH} (upper panel) indicate two α -helices for the protein. T1 and T2 relaxation time, the correlation time τ_c and ^1H - ^{15}N heteronuclear NOE values in panel 2, 3, 4 and 5. Secondary structure elements are indicated on top.

NMR backbone and sidechain experiments for the SF1^{HH} were recorded to assign the residues. $^{13}\text{C}_\alpha - ^{13}\text{C}_\beta$ secondary chemical shifts of SF^{HH} indicate two α -helices (residue from 46

to 67 and from 97 to 119) connected by an unstructured loop region (residue from 68 to 96).

Dynamic properties of SF1^{HH} were analyzed by recording T_1 and T_2 relaxation time as well as the ^1H - ^{15}N heteronuclear NOE (Figure 3.5). Quantitative evaluation was limited to approximately 70% of the peaks due to signal overlap in ^1H , ^{15}N HSQC spectra. T_1 relaxation time of the two α -helices very similar and T_2 relaxation time is increased for the loop region connecting the two α -helices.

These residues in the loop region also have a decreased correlation time (Figure 3.5). The correlation time τ_c for the two helices is similar suggesting the helix hairpin is tumbling as a unit. The overall correlation time τ_c (10 ns) indicates monomeric tumbling of SF1^{HH}. ^1H - ^{15}N heteronuclear NOE data of SF1^{HH} describe a rigid helix hairpin domain with values of approximately 0.75 and a flexible linker in between from residue 68 to 96 with very low rates between 0.3 to -0.3 which is indicative for dynamics on a picosecond to nanosecond timescale. Based on distance, dihedral angle, and residual dipolar coupling restraints (Table 3.1, Figure 3.7), we determined the three-dimensional structure of SF1^{HH}.

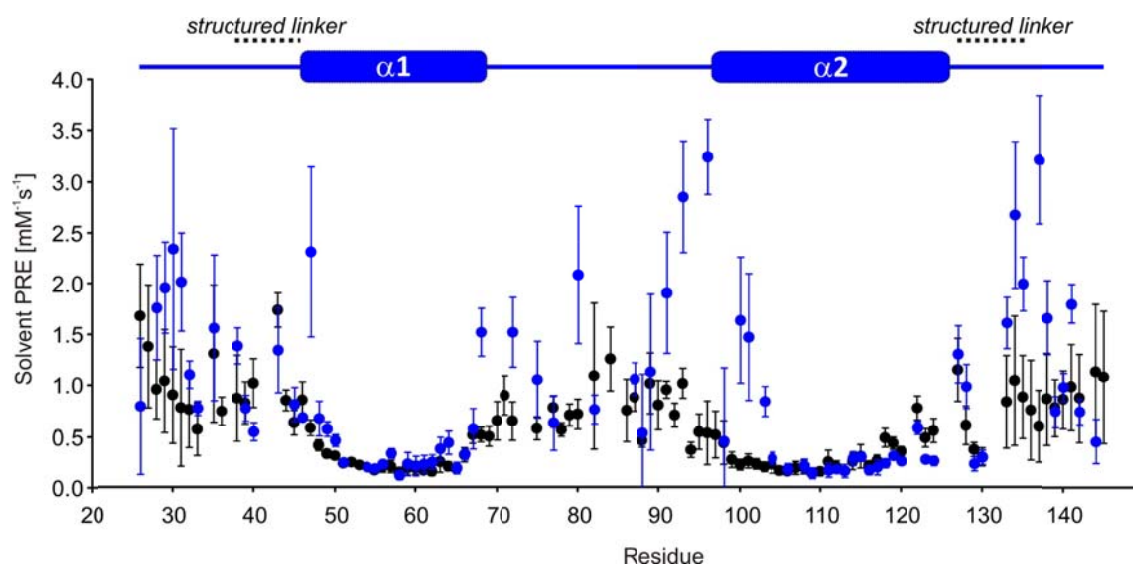


Figure 3.6 Comparison of back-calculated and experimental solvent PREs. Experimental solvent PRE rates of backbone amide protons of SF1^{HH} (blue) were compared to values back-calculated from the ensemble of NMR structures (black). Values for the structured region including helices α 1- α 2 are in excellent agreement. For flexible regions, experimental values are higher than the back-calculated values due to chemical exchange between amide protons in these parts with water molecules which experience a large relaxation enhancement because of transient binding to Gd(DTPA-BMA). The secondary structure is depicted above the diagram.

Dr. T. Madl (Technische Universität München) performed the structure calculations and further validated the structure by comparison of measured and back-calculated relaxation rate enhancements upon addition of the paramagnetic co-solvent Gd(DTPA-BMA) (Figure 3.6) (Madl et al., 2009). SF1^{HH} comprises a helix hairpin motif with two α -helices in an anti-parallel arrangement that are connected by a flexible linker (Figure 3.7), which contains the Ser80-Pro81-Ser82-Pro83 tandem phosphorylation motif. Hydrophobic residues within the two helices are spaced every 3–4 residues, i.e. Ala51, Val54, Ile58, Leu61, and Leu65 in α 1 contact Leu105, Leu112, Met116, and Leu119 in α 2, respectively, and thereby stabilize the arrangement of the two helices. Arg109 in helix α 2 is engaged in a potential salt bridge with Asp128 C-terminal of helix α 2 (Figure 3.7).

Table 3.1 NMR and refinement statistics for SF1^{HH} ensembles*

	SF1 ^{HH}
NMR distance, dihedral and RDC restraints	
Distance restraints	
Hydrogen bonds [†]	38
Intra-molecular	2021
short-range i-j ≤ 1	1140
medium-range 1 < i-j < 5	479
long-range i-j ≥ 5	366
Inter-molecular	-
Dihedral angle restraints [‡]	188
Protein φ/ψ	94/94
Residual dipolar couplings	122
Structure statistics	
Violations (mean and s.d.)	
Dihedral angle violations > 5°	0
Distance violations > 0.5 Å	0
Q ^{RDC}	0.27 ± 0.00
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.000
Bond angles (°)	1.190 ± 0.028
Improper (°)	1.362 ± 0.109
Average pairwise r.m.s.d. (Å) [§]	
Heavy	0.94 ± 0.09
Backbone	0.40 ± 0.06
Ramachandran	
Core	98.5
Allowed	1.5
Generous	0.0
Disallowed	0.0
Fit to SAXS data (c)	-

*Statistics are calculated for ten lowest-energy refined structures of a total of 100 structures.

[†]Hydrogen bond restraints seen in the input structures were maintained by distance restraints, if they were further supported by secondary chemical shifts.

[‡]Dihedral angle restraints were derived from secondary chemical shifts using TALOS+

[§]Backbone atoms within SF1^{HH} (45-69, 97-127)

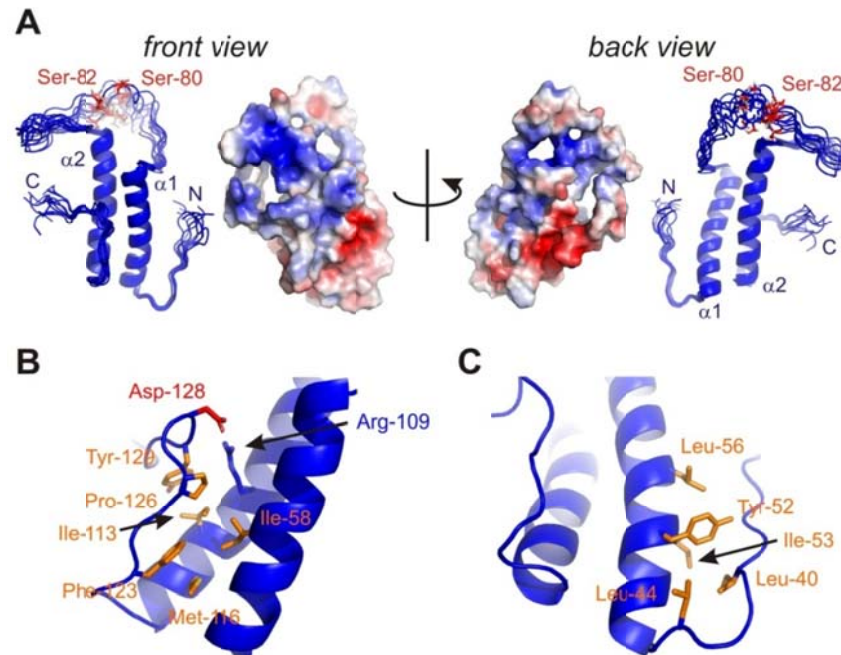


Figure 3.7 Structure of a novel helix hairpin in the N-terminal domain of splicing factor 1. (A) Ribbon representation of the ensemble of the 20 lowest energy structures and surface representation coloured according to electrostatic surface potential at 3 kT/e for positive (blue) or negative (red) charge potential using the program APBS (Baker et al., 2001). Close-up view of the N-terminal (B) and C-terminal (C) folded extensions. Side chains of key residues mediating the interactions between the α -helices and the N/C-termini are shown as sticks.

The C-terminal and N-terminal residues are structurally flexible. However, the C-terminal region adopts an extended conformation and packs against the helix hairpin by hydrophobic interactions of Phe123, Pro126, Tyr129 with residues in both α -helices (Ile58, Ile113, Met116) (Figure 3.7B). In addition, residues N-terminal of $\alpha 1$ (Ile40, Leu44) form hydrophobic contacts with residues located within $\alpha 1$ (Tyr52, Ile53, Leu56; Figure 3.7C). The additional interactions involving regions preceding the N- and C-terminal ends of the helix hairpin therefore increases ^1H - ^{15}N heteronuclear NOE values, which show that these extensions are rigid and thus stably interact with the helix hairpin. Thus, SF1^{HH} comprises a helix hairpin fold, which exposes the SPSP phosphorylation motif in a flexible linker.

3.1.3 SF1^{NTD} provides a secondary interface in the SF1^{NTD}-U2AF65^{UHM} complex

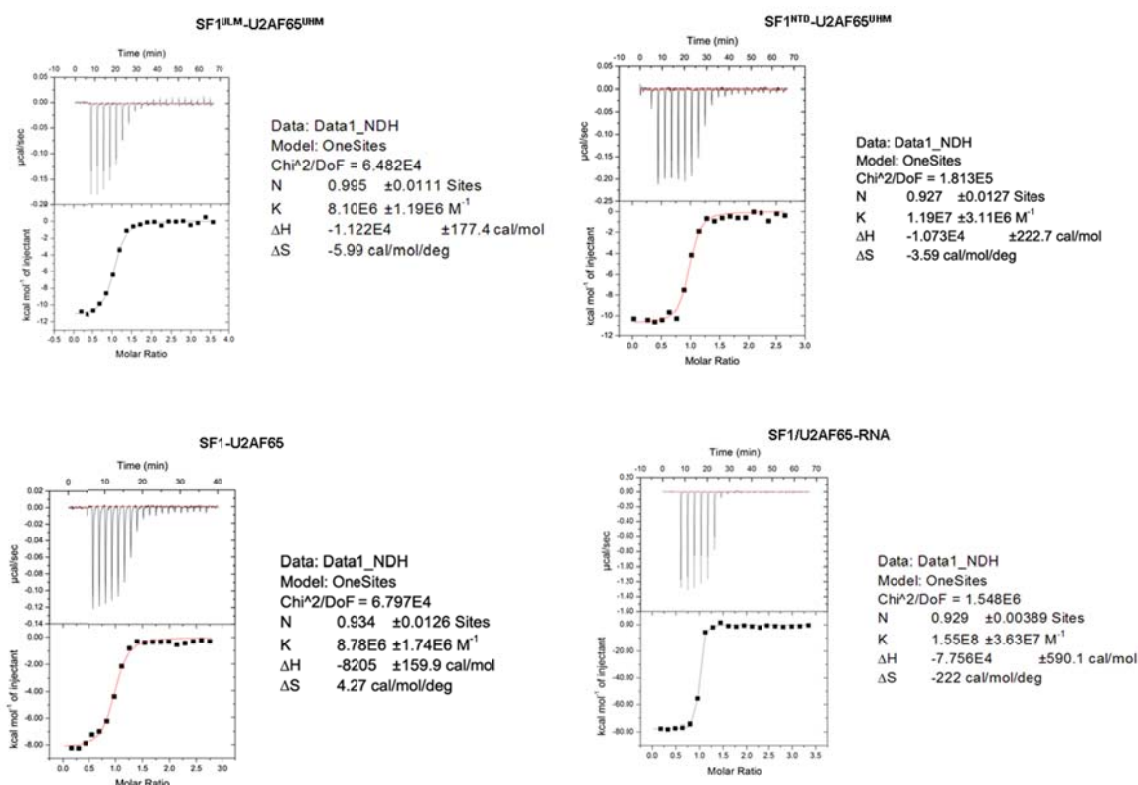


Figure 3.8 Isothermal titration calorimetry data. Measurements were carried out at 25 °C. Dilution heats determined by titrating proteins or RNA into the corresponding buffer were in the range of the heat effects observed at the end of the titration (data not shown) and were subtracted for the analysis.

Table 3.2 Dissociation Constants of SF1-U2AF65 determined by Isothermal Titration Calorimetry

Sample	Binding affinity [K_d]
SF1 ^{ULM} -U2AF65 ^{UHM}	127 ± 48 nM
SF1 ^{HH} -U2AF65 ^{UHM}	n.d.
SF1 ^{NTD} -U2AF65 ^{UHM}	84 ± 23 nM
SF1-U2AF65 ^{RRM123}	114 ± 23 nM
SF1-U2AF65 ^{RRM123} -RNA	12 ± 6 nM

To determine whether the helix hairpin contributes to the SF1-U2AF65 interaction we determined binding affinities of SF1 and U2AF65 fragments using isothermal titration calorimetry (Table 3.2). As reported previously (Selenko et al., 2003), U2AF65^{UHM} binds SF1^{ULM} with $K_d = 127 \pm 48$ nM. Inclusion of the helix hairpin in SF1 (SF1^{NTD}) shows a slightly increased affinity ($K_d = 84 \pm 24$ nM), indicating that this region contributes to the protein-protein interaction. No further increase of binding affinity is detected when studying the SF1-U2AF65 complex using almost full-length U2AF65 (U2AF65^{RRM123}) and a region comprising all structural domains in SF1 (residues 1-260) ($K_d = 114 \pm 24$ nM). We therefore conclude that SF1^{NTD} and U2AF65^{UHM} harbor all relevant binding sites for the U2AF65-SF1 interaction and represent a minimal complex.

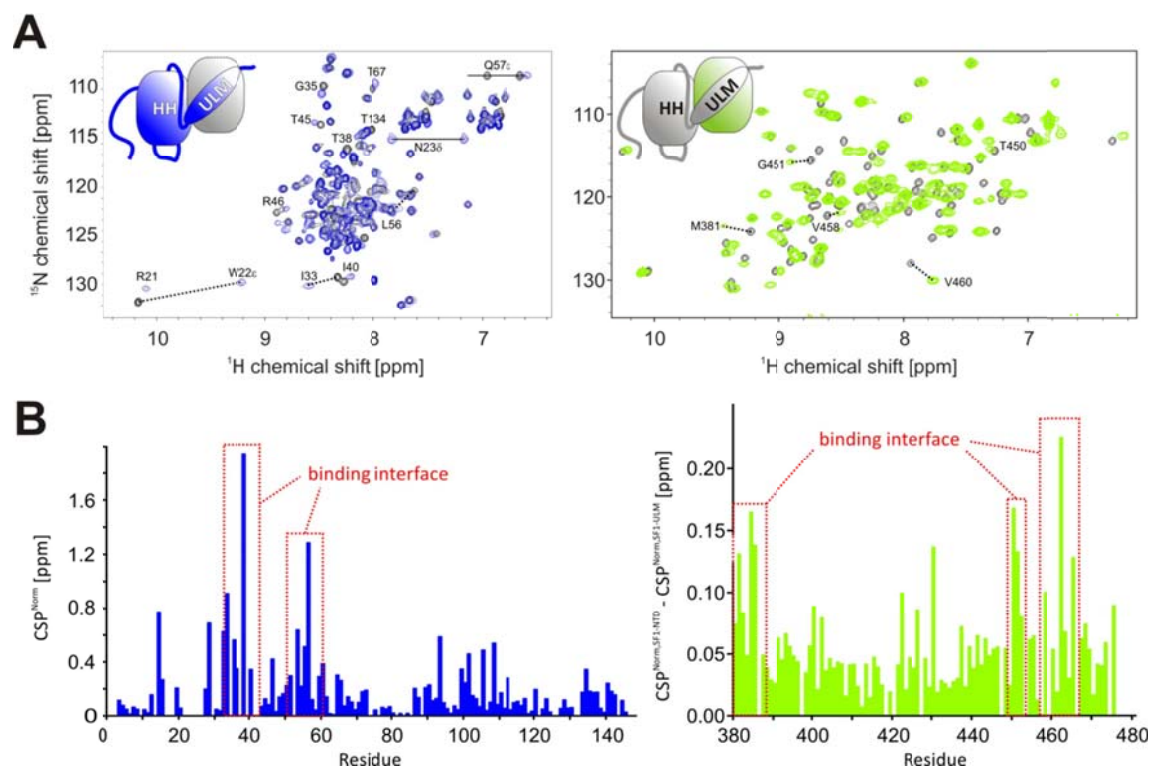


Figure 3.9 NMR analysis of the SF1^{NTD}-U2AF65^{UHM} interaction. (A) Superposition of ¹H, ¹⁵N HSQC NMR spectra of labelled SF1^{NTD} and U2AF65^{UHM} free (black) and when bound to unlabeled U2AF65^{UHM} (blue) or SF1^{NTD} (green), respectively. Selected residues which are shifted upon formation of the SF1^{NTD}-U2AF65^{UHM} complex are annotated. (B) Chemical shift perturbations (CSPs) of amides linked to U2AF65^{UHM} (blue)- and SF1^{NTD} (green)-binding. To selectively analyze the contributions of the secondary binding interface, the CSPs obtained in a titration of labelled SF1^{ULM} with unlabeled U2AF65^{UHM} were subtracted from the SF1^{NTD} CSPs. Residues with strong CSPs that are therefore located in the secondary binding interface are annotated.

To identify the binding interface between SF1^{NTD} and U2AF65^{UHM}, we carried out NMR chemical shift titrations of the isotope-labeled recombinant proteins with the unlabeled binding partner. The NMR signals of both SF1^{NTD} and U2AF65^{UHM} are shifted upon addition of unlabeled U2AF65^{UHM} or SF1^{NTD}, respectively (Figure 3.9A). Some of the NMR signals in the interface exhibit line broadening, which is characteristic for medium- to high-affinity complexes with micromolar dissociation constants, and thus consistent with the ITC data. We observed large chemical shift perturbations (CSPs) of SF1^{NTD} and residues in U2AF65^{UHM} including numerous residues, which are not in contact with SF1^{ULM} in the previously reported NMR structure (Figure 3.9B).

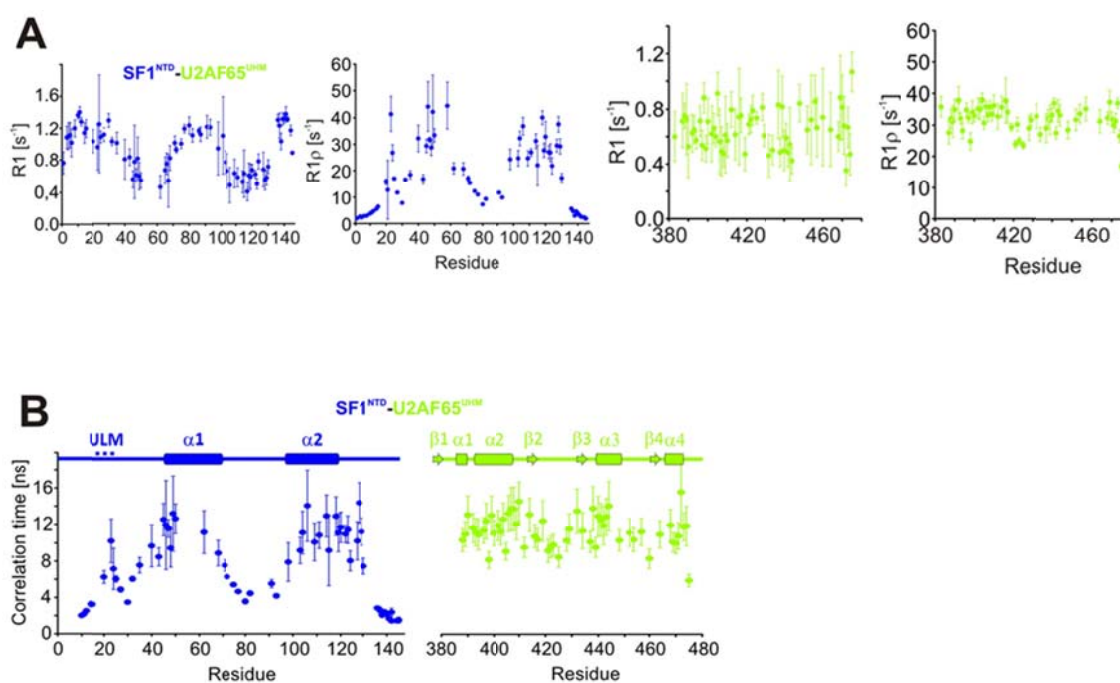


Figure 3.10 NMR relaxation data for the SF1^{NTD}-U2AF65^{UHM} complex. **(A)** ¹⁵N backbone amide longitudinal (R_1) and transverse relaxation rates $R_{1\rho}$ was measured at 750 MHz proton larmor frequency for SF1^{NTD} (blue) and U2AF65^{UHM} (green) in the SF1^{NTD}-U2AF65^{UHM} complex. **(B)** Residue-specific local correlation times τ_c for SF1^{NTD} (blue) and U2AF65^{UHM} (green) in the SF1^{NTD}-U2AF65^{UHM} complex. Correlation time values were determined from ¹⁵N relaxation data as described in Methods. The secondary structures are depicted above the diagram.

In NMR relaxation measurements of the complex, both subunits show similar ¹⁵N R_1 and $R_{1\rho}$ relaxation rates (Figure 3.10A) and tumbling correlation times (τ_c , Figure 3.10B), thus indicating that they tumble as a single entity. The average τ_c of 14 ns is in good agreement with the correlation time expected for the molecular mass of the complex (29 kDa; $\tau_c^{\text{calculated}}=17$ ns)(Daragan and Mayo, 1997).

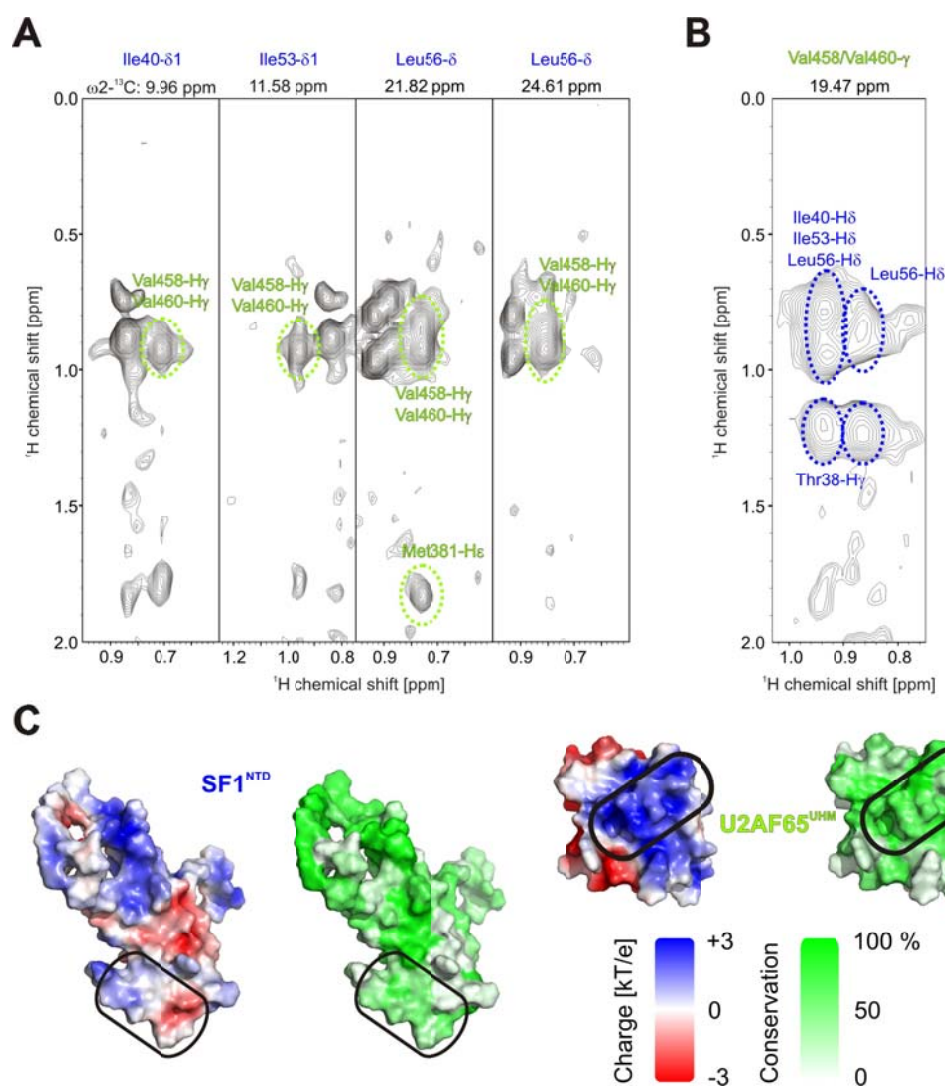


Figure 3.11 Analysis of the SF1^{NTD}-U2AF65^{UHM} complex. Strip plots from filtered/edited NOE experiments to detect SF1^{NTD}-U2AF65^{UHM} NOEs (A,B). Strip plots are shown for ILV methyl-protonated [^2H , ^{13}C , ^{15}N] SF1^{NTD} in complex with unlabeled U2AF65^{UHM} (A) and *vice versa* (B). The methyl group of the labelled subunit is shown on top and the cross peaks to the unlabeled binding partner are shown in the strip plots. (C) Surface representation of the individual subunits of the SF1^{NTD}-U2AF65^{UHM} complex coloured according to electrostatic surface potential and conservation. The electrostatic surface potential is coloured at 3 kT/e⁻ for positive (blue) or negative (red) charge potential using the program APBS (Baker et al., 2001). For the surface representations of the individual subunits, they are rotated by +/- 90° with respect to the orientation to expose the binding interface (black ellipse). Sequence conservation of SF1^{NTD} and U2AF65^{UHM} was generated using the ConSurf and ConSeq webserver (Ashkenazy et al., 2010; Berezin et al., 2004).

To determine the three-dimensional structure of the SF1^{NTD}-U2AF65^{UHM} complex we employed inter- and intra-molecular distance restraints derived from ^{15}N - and ^{13}C -edited NOESY spectra. To unambiguously identify intermolecular NOEs a set of isotope-edited and

filtered NOESY spectra was recorded (Otting and Wuthrich, 1990; Sattler et al., 1999) and specifically isotope-labeled protein complexes were used where one of the subunits (either SF1^{NTD} or U2AF65^{UHM}) was deuterated and methyl-protonated. Several NOEs were detected for ¹H,¹³C-labeled methyl groups in the binding interface (SF1^{NTD} V39, I40, I53, L56 and U2AF65^{UHM} V458, V460; Figure 3.11).

Table 3.3 NMR and refinement statistics for SF1^{NTD}-U2AF65^{UHM} ensembles*

SF1 ^{NTD} -U2AF65 ^{UHM}	
NMR distance, dihedral and RDC restraints	
Distance restraints	
Intra-molecular	14
short-range i-j ≤1	-
medium-range 1< i-j <5	9
long-range i-j ≥5	5
Inter-molecular	22
Dihedral angle restraints [‡]	
Protein φ/ψ	65/65
Structure statistics	
Deviations from idealized geometry	
Bond lengths (Å)	0.004 ± 0.000
Bond angles (°)	0.492 ± 0.005
Impropers (°)	0.800 ± 0.058
Average pairwise r.m.s.d. (Å) [§]	
Heavy	1.38 ± 0.63
Backbone	1.27 ± 0.62
Ramachandran	
Core	84.7
Allowed	14.6
Generous	0.7
Disallowed	0.0
Fit to SAXS data (c)	1.29 ± 0.11

*Statistics are calculated for ten lowest-energy refined structures of a total of 100 structures.

[§] Backbone atoms within SF1^{NTD} (38-40, 45-69, 97-127) - U2AF65^{UHM} (375-473)

Dr. T. Madl did the structure calculation of the SF1^{NTD}-U2AF65^{UHM} complex, using a protocol described recently (Simon et al., 2010). Shortly, semi-rigid body modeling was performed using the previously determined structures of SF1^{ULM}-U2AF65^{UHM} and SF1^{HH} as input. Comparison of secondary chemical shifts between the free and bound proteins and overall similarity of NMR spectra confirm that both binding partners do not undergo substantial structural changes upon complex formation compared to the input structure. Structures were calculated based on inter-molecular NOE distances and dihedral angle restraints derived from TALOS+ (Shen et al., 2009). The final ensemble of structures was refined in a shell of water molecules (Linge et al., 2003) and validated by comparing experimental and back-calculated SAXS data of the complex (Table 3.3).

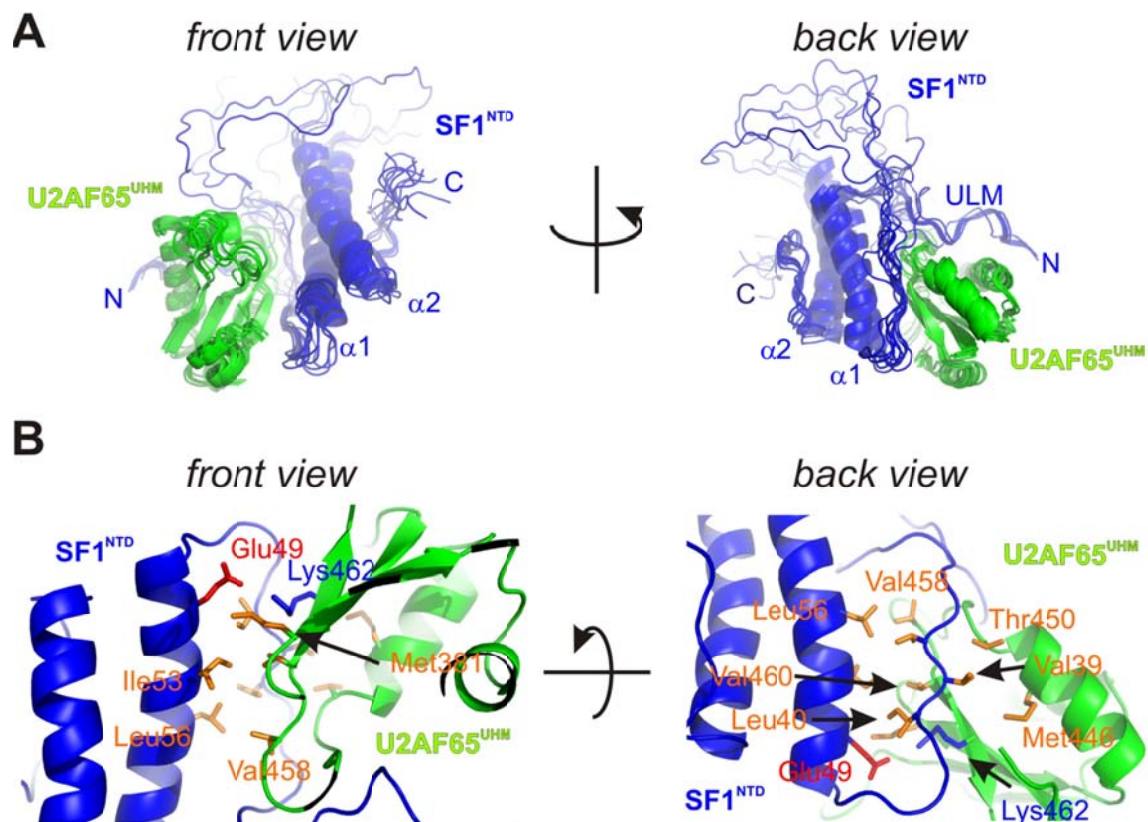


Figure 3.12 Structure of the SF1^{NTD} and U2AF65^{UHM} complex. (A) Ribbon representation of the ensemble of the 10 lowest energy structures and surface representation coloured according to electrostatic surface potential at 3 kT/e for positive (blue) or negative (red) charge potential using the program APBS(Baker et al., 2001). (B) Close-up view of SF1^{NTD}-U2AF65^{UHM} complex interface. Side chains of key residues mediating the interactions between the α -helices and the N/C-termini are shown as sticks.

The final ensemble of structures of the SF1^{NTD}-U2AF65^{UHM} complex is shown in Figure 3.12A and Figure 3.11, SF1^{NTD} and U2AF65^{UHM} form an additional hydrophobic interface involving U2AF65^{UHM} Met381, Val458, Val460 and SF1^{NTD} Val39, Ile40, Ile53, Leu56, which is further stabilized by a potential salt bridge between SF1^{NTD} Glu49 and U2AF65^{UHM} Lys462 (Figure 3.12). This secondary interface buries an additional solvent-exposed surface of 1267 Å² (determined with PDBePISA (Krissinel and Henrick, 2007)). Notably, the additional interface in SF1^{HH} is remote from the ULM; it is located on the opposite side of the SF1^{ULM} binding site of U2AF65^{UHM} and therefore does not interfere with binding of SF1^{ULM}. The additional hydrophobic interface moderately strengthens the SF1-U2AF65 interaction, with ITC-derived dissociation constants of U2AF65^{UHM} for SF1^{ULM} and SF1^{NTD} of $K_d = 127$ and 85 nM, respectively. Consistent with this moderate contribution to the overall affinity, the interaction of the helix hairpin (SF1^{HH}) alone with U2AF65^{UHM} is weak and hardly determined by ITC ($K_d \gg 100$ μM) (Table 3.2, Figure 3.8).

3.1.4 Interactions between SF1 and RNA

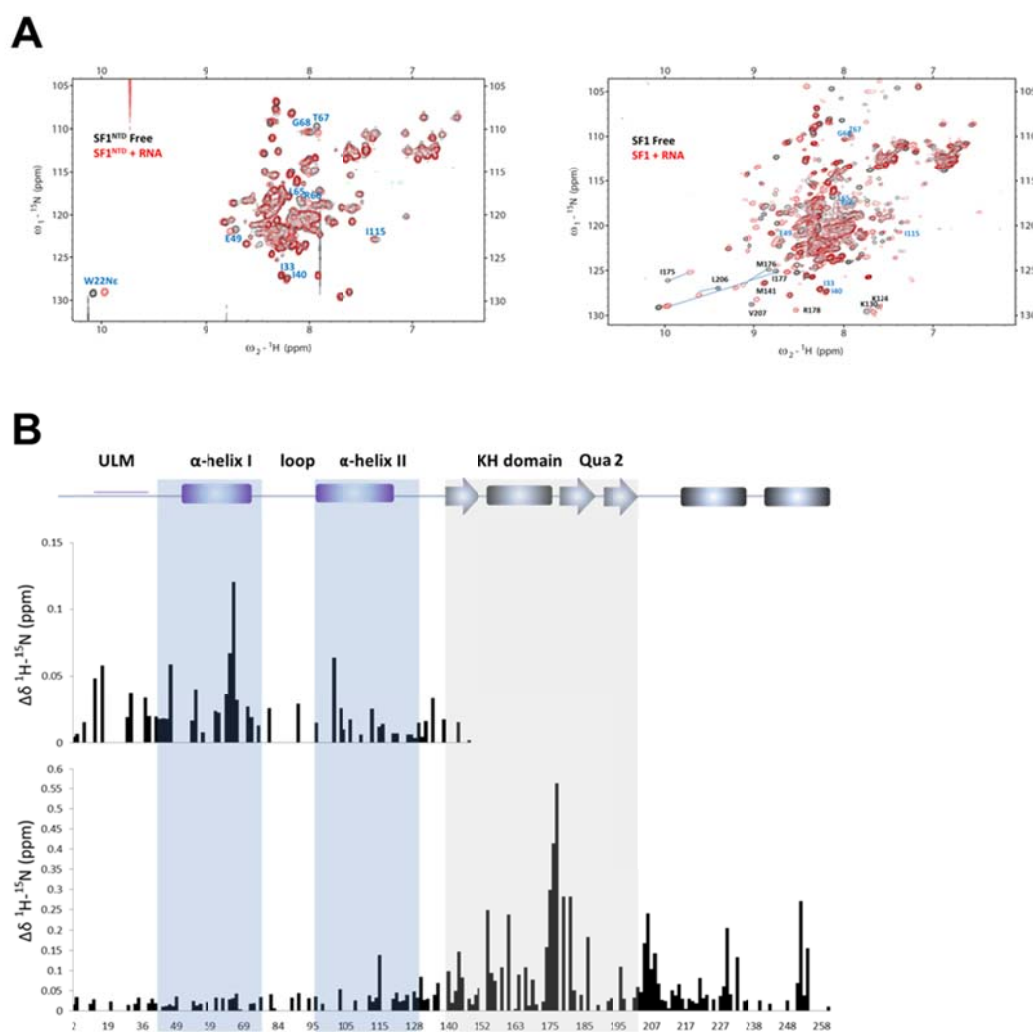


Figure 3.13 NMR analysis of SF1 (SF1^{NTD}) and RNA interactions. **(A)** Overlay of ¹H, ¹⁵N HSQC NMR spectra of labelled SF1 (SF1^{NTD}) free (black) and in the presence of unlabeled 20mer RNA (red). Selected residues which are shifted upon interactions between SF1 (SF1^{NTD}) and RNA are annotated, in the SF1 spectrum, the affected residues in the KH-QUA2 region as previously reported (Liu et al., 2001) are indicated with black letters, the additional residues undergo chemical shift perturbations within N-terminal region are shown with light blue letters. **(B)** Amide proton chemical shift differences of SF1 (SF1^{NTD}) with and without RNA plotted against the amino acid sequence. The secondary structure is indicated on top and residues which undergo the most prominent changes in the N-terminal region and in the KH-QUA2 region are highlighted with light blue boxes and a grey box respectively.

It has been evidenced that the SF1^{KH-QUA2} domain is essential for the RNA recognition at branch point sequence (Liu et al., 2001), while the question how is the SF1^{NTD} positioned in

the SF1-RNA complex remains unclear. To address this question, we did the NMR titrations of SF1^{NTD} with 20mer RNA containing BPS and poly-pyrimidine track (Figure 3.13A). Residues with small chemical shift perturbations (CSP) suggest that although it is weak, the interactions exist between SF1^{NTD} and RNA. The affected residues in SF1^{NTD} are easily reassigned by tracking to the small shifting. The chemical shift differences against the amino acid sequence are shown in Figure 3.13B, and the residues having CPS larger than 0.02 ppm are plotted on the 3D structure (Figure 3.14), displaying most of residues in SF1^{HH} region are involved in the RNA binding. To investigate the influences of the binding of RNA to SF1^{NTD} upon the SF1KH domain-BPS recognition, we repeated the NMR titrations using C-terminal extended SF1 (residues from 1 to 260) including KH-QUA2 domain. In comparison to the previously published NMR ¹H, ¹⁵N HSQC spectra of RNA-bound SF1^{KH-QUA2} (Liu et al., 2001), the spectra of RNA-bound SF1 present highly similar CPS for the residues in the KH-QUA2 region, indicating the interface between SF1^{KH-QUA2} domain and BPS is not affected by the binding of SF1^{NTD} to the RNA.

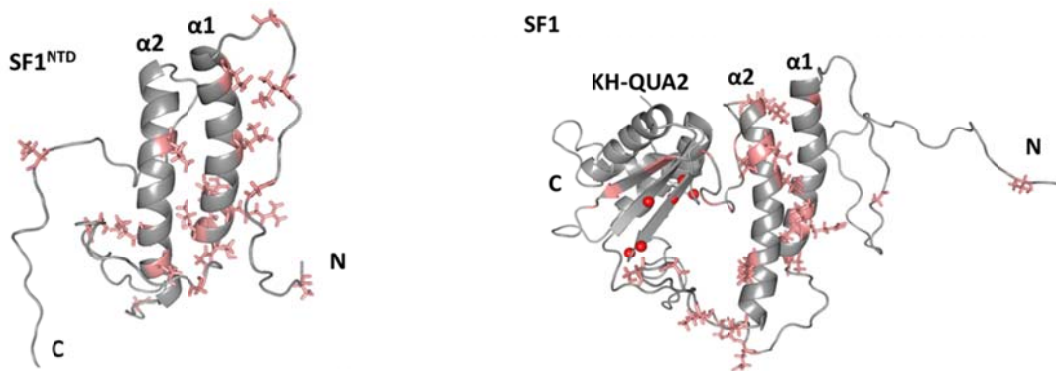


Figure 3.14 Mapping of CSPs induced by RNA binding on the NMR solution structure of SF1^{NTD} (left) and SF1 (right). The affected residues in the N-terminal region are shown in stick representation and highlighted in pink. Intensively changed residues in the KH-QUA2 domain are indicated in sphere representation and coloured in red.

3.1.5 Cooperative recognition of RNA by SF1-U2AF65 complex

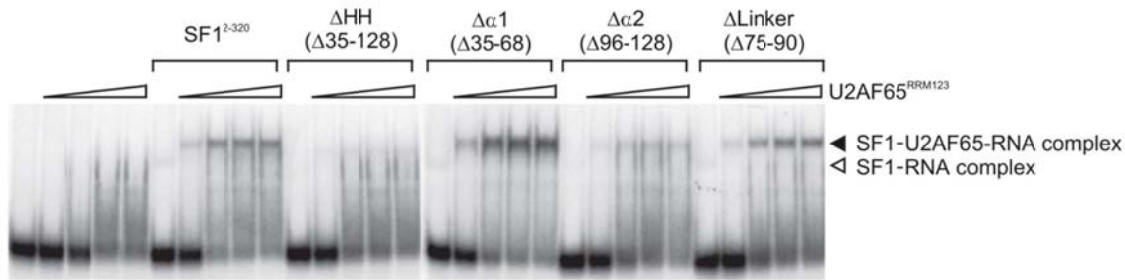


Figure 3.15 EMSA assay of Cooperative binding of U2AF65^{RRM123} and SF1 to a 3' splice site RNA. RNA was incubated with buffer or U2AF65^{RRM123} (0.2, 0.5, 1 and 2 μ M; indicated by triangles) in the absence or presence of 1.5 μ M His₆-tagged SF1²⁻³²⁰ or internal deletion mutants. Reaction products were separated by native PAGE and visualized by autoradiography. The migration of SF1-U2AF65-RNA complexes (closed arrowhead) and SF1-RNA complexes (open arrowhead) is indicated. Role of SF1^{HH} and the SF1^{NTD}-U2AF65^{UHM} interaction in the formation of the 3' splice site recognition complex. SF1^{HH} may establish an optimal orientation of the SF1 (KH-QUA2) and U2AF65^{RRM1,2} RNA-binding subunits in the complex and thereby support cooperative RNA binding.

To understand the role of SF1^{HH} and the SF1^{NTD}-U2AF65^{UHM} interaction in the formation of the 3' splice site recognition complex, the cooperative binding of SF1 and U2AF65 to the pre-mRNA was tested in EMSAs. Increasing concentrations of U2AF65^{RRM123} added to a 3' splice site RNA result in a smear of U2AF65-RNA complexes (Figure 3.15). His₆-tagged SF1²⁻³²⁰ also binds the RNA, although weakly. In the presence of both proteins a ternary SF1-U2AF65-RNA complex forms at lower U2AF65 concentrations, consistent with cooperative binding. The ternary complex is barely evident in the presence of SF1- Δ^{HH} , which indicates that the SF1^{HH} might act as a spacer defining the relative orientation of the subunits in the ternary complex. Deletion of helix strongly reduces ternary complex formation, whereas deletion of helix $\alpha 1$ doesn't. The results suggest that residues located in helix $\alpha 2$ might be involved in stabilization of the ternary complex. In addition, deletion of the linker does not have any effect. Thus, in agreement with the data shown above, the helix hairpin domain in the N terminus of SF1 is necessary for cooperative RNA binding, with helix $\alpha 2$ playing a more important role than helix $\alpha 1$.

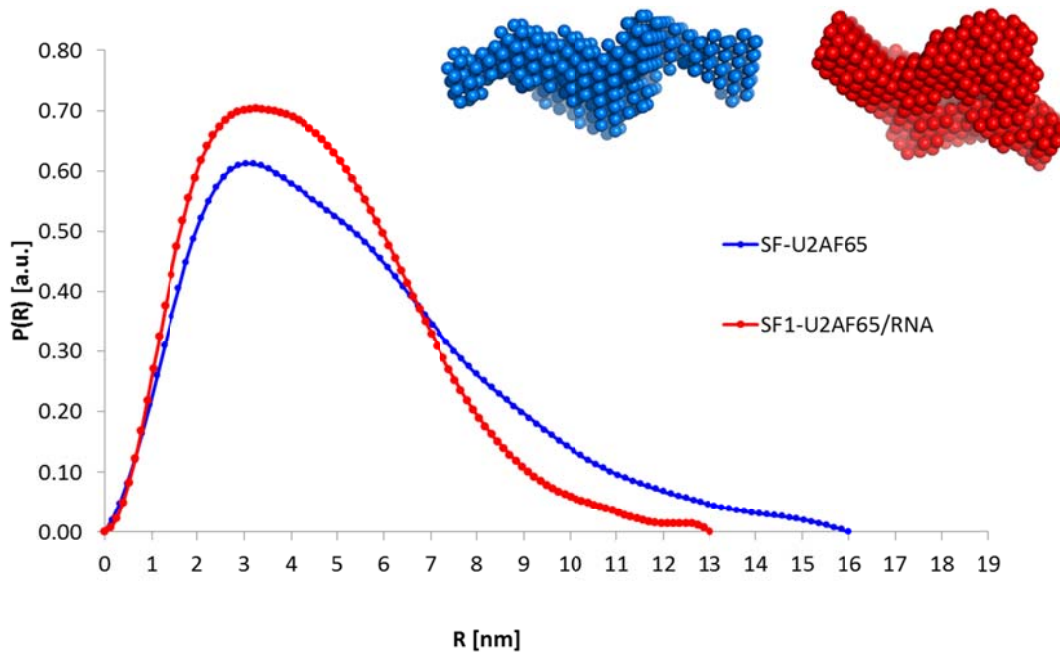


Figure 3.16 SAXS analysis of cooperative binding of SF1-U2AF65 to the pre-mRNA. SAXS data showing comparisons of radial density distributions (R_g) and maximum intraparticle sizes (D_{max}) of U2AF65^{RRM123}-SF1 complex with (red) or without (blue) RNA binding. Averaged and filtered 3D ab initio molecular envelopes are shown on the top right.

We then studied the binding of SF1-U2AF65 to the RNA by SAXS, showing that binding of a 20-mer RNA containing the BPS and the Py tract regions induces large overall changes for the structure and/or dynamics of the SF1-U2AF65^{RRM123} complex and leads to formation of a compact SF1-U2AF65^{RRM123}-RNA arrangement. The change of radial density distributions (R_g) from 34.2 ± 0.1 to 39.4 ± 0.4 Å and maximum intraparticle sizes (D_{max}) from 16 to 13 nm in the presence of the RNA ligand indicates that RNA binding leads to a large change in the overall structure and/or dynamics of the U2AF65-SF1 complex. (Figure 3.16)

3.1.6 Discussion: Cooperative binding of SF1-U2AF65 to Pre-mRNA

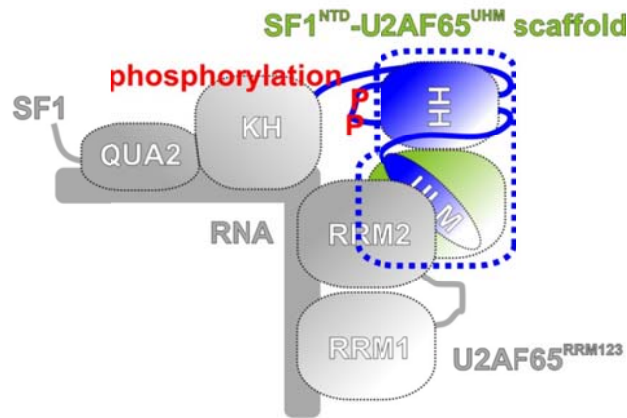


Figure 3.17 Schematic model of the SF1/U2AF65/3' splice site of pre-mRNA complex. Role of SF1^{HH} and the SF1^{NTD}-U2AF65^{UHM} interaction in the formation of the 3' splice site recognition complex. SF1^{HH} may establish an optimal orientation of the SF1 (KH-QUA2) and U2AF65^{RRM1,2} RNA-binding subunits in the complex and thereby support cooperative RNA binding. Tandem phosphorylation of SF1 might contribute to the formation of the ternary complex by stabilizing a yet unknown interface.

Here we show that the N-terminal region of SF1 (SF1^{NTD}) adopts a helix hairpin fold. RNA binding assays show that SF1^{HH} is essential for formation of the ternary SF1-U2AF65-RNA complex, as deletion of the helix hairpin abolishes cooperative RNA binding. Interestingly, deletion of either of the two α -helices does not abrogate formation of the ternary complex, although complex formation is reduced in the absence of helix $\alpha 2$ (Figure 3.15). This suggests that first, SF1^{HH} mainly acts as spacer between the U2AF65^{UHM}-bound SF1^{ULM} and the SF1^{KHQUA2} region and may thus provide an optimal orientation of the proteins within the ternary complex (Figure 3.17). While deletion of both α -helices strongly shortens the distance between the ULM and KH-QUA2 regions in SF1, and therefore may not allow proper arrangement of these regions needed for cooperative RNA binding, the presence of one of the two helices is able to rescue this effect. Second, residues located in helix 2 might be involved in stabilization of the ternary complex. This is based on our observation that deletion of helix 1 did not appear to effect formation of the SF1-U2AF65-RNA complex, whereas deletion of helix 2 did (Figure 3.15).

Our NMR structural studies reveal that U2AF65^{UHM} forms a secondary interface with SF1^{NTD} in addition to the previously reported interaction with SF1^{ULM}. The additional interface locks the relative orientation of U2AF65^{UHM} and SF1^{NTD} and is thus likely to be important for the

specific quaternary arrangement of the proteins in the SF1-U2AF65 complex. In addition to providing the proper geometry of SF1 and U2AF65 for RNA binding, the secondary interface involving the SF1^{HH} and U2AF65^{UHM} may reduce entropic costs linked to SF1-U2AF65-RNA complex formation by providing a pre-arranged protein-protein scaffold for RNA binding. Thus, although the secondary interface has a rather small contribution to the overall affinity between SF1 and U2AF65 (Table 3.2), it is an essential feature for formation of the ternary complex. A specific arrangement of SF1 and U2AF enforced by the helix hairpin of SF1 may be required to accommodate variations in the distance between the BPS and Py tract regions of introns by providing a prearranged scaffold of the SF1^{KH-QUA2} and U2AF65^{RRM1,2} regions.

3.2 Structural analysis of SF1 phosphorylation

3.2.1 Phosphorylation mimics of SF1 (S80E, S82E)

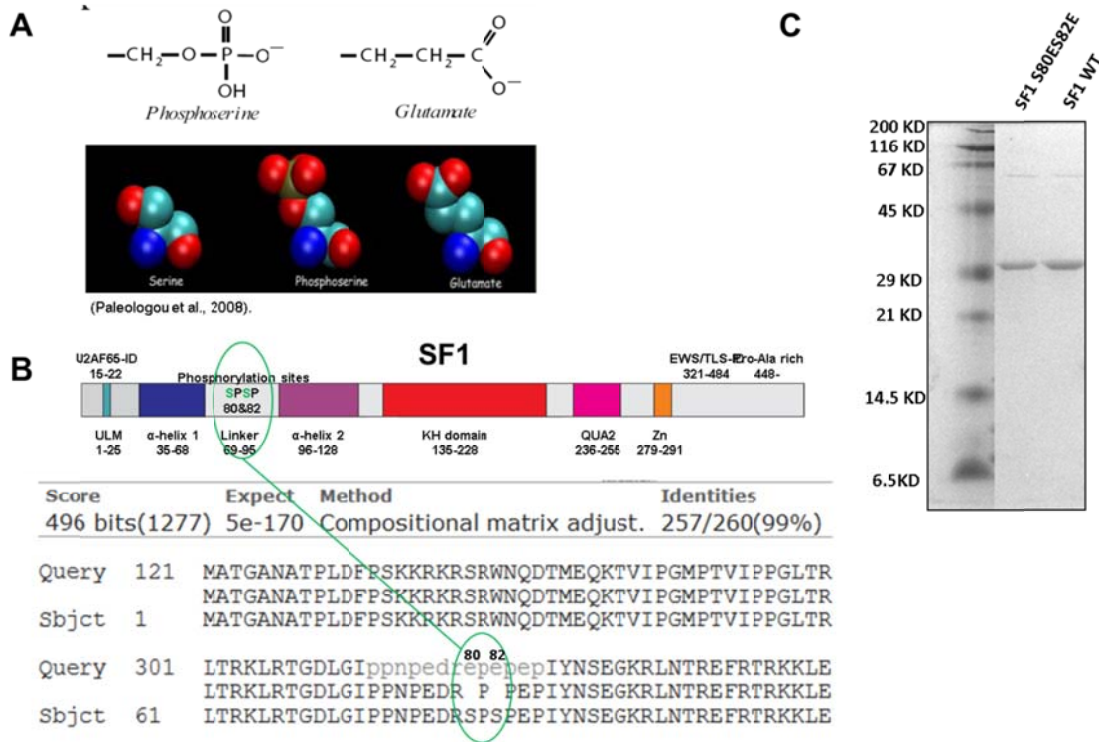


Figure 3.18 Mimicking phosphorylation at Ser80 and Ser82 of SF1. (A) Illustration of structural and charge similarities between phosphoserine and glutamate (Paleologou et al., 2008). Both phosphoserine and glutamate are shown negatively charged. Cartoon representation of molecular structures of phosphoserine and glutamate indicates they are highly similar. (B) Design of SF1 mutant (S80E/S82E). The DNA sequencing results are translated and then aligned in the protein database (swissprot) using BLASTx program (Altschul et al., 1997). The construct of SF1 is shown in cartoon representation on top, the mutation sites are highlighted in green and circled. (C) SDS-PAGE analysis of SF1 mutant. The SF1 wildtype was used as a control.

Phosphorylation of two serine residues (Ser80 and Ser82) within the linker connecting the two helices in SF1^{NTD} has been reported recently to enhance formation of the ternary SF1-U2AF65-RNA complex (Manceau et al., 2006). To study potential structural effects linked to phosphorylation, we employed a phosphorylation-mimics method, which replaces SF1 Ser80 and Ser82 with glutamates to simulate the phosphorylation effects based on the conformational and charge similarities between phosphoserine and glutamate (Figure 3.18A). The replacements for Ser80 and Ser82 were achieved by the site-directed mutagenesis, and the products were checked with DNA gel electrophoresis and further

confirmed by DNA sequencing (Figure 3.18B). The protein expression of the mutant (S80E&S82E) was similar to the wide type, no degradation problem was presented on the SDS-PAGE gel picture (Figure 3.18C).

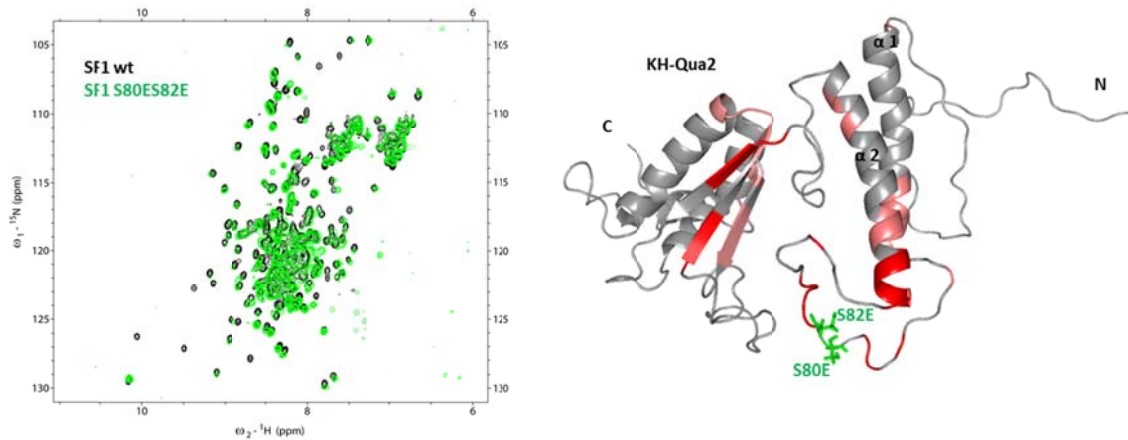


Figure 3.19 Superposition of the ^1H ^{15}N HSQC spectra of SF1 wildtype (black) and SF1 mutant (S80E&S82E; green) and mapping of the CSP on the structure. The NMR signals of SF1 mutant show the overall similarity to that of SF1 wildtype, however NMR line broadening of SF1 mutant was observed, which indicates local structural changes. And the affected residues are highlighted in red on the 3d structure.

NMR spectra of the SF1 mutant displayed a similar overall signal pattern to that of SF1 wide type (Figure 3.20), indicating that the replacement of Ser80 and Ser82 with glutamates did not destroy the overall confirmation of SF1. However, several peaks exhibited NMR line broadening or chemical shift perturbations, implying local structural changes of SF1 mutant. Using a set of standard triple resonance NMR experiments (Sattler et al., 1999), the residues were re-assigned and then the strongly affected ones were plotted onto 3D structure (Figure 3.19). Most chemical shift changes are observed in the flexible loop linker region close to the mutation sites at Ser80 and Ser82 as well as at the N-terminal end of helix $\alpha 2$ (Figure 3.19). The results indicate that the phosphorylation-mimics drive an intra-molecular interaction within SF1.

We also recorded NMR spectra of single mutant of SF1 (S80E or S82E) to investigate the simulation of individual phosphorylation of Ser80 or Ser82. The chemical shift of Ser80 is not affected in SF1 (S82E) mutant, single mutation of SF1 (S80E) changes the chemical shift of Ser82; however both CSP are similar to that of SF1 (S80ES82E) mutant, indicating mutation of S80E and S82E have equivalent influences on conformational changes of SF1. These

results are in good agreement with the biochemical assays in previous reports (Manceau et al., 2006).

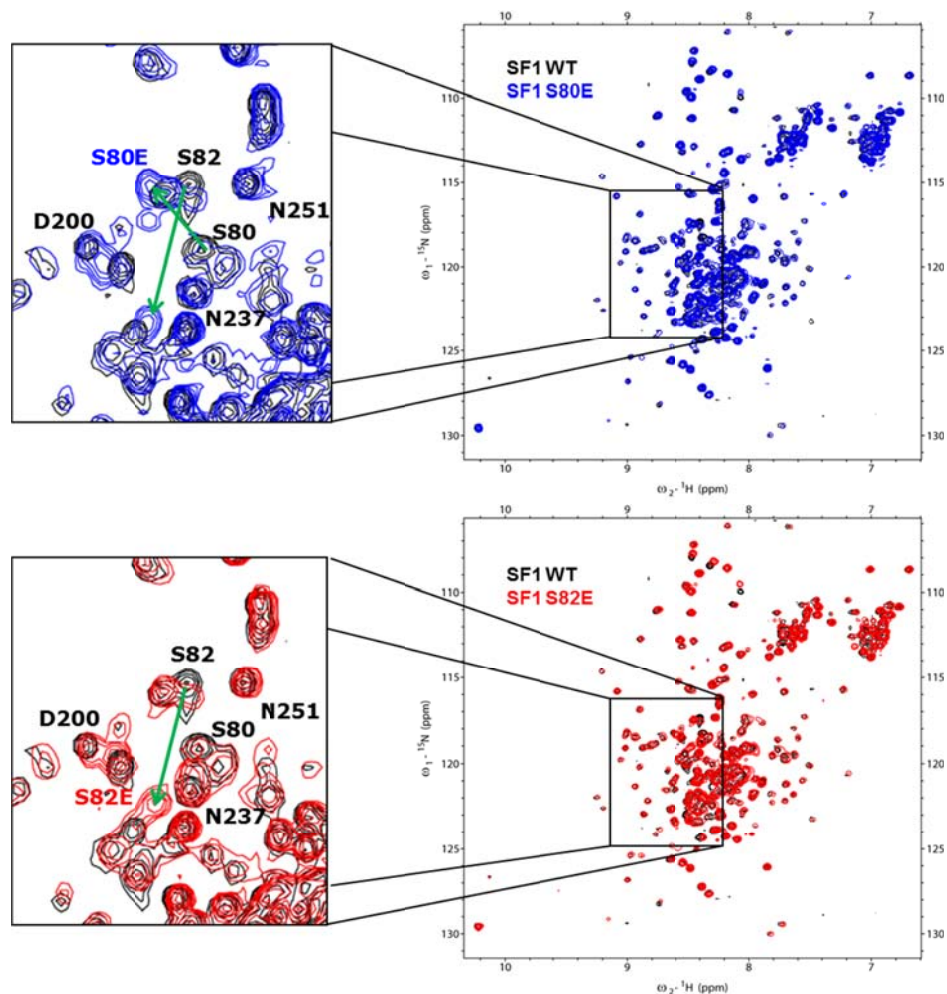


Figure 3.20 Overlay of the ^1H ^{15}N HSQC spectra of SF1 mutants (S80E; blue and S82E; red) with SF1 wildtype (black). The overall signal pattern are similar between SF1 mutants and SF1wt. Single mutation of Ser80 changes chemical shift of Ser82, however single mutation of Ser82 doesn't affect Ser80. The shift of Ser82 in the SF1 S80E is similar to that of SF1 S82E mutant.

3.2.2 Phosphorylation of SF1 on S80 and S82

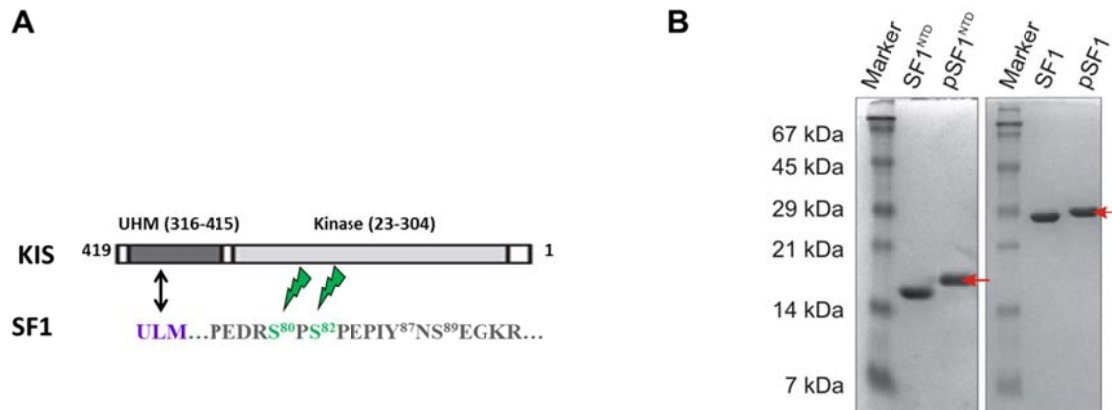


Figure 3.21 *in vitro* phosphorylation of SF1^{NTD} and SF1 with recombinant UHM-KIS kinase. **(A)** The illustration of *in vitro* phosphorylation of SF1 with UHM-KIS kinase, the UHM domain of KIS recognizes SF1 ULM domain, which facilitates the double-phosphorylation of SF1 upon Ser80 and Ser82 by the kinase domain. **(B)** SDS-PAGE gel analysis of phosphorylation of SF1^{NTD} and SF1. The phosphorylated protein (red arrow) migrates more slowly on the gel than the non-phosphorylated.

To understand more structural details related to the phosphorylation upon SF1 ser80 and ser82, we prepared phosphorylated SF1^{NTD} (pSF1^{NTD}) and SF1 (pSF1) by *in vitro* phosphorylation with recombinant UHM-KIS kinase (Figure 3.21A). The UHM-KIS kinase was produced in our lab according to the protocols, describing in Methods and Materials section 2.2.3. The phosphorylated SF1 were identified by SDS-PAGE (Figure 3.21B), MS (Figure 3.22A) as well as NMR (Figure 3.24). SDS-PAGE shows different migration pattern for the phosphorylated SF1, phosphorylation-related conformational changes in SF1 (that persist in detergent-containing solutions) are thought to underlie this phenomenon. Most of the phosphorylation sites for which such a mobility shift has been described fall in the category of SP and TP sites (i.e. a proline residue follows the phosphorylated serine or threonine residue), therefore SPSP (Ser80-Pro81-Ser82-Pro82) motif in SF1 provides suitable acceptors for the specific phosphorylation.

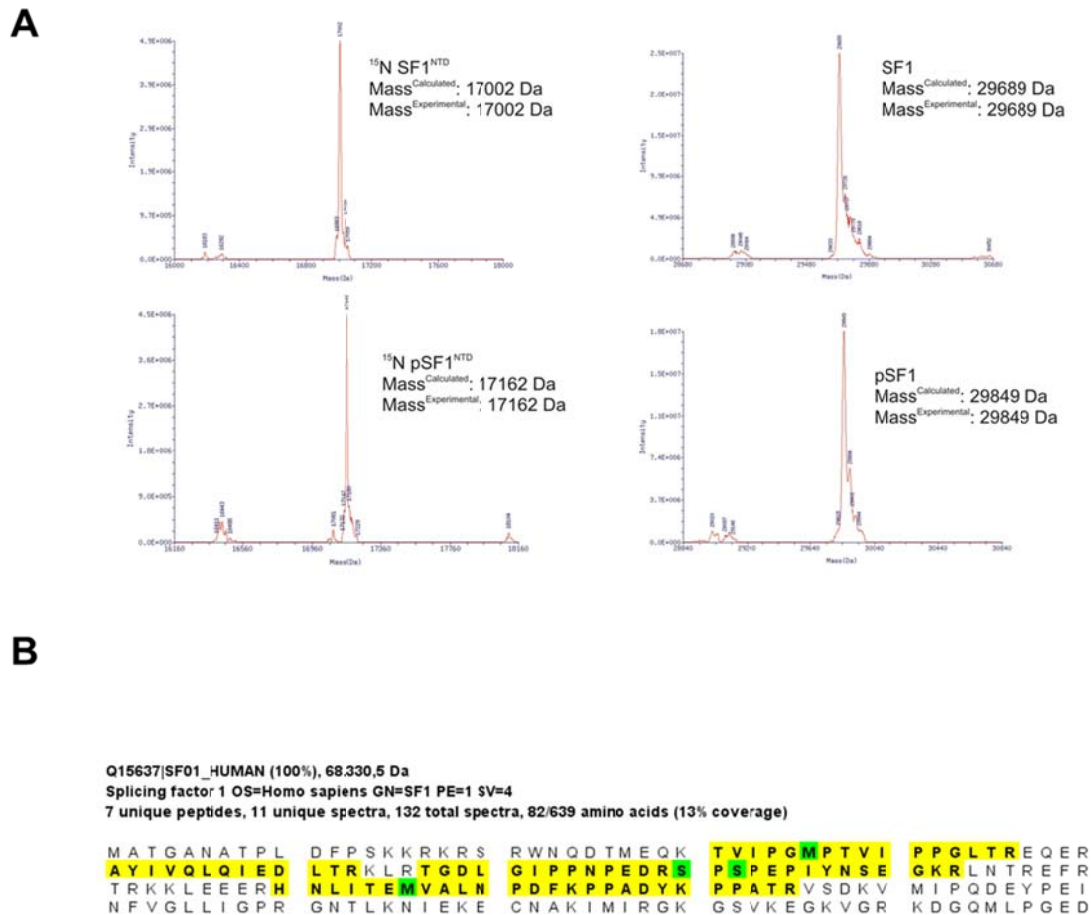


Figure 3.22 Mass spectrometry analysis of SF1 phosphorylation. (A) Deconvoluted spectra of non-phosphorylated ^{15}N -SF1^{NTD} and SF1 (top left and top right) are shown in comparison to their phosphorylated state (bottom left and right). The mass difference of 160 Da confirms the addition of exactly two phosphate groups. (B) Analysis of SF1^{NTD} peptides with MS. The modified residues are highlighted in green. The results indicate that the phosphorylation is restricted to Ser-Pro-Ser-Pro motif. The modification of several methionine residues is due to the oxidation during the MS measurement.

To quantify the phosphorylated residues within SF1 and to confirm the KIS kinase specifically phosphorylated Ser-Pro-Ser-Pro motif, we measured the mass difference between phosphorylated and unphosphorylated SF1^{NTD} (SF1) (Figure 3.22A) and we further analyzed the peptides of phosphorylated SF1^{NTD} with Mass spectrometry (Figure 3.22B). The phosphorylated SF1^{NTD} was digested with trypsin, and the peptides/modified residues were determined by MS. The results indicate double-phosphorylation of SF1 with KIS kinase, which is specifically restricted to the Ser-Pro-Ser-Pro motif.

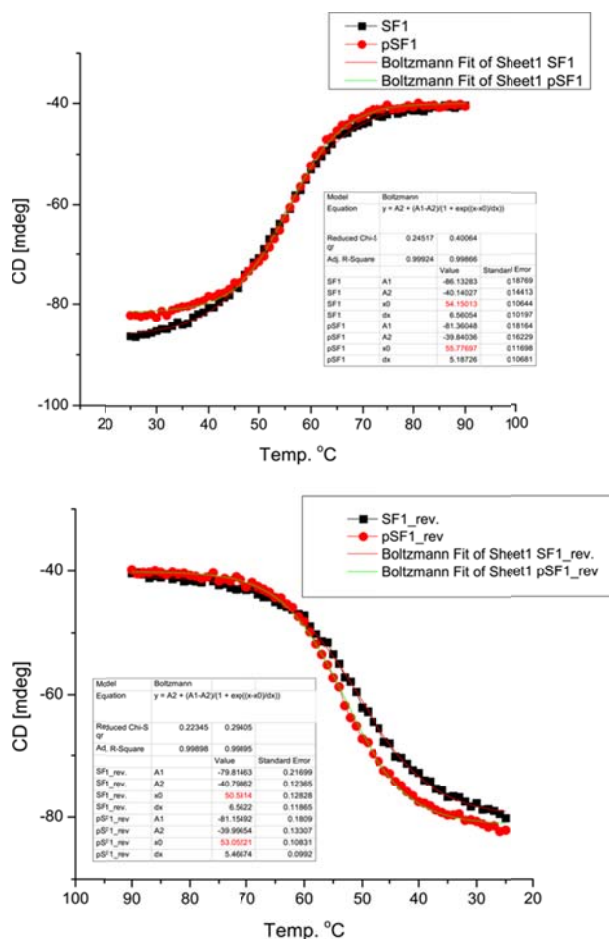


Figure 3.23 Analysis of thermodynamics of phosphorylated SF1 by Circular Dichroism (CD) spectroscopy.

Circular dichroism was measured on non-phosphorylated and phosphorylated SF1. The temperature was varied between 25 and 90 degree centigrade. The melting temperature (T_m) of SF1 (top, black curve) slightly increases from 54.2 to 55.8 degree centigrade of the phosphorylated SF1 (top, red curve). Reverse measurements display the small difference between non-phosphorylated (bottom, red curve) and phosphorylated SF1 (bottom, black curve).

The effects of the phosphorylation on the SF1 protein stability was addressed by circular dichroism (CD) assays in combination with thermal denaturation. The phosphorylated and unphosphorylated SF1 were unfolded by increasing the temperature. The results showed a slightly increase of T_m of pSF1, indicating that the phosphorylation might stabilize some local regions of SF1. The reverse measurement of CD shows that the pSF1 recovered faster than SF1.

3.2.3 NMR analysis of SF1^{NTD} and SF1 phosphorylation

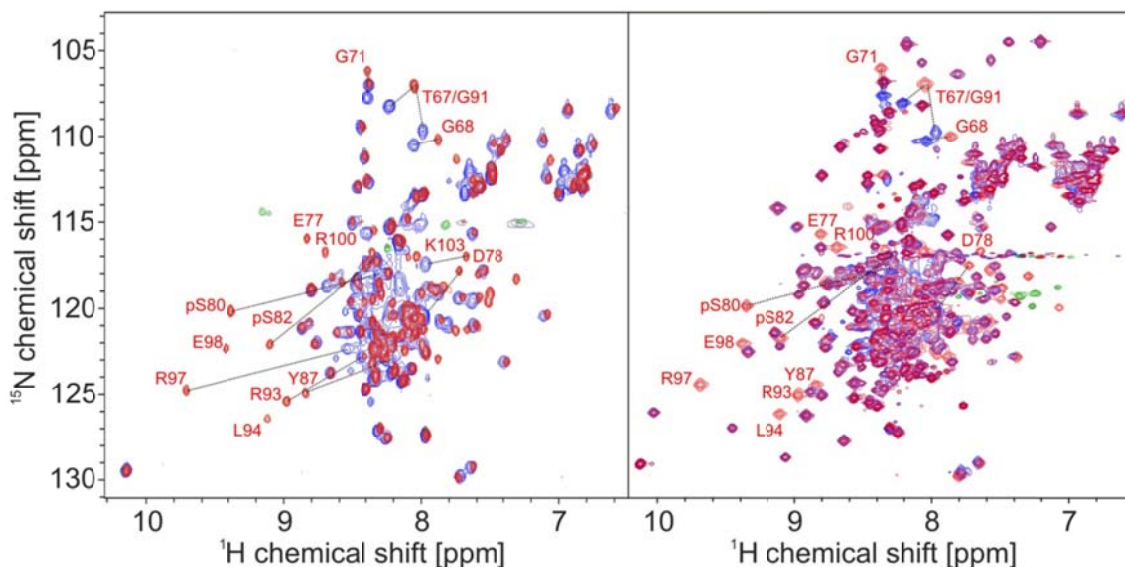


Figure 3.24 Comparison of the ^1H ^{15}N HSQC spectra of SF1^{NTD} and SF1 with their phosphorylated states. NMR spectra of phosphorylated SF1^{NTD} and SF1 (red) exhibit obvious differences with non-phosphorylated samples. The signal of phosphorylated serines (pS80 and pS82) migrated to the downfield region, in a good agreement with the literature results. Selected residues, which are shifted upon phosphorylation are annotated.

The NMR spectra of phosphorylated SF1^{NTD} (pSF1^{NTD}) and SF1 (pSF1) further verify the specific phosphorylation upon Ser80 and Ser82 in SF1. An overlay of NMR spectra comparing non-phosphorylated and phosphorylated proteins (Figure 3.24) reveals large chemical shift changes linked to Ser80/Ser82 tandem phosphorylation. NMR chemical shifts were re-assigned based on a set of standard triple resonance NMR experiments (Sattler et al., 1999). Similar to phosphorylation mimicking experiments, most CPSs are observed in close proximity to the phosphorylation sites at Ser80 and Ser82 and at the N-terminal end of helix $\alpha 2$ (Figure 3.25). Strongly affected signal include many positively charged residues in helix $\alpha 2$ such as Arg97 and Arg100, indicating that the electrostatic effect might mediate the interactions between the phosphate groups and the arginine residues.

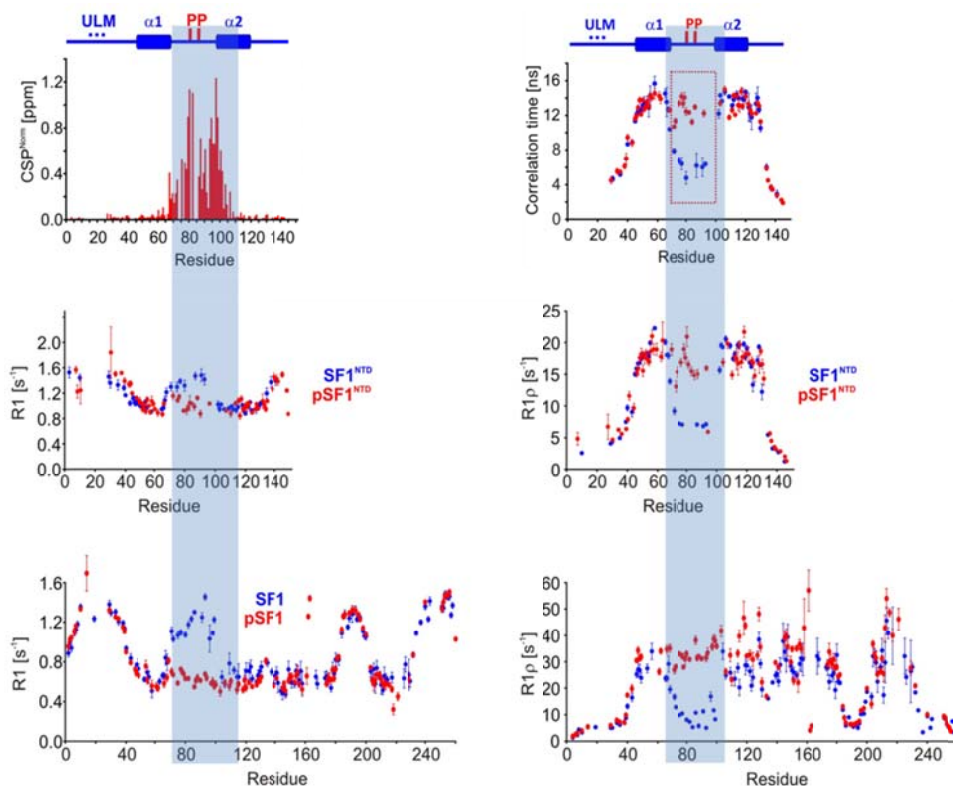


Figure 3.25 NMR and dynamic analysis of the effect of tandem serine phosphorylation of SF1. Chemical shift perturbations (top left, red) and residue-specific local correlation times τ_c are shown for SF1^{NTD} (top right, blue) and pSF1^{NTD} (top right, red), large chemical shift perturbations are observed in close proximity to the phosphorylation sites at Ser80 and Ser82 and at the N-terminal end of helix α_2 . ¹⁵N backbone amide longitudinal (R_1) and transverse relaxation rates $R_{1\rho}$ for non-phosphorylated (blue) and phosphorylated (red) SF1^{NTD} (middle) and SF1 (bottom), respectively. Relaxation data was measured at 600 and 800 MHz proton larmor frequency for (p)SF1^{NTD} and (p)SF1, respectively. The secondary structures and the phosphorylation sites are depicted above the diagram. The tandem phosphorylated linker region which rigidifies and alters R_1 and $R_{1\rho}$ relaxation rates upon phosphorylation is highlighted with grey boxes.

Although strong changes for the residues close to Ser80 and Ser82 are observed, phosphorylation has limited effects on the overall conformation of SF1. For the phosphorylated SF1, residue-specific local correlation times τ_c of the linker region is similar to that of the two helices, suggesting phosphorylation induces the linker region and the two helices tumbling as a unit. ¹⁵N R_1 and $R_{1\rho}$ relaxation rates and local tumbling correlation times indicate that the linker connecting helices α_1 and α_2 becomes rigid upon phosphorylation (Figure 3.25).

3.2.4 Spin labeling

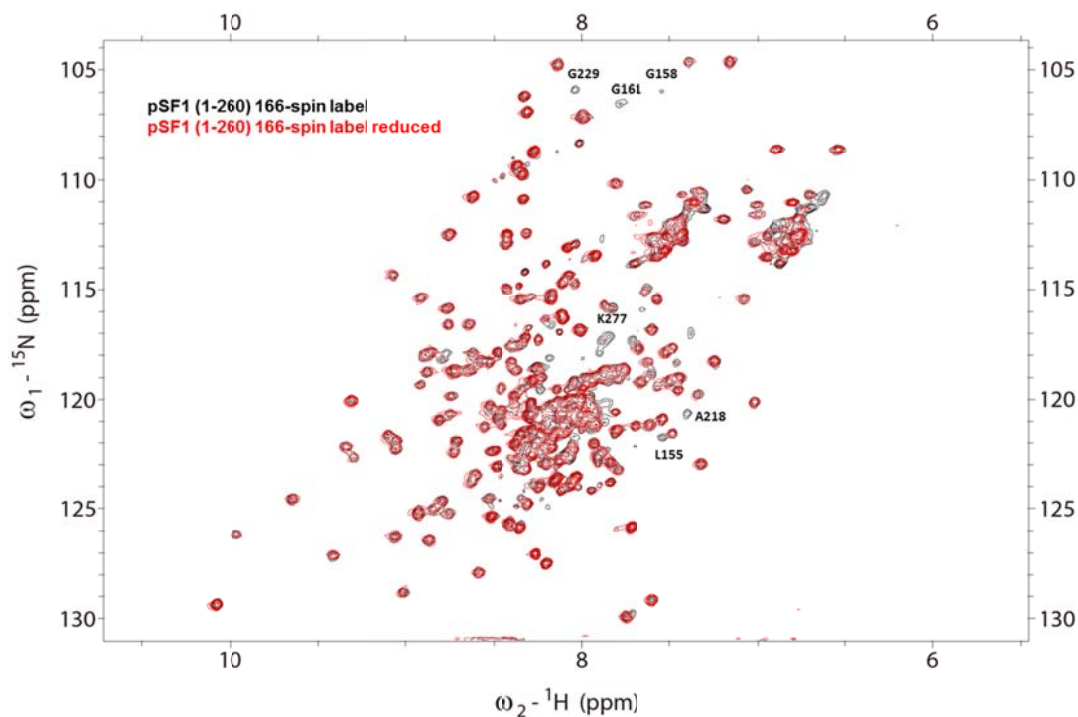


Figure 3.26 Overlay of the ^1H ^{15}N HSQC spectra of phosphorylated SF1 with a spin label at position 166 in the oxidized (red) and reduced state (black). The single cysteine variant SF1 $\Delta\text{c}/\text{N166C}$ with iodoacetamide proxyl attached to 166C shows line broadening compared to the reduced protein.

Paramagnetic relaxation enhancement was measured for an iodoacetamide proxyl spin label attached to the phosphorylated and non-phosphorylated single cysteine variants S20C, S137C and N166C in SF1. For the spin label at N-terminal position 20, the $I_{\text{ox}}/I_{\text{red}}$ ratio is reduced only for amide protons of residues in the proximity to 20, indicating the N-terminus is spatially away from the other domains of SF1. The pattern is highly similar for phosphorylated and non-phosphorylated SF1 (Figure 3.27). Addition of the spin label to position 137 lowers the $I_{\text{ox}}/I_{\text{red}}$ ratio for the residues 128 to 146 adjacent to the spin-label position and for amide protons in helix α_2 as well as in KH-QUA2 domain. In this position, the PRE effects are similar for phosphorylated and non-phosphorylated SF1 with the exception of helix α_1 with a slightly reduced ratio when SF1 is phosphorylated (Figure 3.27). Attachment of the spin label to position 166, the $I_{\text{ox}}/I_{\text{red}}$ ratio is strongly reduced between residue 147 and 179 and additionally in the C-terminal of KH-QUA2 domain. For this spin

label position, the differences of the PRE effects for phosphorylated and non-phosphorylated SF1 are mainly in the helix $\alpha 2$ (Figure 3.27).

Position 137 and 166 are located in the N-terminal and C-terminal boundary of KH-QUA2 domain respectively, thus analysis of PRE effects of the spin label at the two positions indicates the relative movements between the SF1^{NTD} and KH-QUA2 domains driven by phosphorylation. Mapping of the PRE effects on the solution structure of SF1 shows that the affected residues upon phosphorylation are located within the expected distance to the spin label (Figure 3.27).

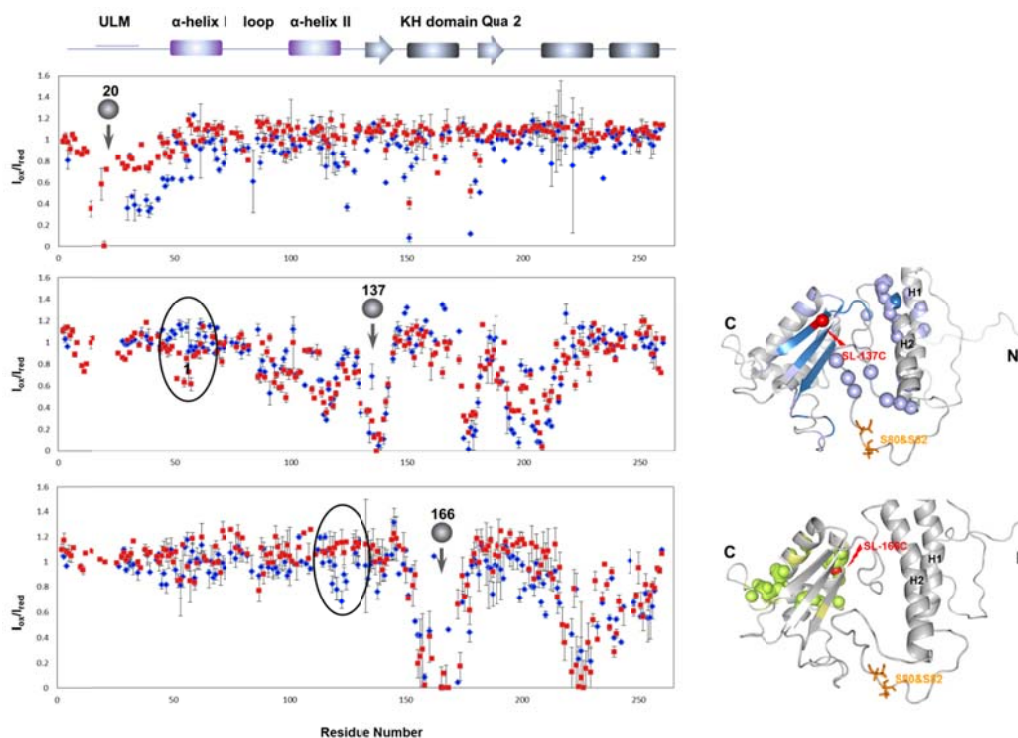


Figure 3.27 Paramagnetic relaxation enhancement of SF1 with and without phosphorylation with a spin-label at residue 20, 137 and 166. I_{ox}/I_{red} ratios of amide protons with the spin label at position 20 (top), 137 (center) and 166 (bottom) of non-phosphorylated (blue) and phosphorylated (red) SF1 plotted against the amino acid sequence. The position of the spin label is indicated with a sphere and an arrow. The residues with different PRE effects for the phosphorylated and non-phosphorylated SF1 are circled. Mapping of I_{ox}/I_{red} ratios of phosphorylated SF1 with a spin label at 137 (center right) and at 166 (bottom right) on the NMR solution structure of phosphorylated SF1 using a purple-blue and a yellow-green gradient respectively. The spin label sites are highlighted in red and the serine residues at the phosphorylation sites are shown in stick representation and highlighted in yellow. The secondary structures and the phosphorylation sites are depicted above the diagram.

3.2.5 Effects of phosphorylation on SF1 structure

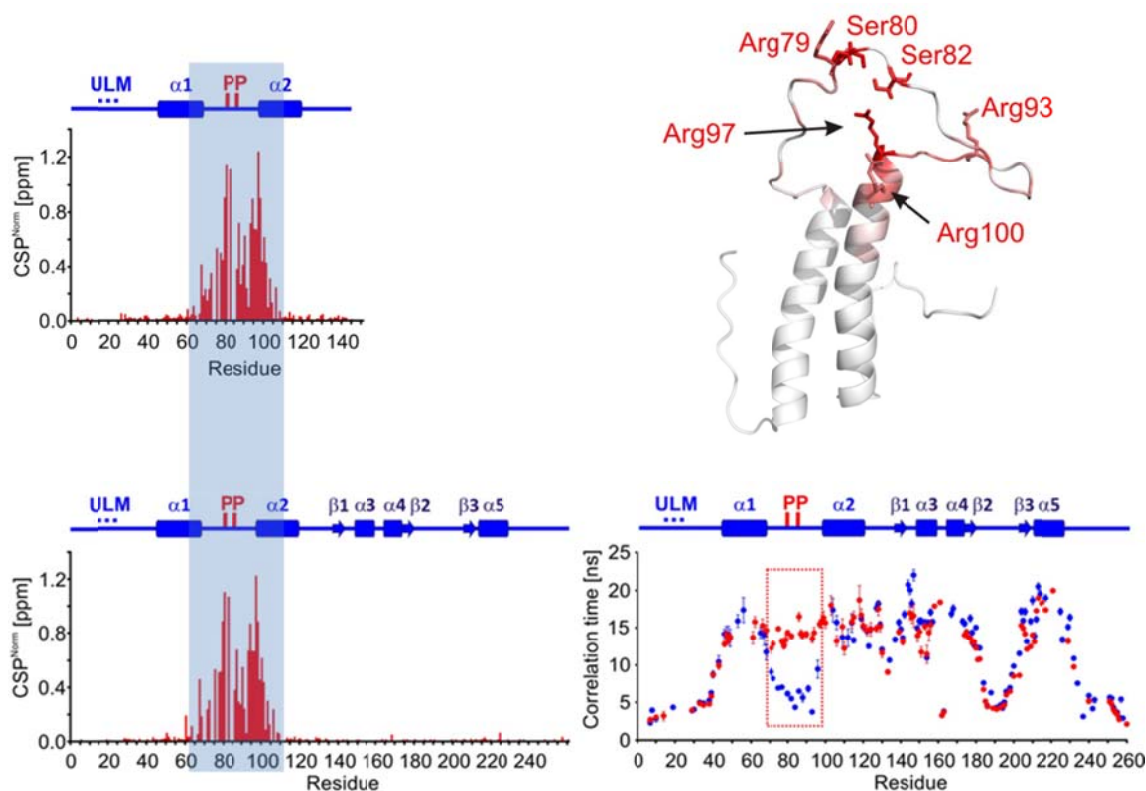


Figure 3.28 Impact of phosphorylation on SF1 structure. Chemical shift perturbations induced by phosphorylation are shown for SF1^{NTD} (top left) and SF1 (bottom left). CSPs are very similar for SF1^{NTD} and SF1 and are mainly localized within the SF1 helix hairpin domain (highlighted in gray box). A ribbon representation of SF1^{HH} color-coded with the phosphorylation-induced CSPs is shown (top right). Residue-specific local correlation times τ_c (bottom right) are compared between SF1 (blue) and pSF1 (red). The secondary structures and the phosphorylation sites are depicted above the diagram. The tandem phosphorylated linker region which rigidifies upon phosphorylation is framed.

We next tested the structural impact of phosphorylation on SF1, as well as on the SF1^{NTD}-U2AF65^{UHM} and SF1-U2AF65^{RRM123} complexes using NMR and SAXS. NMR assignments of pSF1 spectra and dynamic analyses suggest that the side chains of Arg93 (linker), Arg97 and Arg100 ($\alpha 2$) mediate salt bridges with the two phosphate groups in the SF1^{HH} linker and thereby strongly reduce the conformational flexibility of this linker. CSPs induced by phosphorylation are very similar for SF1^{NTD} and SF1 (Figure 3.28), and are mainly localized within the SF1 helix hairpin domain. This indicates that phosphorylation does not induce strong intra-molecular contacts between the NTD and the KH-QUA2 regions of SF1. A

comparison of local mobility along the protein backbone derived from ^{15}N NMR relaxation data of phosphorylated SF1 with the non-phosphorylated protein (Figure 3.28) reveals that the overall backbone dynamics of the protein does not significantly change upon phosphorylation. SAXS data further corroborate the NMR results. The radii of gyration (R_g) of SF1 and pSF1 are comparable (32.5 ± 0.3 vs. 29.1 ± 0.3 Å) and only minor differences are seen in the pairwise distance distribution functions ($P(R)$, Figure 3.29). Although these changes might indicate compaction of a minor fraction of “open” species in the non-phosphorylated protein, the NMR relaxation data and the absence of phosphorylation-induced CSPs in SF1^{KHQUA2} indicate that phosphorylation does not significantly affect the conformation of SF1.

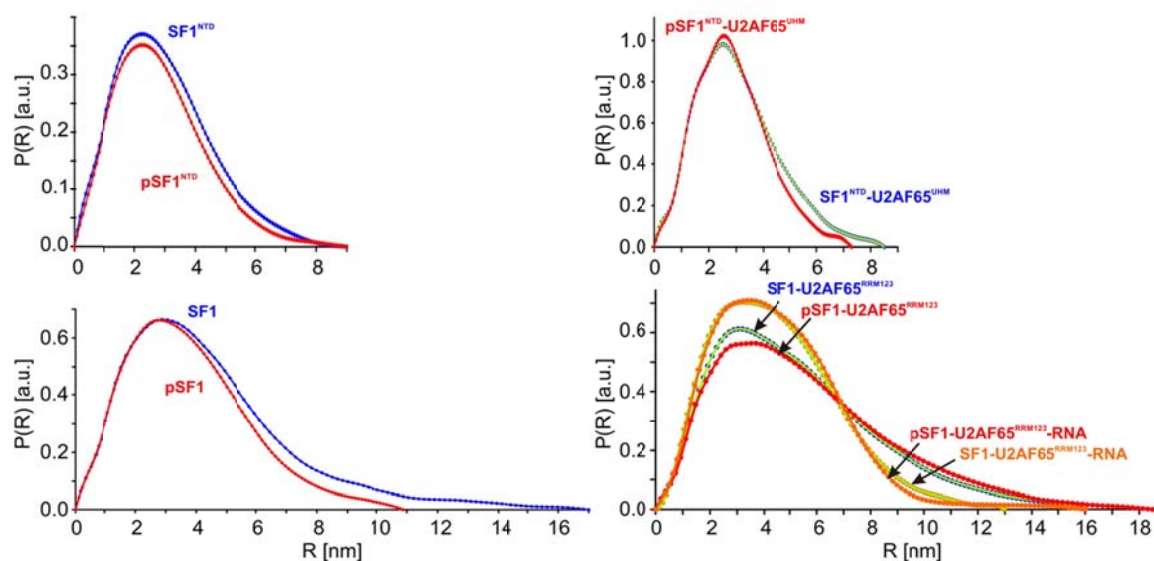


Figure 3.29 SAXS analysis of the effect of tandem serine phosphorylation of SF1. SAXS data showing comparisons of radial density distributions of non-phosphorylated and phosphorylated SF1^{NTD} and SF1, in complex with U2AF65^{UHM}, U2AF65^{RRM123}, and with U2AF65^{RRM123}-RNA, respectively. Tandem serine phosphorylation of SF1 introduces only minor differences to the overall arrangement of the SF1-U2AF65 complexes (with or without RNA).

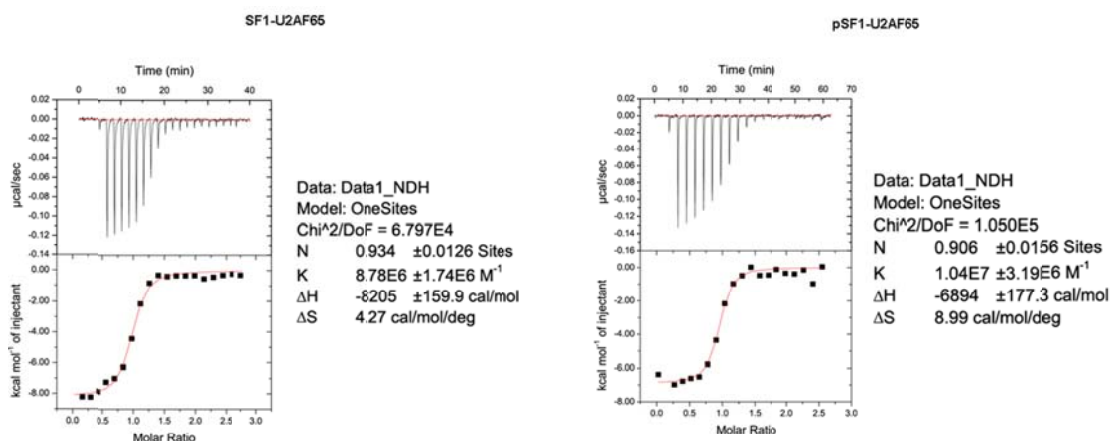
We next studied the impact of phosphorylation on the overall structure of the SF1^{NTD}-U2AF65^{UHM}, SF1-U2AF65^{RRM123} and the ternary SF1-U2AF65^{RRM123}-RNA complexes using SAXS analysis. Only minor differences are observed for the pairwise distribution functions of the two protein complexes (Figure 3.29) and the derived radii of gyration (Table 3.4). In contrast, and similar to a recent report (Gupta et al., 2011) binding of a 20-mer RNA induces large overall changes for the structure and/or dynamics of the SF1-U2AF65^{RRM123}

complex and leads to formation of a compact SF1-U2AF65^{RRM123}-RNA arrangement with $R_g = 39.4 \pm 0.4$ and 34.2 ± 0.1 Å in the absence and presence of the RNA ligand, respectively (Table 2). Notably, tandem serine phosphorylation of SF1 introduces only minor differences to the overall arrangement of the SF1-U2AF65 complexes (with or without RNA) (Figure 3.29). This indicates that SF1 Ser80/82 phosphorylation is mainly limited to minor conformational changes within SF1 while RNA binding leads to a large change in the overall structure and/or dynamics of the U2AF65-SF1 complex.

Table 3.4 SAXS data and analysis

Sample	R_g [nm]	D_{max} [nm]
SF1 ^{NTD}	2.24 ± 0.11	7.8
pSF1 ^{NTD}	2.18 ± 0.11	7.6
SF1	3.25 ± 0.16	10.8
pSF1	2.91 ± 0.09	10.2
SF1 ^{NTD} -U2AF65 ^{UHM}	2.48 ± 0.03	8.7
pSF1 ^{NTD} -U2AF65 ^{UHM}	2.32 ± 0.02	7.6
SF1-U2AF65 ^{RRM123}	3.94 ± 0.04	14.0
pSF1-U2AF65 ^{RRM123}	4.18 ± 0.04	14.0
SF1-U2AF65 ^{RRM123} -RNA	3.42 ± 0.01	11.0
pSF1-U2AF65 ^{RRM123} -RNA	3.36 ± 0.01	11.0

3.2.6 Role of phosphorylation of SF1 for cooperative RNA binding



Sample	Binding affinity [K _d]
SF1-U2AF65 ^{RRM123}	114 ± 23 nM
pSF1-U2AF65 ^{RRM123}	96 ± 32 nM

Figure 3.30 ITC data of (pSF1) SF1 in complex with U2AF65. Measurements were carried out at 25 °C. Dilution heats determined by titrating proteins into the corresponding buffer were in the range of the heat effects observed at the end of the titration (data not shown) and were subtracted for the analysis.

A comparison of ITC data suggests that tandem serine phosphorylation of SF1 induces no changes on binding affinity to U2AF65, with the dissociation constants of U2AF65 for SF1 and pSF1 of $K_d = 114 \pm 23$ and 95 ± 32 nM, respectively (Figure 3.30).

The role of phosphorylation of SF1 for cooperative binding of SF1 and U2AF65 to the pre-mRNA was tested in EMSAs. Increasing concentrations of U2AF65^{RRM123} added to a 3' splice site RNA result in a smear of U2AF65-RNA complexes (Figure 3.31). SF1 and pSF1 also binds the RNA, although weakly. In the presence of both proteins a ternary SF1-U2AF65-RNA complex forms at lower U2AF65 concentrations, consistent with cooperative binding. EMSA results of phosphorylated SF1 show a higher efficiency of ternary complex formation with U2AF65^{RRM123} than non-phosphorylated SF1 (Figure 3.31), consistent with the results of Manceau et al. (Manceau et al., 2006).

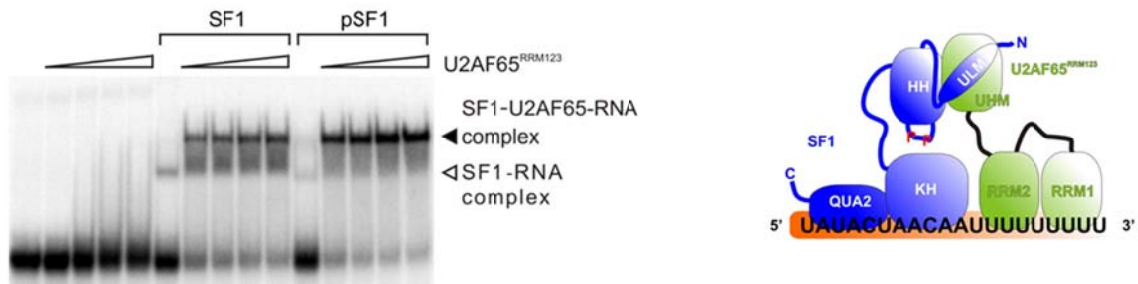


Figure 3.31 Cooperative binding of U2AF65^{RRM123} and (p)SF1 to a 3' splice site RNA. RNA was incubated with buffer or U2AF65^{RRM123} (0.2, 0.5, 1 and 2 μ M; indicated by triangles) in the absence or presence of 6.6 μ M SF1 or pSF1. Reaction products were separated by native PAGE and visualized by autoradiography. The migration of SF1-U2AF65-RNA complexes (closed arrowhead) and SF1-RNA complexes (open arrowhead) is indicated. Assembly of the ternary complex is illustrated in cartoon representation on the right. The tandem phosphorylation sites are indicated in red.

3.2.7 Discussion: The influence of phosphorylation of SF1 on cooperative assembly of the SF1-U2AF65-RNA complex

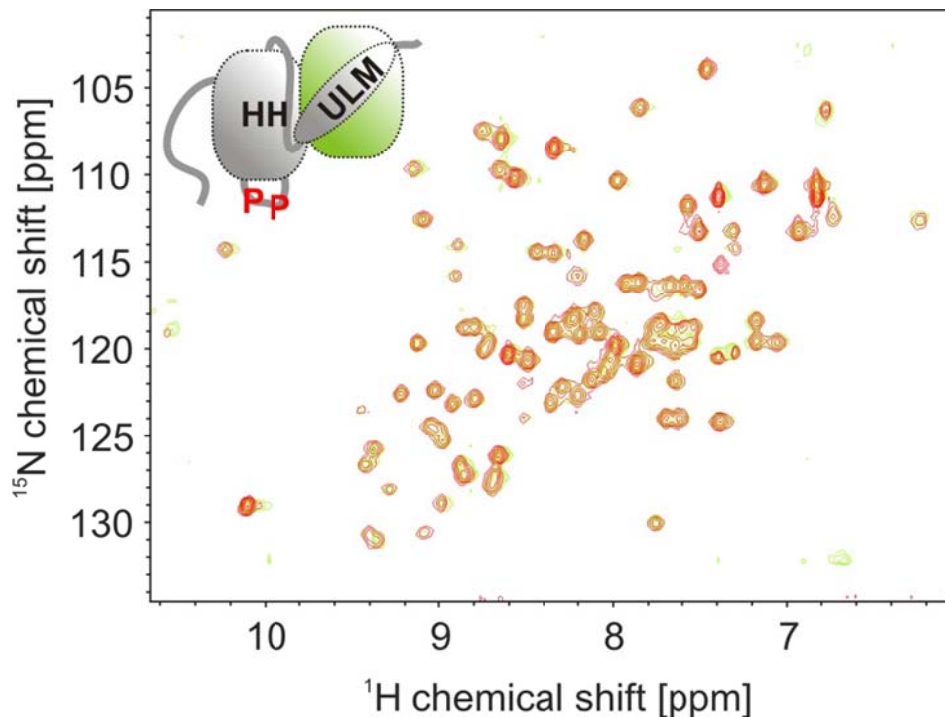


Figure 3.32 Effect of SF1^{NTD} phosphorylation on U2AF65^{UHM}. Superposition of ¹H, ¹⁵N HSQC NMR spectra of U2AF65^{UHM} in complex with unlabeled non-phosphorylated (green) and phosphorylated SF1^{NTD} (red), respectively. No residues are shifted upon SF1^{NTD} phosphorylation, indicating that the phosphorylated residues in SF1 are remote from the SF1^{NTD}/U2AF65^{UHM} interface.

Our data show that tandem serine phosphorylation of the conserved SPSP site within the linker connecting SF1^{NTD} helices α 1 and α 2 has little effect on the conformation and overall structure of SF1 alone, or the SF1^{NTD}-U2AF65^{UHM} and the SF1-U2AF65^{RRM123} complexes. NMR and ITC binding data also suggest that phosphorylation does not alter the protein-protein binding affinities and interactions (Figure 3.30, Figure 3.32) and thus further corroborate these results. Nevertheless, the EMSA data indicate that SF1 phosphorylation enhances cooperative assembly of the SF1-U2AF65-RNA complex (Figure 3.31) consistent with a previous report (Manceau et al., 2006). As SAXS data do not indicate large conformational rearrangements linked to phosphorylation of SF1 alone or bound to U2AF65 (Figure 3.29, Table 3.4), the contribution to RNA binding must be small. The slight improvement in cooperative assembly could reflect either additional direct contacts with U2AF65 involving the serine-phosphorylated SF1^{HH}-linker or contribute to the overall stability (and rigidity) of

a pre-arranged SF1-U2AF65 complex, by reducing entropy loss linked to RNA binding. Structural studies to analyze these effects in the ternary complex are currently on-going in our lab.

Tandem serine phosphorylation of SF1 may have additional roles beyond a function in cooperative assembly of the constitutive splicing complex. Previous studies in mammalian and yeast systems have established that SF1 mainly affects the kinetics of spliceosome assembly, as genetic or biochemical depletion of SF1 does not abolish splicing (Guth and Valcarcel, 2000; Rutz and Seraphin, 1999). In *S. cerevisiae* it has been shown that SF1 is involved in removing introns with suboptimal splice sites and in nuclear pre-mRNA retention (Galy et al., 2004; Rutz and Séraphin, 2000), whereas knockdown of human SF1 in HeLa cells did not result in a general splicing phenotype (Tanackovic and Kramer, 2005). However, Corioni *et al.* (Corioni et al., 2011) have recently shown that SF1 is not a constitutive splicing factor, but is required for the splicing of certain introns and affects alternative splice site choice. In this respect, SF1 phosphorylation may mediate interactions with other factors that could regulate alternative splicing by modulating 3' splice site recognition. For example, it has been suggested that the phosphorylated SPSP motif of SF1 is recognized by the tandem WW domains of FBP11 (Ingham et al., 2005) and the interaction of the SF1 and FBP11 orthologues of *S. cerevisiae* has been implicated in mediating intron definition by connecting the 5' and 3' splice sites (Abovich and Rosbash, 1997). However, we have not detected any interaction of tandem WW domains with phosphorylated SF1 (data not shown). Nevertheless, the substantial conformational changes in the SF1^{NTD} linker associated with tandem serine phosphorylation by KIS kinase suggest that interactions with so far unknown factors are modulated that may contribute to regulate alternative splicing by SF1.

3.3 NMR characterization of the U2AF65 arginine-serine-rich (RS) domains

3.3.1 Preparation of U2AF65 RS domain

Although RS domains share a similar content of arginine and serine residues, complexes are presumably disordered and do not adopt an ordered tertiary fold. Recent work has shown that RS domain in U2AF65 contacts the pre-mRNA directly via the BPS in the complex E to promote the spliceosome assembly (Hertel and Graveley, 2005; Shen et al., 2004). The molecular mechanisms in the complex, however, are still not clear. Therefore, we have set out to study the structural details of the U2AF65 RS domain and its interactions using NMR spectroscopy.

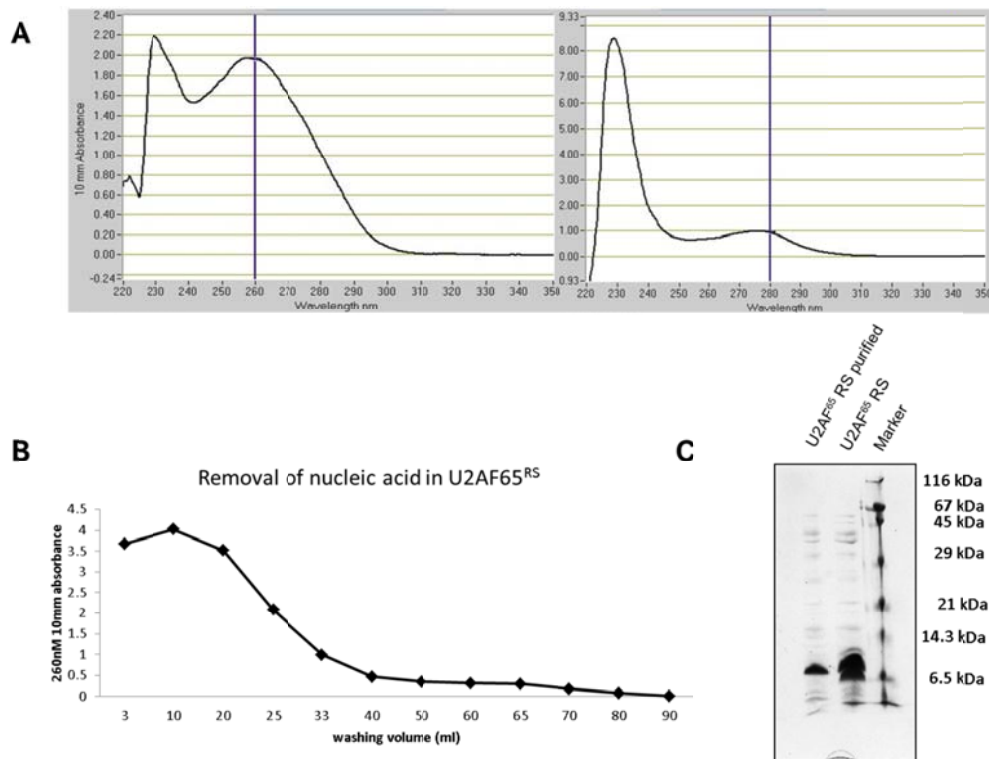


Figure 3.33 Purification of U2AF65 RS protein. (A) U2AF65 RS protein was measured by NanoDrop 1000. The results of RS protein show intensive absorption at 260 nm, suggesting strongly nucleic acid contamination. An additional purification step using RS-washing buffer (comprising 4M Urea and 1M NaCl) is required to clean the protein. (B) The correlation between the washing volume and the removal of nucleic acid is shown. (C) SDS-PAGE of U2AF65 RS protein with nucleic acid contamination removed.

Due to being rich in arginine, the U2AF65 RS protein is highly positively charged and thus strongly binds to negatively charged nucleic acid. The binding of nucleic acid to RS protein is too tight to be removed by the Ni-NTA purification. We then applied an additional purification step to clean RS protein by using RS-washing buffer (comprising 4M Urea and 1M NaCl). Although a small amount of protein was washed away, the nucleic acids contamination was removed, indicated by a decreased absorption at 260 nm (Figure 3.33A). Almost 100 mL RS-washing buffer is required to completely remove the nucleic acids contamination (Figure 3.33B), resulting in loss of the RS protein sample. Therefore, a large amount of RS was prepared to have enough samples for the NMR measurement (Figure 3.33C).

3.3.2 NMR assignment and NMR analysis of U2AF65 RS domain

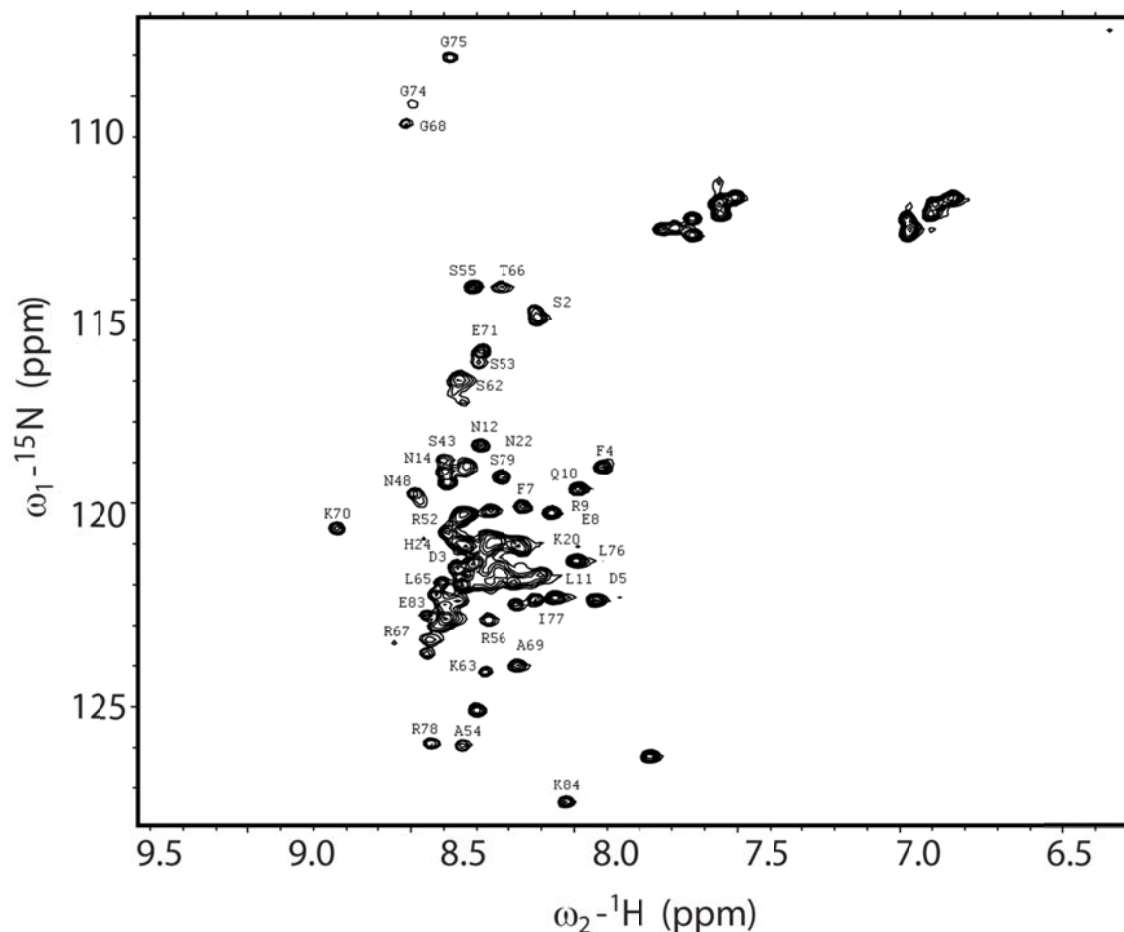


Figure 3.34 ^1H , ^{15}N HSQC spectrum of U2AF65 RS. Signals from the U2AF65 RS (residues 1-84) cluster in the region of the spectrum between 8 and 8.5 ppm, which is a typical pattern for unstructured proteins. NMR assignments of RS were done using a series of standard NMR measurements.

A ^1H , ^{15}N HSQC NMR spectrum was recorded on isotope-labelled U2AF65 RS domain with residues from 1 to 84. The majority of the signals in the ^1H , ^{15}N HSQC NMR spectrum of RS cluster in the region of the spectrum between 8 and 8.5 ppm (Figure 3.34), which is a typical pattern for unstructured protein. NMR assignments of RS were done using a series of standard NMR measurements. A large amount of similar Ser-Arg motifs lead to serious signal overlap. Thus, approximately 30% of NMR signals could not be unambiguously assigned to a certain amide proton in the ^1H , ^{15}N HSQC (Figure 3.34). We then calculated the secondary chemical shifts using the assignments of U2AF65 RS.

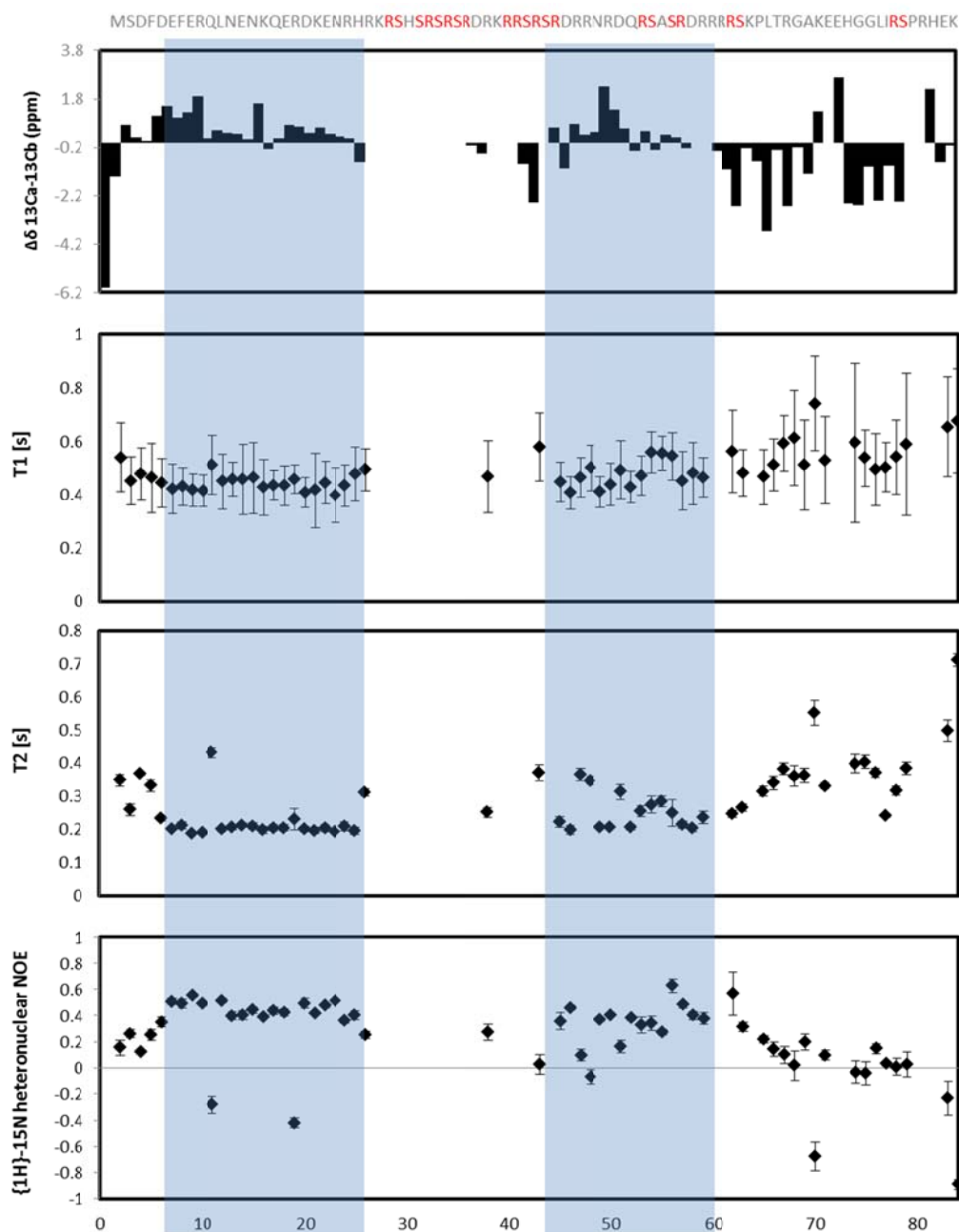


Figure 3.35 Secondary chemical shift analysis and ^{15}N relaxation studies of U2AF65 RS domain plotted against the amino acid sequence. The $^{13}\text{C}_\alpha$ - $^{13}\text{C}_\beta$ secondary chemical shifts of PEX19 CTD (upper panel) indicate that the N-terminal region with residues from 5 to 25 and the central region with residues from 45 to 60 (highlighted in grey boxes) present potential α -helix secondary structure. T_1 and T_2 rates and ^1H - ^{15}N heteronuclear NOE values in panel 2, 3 and 4 demonstrate that the two regions highlighted in grey boxes are less dynamic and possess more rigid properties. The protein sequences are shown in letter representation on top, with RS motifs are colored in red.

$^{13}\text{C}_\alpha$ - $^{13}\text{C}_\beta$ secondary chemical shifts of U2AF65 RS indicate that although the free U2AF65 RS is unfolded in solution, there are some potential structural elements. The N-terminal region with residues from 5 to 25 and the central region with residues from 45 to 60 exhibit continuous positive values in secondary chemical shift (Figure 3.35), which is a typical pattern for α -helix secondary structure. Consistent with these results, although the T1 and T2 relaxation time and ^1H - ^{15}N heteronuclear NOE demonstrate the overall dynamic properties of U2AF65 RS domain in solution, the two potential structure regions which are highlighted in grey boxes show less dynamic and more rigid. For U2AF65 RS ^1H - ^{15}N heteronuclear NOE data describe the two regions as less flexible domains with values of approximately 0.5 and overall flexibility with low rates between 0.3 to -0.9.

3.3.3 U2AF65 RS and SF1-BPS interaction

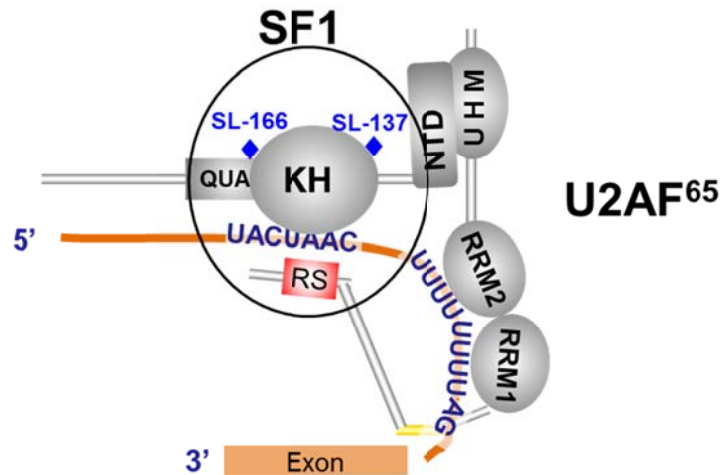


Figure 3.36 Interactions between U2AF65 RS domain and SF1-BPS. Binding of U2AF65 RRM1 and RRM2 to the polypyrimidine tract directed the RS domain (red) to contact the branch point and SF1^{KH-QUA2} domain. U2AF65 RS domain plays a direct role in modulating spliceosomal RNA-RNA interactions. Intermolecular PREs were recorded with a spin-label at 137 and at 166 (blue rhombus) of SF1. The RS-SF1-BPS complex is indicated in a circle.

It has been reported that binding of U2AF65 RRM1 and RRM2 to the polypyrimidine tract directed the RS domain to contact the branch point and promoted U2 snRNA-branch point base pairing even in the absence of other splicing factors. (Valcarcel et al., 1996). As U2AF65 RS domain is rich in arginine and thereby is highly positively charged, electrostatic effects mainly drive the binding of RS domain to the BPS. SF1 recognizes BPS as well and may play a role as an adjustor of interaction between RS domain and the BPS. The ternary complex is formed, how is the RS domain positioned in the complex still requires structural analysis. To provide an insight into conformational information of the interaction, intermolecular PREs were employed to investigate RS-SF1-BPS complex with a molecular weight of 50 kDa. Paramagnetic relaxation enhancement was measured for the isotope labelled RS domain in the complex with an iodoacetamide proxyl spin label attached to single cysteine variants S137C and N166C in SF1 and BPS. Position 137 and 166 are the N-terminal and C-terminal boundary of SF1 KH domain respectively and thereby information acquired from these two positions is able to cover the entire complex.

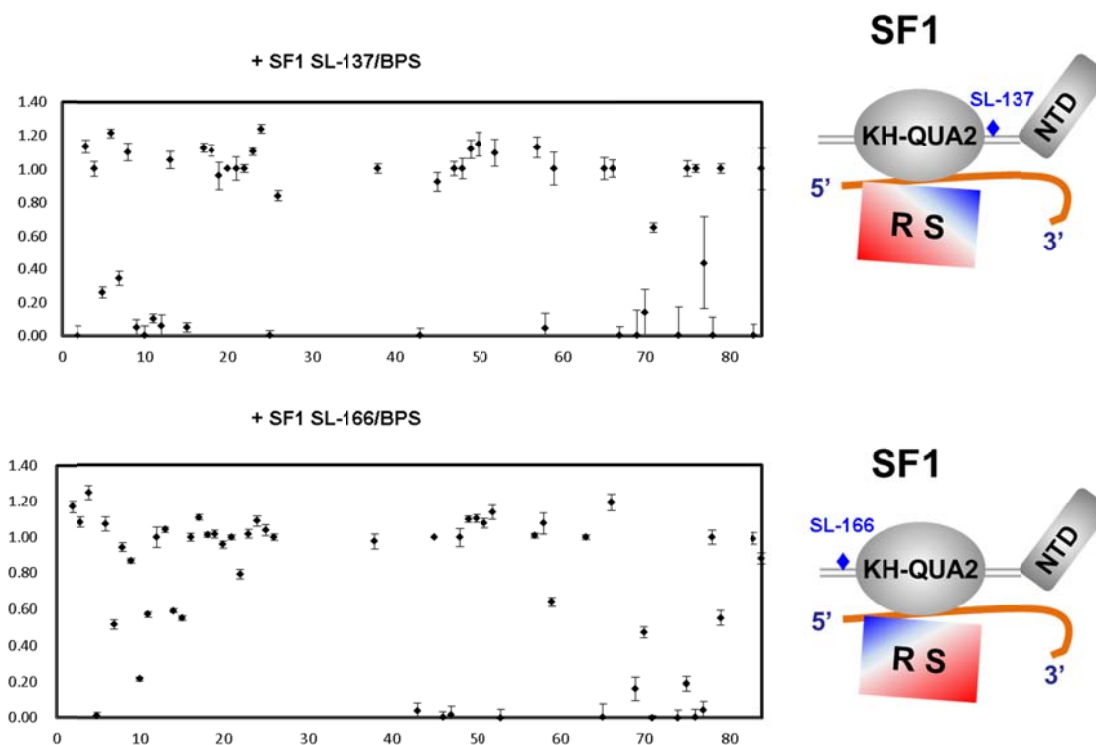


Figure 3.37 Measurement of intermolecular paramagnetic relaxation enhancement of U2AF65 RS domain in complex with the spin label attached at position 137 (upper panel) and 166 (lower panel) of BPS-bound SF1. I_{ox}/I_{red} ratios of amide protons of U2AF65 RS domain in the complex with a spin-label at residue 137 and 166 of BPS-bound SF1 plotted against the amino acid sequence. Isotope labeled RS domain (red) are in complex with SF1/RNA with the spin label (blue rhombus) at position 137 and 166, respectively. The residues of RS in close proximity to the spin label can have PREs, which are shown in blue. Schematic representations are indicated correspondingly on the right.

For the U2AF65 RS domain in the complex with a spin-label at residue 137 and 166 of BPS-bound SF1 label, the I_{ox}/I_{red} ratio is strongly reduced for amide protons between residue 8 and 11, 41 and 49 and additionally from C-terminal residue 68 to 78. Due to signal overlap in ^1H , ^{15}N HSQC spectra quantitative evaluation was limited to approximately 70% of the peaks, the analysis of the region from residue 28 to 39 requires further updates and therefore we are not able to investigate the interactions in this region. Both 137 and 166 spin-labelled SF1/BPS lower the I_{ox}/I_{red} ratio of U2AF65 RS domain and present similar patterns, demonstrating the interactions between U2AF65 RS domain and SF1/BPS. Reductions of the I_{ox}/I_{red} ratio were observed on C- and N-terminus of RS domain, indicating SF1/BPS contact the entire RS domain. The information of central region of RS domain is not completely gained due to the signal overlap of arginine-serine motifs, low I_{ox}/I_{red} ratio of the

potential structure region from residue 45 to 50 suggests this less flexible α -helix-like region is involved in the RS domain and SF1/BPS interactions.

3.3.4 Discussion: RS domain contacts SF1 and RNA at the branch point

Previous studies have shown that the RS domain present in splicing factors of SR family (Zahler et al., 1992) mediate protein-protein interactions (Kohtz et al., 1994; Wu and Maniatis, 1993). Whereas, it is reported the target of the U2AF65 RS domain is the pre-mRNA branch point, and not a protein (Valcarcel et al., 1996). Due to intrinsically unstructured and highly dynamic feature of U2AF65 RS domains, structural insight into molecular mechanisms of the interactions remains elusive. Because U2AF65 RS domain is rich in positively charged residues (arginine), resulting in intensively binding to negatively charged nucleic acid. Thereby, purification of RS protein, removal of nucleic acids contamination, could prove a challenge. In this study, RS protein containing an N-terminal Histidine-tag bound to Ni-NTA column and was washed with high urea concentration and high salt concentration buffer. High concentration of urea is able to unfold the protein to expose the nucleic acids and salt plays role as an eliminator of electrostatic effects between RS protein and nucleic acids. The RS domain contacts the branch point and promoted U2 snRNA branch point base pairing, which might be explained by highly positively charged property of RS domain. The binding of RS domain can reduce the negative charges of the branch point, might further facilitate RNA-RNA interactions by indirectly diminish the electrostatic repulsion effects. In addition, SF1^{KH-QUA2} domain has been reported recognize branch point (Liu et al., 2001),

Different from conventional view of intrinsically unstructured and highly dynamic feature, in this study we propose that RS domain tends to adopt structural elements. The N-terminal region with residues from 5 to 25 and the central region with residues from 45 to 60 exhibit typical pattern for α -helix secondary structure and show less flexibility and less dynamics. Intermolecular paramagnetic relaxation enhancements demonstrate that the α -helix-like region might mediate the RS domain-SF1^{KH-QUA2} interactions, in good agreement with our hypothesis of formation of RS-SF1-BPS complex.

In vivo, RS domains are found to be extensively phosphorylated on serine residues enabling post-translational modification of the interaction network. It is believed that the subcellular localization as well as the activity is regulated on the level of phosphorylation (Lin and Fu, 2007). Inclusion of phosphorylation in the structural studies is crucial in order to understand how the binding-strength and -specificity of RS domains for their interaction partners is

modulated. The impact of the phosphorylation patterns will reveal how the dynamic interaction network is modulated on a molecular level. Study of *in vivo* phosphorylation of U2AF65 RS domain will further describes its functional roles in regulation of the formation of SF1-U2AF65-mRNA complex. However, limited amount of RS protein was purified from our protocol, which to some extent increases the difficulty in preparation of phosphorylated RS protein in NMR-scale amount. There is no denying that structural analysis of phosphorylated U2AF65 RS domain will brighten the insight into molecular mechanisms of regulation of pre-mRNA splicing.

4 DISCUSSION

4.1 A model for the role of phosphorylation and U2AF65-binding of SF1 during 3' splice site recognition

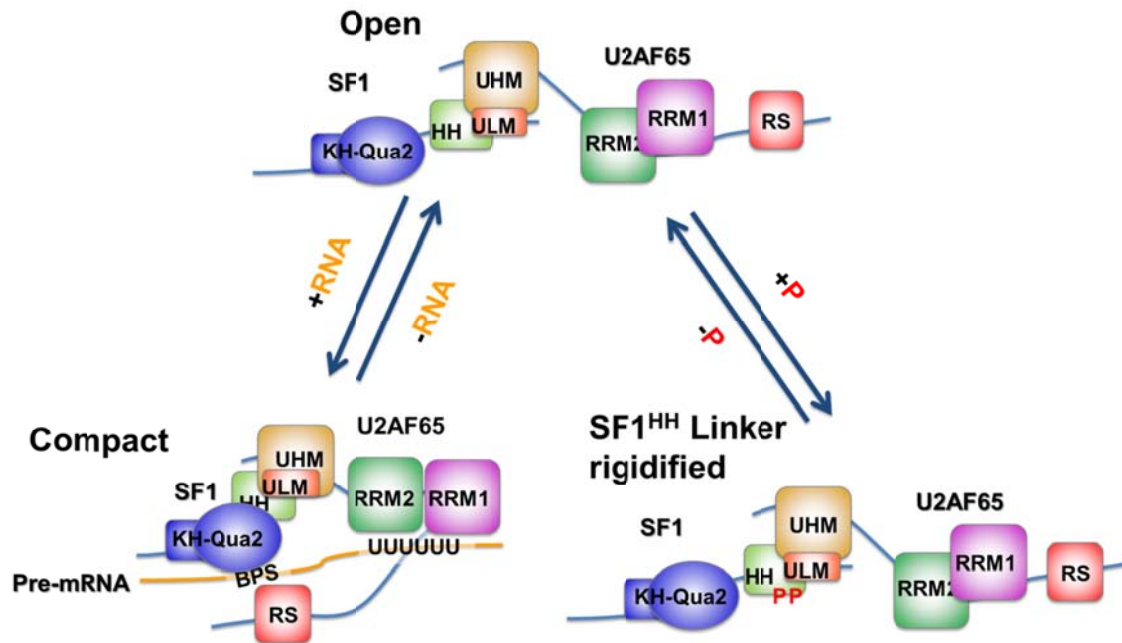


Figure 4.1 Model for the role of phosphorylation and U2AF65 binding of SF1 during 3' splice site recognition. Binding of a 20-mer RNA containing the BPS and the polypyrimidine (Py) tract regions induces large overall changes for the structure and/or dynamics of the SF1-U2AF65^{RRM123} complex, and RS domain is directed to contact the branch point and SF1^{KH-QUA2}. These changes result in formation of a compact SF1-U2AF65^{RRM123}-RNA arrangement. Tandem serine phosphorylation can structures and rigidifies the SF1^{HH} linker, which can locally lock the relative position between the two α -helices within SF1^{HH}. However, our data support that phosphorylation doesn't alter the SF1-U2AF65 binding affinities and interactions.

Experimental data obtained in this study together with previous findings can be combined to a model for the role of phosphorylation and U2AF65 binding of SF1 during 3' splice site recognition. For the U2AF65-SF1 recognition, RRM3-ULM (U2AF Homology Motif-UHM Ligand Motif) interactions form the primary interface (Selenko et al., 2003), SF1^{HH} domain additionally provides a secondary interface and mainly acts as spacer between U2AF65^{RRM3}-bound SF1^{ULM} and the SF1^{KHQUA2} region and may thus provide an optimal pre-arranged protein-protein scaffold for the further RNA binding.

Our study has shown that tandem serine phosphorylation of the conserved SPSP motif rigidifies the linker connecting SF1^{HH} helices $\alpha 1$ and $\alpha 2$, and thereby lock the protein-protein scaffold, however our data also suggest that it has little effect on the conformation and overall structure of SF1 alone and the SF1-U2AF65^{RRM123} complex. The crystal structure of pSF1^{NTD}-U2AF65^{UHM} proved that the SF1^{HH} mediates an interface with U2AF65 and phosphorylation induces an arginine claw surrounding the phosphorylated SPSP motif (Wang et al., 2013). Their findings show a good agreement with the results in this thesis. However, although local structural changes linked to phosphorylation of SF1 are shown in their paper and in this thesis, we hardly observed global conformational changes (C-shape formation) in the pSF1-U2AF65-RNA complex, which is claimed in their paper. As they only employed SAXS data to generate the C-shape structure, the limit of resolution on atomic level leads to this controversial conclusion. To clarify the conformational arrangements of pSF1-U2AF65-RNA complex, more experiments and high-resolution techniques are required. Nevertheless, the EMSA data indicate that SF1 phosphorylation enhances cooperative assembly of the SF1-U2AF65-RNA complex (Figure 3.31); although there are no large conformational rearrangements, our SAXS data to some extent corroborate these results. RNA recognition by the tandem RNA recognition motifs (RRM1-RRM2) of U2AF65 involves equilibrium between two distinct domain arrangements (Mackereth et al., 2011). Without binding of RNA to U2AF65 RRM1-2 keeps the conformational equilibrium in close state. The equilibrium is shifted from the closed state towards the open state in the presence of RNA. From unpublished data in our group, the relative domain arrangements in the minimal U2AF heterodimer are modulated by RNA binding and the conformational equilibrium in U2AF65 is affected by the presence of the U2AF35 RRM.

Binding of U2AF65^{RRM1-2} to the Py tract directed the RS domain to contact the branch point and SF1^{KH-QUA2}. The changes of electrostatic effects may interpret the promotions of U2 snRNA branch point base pairing upon on RS binding. Although RS domain is intrinsically unstructured and highly dynamic feature, in this study we propose that it tends to have potential structural elements. The NMR relaxation results and intermolecular PRE reveal that these potential structure regions are involved in the RS-SF1-BPS interactions. *In vivo*, RS domains are able to be extensively phosphorylated on serine residues and enabling post-translational modification of the interaction network. Therefore, structural studies of

the phosphorylated U2AF65 RS domain will give a detail insight into the molecular mechanisms of regulation of pre-mRNA splicing.

Similar to a recent report (Gupta et al., 2011), SAXS results demonstrate binding of RNA induces large overall changes for the structure of the SF1-U2AF65^{RRM123} complex and leads to formation of a compact SF1-U2AF65^{RRM123}-RNA arrangement. The conformational information of SF1-U2AF65^{RRM123}-RNA will describe the mechanism of 3' splice site recognition.

4.2 Conclusions

This study has addressed open questions on the molecular details of 3' splice site recognition in the complex E with a focus on the structure of SF1^{NTD}, the role of SF1 phosphorylation and the arrangements of SF1-U2AF65-RNA complex. NMR studies of SF1^{NTD} have revealed a hairpin scaffold structure, which locks the relative orientation of U2AF65^{UHM} and SF1^{NTD} and provides the proper geometry of SF1 and U2AF65 for RNA binding. The structural biology of SF1 phosphorylation have elucidated that although the tandem serine phosphorylation of SF1 on SPSP motif rigidifies the linker connecting SF1^{HH} α 1 and α 2, it has little effects on the overall conformational arrangement of the SF1-U2AF65-RNA complex. However, EMSA results imply that phosphorylated SF1 plays roles in the formation of SF1-U2AF65-RNA complex. It has been assumed that interactions of phosphorylated SF1 with so far unknown factors may contribute to regulate alternative splicing.

Binding of U2AF65 RRM1-2 to Py trace drives the RS domain to contact BPS and SF1^{KH-QUA2}. These interactions may be mediated by some potential structure regions and promotes of U2 snRNA branch point base pairing. A compact SF1-U2AF65^{RRM123}-RNA is induced in the presence of 20mer RNA containing of BPS and Py trace.

In this thesis, the NMR structure of SF1^{NTD} was determined, revealing that the SF1^{HH} might act as a spacer defining the relative orientation of the subunits in the ternary complex. These data contribute to studies of the detailed structural information of SF1-U2AF65-RNA. In addition, investigations of post-translational modifications, such as phosphorylation, of splicing factors exhibit significances in understanding of regulation of pre-mRNA alternative splicing. Furthermore, NMR studies of the U2AF65 RS domain show that it may adopt secondary structure elements and may interact with the protein-RNA complex. All together, these findings are a major step forward in understanding the molecular mechanism of 3' splicing site recognition and deciphering the splicing code that regulates alternative pre-mRNA splicing.

5 LITERATURE REFERENCES

Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* *89*, 403-412.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* *25*, 3389-3402.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* *38*, W529-533.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* *98*, 10037-10041.

Battiste, J.L., and 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*, 5355-5365.

Beausoleil, S.A., Jedrychowski, M., Schwartz, D., Elias, J.E., Villen, J., Li, J., Cohn, M.A., Cantley, L.C., and Gygi, S.P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* *101*, 12130-12135.

Bedford, M.T., Reed, R., and Leder, P. (1998). WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. *Proc Natl Acad Sci USA* *95*, 10602-10607.

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., and Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics* *20*, 1322-1324.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* *270*, 2411-2414.

Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Gene Dev* *12*, 858-867.

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

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* *16*, 2384-2403.

- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **72**, 291-336.
- Boudrez, A., Beullens, M., Waelkens, E., Stalmans, W., and Bollen, M. (2002). Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPP1. *J Biol Chem* **277**, 31834-31841.
- Caceres, J.F., Misteli, T., Sreaton, G.R., Spector, D.L., and Krainer, A.R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* **138**, 225-238.
- Cass, D.M., and Berglund, J.A. (2006). The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry* **45**, 10092-10101.
- Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature reviews Molecular cell biology* **10**, 741-754.
- Chiara, M.D., Gozani, O., Bennett, M., Champion-Arnaud, P., Palandjian, L., and Reed, R. (1996). Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly. *Mol Cell Biol* **16**, 3317-3326.
- Clark, F., and Thanaraj, T.A. (2002). Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Human molecular genetics* **11**, 451-464.
- Collins, C.A., and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nature structural biology* **7**, 850-854.
- Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* **136**, 777-793.
- Corioni, M., Antih, N., Tanackovic, G., Zavolan, M., and Kramer, A. (2011). Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res* **39**, 1868-1879.
- Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J., and Sattler, M. (2007). U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat Struct Mol Biol* **14**, 620-629.
- Daragan, V.A., and Mayo, K.H. (1997). Motional model analyses of protein and peptide dynamics using ¹³C and ¹⁵N NMR relaxation. *Prog NMR Spectrosc* **31**, 63-105.
- Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* **5**, 779-787.
- Day, I.S., Golovkin, M., Palusa, S.G., Link, A., Ali, G.S., Thomas, J., Richardson, D.N., and Reddy, A.S. (2012). Interactions of SR45, an SR-like protein, with spliceosomal proteins and an intronic sequence: insights into regulated splicing. *The Plant journal : for cell and molecular biology* **71**, 936-947.

- Del Gatto-Konczak, F., Bourgeois, C.F., Le Guiner, C., Kister, L., Gesnel, M.C., Stevenin, J., and Breathnach, R. (2000). The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Mol Cell Biol* 20, 6287-6299.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6, 277-293.
- Dietrich, R.C., Incurvaia, R., and Padgett, R.A. (1997). Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol Cell* 1, 151-160.
- Doreleijers, J.F., Vranken, W.F., Schulte, C., Markley, J.L., Ulrich, E.L., Vriend, G., and Vuister, G.W. (2012). NRG-CING: integrated validation reports of remediated experimental biomolecular NMR data and coordinates in wwPDB. *Nucleic Acids Res* 40, D519-524.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Formankay, J.D., and Kay, L.E. (1994). Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology-2 Domain Studied by N-15 Nmr Relaxation. *Biochemistry* 33, 5984-6003.
- Forch, P., Merendino, L., Martinez, C., and Valcarcel, J. (2001). Modulation of msl-2 5' splice site recognition by Sex-lethal. *RNA* 7, 1185-1191.
- Forch, P., Merendino, L., Martinez, C., and Valcarcel, J. (2003). U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5' splice sites. *The Biochemical journal* 372, 235-240.
- Forch, P., Puig, O., Kedersha, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P., and Valcarcel, J. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol Cell* 6, 1089-1098.
- Galy, V., Gadai, O., Fromont-Racine, M., Romano, A., Jacquier, A., and Nehrbass, U. (2004). Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* 116, 63-73.
- Garrey, S.M., Voelker, R., and Berglund, J.A. (2006). An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *J Biol Chem* 281, 27443-27453.
- Ghosh, G., and Adams, J.A. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J* 278, 587-597.
- Golas, M.M., Sander, B., Will, C.L., Luhrmann, R., and Stark, H. (2003). Molecular architecture of the multiprotein splicing factor SF3b. *Science* 300, 980-984.
- Goldstrohm, A.C., Albrecht, T.R., Sune, C., Bedford, M.T., and Garcia-Blanco, M.A. (2001). The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol* 21, 7617-7628.

- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* *18*, 4752-4760.
- Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA-Publ RNA Soc* *6*, 1197-1211.
- Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends in genetics : TIG* *17*, 100-107.
- Guntert, P. (2009). Automated structure determination from NMR spectra. *Eur Biophys J Biophys* *38*, 129-143.
- Gupta, A., Jenkins, J.L., and Kielkopf, C.L. (2011). RNA induces conformational changes in the SF1/U2AF65 splicing factor complex. *J Mol Biol* *405*, 1128-1138.
- Guth, S., and Valcarcel, J. (2000). Kinetic role for mammalian SF1/BBP in spliceosome assembly and function after polypyrimidine tract recognition by U2AF. *J Biol Chem* *275*, 38059-38066.
- Hansen, M.R., Mueller, L., and Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* *5*, 1065-1074.
- Hertel, K.J., and Graveley, B.R. (2005). RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends in biochemical sciences* *30*, 115-118.
- Hong, W., Bennett, M., Xiao, Y., Feld Kramer, R., Wang, C., and Reed, R. (1997). Association of U2 snRNP with the spliceosomal complex E. *Nucleic Acids Res* *25*, 354-361.
- Hore, P.J., Jones, J.A., and Wimperis, S. (2001). *NMR: The Toolkit* (Oxford University Press, USA).
- Ingham, R.J., Colwill, K., Howard, C., Dettwiler, S., Lim, C.S., Yu, J., Hersi, K., Raaijmakers, J., Gish, G., Mbamalu, G., *et al.* (2005). WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol* *25*, 7092-7106.
- Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* *28*, 8972-8979.
- Keeler, J. (2005). *Understanding NMR Spectroscopy*. (John Wiley & Sons).
- Kent, O.A., and MacMillan, A.M. (2002). Early organization of pre-mRNA during spliceosome assembly. *Nature structural biology* *9*, 576-581.
- Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* *18*, 1513-1526.

- Kielkopf, C.L., Rodionova, N.A., Green, M.R., and Burley, S.K. (2001). A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 106, 595-605.
- Kim, N., Alekseyenko, A.V., Roy, M., and Lee, C. (2007). The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res* 35, D93-98.
- Klippel, S., Wieczorek, M., Schumann, M., Krause, E., Marg, B., Seidel, T., Meyer, T., Knapp, E.W., and Freund, C. (2011). Multivalent binding of formin-binding protein 21 (FBP21)-tandem-WW domains fosters protein recognition in the pre-spliceosome. *J Biol Chem* 286, 38478-38487.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garciblanco, M.A., and Manley, J.L. (1994). Protein-Protein Interactions and 5'-Splice-Site Recognition in Mammalian Messenger-Rna Precursors. *Nature* 368, 119-124.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36, 1277-1282.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Lin, S.R., and Fu, X.D. (2007). SR proteins and related factors in alternative splicing. In *Alternative Splicing in the Postgenomic Era* (Berlin: Springer-Verlag Berlin), pp. 107-122.
- Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M., and Nilges, M. (2003). Refinement of protein structures in explicit solvent. *Proteins* 50, 496-506.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Krämer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294, 1098-1102.
- Long, J.C., and Cáceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* 417, 15-27.
- Lundberg, K.S., Shoemaker, D.D., Adams, M.W.W., Short, J.M., Sorge, J.A., and Mathur, E.J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 108, 1-6.

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

Madl, T., Bermel, W., and Zangger, K. (2009). Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy. *Angew Chem Int Ed Engl* 48, 8259-8262.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298, 2205-2208.

Manceau, V., Kielkopf, C.L., Sobel, A., and Maucuer, A. (2008). Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. *J Mol Biol* 381, 748-762.

Manceau, V., Swenson, M., Le Caer, J.P., Sobel, A., Kielkopf, C.L., and Maucuer, A. (2006). Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. *FEBS J* 273, 577-587.

Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. *Nature reviews Molecular cell biology* 6, 386-398.

McPheeters, D.S., and Muhlenkamp, P. (2003). Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast. *Mol Cell Biol* 23, 4174-4186.

Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nature genetics* 30, 13-19.

Myung, J.K., and Sadar, M.D. (2012). Large scale phosphoproteome analysis of LNCaP human prostate cancer cells. *Molecular bioSystems* 8, 2174-2182.

Nilges, M., Macias, M.J., O'Donoghue, S.I., and Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* 269, 408-422.

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

Nolen, B., Yun, C.Y., Wong, C.F., McCammon, J.A., Fu, X.D., and Ghosh, G. (2001). The structure of Sky1p reveals a novel mechanism for constitutive activity. *Nature structural biology* 8, 176-183.

Olsen, J.V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.

Otting, G. (2008). Prospects for lanthanides in structural biology by NMR. *J Biomol NMR* 42, 1-9.

- Otting, G., and Wuthrich, K. (1990). Heteronuclear filters in two-dimensional $[^1\text{H},^1\text{H}]$ -NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q Rev Biophys* 23, 39-96.
- Paleologou, K.E., Schmid, A.W., Rospigliosi, C.C., Kim, H.Y., Lamberto, G.R., Fredenburg, R.A., Lansbury, P.T., Jr., Fernandez, C.O., Eliezer, D., Zweckstetter, M., *et al.* (2008). Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of alpha-synuclein. *J Biol Chem* 283, 16895-16905.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* 40, 1413-1415.
- Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89, 1237-1250.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458, 475-480.
- Ptacek, J., and Snyder, M. (2006). Charging it up: global analysis of protein phosphorylation. *Trends in genetics : TIG* 22, 545-554.
- Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P., and Krämer, A. (1998). Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* 4, 551-565.
- Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Current opinion in cell biology* 12, 340-345.
- Roest Crolius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., *et al.* (2000). Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature genetics* 25, 235-238.
- Rutz, B., and Seraphin, B. (1999). Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* 5, 819-831.
- Rutz, B., and Séraphin, B. (2000). A dual role for BBP/ScSF1 in nuclear pre-mRNA retention and splicing. *EMBO J* 19, 1873-1886.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* 34, 93-158.
- Schmucki, R., Yokoyama, S., and Guntert, P. (2009). Automated assignment of NMR chemical shifts using peak-particle dynamics simulation with the DYNASSIGN algorithm. *J Biomol NMR* 43, 97-109.

Seghezzi, W., Chua, K., Shanahan, F., Gozani, O., Reed, R., and Lees, E. (1998). Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. *Mol Cell Biol* *18*, 4526-4536.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural Basis for the Molecular Recognition between Human Splicing Factors U2AF(65) and SF1/mBBP. *Mol Cell* *11*, 965-976.

Sharp, P.A., and Burge, C.B. (1997). Classification of introns: U2-type or U12-type. *Cell* *91*, 875-879.

Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* *13*, 367-376.

Shen, H.H., and Green, M.R. (2004). A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell* *16*, 363-373.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular Nmr* *44*, 213-223.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. *Mol Cell* *23*, 819-829.

Shu, H., Chen, S., Bi, Q., Mumby, M., and Brekken, D.L. (2004). Identification of phosphoproteins and their phosphorylation sites in the WEHI-231 B lymphoma cell line. *Molecular & cellular proteomics : MCP* *3*, 279-286.

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

Simon, B., Madl, T., Mackereth, C.D., Nilges, M., and Sattler, M. (2010). An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew Chem Int Ed Engl* *49*, 1967-1970.

Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in biochemical sciences* *25*, 381-388.

Solomon, I. (1955). Relaxation Processes in a System of Two Spins. *Physical Review* *99*, 559.

Solomon, I., and Bloembergen, N. (1956). Nuclear Magnetic Interactions in the HF Molecule *J Chem Phys* *25*, 261-266.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Luhrmann, R., and Sattler, M. (2006). Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA* *12*, 410-425.

- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315-326.
- Stamm, S. (2008). Regulation of alternative splicing by reversible protein phosphorylation. *J Biol Chem* 283, 1223-1227.
- Stark, H., Dube, P., Luhrmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539-542.
- Su, X.C., and Otting, G. (2010). Paramagnetic labelling of proteins and oligonucleotides for NMR. *J Biomol NMR* 46, 101-112.
- Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRY SOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28, 768-773.
- Svergun, D.I. (1992). Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* 25, 495-503.
- Tanackovic, G., and Kramer, A. (2005). Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol Biol Cell* 16, 1366-1377.
- Tarn, W.Y., and Steitz, J.A. (1997). Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends in biochemical sciences* 22, 132-137.
- Tugarinov, V., Kanelis, V., and Kay, L.E. (2006). Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1, 749-754.
- Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* 413, 701-707.
- Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* 273, 1706-1709.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The sequence of the human genome. *Science* 291, 1304-1351.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.
- Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev* 12, 1409-1414.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.

- Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K.R., Bauer, W.J., Kennedy, S.D., Wedekind, J.E., Green, M.R., and Kielkopf, C.L. (2013). Structure of Phosphorylated SF1 Bound to U2AF(65) in an Essential Splicing Factor Complex. *Structure* 21, 197-208.
- Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P.J., Krämer, A., and Robinson, P.J. (1999). Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J* 18, 4549-4559.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.
- Weber, G., Trowitzsch, S., Kastner, B., Luhrmann, R., and Wahl, M.C. (2010). Functional organization of the Sm core in the crystal structure of human U1 snRNP. *The EMBO journal* 29, 4172-4184.
- Weiner, L.M. (1986). Magnetic resonance study of the structure and functions of cytochrome P450. *CRC critical reviews in biochemistry* 20, 139-200.
- Wiesner, S., Stier, G., Sattler, M., and Macias, M.J. (2002). Solution structure and ligand recognition of the WW domain pair of the yeast splicing factor Prp40. *J Mol Biol* 324, 807-822.
- Will, C.L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current opinion in cell biology* 13, 290-301.
- Will, C.L., Schneider, C., Reed, R., and Luhrmann, R. (1999). Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284, 2003-2005.
- Wishart, D.S., and Sykes, B.D. (1994). The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* 4, 171-180.
- Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75, 1061-1070.
- Xu, Y.Z., Newnham, C.M., Kameoka, S., Huang, T., Konarska, M.M., and Query, C.C. (2004). Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *The EMBO journal* 23, 376-385.
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., *et al.* (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64-69.
- Zahler, A.M., Lane, W.S., Stolk, J.A., and Roth, M.B. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* 6, 837-847.
- Zhang, Y., Madl, T., Bagdiul, I., Kern, T., Kang, H.S., Zou, P., Mausbacher, N., Sieber, S.A., Kramer, A., and Sattler, M. (2013). Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3'-splice site recognition. *Nucleic Acids Res* 41, 1343-1354.

6 LIST OF FIGURES

Figure 1.1	Elementary alternative splicing patterns and regulatory elements (Matlin et al., 2005).....	2
Figure 1.2	Schematic representation of the spliceosome assembly (adapted from P. Selenko & M. Sattler).	4
Figure 1.3	Arrangement of proteins and pre-mRNA in complex E (adapted from P. Selenko & M. Sattler).	7
Figure 1.4	Schematic representation of the structural arrangement of protein and RNA components of the U1 snRNP at the 5' splice site	8
Figure 1.5	Schematic overview and topology of molecular interactions that define the 3' splice site.	10
Figure 1.6	Two energy levels for nuclei with $I = 1/2$ in an external homogeneous magnetic field.....	18
Figure 1.7	The vector model of NMR.....	20
Figure 1.8	Protein NMR spectra.	22
Figure 1.9	Paramagnetic relaxation enhancement.	25
Figure 3.1	Proteins and domains involved in 3' splice site recognition..	49
Figure 3.2	Multiple sequence alignment of SF1 N-terminal domains (SF1 ^{NTD}).....	50
Figure 3.3	Diagram of the protein constructs used in our experiments.	51
Figure 3.4	Overlay of the ¹ H ¹⁵ N HSQC spectra of SF1 ^{NTD} (1-145; black) with SF1 ^{HH} (27-145, blue).....	52
Figure 3.5	Secondary chemical shift and dynamic analysis of SF1 ^{HH} plotted against the amino acid sequence.....	53
Figure 3.6	Comparison of back-calculated and experimental solvent PREs.....	55
Figure 3.7	Structure of a novel helix hairpin in the N-terminal domain of splicing factor 1.....	57
Figure 3.8	Isothermal titration calorimetry data.....	58
Figure 3.9	NMR analysis of the SF1 ^{NTD} -U2AF65 ^{UHM} interaction.....	59
Figure 3.10	NMR relaxation data for the SF1 ^{NTD} - U2AF65 ^{UHM} complex.....	60
Figure 3.11	Analysis of the SF1 ^{NTD} -U2AF65 ^{UHM} complex.....	61
Figure 3.12	Structure of the SF1 ^{NTD} and U2AF65 ^{UHM} complex.....	63
Figure 3.13	NMR analysis of SF1 (SF1 ^{NTD}) and RNA interactions.	65
Figure 3.14	Mapping of CPSs induced by RNA binding on the NMR solution structure of SF1 ^{NTD} (left) and SF1 (right).....	66

Figure 3.15	EMSA assay of Cooperative binding of U2AF65 ^{RRM123} and SF1 to a 3' splice site RNA.....	67
Figure 3.16	SAXS analysis of cooperative binding of SF1-U2AF65 to the pre-mRNA.....	68
Figure 3.17	Schematic model of the SF1/U2AF65/3' splice site of pre-mRNA complex.....	69
Figure 3.18	Mimicking phosphorylation at Ser80 and Ser82 of SF1.....	71
Figure 3.19	Superposition of the ¹ H ¹⁵ N HSQC spectra of SF1 wildtype (black) and SF1 mutant (S80E&S82E; green) and mapping of the CSP on the structure.....	72
Figure 3.20	Overlay of the ¹ H ¹⁵ N HSQC spectra of SF1 mutants (S80E; blue and S82E; red) with SF1 wildtype (black).....	73
Figure 3.21	<i>in vitro</i> phosphorylation of SF1 ^{NTD} and SF1 with recombinant UHM-KIS kinase.....	74
Figure 3.22	Mass spectrometry analysis of SF1 phosphorylation.....	75
Figure 3.23	Analysis of thermodynamics of phosphorylated SF1 by Circular Dichroism (CD) spectroscopy.....	76
Figure 3.24	Comparison of the ¹ H ¹⁵ N HSQC spectra of SF1 ^{NTD} and SF1 with their phosphorylated states.....	77
Figure 3.25	NMR and dynamic analysis of the effect of tandem serine phosphorylation of SF1.....	78
Figure 3.26	Overlay of the ¹ H ¹⁵ N HSQC spectra of phosphorylated SF1 with a spin label at position 166 in the oxidized (red) and reduced state (black).....	79
Figure 3.27	Paramagnetic relaxation enhancement of SF1 with and without phosphorylation with a spin-label at residue 20, 137 and 166.....	80
Figure 3.28	Impact of phosphorylation on SF1 structure.....	81
Figure 3.29	SAXS analysis of the effect of tandem serine phosphorylation of SF1.....	82
Figure 3.30	ITC data of (pSF1) SF1 in complex with U2AF65.....	84
Figure 3.31	Cooperative binding of U2AF65 ^{RRM123} and (p)SF1 to a 3' splice site RNA.....	85
Figure 3.32	Effect of SF1 ^{NTD} phosphorylation on U2AF65 ^{UHM}	86
Figure 3.33	Purification of U2AF65 RS protein.....	88
Figure 3.34	¹ H, ¹⁵ N HSQC spectrum of U2AF65 RS.....	90
Figure 3.35	Secondary chemical shift analysis and ¹⁵ N relaxation studies of U2AF65 RS domain plotted against the amino acid sequence.....	91
Figure 3.36	Interactions between U2AF65 RS domain and SF1-BPS.....	93
Figure 3.37	Measurement of intermolecular paramagnetic relaxation enhancement of U2AF65 RS domain in complex with the spin label attached at position 137 (upper panel) and 166 (lower panel) of BPS-bound SF1. I.....	94

III. APPENDIX

III.I DNA sequences

H. sapiens Splicing Factor 1 (SF1). Accession number NCBI NM_201998.2

```

1 gagctcgccg tcgctccgctc atagagtteg cccaccacca tcccctcctt tctggactcg
61 gagctcagtt cacgcagtaa caaatgaagt gcgcgctgcy acacctccca gccaccgaa
121 ctccgcccgc atttccctgc ttggcctaac ggttcggcca atcccagcgc gcatcaagaa
181 ggactgaggc tccgccaatc ggaggccgcc gatttcgacc cttcgctctg gcccgccca
241 atccaggccc cggccccgcc gccccggcc cgcgcccgcy gtgcctctc tctcctctc
301 ttgtgcgtct cgcgcccgcg ccgcccgcg cgtgagagga cgggctccgc gcgctccggc
361 agcgcattcg ggtcccctcc cccggggagg cttgcgaagg agaagccgcc gcagaggaaa
421 agcaggtgcc ggtgcctgtc cccggggggc ccatggcgac cggagcgaac gccacgccgt
481 tggacttccc aagtaagaag cggaagagga gccgctggaa ccaagacaca atggaacaga
541 agacagtgat tccaggaatg cctacagtta ttcccctgg acttactcga gaacaagaaa
601 gagcttata atgtgcaactg cagatagaag acctgactcy taaactgcy acaggagacc
661 tgggcatccc ccctaaccct gaggacaggt ccccttcccc tgagccatc tacaatagcy
721 aggggaagcy gcttaacacc cgagagtcc gcaccgcga aaagctggaa gaggagcggc
781 acaacctcat cacagagatg gttgactca atccggattt caagccacct gcagattaca
841 aacctccagc aacacgtgtg agtgataaag tcatgattcc acaagatgag taccagaaa
901 tcaactttgt ggggctgctc atcgggcccc gagggaaacac cctgaagaac atagagaagg
961 agtgcaatgc caagattatg atccggggga aagggtctgt gaaagaaggg aaggttgggc
1021 gcaaagatgg ccagatgttg ccaggagaag atgagccact tcatgcctg gttactgcca
1081 atacaatgga gaacgtcaaa aaggcagtgg aacagataag aaacatccty aagcagggta
1141 tcgagactcc agaggaccag aatgatctac ggaagatgca gcttcgggag ttggtcgct
1201 taaatgggac ccttcgggaa gacgataaca ggatcttaag acctggcag agctcagaga
1261 cccgcagcat taccaacacc acagtgtgta ccaagtgtgg aggggctggc cacattgctt
1321 cagactgtaa attccaaagg cctggtgatc ctcagtcagc tcaggataaa gcacggatgg
1381 ataaagaata tttgtccctc atggctgaac tgggtgaagc acctgtccca gcatctgtgg
1441 gctccacctc tgggctgccc accacacccc tggccagcgc acctcgctct gctgctcccg
1501 ccaacaaccc acctccaccg tctctcatgt ctaccacca gagccgcca cctggatga
1561 attctggccc ttcagagagt cggccctacc acggcatgca tggaggtggc cctggtgggc
1621 ccgaggtggy ccccacagc ttcccacacc cattaccagc cctgacaggt gggcaggtg
1681 gacatcccat gcagcacaac ccaatggac cccaccccc ttggatgcag ccaccaccac
1741 caccgatgaa ccagggcccc caccctcctg ggcaccatgg cctcctcca atggatcagt
1801 acctgggaag taogcctgtg ggctctgggg tctatcgctt gcatcaagga aaaggtatga
1861 tgccgccacc acctatgggc atgatccgc cgcgcccgc gcctcccagt gggcagcccc
1921 caccctcctc ctctggctct cttcccccat ggcaacaaca gcagcagcag cctccgccac
1981 cccctccgcc cagcagcagt atggcttcca gtacccctt gccatggcag caaagatccc
2041 tcccgcggc ggcgatggcc cgagccatga gagtgaggac tttccgcgcc cattggtgac
2101 cttccaggc agacagcctc agcaacgccc ctggtggaca ggatggttcg gcaaagcagc
2161 ctgagttatt tttgtggacg gaatcggaac acgctggctc catatcgtga aatttttatt
2221 aatttttttc ttttccctt gttactctct tatcttttcc tttcttcaga ctccgtccaa
2281 ggagatgctc tcccgggtct tctgctgcaa tttagattcc tttcccttct ctccagttct
2341 ctttccctta ccaaggagag gggagcaaat ggttttgggc aagggtcttg gccattcatg
2401 tcaagctggt tgtgggtttt tcaaggtgcc atagccacc ccaaagatgt ttgtttaaag
2461 cgtggggttt tttaatctct gccacccttg tcaagggagt cttgtaaagt tgccaggggt
2521 aggttcatct ccagggtttcg ggattcccat cgtcctggc gatcctgcca gcagtgggtg
2581 ggcagcctga gctccctcgg gctcgctgc cagcctggag ttcttctctg gctccttgat
2641 cacctgagct gcctcagatt ccatttggct ctctccttcc tgggaaggctt ccttttatgt
2701 tttgttttaa tcccaaatgt ctgaatgttt tgcaagtgtg aggggtttga gcccttgtt
2761 cattctcctt ccttttctct ccgcttccc tctccatgaa gtgattctgt tgacaataat
2821 gtatactgcy cgttctcttc actggtttat ctgcagaaat ttctctgggc tttttcggc
2881 gttagattca aactgcgct aaagcgggga tgttccattg aataaaagag cagtgtggtt
2941 ttctgggtc

```

H. sapiens U2AF65. Accession number GenBank AB527505.1

```

1  gcgatcgcca  tgtcggactt  cgacgagttc  gagcggcagc  tcaacgagaa  taaacaagag
61  cgggacaagg  agaaccggca  tcggaagcgc  agccacagcc  gctctcggag  ccgggaccgc
121 aaacgccgga  gccggagccg  cgaccggcgc  aaccgggacc  agcggagcgc  ctcccgggac
181 aggcgacgac  gcagcaaacc  tttgaccaga  ggcgctaaag  aggagcacgg  tggactgatt
241 cgttcccccc  gccacgagaa  gaagaagaag  gtccgtaaag  actgggacgt  gccaccccca
301 ggctttgagc  acatcaccce  aatgcagtac  aaggccatgc  aagctgcggg  tcagattcca
361 gccactgctc  ttctccccac  catgaccctt  gacggctctg  ctgtgacccc  aacgccggtg
421 cccgtggtcg  ggagccagat  gaccagacaa  gcccggcgcc  tctacgtggg  caacatcccc
481 tttggcatca  ctgaggaggc  catgatggat  ttcttcaacg  cccagatgcg  cctggggggg
541 ctgaccagg  cccctggcaa  cccagtgttg  gctgtgcaga  ttaaccagga  caagaatddd
601 gcctttttgg  agttccgctc  agtggacgag  actaccagg  ctatggcctt  tgatggcatc
661 atcttccagg  gccagtcact  aaagatccgc  aggcctcag  actaccagcc  gcttctggc
721 atgtcagaga  acccctccgt  ctatgtgcct  ggggttgtgt  ccaactgtgg  ccccgactct
781 gccacaagc  tgttcatcgg  gggcttacc  aactacctga  acgatgacca  ggtcaaagag
841 ctgctgacat  cctttgggcc  cctcaaggcc  ttcaacctgg  tcaaggacag  tgccacgggg
901 ctctccaagg  gctacgcctt  ctgtgagtac  gtggacatca  acgtcacgga  tcaggccatt
961 gcggggctga  acggcatgca  gctgggggat  aagaagctgc  tggccagag  ggcgagtgtg
1021 ggagccaaga  atgccacgct  gagcaccatc  aatcagacgc  ctgtgaccct  gcaagtgcg
1081 ggcttgatga  gctcccaggt  gcagatgggc  gggcaccgga  ctgaggtcct  gtgcctcatg
1141 aacatggtgc  tgccctgagga  gctgctggac  gacgaggagt  atgaggagat  cgtggaggat
1201 gtgcgggacg  agtgcagcaa  gtacgggctt  gtcaagtcca  tcgagatccc  ccggcctgtg
1261 gacggcgtcg  aggtgcccgg  ctgcggaaag  atctttgtgg  agttcacctc  tgtgtttgac
1321 tgccagaaag  ccatgcaggg  cctgacgggc  cgcaagtctg  ccaacagagt  ggttgtcaca
1381 aaatactgtg  accccgactc  ttatcaccgc  cgggacttct  gggtttaaac

```

III.II Protein sequences

SF1 (*H. sapiens* SF1₁₋₂₆₀)

```

1  MATGANATPL DFPSKKRKRS RWNQDTMEQK TVIPGMPTVI PPGLTREQER AYIVQLQIED      60
61  LTRKLRGTDL GIPPNPEDRS PSPEPIYNSE GKRLNTRFR TRKKLEEERH NLITEMVALN      120
121 PDFKPPADYK PPATRVSDKV MIPQDEYPEI NFVGLLIGPR GNTLNKIEKE ANAKIMIRGK      180
181 GSVKEGKVGK KDGQMLPGED EPLHALVTAN TMENVKKAQE QIRNILKQGI ETPEDQNDLR      240
241 KMQLRELARL NGTLREDDNR

```

U2AF65 UHM (*H. sapiens* U2AF65₃₆₇₋₄₇₅)

```

361          QGAM GGHPTVLCL MNMVLPEELL DDEEYEEIVE DVRDECSKYG LVKSIEIPRP      420
421 VDGVEVPGCG KIFVEFTSVF DCQKAMQGLT GRKFANRVVV TKYCDPDSYH RRDFW

```

Residues different from wildtype are highlighted red.

III.III Resonance assignments of SF1^{NTD}

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
S2	N	15N	116.177	M27	CB	13C	32.6638	K30	HA	1H	4.33791
D3	HN	1H	8.2865	M27	CG	13C	32.0586	K30	HB1	1H	1.69388
D3	N	15N	121.738	M27	HA	1H	4.38855	K30	HB2	1H	1.76989
F4	N	15N	118.777	M27	HB1	1H	1.97725	K30	HB3	1H	1.787
F7	HN	1H	8.082	M27	HB2	1H	2.0168	K30	HN	1H	8.30607
F7	N	15N	120.207	M27	HB3	1H	2.0595	K30	N	15N	122.918
E8	HN	1H	8.147	M27	HG1	1H	2.50488	K30	QD	1H	1.6235
E8	N	15N	120.562	M27	HG2	1H	2.5585	K30	QE	1H	2.93812
R9	HN	1H	8.112	M27	HG3	1H	2.575	K30	QG	1H	1.38033
R9	N	15N	120.674	M27	HN	1H	8.43861	K30a	C	13C	176.062
Q10	HN	1H	8.17	M27	N	15N	119.38	K30a	CA	13C	62.926
Q10	N	15N	119.831	E28	C	13C	176.312	K30a	CB	13C	34.6625
L11	N	15N	122.075	E28	CA	13C	56.945	T31	?	1H	1.108
N12	HN	1H	8.204	E28	CB	13C	30.2517	T31	C	13C	174.071
N12	N	15N	118.093	E28	HA	1H	4.20133	T31	CA	13C	62.0495
E13	N	15N	120.867	E28	HN	1H	8.31977	T31	CB	13C	69.8278
K15	N	15N	121.568	E28	N	15N	122.259	T31	CG2	13C	21.7775
Q16	HN	1H	8.2675	E28	QB	1H	1.9285	T31	HA	1H	4.24914
Q16	N	15N	119.692	E28	QG	1H	2.193	T31	HB	1H	4.07183
E17	N	15N	121.564	Q29	?	1H	2.1165	T31	HN	1H	8.1286
K20	HN	1H	8.066	Q29	C	13C	175.761	T31	N	15N	116.711
K20	N	15N	120.885	Q29	CA	13C	56.0324	T31	QG2	1H	1.127
E21	HN	1H	8.2365	Q29	CB	13C	29.5154	T31a	C	13C	174.11
E21	N	15N	119.992	Q29	CD	13C	180.466	T31a	CA	13C	62.6463
N22	HN	1H	8.2495	Q29	CG	13C	33.8432	T31a	CB	13C	69.7443
N22	N	15N	118.74	Q29	HA	1H	4.28943	T31a	HN	1H	8.25857
R23	HN	1H	8.1825	Q29	HB1	1H	1.93233	T31a	N	15N	116.697
R23	N	15N	120.633	Q29	HB2	1H	2.03071	V32	C	13C	175.57
H24	HN	1H	8.244	Q29	HB3	1H	2.046	V32	CA	13C	62.2916
H24	N	15N	121.397	Q29	HE21	1H	6.78075	V32	CB	13C	32.8996
G25	C	13C	170.193	Q29	HE22	1H	7.458	V32	HA	1H	4.05867
G25	CA	13C	43.334	Q29	HN	1H	8.31823	V32	HB	1H	1.96583
A26	C	13C	176.813	Q29	N	15N	122.006	V32	HN	1H	8.16959
A26	CA	13C	52.9106	Q29	NE2	15N	112.219	V32	N	15N	124.003
A26	CB	13C	19.387	Q29	QG	1H	2.31183	V32	QG1	1H	0.81
A26	HA	1H	4.2855	K30	C	13C	176.471	V32	QG2	1H	0.8705
A26	QB	1H	1.3535	K30	CA	13C	56.2597	V32a	C	13C	174.383
M27	C	13C	176.298	K30	CB	13C	33.1674	V32a	CA	13C	62.322
M27	CA	13C	55.8304	K30	CE	13C	42.2556	V32a	CB	13C	32.7825

APPENDIX

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
V32a	HA	1H	4.057	P34	HD1	1H	3.653	P37	CB	13C	32.1332
V32a	HB	1H	1.975	P34	HD2	1H	3.85562	P37	CD	13C	50.8104
V32a	HN	1H	8.1122	P34	HD3	1H	3.867	P37	CG	13C	27.4244
V32a	N	15N	124.019	P34	HG1	1H	1.92542	P37	HA	1H	4.47183
V32a	QG1	1H	0.8635	P34	HG2	1H	2.0235	P37	HB1	1H	1.8858
V32a	QG2	1H	0.822	P34a	C	13C	177.1	P37	HB2	1H	2.22791
I33	?	1H	0.847	P34a	CA	13C	63.541	P37	HB3	1H	2.267
I33	C	13C	174.335	P34a	CB	13C	32.142	P37	HD1	1H	3.6626
I33	CA	13C	58.3837	G35	C	13C	173.906	P37	HD2	1H	3.8073
I33	CB	13C	38.6994	G35	CA	13C	45.2428	P37	QG	1H	1.9795
I33	CD1	13C	12.705	G35	HA1	1H	3.8145	P37a	CB	13C	32.053
I33	CG1	13C	26.9874	G35	HA2	1H	3.865	T38	C	13C	174.16
I33	CG2	13C	17.116	G35	HA3	1H	3.921	T38	CA	13C	62.3123
I33	HA	1H	4.42609	G35	HN	1H	8.3759	T38	CB	13C	69.8275
I33	HB	1H	1.8124	G35	N	15N	109.539	T38	CG2	13C	21.7405
I33	HD11	1H	0.804	G35	QA	1H	3.868	T38	HA	1H	4.25129
I33	HG11	1H	1.1215	G35a	C	13C	173.99	T38	HB	1H	4.07933
I33	HG12	1H	1.446	G35a	CA	13C	45.2213	T38	HN	1H	8.16411
I33	HG21	1H	0.885	G35a	HN	1H	8.31033	T38	N	15N	115.427
I33	HN	1H	8.27164	G35a	N	15N	109.337	T38	QG2	1H	1.13817
I33	N	15N	127.335	M36	C	13C	174.235	T38a	HN	1H	7.925
I33	QB	1H	1.815	M36	CA	13C	53.217	T38a	N	15N	115.272
I33	QD	1H	0.801333	M36	CB	13C	32.7813	V39	C	13C	175.178
I33	QG11	1H	1.11537	M36	CG	13C	32.2252	V39	CA	13C	62.1843
I33	QG12	1H	1.44287	M36	HA	1H	4.7756	V39	CB	13C	32.9404
I33	QG2	1H	0.887111	M36	HB1	1H	1.92362	V39	HA	1H	4.06333
I33a	C	13C	174.925	M36	HB2	1H	2.0153	V39	HB	1H	1.94933
I33a	CA	13C	57.811	M36	HB3	1H	2.0365	V39	HN	1H	8.0125
I33a	CB	13C	40.712	M36	HG1	1H	2.50144	V39	N	15N	123.63
I33a	HA	1H	4.275	M36	HG2	1H	2.5646	V39	QG1	1H	0.815333
I33a	HN	1H	7.906	M36	HG3	1H	2.587	V39	QG2	1H	0.86725
I33a	N	15N	124.29	M36	HN	1H	7.98531	I40	?	1H	0.79
P34	C	13C	177.255	M36	N	15N	120.813	I40	C	13C	173.894
P34	CA	13C	63.4796	M36a	C	13C	174.333	I40	CA	13C	57.8371
P34	CB	13C	32.1418	M36a	CA	13C	53.202	I40	CB	13C	38.5886
P34	CD	13C	51.22	M36a	CB	13C	32.814	I40	CD1	13C	12.5322
P34	CG	13C	27.4997	M36a	HN	1H	8.1936	I40	CG1	13C	27.5905
P34	HA	1H	4.3035	M36a	N	15N	120.925	I40	CG2	13C	16.8998
P34	HB1	1H	1.87342	P37	C	13C	176.739	I40	HA	1H	4.17327
P34	HB2	1H	2.24425	P37	CA	13C	63.2205	I40	HB	1H	1.71091

APPENDIX

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
I40	HN	1H	8.2144	L44	CG	13C	26.452	E47	HG1	1H	2.21325
I40	N	15N	127.77	L44	HA	1H	4.5448	E47	HG2	1H	2.397
I40	NH	1H	8.213	L44	HB1	1H	1.52171	E47	HG3	1H	2.431
I40	QD	1H	0.7752	L44	HB2	1H	1.72887	E47	HN	1H	8.73047
I40	QG1	1H	1.1105	L44	HD11	1H	0.6622	E47	N	15N	118.874
I40	QG11	1H	1.10525	L44	HD21	1H	0.5981	R47	HN	1H	8.34
I40	QG12	1H	1.4507	L44	HG	1H	1.53483	R47	N	15N	122.523
I40	QG2	1H	0.8021	L44	HN	1H	8.04027	N48	N	15N	119.421
P41	CA	13C	61.5924	L44	N	15N	120.886	Q48	C	13C	179.278
P41	CB	13C	31.0477	T45	C	13C	175.375	Q48	CA	13C	58.2097
P41	CD	13C	50.886	T45	CA	13C	61.1702	Q48	CB	13C	29.5393
P41	CG	13C	27.3922	T45	CB	13C	71.0048	Q48	CD	13C	179.961
P41	HA	1H	4.4938	T45	CG2	13C	21.527	Q48	CG	13C	34.4484
P41	HB1	1H	1.6619	T45	HA	1H	4.38771	Q48	HA	1H	4.06088
P41	HB2	1H	2.247	T45	HB	1H	4.6694	Q48	HB1	1H	1.958
P41	HD1	1H	2.877	T45	HN	1H	8.40482	Q48	HB2	1H	2.332
P41	HD2	1H	3.4682	T45	N	15N	112.949	Q48	HE21	1H	6.70075
P41	QG	1H	1.7473	T45	QG2	1H	1.2778	Q48	HE22	1H	7.4355
P42	C	13C	177.714	R46	C	13C	179.306	Q48	HN	1H	7.677
P42	CA	13C	62.6523	R46	CA	13C	59.5179	Q48	N	15N	119.393
P42	CB	13C	32.3537	R46	CB	13C	29.7922	Q48	NE2	15N	110.35
P42	CD	13C	50.3206	R46	CD	13C	43.2992	Q48	QG	1H	2.4388
P42	CG	13C	27.3	R46	CG	13C	26.9514	E49	C	13C	177.876
P42	HA	1H	4.25475	R46	HA	1H	4.03617	E49	CA	13C	59.937
P42	HB1	1H	1.8514	R46	HB1	1H	1.784	E49	CB	13C	29.748
P42	HB2	1H	2.1975	R46	HB2	1H	1.859	E49	CG	13C	37.304
P42	HD1	1H	3.55444	R46	HB3	1H	1.874	E49	HA	1H	3.8757
P42	HD2	1H	3.68611	R46	HG1	1H	1.6252	E49	HB1	1H	2.03887
P42	QG	1H	1.9059	R46	HG2	1H	1.70264	E49	HB2	1H	2.09725
G43	C	13C	174.648	R46	HG3	1H	1.71	E49	HG1	1H	2.07212
G43	CA	13C	45.84	R46	HN	1H	8.76815	E49	HG2	1H	2.332
G43	HA1	1H	3.724	R46	N	15N	121.085	E49	HN	1H	8.70437
G43	HA2	1H	3.902	R46	QD	1H	3.20036	E49	N	15N	122.059
G43	HN	1H	8.32042	E47	C	13C	178.93	D50	HN	1H	8.253
G43	N	15N	107.151	E47	CA	13C	60.3644	D50	N	15N	121.414
G43	QA	1H	3.749	E47	CB	13C	29.166	R50	C	13C	177.567
L44	C	13C	178.575	E47	CG	13C	37.3735	R50	CA	13C	58.241
L44	CA	13C	54.4949	E47	HA	1H	4.0053	R50	CB	13C	29.0729
L44	CB	13C	42.5297	E47	HB1	1H	1.95663	R50	CD	13C	43.012
L44	CD1	13C	22.3155	E47	HB2	1H	2.00644	R50	CG	13C	27.199
L44	CD2	13C	25.5225	E47	HB3	1H	2.016	R50	HA	1H	4.08594

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
R50	HB1	1H	1.89975	I53	HA	1H	3.60243	L56	CB	13C	41.8352
R50	HB2	1H	1.99364	I53	HB	1H	2.01833	L56	CD1	13C	23.1857
R50	HB3	1H	2.03	I53	HG11	1H	1.3391	L56	CD2	13C	25.508
R50	HD1	1H	3.00931	I53	HG12	1H	1.7585	L56	CG	13C	26.731
R50	HD2	1H	3.1812	I53	HN	1H	7.81711	L56	HA	1H	3.94479
R50	HD3	1H	3.204	I53	N	15N	119.185	L56	HB1	1H	1.45014
R50	HG1	1H	1.54358	I53	QD	1H	0.8748	L56	HB2	1H	1.77414
R50	HG2	1H	1.62146	I53	QG2	1H	0.8998	L56	HD11	1H	0.73325
R50	HG3	1H	1.756	S53	HN	1H	8.233	L56	HD21	1H	0.693636
R50	HN	1H	7.89452	S53	N	15N	116.34	L56	HG	1H	1.515
R50	N	15N	118.858	A54	HN	1H	8.287	L56	HN	1H	7.54808
A51	C	13C	178.737	A54	N	15N	125.695	L56	N	15N	119.295
A51	CA	13C	55.2138	V54	C	13C	177.99	R56	HN	1H	8.1925
A51	CB	13C	18.183	V54	CA	13C	67.0114	R56	N	15N	122.716
A51	HA	1H	3.8062	V54	CB	13C	31.2341	D57	HN	1H	8.192
A51	HN	1H	7.94984	V54	CG1	13C	23.4197	D57	N	15N	120.663
A51	N	15N	119.19	V54	CG2	13C	23.7173	Q57	C	13C	179.167
A51	QB	1H	1.3912	V54	HA	1H	3.47827	Q57	CA	13C	59.544
Q51	HN	1H	8.271	V54	HB	1H	1.98286	Q57	CB	13C	29.5043
Q51	N	15N	120.751	V54	HN	1H	8.54118	Q57	CD	13C	179.343
R52	HN	1H	8.293	V54	N	15N	121.337	Q57	CG	13C	34.8288
R52	N	15N	121.336	V54	QG1	1H	0.822857	Q57	HA	1H	4.10767
Y52	C	13C	177.363	V54	QG2	1H	0.765857	Q57	HB1	1H	2.013
Y52	CA	13C	61.6448	Q55	C	13C	178.599	Q57	HB2	1H	2.22308
Y52	CB	13C	38.6427	Q55	CA	13C	60.2329	Q57	HB3	1H	2.277
Y52	CD1	13C	132.91	Q55	CB	13C	28.068	Q57	HE21	1H	6.56333
Y52	CE1	13C	118.369	Q55	CD	13C	179.96	Q57	HE22	1H	6.88475
Y52	HA	1H	4.05788	Q55	CG	13C	34.4069	Q57	HG1	1H	2.27712
Y52	HB1	1H	3.08425	Q55	HA	1H	3.7029	Q57	HG2	1H	2.56389
Y52	HB2	1H	3.171	Q55	HE21	1H	6.797	Q57	HG3	1H	2.6
Y52	HB3	1H	3.199	Q55	HE22	1H	7.386	Q57	HN	1H	8.014
Y52	HD1	1H	7.024	Q55	HG1	1H	2.2414	Q57	N	15N	120.101
Y52	HE1	1H	6.754	Q55	HG2	1H	2.4232	Q57	NE2	15N	108.376
Y52	HN	1H	7.81919	Q55	HN	1H	8.15083	I58	C	13C	179.361
Y52	N	15N	118.937	Q55	N	15N	119.794	I58	CA	13C	66.4979
I53	C	13C	178.713	Q55	NE2	15N	110.808	I58	CB	13C	38.1593
I53	CA	13C	64.2101	Q55	QB	1H	2.1025	I58	CD1	13C	15.2483
I53	CB	13C	38.2898	S55	HN	1H	8.1555	I58	CG1	13C	31.3707
I53	CD1	13C	12.8386	S55	N	15N	114.448	I58	CG2	13C	17.4597
I53	CG1	13C	29.0441	L56	C	13C	180.436	I58	HA	1H	3.41113
I53	CG2	13C	17.3308	L56	CA	13C	58.1649	I58	HB	1H	1.97925

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
I58	HG11	1H	0.645333	L61	HA	1H	3.92713	K64	CE	13C	42.17
I58	HG12	1H	1.63	L61	HB1	1H	1.224	K64	CG	13C	26.2318
I58	HG13	1H	1.953	L61	HB2	1H	1.68575	K64	HA	1H	3.84855
I58	HN	1H	8.74852	L61	HB3	1H	1.83	K64	HB1	1H	1.75856
I58	N	15N	120.705	L61	HD11	1H	0.9274	K64	HB2	1H	1.8604
I58	QD	1H	0.502778	L61	HD21	1H	0.8196	K64	HB3	1H	1.872
I58	QD1	1H	0.511	L61	HG	1H	1.9286	K64	HG1	1H	1.4365
I58	QG2	1H	0.8215	L61	HN	1H	8.40473	K64	HG2	1H	1.66017
R58	HN	1H	8.181	L61	N	15N	118.412	K64	HG3	1H	1.681
R58	N	15N	121.59	S62	HN	1H	8.288	K64	HN	1H	8.10765
E59	C	13C	178.897	S62	N	15N	117.317	K64	N	15N	120.589
E59	CA	13C	59.9212	T62	C	13C	176.175	K64	QD	1H	1.5354
E59	CB	13C	29.3718	T62	CA	13C	67.5328	K64	QE	1H	2.90373
E59	CG	13C	36.5214	T62	CB	13C	68.899	L65	C	13C	179.011
E59	HA	1H	4.05285	T62	CG2	13C	21.217	L65	CA	13C	57.3054
E59	HB1	1H	2.0735	T62	HA	1H	3.78037	L65	CB	13C	42.3485
E59	HB2	1H	2.1555	T62	HB	1H	4.4024	L65	CD1	13C	24.0723
E59	HB3	1H	2.1835	T62	HN	1H	8.31864	L65	CD2	13C	24.9288
E59	HG1	1H	2.29013	T62	N	15N	115.548	L65	CG	13C	27.507
E59	HG2	1H	2.45044	T62	QG2	1H	1.2386	L65	HA	1H	3.968
E59	HG3	1H	2.473	K63	HN	1H	8.239	L65	HB1	1H	1.4885
E59	HN	1H	8.26796	K63	N	15N	123.99	L65	HB2	1H	1.8256
E59	N	15N	122.346	R63	C	13C	179.71	L65	HB3	1H	1.906
R59	HN	1H	8.33	R63	CA	13C	60.1872	L65	HD11	1H	0.9624
R59	N	15N	122.437	R63	CB	13C	29.8568	L65	HD21	1H	0.680222
D60	C	13C	178.316	R63	CD	13C	43.32	L65	HG	1H	1.70433
D60	CA	13C	58.0796	R63	CG	13C	27.6531	L65	HN	1H	8.07836
D60	CB	13C	42.1938	R63	HA	1H	3.99383	L65	N	15N	118.578
D60	CG	13C	178.873	R63	HB1	1H	2.0124	R66	C	13C	178.066
D60	HA	1H	4.43712	R63	HB2	1H	2.05642	R66	CA	13C	58.3106
D60	HB1	1H	2.71275	R63	HB3	1H	2.05	R66	CB	13C	30.7364
D60	HB2	1H	2.75775	R63	HG1	1H	1.5084	R66	CD	13C	43.4703
D60	HB3	1H	2.749	R63	HG2	1H	1.72145	R66	CG	13C	27.4369
D60	HN	1H	8.11005	R63	HG3	1H	1.759	R66	HA	1H	4.21555
D60	N	15N	121.214	R63	HN	1H	7.98953	R66	HG1	1H	1.63213
L61	C	13C	178.86	R63	N	15N	121.82	R66	HG2	1H	1.76667
L61	CA	13C	58.3831	R63	QD	1H	3.23073	R66	HG3	1H	1.782
L61	CB	13C	42.5935	K64	C	13C	178.706	R66	HN	1H	8.07721
L61	CD1	13C	24.021	K64	CA	13C	60.5767	R66	N	15N	118.719
L61	CD2	13C	26.2848	K64	CB	13C	33.1501	R66	QB	1H	1.91556
L61	CG	13C	27.5085	K64	CD	13C	29.668	R66	QD	1H	3.13656

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
T66	HN	1H	8.067	L70	HA	1H	4.1395	I72a	QD1	1H	0.763
T66	N	15N	115.229	L70	HB1	1H	1.54825	P73	CA	13C	61.5972
R67	N	15N	123.665	L70	HB2	1H	1.69614	P73	CB	13C	31.1755
T67	C	13C	175.832	L70	HB3	1H	1.715	P73	CD	13C	51.132
T67	CA	13C	62.7298	L70	HD11	1H	0.698556	P73	CG	13C	27.433
T67	CB	13C	70.7298	L70	HD21	1H	0.873222	P73	HA	1H	4.61967
T67	CG2	13C	21.563	L70	HG	1H	1.539	P73	HB1	1H	1.84967
T67	HA	1H	4.31686	L70	HN	1H	7.94538	P73	HB2	1H	2.29667
T67	HB	1H	4.285	L70	N	15N	120.544	P73	HD1	1H	3.59342
T67	HN	1H	7.92641	L70a	C	13C	177.633	P73	HD2	1H	3.89017
T67	N	15N	109.734	L70a	CA	13C	55.316	P73	HG1	1H	1.949
T67	QG2	1H	1.27517	L70a	CB	13C	42.129	P73	HG2	1H	2.01233
T67a	C	13C	175.61	G71	C	13C	173.93	G74	HN	1H	8.464
T67a	CA	13C	62.717	G71	CA	13C	45.4485	G74	N	15N	110.031
T67a	CB	13C	70.566	G71	HA1	1H	3.7775	P74	C	13C	176.421
G68	C	13C	173.963	G71	HA2	1H	3.82375	P74	CA	13C	62.9654
G68	CA	13C	46.0282	G71	HA3	1H	3.865	P74	CB	13C	32.035
G68	HA1	1H	3.69	G71	HN	1H	8.32744	P74	CD	13C	50.5547
G68	HA2	1H	3.85675	G71	N	15N	107.837	P74	CG	13C	27.4388
G68	HA3	1H	4.012	G71	QA	1H	3.822	P74	HA	1H	4.38
G68	HN	1H	7.98558	G71a	C	13C	172.796	P74	HB1	1H	1.83875
G68	N	15N	110.597	G71a	CA	13C	45.4675	P74	HB2	1H	2.20678
G68a	HN	1H	8.053	G71a	HN	1H	8.2926	P74	HD1	1H	3.6175
G68a	N	15N	110.66	G71a	N	15N	108.439	P74	HD2	1H	3.74371
A69	HN	1H	8.112	I72	C	13C	174.227	P74	QG	1H	1.9591
A69	N	15N	123.643	I72	CA	13C	58.5465	G75	N	15N	108.857
D69	C	13C	176.034	I72	CB	13C	38.7455	N75	C	13C	174.066
D69	CA	13C	41.1427	I72	HA	1H	4.394	N75	CA	13C	51.1399
D69	CB	13C	40.553	I72	HB	1H	1.794	N75	CB	13C	38.8485
D69	HA	1H	4.6634	I72	HG12	1H	1.06	N75	CG	13C	177.178
D69	HB1	1H	2.5385	I72	HG13	1H	1.444	N75	HA	1H	4.92967
D69	HB2	1H	2.66	I72	HN	1H	7.75626	N75	HB1	1H	2.66975
D69	HB3	1H	2.698	I72	N	15N	120.857	N75	HB2	1H	2.79967
D69	HN	1H	8.019	I72	QD1	1H	0.811	N75	HB3	1H	2.862
D69	N	15N	121.217	I72	QG2	1H	0.8675	N75	HD21	1H	6.93675
K70	N	15N	120.321	I72a	C	13C	174.297	N75	HD22	1H	7.6355
L70	C	13C	177.525	I72a	CA	13C	57.821	N75	HN	1H	8.39213
L70	CA	13C	55.2635	I72a	CB	13C	40.4405	N75	N	15N	119.421
L70	CB	13C	42.0777	I72a	HA	1H	4.161	N75	ND2	15N	113.17
L70	CD1	13C	23.0223	I72a	HB	1H	1.702	P76	C	13C	177.318
L70	CD2	13C	25.9653	I72a	HN	1H	7.477	P76	CA	13C	64.1869
L70	CG	13C	27	I72a	N	15N	118.767	P76	CB	13C	32.2242

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
P76	CD	13C	50.977	R79	HG2	1H	1.605	S82	HA	1H	4.688
P76	CG	13C	27.3405	R79	HN	1H	8.04869	S82	HN	1H	8.26475
P76	HA	1H	4.29029	R79	N	15N	120.902	S82	N	15N	117.325
P76	HB1	1H	1.9439	R79	QD	1H	3.13842	S82	QB	1H	3.7895
P76	HB2	1H	2.2775	R79a	C	13C	175.108	E83	HN	1H	8.426
P76	HD1	1H	3.73744	R79a	CA	13C	55.73	E83	N	15N	122.88
P76	HD2	1H	3.79411	R79a	CB	13C	31.2677	P83	CA	13C	63.266
P76	QG	1H	1.99644	R79a	HN	1H	7.9555	P83	CB	13C	32.0758
E77	C	13C	176.4	R79a	N	15N	121.073	P83	CD	13C	50.816
E77	CA	13C	57.0204	S80	C	13C	172.577	P83	CG	13C	27.324
E77	CB	13C	29.7424	S80	CA	13C	56.7438	P83	HA	1H	4.303
E77	CG	13C	36.4923	S80	CB	13C	63.2745	P83	HB1	1H	1.6565
E77	HA	1H	4.17513	S80	HA	1H	4.66267	P83	HB2	1H	2.044
E77	HB1	1H	1.87633	S80	HB1	1H	3.78425	P83	HD1	1H	3.646
E77	HB2	1H	2.03933	S80	HB2	1H	3.8298	P83	HD2	1H	3.756
E77	HB3	1H	2.04	S80	HB3	1H	3.846	P83	QG	1H	1.98367
E77	HN	1H	8.40246	S80	HN	1H	8.24662	E84	C	13C	174.546
E77	N	15N	118.41	S80	N	15N	118.336	E84	CA	13C	54.4726
E77	QG	1H	2.207	S80a	C	13C	173.006	E84	CB	13C	29.8373
D78	C	13C	176.085	S80a	CA	13C	55.7875	E84	CG	13C	36.0753
D78	CA	13C	54.5009	S80a	CB	13C	64.2815	E84	HA	1H	4.49412
D78	CB	13C	41.1118	S80a	HA	1H	4.554	E84	HB1	1H	1.8325
D78	CG	13C	179.982	S80a	HN	1H	8.05389	E84	HB2	1H	1.95771
D78	HA	1H	4.54167	S80a	N	15N	116.815	E84	HB3	1H	1.986
D78	HB1	1H	2.597	S80a	QB	1H	3.678	E84	HN	1H	8.20817
D78	HB2	1H	2.71075	P81	C	13C	176.725	E84	N	15N	122.184
D78	HN	1H	7.91864	P81	CA	13C	63.2841	E84	QG	1H	2.256
D78	N	15N	120.608	P81	CB	13C	32.0516	K84	N	15N	127.065
D78	QB	1H	2.652	P81	CD	13C	50.744	P85	C	13C	176.022
D78a	CA	13C	53.9935	P81	CG	13C	27.449	P85	CA	13C	63.2058
D78a	CB	13C	40.812	P81	HA	1H	4.42118	P85	CB	13C	32.078
R79	C	13C	176.121	P81	HB1	1H	1.8826	P85	CD	13C	50.7787
R79	CA	13C	55.8726	P81	HB2	1H	2.2443	P85	CG	13C	27.2778
R79	CB	13C	30.655	P81	HD1	1H	3.6631	P85	HA	1H	4.4152
R79	CD	13C	43.4434	P81	HD2	1H	3.7469	P85	HB1	1H	1.86625
R79	CG	13C	27.1424	P81	QG	1H	1.9734	P85	HB2	1H	2.22289
R79	HA	1H	4.3295	H82	N	15N	120.268	P85	HD1	1H	3.64237
R79	HB1	1H	1.7057	S82	C	13C	172.803	P85	HD2	1H	3.7476
R79	HB2	1H	1.8709	S82	CA	13C	56.3253	P85	QG	1H	1.95944
R79	HG1	1H	1.55944	S82	CB	13C	63.5847	I86	?	1H	1.003

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
I86	C	13C	175.415	N88	HA	1H	4.7498	K92	HN	1H	7.92207
I86	CA	13C	61.0083	N88	HB1	1H	2.8	K92	N	15N	120.534
I86	CB	13C	39.3097	N88	HB2	1H	3.0556	K92	QE	1H	2.9175
I86	CD1	13C	13.0898	N88	HB3	1H	3.098	L92	C	13C	176.453
I86	CG1	13C	27.1892	N88	HD21	1H	6.905	R93	C	13C	176.323
I86	CG2	13C	17.6484	N88	HD22	1H	7.60775	R93	CA	13C	56.4031
I86	HA	1H	4.06246	N88	HN	1H	8.37593	R93	CB	13C	30.6711
I86	HB	1H	1.63315	N88	N	15N	119.95	R93	CD	13C	43.4283
I86	HG11	1H	1.0319	N88	ND2	15N	111.898	R93	CG	13C	27.4863
I86	HG12	1H	1.30845	S89	C	13C	175.066	R93	HA	1H	4.18736
I86	HG13	1H	1.327	S89	CA	13C	60.0073	R93	HB1	1H	1.66375
I86	HN	1H	7.95342	S89	CB	13C	63.337	R93	HB2	1H	1.7295
I86	N	15N	120.251	S89	HA	1H	4.241	R93	HN	1H	8.38733
I86	QD	1H	0.7532	S89	HB1	1H	3.84	R93	N	15N	123.017
I86	QD1	1H	0.705	S89	HB2	1H	3.929	R93	QD	1H	3.06712
I86	QG2	1H	0.7205	S89	HN	1H	8.3616	R93	QG	1H	1.46162
I86a	C	13C	175.401	S89	N	15N	115.767	L94	?	1H	1.4325
I86a	CA	13C	61.4455	S89	QB	1H	3.849	L94	C	13C	176.974
I86a	CB	13C	38.9825	E90	C	13C	176.935	L94	CA	13C	55.0334
Y87	C	13C	175.92	E90	CA	13C	56.7728	L94	CB	13C	42.7727
Y87	CA	13C	57.2147	E90	CB	13C	29.9801	L94	CD1	13C	23.477
Y87	CB	13C	40.1935	E90	CG	13C	36.6733	L94	CD2	13C	25.059
Y87	CD1	13C	133.11	E90	HA	1H	4.27762	L94	CG	13C	27.005
Y87	CE1	13C	118.016	E90	HB1	1H	1.94667	L94	HA	1H	4.26
Y87	HA	1H	4.82517	E90	HB2	1H	2.11833	L94	HB1	1H	1.4318
Y87	HB1	1H	2.78375	E90	HN	1H	7.98283	L94	HB2	1H	1.4912
Y87	HB2	1H	2.8904	E90	N	15N	120.735	L94	HB3	1H	1.491
Y87	HB3	1H	2.928	E90	QG	1H	2.21067	L94	HD11	1H	0.753222
Y87	HD1	1H	6.965	G91	C	13C	174.114	L94	HD21	1H	0.785222
Y87	HE1	1H	6.72	G91	CA	13C	45.7047	L94	HG	1H	1.5044
Y87	HN	1H	8.16161	G91	HA1	1H	3.7215	L94	HN	1H	8.25671
Y87	N	15N	122.893	G91	HA2	1H	3.8505	L94	N	15N	124.095
Y87a	CA	13C	57.01	G91	HA3	1H	3.98233	N95	C	13C	175.594
Y87a	CB	13C	40.089	G91	HN	1H	8.16257	N95	CA	13C	52.9215
Y87a	HN	1H	8.13683	G91	N	15N	108.332	N95	CB	13C	38.6364
Y87a	N	15N	123.04	K92	?	1H	1.668	N95	CG	13C	177.171
N88	C	13C	176.277	K92	C	13C	176.387	N95	H	1H	8.438
N88	CA	13C	52.144	K92	CA	13C	55.8407	N95	HA	1H	4.7766
N88	CB	13C	38.9388	K92	CB	13C	33.1205	N95	HB1	1H	2.76475
N88	CG	13C	176.907	K92	HA	1H	4.3095	N95	HB2	1H	2.81725

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
N95	HB3	1H	2.8365	F99	CB	13C	39.4636	K103	HG3	1H	1.35
N95	HD21	1H	6.88475	F99	CD1	13C	131.768	K103	HN	1H	7.90021
N95	HN	1H	8.42811	F99	CE1	13C	131.64	K103	N	15N	117.425
N95	N	15N	119.575	F99	HA	1H	4.32267	K103	QB	1H	1.69262
N95	ND2	15N	112.471	F99	HD1	1H	7.18	K103	QD	1H	1.56421
N95	ND22	1H	7.58375	F99	HE1	1H	7.298	K104	C	13C	179.374
T96	C	13C	175.189	F99	HN	1H	8.14325	K104	CA	13C	58.9671
T96	CA	13C	62.65	F99	N	15N	121.235	K104	CB	13C	32.1836
T96	CB	13C	69.9413	F99	QB	1H	3.13033	K104	CD	13C	29.0684
T96	CG2	13C	21.528	R100	C	13C	178.181	K104	CE	13C	42.2491
T96	HA	1H	4.22117	R100	CA	13C	58.533	K104	CG	13C	24.9062
T96	HB	1H	4.1008	R100	CB	13C	30.5883	K104	HA	1H	4.14821
T96	HN	1H	8.00777	R100	HN	1H	8.37288	K104	HB1	1H	1.93164
T96	N	15N	114.069	R100	N	15N	120.193	K104	HB2	1H	1.99357
T96	QG2	1H	1.1428	T101	C	13C	175.294	K104	HD1	1H	1.68453
R97	C	13C	177.938	T101	CA	13C	65.022	K104	HD2	1H	1.72631
R97	CA	13C	58.453	T101	CB	13C	69.3037	K104	HG1	1H	1.4545
R97	CB	13C	30.2871	T101	CG2	13C	22.146	K104	HG2	1H	1.54193
R97	CD	13C	43.4952	T101	HA	1H	4.0208	K104	HN	1H	7.57275
R97	CG	13C	27.092	T101	HB	1H	4.2992	K104	N	15N	119.701
R97	HA	1H	4.11244	T101	HN	1H	7.96092	K104	QE	1H	2.96314
R97	HG1	1H	1.5465	T101	N	15N	115.298	L105	C	13C	179.476
R97	HG2	1H	1.63389	T101	QG2	1H	1.2352	L105	CA	13C	58.2962
R97	HG3	1H	1.653	R102	C	13C	177.642	L105	CB	13C	41.9847
R97	HN	1H	8.48877	R102	CA	13C	60.6112	L105	CD1	13C	23.206
R97	N	15N	122.442	R102	CB	13C	29.8078	L105	CD2	13C	26.2135
R97	QB	1H	1.825	R102	HA	1H	4.307	L105	CG	13C	27.312
R97	QD	1H	3.16967	R102	HN	1H	8.347	L105	HA	1H	4.11829
E98	C	13C	177.111	R102	N	15N	122.207	L105	HB1	1H	1.3044
E98	CA	13C	58.7121	R102	QD	1H	3.469	L105	HB2	1H	1.85625
E98	CB	13C	29.8598	K103	C	13C	178.91	L105	HD11	1H	0.802
E98	CG	13C	36.627	K103	CA	13C	59.0436	L105	HD21	1H	0.79275
E98	HA	1H	4.01512	K103	CB	13C	32.043	L105	HG	1H	1.86167
E98	HB1	1H	1.91783	K103	CD	13C	29.0809	L105	HN	1H	8.02777
E98	HB2	1H	1.96529	K103	CE	13C	42.0541	L105	N	15N	120.312
E98	HB3	1H	1.985	K103	CG	13C	24.6367	L105	QD	1H	0.804
E98	HN	1H	8.49769	K103	HA	1H	3.954	E106	C	13C	180.566
E98	N	15N	121.543	K103	HE1	1H	2.836	E106	CA	13C	59.792
E98	QG	1H	2.20857	K103	HE2	1H	2.91275	E106	CB	13C	29.7745
F99	C	13C	177.07	K103	HG1	1H	1.24317	E106	HA	1H	3.988
F99	CA	13C	60.063	K103	HG2	1H	1.33177	E106	HN	1H	8.63117

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
E106	N	15N	118.78	N111	CB	13C	37.6488	T114	CB	13C	68.7838
E107	C	13C	178.455	N111	CG	13C	175.752	T114	CG2	13C	21.486
E107	CA	13C	59.821	N111	HA	1H	4.27886	T114	HA	1H	3.718
E107	CB	13C	29.5083	N111	HB1	1H	2.79425	T114	HB	1H	4.18167
E107	HA	1H	4.121	N111	HB2	1H	2.88914	T114	HN	1H	8.03795
E107	HB2	1H	2.035	N111	HB3	1H	2.96333	T114	N	15N	114.805
E107	HB3	1H	2.18	N111	HD21	1H	7.0225	T114	QG2	1H	1.0996
E107	HG2	1H	2.254	N111	HD22	1H	7.603	E115	C	13C	178.189
E107	HG3	1H	2.47	N111	HN	1H	8.292	E115	CA	13C	59.4105
E107	HN	1H	8.2391	N111	N	15N	119.387	E115	CB	13C	29.2375
E107	N	15N	122.094	N111	ND2	15N	112.211	E115	HA	1H	3.9325
E108	C	13C	179.336	L112	C	13C	178.548	E115	HB2	1H	1.945
E108	CA	13C	59.6347	L112	CA	13C	57.9804	E115	HB3	1H	2.101
E108	CB	13C	29.555	L112	CB	13C	42.5418	E115	HG2	1H	1.993
E108	HA	1H	4.107	L112	CD1	13C	22.7	E115	HG3	1H	2.277
E108	HB2	1H	2.053	L112	CD2	13C	27.5893	E115	HN	1H	7.33556
E108	HB3	1H	2.165	L112	CG	13C	27.319	E115	N	15N	123.051
E108	HG2	1H	2.18	L112	HA	1H	4.12125	M116	C	13C	177.881
E108	HG3	1H	2.416	L112	HD11	1H	0.8674	M116	CA	13C	57.0352
E108	HN	1H	8.305	L112	HD21	1H	0.8104	M116	CB	13C	32.203
E108	N	15N	120.79	L112	HG	1H	1.80167	M116	CG	13C	29.518
R109	C	13C	177.178	L112	HN	1H	7.92822	M116	HA	1H	4.11637
R109	CA	13C	60.4717	L112	N	15N	122.059	M116	HB1	1H	1.914
R109	CB	13C	30.015	I113	C	13C	176.961	M116	HB2	1H	1.9735
R109	HN	1H	8.5698	I113	CA	13C	66.1216	M116	HB3	1H	2.092
R109	N	15N	118.571	I113	CB	13C	37.5469	M116	HG1	1H	2.1865
H110	C	13C	177.577	I113	CD1	13C	14.2207	M116	HG2	1H	2.252
H110	CA	13C	60.6048	I113	CG1	13C	31.0972	M116	HG3	1H	2.269
H110	CB	13C	30.43	I113	CG2	13C	17.123	M116	HN	1H	8.34875
H110	CD2	13C	119.445	I113	HA	1H	3.23825	M116	N	15N	119.712
H110	CE1	13C	138.556	I113	HB	1H	1.67738	V117	C	13C	177.561
H110	HA	1H	4.34857	I113	HG11	1H	0.6756	V117	CA	13C	64.8373
H110	HB1	1H	3.404	I113	HG12	1H	1.46125	V117	CB	13C	31.9376
H110	HB2	1H	3.45267	I113	HG13	1H	1.725	V117	CG1	13C	21.455
H110	HB3	1H	3.487	I113	HN	1H	8.44138	V117	CG2	13C	22.6357
H110	HD2	1H	7.116	I113	N	15N	121.26	V117	HA	1H	3.897
H110	HE1	1H	7.7855	I113	QD	1H	0.225667	V117	HB	1H	2.08229
H110	HN	1H	8.432	I113	QD1	1H	0.242	V117	HN	1H	8.24782
H110	N	15N	119.428	I113	QG2	1H	0.5381	V117	N	15N	117.446
N111	C	13C	178.624	T114	C	13C	176.558	V117	QG1	1H	1.00729
N111	CA	13C	56.3155	T114	CA	13C	66.7442	V117	QG2	1H	1.116

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
A118	C	13C	179.594	P121	CG	13C	26.9976	K124	N	15N	129.209
A118	CA	13C	54.1574	P121	HA	1H	4.6571	K124	QB	1H	1.46867
A118	CB	13C	18.1984	P121	HB1	1H	1.9599	K124	QE	1H	2.932
A118	HA	1H	4.1338	P121	HB2	1H	2.4132	P125	CA	13C	60.9455
A118	HB1	1H	1.449	P121	HB3	1H	2.431	P125	CB	13C	31.29
A118	HN	1H	7.04606	P121	HD1	1H	3.5416	P125	CD	13C	51.087
A118	N	15N	120.403	P121	HD2	1H	3.79575	P125	CG	13C	27.0251
A118	QB	1H	1.436	P121	HD3	1H	3.852	P125	HA	1H	4.08758
L119	?	1H	1.868	P121	QG	1H	2.0206	P125	HB1	1H	1.85764
L119	C	13C	177.859	D122	C	13C	176.001	P125	HB2	1H	1.89718
L119	CA	13C	56.6721	D122	CA	13C	54.2729	P125	HD1	1H	3.55945
L119	CB	13C	43.2458	D122	CB	13C	41.3826	P125	HD2	1H	3.8335
L119	CD1	13C	22.757	D122	CG	13C	179.922	P125	HG1	1H	1.79986
L119	CD2	13C	25.0784	D122	HA	1H	4.67443	P125	HG2	1H	2.0685
L119	CG	13C	26.501	D122	HB1	1H	2.542	P126	C	13C	177.667
L119	HA	1H	4.13817	D122	HB2	1H	2.67657	P126	CA	13C	62.5461
L119	HB1	1H	1.3915	D122	HB3	1H	2.762	P126	CB	13C	32.2871
L119	HB2	1H	1.73114	D122	HN	1H	7.56567	P126	CD	13C	49.7807
L119	HB3	1H	1.8735	D122	N	15N	115.588	P126	CG	13C	27.9706
L119	HD11	1H	0.8562	F123	C	13C	173.949	P126	HA	1H	4.37823
L119	HD21	1H	0.778	F123	CA	13C	59.0162	P126	HB1	1H	2.00958
L119	HG	1H	1.79967	F123	CB	13C	39.6014	P126	HB2	1H	2.34931
L119	HN	1H	7.49048	F123	CD	13C	132.233	P126	HB3	1H	2.388
L119	N	15N	117.735	F123	CE1	13C	130.9	P126	HD1	1H	3.2055
N120	C	13C	172.51	F123	H	1H	7.899	P126	HD2	1H	3.60858
N120	CA	13C	50.068	F123	HA	1H	4.13483	P126	HG1	1H	1.83425
N120	CB	13C	39.2731	F123	HB1	1H	2.81125	P126	HG2	1H	2.0605
N120	CG	13C	177.943	F123	HB2	1H	3.02417	A127	C	13C	177.643
N120	HA	1H	5.02043	F123	HB3	1H	3.142	A127	CA	13C	54.6396
N120	HB1	1H	2.5825	F123	HD1	1H	7.016	A127	CB	13C	18.8356
N120	HB2	1H	2.8186	F123	HE1	1H	7.106	A127	HA	1H	4.0626
N120	HB3	1H	2.878	F123	HN	1H	7.90411	A127	HN	1H	8.5988
N120	HD21	1H	7.07075	F123	N	15N	122.446	A127	N	15N	123.67
N120	HD22	1H	7.34167	K124	?	1H	1.472	A127	QB	1H	1.40125
N120	HN	1H	8.43429	K124	C	13C	172.622	D128	C	13C	175.718
N120	N	15N	115.05	K124	CA	13C	51.69	D128	CA	13C	53.5939
N120	ND2	15N	110.02	K124	CB	13C	32.5793	D128	CB	13C	40.3428
P121	C	13C	177.02	K124	HA	1H	4.55267	D128	CG	13C	181.265
P121	CA	13C	64.2542	K124	HG2	1H	1.287	D128	HA	1H	4.45214
P121	CB	13C	32.3412	K124	HG3	1H	1.332	D128	HB1	1H	2.47425
P121	CD	13C	50.6638	K124	HN	1H	7.58779	D128	HB2	1H	2.75

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
D128	HB3	1H	2.819	P132	CD	13C	50.3567	S137	C	13C	174.257
D128	HN	1H	8.307	P132	CG	13C	27.16	S137	CA	13C	58.3104
D128	N	15N	112.716	P132	HA	1H	4.23491	S137	CB	13C	63.9288
Y129	C	13C	174.188	P132	HB1	1H	1.79025	S137	HA	1H	4.40783
Y129	CA	13C	58.3705	P132	HB2	1H	2.14333	S137	HB1	1H	3.773
Y129	CB	13C	38.9879	P132	HD1	1H	3.17533	S137	HB2	1H	3.80225
Y129	CD1	13C	133.5	P132	HD2	1H	3.40964	S137	HB3	1H	3.8555
Y129	CE1	13C	118.016	P132	HG1	1H	1.68345	S137	HN	1H	8.25557
Y129	HA	1H	4.36343	P132	HG2	1H	1.73764	S137	N	15N	118.706
Y129	HB1	1H	2.6055	A133	C	13C	177.865	S137	QB	1H	3.853
Y129	HB2	1H	2.8874	A133	CA	13C	52.5254	D138	C	13C	176.051
Y129	HB3	1H	2.9605	A133	CB	13C	19.2418	D138	CA	13C	54.537
Y129	HD1	1H	6.941	A133	HA	1H	4.2706	D138	CB	13C	41.1178
Y129	HE1	1H	6.738	A133	HN	1H	8.28623	D138	HA	1H	4.54017
Y129	HN	1H	7.5649	A133	N	15N	124.193	D138	HB1	1H	2.609
Y129	N	15N	120.912	A133	QB	1H	1.3154	D138	HB2	1H	2.71525
K130	?	1H	1.3335	T134	C	13C	174.28	D138	HN	1H	8.21893
K130	C	13C	172.314	T134	CA	13C	61.9275	D138	N	15N	122.703
K130	CA	13C	52.539	T134	CB	13C	69.9262	D138	QB	1H	2.621
K130	CB	13C	33.5705	T134	CG2	13C	21.738	K139	?	1H	1.695
K130	HA	1H	4.305	T134	HA	1H	4.22533	K139	C	13C	176.386
K130	HB2	1H	1.42	T134	HB	1H	4.0955	K139	CA	13C	56.3695
K130	HB3	1H	1.529	T134	HN	1H	7.99179	K139	CB	13C	33.0985
K130	HG2	1H	1.197	T134	N	15N	113.662	K139	HA	1H	4.258
K130	HG3	1H	1.264	T134	QG2	1H	1.128	K139	HG2	1H	1.346
K130	HN	1H	7.66437	R135	?	1H	1.524	K139	HN	1H	8.06545
K130	N	15N	129.691	R135	C	13C	176.217	K139	N	15N	120.957
K130	QE	1H	2.925	R135	CA	13C	55.993	K139	C	13C	176.234
P131	CA	13C	61.0622	R135	CB	13C	30.9925	K139	CA	13C	56.302
P131	CB	13C	31	R135	HA	1H	4.349	K139	CB	13C	33.2155
P131	CD	13C	50.5296	R135	HN	1H	8.28292	V140	C	13C	175.842
P131	CG	13C	27.0195	R135	N	15N	124.061	V140	CA	13C	62.4699
P131	HA	1H	4.2703	V136	C	13C	176.175	V140	CB	13C	32.7944
P131	HB1	1H	1.7865	V136	CA	13C	62.6246	V140	H	1H	8.002
P131	HB2	1H	2.29856	V136	CB	13C	32.7338	V140	HA	1H	4.01314
P131	HD1	1H	3.2936	V136	HA	1H	4.07029	V140	HB	1H	1.9945
P131	HD2	1H	3.4319	V136	HB	1H	2.0332	V140	HN	1H	7.99407
P131	QG	1H	1.95956	V136	HN	1H	8.21206	V140	N	15N	121.547
P132	C	13C	176.459	V136	N	15N	122.119	V140	QG1	1H	0.859
P132	CA	13C	62.8781	V136	QG1	1H	0.8765	V140	QG2	1H	0.87075
P132	CB	13C	31.9929	V136	QG2	1H	0.883333	M141	C	13C	175.558

Res.	Atom	Nuc.	Shift	Res.	Atom	Nuc.	Shift
M141	CA	13C	55.1608	Q144	CG	13C	33.8939
M141	CB	13C	32.9152	Q144	H	1H	8.387
M141	CG	13C	32.0355	Q144	HA	1H	4.30129
M141	HA	1H	4.4559	Q144	HB1	1H	1.916
M141	HB1	1H	1.92767	Q144	HB2	1H	2.09883
M141	HB2	1H	1.97222	Q144	HE21	1H	6.7775
M141	HB3	1H	2.003	Q144	HE22	1H	7.548
M141	HG1	1H	2.44462	Q144	HN	1H	8.397
M141	HG2	1H	2.51544	Q144	N	15N	121.555
M141	HG3	1H	2.528	Q144	NE2	15N	112.861
M141	HN	1H	8.35071	Q144	QB	1H	2.098
M141	N	15N	124.82	Q144	QG	1H	2.35014
I142	?	1H	0.881	D145	C	13C	180.743
I142	C	13C	174.331	D145	CA	13C	55.75
I142	CA	13C	58.538	D145	CB	13C	42.2672
I142	CB	13C	38.648	D145	HA	1H	4.3498
I142	CD1	13C	12.742	D145	HB1	1H	2.5152
I142	HA	1H	4.421	D145	HB2	1H	2.5986
I142	HB	1H	1.807	D145	HN	1H	7.9193
I142	HN	1H	8.09355	D145	N	15N	127.536
I142	N	15N	124.525				
I142	QD	1H	0.796				
I142	QG1	1H	1.052				
P143	C	13C	176.617				
P143	CA	13C	63.4066				
P143	CB	13C	32.2052				
P143	CD	13C	51.2324				
P143	CG	13C	27.5848				
P143	HA	1H	4.34675				
P143	HB1	1H	1.89655				
P143	HB2	1H	2.24918				
P143	HD1	1H	3.6436				
P143	HD2	1H	3.82925				
P143	HG1	1H	1.92142				
P143	HG2	1H	2.02817				
Q144	C	13C	174.752				
Q144	CA	13C	55.6327				
Q144	CB	13C	30.1147				
Q144	CD	13C	180.931				

III.IV Abbreviations

1D, 2D, 3D	one-, two-, three-dimensional	NMR	nuclear magnetic resonance
aa	amino acid	NTP	Nucleoside Triphosphate
ATP	Adenosintriphosphat	PAGE	Polyacrylamide Gel Electrophoresis
ATPase	Adenosintriphosphatase	PCR	polymerase chain reaction
BMRB	Biological Magnetic Resonance Bank	PCS	pseudo contact shift
CD	Circular Dichroism	pre-mRNA	precursor messenger RNA
CSP	chemical shift perturbation	ppm	parts per million
DHAP	dihydroxyacetone phosphate	PRE	paramagnetic relaxation enhancement
DMSO	Dimethyl Sulfoxide	PTM	posttranslational modification
DNA	Deoxyribonucleic Acid	RDC	residual dipolar coupling
DTT	dithiothreitol	rmsd	root mean square deviation
EMSA	Electrophoretic Mobility Shift Assay	RNA	Ribonucleic Acid
FID	free induction decay	RNase	Ribonuclease
fl	full-length	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
FPP	farnesyl pyrophosphate	SF1	Splicing Factor 1
FT-NMR	Fourier-Transform NMR	TEV	tobacco etch virus
Gd(DTPA-B MA)	diethylenetriaminepentaaceti cacid bismethylamide	TMS	tetramethylsilane
HSQC	heteronuclear single quantum correlation	TOCSY	total correlation spectroscopy
KIS	kinase interacting with stathmin	TROSY	transverse relaxation optimized spectroscopy
LB	lysogeny broth	U2AF	U2 snRNP auxiliary factor
MS	Mass Spectrometry	ULM	UHM ligand motifs
MWCO	molecular weight cut-off	UHM	U2AF homology motifs
NOE	nuclear Overhauser effect		
NOESY	nuclear Overhauser effect spectroscopy		

Amino acids		Ribonucleotides	
Ala, A	alanine	A	Adenine
Arg, R	arginine	G	Guanine
Asn, N	asparagine	C	Cytosine
Asp, D	aspartic acid	U	Uracil
Cys, C	cysteine	mC	5 Methylcytosine
Glu, E	glutamic acid		
Gln, Q	glutamine	ATP	Adenosine Triphosphate
Gly, G	glycine	GTP	Guanosine Triphosphate
His, H	histidine	CTP	Cytidine Triphosphate
Ile, I	isoleucine	UTP	Uridine Triphosphate
Leu, L	leucine		
Lys, K	lysine	Concentrations	
Met, M	methionine	v/v	volume to volume
Phe, F	phenylalanine	w/v	weight to volume
Pro, P	proline	w/w	weight to weight
Ser, S	serine		
Thr, T	threonine		
Trp, W	tryptophane		
Tyr, Y	tyrosine		
Val, V	valine		

ACKNOWLEDGEMENTS

First and foremost, I would like to show my deepest gratitude to my supervisor Michael Sattler, a scholar full of enthusiasm about work and with passion for science. I would like to thank him for taking me up in his lab and giving me the opportunity to work on this interesting and exiting project. When I was about to lose confidence in phosphorylation experiments, his suggestions and encouragements pulled me through.

I am greatly indebted to my colleague Tobias Madl, whom I would like to consider as my second supervisor. Tobias is a brilliant and hardworking scientist. To work with him was a challenging, informative and fruitful experience. I am very lucky to know him as a colleague, a supervisor as well as a reliable and kind friend.

My sincere regards go to the members of my Thesis Advisory Committee Prof. Dr. Marius Ueffing from the University Eye Hospital Tuebingen and Prof. Dr. Christian Becker from University of Vienna, for their guidance and support during my PhD.

I would like to thank Fatiha Kateb and Konstantinos Tripsianes, who helped me a lot with NMR and NMR spectrometers. Peijian Zou contributed to the wet lab works. I will never forget his help when I was looking for an apartment in Munich. Hanso's carefulness and brilliance made the final success of SF1 phosphorylation. Ana Messias, without her valuable advices, I would never finish this project. I would like to acknowledge Thomas Kern, who helped me with analysis of relaxation data, and Waltraud Wolfson for taking care of my incredibly large amounts of paperwork.

With particular gratitude, I would like to thank all of my former and present colleagues in the Sattler group for interesting 4.5 years. My special thanks go to Hamed Kooshapur, Alex Beribisky, Elke Prade and Diana Rodriguez. To work with them made life much more fun, and because of them, I had a lot of fun in the group retreats. I would like to thank André Mourão, who picked me up at the airport when I came to Munich. Going to Oktoberfest in my first working day was a good beginning of the doctoral studying.

Last but not least, I owe my deepest gratitude to my love, Yuxiang Zhou, who has incessantly supported me all along this challenging 4.5 years. Life means nothing to me without you. And I would like to thank my parents for their understanding of my studying abroad. I know it was not easy for you without the only son's company during the Spring Festival in the past three years. And now I am coming back home ...

CURRICULUM VITAE



NAME	Zhang Yun
DATE OF BIRTH	30.10.1981
NATIONALITY	Chinese
ADDRESS (work)	Technical University of Munich Biomolecular NMR Spectroscopy Lichtenbergstr. 4 85747 Garching Germany E-mail: zhangyun@mytum.de

EDUCATION

09/2008 to present	Doctoral student, Technische Universität München
09/2004 – 07/2007	Master of Science in Biology, Beijing Normal University
09/1999 – 07/2003	Bachelor of Science in Chemistry, Xiangtan University
09/1996 – 07/1999	High School at Taoyuan, Hunan, China

WORK EXPERIENCE

09/2008 to present	Doctoral Thesis Research at the Chair of Biomolecular NMR Spectroscopy in the Department of Chemistry, Technische Universität München, and Institute of Structural Biology, Helmholtz Zentrum München Title: Structural study of SF1/U2AF65 in cooperative recognition of 3' splice site and its regulation. Supervisor: Prof. Dr. M. Sattler
09/2004 – 08/2008	Master Thesis Research at The National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences. Title: Using cadmium telluride quantum dots as a proton flux sensor and applying to detect H9 avian influenza virus. Supervisor: Prof. Dr. J.C. Yue

PUBLICATIONS

1. **Zhang, Y.**, Madl, T. (shared first author), Bagdiul, I., Kern, T., Kang, H.S., Zou, P., Mausbacher, N., Sieber, S.A., Kramer, A., and Sattler, M. (2013). Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3'-splice site recognition. *Nucleic Acids Res* **41**, 1343-1354.
2. Deng, Z., **Zhang, Y. (shared first author)**, Yue, J., Tang, F., and Wei, Q. (2007). Green and orange CdTe quantum dots as effective pH-sensitive fluorescent probes for dual simultaneous and independent detection of viruses. *The journal of physical chemistry B* **111**, 12024-12031. (Highlighted by Nature Nanotechnology. doi:10.1038/nnano.2007.361)
3. **Zhang, Y.**, Deng, Z., Yue, J., Tang, F., and Wei, Q. (2007). Using cadmium telluride quantum dots as a proton flux sensor and applying to detect H9 avian influenza virus. *Analytical biochemistry* **364**, 122-127.
4. Liu, X., **Zhang, Y.**, Yue, J., Jiang, P., and Zhang, Z. (2006). FOF1-ATPase as biosensor to detect single virus. *Biochemical and biophysical research communications* **342**, 1319-1322.
5. Liu, X., Zhao, R., **Zhang, Y.**, Jiang, X., Yue, J., Jiang, P., and Zhang, Z. (2007). Using giant unilamellar lipid vesicle micro-patterns as ultrasmall reaction containers to observe reversible ATP synthesis/hydrolysis of FOF1-ATPase directly. *Biochimica et biophysica acta* **1770**, 1620-1626.