TABLE of CONTENTS

Summary

This thesis describes two novel three-dimensional structures and the functional characterization of proteins that play important roles in eukaryotic RNA splicing. These results are discussed in Chapters 1 and 2, while biomolecular NMR techniques that were employed for the structure determination are outlined in Chapter 3. Materials and methods are described in Chapter 4.

Chapter 1 presents the solution structure of the Tudor domain of the human Survival of Motor Neuron (SMN) protein and its molecular interaction with the spliceosomal Sm proteins. Sm proteins are common components of small nuclear ribonucleoprotein particles (snRNPs), which are assembled by a protein complex that contains SMN. The structure of the SMN Tudor domain exhibits a five stranded β-barrel, which resembles the fold of Sm proteins. The Tudor domain of SMN binds to arginine and glycine-rich C-terminal regions of Sm proteins, where it specifically recognizes symmetrically di-methylated arginine residues. The E134K mutant Tudor domain, which corresponds to a human mutation associated with Spinal Muscular Atrophy (SMA), is structurally intact but fails to interact with Sm proteins. This provides an explanation for a molecular defect underlying SMA.

In Chapter 2, the structural basis for the molecular recognition between the essential splicing factors SF1 and U2 auxiliary factor 2 (U2AF) is provided. This interaction involves the third RNA recognition motif (RRM3) of the large subunit of U2AF ($U2AF^{65}$) and the N-terminal 25 residues of SF1. The structure of RRM3 exhibits the classical RNP-type fold, but contains an additional C-terminal helix. SF1 is bound by the helical surface of RRM3, opposite of the canonical RNA binding site. The molecular recognition involves insertion of a conserved tryptophan of SF1 into a hydrophobic binding pocket of RRM3. This interaction is complemented by electrostatic contacts that are mediated by acidic residues of RRM3 and basic amino acids of SF1. Surprisingly, the molecular interface is highly similar to that between the large ($U2AF^{65}$) and small ($U2AF^{35}$) subunits of U2AF. This RRM-mediated protein interaction provides an example of how conserved structural folds have evolved different molecular functions.

Zusammenfassung

In der vorliegende Arbeit werden neue drei-dimensionale Strukturen sowie die funktionelle Charakterisierung von Proteinen beschrieben, welche wichtige Funktionen für das RNA Spleissen in Eukaryonten ausführen. Diese Ergebnisse werden in Kapitel 1 und 2 diskutiert, während die biomolekularen NMR Techniken, welche für die Strukturbestimmung verwendet wurden, in Kapitel 3 erläutert werden. Experimentelle Methoden sind in Kapitel 4 beschrieben.

In Kapitel 1 wird die Struktur der Tudor Domäne des menschlichen 'Survival of Motor Neuron' (SMN) Proteins, sowie dessen Wechselwirkung mit den spleissosomalen Sm Proteinen vorgestellt. Sm Proteine sind gemeinsame Bestandteile der 'small nuclear' Ribonukleoprotein Partikel (snRNP), die von einem Proteinkomplex assembliert werden, welcher SMN enthält. Die Struktur der SMN Tudor Domäne besteht aus einem 5-strängigen β-Faltblatt, das der Faltung der Sm Proteine ähnelt. Die Tudor Domäne des SMN Proteins bindet an Arginin- und Glycin-reiche Sequenzen im C-Terminus der Sm Proteine. Dort erkennt es spezifisch symmetrisch dimethylierte Arginine. Die Struktur der E134K mutanten Tudor Domäne, welche einer genetischen Mutation der spinalen Muskelatrophy (SMA) entspricht, ist nicht beeinträchtigt, kann aber keine Sm Protein Bindung mehr vermitteln. Dies liefert eine Erklärung für einen molekularen Defekt welcher der SMA Krankheit zugrunde liegt.

Kapitel 2 beschreibt die strukturelle Grundlage für die molekulare Erkennung zwischen den essentiellen Spleissfaktoren SF1 und 'U2 auxillary factor' (U2AF). Diese Interaktion wird durch das dritte 'RNA recognition motif' (RRM3) der grossen Untereinheit von U2AF ($U2AF^{65}$) und den ersten 25 Aminosäuren von SF1 vermittelt. Die RRM3 Struktur entspricht der klassischen RNP Faltung, enthält jedoch eine zusätzliche C-terminale Helix. SF1 wird an einer helikalen Oberfläche gebunden, welche sich auf der Rückseite der kanonischen RNA Bindungsstelle befindet. Die molekulare Erkennung wird über einen Tryptophan Rest von SF1 koordiniert, welcher in eine hydrophobe Tasche der RRM3 Domäne bindet. Diese Interaktion wird zusätzlich durch komplementäre elektrostatische Kontakte der sauren Reste von RRM3 und der basischen Reste von SF1 verstärkt.

Überraschenderweise, ist diese molekulare Erkennung fast identisch mit der zwischen der grossen ($U2AF^{65}$) und kleinen ($U2AF^{35}$) Untereinheit von U2AF. Diese RRM3-vermittelte Interaktion ist daher ein Bespiel für die Evolution unterschiedlich molekularer Funktionen einer weit verbreiteten Protein Domäne.

Preface

This thesis is divided into three main chapters. The first two parts (Chapter 1 and Chapter 2) comprise an introduction to the biological area of research and present two novel three-dimensional structures; The Tudor domain of the human Survival of Motor Neuron (SMN) protein and the protein-peptide complex of the third RNA recognition domain (RRM) of the large subunit of human U2 auxiliary factor 2 ($U2AF^{65}$-RRM3) and the N-terminus of human Splicing Factor 1 (SF1). Both three-dimensional structures were determined during the course of this doctoral work. The third part (Chapter 4) will focus on the biophysical method used to determine these structures and evaluate the physical basis and state of the art methods used in the field of biomolecular Nuclear Magnetic Resonance Spectroscopy (NMR).

The first two parts will be more thoroughly evaluated than the third part. No detailed introduction into the quantum mechanical basis of Nuclear Magnetic Resonance Spectroscopy nor into the mathematical formalism underlying this method for protein structure determination will be presented. The interested reader is referred to excellent and extensive literature on this subject, for a more detailed introduction (Chapter 3.7).

1. CHAPTER 1

1.1. INTRODUCTION

1.1.1. Eukaryotic Splicing and Spliceosome Assembly

Most eukaryotic genes contain non-coding intervening sequences (introns) that have to be removed from the primary mRNA transcripts prior to translation into protein. In the nucleus, introns are excised by two successive trans-esterification reactions within a macromolecular assembly called the spliceosome. In the first step, the 5' splice site is attacked by the 2' hydroxyl group of a conserved adenosine at a position known as the intron branch point. The 5' exon is cleaved off and the 5' end is ligated to the 2' hydroxyl group of the branch point adenosine (or BP). This results in a circular lariat intron intermediate. In the second step, the 3' hydroxyl group of the 5' exon attacks the phosphodiester bond at the 3' intron-exon junction thus ligating the two exons and liberating the intron sequence (reviewed in (Staley and Guthrie, 1998) and schematically outlined in Figure 1, left panel).

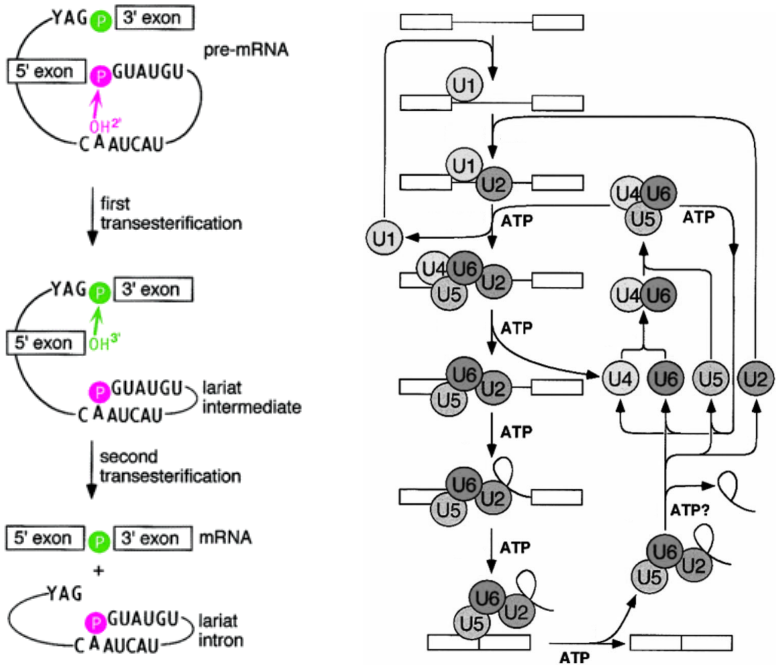


Figure 1
Left panel: Schematic representation of the two trans-esterification reactions leading to exon joining during the splicing reaction
Right panel: stepwise assembly of snRNP onto pre-mRNA. Reproduced from Staley and Guthrie, 1998

The major components of the spliceosome are four RNA-protein complexes, the U1, U2, U4/U6 and U5 U snRNPs (Uridine-rich small nuclear ribonucleoprotein particles). These snRNPs assemble onto the pre-mRNA through an ordered pathway (reviewed in (Kambach et al., 1999a; Will and Luhrmann, 1997; Will and Luhrmann, 2001) and schematically outlined in Figure 1, right panel). In a first step, the U1 snRNP binds to the 5' splice site, while the 3' intron-exon boundary is specifically recognized by the U2 specific auxiliary factor (U2AF) that interacts with Splicing Factor 1 (SF1) (see Chapter 2.1.1 for a more detailed introduction). U2AF consists of two subunits $U2AF^{65}$ and $U2AF^{35}$. $U2AF^{65}$ binds to the conserved poly-pyrimidine stretch (PPT) on pre-mRNA downstream of the BP adenosine. It also interacts with SF1 to enable specific recognition of the BP. Furthermore, the $U2AF^{65}$/$U2AF^{35}$ interaction allows $U2AF^{35}$ to recognize the AG di-nucleotide immediately preceding the exon sequence and results in cooperative recognition of the intron-exon boundary (reviewed in (Hastings and Krainer, 2001; Reed, 1996; Reed, 2000). This early structural arrangement is termed the 'commitment' complex or complex E. Consecutively, SF1 is replaced by SAP155, a protein component of the U2 snRNP, which facilitates further assembly of U2 snRNPs and formation of complex A. The pre-assembled U4/U6 U5 tri-snRNP then joins the macromolecular assembly, now termed complex B. After ATP hydrolysis dependent release of U4 snRNP the spliceosome has achieved its splicing compatible conformation (complex C) and the actual splicing reaction is carried out (Figure 1, right panel).

Genetic and biochemical experiments have revealed an intricate network of interactions between pre-mRNA, snRNAs and protein factors which undergoes an extensive rearrangement during the course of the splicing reaction. In the spliceosome the base pairing between U4 and U6 snRNAs is unwound and the U6 snRNA subsequently base pairs with both U2 snRNA and the 5' splice site. A highly conserved loop in the U5 snRNA interacts with the exon sequences at the 5' and 3' splice sites and these interactions are important for the second trans-esterification step (Staley and Guthrie, 1998). Thus, nuclear pre-mRNA splicing is a highly dynamic process with protein and RNA components playing important regulatory roles in the assembly of snRNPs and the rearrangement of the complex network of RNA-RNA interactions.

1.1.2. Relevance of this Work to Splicing and U snRNP Biogenesis

Work described in this thesis involves determination of two novel three-dimensional structures. The solution structure of the Tudor domain of the Survival of Motor Neuron protein (SMN), a protein essential in snRNP core domain assembly (Chapter 1.2.1) and the protein-peptide complex of human U2AF^{65} and human SF1 (Chapter 2.2.1). This early macromolecular assembly is important for correct 3' splice site recognition in the 'commitment' or E complex (see Chapter 2.1.1). Thus, both structures provide important functional insight into essential mechanisms during snRNP biogenesis and pre-mRNA splicing.

1.1.3. U snRNP Biogenesis and Assembly

U snRNPs contain two classes of proteins: those specific to a given snRNP and those who are common to the U1, U2, U4/U6 and U5 snRNPs. The latter are called core or Sm proteins and they assemble on snRNAs into a globular structure called the core snRNP domain. The Sm-protein binding site (or Sm-site) on snRNAs is a short, conserved uridine-rich sequence in U1, U2, U4 and U5 snRNAs. Eight generic Sm proteins have been identified in snRNPs (Luhrmann et al., 1990). They are named, in order of decreasing size, B'/B, D_3, D_2, D_1, E, F and G. The B and B' proteins arise from a single gene by alternative splicing and differ only in 11 residues at their C termini (Chu and Elkon, 1991; van Dam et al., 1989). These Sm proteins contain conserved sequence motifs in two segments, Sm1 and Sm2, which are connected by a linker of variable length (Hermann et al., 1995). The Sm motif is related to no known protein sequence and hence, these proteins form a distinct protein family (Seraphin, 1995).

Core domain assembly is marked by several distinct intermediates. In the absence of snRNA, Sm proteins exist as three subcomplexes, D_1D_2, D_3B (or D_3B') and EFG. The EFG complex binds, together with the D_1D_2 subcomplex, to U snRNA to form a stable subcore, which is then joined by the D_3B (or D_3B') heterodimer to complete core domain assembly (Raker et al., 1996). Neither the individual Sm proteins nor individual Sm subcomplexes alone bind to snRNA. Recently, the crystal structures of two Sm protein subcomplexes, D_1D_2 and D_3B have been solved (Kambach et al., 1999b). The four Sm proteins show a common fold

containing a short, N-terminal α-helix followed by a five-stranded, anti-parallel β-sheet (Figure 2). Strands 1-3 of the β-sheet are made of residues within the Sm1 motif, the linker of variable length between the two motifs forms a connecting loop and Sm2 motif residues constitute β-strands 4 and 5. Strands 2, 3, and 4 are heavily bent and strand 5 loops back to interact with strand 1. The main interaction interface in both complexes comprises β-strand 4 of one partner (D_2 or B) pairing with β-strand 5 of the other (D_1 or D_3, respectively), thereby continuing the β-sheet throughout the complex (Figure 2). Thus, Sm D_1D_2 and D_3B reveal a high degree of structural similarity at the level of both the individual protein fold and the dimer architecture.

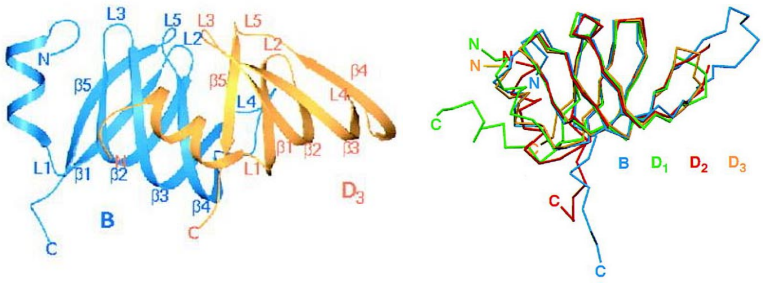


Figure 2:
Ribbon representation of the Sm B and Sm D_3 heterodimer. Secondary structure elements are indicated and presence of the conserved Sm fold is clearly visible in both domains. A superposition of the backbone trace of Sm B, D_1, D_2 and D_3 is outlined at the right. The Sm folds of these four snRNP proteins are super imposable to an overall rmsd of 0.3 Å. Reproduced from Kambach et al., 1999b

A model of a higher order structure could be built by consecutively adding monomers one by one, using identical subunit interactions (Kambach et al., 1999b). Such a model suggests that the seven core snRNP proteins could assemble in a doughnut-shaped, ring-like structure (Figure 3).

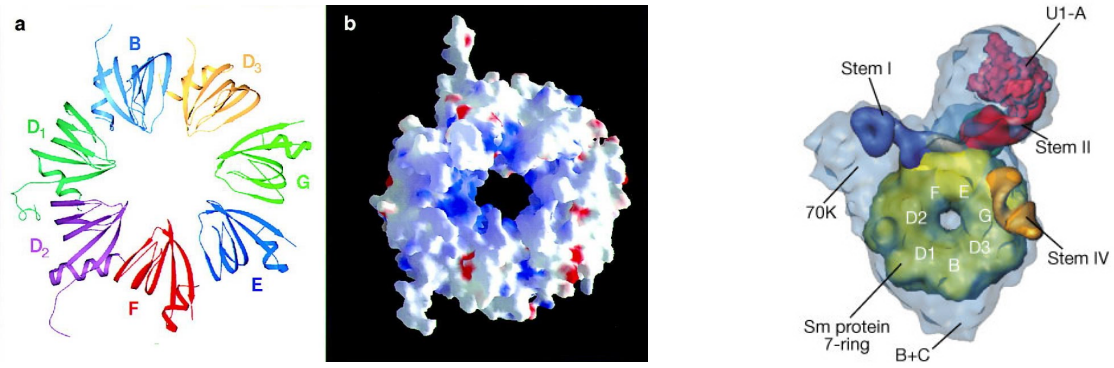


Figure 3:
a. Consecutive alignment of Sm folds in a repeating β4/β5 interactive fashion between different subunits results in a heptameric ring model for fully assembled Sm proteins. The proposed electrostatic surface of such a model-assembly is depicted in b. together with an experimental electron density envelope of the complete U1 snRNP (at the right). Reproduced from (Kambach et al., 1999b and Stark et al., 2001)

The heptameric circular arrangement represents a core domain model that is in agreement with all biochemical and genetic data currently available. Additionally, cryo electron-microscopy (cryoEM) data of different snRNPs (Kastner et al., 1990a; Kastner et al., 1990b) indicates an overall structural assembly which correlates well with the proposed heptameric model. Recent experimental EM evidence of the arrangement of RNA and proteins in the spliceosomal U1 snRNP particle confirms the presence of the 7-membered ring-like structure (Stark et al., 2001) (Figure 3, right panel).

SnRNA is thought to bind to or thread through the central, positively charged, hole of the doughnut, which is 20Å in diameter and could potentially accommodate single stranded RNA (Kambach et al., 1999a). Such a mode of RNA binding is in agreement with recent interaction mapping studies of snRNA contact sites on the Sm fold of Sm B and G (Urlaub et al., 2001), Direct structural evidence for this mode of RNA coordination by Sm proteins is provided by a recent paper from the Suck laboratory (Toro et al., 2001). They report the crystal structure of the *Archaeoglobus fulgidus* AF-Sm1 protein assembly in complex with an uridine (U_5) oligonucleotide (Figure 4). The reported structure shows that the nucleotides are coordinated by residues lined along the inner rim of the central hole in the heptameric AF-Sm1 protein complex (Figure 4).

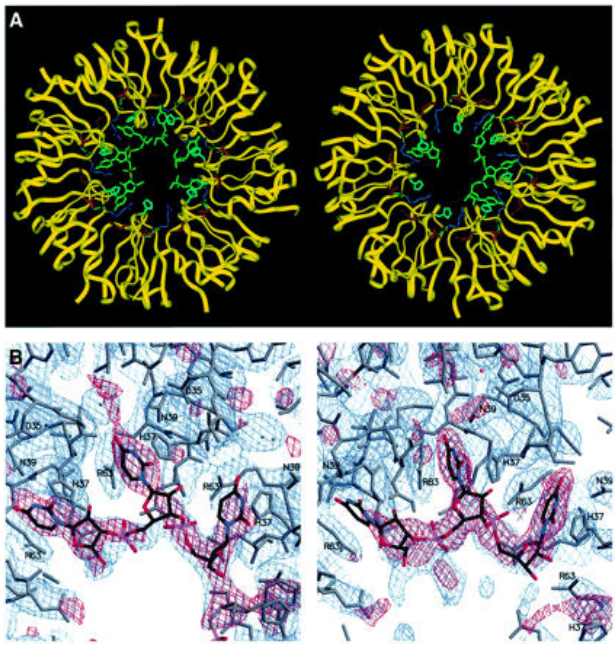


Figure 4:

Top panel: The two AF-Sm1 heptamers in the asymmetric unit are shown in a yellow ribbon representation with the bound oligonucleotides shown as green sticks.
Bottom panel: Electron density maps for the AF-Sm1 protein (blue) and for the bound U_5 oligonucleotide (red). Reproduced from Toro et al., 2001

Core domain formation is an essential step in U snRNP biogenesis and critically depends on the function of a protein termed SMN (for Survival of Motor Neuron) (Buhler et al., 1999) (Chapter 1.1.4). In humans, defects in SMN function lead to a genetic disease called Spinal Muscular Atrophy or SMA (Sendtner, 2001), the most common hereditary cause for infant mortality (Chapter 1.1.6). Core domain biogenesis occurs in the cytoplasm after nuclear export of newly transcribed U snRNAs, containing the N^7-monomethylguanosine (m^7G) cap (Mattaj and De Robertis, 1985) and triggers hypermethylation of the m^7G cap to a 2,2,7-trimethylguanosine (m^3G) cap structure. The core domain and the m^3G cap act as a bipartite nuclear import signal and pre-snRNPs mature in the nucleus by association with specific proteins (Fischer et al., 1991; Palacios et al., 1997).

1.1.4. SMN and U snRNP Biogenesis

Pre-mRNA splicing requires the action of snRNPs, RNA-protein complexes containing the U1, U2, U4, U5 and U6 snRNAs, Sm proteins B/B', D_{1-3}, EFG and a set of more specific proteins. Ultimately, the individual mature snRNPs organize into spliceosomes. Biogenesis of functionally intact snRNP particles critically depends on the function of a protein complex that assembles around the Survival of Motor Neuron (SMN) protein (Buhler et al., 1999; Fischer et al., 1997; Meister et al., 2001a). Proteins in this complex include SMN, the core Sm proteins, Sm-like proteins (LSms) (Buhler et al., 1999; Friesen and Dreyfuss, 2000), the snoRNP proteins GAR1 and fibrillarin (Pellizzoni et al., 2001a), RNA helicase A (Pellizzoni et al., 2001b), the hnRNP proteins hnRNP Q (Mourelatos et al., 2001), hnRNP U and hnRNP R (Rossoll et al., 2002) and proteins collectively referred to as Gemins. Gemin2/SIP1 (Fischer et al., 1997; Liu et al., 1997) Gemin3/dp103 (Charroux et al., 1999),Gemin4 (Charroux et al., 2000) Gemin5 (Gubitz et al., 2002), Gemin6 (Pellizzoni et al., 2002) and recently Gemin7 (Baccon et al., 2002) have been shown to interact either directly or indirectly with SMN in the SMN-protein complex. Additionally, direct and sequence-specific SMN complex interactions to U1, U4 and U5 snRNAs have been reported recently (Yong et al., 2002). Functional impairment of SMN to interact with components of this macromolecular assembly results in disruption of snRNP biogenesis and consequently loss of pre-mRNA

splicing (Pellizzoni et al., 1998). Thus, SMN exerts a critical and essential role in the assembly pathway of functionally intact snRNPs.

1.1.5. SMN Protein Organization, Interaction and Localization

SMN is a 294 amino acid residue protein which contains a central Tudor domain (aa 91-158) as the only readily identified sequence motif (Ponting, 1997) (see Figure 5 for domain overview). The Tudor domain is preceded by a short stretch of conserved positively charged amino acids (~aa 65-80) which have been implicated in nucleic acid binding (Bertrandy et al., 1999; Lorson and Androphy, 1998). However, no other confirmation of these findings nor the identification of a specific DNA or RNA target has been reported since then. Amino acids N-terminal to this hypothetical DNA/RNA binding site (aa 13-44) are shown to directly interact with SIP1 (SMN interacting protein 1), now termed Gemin2 (Fischer et al., 1997; Liu et al., 1997). Gemin3 binds to a SMN region following the Tudor domain C-terminally (Campbell et al., 2000), but also requires the presence of residues encoded by exon 7 for interaction (Charroux et al., 1999). Interestingly, Gemin3/dp103 contains a putative RNA helicase motif and thus points to a role of how the SMN complex could assist in the assembly of spliceosomal complexes. Residues 200-240 in SMN are proline-rich and interact with the profilins PFN1 and PFN2 (Giesemann et al., 1999), although the physiological relevance of these interactions is not clear. The C-terminal 50 residues in SMN (aa 251-294) are essential for self-oligomerisation (Lorson et al., 1998). These homo-aggregates confer protein stability and protect the SMN protein from degradation.

SMN is encoded by two genes, the telomeric *smn1*- and the centromeric *smn2*-gene [Monani, 1999 #30]. These two genes are virtually identical. A single-nucleotide difference in the *smn2* gene leads to preferential alternative splicing and exon 7 skipping of the respective *smn2* mRNA [Monani, 1999 #30][Lorson, 1999 #29]. Full-length SMN only comprises ~10% of the gene product produced from the *smn2* gene, whereas the mainly produced shorter protein (SMN2) lacks the important C-terminal protein-stability conferring oligomerisation domain and is rapidly degraded. No detectable levels of the shorter SMN2 version of the SMN protein are found in any cell type [Lefebvre, 1998 #14].

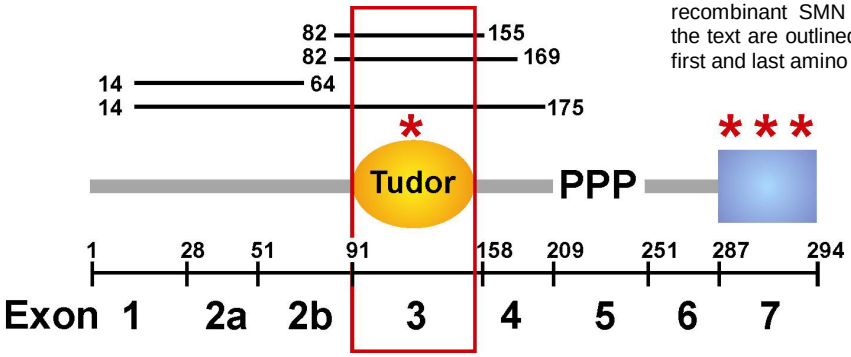


Figure 5:
Schematic organization of the SMN gene/protein. The central Tudor domain, the proline-rich sequence stretch and the C-terminal oligomerisation domain (corresponding to Exon 7) are indicated. Red stars denote the positions of reported SMA causing point mutations. Portions of recombinant SMN protein constructs discussed in the text are outlined above and residue numbers for first and last amino acids are given.

The Tudor domain has been initially identified as a conserved sequence motif in proteins that function in RNA transport and localization during early embryonic development (Ponting, 1997). A more general role for Tudor domains as a protein-protein interaction motif during RNA related metabolic events has been deducted since then. A multiple sequence alignment of Tudor domain containing proteins is shown in Figure 6.

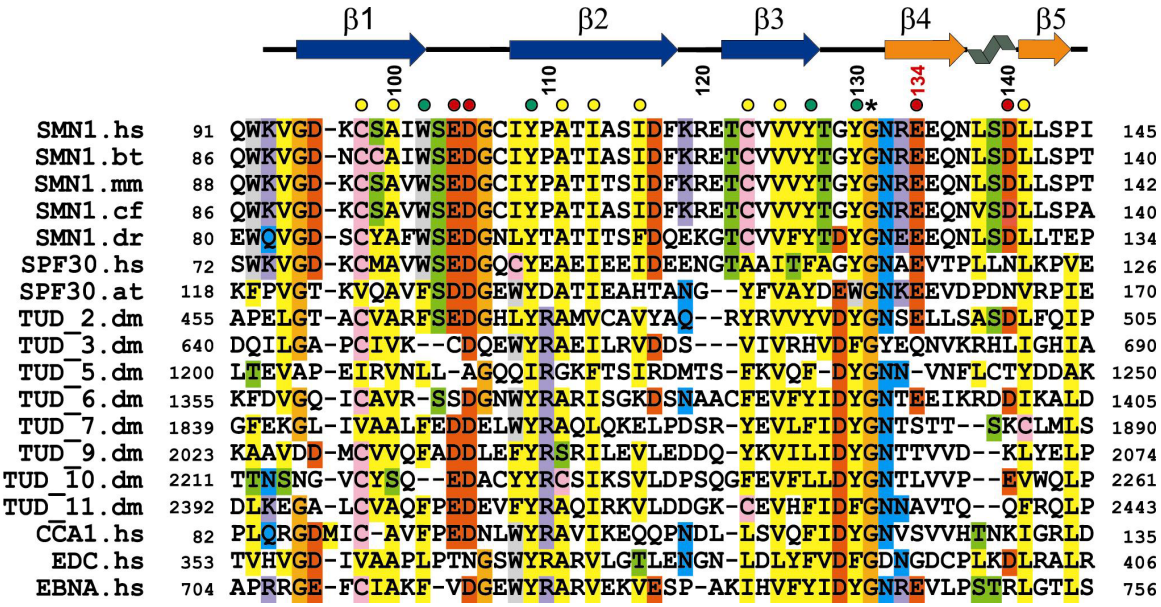


Figure 6:

Multiple sequence alignment of SMN orthologous and other Tudor domain containing proteins. Conserved hydrophobic, aromatic and negatively charged amino acids are marked by yellow, green and red circles, respectively. Secondary structure elements defined by the NMR structure of the Tudor domain of SMN are indicated above the alignment.

The Tudor domain of SMN binds to the C-terminal RG-rich tails of the Sm proteins D_1, D_2, D_3, B and E (Buhler et al., 1999; Friesen and Dreyfuss, 2000; Selenko et al., 2001). Furthermore, the Tudor domain of SMN interacts with RG-rich regions of various other target proteins including coilin, fibrillarin and GAR1 (Fischer et al., 1997; Hebert et al., 2001; Jones et al., 2001). These RG domains are subjected to post-translational modifications and are methylated in the cytoplasm at arginine residues to form symmetric dimethylarginines, sDMA (Brahms et al., 2001; Brahms et al., 2000). Interestingly, these modifications are selectively recognized by the Tudor domain of SMN and bound with higher affinity than unmodified RG peptides (Friesen et al., 2001) (and this work).

The SMN protein is found in both the cytoplasm and the nucleus of most tissues. The abundance of expression varies among different tissues, with the greatest levels of expression in the brain, spinal cord, and muscle, and the lowest relative level of expression occurring in lymphocytes and fibroblasts (Coovert et al., 1997; Lefebvre et al., 1997). In the nucleus, SMN is concentrated in aggregates called ‘gemini of coiled bodies’ or ‘gems’, so named because they are found in close association with coiled bodies (Liu and Dreyfuss, 1996). Certain tissues, for example cardiac muscle and smooth muscle, do not have either coiled bodies or gems, indicating that nuclear bodies may be storage sites for reserves of essential proteins and snRNPs (Young et al., 2000).

1.1.6. SMN and Spinal Muscular Atrophy (SMA)

Proximal spinal muscular atrophies (SMAs) are a group of inherited neuromuscular disorders characterized by the degeneration of spinal motor neurons leading to muscular paralysis with muscular atrophy. They form the second most common fatal autosomal recessive disease after cystic fibrosis, with an incident of 1 in 10000 newborns (Pearn, 1978; Pearn and Wilson, 1973; Roberts et al., 1970).

The clinical SMA phenotype exhibits a broad spectrum of manifestations ranging from severe infantile to mild chronic forms of the disease. SMA is subdivided into three types. Type 1 SMA is the severe form of Werdnig-Hoffman disease with onset at birth or before 6 months of age and death of respiratory distress usually

within 2 years. Type 1 SMA patients are unable to sit or walk due to profound muscular weakness. Children with Type 2 SMA (intermediate form) can sit but cannot stand or walk unaided. Type 3 SMA (Kugelberg-Welander disease) patients show first clinical signs 18 months after birth and evolve a chronic course of the disease (Schmalbruch and Haase, 2001). The pathological hallmark of SMA is the loss of motor neurons in the anterior horn of the spinal cord and often in the brainstem. Irrespective of the clinical severity, weakness of proximal muscles and the absence of or a marked decrease in deep reflexes are observed (see Figure 7)

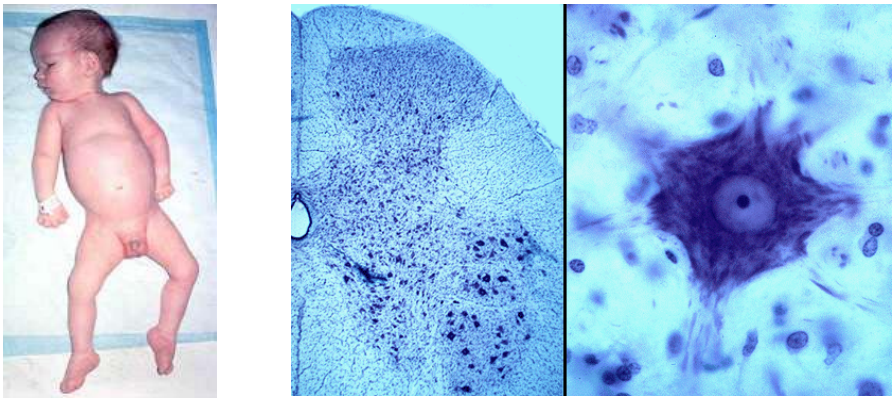


Figure 7:
Left: an infant SMA Type1 patient. This individual is unable to flex or move its extremities or control the position of its head. *Right:* section through a healthy spinal chord tissue, Stained in blue and shown enlarged are neuronal cells of the anterior horn. These cell type protrudes from the spinal chord and forms contact to surrounding muscle tissue. SMA affected individuals are physiologically characterized by sever loss of this neuronal cell type and inability to transmit nerve pulses to the otherwise healthy muscle tissue.

All forms of SMA are caused by mutation in the telomeric *smn1* gene, which is usually lost by deletion or gene conversion. 98% of all SMA patients have no detectable *smn1* gene The remaining 2% of cases retain a copy of the *smn1* gene, which is corrupted by deletions, insertions or point mutations. Out of these, C-terminal deletions or premature reading frame terminations are most common. These mutations result in protein products which are, like the SMN2 gene product (see above), unstable and rapidly degraded from cells. Thus, overall SMN protein levels are severely decreased in most cases of SMA and often the only functional version of full-length of SMN stems from the 10% of total protein encoded by *smn2*. In fact, protein levels of full-length SMN can be directly correlated to severeness of the Spinal Muscular Atrophy phenotype (Lefebvre et al., 1997; Lorson et al., 1998).

Only a few cases of SMA patients harboring point mutations within the *smn1* gene have been identified and again, for mutations affecting C-terminal residues important for oligomerisation, a direct correlation between ability for self-assembly, levels of endogenous full-length SMN protein concentrations and severeness of SMA could be deducted (Lefebvre et al., 1997).

1.1.7. The E134K Mutation of the Tudor Domain and Implication for SMA

A patient carrying a single point mutation (E134K) within the Tudor domain of the SMN protein has been identified . This individual suffers from a severe Type 1 form of SMA albeit protein levels of full-length SMN appear normal. This case is believed to represent a rare example of a 'functional' mutant (whereas most other cases are 'null' or 'knockout' mutants due to their extremely low levels of full-length protein). Studying the molecular and cellular functions affected by this point mutation should yield novel insight into the biological mechanisms underlying SMA. It has been shown that E134K within the Tudor domain severely reduces SMNs ability to interact with Sm proteins and thereby abolishes assembly of U snRNPs *in vivo* (Buhler et al., 1999). This seems, in part, to be a common molecular basis underlying SMA as other SMA causing mutations have also been shown to disrupt the critical SMN/Sm interactions and thereby prevent snRNP assembly and consequently pre-mRNA splicing (Pellizzoni et al., 1999; Pellizzoni et al., 1998). The question as to why the impairment of such a fundamental biological process allows proceeding through embryonic development and only manifests itself in such a narrow section of a particular neuronal cell type within the spinal chord (SMA, most prominently, leads to a great reduction in anterior horn cell number, see above), remains enigmatic.

1.2. RESULTS and DISCUSSION

1.2.1. The 3D Structure of the human SMN Tudor Domain

To reveal novel insight into the functional role of the human SMN Tudor domain and to yield three-dimensional information of this, by then, uncharacterized structural motif, we determined the solution structure of the Tudor domain of SMN by heteronuclear multidimensional NMR spectroscopy.

Protein Constructs: A number of recombinant protein constructs were initially used for NMR analysis (see Figure 5). One comprised residues 14-175 and included most of the region N-terminal to the Tudor domain, plus a short C-terminal stretch corresponding to a fragment of exon 4. Most residues of this protein construct appeared poorly dispersed in ^{1}H, ^{15}N correlation NMR spectra, indicating that a large region is unfolded. A similar unfolded conformation was observed for a recombinant protein construct comprising residues 14-64 of human SMN. Shorter constructs, with residues 83 to 169 and 83 to 155, displayed an identical subset of folded residues as observed for the longer version of SMN, but a smaller degree of the unfolded state. Only residues 92 to 144, which correspond to most of exon 3 in the SMN gene, adopt a well-defined tertiary structure (see below). Amino acids N- and C-terminal to these residues are disordered as additionally indicated by the paucity of nuclear Overhauser effects (NOEs) and small heteronuclear $\{^{1}H\}$-^{15}N-NOE values. This indicates that in a protein construct comprising residues 14-175 of human SMN, the Tudor domain (aa 92-144) is, in the absence of a ligand, the only folded fragment within such a polypeptide.

Nucleic Acid Binding: We tested the longer construct (aa 14-175) for its nucleic acid binding ability by adding excess of an unlabeled poly-G RNA 10-mer to a ^{15}N labeled protein sample. This ribonucleotide was reported to strongly interact with a region in human SNM corresponding to amino acid residues 1-76 (Bertrandy et al., 1999). No changes in chemical shifts, indicating changes in chemical enviroment for residues involved in RNA binding, could be observed for any SMN protein residue in the NMR spectra. We were thus unable to reproduce RNA binding activity of this fragment of human SMN.

NMR Methods: The three-dimensional structure of the human SMN Tudor domain was solved by a combination of NMR experiments typically employed for structure determination of proteins (Chapter 3.2). Hydrogen-bond restraints were derived from H_2O/D_2O exchange experiments and characteristic ^{15}N-edited NOESY derived sequential NOEs (Chapter 3.5.1). Dihedral-angle restraints were used based on experimentally obtained ^{3}J*-coupling* values from an HNHA-*J* NMR experiment (Chapter 3.5). Distance restraints for structure calculations were derived from three-dimensional ^{15}N- and ^{13}C-edited NOE spectra. The structure

was calculated using a mixed torsion and Cartesian angles dynamics simulated annealing protocol with the programs CNS and ARIA. The structure is well-defined by the NMR data which provide more than 28 restraints per residue. The overall root mean square deviation (rmsd) for back-bone N, Cα and C' atoms is 0.42Å and 0.88Å for all heavy atoms. A more detailed summary of structural statistics is given in Table 1. An ensemble of the 20 lowest energy NMR structures and a ribbon representation of the human SMN Tudor domain, are shown in Figure 8.

Table 1 Structural statistics for the SMN Tudor domain

		<SA>[1]
Experimental restraints		
	Number of restraints	R.m.s. deviations
Distance restraints (Å)[2]		
Unambiguous	1,402	0.0045 ± 0.0016
Hydrogen bonds	50	0.005 ± 0.002
Dihedral angle restraints (°)[3]	17 ϕ, 6 χ_1	0.09 ± 0.10
Residual dipolar coupling restraints (Hz)		
H–N	44	0.56 ± 0.12
Coordinate precision (Å; residues 92–144)[4]		
N, Cα, C′		0.42 ± 0.09
All heavy atoms		0.88 ± 0.09
Structural quality		
$E_{L.-J.}$[5] (kcal mol^{-1})		-481 ± 6
Ramachandran plot (%)		
Most favored region		80.0 ± 4.1
Additionally allowed region		19.8 ± 4.1

[1]<SA> is the ensemble of the 20 lowest energy structures out of 100 calculated. Root mean square (r.m.s.) deviations for bond lengths, bond angles and improper dihedral angles were 0.00092 ± 0.00004 Å, 0.240 ± 0.003° and 0.119 ± 0.004°, respectively.
[2]No distance restraint was violated by >0.3 Å in any of the final structures.
[3]No dihedral angle restraint was violated by >2.1°.
[4]Coordinate precision is given as the Cartesian coordinate r.m.s. deviation of the 20 lowest energy structures with respect to their mean structure.
[5]$E_{L.-J.}$ is the Lennard-Jones van der Waals energy calculated using the CHARMM PARMALLH6 parameters. $E_{L.-J.}$ was not included in the target function during the structure calculations.

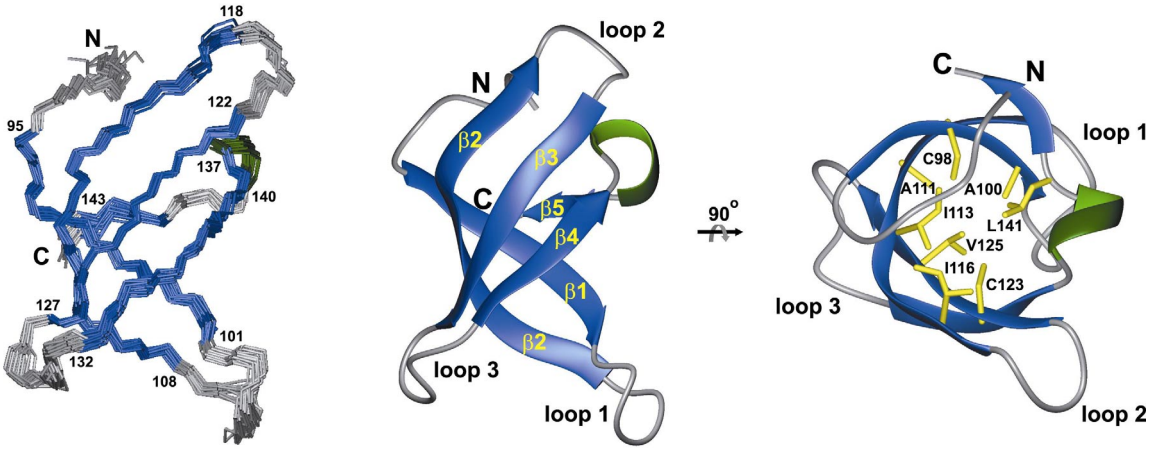


Figure 8:
Ensemble and ribbon representation of the Tudor domain of SMN. Secondary structure elements are numbered and indicated. The perpendicular view at the far right shows conserved residues making up the hydrophobic core of the β-barrell.

1.2.2. Description of the Tudor Domain Structure

The three-dimensional structure of the SMN Tudor domain adopts a strongly bent antiparallel β-sheet. Five strands, β1 to β5, form a barrel-like fold which is lined at the bottom by a long curved strand β2 and closed by an antiparallel interaction between β1 and Leu142 within the short strand β5. Strands β1 to β4 are connected by short turns, while strand β4 and β5 are linked by a helical turn which leads to a ~90° angle between the direction of these two strands. Conserved hydrophobic residues, Cys98, Ala100, Ala111, Ile113, Ile116, Cys123, Val125 and Leu141, stabilize the structure through formation of a hydrophobic core. Based on the conservation of these structurally important residues, a similar three-dimensional fold can be expected for other Tudor domains. In addition, conserved aromatic residues, Trp102 (Loop 1), Tyr109 (β2), Tyr127 (β3) and Tyr130 (Loop 3) form a cluster of hydrophobic side chains between loops 1 and 3. Therefore, both loops adopt a well-defined structure and are not disordered (Figure 8). The electrostatic surface representation of the Tudor domain exhibits a hydrophobic patch in this region which may be involved in ligand interactions. Next to this hydrophobic surface, a number of negatively charged amino acids are located in Loop 1 (Glu104, Asp105), β4 (Glu134) and the helical turn connecting β4 and β5 (Asp140). These residues are conserved in all SMN homologues and in many

other Tudor domain containing proteins (Figure 6) leading to an overall negatively charged surface (Figure 9). These unique structural features suggest that the Tudor domain is more likely to represent a protein interaction domain than to directly bind to RNA or DNA (Bertrandy et al., 1999; Lorson and Androphy, 1998).

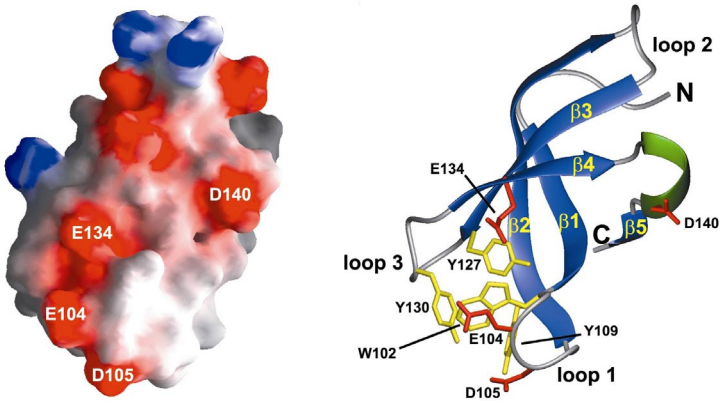


Figure 9:
Electrostatic surface representation of the Tudor domain of SMN. Basic to acidic surfaces areas are colored from blue to red, respectively. A prominent stretch of negatively charged amino acids, comprising residue Glu134, lining one side of the Tudor surface is shown. The ribbon presentation at the right is shown in the same orientation as the space filled molecule. Side chains of residues forming the hydrophobic, aromatic cluster between loop1 and loop3 are shown in yellow and acidic amino acids responsible for the overall negative charge (including E134), in red.

1.2.3. Structure of the E134K Mutant Tudor Domain

We next focused on the reported spinal muscular atrophy (SMA) causing point mutation (E134K) within the Tudor domain of human SMN and investigated the effect of such a point mutation on the overall Tudor domain structure. It is possible that this mutation would lead to disruption of the structural integrity of the Tudor domain and thereby destroy a potential interaction module and lead to SMA. We cloned and expressed an identical version of the Tudor domain containing protein construct (aa 83-155), bearing the single amino acid exchanged Lys134 instead of the wild type Glu134. Comparison of $^{1}H,^{15}N$ correlation spectra of wildtype and mutant proteins revealed only minor chemical shift changes for residues in close spatial proximity to Glu134 (Figure 10). In addition, $\{^{1}H\}$-^{15}N NOE experiments confirmed that the E134K mutant Tudor domain remains structured and also exclude the possibility that strand β4 becomes locally unfolded upon amino acid substitution (Figure 10). We thus proved, that the SMA causing single point

mutation (E134K) does not affect the overall structural integrity of the SMN Tudor domain.

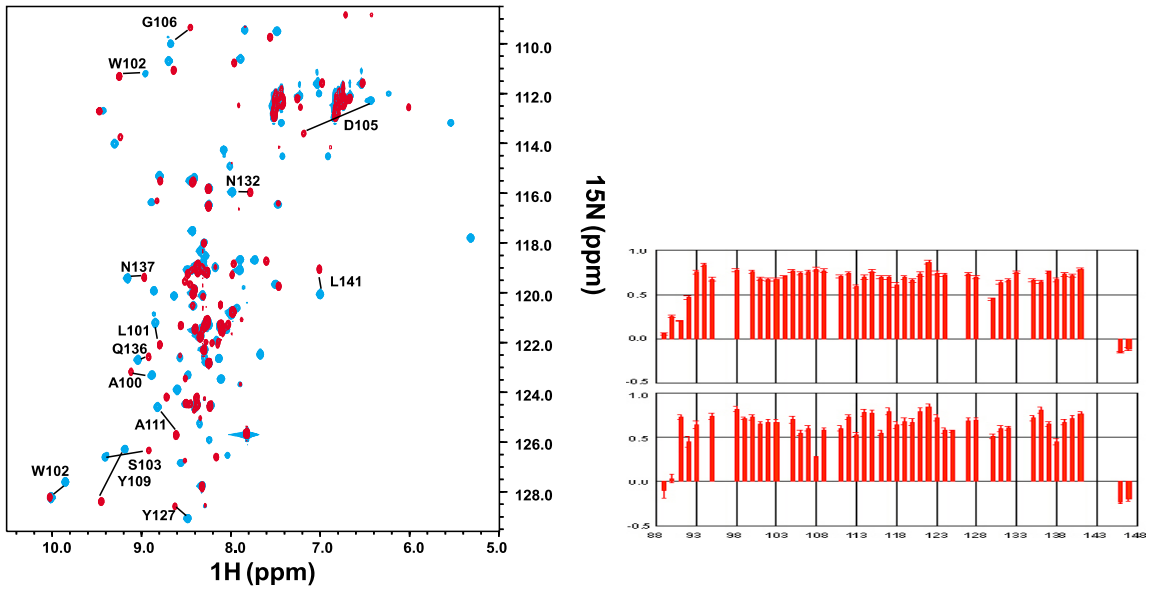


Figure 10:
Overlay of ^{15}N-HSQC correlation spectra of wild-type (red) and the mutant E134K Tudor domain (blue). Largest differences for residues in close spatial proximity to the site of mutation are indicated. Hetero-nuclear NOE measurements (above) for wild-type (top) and mutant (bottom) Tudor domain display similar characteristics throughout the Tudor domain sequence and thus indicate no local unfolding in the mutant protein.

1.2.4. Structure of the Tudor Domain and the Sm Protein Fold

We have compared the newly determined Tudor domain structure with known three-dimensional structures in the Brookhaven Protein Data Base (PDB) employing a search routine used by the DALI package (Holm and Sander, 1995). Surprisingly, we found that the three-dimensional structure of the SMN Tudor domain resembles the fold of the Sm core proteins even though the amino acid sequences do not share any detectable similarity (Kambach et al., 1999b) (Figure 11). Compared to the Sm fold, the Tudor domain lacks an N-terminal helix, and strands β3, β4 and β5 are shorter. Therefore, β3 and β4 are much less curved than the corresponding strands in the Sm fold. However, the lengths of β3 and β4 are also variable within the Sm protein family (Kambach et al., 1999b) and the Tudor domain structure can thus be considered a truncated Sm protein fold. Sm proteins exist as heteromeric complexes in solution (Brahms et al., 2000; Raker et al., 1996). The binding interface in the Sm D_3B and Sm D_1D_2 heterodimers consists of

an antiparallel β-sheet formed between strand β4 of one monomer and strand β5 of the neighboring Sm protein (Kambach et al., 1999b). This interface is further stabilized by hydrophobic and electrostatic interactions. The known biochemical interaction of the SMN protein with different members of the Sm and LSm protein family (Buhler et al., 1999; Friesen and Dreyfuss, 2000; Pellizzoni et al., 1999) and the intriguing structural similarity between the Tudor domain and the Sm fold suggests that the Tudor domain may form a similar intermolecular β4/β5 interface with a cognate Sm protein. This could point to a direct involvement of the SMN protein in the snRNP core domain assembly and would suggest a chaperone-like, regulatory function for the Tudor domain in core domain biogenesis (MacKenzie and Gendron, 2001).

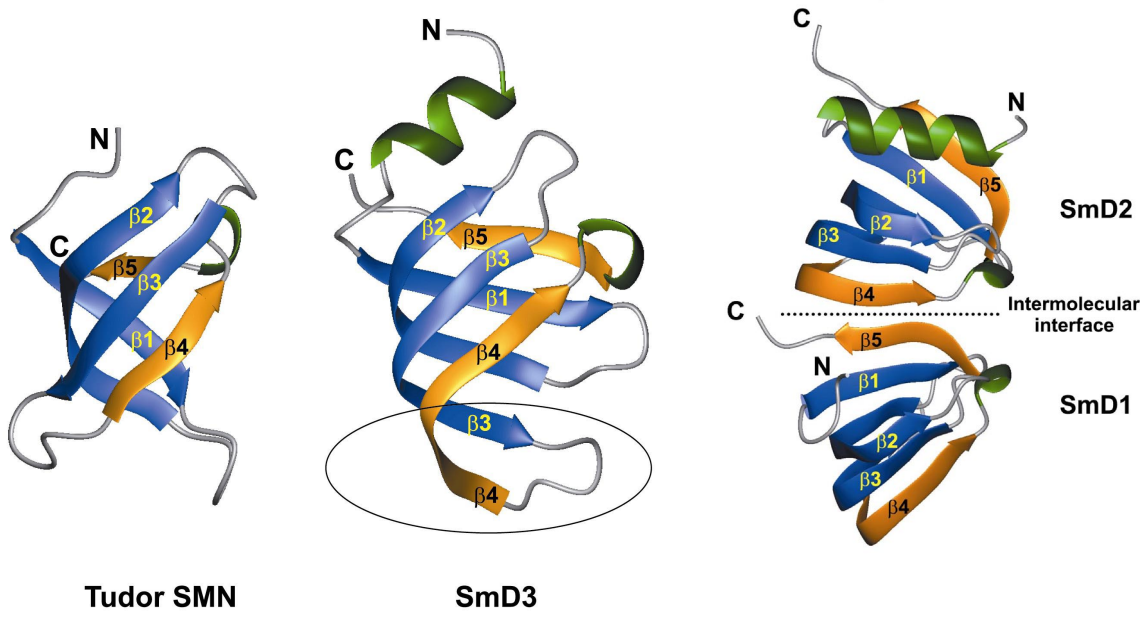


Figure 11:
Ribbon representation of the similarities of the Tudor domain and the Sm fold. Secondary structure elements making up the typical Sm hetero-dimer interface (β4/β5) are shown in gold. The circle around β3/β4 within the SmD_3 fold denotes the elongation of this structural region compared to the Tudor domain, leading to an additional, kinked extension in SmD_3.

1.2.5. Biochemical Characterization of the SMN/Sm Protein Interaction

Based on the reported biochemical interaction of human SMN and members of the Sm and LSm protein family (Buhler et al., 1999), we analyzed, in collaboration with Utz Fischers' group at the Max Planck Institute (MPI) in Martinsried, the molecular basis for this interaction. Sm proteins are characterized by a common globular N-terminal domain, which comprises the typical Sm sequence motif (Hermann et al., 1995) and makes the characteristic Sm fold (Kambach et al., 1999b).This global fold is followed by an extended C-terminal region, bearing a degenerate repetition of arginine and glycine residues (Chapter 1.2.7). These carboxy-terminal regions are conserved to different extents within SmD_1, D_2 and D_3 (Hermann et al., 1995) (see also Figure 15). It was shown recently that SMN binds to the arginine- and glycine-rich carboxy-terminal tails of the Sm proteins D_1 and D_3 (Friesen and Dreyfuss, 2000). In order to investigate whether this interaction could involve the SMN Tudor domain, we performed *in vitro* binding assays (see materials and methods). Tails of Sm D_1 and Sm D_3 were expressed as GST-fusion proteins and used in pull-down experiments with *in vitro* translated full-length SMN, the Tudor domain or the Tudor domain harboring the E134K mutation (see materials and methods). As shown in Figure 12, full-length SMN and the Tudor domain efficiently bound to either tail whereas no binding of the mutated Tudor domain (E134K) was observed. Conversely, binding to recombinant GST-Tudor domain and SMN could be observed for *in vitro* translated Sm D_1 but not for a truncated version lacking the C-terminal tail region (Figure 13). Again, this interaction was abolished by the E134K mutation. Similar results were obtained using Sm D_3. Thus, the Tudor domain of human SMN binds to the tails of Sm D_1 and D_3.

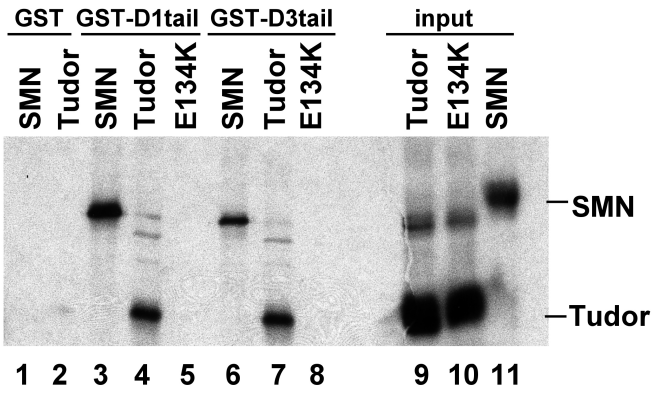


Figure 12:
Biochemical pull-down experiments using GST-tagged versions of SmD_1 and D_3 carboxy-terminal tail regions and *in vitro* translated full length SMN and wild type or mutant (E134K) Tudor domain.

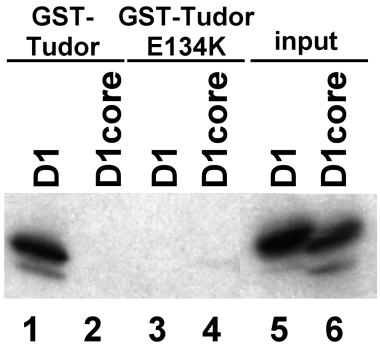


Figure 13:
Reciprocal biochemical pull-down experiments using GST-tagged versions of wild-type and mutant (E134K) Tudor-domain and in *vitro translated* full-length D_1 and the globular N-terminal Sm-domain of D_1 (D1core).

1.2.6. NMR of Wild-type and Mutant Tudor/Sm Interactions

Since SMN binding to Sm proteins is essential for its function, we wanted to further characterize the role of the Tudor domain for this interaction. To localize the binding surface on the Tudor domain structure we performed NMR titrations with a 23-mer peptide, $(GR)_9GGPRR$, derived from the C-terminal tail of Sm D_1. Heteronuclear correlation experiments were recorded on a ^{15}N-labeled Tudor domain to monitor ^{1}H and ^{15}N chemical shift changes upon addition of the peptide (Figure 14). The exchange between free and bound protein conformations is fast on the NMR time scale indicative of a micromolar dissociation constant. The spectral changes show that the Sm D_1 tail binds to a region consisting of loops 1 and 3 and neighboring parts of strands β1, β2, β3 and β4. This surface involves the conserved negatively charged and aromatic residues of the Tudor domain. Thus, a region of the Tudor domain comprising Glu134 is part of the binding site that recognizes the C-terminal tail of Sm D_1. Furthermore, we tested binding of the same SmD_1 peptide to the mutant (E134K) from of the Tudor domain, in an identical experimental set-up. No binding of the carboxy-terminal D_1 tail peptide was observed to the mutant Tudor domain. A 10-fold excess of peptide was necessary to induce slight differences in chemical shift of Tudor domain residues, clearly indicating that the single charge reverting point mutation within the Tudor domain of human SMN greatly reduces its ability to interact with the Sm tail region. We have thus shown that the presence of a conserved negatively charged residues (Glu134), located in the center of a electrostatically negative surface area, is essential and required for interaction with the positively charged carboxy-

terminal tails of Sm proteins. An interaction which appears to critically depend on favorable, reciprocal charge distributions of the Tudor domain and its Sm ligands.

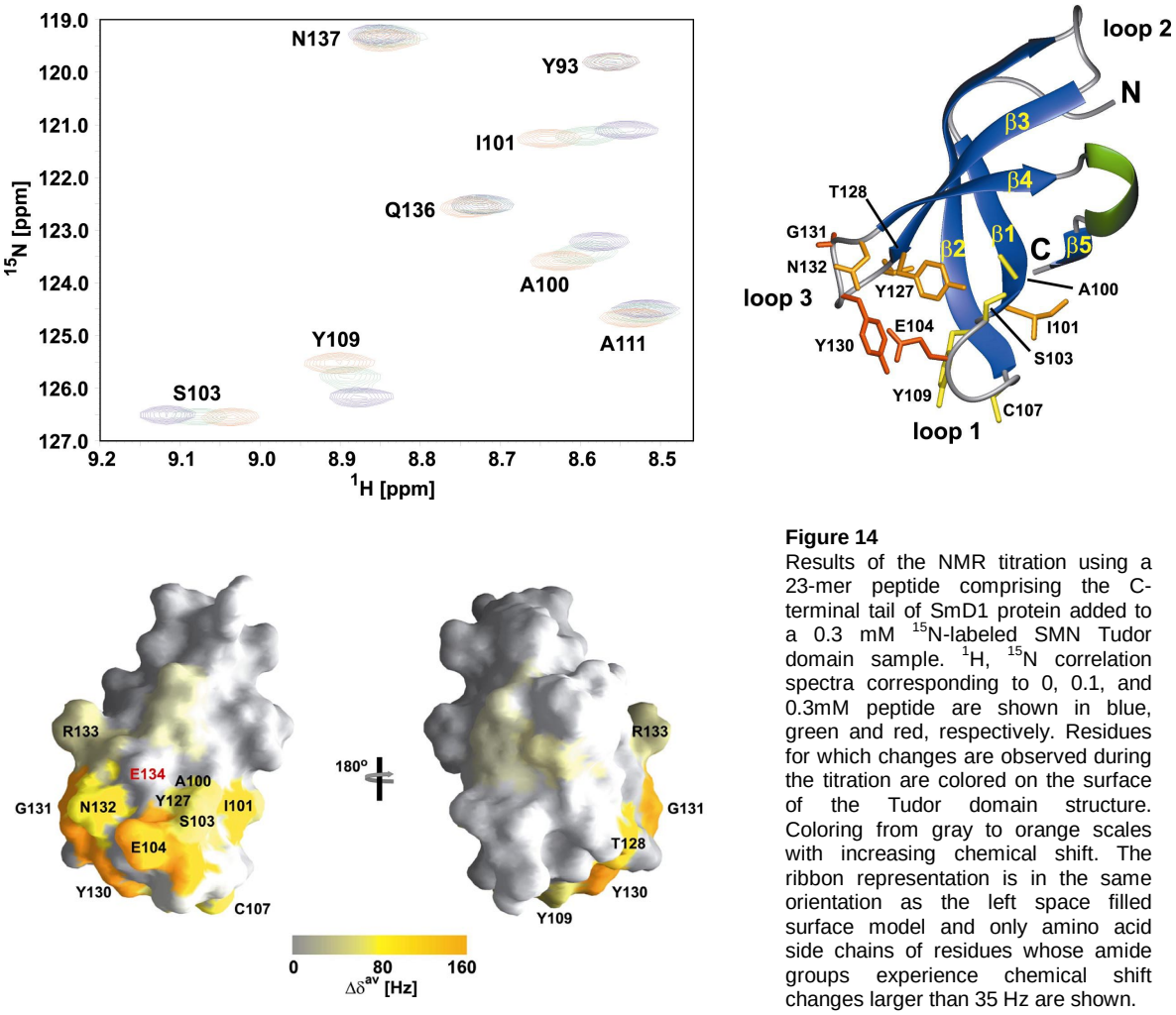


Figure 14
Results of the NMR titration using a 23-mer peptide comprising the C-terminal tail of SmD1 protein added to a 0.3 mM ^{15}N-labeled SMN Tudor domain sample. ^{1}H, ^{15}N correlation spectra corresponding to 0, 0.1, and 0.3mM peptide are shown in blue, green and red, respectively. Residues for which changes are observed during the titration are colored on the surface of the Tudor domain structure. Coloring from gray to orange scales with increasing chemical shift. The ribbon representation is in the same orientation as the left space filled surface model and only amino acid side chains of residues whose amide groups experience chemical shift changes larger than 35 Hz are shown.

1.2.7. Role of RG-Methylation for Sm/SMN Interaction

The carboxy-terminal tails of Sm D1, Sm D3, Sm B/B' and LSm4 are post-translationally modified *in vivo* (Brahms et al., 2001; Brahms et al., 2000). For these Sm and Sm-like proteins, all arginine residues within the RG-rich C-terminal tail regions contain symmetrical di-methylated arginines (sDMAs) (see Figure 15). In the case of Sm D_1 and Sm D_3, these are antigenic sites that are specifically recognized by auto-immune antibodies from patients suffering from systemic lupus erythematosus (SLE) (Brahms et al., 2000). Methylation of Sm proteins is exerted by a protein complex containing PRMT5 and the putative U snRNP assembly

factor plCln (Meister et al., 2001b). Furthermore, SMN has been shown to preferentially bind to dimethylarginine containing protein targets (Friesen et al., 2001) and that symmetrically di-methylated versions of Sm tails would compete binding to an N-terminal SMN fragment (1-160) over asymmetrically di-methylated and non-methylated Sm tail peptides (Brahms et al., 2001). It appears that methylation of arginine residues within the C-terminal regions of Sm proteins functions as a tunable post-translational modification to regulate the affinity of interaction to the SMN protein.

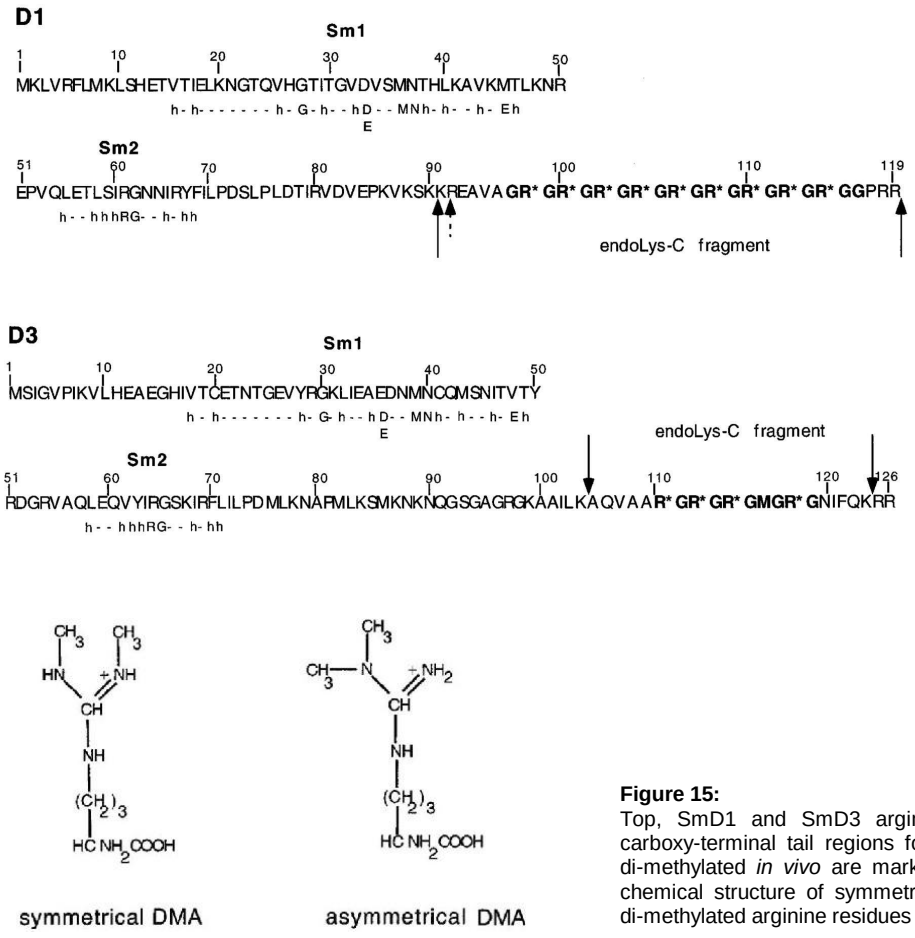


Figure 15:
Top, SmD1 and SmD3 arginine residues within the carboxy-terminal tail regions found to be symmetrically di-methylated *in vivo* are marked with an asterisk. The chemical structure of symmetrically and asymmetrically di-methylated arginine residues is outlined at the left.

To investigate the effect of these modifications on the binding behavior of Sm peptides to the Tudor domain of SMN, we used asymmetrically and symmetrically di-methylated versions of the aforementioned Sm D_1 C-terminal tail peptide (see above) and tested their mode of binding to the Tudor domain by NMR. We performed an identical set of NMR titrations with these modified peptides and

observed, firstly, identical changes in chemical shifts for all three versions of the peptide, indicating identical binding sites irrespective of the state of methylation, secondly, the same micromolar affinities (fast exchange rates between free and bound conformations) for non-methylated and asymmetrically di-methylated versions of the peptide, and thirdly, that only the symmetrically di-methylated Sm D_1 tail peptide would interact in a slow exchange regime on the NMR time scale and therefore exhibit an increase in affinity towards the Tudor domain by, at least, an order of magnitude (hence, low micromolar to nanomolar in Kd). In summary we conclude that the Tudor domain of human SMN binds with an order of magnitude higher affinity to the symmetrically di-methylated carboxy-terminal RG-rich tail of Sm D_1, as compared to an unmodified version of such a peptide. Nevertheless, residues of the Tudor domain forming the binding site for Sm interaction are identical irrespective of the state of methylation of the target ligand. Similar results for other modified/unmodified Sm proteins are expected and we believe that this mechanism of regulation may be general to most RG-rich ligand SMN interactions.

1.3. SUMMARY and CONCLUSIONS

In summary, this section of the thesis describes the determination of the hitherto unknown three-dimensional structure of the Tudor domain of human SMN. This structural motif adopts a five stranded β-barrel conformation, which, based on sequence conservation of hydrophobic residues forming the core of this structure, is expected to be conserved in all Tudor domain containing proteins. We pointed out the structural similarity of this fold with the previously determined Sm protein fold, characteristically encountered in the snRNP contained Sm proteins and have furthermore confirmed interaction of Sm D_1, D_2 and D_3 with the Tudor domain of human SMN. A binding site for these interactions was delineated to the carboxy-terminal RG-rich tail regions within the Sm proteins. We have mapped the binding interface onto the Tudor domain structure and found that it involves the negatively charged surface area, comprising the conserved glutamate residue 134 and showed that the charge-reverting point mutation, E134K, corresponding to a ‘natural’ SMA disease causing mutation, does not disrupt the Tudor domain structure but severly affects the Tudor domains’ ability to interact with Sm proteins. Furthermore, the *in vivo* occurring post-translational modification, sMDA, within the

RG-rich C-terminal tails of Sm proteins, greatly enhances the affinity of these ligands towards the Tudor domain of SMN.

Conclusively, we have elucidated in detail one aspect of the structural and molecular basis underlying the essential biogenesis of functional snRNP particles and. offered a plausible explanation for a molecular defect underlying a specific case of the genetic disease of spinal muscular atrophy (SMA).

1.4. PERSPECTIVES

The Tudor domain preferentially binds to symmetrically di-methylated RG peptides. From a structural point of view, it is intriguing to ask how the Tudor domain selectively discriminates between modified and unmodified arginine amino acids. A special three-dimensional structural arrangement must be present to allow such a selective mode of binding and obviously, elucidation of the atomic details of how the Tudor domain accomplishes this task, represents a challenging question for future investigations. We have begun to experimentally address this problem employing both NMR spectroscopy and X-ray crystallography.

1.5. REFERENCES

Baccon, J., Pellizzoni, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Identification and characterization of Gemin7, a novel component of the SMN complex. J Biol Chem *13*, 13.

Bertrandy, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D., Munnich, A., and Lefebvre, S. (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. Hum Mol Genet *8*, 775-782.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Luhrmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. Rna *7*, 1531-1542.

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., and Luhrmann, R. (2000). The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. J Biol Chem *275*, 17122-17129.

Buhler, D., Raker, V., Luhrmann, R., and Fischer, U. (1999). Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. Hum Mol Genet *8*, 2351-2357.

Campbell, L., Hunter, K. M., Mohaghegh, P., Tinsley, J. M., Brasch, M. A., and Davies, K. E. (2000). Direct interaction of Smn with dp103, a putative RNA helicase: a role for Smn in transcription regulation? Hum Mol Genet *9*, 1093-1100.
Charroux, B., Pellizzoni, L., Perkinson, R. A., Shevchenko, A., Mann, M., and Dreyfuss, G. (1999). Gemin3: A novel DEAD box protein that interacts with SMN, the spinal muscular atrophy gene product, and is a component of gems. J Cell Biol *147*, 1181-1194.

Charroux, B., Pellizzoni, L., Perkinson, R. A., Yong, J., Shevchenko, A., Mann, M., and Dreyfuss, G. (2000). Gemin4. A novel component of the SMN complex that is found in both gems and nucleoli. J Cell Biol *148*, 1177-1186.

Chu, J. L., and Elkon, K. B. (1991). The small nuclear ribonucleoproteins, SmB and B', are products of a single gene. Gene *97*, 311-312.

Coovert, D. D., Le, T. T., McAndrew, P. E., Strasswimmer, J., Crawford, T. O., Mendell, J. R., Coulson, S. E., Androphy, E. J., Prior, T. W., and Burghes, A. H. (1997). The survival motor neuron protein in spinal muscular atrophy. Hum Mol Genet *6*, 1205-1214.

Fischer, U., Darzynkiewicz, E., Tahara, S. M., Dathan, N. A., Luhrmann, R., and Mattaj, I. W. (1991). Diversity in the signals required for nuclear accumulation of U snRNPs and variety in the pathways of nuclear transport. J Cell Biol *113*, 705-714.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell *90*, 1023-1029.

Friesen, W. J., and Dreyfuss, G. (2000). Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). J Biol Chem *275*, 26370-26375.

Friesen, W. J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001). SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. Mol Cell *7*, 1111-1117.

Giesemann, T., Rathke-Hartlieb, S., Rothkegel, M., Bartsch, J. W., Buchmeier, S., Jockusch, B. M., and Jockusch, H. (1999). A role for polyproline motifs in the spinal muscular atrophy protein SMN. Profilins bind to and colocalize with smn in nuclear gems. J Biol Chem *274*, 37908-37914.

Gubitz, A. K., Mourelatos, Z., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Gemin5, a novel WD repeat protein component of the SMN complex that binds Sm proteins. J Biol Chem *277*, 5631-5636.
Hastings, M. L., and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. Curr Opin Cell Biol *13*, 302-309.

Hebert, M. D., Szymczyk, P. W., Shpargel, K. B., and Matera, A. G. (2001). Coilin forms the bridge between Cajal bodies and SMN, the spinal muscular atrophy protein. Genes Dev *15*, 2720-2729.

Hermann, H., Fabrizio, P., Raker, V. A., Foulaki, K., Hornig, H., Brahms, H., and Luhrmann, R. (1995). snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein-protein interactions. Embo J *14*, 2076-2088.

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem Sci *20*, 478-480.

Jones, K. W., Gorzynski, K., Hales, C. M., Fischer, U., Badbanchi, F., Terns, R. M., and Terns, M. P. (2001). Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. J Biol Chem *276*, 38645-38651.

Kambach, C., Walke, S., and Nagai, K. (1999a). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. Curr Opin Struct Biol *9*, 222-230.
Kambach, C., Walke, S., Young, R., Avis, J. M., de la Fortelle, E., Raker, V. A., Luhrmann, R., Li, J., and Nagai, K. (1999b). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell *96*, 375-387.

Kastner, B., Bach, M., and Luhrmann, R. (1990a). Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common

structure-determining principle in the major U snRNP family. Proc Natl Acad Sci U S A *87*, 1710-1714.
Kastner, B., Bach, M., and Luhrmann, R. (1990b). Electron microscopy of snRNPs U2, U4/6 and U5: evidence for a common structure-determining principle in the major UsnRNP family. Mol Biol Rep *14*, 171.

Lefebvre, S., Burlet, P., Liu, Q., Bertrandy, S., Clermont, O., Munnich, A., Dreyfuss, G., and Melki, J. (1997). Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet *16*, 265-269.

Liu, Q., and Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. Embo J *15*, 3555-3565.

Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell *90*, 1013-1021.

Lorson, C. L., and Androphy, E. J. (1998). The domain encoded by exon 2 of the survival motor neuron protein mediates nucleic acid binding. Hum Mol Genet *7*, 1269-1275.

Lorson, C. L., Strasswimmer, J., Yao, J. M., Baleja, J. D., Hahnen, E., Wirth, B., Le, T., Burghes, A. H., and Androphy, E. J. (1998). SMN oligomerization defect correlates with spinal muscular atrophy severity. Nat Genet *19*, 63-66.

Luhrmann, R., Kastner, B., and Bach, M. (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim Biophys Acta *1087*, 265-292.
MacKenzie, A. E., and Gendron, N. H. (2001). Tudor reign. Nat Struct Biol *8*, 13-15.

Mattaj, I. W., and De Robertis, E. M. (1985). Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. Cell *40*, 111-118.
Meister, G., Buhler, D., Pillai, R., Lottspeich, F., and Fischer, U. (2001a). A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. Nat Cell Biol *3*, 945-949.
Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., and Fischer, U. (2001b). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. Curr Biol *11*, 1990-1994.

Mourelatos, Z., Abel, L., Yong, J., Kataoka, N., and Dreyfuss, G. (2001). SMN interacts with a novel family of hnRNP and spliceosomal proteins. Embo J *20*, 5443-5452.

Palacios, I., Hetzer, M., Adam, S. A., and Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. Embo J *16*, 6783-6792.

Pearn, J. (1978). Incidence, prevalence, and gene frequency studies of chronic childhood spinal muscular atrophy. J Med Genet *15*, 409-413.

Pearn, J. H., and Wilson, J. (1973). Acute Werdnig-Hoffmann disease: acute infantile spinal muscular atrophy. Arch Dis Child *48*, 425-430.

Pellizzoni, L., Baccon, J., Charroux, B., and Dreyfuss, G. (2001a). The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1. Curr Biol *11*, 1079-1088.

Pellizzoni, L., Baccon, J., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). Purification of native survival of motor neurons complexes and identification of Gemin6 as a novel component. J Biol Chem *277*, 7540-7545.

Pellizzoni, L., Charroux, B., and Dreyfuss, G. (1999). SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. Proc Natl Acad Sci U S A *96*, 11167-11172.

Pellizzoni, L., Charroux, B., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001b). A functional interaction between the survival motor neuron complex and RNA polymerase II. J Cell Biol *152*, 75-85.
Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell *95*, 615-624.

Ponting, C. P. (1997). Tudor domains in proteins that interact with RNA. Trends Biochem Sci *22*, 51-52.

Raker, V. A., Plessel, G., and Luhrmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. Embo J *15*, 2256-2269.

Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. Curr Opin Genet Dev *6*, 215-220.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-345.

Roberts, D. F., Chavez, J., and Court, S. D. (1970). The genetic component in child mortality. Arch Dis Child *45*, 33-38.

Rossoll, W., Kroning, A. K., Ohndorf, U. M., Steegborn, C., Jablonka, S., and Sendtner, M. (2002). Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Hum Mol Genet *11*, 93-105.

Schmalbruch, H., and Haase, G. (2001). Spinal muscular atrophy: present state. Brain Pathol *11*, 231-247.

Selenko, P., Sprangers, R., Stier, G., Buhler, D., Fischer, U., and Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat Struct Biol *8*, 27-31.

Sendtner, M. (2001). Molecular mechanisms in spinal muscular atrophy: models and perspectives. Curr Opin Neurol *14*, 629-634.
Seraphin, B. (1995). Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. Embo J *14*, 2089-2098.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

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.

Toro, H., Thore, S., Mayer, C., Basquin, J., Seraphin, B., and Suck, D. (2001). RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. Embo J. 9, 2293-2303

Urlaub, H., Raker, V. A., Kostka, S., and Luhrmann, R. (2001). Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. Embo J *20*, 187-196.

van Dam, A., Winkel, I., Zijlstra-Baalbergen, J., Smeenk, R., and Cuypers, H. T. (1989). Cloned human snRNP proteins B and B' differ only in their carboxy-terminal part. Embo J *8*, 3853-3860.

Will, C. L., and Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. Curr Opin Cell Biol *9*, 320-328.

Will, C. L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol *13*, 290-301.

Yong, J., Pellizzoni, L., and Dreyfuss, G. (2002). Sequence-specific interaction of U1 snRNA with the SMN complex. Embo J *21*, 1188-1196.

Young, P. J., Le, T. T., thi Man, N., Burghes, A. H., and Morris, G. E. (2000). The relationship between SMN, the spinal muscular atrophy protein, and nuclear coiled bodies in differentiated tissues and cultured cells. Exp Cell Res *256*, 365-374.

2. CHAPTER 2

2.1. INTRODUCTION

2.1.1. Pre-Spliceosome Assembly

To allow for correct joining of exon sequences during the splicing reaction, the exon-intron boundaries at the 5' and 3' splice-sites need to be defined by protein factors binding to specific RNA sequence elements and serve as marks for proper assembly of components of the spliceosome. The first ATP-dependent step in spliceosome assembly is the stable association of U2 snRNP with the 3' region of the intron (Nelson and Green, 1989). This region of the intron contains three sequence elements which are important for the splicing process: the branch-point region (BP or BPS) containing the invariant branch-point adenosine nucleotide, the polypyrimidine tract (Py tract or PPT) and the conserved di-nucleotide AG at the 3' splice site. The BP is highly conserved in yeast and more degenerate in higher eukaryotes (Rain et al., 1998). It establishes base pairing interactions with a specific sequence of U2 snRNA, bulging out the adenosine nucleotide that forms a 2'-5' phosphodiester bond with the 5' end of the intron. The Py-tract is a pyrimidine-rich sequence located immediately downstream of the BP and upstream of the AG di-nucleotide. Two proteins are instrumental for the initial recognition of these pre-mRNA sequence elements. U2 auxilliary factor, U2AF, comprising a large ($U2AF^{65}$) and small subunit ($U2AF^{35}$) (Zamore and Green, 1989) and SF1 (Splicing factor 1) or branch-point binding protein (BBP) (Berglund et al., 1997; Kramer, 1992). $U2AF^{65}$ employs two 'classical' RNP-type RNA binding domains to interact with the conserved Py-tract (Zamore et al., 1992). SF1 binds to the branch point adenosine upstream of it, via its KH-QUA2-domain (Berglund et al., 1997). Additionally, $U2AF^{65}$ and SF1 interact with each other in an regulated fashion (Berglund et al., 1998; Wang et al., 1999) and thus introduce an additional tunable parameter in this cooperative RNA recognition. $U2AF^{65}$ and $U2AF^{35}$ stably interact with one another and $U2AF^{35}$ recognizes the 3' splice site AG and helps to stabilize the U2AF Py-tract interaction (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). This is the initial architectural arrangement at the 3' splice-site, which, together with the U1snRNP specifically recognizing the 5' splice-site, forms the commitment or E-complex (Figure 16). Several 3' splice-site

subunits are subsequently exchanged by 'later' U2-specific proteins, SF1 is substituted by the SAP155 protein for example, and the commitment complex matures into complex A.

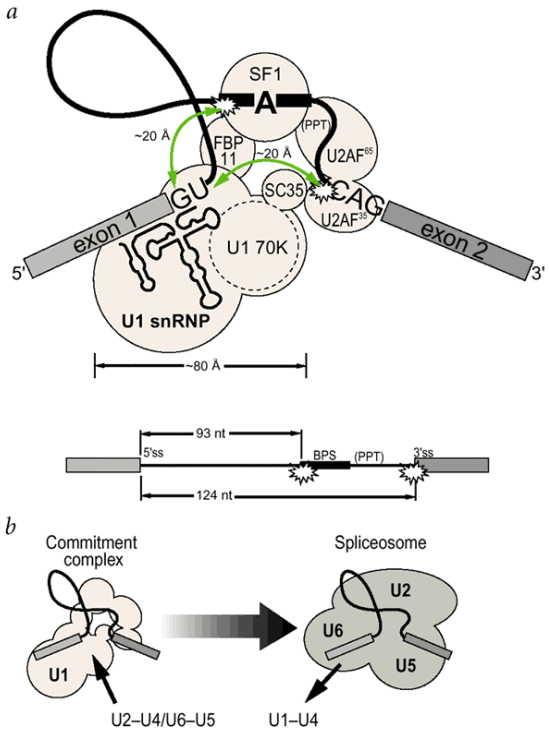


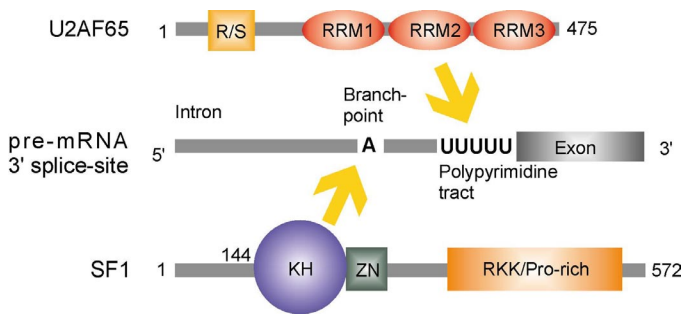
Figure 16:
a.) Schematic representation of the 'commitment' complex (complex E). Protein factors of the U1 snRNP together with U1 snRNA bind to the 5' splice site, whereas RNA sequence elements of the 3' splice-site (the AG di-nucleotide, the poly-pyrimidine tract (PPT), and the branch-point adenosine) are specifically bound by the U2AF heterodimer ($U2AF^{65}$ and $U2AF^{35}$) and the SF1 protein. Additional protein factors bridging the 5' and 3' splice sites are indicated and the respective distances between the reactive nucleotides in the 'commitment' complex are outlined by green arrows. Panel b.) displays maturation of the commitment complex into the splicing compatible macromolecular assembly of the spliceosome (Reproduced from Kent and MacMillan, 2002)

2.1.2. Domain Organization of $U2AF^{65}$ and SF1

U2AF is a hetero-dimeric protein consisting of a large ~65kD ($U2AF^{65}$) and a small ~35kD ($U2AF^{35}$) subunit. Human $U2AF^{65}$ is a 475-residue protein that contains an N-terminal arginine-rich RS domain (aa 30-60) followed by a positively charged stretch of residues (aa 85-90) and a poly-proline rich sequence (aa 102-112). It employs these later sequence elements to specifically interact with the central RNA-binding domain (RRM) of $U2AF^{35}$ (see Chapter 2.2.5). $U2AF^{65}$ also contains three consecutive, classical RNP-type RNA binding motifs (RRM1 aa 148-237, RRM2 aa 258-342, RRM3 aa 367-475) at the protein's C-terminus (Zamore and Green, 1991; Zamore et al., 1992). Earlier data suggested that all three RNA binding domains are necessary for high affinity poly-pyrimidine tract recognition (Zamore et al., 1992). Availability of the recently determined solution structures of the first two RRMs of human $U2AF^{65}$ (Ito et al., 1999), indicates that a construct used in a experiment described in (Zamore et al., 1992)., lacking the very C-

terminal third RRM (U2AF^{65}ΔRRM3, deletion of residues 325-475), was also corrupted in the second RRM as it completely lacked the last β-strand of RRM2. Therefore, any conclusions on the requirement of the third RRM for RNA binding drawn there, need to be considered with great caution.

U2AF^{65} employs its first two RRMs (RRM1 and RRM2) to specifically interact with the poly-pyrimidine tract RNA and binds to the N-terminus of SF1 via its third RRM (RRM3), thus employing a RNA binding motif for protein-protein interaction (Berglund et al., 1998; Rain et al., 1998). Human SF1 is a 572 residue protein which exhibits a N-terminal KH-domain (aa 134-227) immediately followed by a QUA2-domain (aa 235-260) and a distant Zn-knuckle type of metal coordinating sequence (aa 287-294). The KH- and QUA2-elements form a joint structural integrity that interacts with the branch-point adenosine nucleotide (Liu et al., 2001). SF1 furthermore contains an extended proline-rich sequence element at its C-terminus (aa 335-448). A domain overview of the U2AF^{65} and SF1 proteins and how they cooperatively reconstitute with their respective pre-mRNA recognition sequences is given in Figure 17.



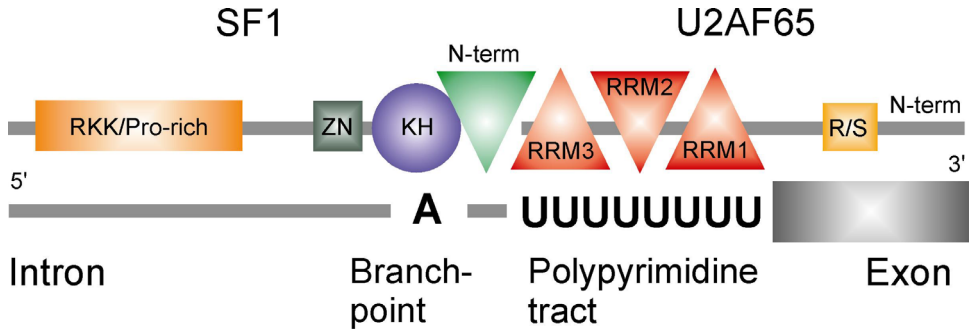


Figure 17:
Schematic domain organization of the U2AF^{65} and SF1 proteins and their respective recognition sequences on pre-mRNA. The lower panel depicts the cooperative assembly of these early splicing factors at the 3' splice-site, involving multiple protein-RNA and protein-protein contacts

2.1.3. RNP-type RNA Binding Domains (RRM)

RNP-type RNA binding domains (RRMs) are amongst the most prominent sequence motifs in the genomic database (Kenan et al., 1991; Nagai et al., 1995). There is a wealth of biochemical and structural information of free form RRMs (Avis et al., 1996; Ito et al., 1999; Xu et al., 1997), RRMs in complex with their cognate RNAs (Deo et al., 1999; Ding et al., 1999; Handa et al., 1999; Wang and Tanaka Hall, 2001) and mapped interactions of RNA onto RRM folds (Ito et al., 1999; Yuan et al., 2002), available. The RNP-type fold comprises two conserved sequence motifs (RNP1 and RNP2) which consist of aromatic and basic residues. It exhibits a four-stranded antiparallel β-sheet which is unilaterally packed by two α-helices leading to the typical RNP-type topology (β1α1β2β3α2β4). One side of the β-surface remains accessible for RNA binding, whereas the other side is shielded from solvent by the presence of the two α-helices (Figure 18).

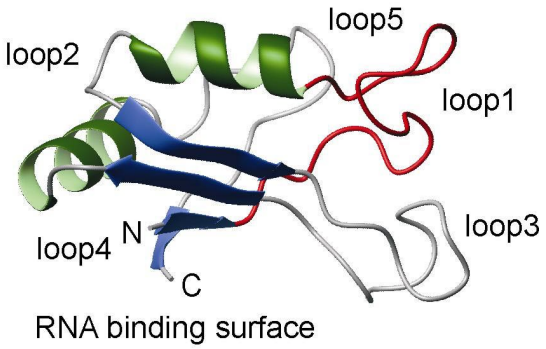


Figure 18:
Ribbon diagram of the 'classical' RNP-type RRM domain fold. The 4 anti-parallel β-strands comprise the typical RNA binding surface and are shown in blue. Two α-helices (helix A and B) pack against the back-side of this β-sheet and are depicted in green. The loop region connecting the first two secondary structure elements (Loop 1) is shown in red

Exceptions to the 'classical' RNP-type fold are observed for the third RRM of the poly-pyrimidine tract binding protein PTB, which contains an additional fifth β-strand (Conte et al., 2000), extending its RNA binding surface, and for U1A and U2B'', exhibiting an additional short, C-terminal helix (Oubridge et al., 1994; Price et al., 1998). Common to all known structures of RRM-RNA complexes to date is the usage of the accessible β-surface to accommodate cognate RNA binding (Antson, 2000; Perez-Canadillas and Varani, 2001). RNA substrates can vary from duplex RNA sequences in hairpin-loop types of conformations, as seen for the U1A RNA complexes and for the U2A' U2B'' protein-RNA assembly (Allain et al., 1997; Oubridge et al., 1994; Varani et al., 2000) and (Price et al., 1998) respectively, to extended single stranded RNAs in kinked-, *tra* mRNA to Sex-lethal protein and AU-

rich element to HuD protein interaction (Handa et al., 1999; Wang and Tanaka Hall, 2001), and planar-conformations, like for the polyadenylate RNA Poly(A)-binding protein complex (Deo et al., 1999). In these last examples, two RRM domains are used to facilitate RNA binding and due to the preferential occurrence of multiple, consecutive RRMs in a protein sequence, a concerted mode of RNA binding can be envisaged for many RRM containing proteins (Shamoo et al., 1995).

2.1.4. Characteristics of RRM Domains in U2AF^{65}

For U2AF^{65}, initial sequence comparison between U2AF^{65}-RRM1, -RRM2 and RRM3 indicated a sequence insertion of negatively charged residues for the third, most C-terminal RRM, which we initially predicted to extend the loop region between the first two secondary structure elements (β1 and α1) (Figure 19). Due to these residues, structure based homology modeling suggested a highly negative electrostatic surface potential for U2AF^{65}-RRM3. This is in contrast to most RRMs with known RNA binding activity which have an neutral overall potential .

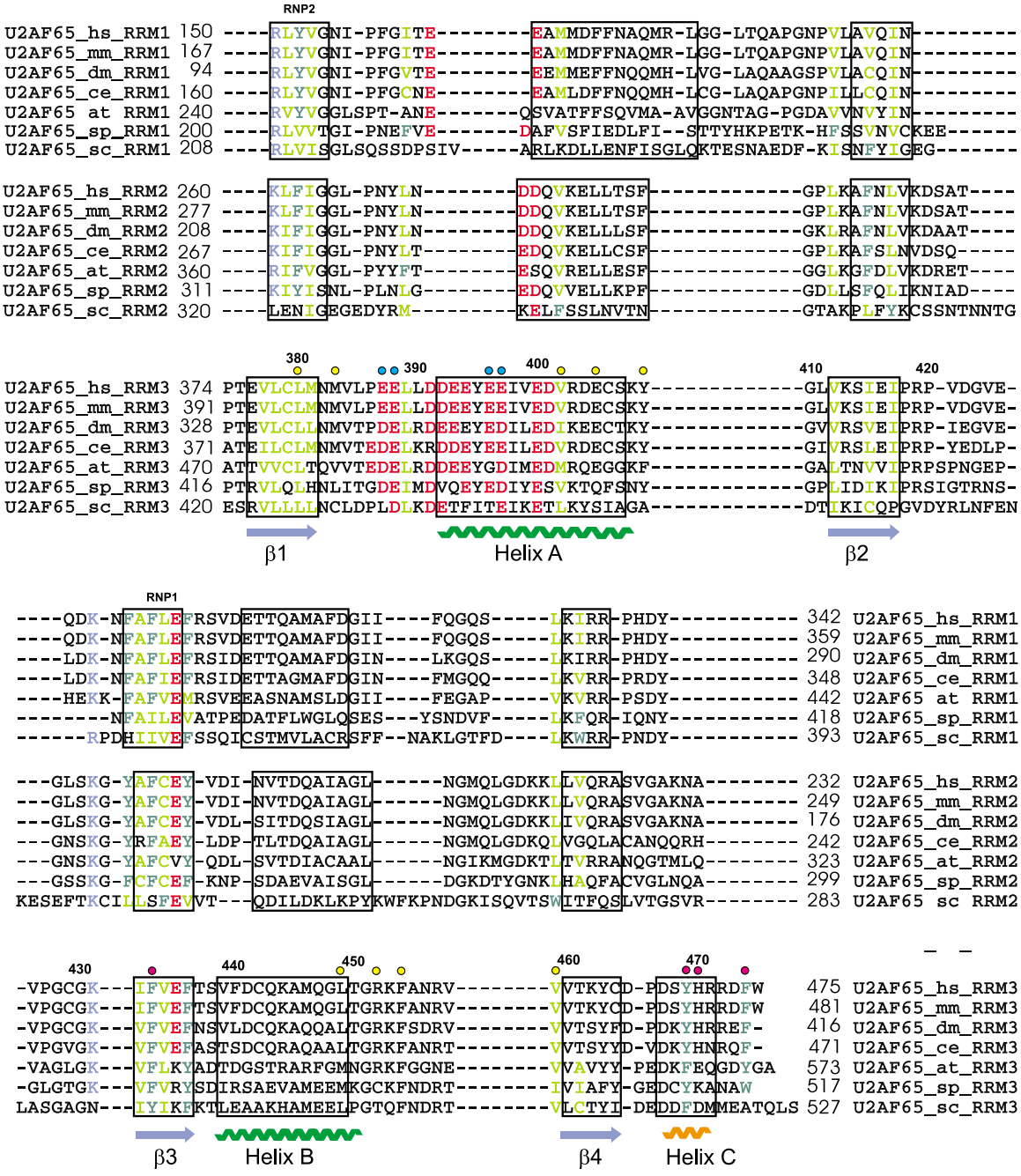


Figure 19:
Structure based multiple sequence alignment of U2AF^{65} RRM1, 2 and RRM3. U2AF^{65} Homo sapiens (hs) acc.# 1805352A, Mus musculus (mm) acc.# S22646, Drosophila melanogaster (dm) acc.# NP_476891, Caenorhabditis elegans (ce) acc.# AAM44400, Arabidopsis thaliana (at) acc.# CAB16828, Schizosaccharomyces pombe (sp) acc.# AAA03578 and Mud2p/U2AF^{65} of Saccharomyces cerevisiae (sc) acc.# AAA64215. The respective RNP1 and RNP2 motifs are indicated and boxed in black. Conserved residues in all three RRMs are colored green for hydrophibic, blue for basic, turquoise for aromatic and red for acidic residues, respectively. The RRM3 characteristic stretch of negatively charged amino acids within the first α-helix (helixA) are shown in red. Secondary structure elements for human U2AF^{65} RRM3 are indicated below the respective alignment block and sequence numbering is given. Yellow circles above the human U2AF^{65} primary sequence denote residues forming the hydrophobic pocket accommodating Trp22 of human SF1 (see Chapter 2.2.3). Blue dotes indicate positions of charge-reverting residues discussed in Chapter 2.2.3 and pink circles point to aromatic residues within the C-terminal additional helix C forming the hydrophobic cage around the β-sheet protruding Phe433 (Chapter 2.2.4)

Additionally, the first 28 residues in SF1, shown to be essential and sufficient for binding to U2AF^{65}-RRM3 (Rain et al., 1998). These residues contain a reversibly phosphorylated serine residue (Wang et al., 1999) and exhibit a conserved stretch of positively charged amino acids, suggesting an electrostatic driven interaction between U2AF^{65}-RRM3 and the N-terminus of SF1 (Figure 20). The biological role of phosphorylation in this part of the SF1 sequence has been demonstrated to regulate U2AF^{65}-RRM3 SF1 interaction, as a serine-phosphorylated version of the N-terminus of SF1 can no longer bind to U2AF^{65} (Wang et al., 1999).

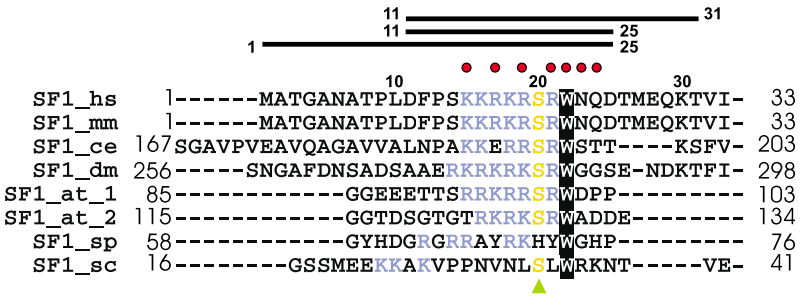


Figure 20:
Multiple sequence alignment of N-terminal residues of SF1. Homo sapiens (hs) acc.# CAA03883, Mus musculus (mm) acc.# CAA59797, Caenorhabditis elegans (ce) acc.# CAB64866, Drosophila melanogaster (dm) acc.# CAB64937, Arabidopsis thaliana (at_1 and at_2) acc.# BAA97393 (note that the N-terminus of this homologue of SF1 contains an repeated version the U2AF^{65} interaction sequence), Schizosaccharomyces pombe (sp) acc.# AAF02214 and Saccharomyces cerevisiae (sc) acc.# NP_013217. Conserved positively charged residues are colored in blue and positions of charge-reverting point mutations discussed in Chapter 2.2.3 are indicated with red circles. The potentially phosphorylated Serine, residue number 20 in human SF1, is colored in yellow and marked with a green triangle. The invariant Tryptophan (Trp22 in human SF1), forming the main interaction site in the U2AF^{65} SF1 interface, is boxed in black (Chapter 2.2.3). Black bars above the alignment display synthetic peptides of human SF1 reported in this study and indicate their respective sequences (Chapter 2.2.2).

To study the molecular and structural mechanism underlying the U2AF^{65}-RRM3 SF1 interaction, we solved the three-dimensional solution structure of human U2AF^{65}-RRM3 in complex with a peptide fragment corresponding to the first 25 residues of SF1 by NMR and biochemically investigated their mode of binding.

2.2. RESULTS and DISCUSSION

2.2.1. Structural Characteristics of U2AF^{65}-RRM3

The solution structure of human U2AF^{65}-RRM3 in complex with a N-terminal peptide of human SF1 was solved using standard methods for biomolecular structure determination by NMR (Chapter 3.2). ^{15}N and ^{15}N, ^{13}C labeled versions of recombinant human U2AF^{65}-RRM3 (aa 367-475) were produced (see materials and methods) and the U2AF/SF1 complex was reconstituted by addition of a synthetic peptide fragment corresponding to residues 11-25 of human SF1. Furthermore, a second U2AF/SF1 complex was assembled by recombinant protein production of a doubly-labeled N-terminal fragment of human SF1 (aa 1-25) to which an unlabeled version of recombinant U2AF^{65}-RRM3 (aa 367-475) was added. Thus, two equivalent, differentially labeled protein-peptide complexes were available for structural studies. An identical set of heteronuclear triple-resonance NMR experiments was recorded for both assemblies (materials and methods), NOE- and edited-filtered NOE-experiments (Chapter 3.5.1) were finally employed to structurally arrange complex components with respect to each other.

The U2AF^{65}-RRM3 SF1 complex is well defined by the NMR data, structural statistics are summarized in Table 2.

Table 2 structural statistics for the U2AF SF1 complex

Experimental restraints		Number of restraints	R.m.s. deviation
Distance restraints (Å)			
Unambigous NOEs (total)		3232	0.0106 +/- 0.009
	$U2AF^{65}$ Total	2910	
	$U2AF^{65}$ Short range	985	
	$U2AF^{65}$ Long range	642	
	SF1 Total	258	
	SF1 Short range	46	
	SF1 Long range	0	
	Intermolecular	64	
No NOE viol > 0.3Å			
Hydrogen Bonds		34	0.0196 +/- 0.026
Dihedral angle restraints Φ(º)		35	0.10 +/- 0.13
No dihedral viol > 0.13º			
Coordinate precision (Å)			
RRM3 secondary struct., backbone			0.50 +/- 0.08
RRM3 secondary struct., heavy atoms			1.27 +/- 0.15
RRM3 all, backbone			0.87 +/- 0.29
RRM3 all, heavy atoms			1.72 +/- 0.44
SF1 19-23 all, backbone			1.09 +/- 0.37
SF1 19-23 all, heavy atoms			2.97 +/- 0.78
Structural quality (Lennard-Jones Energy, kcal mol^{-1})			-1176.94 +/- 10.7
Ramachandran plot (%)			
RRM3_all	most favored		76.5
	Additionally allowed		20.8
	Generously allowed		2.8
	Disallowed		0.0

The solution structure of $U2AF^{65}$-RRM3 exhibits the classical RNP-type fold. Residues 376-381, 412-417, 432-436 and 460-464 form β1, β2, β3 and β4 of the four-stranded anti-parallel β-sheet, with amino acids 392-407 and 439-449 adopting α-helical conformations and thus making the characteristic two helices A and B in the typical β1α1β2β3α2β4 topology. Additionally, the very C-terminal set of residues (aa 467-475), including conserved aromatic amino acids (Tyr469, His 470, Phe474, Trp475), give rise to an extra helix, 467-471 (helix C), that closely packs against the normally solvent exposed surface area of the β-sheet (Figure 21).

This unusual extension to the RNP-type fold is reminiscent of the N-terminal RRM of the human U1A protein, which also exhibits a similarly coordinated additional helix in its ligand free conformation (see also Chapter 2.2.4, Figure 31) (Allain et al., 1997; Avis et al., 1996). Initial NMR experiments indicated that the presence of these helical residues in U2AF^{65}-RRM3 is important for proper folding of the domain as a shorter version of this construct, Δaa 468-475, (lacking the very C-terminal residues) fails to give single conformation NMR spectra and is structurally unstable. A more detailed analysis of the importance of this additional helix and a detailed comparison to the U1A structure is given in Chapter 2.2.4. Also, the first α-helix (helix A) in U2AF^{65}-RRM3 extends N-terminally for two turns, as compared to the majority of known RNP-type RRM structures to date. Instead of the usual 10 helical residues (3 turns), common to most RRMs, U2AF^{65}-RRM3 accommodates 16 amino acids in helix A (4 turns) and places its characteristic negatively charged residues in an α-helical, rather than in an extended loop-like, conformation. In the present α-helical conformation all negatively charged amino acids, which occur in a repeating –XX-(n)$_2$- [X being Glu or Asp] pattern, adopt a structural arrangement in which all side-chain carboxyl-groups are at equivalent positions, separated by single helical turns, along one side of this structural element. Thus creating a continuous, negatively charged ridge along one side of the first α-helix (Figure 28 and Chapter 2.2.3).

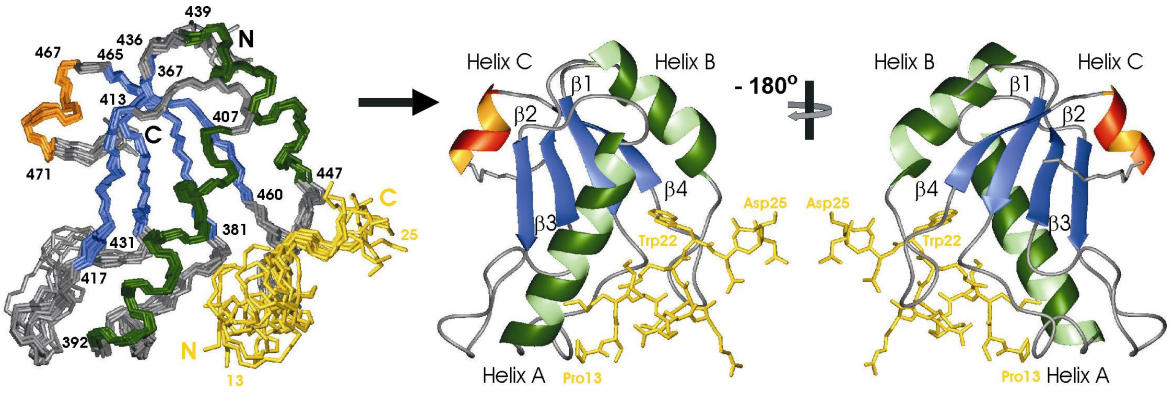


Figure 21:
Left: Backbone atoms (N, Cα, C') of residues 376-475 of U2AF^{65}-RRM3 and 13-25 of SF1 (1-25) from an NMR ensemble of 10 superimposed lowest-energy structures of the U2AF SF1 protein peptide complex. Secondary structure elements are colored blue for β-strands, green for helix A, B and orange for the additional C-terminal helix C. The SF1 peptide is shown in yellow. The first and last residues within the respective secondary structure elements are indicated.
Middle: Ribbon representation of the U2AF^{65}-RRM3/SF1 complex structure closest to the average conformation of the NMR ensemble in the same orientation as the NMR ensemble. Corresponding β-strands and helices are indicated.
Right: rotated view of the complex structure, displaying the β-sheet surface classically employed by RRM-domains for RNA binding.

The recently determined X-ray structure of the hetero-dimeric interface of the small ($U2AF^{35}$) and large ($U2AF^{65}$) subunit of human U2AF (Kielkopf et al., 2001), yielded the first example of an unusually long, N-terminally extended helix A in a RNP-type RRM domain and also proved to be an example of an RRM-mediated protein-protein interaction. A more detailed evaluation of this work and how it relates to our findings is given in Chapter 2.2.5. $U2AF^{65}$-RRM3 contains two large loop conformations between β2 and β3 and α2 and β4 (residues 418-431 and 447-460), Loop 2 and 5, which appear flexible based on heteronuclear NOE data and less well defined in the NMR ensemble structure. Loops 1 and 5 participate in SF1 interaction and become more rigid upon peptide binding, as observed in heteronuclear NOE experiments for the complex (Figure 22). Upon SF1 coordination, Loop 1 and Loop 5 of $U2AF^{65}$-RRM3 become solvent protected as seen from differences in H_2O/D_2O exchange experiments.

A summary of NMR experimental data, illustrating structural differences between free and SF1 complexed $U2AF^{65}$-RM3 is given in Figure 22. In general, the overall fold of $U2AF^{65}$-RRM3 remains unchanged upon complex formation. No additional secondary structure elements are being formed and the overall flexibility of the RRM domain is not altered.

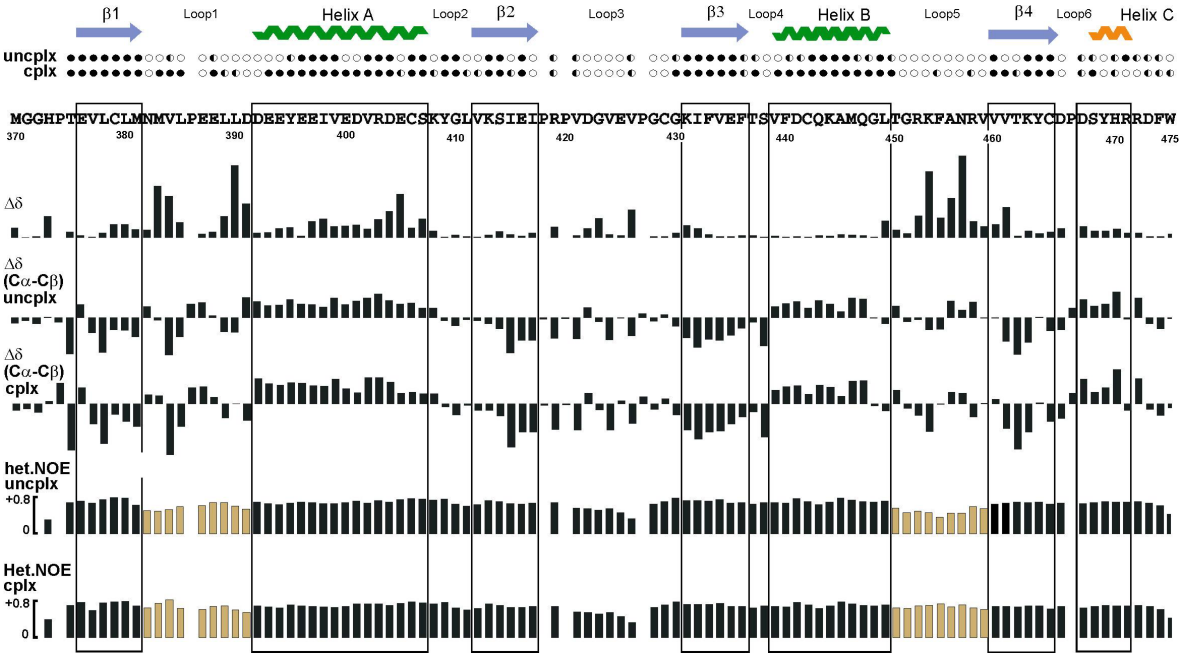


Figure 22:
Summary of NMR data for U2AF^{65}-RRM3 in its free and SF1 peptide complexed form. Circles above the respective amino acid sequence of U2AF^{65}-RRM3 display results from an H_2O/D_2O exchange experiment for free and peptide-bound forms of RRM3. Rapidly solvent exchanging amide groups are indicated with empty circles, whereas structurally buried residues, exchanging slowly, are marked by filled dots. Different degrees of solvent exposure of free and SF1 peptide complexed U2AF^{65}-RRM3 are discernable. Residues within Loop 1, parts of helix A and Loop 5 of RRM3 are shielded from solvent by SF1 peptide coordination. Additionally, Cα, Cβ chemical shift difference plots of uncomplexed and complexed U2AF^{65}-RRM3 are similar, indicating no changes in secondary structure upon SF1 (1-25) peptide addition. Positive values correspond to α-helical conformations, negative values indicate β-strands. Residues of U2AF^{65}-RRM3 within loop regions display small deviations from the mean chemical shift values characteristic for random conformational states. Heteronuclear NOE data for free and bound U2AF^{65}-RRM3 display positive values, corresponding to a rigid overall conformation, for this RRM domain with slightly lower numbers for more flexible loop regions. Upon SF1 binding, residues within Loop 1 and Loop 5 become dynamically restricted and exhibit larger values than in the free form (these regions are colored in yellow for clarity).

2.2.2. NMR of U2AF^{65}-RRM3 and SF1

NMR titration experiments with ^{15}N-labeled U2AF^{65}-RRM3 and SF1 peptides corresponding to residues 1-25, 11-25, 11-31 of human SF1, demonstrate tight (low micromolar to nanomolar affinity) binding, in the slow exchange regime on the NMR time scale, for all three peptides (Figure 23). Furthermore, chemical shift changes for U2AF^{65}-RRM3 in complex with SF1 are identical for all versions of SF1 peptides, thus indicating that residues N-terminal to Asp11 and C-terminal of amino acid Asp25 in human SF1 do not contribute to the U2AF-SF1 interaction. The HN chemical shift dispersion pattern remains unchanged after equimolar addition of excess peptide, strongly suggesting a stochiometric 1:1 complex between U2AF^{65}-RRM3 and SF1. Quantitative chemical shift perturbation mapping suggests that amide groups of residues comprising Loop 1, the negatively charged surface of helix A, and thus the U2AF^{65}-RRM3 specific negative amino acids and amino acids within Loop 5, experience the greatest changes in chemical environment upon peptide interaction and are therefore either directly or most closely involved in binding (Figure 24). Mapping chemical shift changes upon peptide binding onto the U2AF^{65}-RRM3 surface indicates that regions involved in peptide coordination overlap with negatively charged areas of the RRM domain (Figure 24). This points to an electrostatic contribution in binding and is further investigated by biochemical interaction studies (Chapter 2.2.3).

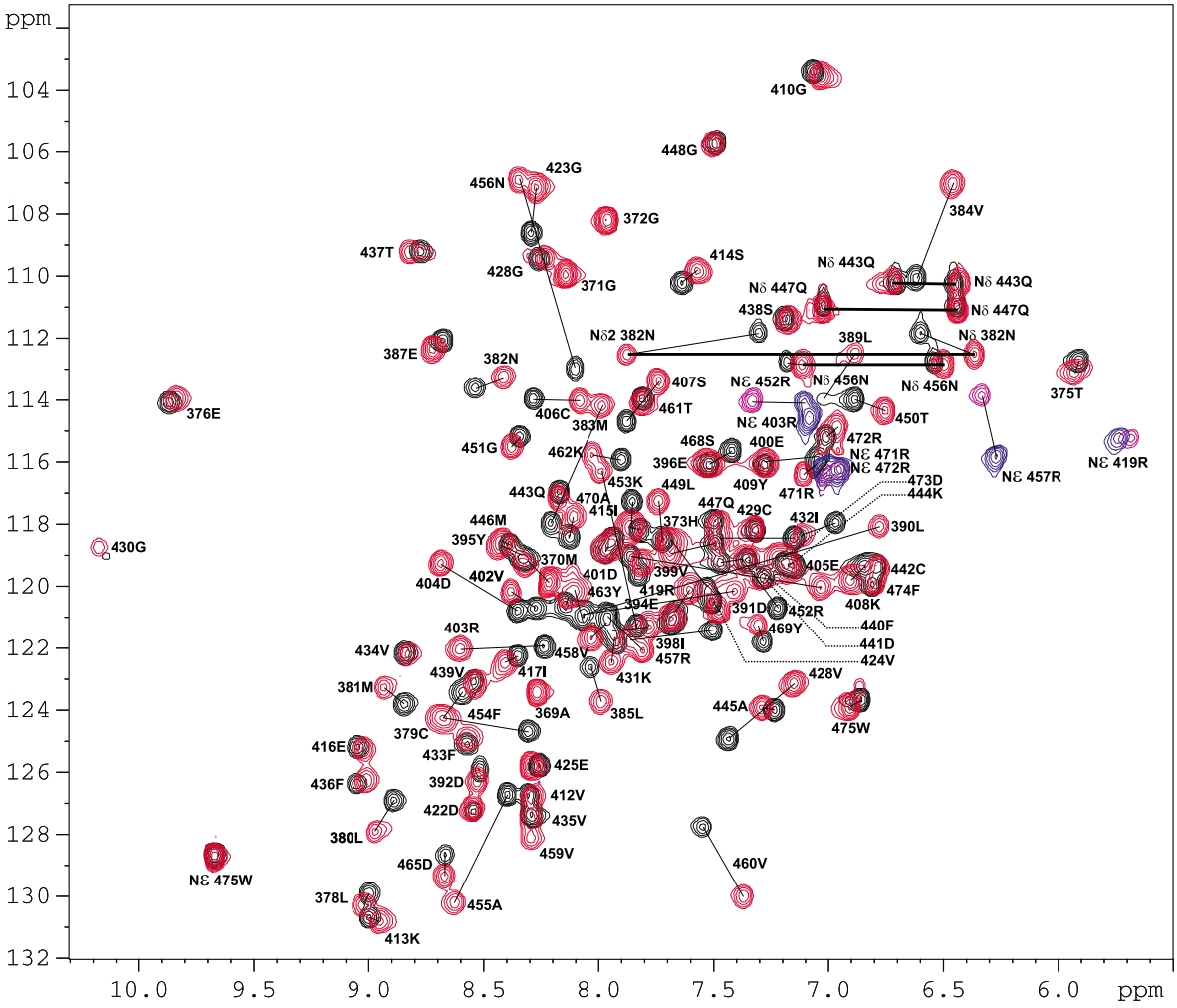


Figure 23:
Fully assigned ^{1}H, ^{15}N correlation spectrum of 0.5 mM of labeled, recombinant U2AF^{65}-RRM3 in the uncomplexed form (black cross-peaks) and upon addition of 2 mM of unlabeled, SF1 (aa 1-25) peptide (red cross-peaks). Each cross-peak corresponds to a backbone-amide group of the protein. Corresponding assignments are indicated.

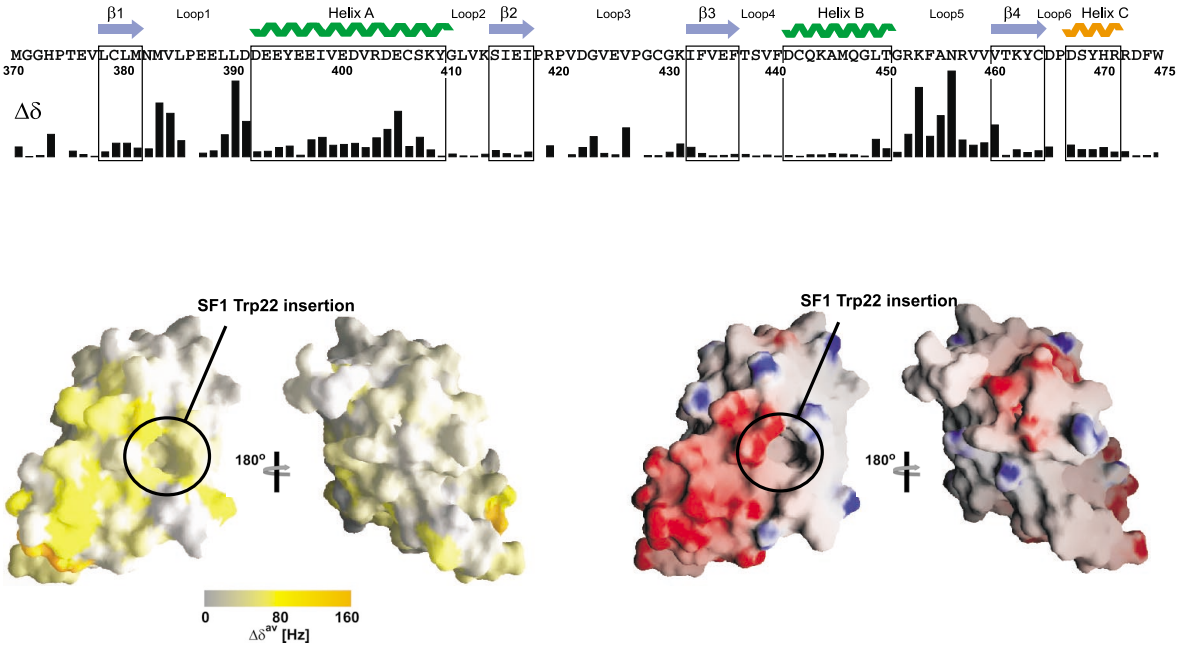


Figure 24:
Top panel: Quantitative representation of experimental results shown in Figure 23. Changes in chemical shifts for amide-groups in U2AF^{65}-RRM3 upon SF1 (1-25) peptide binding. Δδ values are shown for the primary sequence of RRM3 residues. Secondary structure elements are indicated and boxed. Regions that experience largest changes in chemical environment correspond to residues within Loop 1, parts of helix A and Loop 5.
Lower panel: A surface representation of the RRM domain without peptide is depicted below. Orientation of the molecule is the same as the NMR ensemble in Figure 21. Chemical shift changes upon SF1 binding are mapped onto the RRM surface and range from white to orange. The corresponding electrostatic charge distribution is outlined at the right. Surface areas involved in peptide binding overlap with conserved negatively charged surface regions. The hydrophobic pocket accommodating Trp22 of SF1is shown (see Chapter 2.2.3 and Figure 27 for details).

Chemical shift dispersion of ^{15}N-HSQC spectra of unbound, labeled SF1 (residues 1-25) displays characteristics of a random fold conformation with no secondary structure elements (Figure 25). Upon addition of unlabeled U2AF^{65}-RRM3, binding to SF1 is observed in the same slow exchange time regime as in the reciprocal experiment with labeled RRM3 (see above). The overall conformation of the now bound peptide remains randomly folded, no extended secondary structure elements are formed upon complex formation (see also Figure 26). Changes in chemical environment and hence differences in chemical shift, are most dramatic for the amide group of Arg21, which appears to shift downfield ~2 ppm in the proton dimension during complex formation. Also, the water exchangeable NHε of Arg21, unobserved in free SF1 (aa 1-25), becomes visible in the complex and thus slowed down by an order of magnitude in its solvent exchange rate. Furthermore, the NHε in the aromatic ring of Trp22 in SF1 undergoes a large change in chemical environment, whereas its amide group is only slightly shifted upon addition of protein (Figure 25). These results indicate that side-chain specific contacts

involving Arg21 and Trp22 of SF1 are instrumental for peptide-protein recognition. In general, changes in chemical shift for amide groups in SF1 are observed only for residues C-terminal to Ser14. This is in agreement with our previously stated observation that residues N-terminal to Lys15 in SF1 are not involved in U2AF^{65}-RRM3 interaction. Heteronuclear NOE data for complexed SF1 display negative values (i.e. mobile conformations) for residues 1-15 of SF1 and only indicate less flexible behavior (positive values) for amino acids 16-24, with Trp22 exhibiting the most dynamically restricted conformation (Figure 25). Residues directly participating in binding, or spatially close to the binding site, involve conserved positively charged amino acids immediately N-terminal to the invariant Ser20, Arg21, Trp22 in SF1 as judged by the chemical shift changes of HN frequencies of these amide groups. Hence, indicating the importance of the reciprocal electrostatic nature of charged residues in U2AF^{65}-RRM3 and SF1 for interaction.

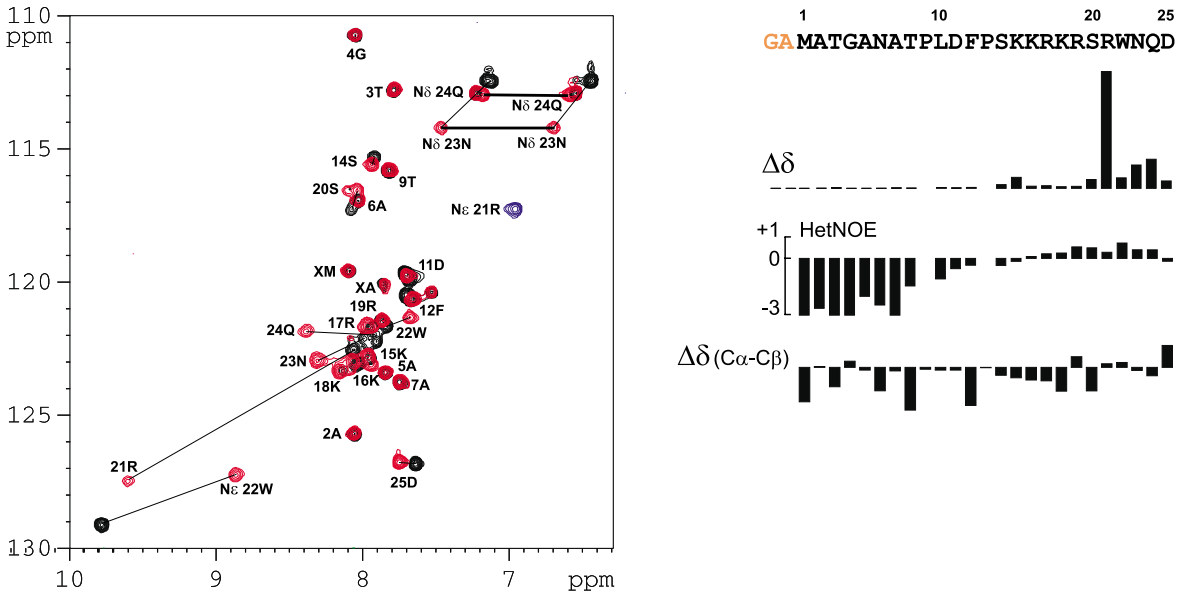


Figure 25:
Left: $^{15}N^{1}H$ Correlation spectrum at 295K of 0.5mM of labeled recombinant human SF1 (1-25); in the free form (black cross-peaks) and complexed with 2mM of unlabeled U2AF^{65}-RRM3 (red cross-peaks). Cross-peaks labeled XM and XA denote additional non-SF1 residues resulting from the TEV cleavage reaction.
Right, top: Sequence of the first 25 residues in human SF1 and changes in SF1 NH chemical shifts upon complex formation. Only residues C-terminal of Ser14 experience significant chemical shift changes. Heteronuclear NOE values of complexed SF1 (1-25). Only residues 16-24 display positive het-NOEs corresponding to dynamically restricted amino acids in the SF1 U2AF^{65}-RRM3 complex. A secondary chemical shift plot of Cα- and Cβ-SF1 resonances in the protein-peptide complex is shown below. Values are distributed evenly around the mean random-conformation size, indicating an extended conformation of SF1 (1-25) in complex with U2AF^{65}-RRM3 (see also Figure 26) with no secondary structure elements.

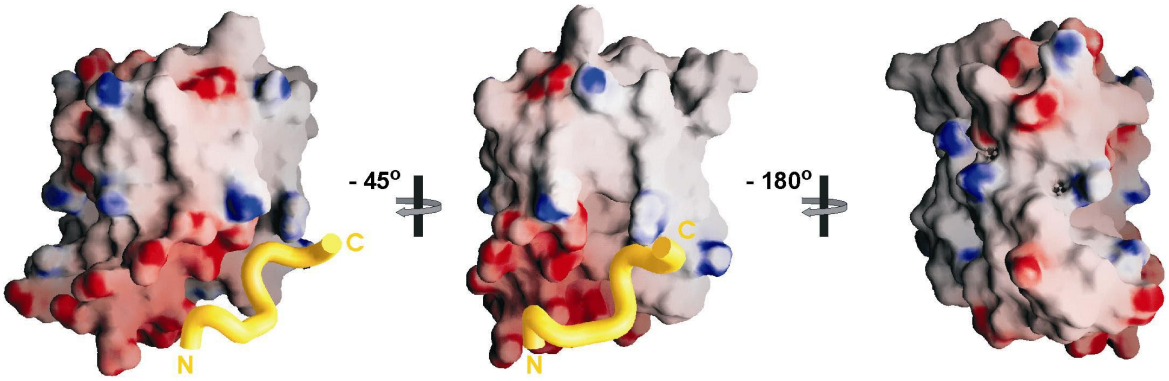


Figure 26:
Electrostatic surface representation of U2AF^{65}-RRM3 in complex with the SF1 peptide (aa1-25 and rotated views. Only residues 13-25 of SF1 are depicted as a backbone worm (yellow). The elongated random conformation of SF1 positions the positively charged N-terminal region immediately above the acidic RRM surface area. A more detailed picture and investigation of the requirements of these electrostatic complementarities for interaction is presented in Chapter 2.2.3, Figure 28.

2.2.3. Structure and Biochemistry of the U2AF^{65}-RRM3/SF1 Interaction

Interaction of human SF1 and U2AF^{65}-RRM3 involves the insertion of a conserved tryptophan residue of SF1 (Trp22) into a hydrophobic groove formed between the two RNP-characteristic α-helices (helix A and B) of human U2AF^{65}-RRM3 as the major 'anchor' for molecular recognition (Figure 27). It forms hydrophobic contacts to residues both at the 'inner' side of helix A (Val402, Glu405), a residue in an extended helical conformation following helix B (Leu449), two adjacent residues within Loop 4 (Arg452 and Leu459) and two residues at the 'base' of the hydrophobic core (Leu380 and Met383). Trp22 in the SF1 peptide shows most inter-molecular NOEs and appears as the most tightly coordinated residue by U2AF^{65}-RRM3.

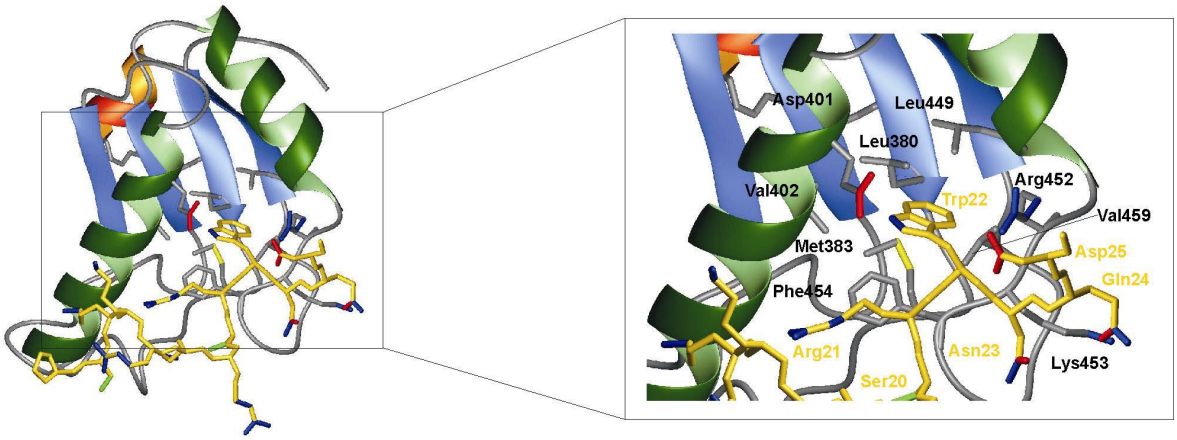


Figure 27:
Structural analysis of the hydrophobic contribution to the U2AF^{65}-RRM3/SF1 interaction. The ribbon representation on the left displays the overall coordination of the SF1 peptide (aa13-27) by the RRM domain of U2AF^{65}. The enlarged view on the right shows the coordination of the conserved Trp22 of SF1 within a hydrophobic pocket formed between helix A and B of the RRM domain. Side-chains of residues involved in binding and discussed in the text are outlined and labeled. Note, the similarity of this interaction with part of the heterodimeric U2AF^{35}-RRM/U2AF^{65} interaction discussed in Chapter 2.2.5 and shown in Figure 32.

In agreement with this, edited-filtered NOE experiments of labeled U2AF^{65}-RRM3 and unlabeled SF1 peptide yielded intermolecular NOEs only to Trp22, whereas intermolecular NOEs to other peptide residues were not observed. Residues in the SF1 peptide C-terminal to Trp22 (Asn23 and Gln24), display observable NOEs to side-chain hydrogen atoms of Lys453, whereas Asp25 of the peptide shows no inter-molecular NOEs and appears flexible based on its negative value in the heteronuclear NOE experiment (Figure 25).
The arginine residue preceding Trp22 (Arg21) in SF1 is structurally placed above the aromatic ring of Phe454 within Loop 4 of U2AF and thus experiences aromatic ring currents, explaining the great changes in chemical shift upon binding (Figure 25).

No intermolecular NOEs are observed for Ser20 in SF1, the side-chain is solvent-exposed and accessible for potential phosphorylation. A phosphorylated version of this side-chain would not be sterically restrained, suggesting that loss of binding upon phosphorylation may be due to unfavorable electrostatic behavior rather then to steric incompatibility (Wang et al., 1999). Basic residues N-terminal to this conserved triad of directly interacting amino acids, display contacts to acidic side-chains within the conserved region of negative electrostatic charges in the

beginning of helix A of U2AF^{65}-RRM3 (Figure 28). Due to the high degree of sequence redundancy, both for basic residues in the SF1 peptide and for acidic amino acids in this part of U2AF^{65}, the proton and carbon chemical shift dispersion is poor and unambiguous inter-molecular distance restraints cannot be assigned for these residues. Nevertheless, the directionality and orientation of SF1 binding to U2AF^{65}-RRM3, strictly places areas of opposite charge in close spatial proximity. This enforces the importance of inverse electrostatics at these positions for molecular interaction (Figure 28).

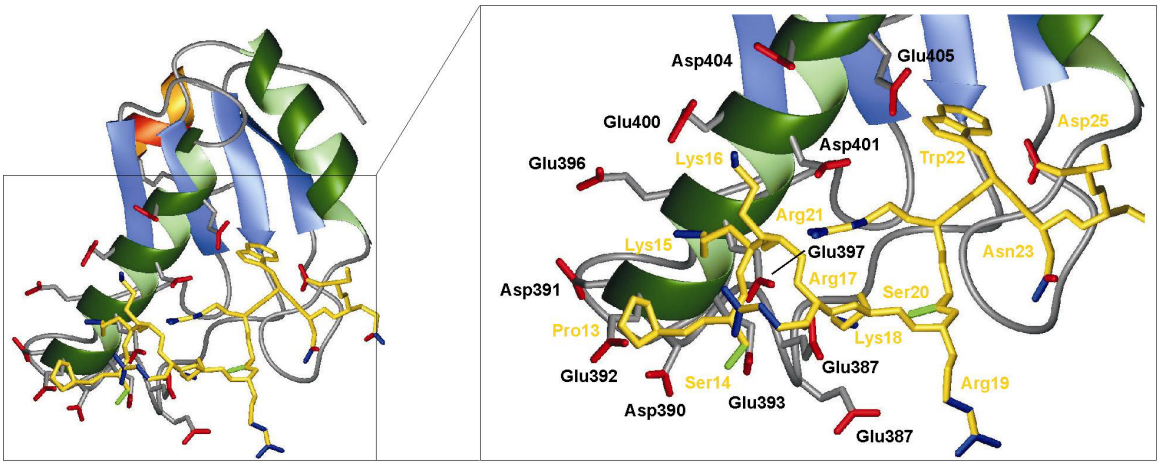


Figure 28:
Structural analysis of the electrostatic contribution to the U2AF^{65}-RRM3/SF1 interaction. Side-chains of negatively charged residues within Loop1 and along one surface side of helix A of the RRM domain are shown and labeled, indicating the close spatial proximity with positively charged amino acids of SF1. Note that this depiction only represents a tentative picture of this part of the structural arrangement of the U2AF^{65}-RRM3/SF1 interface, because no NOEs between this region of SF1 and the negatively charged residues within the RRM domain could be unambiguously assigned (see text).

To further test the requirement of opposed charges at the interface between U2AF^{65}-RRM3 and SF1 and to confirm our functional prediction for these residues, we collaborated with Prof. Angela Kramers group at the University of Geneva, Switzerland, who performed mutational experiments. Individual amino acids within U2AF^{65}-RRM3 and SF1 were replaced and the effect on binding experimentally probed (Figure 29). Introduction of double or triple amino acid replacements for conserved negatively charged residues within helix A of U2AF^{65}-RRM3 only abolished binding to SF1 when the electrostatic charge was inverted (i.e. changed from negative to positive). Binding was unaffected for neutral substitutions (negative to Ala). This effect was pronounced for multiple substitutions whereas

single charge reverting point mutations appear more tolerable (Figure 29, upper panel). Furthermore, changing the aromatic Phe454 within the binding pocket to Tyr (F454Y) had no effect on SF1 interaction, but changing this residue to Ala (F454A) reduced binding. A similar result was obtained probing conserved positive charges within the SF1 peptide. All residues, except Arg21, of the conserved basic stretch, tolerate single charge reverting mutations in their binding behavior to U2AF. The aromatic property of Trp22 is required for binding. A Trp to Phe (W22F) mutation is tolerated whereas a Trp to Ala (W22A) mutation abolishes binding to the RRM3 domain. Triple substitutions with negatively charged amino acids result in loss of binding whereas Ala mutants appear functionally intact. Point mutations of residues C-terminal to Trp22 had no effect on SF1 interaction (Figure 29, lower panel).

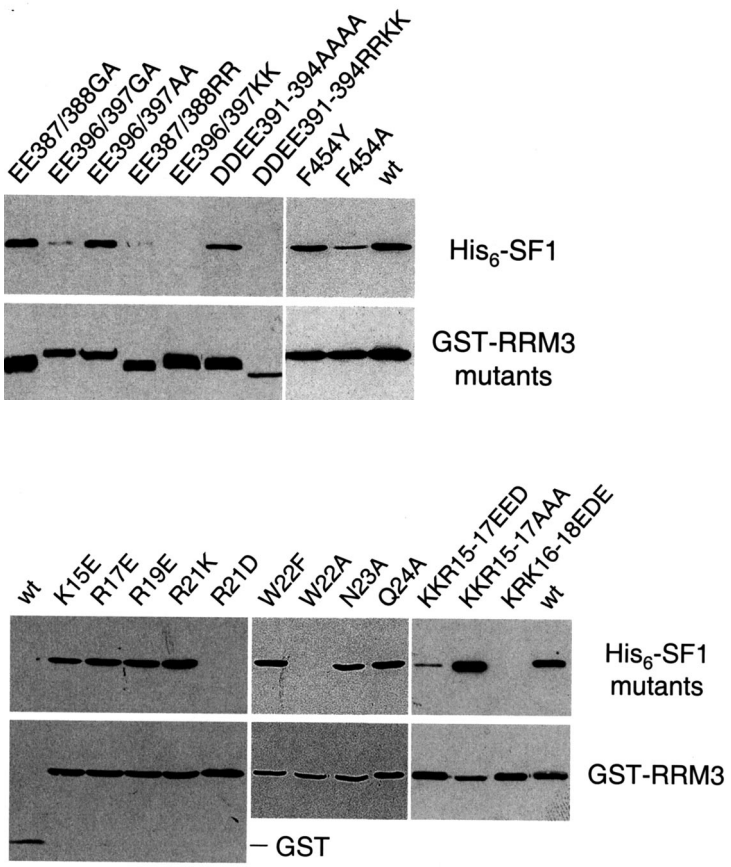


Figure 29:
Upper panel: Biochemical *in vitro* pull-down assay probing the effect of charge-reverting mutations within the conserved acidic stretch of residues of U2AF^{65}-RRM3 on SF1 binding (see text). Recombinant mutant versions of GST tagged U2AF65-RRM3 (aa 367-475) was bound to Glutathion-Sepharose and incubated with wild-type His-tagged SF1 protein (aa 2-320), separated by SDS-PAGE, transferred onto Nitrocellulose and Western-blotted with α-His and α-GST antibodies (see materials and methods).
Lower panel: Reciprocal biochemical pull-down experiments analyzing the requirement of basic residues of SF1 for binding to U2AF^{65}-RRM3. Wild-type GST-RRM3 (aa 367-475) was probed for interaction with mutant versions of His-tagged SF1 (aa 2-320). Experimental procedures as described above.

These results stress the importance of the collective electrostatic nature of these residues for interaction. A property which is favorable but not absolutely required for binding, as a more specific set of additional hydrophobic contacts is present in the U2AF^{65}-RRM3 SF1 molecular interface.

2.2.4. Helix C in U2AF^{65}-RRM3 and U1A

The solution structure of the third RRM of human U2AF^{65} reveals the presence of an additional secondary structure element at the very C-terminus of the protein. Hydrophilic residues (469-HRRD-473), flanked by invariant aromatic amino acids at the C-terminus of U2AF^{65} (Tyr469, His 470, Phe474, Trp475), form an extra helix (helix C). Helix C tightly contacts hydrophobic residues lined at the classical RNA binding surface of the β-sheet (Figure 30). Residues Tyr469, His 470 and Phe474 form an aromatic 'cage' that engulfs the β-surface protruding aromatic Phe433, which is an invariant RNP1-motif aromatic residue, most commonly employed for specific RNA contacts in RRM-RNA complexes.

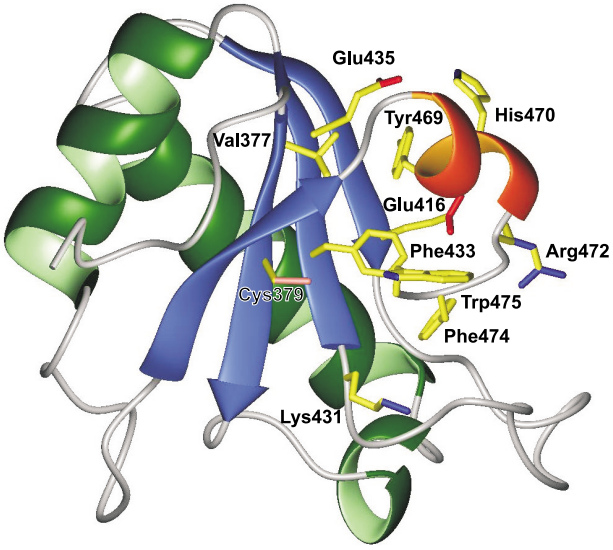


Figure 30:
Ribbon diagram of the U2AF^{65}-RRM3. Secondary structure elements are colored blue for the 4-stranded anti-parallel β-sheet, green for helix A and B and orange for the additional C-terminal helix C. Side-chains of aliphatic and aromatic residues hydrophobically coordinating helix C are drawn in yellow and labeled.

This structural arrangement appears tightly coordinated, capping a possible RNA interaction site, rendering Phe433 inaccessible for binding. Sequence conservation among all orthologues of U2AF^{65}-RRM3 indicates similar C-terminal extensions to the RNP-type fold for different organisms. Two other structural examples exhibit a similarly arranged helical extension to the characteristic RNP-type motif. The N-terminal RRMs of human U1A (Avis et al., 1996) and U2B'' (Price et al., 1998). For U1A, structural data is available for the free protein (Avis et al., 1996; Nagai et al., 1990) as well as for human U1A in complex with two different, natural RNA ligands. Stem-loop II of U1snRNA (Allain et al., 1997; Oubridge et al., 1994) and the U1A 3' UTR or PIE RNA element (for polyadenylation inhibition element)

(Varani et al., 2000). Whereas the crystal structure of the free form of the first RRM of U1A did not reveal the presence of helix C, the protein construct lacked residues important for helix formation (Nagai et al., 1990), an NMR derived solution structure of a longer U1A construct indicated the presence of the flexible additional α-helix (Avis et al., 1996). In the RNA free form, helix C of U1A packs against the 'front-side' of the β-sheet, the surface classically employed for RNA binding, and protects conserved RNP-type hydrophobic and aromatic residues from solvent exposure. U1A residues employed for helical coordination are Ile93, Ile94 and Met97, on the helical surface and Leu44, Phe56 and Ile58 within the RRM fold (Figure 31).

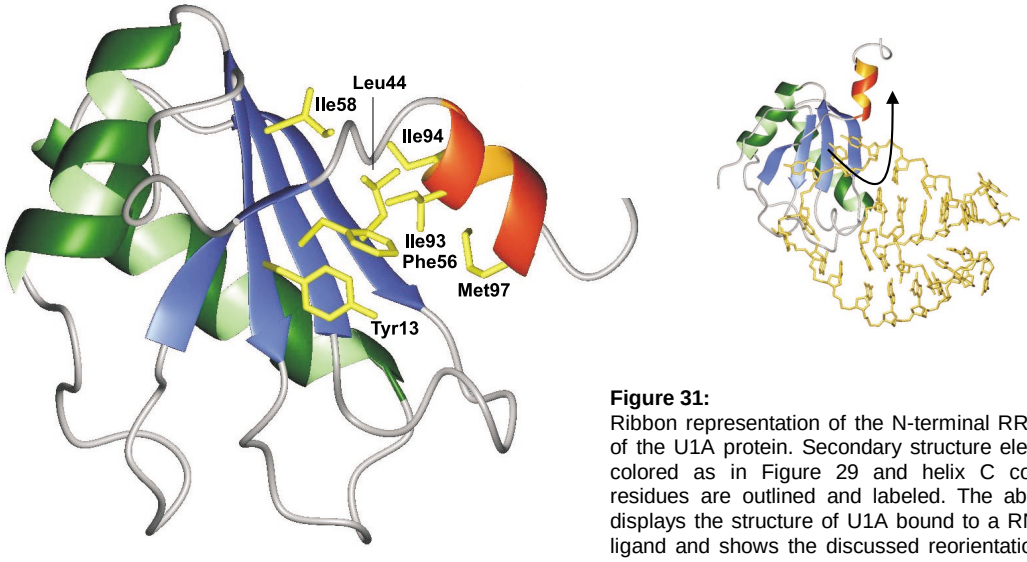


Figure 31:
Ribbon representation of the N-terminal RRM domain of the U1A protein. Secondary structure elements are colored as in Figure 29 and helix C coordinating residues are outlined and labeled. The above insert displays the structure of U1A bound to a RNA hairpin ligand and shows the discussed reorientation of helix C.

Both, the X-ray and NMR structure of an U1A-RRM1 complex with its cognate RNA ligand, stem-loop II of U1snRNA (Allain et al., 1997; Oubridge et al., 1994), indicate a dramatically altered orientation of helix C. To facilitate RNA binding, helix C has swung away from the β-sheet by ~130 degree, allowing aromatic stacking interactions and hydrophobic contacts of RNA moieties with β-surface amino acid residues. Furthermore, the recently determined solution structure of the N-terminal RRM of human U1A, in complex with its cognate RNA sequence of PIE (Varani et al., 2000), indicates a similar open conformation for helix C of the two molecules of U1A-RRM1 when bound to their closely spaced RNA stem-loop targets. Additional interactions of the two protein subunits by helix C mediated

protein-protein contacts are observed. Thus, in the case of human U1A, helix C not only provides a switch between the free and RNA bound form, but also functions as an interaction platform for homodimer formation upon RNA binding. The predicted helix C for the U2B'' N-terminal RRM was confirmed by the crystal structure of the U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA (snRNA) (Price et al., 1998). Again, an open conformation was observed for the protein RNA complex, with helix C swung away from the surface β-sheet RNA binding platform, allowing RNP specific nucleotide recognition.

For U1A and U2B'', protein sequences extend C-terminally to the additionally formed helix C. In the case of the third RRM of $U2AF^{65}$, the protein ends 4 residues after the last helical coordinated amino acid (Arg471). Helix C thus forms a structural 'cap' for the U2AF protein and allows incorporation of unusually hydrophobic residues at the C-terminus of this protein. Moreover, helix C appears much tighter coordinated in $U2AF^{65}$-RRM3 as compared to the RNA free form of U1A (compare Figure 30 and 31). Aromatic, helical residues Tyr469, His 470, together with hydrophobic contributions of Phe433, Lys431, Pro418 and Val377 form a closely spaced hydrophobic network, which is more sophisticated than for U1A. A similar re-orientation of helix C to accommodate cognate RNA binding, seems energetically less favorable for U2AF and whether it is actually performed remains to be determined. In this regard, the additional C-terminal helix C in $U2AF^{65}$ seems to present a conformational lock, assuring structural integrity of the RNP-fold and protecting β-surface protruding hydrophobic residues from solvent exposure, rather than a tunable switch to accommodate sequence specific cognate RNA recognition, like in the case of U1A and U2B''.

2.2.5. Comparison of $U2AF^{35}$/$U2AF^{65}$ with $U2AF^{65}$/SF1

U2AF consists of a large (~65kD, $U2AF^{65}$) and a small (~35kD, $U2AF^{35}$) subunit. $U2AF^{35}$ contains a single, central RRM domain (residues 43-146) which binds to a proline-rich sequence stretch within the N-terminus of $U2AF^{65}$ (residues 85-112). The recently solved crystal structure of the $U2AF^{35}$/$U2AF^{65}$ hetero-dimeric protein complex revealed a novel mode of RRM mediated peptide recognition (Kielkopf et al., 2001) (Figure 32). The RRM domain of human $U2AF^{35}$ employs the aromatic

side-chain of a tryptophan residue (Trp134) to specifically dock into a poly-proline pocket of an extended region in U2AF^{65}. U2AF^{65} reciprocally buries a conserved tryptophan (Trp92) within the hydrophobic fault in-between helix A and B of U2AF^{35}, at the protein surface opposite the 'classical' RNA binding site. The interaction surface is thus comprised of a double docking module, resulting in a two-fold 'tongue into groove' mode of reciprocal binding. Additionally, a conserved cluster of positively charged residues, upstream of the buried Trp92 within U2AF^{65}, is placed immediately adjacent to acidic residues located within the N-terminally extended first α-helix (helix A) of U2AF^{35}-RRM and although these residues appear flexible, no electron density can be observed for this stretch of amino acids in the crystal structure, it is likely that they also contribute to binding.

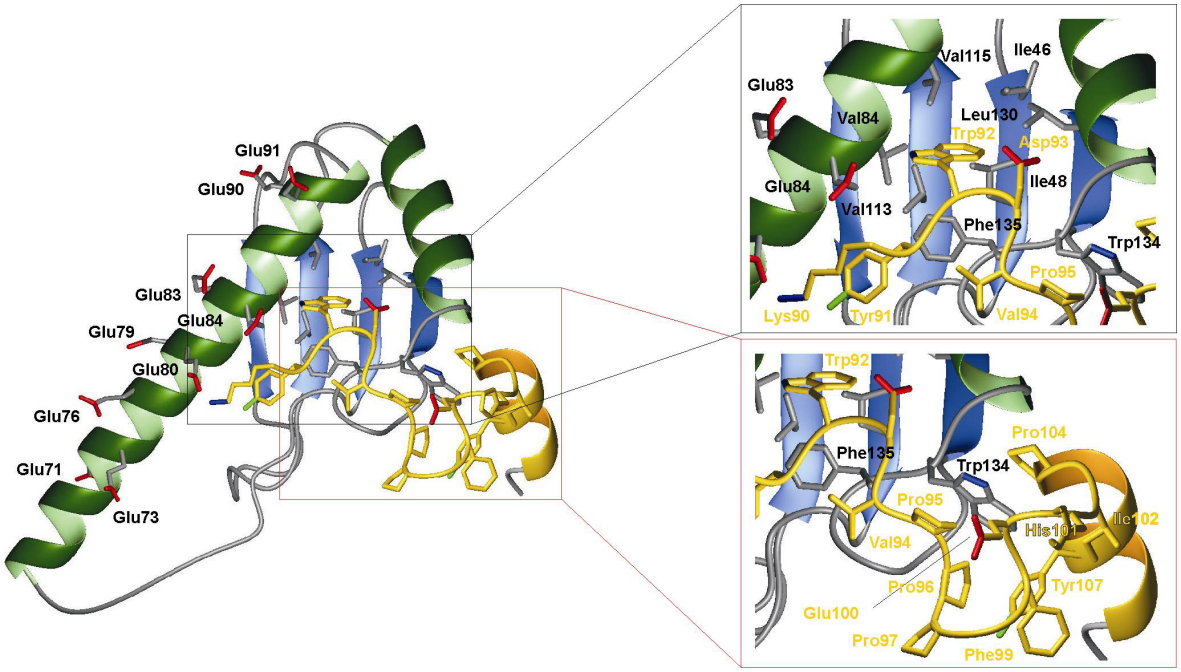


Figure 32:
Structural analysis of the U2AF^{35}-RRM with the proline-rich peptide of U2AF^{65}, corresponding to the minimal interaction region. A ribbon diagram of the U2AF^{35}-RRM is shown at the left and characteristic acidic residues the N-terminally elongated helix A are displayed. The U2AF^{65} peptide ligand is colored in yellow and side-chains of residues involved in binding are outlined. The double-docking modules (boxed in black and red) are enlarged and the reciprocal tryptophan insertions are shown. Side-chains of residues forming the respective binding pockets are labeled and their contribution to binding is discussed in the text. Note that one part of this interaction surface, the hydrophobic insertion of the conserved Trp92 of the U2AF^{65} ligand into the pocket formed between helix A and B (top insert) is structurally equivalent to the mode of interaction of U2AF^{65}-RRM3 and SF1 (see Chapter 2.2.3 and Figure 27).

The mode of interaction of this part of the binding interface closely resembles the mechanism employed by U2AF^{65}-RRM3 to interact with the N-terminus of SF1 (compare Figure 32 with Figure 27). For the first part of this interaction surface, the tryptophan residue within the U2AF^{35}-RRM, used for the proline pocket

recognition, does not have an aromatic residue at an equivalent position within U2AF^{65}-RRM3, as this residue is not absolutely conserved within RRM sequences (see Figure 33). Furthermore, no equivalent proline residues are present within close proximity in SF1, although a degenerate P-(X_2)-P-(X_3)-PP- motif appears later in sequence for some SF1 orthologues. It is questionable whether these could accommodate an aliphatic equivalent of U2AF^{65}-RRM3 because of their remote position in sequence and the absence of any solvent exposed hydrophobic residue on this part of the surface of the U2AF^{65} RRM. This first 'half' of the hetero-dimeric interaction domain appears to be unique and U2AF^{35}-RRM specific. No such structural elements are discernable in the U2AF^{65}-RRM3 SF1 interface.

The second 'half' of the U2AF^{35}-RRM U2AF^{65} specific hetero-dimeric interface however, is structurally equivalent to the U2AF^{65} SF1 interaction surface. SF1 buries its conserved Trp22 within the RRM fold of U2AF^{65} in a conformation which is almost identical to U2AF^{65}s own insertion of Trp92 within the RRM fold of U2AF^{35}. Structurally equivalent hydrophobic residues in both RRMs accommodate the respective tryptophan side chains. Moreover, phenylalanine residues (Phe35 in U2AF^{35} and Phe454 in U2AF^{65}, respectively), at identical positions of the RRM fold, adopt similar conformations in both RRMs, stabilizing the 'inserted' tryptophans in similar manners (compare Figure 32 and 27). The first helices (helix A) of U2AF^{35}-RRM and of U2AF^{65}-RRM3 are both N-terminally extended, although U2AF^{35} to a greater degree. They contain conserved negatively charged residues, which are aligned in close spatial proximity to basic amino acids of the respective ligands. A similar electrostatic contribution in peptide binding is thus present in these two examples of RRM mediated protein recognition.

In summary, U2AF^{35}-RRM and U2AF^{65}-RRM3 employ the same molecular surface for peptide coordination, opposite the classical RRM RNA binding platform. Secondly, both RRMs hydrophobically accommodate tryptophan residues of their respective peptide ligands in structurally equivalent positions and both interactions result in the favorable arrangement of conserved patches of opposed charges in close spatial proximity.

2.2.6. Electrostatic Properties of RRMs

RRMs represent the classical prototype for RNA binding domains. However, recent evidence suggests that RRMs can simultaneously bind to RNA and protein, as seen for the U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA (snRNA) (Price et al., 1998). In this case, RRM-mediated protein-protein interaction is necessary for specific RNA binding. U2B'' can not selectively interact with U2 snRNA in the absence of U2A' (Scherly et al., 1990a; Scherly et al., 1990b). The crystal structure of the $U2AF^{35}$-RRM/$U2AF^{65}$ interaction surface and work described here, provide examples of RRM-mediated protein-peptide interactions, with no RNA contribution (Kielkopf et al., 2001). Thus giving direct structural evidence for three possible RRM-mediated modes of target recognition. Selective and direct RRM-RNA recognition, RRM-protein interaction driven RNA interaction and RRM-protein binding in the absence of RNA. Are these ways of binding mutually exclusive? In other words, do certain RRMs solely bind protein, rather than RNA? Analyzing electrostatic properties of RRM motifs might provide a useful tool in answering this question.

In the case of $U2AF^{65}$, a protein construct comprising RRM1 and RRM2 (aa 148-232) can band shift poly-pyrimidine tract (PPT) RNA in a gel based mobility retardation assay. Isolated $U2AF^{65}$-RRM3 fails to show binding in such an experimental, both in a free and SF1 complexed form (L. Soares and J. Valcarcel, personal communication). NMR titration experiments, using ^{15}N-labeled $U2AF^{65}$-RRM3 and unlabelled PPT RNA confirm these biochemical experiments, as no significant changes in chemical shifts are observed upon addition of excess Py-tract RNA, both for free RRM3 and for RRM3 bound to SF1. These results demonstrate that $U2AF^{65}$-RRM3 is unable to interact with polypyrimidine tract RNA in the absence of its N-terminal two RNA binding domains.

Both, the $U2AF^{35}$-RRM and the third most C-terminal RRM in $U2AF^{65}$, exhibit a highly negative electrostatic potential over a large extent of their surface area. This electrostatic property is clearly unfavorably for RNA binding, as RNA is intrinsically negatively charged due to its phosphodiester backbone. At the same time, peptide ligand partners for the aforementioned RRMs both contain conserved patches of positive charge, which are structurally arranged in electrostatic most favorable

conformations in the respective complexes and a clear electrostatic contribution to binding is evident and has been confirmed for the U2AF/SF1 interaction (Chapter 2.2.3).

The presence of multiple acidic residues in between the first two secondary structure elements and within helix A of an RRM is not a rare feature of only a few RRM domains. A profile weighted BLAST search on an RRM containing protein database, yielded multiple examples for this kind of acidic amino acid conservation. A multiple sequence alignment of human proteins containing related negatively charged RRM domains is depicted in Figure 33.

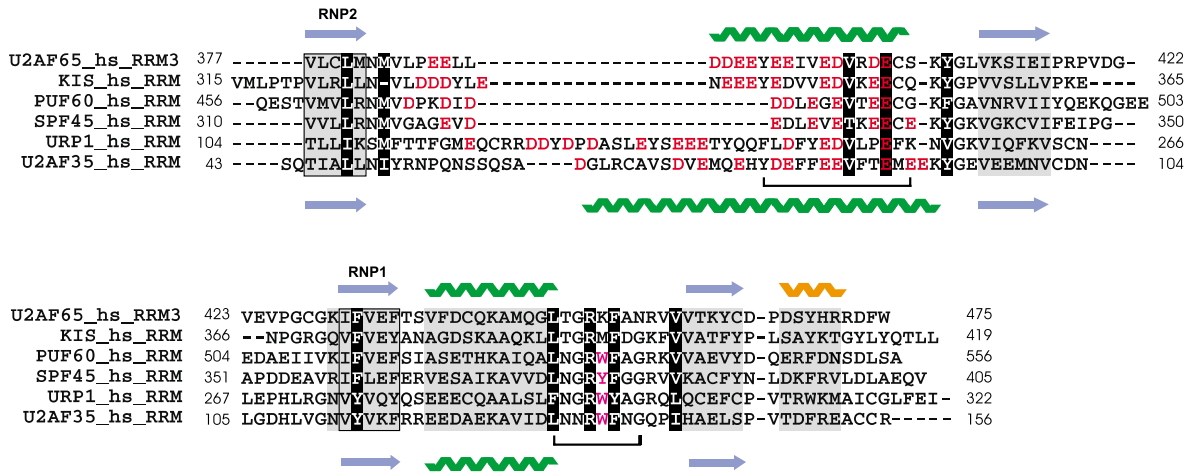


Figure 33:
Multiple sequence alignment of human proteins containing RRM domains with similar negative electrostatic characteristics as observed for U2AF^{65}-RRM3 and U2AF^{35}. KIS, a protein kinase with auto- and stathmin phosphorylating activity, acc.# XP_059194, PUF60, a poly-U binding splicing factor, acc.# AAF05605, SPF45, another splicing factor and component of the spliceosome, acc.# AAC64085, URP1, for U2AF^{35} related protein 1, acc.# BAA08532 and splicing factor U2AF^{35}, acc.# NP_006749. Only the first ~100 residues are shown. The respective RNP2 and RNP1 sequence motifs are indicated and experimentally determined secondary structure elements of U2AF^{65} (this work) and U2AF^{35} are drawn. Conserved residues of both U2AF^{65} and U2AF^{35} mediating ligand recognition are shown in white on black background. The Trp134 residue of U2AF^{35}, inserted into the poly-proline pocket of U2AF^{65} (Figure 31) is colored in pink. Gray boxes denote predicted secondary structure elements. Acidic residues within the first a-helix (helix A) are colored in red. Similar N-terminal extended helices A are predicted and may be expected for the URP1 RRM and, to a lesser extent, for KIS.

Molecular modeling of some of these RRMs, assuming ‘standard’ RNP-fold topology, and subsequent potential calculations, indicates similar ranges of negative electrostatic surface potentials for these selected RRM models. Analyzing domain organization of this set of proteins, shows great variability in modular composition (Figure 34). These highly negatively charged RRMs can occur as isolated domains, as seen in human SPF45 (Neubauer et al., 1998) and it’s

Arabidopsis homologue DRT111 (Pang et al., 1993) or in conjunction with other, more conventional RRMs like in the case of PUF60 (Page-McCaw et al., 1999) or in combination with RS motifs, like for U2AF^{35} and URP1 and URP2 (Tronchere et al., 1997). A serine/threonine/tyrosine kinase domain is found together with the RRM motif in the family of KIS related proteins (Maucuer et al., 1997).

It has been suggested that the PUF60-RRM3 functions in protein/protein interaction (Page-McCaw et al., 1999) and for the atypical RRMs in URP1 and URP2 (Tronchere et al., 1997). For human PUF60 the two N-terminal classical RRMs (RRM1 and RRM2) confer poly-pyrimidine RNA binding activity whereas the C-terminal negatively charged RRM3, located more than 140 amino acids downstream of RRM1 and RRM2, is essential for PUF60 homo-dimerisation (Page-McCaw et al., 1999).

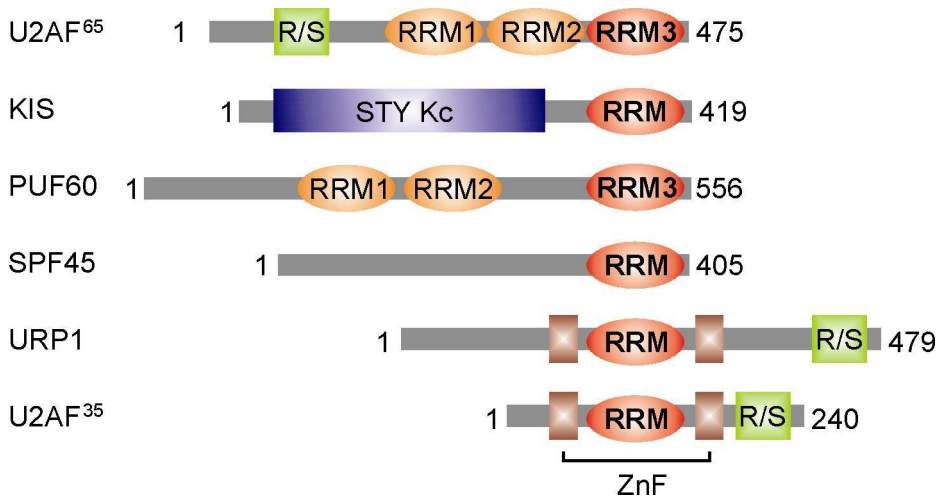


Figure 34:
Domain organization of the human proteins bearing atypical, negatively charged RRM domains (see text). RRM stands for RNA recognition motif, R/S for arginine/serine rich domain, STY Kc denotes a predicted serine/threonine/tyrosine kinase domain and ZnF indicates a zinc finger-like arrangement of cystein and histidine residues.

Some of the proteins bearing negatively charged RRMs are functionally connected. URP1 and URP2 have been shown to be associated with the U2AF^{65}/U2AF^{35} heterodimer (Tronchere et al., 1997). PUF60 strongly binds to another 'conventional' RRM containing splice factor, p54, which by itself interacts with U2AF^{65}, but not U2AF^{35} (Zhang and Wu, 1996). SPF45 has been co-purified

together with U2AF^{65} and U2AF^{35} (Neubauer et al., 1998) . Thus, most of these non-canonical RRM-containing proteins are present within the macromolecular assembly of the spliceosome. Whether they employ their characteristic atypical RNA recognition motifs to directly interact with nucleic acids or whether they have evolved their RNP type RNA binding domains into protein-protein interaction platforms to confer additional ligand binding specificities for cooperative target recognition, remains to be determined. At present, no clear evidence for direct RNA binding for any of these RRMs has been reported.

2.3. SUMMARY and CONCLUSIONS

In summary, the U2AF^{65}-RRM3 domains displays a characteristic RNP-type fold. It consists of a four-stranded anti-parallel β-sheet which is covered on one side by two α-helices (helix A and B). Additionally, the U2AF^{65}-RRM3 fold contains a short C-terminal helix (helix C) that packs against the β-sheet surface classically employed by RRM domains for RNA binding. U2AF^{65} interacts with the SF1 protein via this RRM domain and coordinates the conserved Trp22 residue in the N-terminus of SF1 within a hydrophobic pocket formed between helix A and B. It thus employs a surface area for SF1 coordination which is opposite the typical RNA binding site. Additionally, atypical acidic residues in the N-terminally elongated helix A of U2AF^{65}-RRM3 form electrostatic interactions with basic amino acids in the N-terminus of SF1. These electrostatic contributions are favorable but not absolutely required for U2AF^{65}/SF1 binding, as biochemical experiments have demonstrated. Furthermore, this example of a RRM-mediated protein coordination is similar to part of the hetero-dimeric U2AF^{35}/U2AF^{65} interaction. It is conceivable, that other atypical RRM domain containing proteins, employ the same mode of interaction for protein-protein binding.

2.4. PRESPECTIVES

It is interesting to ask how much of the U2AF^{65}/SF1 interaction interface is preformed in unbound U2AF^{65}. Does the free form of U2AF^{65}-RRM3 already exhibit the hydrophobic coordination site for SF1 binding or is it formed by a structural re-arrangement of helices A and B upon SF1 recognition? To answer this

question a more detailed structural analysis of the uncomplexed form of U2AF^{65}-RRM3 is necessary. We are in the process of solving the solution structure of free U2AF^{65}-RRM3 and hope to comparatively analyse structural re-arrangements upon SF1 binding.

2.5. REFERENCES

Allain, F. H., Howe, P. W., Neuhaus, D., and Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. Embo J 16, 5764-5772.

Antson, A. A. (2000). Single-stranded-RNA binding proteins. Curr Opin Struct Biol 10, 87-94.

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

Berglund, J. A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. Genes 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.

Conte, M. R., Grune, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., and Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. Embo J 19, 3132-3141.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98, 835-845.

Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. Genes Dev 13, 1102-1115.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex- lethal protein. Nature 398, 579-585.

Ito, T., Muto, Y., Green, M. R., and Yokoyama, S. (1999). Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)). Embo J 18, 4523-4534.

Kenan, D. J., Query, C. C., and Keene, J. D. (1991). RNA recognition: towards identifying determinants of specificity. Trends Biochem Sci 16, 214-220.

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.

Kramer, A. (1992). Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. Mol Cell Biol 12, 4545-4552.

Liu, Z., Luyten, I., Bottomley, M. J., Messias, A. C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. Science 294, 1098-1102.

Maucuer, A., Ozon, S., Manceau, V., Gavet, O., Lawler, S., Curmi, P., and Sobel, A. (1997). KIS is a protein kinase with an RNA recognition motif. J Biol Chem 272, 23151-23156.

Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. Nature 402, 838-841.

Nagai, K., Oubridge, C., Ito, N., Avis, J., and Evans, P. (1995). The RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. Trends Biochem Sci 20, 235-240.
Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990). Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. Nature 348, 515-520.

Nelson, K. K., and Green, M. R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev 3, 1562-1571.

Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nat Genet 20, 46-50.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994). Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature 372, 432-438.

Page-McCaw, P. S., Amonlirdviman, K., and Sharp, P. A. (1999). PUF60: a novel U2AF65-related splicing activity. Rna 5, 1548-1560.

Pang, Q., Hays, J. B., and Rajagopal, I. (1993). Two cDNAs from the plant Arabidopsis thaliana that partially restore recombination proficiency and DNA-damage resistance to E. coli mutants lacking recombination-intermediate-resolution activities. Nucleic Acids Res 21, 1647-1653.

Perez-Canadillas, J. M., and Varani, G. (2001). Recent advances in RNA-protein recognition. Curr Opin Struct Biol 11, 53-58.

Price, S. R., Evans, P. R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B"-U2A' protein complex bound to a fragment of U2 small nuclear RNA. Nature 394, 645-650.

Rain, J. C., Rafi, Z., Rhani, Z., Legrain, P., and Kramer, 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.

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., and Mattaj, I. W. (1990a). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. Nature 345, 502-506.

Scherly, D., Dathan, N. A., Boelens, W., van Venrooij, W. J., and Mattaj, I. W. (1990b). The U2B'' RNP motif as a site of protein-protein interaction. Embo J 9, 3675-3681.

Shamoo, Y., Abdul-Manan, N., and Williams, K. R. (1995). Multiple RNA binding domains (RBDs) just don't add up. Nucleic Acids Res 23, 725-728.

Tronchere, H., Wang, J., and Fu, X. D. (1997). A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. Nature 388, 397-400.

Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., and Varani, G. (2000). The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. Nat Struct Biol 7, 329-335.

Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P. J., Kramer, 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.

Wang, X., and Tanaka Hall, T. M. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. Nat Struct Biol 8, 141-145.

Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature 402, 832-835.

Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997). Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. Structure 5, 559-570.

Yuan, X., Davydova, N., Conte, M. R., Curry, S., and Matthews, S. (2002). Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. Nucleic Acids Res 30, 456-462.

Zamore, P. D., and Green, M. R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A 86, 9243-9247.

Zamore, P. D., and Green, M. R. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. Embo J 10, 207-214.

Zamore, P. D., Patton, J. G., and Green, M. R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature 355, 609-614.

Zhang, W. J., and Wu, J. Y. (1996). Functional properties of p54, a novel SR protein active in constitutive and alternative splicing. Mol Cell Biol 16, 5400-5408.

Zorio, D. A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. Nature 402, 835-838.

3. INTRODUCTION to Biomolecular NMR

Three experimental methods are generally employed for structure determination of biological macromolecules. X-ray crystallography, electron microscopy (EM) and nuclear magnetic resonance spectroscopy (NMR). NMR has a clear advantage over these other methods because it relies on biological macromolecules in solution and thus is the only method that investigates three-dimensional structures in a 'physiological' environment.

NMR spectroscopy relies on the quantum-mechanical property of the nuclear spin angular momentum (**I**). The spin angular momentum is a vector quantity with magnitude given by Equation 1.1

$$| \mathbf{I}^2 | = \mathbf{I} \cdot \mathbf{I} = \hbar^2 [I(I+1)] \qquad [1.1]$$

in which I is the angular momentum quantum number and $\hbar$ is the Planck's constant divided by 2π. The value of the z component of $\mathbf{I}_z$ is specified by Equation 1.2.

$$\mathbf{I}_z = \hbar \ m \qquad [1.2]$$

in which m is the magnetic quantum number. $m = (-I, -I+1, \ldots, I-1, I)$. Thus $\mathbf{I}_z$ has $2I+1$ possible values. The orientation of the spin angular momentum is quantitized because the magnitude of the vector $\mathbf{I}$ is constant and the z component has a set of discrete values.

Nuclei of atoms with odd mass numbers have half-integral angular momentum quantum numbers, nuclei with an even mass number and an even charge number have spin quantum numbers equal to zero and nuclei with an even mass number and an odd charge number have integral spin numbers. An overview of commonly occurring nuclei, their respective spin numbers, gyromagnetic ratios (γ) and natural abundances is given in Table 3, reproduced from Cavanagh 1997 (Chapter 3.7).

TABLE 3
Properties of Selected Nuclei[a]

Nucleus	I	γ (T · s)$^{-1}$	Natural abundance (%)
^{1}H	½	2.6752×10^{8}	99.98
^{2}H	1	4.107×10^{7}	0.02
^{13}C	½	6.728×10^{7}	1.11
^{14}N	1	1.934×10^{7}	99.64
^{15}N	½	-2.712×10^{7}	0.36
^{17}O	5/2	-3.628×10^{7}	0.04
^{19}F	½	2.5181×10^{8}	100.00
^{23}Na	3/2	7.080×10^{7}	100.00
^{31}P	½	1.0841×10^{8}	100.00
^{113}Cd	½	5.934×10^{7}	12.26

[a] The angular momentum quantum number, I, and the gyromagnetic ratio, γ, and natural isotopic abundance for nuclei of particular importance in biological NMR spectroscopy are shown.

Since NMR relies on the existence of nuclear spin, atoms with a spin number of zero are NMR inactive. Nuclei with spin quantum numbers of greater than ½ also possess electric quadrupole moments arising from non-spherical nuclear charge distributions. Their NMR resonance lines are correspondingly broad and unfavorable for analysis. In general, only nuclei with spin quantum numbers of ½ are employed for biomolecular NMR. These include ^{1}H, ^{15}N and ^{13}C, which are the preferred observable spins in NMR. Unfortunately, the natural abundance of the ^{15}N and ^{13}C nuclei is low (0.36% and 1.1%, respectively), so that isotope enrichment is necessary. Today, this is routinely achieved by recombinant expression technologies in labeled growth media (see materials and methods). The spin angular momentum (**I**) is proportionally linked to the magnetic dipole moment of a nucleus ($\boldsymbol{\mu}$). See Equation 1.3.

$$\boldsymbol{\mu} = \gamma \mathbf{I} \qquad \mu_z = \gamma I_z = \gamma \hbar m \qquad [1.3]$$

The magnetic properties of atomic nuclei ($\boldsymbol{\mu}$) align with the main axis of an external magnetic field $\boldsymbol{B_0}$ resulting in a preferential orientation of the net spin angular momentum either parallel or anti-parallel to this axis. For conventional liquid state NMR spectrometers this corresponds to the main vertical axis of the magnet and is by definition referred to as z-axis. In thermal-equilibrium all spins of a given sample exert a precession motion about the z-axis (Figure 35)

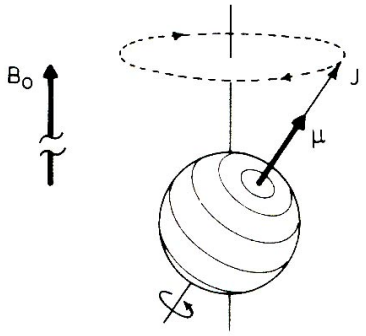


Figure 35:
Schematic representation of the precession movement which the magnetic dipole moment of an atomic nucleus with *I>0* exerts about an axis parallel to $\mathbf{B}_0$

The frequency of precession ($\boldsymbol{\omega_0}$) (termed *Larmor frequency*) is proportional to the strength of the external magnetic field ($\boldsymbol{B_0}$) as seen from Equation 1.4.

$$\omega_0 = -\ \gamma\ B_0 \qquad [1.4]$$

Thus, at a certain external magnetic field each nucleus (^{1}H, ^{15}N, or ^{13}C) precesses with a characteristic frequency. More precisely, the *resonance frequency* of a given spin is proportional to the *effective* magnetic field that this spin experiences. The *effective* magnetic field ($\boldsymbol{B}_{\text{eff}}$) depends on the external magnetic field ($\boldsymbol{B}_0$) and the local magnetic (i.e. chemical) environment ($\boldsymbol{B}_{\text{local}}$), also termed secondary magnetic field, of a given spin system. Hence, Equation 1.5,

$$\boldsymbol{B}_{eff} = \boldsymbol{B}_0 + \boldsymbol{B}_{local} \qquad \boldsymbol{B}_{eff} = (1-\sigma_{kk})\boldsymbol{B}_0 \qquad [1.5]$$

in which $\boldsymbol{\sigma_{kk}}$ is the second-rank nuclear shielding tensor. This local magnetic contribution is mainly governed by the electronic magnetic moments of neighboring atoms. Hence, each spin displays a slightly different frequency of precession, which directly corresponds to different local chemical environments and gives rise to a phenomenon called *chemical shift*. In practice, chemical shifts are measured in parts per million (ppm or $\boldsymbol{\delta}$) relative to a reference resonance signal from a standard molecule. Equation 1.6,

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 = (\sigma_{r_{ef}} - \sigma) \times 10^6 \qquad [1.6]$$

in which Ω and Ω_{ref} are the offset frequencies of the signal of interest and the reference signal, respectively, ω_0 is the *Larmor frequency* and σ is the average, isotropic shielding constant for the nucleus.

Chemical shift frequencies of a specific spin system therefore contain structural information about local geometries and electronic properties. Ideally, each observable nucleus in a given sample is unambiguously characterized by its chemical shift property. To investigate the interaction of various spin systems in a protein sample, chemical shift information directly correlates the identity of one nucleus to another. It is therefore a first step in NMR spectroscopy to assign chemical shift frequencies.

In general, three physical phenomena are exploited in structure determination by NMR. *Chemical shift* (see above), *scalar coupling,* and *cross-relaxation*, which gives rise to the nuclear Overhauser effect (*NOE*). These three parameters will be briefly outlined.

3.1. Chemical Shift, Scalar Coupling and Nuclear Overhauser Effect (NOE)

Chemical Shift: Differences in precession frequencies (chemical shift) for spins of the same kind are intrinsic physical properties which are always present. To develop these properties into observable magnitudes, equilibrium magnetization (along the *z*-axis) is flipped into the transverse plane (perpendicular to the initial alignment axis, i.e. the *xy*-plane of such a coordinate system). To tilt equilibrium magnetization from *z*-axis alignment (*longitudinal magnetization*) into *xy*-magnetization (*transverse magnetization*), a *radio-frequency (RF)* pulse of a certain strength and duration (corresponding to the desired *flip-angle*) is applied. This radio frequency pulse functions as a temporary external magnetic field ($\boldsymbol{B_1}$), which acts perpendicular to the static field of the NMR magnet ($\boldsymbol{B}_0$) and selectively 'flips' magnetization along this new magnetic field vector (into the *xy*-plane) (Figure 36). Transverse magnetization precesses in the *xy*-plane for the time *(t)*. During this time, chemical shift frequencies evolve into observable differences in precession frequencies. Precession motions now describe a rotation in the *xy*-plane, with different radial frequencies for individual spin systems. These are recorded by a detector in the *xy*-plane as an ensemble of time domain amplitudes

or *free induction decays (FIDs)* and eventually de-convoluted by *Fourier transformation (FT)* into frequency domain signals. Once the radio frequency pulse is switched off, magnetization also *relaxes* back to the equilibrium state (i.e. along the z-axis). This is the basic principle underlying every pulse-sequence scheme in NMR spectroscopy.

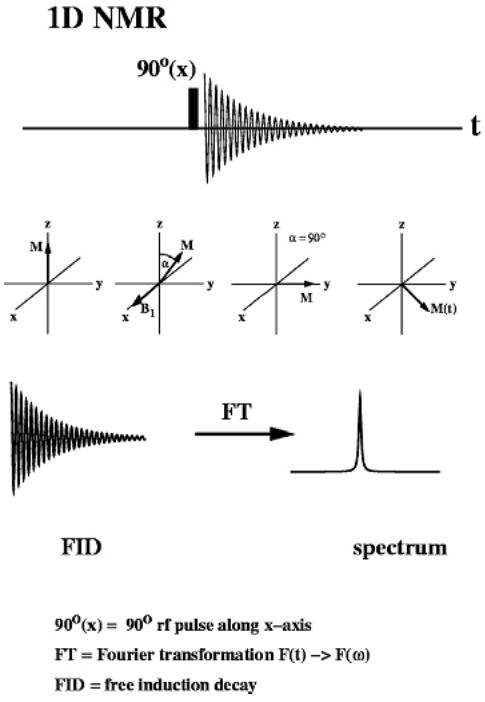


Figure 36:
Description of a single 90° pulse NMR experiment. Equilibrium magnetization along *z* (M) is flipped by a 90° pulse along *–x* to yield transverse magnetization along *y*. Transverse magnetization precesses about the *xy*-plane and induces the observed time-domain FID signal in the detector. Fourier Transformation of the time-domain signal yields the characteristic frequency domain NMR signal (Reproduced from Cavanagh, 1997).

Scalar Coupling: Another principle that is generally employed in NMR spectroscopy is the correlation of different spin systems via *scalar couplings* (or *J-couplings*). Nuclei that are connected by chemical bonds can exhibit spin-spin couplings which are mediated by electrons forming the chemical bonds between nuclei. A coupled two spin system ***(IS)*** displays four energy levels corresponding to possible combinations of spin states. Two of which display parallel orientations (up-up or down-down, $\beta\beta$, $\alpha\alpha$) and two with anti-parallel orientation (up-down and down-up, $\beta\alpha$, $\alpha\beta$). Observable transitions obey the selection rule $\Delta m=\pm 1$ (where m is the magnetic quantum number of each state) and result in two signals at the respective resonance frequencies of the $\boldsymbol{I}$ and $\boldsymbol{S}$ spins. Coupling of these nuclear spin states to the bonding electron spin states (occupied based on the *Pauli-principle* by an equal superposition of the 'down-up' and 'up-down' spin state)

results in an interaction with the anti-parallel αβ, βα nuclear spin states leading to a decrease or increase of the respective energy levels and an unfavorable interaction with the parallel αα and ββ states resulting in an increase or increase in these energy levels. The respective transitions are now modulated by these differences in energy levels and hence, each of the initially observed signals is split into two peaks with inter-peak distances corresponding to the magnitude of the *J-coupling* (see Figure 37 for detail).

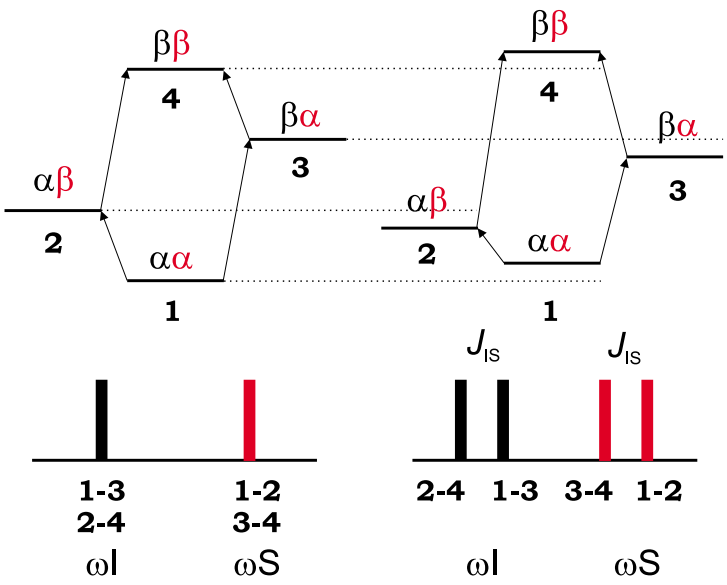


Figure 37:
Energy level diagram of an *IS* spin system. The four possible energy states of an uncoupled IS system are indicated at left, whereas the same *J*-coupled spin system is indicate at the right. Allowed transitions (obeying the selection rule $\Delta m=\pm 1$) are indicated by arrows and yield the respective spectra outlined below. Transitions of the **I** spin are shown in black, transition of the **S** spin are depicted in red. Note that upon *J*-coupling, the energy difference of the **I** spin specific 1-3 transition becomes smaller and the 2-4 transition larger, by the same magnitude. Similarly, the energy difference for the **S** spin specific 1-2 transition decreases and increases for the 3-4 transition. Hence, identical splittings in resonance frequencies are observed for both signals.

J-couplings are given in Hz and values are large (~35-140 Hz) for nuclei separated by a single bond (^{1}*J-coupling*) and become small (~0-10 Hz) for two- and three-bond couplings (^2J, ^3J). An overview of characteristic coupling constants for a ^{13}C, ^{15}N labeled protein backbone is given in Figure 38. For *J-coupling* to be active, one component of the coupled spin systems needs to contain transverse magnetization. To *refocus* scalar couplings, the energy levels of a coupled spin system are 'inverted' (usually by applying a 180° pulse on one of the components of the coupled spin system), *J-coupling* evolves with opposite sign and hence cancels. When transfer of magnetization from one nuclei to another is desired, scalar coupling is allowed to evolve. Therefore, *J-couplings* serve as a fundamental mechanism to transfer magnetization between spins in NMR spectroscopy.

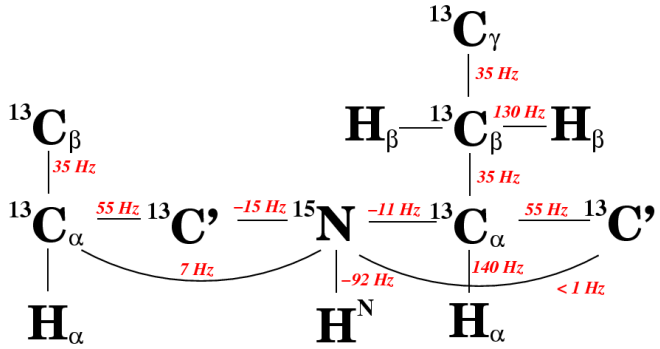


Figure 38:
Overview of one- and two-bond (*^1J, ^2J*) values of scalar coupling rates for a poly-peptide backbone and Cα, Cβ and Hα, Hβ spins. *J-coupling* values are given in Hertz (Hz) indicating the inverse proportionality of this physical property with time. Hence, for small J-couplings to be active, a longer period of time is necessary for coupling evolution, whereas for large values of *J*, short delays are sufficient.

Once resonance frequencies (chemical shifts) of all spin systems in an appropriately labeled biological sample haven been correctly assigned by respective NMR experiments, one tries to obtain geometrical restraints (distance and angular restraints). This procedure will eventually yield the three-dimensional structure of the biological molecule under investigation. Therefore, distance information for non-bonded spin systems and this information is obtained by NMR experiments that exploite the nuclear Overhauser effect (NOE).

NOE: The NOE describes cross-relaxation via 'classically' forbidden double- and zero-quantum transitions of uncoupled (dipole-dipole) spin systems. These transitions are not excited by radio frequency pulses or are directly detectable NMR signals but result as a consequence of *longitudinal-relaxation* and are known as *cross-relaxation* pathways. In short, resonance frequencies of one kind ***S***, perturbed by saturation or inversion, affect the net intensities of resonance frequencies of another kind ***I*** (Figure 39).

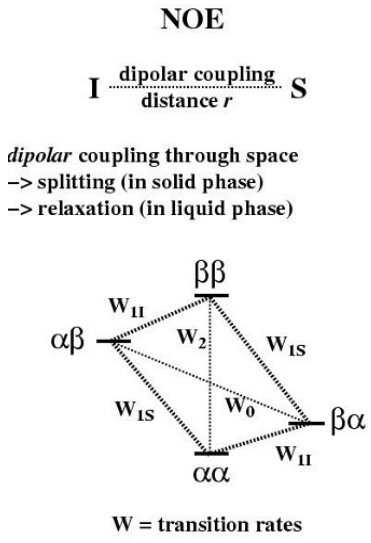


Figure 39:
The NOE is a result of spin-spin cross relaxation via W_2 and W_0 transition rates between two spins I and S. The NOE is inverse proportional to the distance between the two interacting spin. NOE intensity ~ $1/r^{-6}$
This distance dependence of the NOE is used to derive proton-proton distance restraints for the structure determination by NMR.

The importance of the NOE effect lies primarily in the fact that resonances which change their intensities are due to spins *close in space* to those directly affected by the perturbation. If the normal intensity of a resonance at thermal equilibrium and without perturbing the system is $\boldsymbol{I}_0$, then the intensity observed while saturating some other resonance is $\boldsymbol{I}$, the NOE enhancement $\boldsymbol{f_I\{S\}}$ is defined as the fractional change in the intensity of $\boldsymbol{I}$ on saturating $\boldsymbol{S}$, as seen from Equation 1.7.

$$f_I\{S\} = \frac{(I - I_0)}{I_0} \qquad [1.7]$$

Intensities of NOE cross-peaks are proportional to the inverse of the sixth power of the inter-nuclear distance (r^{-6}) and can thus yield distance restraints for protons coupled < 6 Å.

3.2. Resonance Assignment of ^{15}N, ^{13}C labeled Proteins

Modern NMR techniques for protein structure determination, rely on the availability of uniformly isotopically labeled protein samples. Labeling is achieved by recombinant expression of protein constructs in organisms grown in isotope-enriched media which results in uniform incorporation of ^{15}N or ^{15}N/^{13}C. An methodological overview is given in the materials and methods section.

In the next few paragraphs various assignment strategies that depend on the respective isotope labeling scheme will be introduced, basic building blocks and experimental aspects of the implementation of multidimensional triple resonance experiments will be discussed. The nomenclature for these experiments reflects the magnetization transfer pathways. Involved nuclei form the name of an experiment. Spins, whose chemical shifts are not evolved are put in parentheses. For an out-and-back type experiment, where magnetization of a spin is transferred to a remote spin and then brought back the same way, only the first half of the transfer is used for the name. For example, the out-and-back experiment that transfers magnetization from the amide proton (H^N) via the amide nitrogen (N) to the carbonyl C' (CO) is called HNCO. If another magnetization transfer step to the C_α (CA) of the previous residue is included and the corresponding C_α chemical shift is recorded, the experiment is referred to as H(N)COCA. The parentheses

indicate that magnetization is only transferred via the nitrogen spin, without chemical shift evolution taking place. Pulse schemes are depicted in pulse-sequence diagrams. *RF-pulses* for each nucleus are written on separate lines and pulses for each spins in the spin system are depicted as black bars on these lines. Thickness of the bars denote flip-angles induced by the respective pulses, usually 90° and 180° and their spacing corresponds to specific time delays between consecutive pulses. Pulse phases are indicated above the respective pulses. See Figure 41 for an example. Pulse sequence diagrams are read from left to right and all spin systems are considered simultaneously. Delays during which chemical shift evolution takes place are denoted *t1*, *t2*, and *t3*. Product operators describing states of coupled spin systems are given at selected points during the pulse sequence. Many of the discussed pulse schemes contain sensitivity-enhanced back transfers combined with heteronuclear gradient echoes and water-flip-back pulses. These are NMR technical details important for signal-to-noise improvement and water suppression and will not be discussed further in the text (Chapter 3.7).

3.3. NMR Experiments for Backbone Assignment

A number of triple-resonance NMR experiments are available for backbone chemical shift assignment of uniformly ^{13}C, ^{15}N labeled protein samples. A basic set of three of these experiments will be outlined and these allow to obtain connectivities between the amide groups of the backbone with C_α and C_β nuclei of the respective amino acid side-chains (Figure 40).

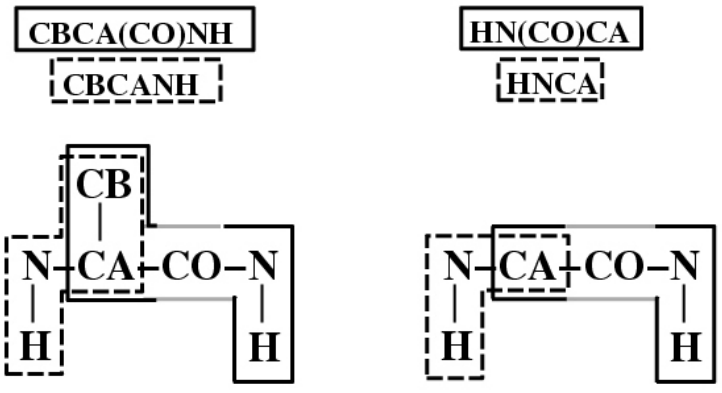


Figure 40: Summary of NMR experiments for assignment of protein backbone resonance frequencies. Correlated spin-systems are boxed for the respective experiments, spin-systems which are employed to transfer magnetization without chemical shift evolution are outlined in gray and indicated in brackets in the respective pulse-sequence names.

Common to these experiments is the initial transfer of magnetization from protons to nitrogen spins by a short pulse sequence element termed INEPT (insensitive nuclei enhanced by polarization transfer). Nitrogen chemical shift evolution is recorded and stored to be used as the second dimension of the spectrum. Correlation frequencies corresponding to nitrogen atoms coupled to a proton nucleus will eventually identify the amide groups. Subsequently, magnetization is jointly transferred via ^{1}J or ^{2}J couplings to carbon atoms (either only C_{α}, in the case of the HNCA- or to both C_{α} and C_{β} nuclei, for the HNCACB-experiment) of the same *(i)* or the preceding residues *(i-1)*. Carbon chemical shift is evolved and recorded resulting in the final third dimension of this 3D experiment. Magnetization is transferred back to protons by the same route it was originally built up (out-and-back-type of transfer). Proton chemical shift is recorded to yield the proton dimension in the three dimensional reconstruction. Hence, the final spectrum will correlate nitrogen nuclei coupled to proton atoms in two dimensions with the coupled carbon atoms of the same and the preceding residues in the third dimension. Therefore, three resonance frequencies can be assigned to any cross-peak observed in the spectrum, yielding initial chemical shift values for ^{1}H and ^{15}N of all amide-groups and ^{13}C frequencies of the corresponding C_{α} and C_{β} atoms.

3.3.1. The 3D HNCO and 3D HNCA Experiment

The pulse sequence for the 3D HNCO and 3D HNCA experiments are identical, except for the interchange of C_{α} and C' pulses. The HNCO/HNCA pulse sequence(s) is displayed in Figure 41, basic sequential combinations of pulses or 'building blocks' are indicated. This served as a simple example of a typical pulse sequence scheme for a 3D NMR experiment.

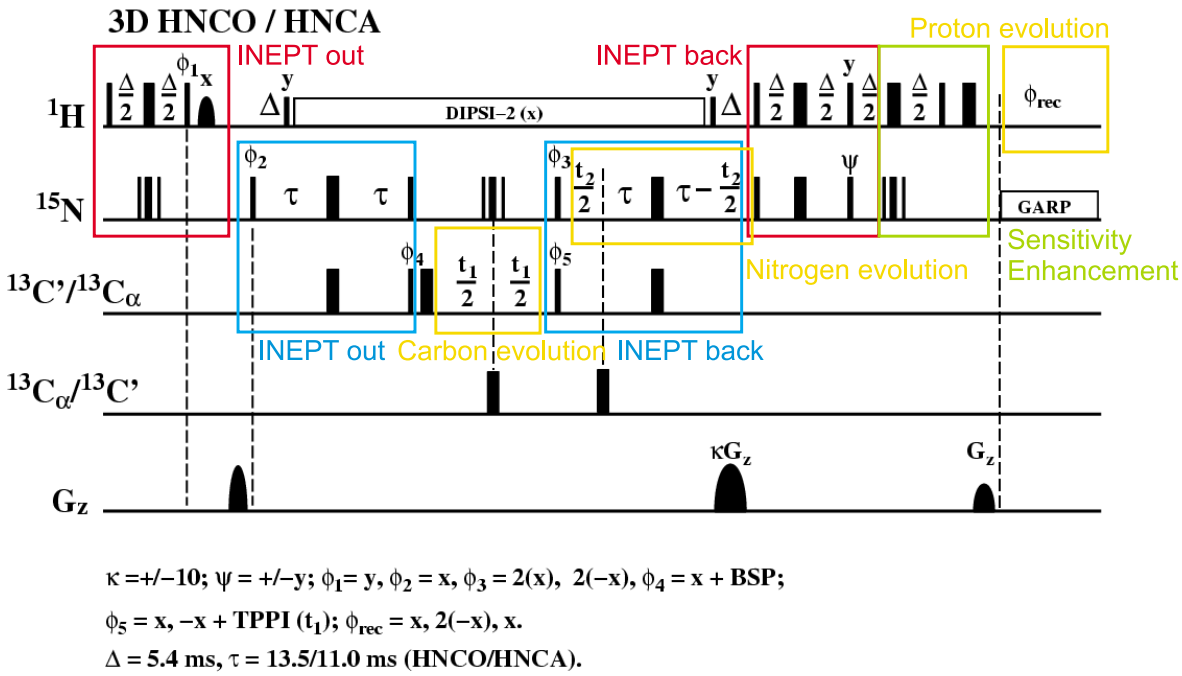


Figure 41:
Pulse sequence diagram of the 3D HNCO or the 3D HNCA triple resonance NMR experiment. INEPT pulse sequence blocks are employed to transfer magnetization from proton to nitrogen and from nitrogen to proton (red boxes) and from nitrogen to carbon and back to nitrogen (blue boxes). Carbon chemical shift (either Cα or C') is evolved during t1 (first yellow box), with composite pulse decoupling for nitrogen and a 180° decoupling pulse for carbon (either Cα or C'). During the INEPT back-transfer from carbon to nitrogen, nitrogen chemical shift is evolved during t2 (second yellow box). Finally, magnetization is transferred back to protons by a sensitivity-enhanced version of the INEPT transfer, proton chemical shift evolves and is recorded during t3, or Φrec. Solvent suppression is achieved by a H_2O selective water flip-back pulse after the second proton 90° hard-pulse and decoupled during carbon and nitrogen chemical shift evolution by the DIPSI decoupling sequence. Nitrogen is decoupled during proton evolution by a GARP decoupling sequence. Gradient pulses along z are depicted on a separate line. Pulses which are phase-cycled are denoted by respective coefficients, defined below the pulse sequence diagram.

Two types of cross-peaks for each amide group are observed in the HNCA experiment, namely the correlations $H^N(i)$, $N(i)$, $C_\alpha(i)$ and $H^N(i)$, $N(i)$, $C_\alpha(i-1)$ because both couplings ${}^1J_{C\alpha(i),N(i)}$ = 11Hz and ${}^2J_{C\alpha(i-1),N(i)}$ = 7Hz are of similar size. In contrast, only the correlation $H^N(i)$, $N(i)$, $C'(i-1)$ are observed in the HNCO, since ${}^1J_{C'(i-1),N(i)}$ = 15Hz, while ${}^2J_{C'(i-1),N(i)}$ ~ 0Hz. Thus, the HNCA experiment yields two cross-peaks for any amide NH position, one for the same residue *(i)*, which because of the large coupling exhibits high intensity and one for the previous amino acid *(i-1)*, which, due to the smaller coupling displays weak intensity. The HNCO experiment only displays one cross peak which correlates every amide group to the carbonyl of the previous residue (Figure 42).

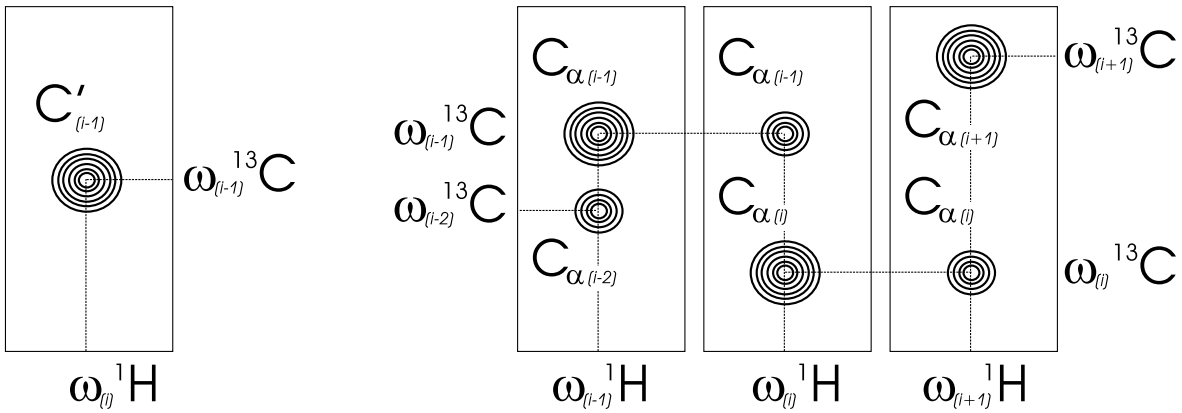


Figure 42:
Schematic appearance of a NMR spectrum obtained by a 3D HNCO- (left panel) and HNCA-experiment (right panel). Stripes denote planes in the 3D spectrum along 15N resonance frequencies. Cross-peaks denote carbon-proton correlations for a respective nitrogen chemical shift. Whereas the HNCO experiment displays a single cross-peak, correlating the carbon frequency of the preceding ***(i-1)*** carbonyl C (C') to the nitrogen and proton chemical shift of residue (***i***), the HNCA experiment allows to sequentially connect resonance frequencies of Cα nuclei of the same residue ***(i)*** to the Cα chemical shift of the preceding residue (***i-1***). It thus correlates two carbon resonance frequencies to every amide proton-nitrogen spin system. Horizontal lines connecting cross-peaks with identical carbon chemical shifts denote sequential connectivities of consecutive residues.

A combination of both experiments can be used to sequentially connect C_α chemical shifts of a protein of interest and to yield C' frequency values for corresponding residues. Unfortunately, it is not be possible to determine the sequence identity of the connected spin-systems because all amino acids contain C_α atoms. Furthermore, C_α chemical shifts are usually not well dispersed and regions of large overlap are normally encountered that make unambiguous sequential assignment difficult. To overcome these difficulties, the following set of experiments is recorded.

3.3.2. The 3D HN(CO)CA, 3D HN(CO)CACB and 3D HNCACB Experiment

The 3D HN(CO)CA and the 3D H(N)COCA experiments correlate the spins H^N*(i)*, N*(i)*, C_α*(i-1)* and H^N*(i)*, C'*(i-1)*, C_α*(i-1)*. Thus, they connect the amide proton of amino acid *i* with resonance frequencies (C_α or C_α and C') of the preceding amino acid *i-1*. To additionally resolve ambiguities, experiments have been devised that simultaneously correlate C_α and C_β carbon shifts for respective amide groups. The HN(CO)CACB experiment directs transfer of magnetization to C_α, C_β carbon nuclei of the preceding residue by transferring magnetization via the *i-1* carbonyl C' (CO) without chemical shift evolution. It correlates H^N*(i)*, N*(i)*, C_α*(i-1)*, C_β*(i-1)* and thus

yields a total of two cross peaks for each amide group, corresponding to the C_α, C_β carbon chemical shifts of the preceding residue. The 3D HN(CO)CACB is more routinely performed as a 3D (H)CBCA(CO)NH (i.e. magnetization is initially created on aliphatic side-chain protons and consequently transferred via aliphatic Cα and Cβ carbon spins to the back-bone amide groups), yet the correlated spins are identical.

The HNCAB version of this experiment correlates the N and the H^N resonances of each amino acid with aliphatic carbon resonances of both the same *(i)* and the preceding residue *(i-1)*. Thus four cross peaks are obtained for each amino acid, which connect the chemical shifts of H^N*(i)*, N*(i)* with the chemical shifts of C_α*(i)*, C_β*(i)* and C_α*(i-1)*, C_β*(i-1)*. Transfer amplitudes are larger for C_α*(i)*, C_β*(i)* spin systems than for C_α*(i-1)*, C_β*(i-1)* and C_α cross peaks of all amino acids but glycine have opposite signs compared to the C_β cross peaks (Figure 43).

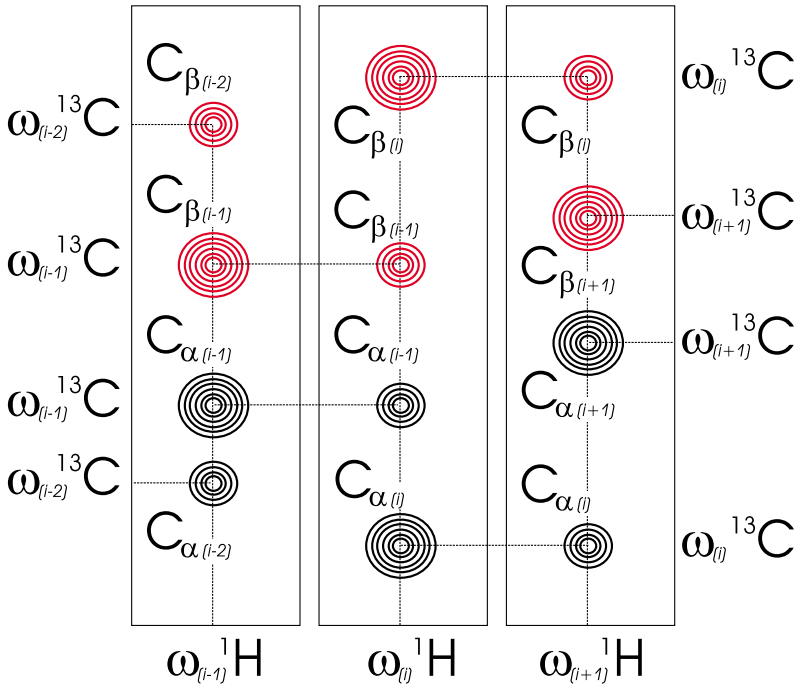


Figure 43:
Schematic appearance of a NMR spectrum obtained by a 3D HNCACB experiment. Stripes denote planes in the 3D spectrum along 15N resonance frequencies. Cross-peaks denote carbon-proton correlations for a respective nitrogen chemical shift. The HNCACB experiment allows sequential alignment of Cα and Cβ resonance frequencies of spin systems ***i*** with those of the preceding residue ***i-1***. Sequential connectivities are indicated by horizontal lines. Ideally, all Cα and Cβ side-chain carbon resonances of a protein can be assigned in that way. Note, that compared to the HNCA experiment, the HNCACB resolves ambiguities for similar Cα chemical shifts by observing Cβ resonance frequencies to resolve spectral overlap.

This greatly helps in unambiguously resolving sequential connectivities and usually allows complete assignment of the back-bone poly-peptide chain. A disadvantage of the HNCACB experiment is that carbon chemical shifts recorded during *t1* are

modulated by C,C couplings. This is especially critical for C_β cross peaks, which can have up to three carbon coupling partners. In general, the HNCACB experiment is suitable for proteins with short ^{13}C transverse-relaxation times, corresponding to small protein size (~15 kD) and for larger proteins provided that their aliphatic side-chains are ^{2}H labeled. For non-deuterated proteins with ^{13}C transverse-relaxation rates of ~20 ms, the 3D CBCANH rather than the HNCACB experiment is performed. Correlated spin-systems are the same as for the HNCAB, however the magnetization transfer involves initial magnetization of the aliphatic H_α, H_β protons and an INEPT transfer to the C_α, C_β spins and chemical shift evolution of the carbons. In the following relay-step, the magnetization transfer $C_\beta \rightarrow C_\alpha$ and $C_\alpha \rightarrow C_\alpha$ is selected. From the C_α the magnetization is transferred in two INEPT sequences to the amide resonances of the same amino acid (via $^{1}J_{C\alpha,N}$) and of the next amino acid (via $^{2}J_{C\alpha,N}$). Amide chemical shift is recorded in a constant time (CT) evolution delay t_2 and then transferred from N to H^N and detected in t_3 (Chapter 3.7)

3.4. NMR Experiments for Side-Chain Assignment

In a next step, triple resonance experiments are employed to assign carbon and proton chemical shifts of respective side-chain amino acids. Information about H^N, N, C_α and C_β chemical shifts, obtained from previously outlined experiments, is taken as a starting-point to yield additional resonances in aliphatic side-chains. In general, magnetization is initially created on aliphatic side-chain protons, transferred to the directly bound carbon nuclei, where it is allowed to 'propagate' or 'mix' along the whole of the aliphatic spin system via a transfer termed TOCSY (total correlation spectroscopy), redirected to the amide nitrogen spin system via the carbonyl C' (CO) and eventually transferred from there to the H^N protons for detection. These experiments are referred to as (H)CC(CO)NH, for correlation of aliphatic carbon resonances (*i-1*) to H^N*(i)*, N*(i)* spin systems, or H(CC)(CO)NH for side-chain proton (*i-1*) to backbone amide *(i)* correlation. Eventually, a 3D HCCH-TOCSY correlates all aliphatic proton resonances with all carbon spin systems within the same amino acid residue (Figure 44).

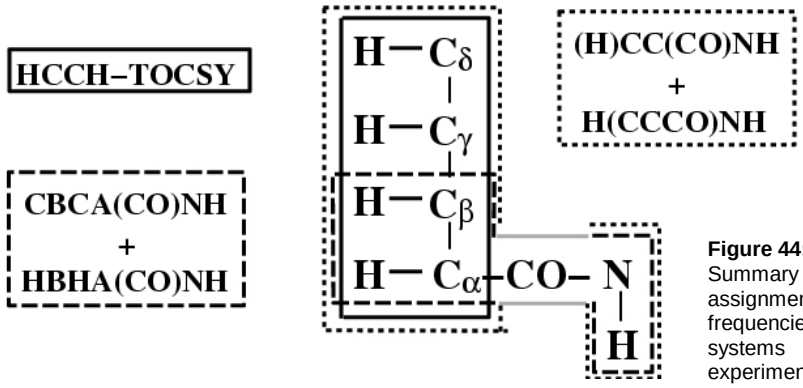


Figure 44:
Summary of used NMR experiments for assignment of protein side-chain resonance frequencies. Experimentally correlated spin-systems are boxed for the respective experiments.

3.4.1. The 3D (H)CC(CO)NH and 3D HCCH-TOCSY Experiment

Several similar (H)CC(CO)NH experiments are available that all transfer magnetization via H → $C_{aliphatic}$ → $C_{aliphatic}$ → C_α , followed by INEPT steps to transfer the magnetization via C' and N to the amide proton. The relay-COSY (correlation spectroscopy) transfer C_β, C_α → C_α, employed in the CBCA(CO)NH and CBCANH experiments (see above) is replaced by the homonuclear C,C-TOCSY transfer. Such transfer steps are less sensitive because only a fraction of the magnetization is finally localized on the C_α. Nevertheless, for proteins < 20 kD TOCSY transfers are routinely employed. The (H)CC(CO)NH experiment will eventually correlate aliphatic carbon resonances of residues *(i-1)* with amide group resonances of residues *(i)*. Hence, cross-peaks for all side-chain carbon atoms of the previous residue are observed for each amide resonance. The H(CC)(CO)NH experiment is closely related to the (H)CC(CO)NH experiment the only difference being that chemical shift is initially evolved on aliphatic protons and not on the respective carbon nuclei. This type of experiment correlates aliphatic proton resonances of residues *(i-1)* to amide group frequencies of residues *(i)*. Cross-peaks for all side-chain protons of the previous residue are observed for each amide group.

The 3D HCCH-TOCSY experiment is used to assign aliphatic ^1H, ^{13}C spin systems and to link them to the sequentially assigned backbone resonances. In an $H_nC(i)$, $H_mC(j)$ spin system, a HCCH-TOCSY experiment connects chemical shifts H*(i)*, H*(j)* and C*(j)* (Figure 45).

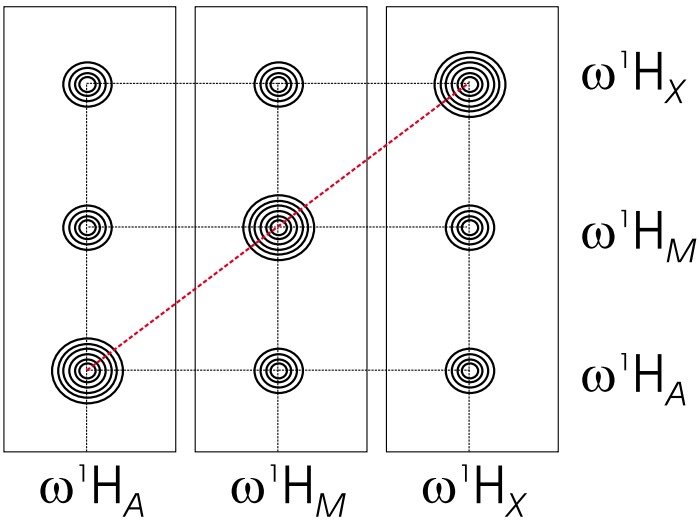
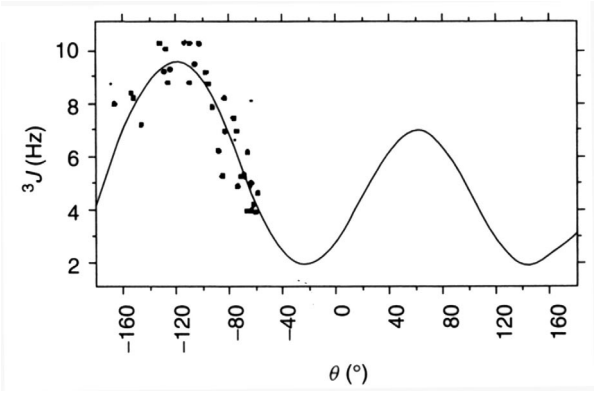


Figure 45:
Schematic appearance of a spectrum obtained by a 3D HCCH-TOCSY experiment. An arbitrary AMX spin system containing three 13C carbon nuclei (A, M, X) bound to three protons (HA, HM, HX) yields a total correlation spectrum of this kind. Strips denote different planes in the indirect carbon dimension with resonance frequencies corresponding to the chemical shifts of carbon atoms A, M and X. Diagonal cross-peaks, indicated by the red line, correspond to 'auto-resonances' of protons directly bound to the respective carbon nuclei. Ideally, each carbon-proton correlation displays cross-peaks to all other proton frequencies of the respective spin system.

This connectivity information yields the complete assignment of the ^{1}H and ^{13}C resonances, because the aliphatic proton chemical shifts of a given spin system are found at the ^{13}C chemical shift of all carbon frequencies involved in that spin system. Hence, for each carbon nucleus of an amino acid side chain, corresponding to a separate plane in the 3D spectrum, there are cross peaks to all hydrogen atoms of the same residue.

3.5. HNHA-*J* and Nuclear Overhauser Effect (NOE)

After the assignment of all or nearly all resonances of a protein, experiments for the extraction of structural parameters are analyzed. The most important parameters for NMR-based structure determination are ^{1}H, ^{1}H distances which are derived from NOE intensities and dihedral ***Φ*** angles which are obtained from ^{3}J couplings (Figure 46, reproduced from Levitt 2002, Chapter 3.7).



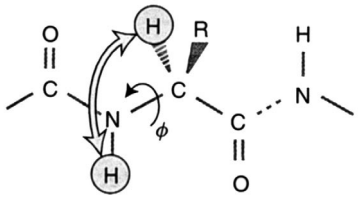


Figure 46:
The three-bond J-couplings between NH protons and Cα protons in a protein have a well-defined dependence on the molecular torsion angle *Φ*. Each circle on the *Karplus curve* represents a single amino acid residue in a protein. The semi-empirical Karplus curve has the mathematical form $^{3}J=(6.4\cos^{2}\theta-1.4\cos\theta+1.9)$ Hz, where θ is the H-N-C-H torsion angle, given in terms of the backbone torsion angle Φ by $\theta=\Phi-\pi/3$.

^{13}C- and ^{15}N-edited NOESY experiments are generally used for the measurement of NOE intensities (see below). A special set of *edited-filtered* NOE experiments are employed to study *inter-molecular* proton-proton distances in protein-peptide complexes of labeled protein and un-labeled peptide (or vice-versa).

3.5.1. NOESY and *edited-filtered* NOESY Experiments

With increasing size of the molecules under investigation, overlap of cross-peaks becomes a problem for the extraction of NOE intensities and intermolecular distances from 2D NOESY spectra. This overlap can be removed by the introduction of additional frequency dimensions. Resolving the NOESY spectrum along a heteronuclear dimension yields improved resolution, since the number of cross-peaks remains constant. The pulse sequence of the ^{15}N-edited NOESY experiment is composed of a 2D NOESY and a sensitivity-enhanced ^{15}N, ^{1}H-HSQC (*heteronuclear single quantum coherence*) experiment using a heteronuclear gradient echo combined with a water-flip-back pulse. The sequence of the ^{13}C-edited NOESY experiment is again composed of a 2D NOESY and a ^{13}C, ^{1}H-HMQC (*heteronuclear multiple quantum coherence*) pulse sequence.

In practice, the resulting pattern of cross-peaks obtained from the ^{13}C-edited NOESY experiment is similar to the 3D HCCH-TOCSY spectrum, only that it additionally contains NOE intensities (cross-peaks) correlating *through-space* proton-proton distances of up to 6Å (Figure 47).

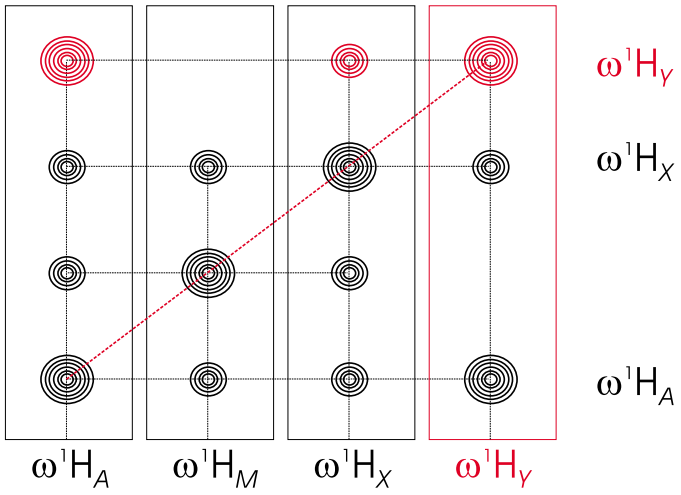


Figure 47:
Schematic appearance of a spectrum obtained by a 3D 13C-edited NOESY experiment. Again, an arbitrary AMX spin system is chosen and as for the 3D HCCH-TOCSY, stripes correspond to different planes, i.e. different resonance frequencies of the respective carbon nuclei of this spin system. Directly bound protons display diagonal auto-correlation cross-peaks to proton spins of the same system, but additionally exhibit NOE cross-peaks to a different spin system Y (colored in red). The HA and HX protons of the AMX spin system display different intensity NOE cross-peaks to the HY nucleus of the Y spin system, which by itself exhibits similar intensity NOE cross-peaks to the AX protons. Together with information obtained by the 3D HCCH-TOCSY experiment, NOE cross-peaks to different spin systems can be clearly distinguished from intra-residual NOE intensities.

Integration of cross-peaks yields NOE-derived distance restraints which are classified according to weak, medium and strong intensities in the NOESY experiment. To discriminate between *intra-* and *inter-molecular* NOEs in a differentially labeled protein-peptide complex (Figure 48), a special type of *edited-filtered* NOE experiment is performed. This kind of experiment allows to either select or repress magnetization of hetero-nucleus bound protons and can thereby provide NOE intensities between protons bound to ^{13}C nuclei and protons bound to ^{14}C (unlabeled) nuclei. Hence, describing *inter-molecular* proton-proton distances, which will help to structurally position the unlabelled component in respect to the labeled part of the complex.

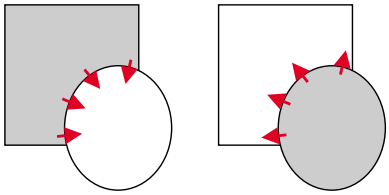


Figure 48:
Reciprocal edited filtered NOE experiments. Only NOEs (red arrows) from labeled components (gray) to unlabeled, bound ligands (white) are observed. Ideally, a complementary set of edited filtered NOE experiments with differentially labeled components allows to spatially position interacting proteins in a complex.

3.6. Structure Calculation

The principal source of structural information in the determination and refinement of three-dimensional structures of proteins and nucleic acids in solution by NMR are the inter-proton distances $\mathbf{r}_{ji}$ calculated from the NOE intensities $\mathbf{I}_{ij}$ using relation 1.8,

$$r_{ij} = r_{ref}\left(\frac{I_{ref}}{I_{ij}}\right)^{1/6} \qquad [1.8]$$

with an appropriate reference for distance calibration. The experimental distance is imprecise because of peak integration errors, spin diffusion and internal dynamics. Therefore the experimental distance **D** is specified between lower (d_{lower}) and upper (d_{upper}) bounds. Since availability of NOE distances does not fully determine the three-dimensional structure, the experimental information has to be extended by chemical knowledge about bond lengths, bond angles and van der Waal radii of the macromolecular system. This knowledge is introduced into the calculation via an *hybrid energy function* $\mathbf{E}_{hybrid}$, which is a weighted sum of experimental and

chemical contributions describing internal geometric interactions (bond lengths, bond angles, chiral centers, planarity of aromatic rings) as well as non-bonded (van der Waals, hydrogen bonding and electrostatic) interactions. The structure calculation is then a search of a global minimum of this energy function $\mathbf{E}_{hybrid}$, which corresponds to a family of three dimensional structures of the atomic model which satisfy both the chemical as well as the experimental information for conformations of the molecule with low values of the hybrid energy. Therefore, NMR derived three-dimensional structures are always depicted as an *ensemble* of lowest energy models. In practice, structures presented in this work were calculated in a mixed torsion and Cartesian angle simulated annealing protocol with the programs CNS and ARIA.

3.7. References

‘Spin Dynamics; Basics of Nuclear Magnetic Resonance’
Malcolm H. Levitt
Wiley 2002

‘Protein NMR Spectroscopy; Principles and Practice’
John Cavanagh, Wayne J. Fairbrother, Arthur G. Palmer III, Nicholas J. Skelton
Academic Press 1997

‘The Nuclear Overhauser Effect in Structural and Conformational Analysis’
David Neuhaus and Michael Williamson
VCH Publishers 1989

‘Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed gradient’
Michael Sattler, Juergen Schleucher, Christian Griesinger
Progress in Nuclear Magnetic Resonance Spectroscopy 1999 (34); 93-158

4. MATERIALS and METHODS

Methods for cloning, expression and purification of recombinant human SMN Tudor domain and recombinant human U2AF^{65}-RRM3 and SF1 are largely identical and will be outlined in parallel. Individual differences will be pointed out when necessary. Protocols for cloning and recombinant protein production and purification are outlined in brief, a very detailed description of the procedures can be obtained from the EMBL Protein Expression and Purification Units homepage:

http://www.embl-heidelberg.de/ExternalInfo/geerlof/draft_frames/frame_protocol_database.html

4.1. Cloning

For the human Survival of Motor Neuron (SMN) Tudor domain, two primers (**SMN5'** and **SMN3'**) were initially designed to allow PCR (Polymerase-chain-reaction) amplification of a full-length SMN (aa 14-294) cDNA construct from a human skeletal muscle cDNA library.

SMN5': 5'-TGCC**ATG***GAGCAGGAGGATTCCGTGCTGTTC*-3'
SMN3': 5'-CATGGTA*CC****TTA****ATTTAAGGAATGTGAGCACCTTCC*-3'

The *NcoI* and *KpnI* restriction sites in the 5' and 3' primers, respectively, are underlined, SMN nucleotides are indicated in italic and the START- and STOP-codons are shown in bold. This PCR product was gel-purified and used as a template for subsequent amplification reactions to yield different N-terminal portions of the human SMN protein, that include the Tudor domain.

SMN3Gr: 5'-GTTTGGTACC**TTA***ACCCGAAGTTTCACAAATGTCACC*-3'
SMN3Tr: 5'-GTTGGTACC**TTA***GGTTTTTGGTTTACCCGAAGTTTC*-3'

SMNTUs: 5'-TTTCCATGG*AAAAGAAGAATACTGCAGCTTCCTTA*-3'

SMN4r: 5'-CGGGGTACC**TTA***TTCACAGATTGGGGAAAGTAG*-3'

SMN2r: 5’-CGGGGTACC**TTA***ATTCTGTTCTATATTATTAGCTAC*-3’

SMNTUr: 5’-GTTTGGTACC**TTA***TTCATCTGTTGAAACTTGGCTTTCA*-3’

SMN1r: 5’-CGGGGTACC**TTA***AGACCTGGAGTTCTCACTTTCATC*-3’

SMN3r: 5’-CGGGGTACC**TTA***TTCGAGAAAAGAGTTCCAAGG*-3’

Combination of **SMN5’** with **SMN3Gr** and **SMN3Tr** yields poly-peptides corresponding to SMN regions (aa 14-64 and aa 14-68, respectively) N-terminal to the Tudor domain, whereas combinations of **SMN5’** with any of the C-terminal reverse primers (marked by an **r** at the end of the primers name) results in protein constructs containing both the Tudor domain and the additional N-terminal region. The Tudor domain can be produced with various different C-termini by a combination of **SMNTUs** with any of the C-terminal reverse primers. The solution structure of the Tudor domain was solved with a minimal protein construct encompassing the folded three-dimensional Tudor structure (aa 95-147) and corresponding to the **SMNTUs**, **SMN4** combination of primers (aa 82-147). Any other combination of primers resulted in poly-peptides with additionally unfolded residues C-terminal to the Tudor domain. The Tudor domain E134K point mutation construct was cloned by PCR amplification with a combination of **SMNTUs**, **SMN4** and mutant primers (**tud_mut2** and **tud_mut1**, respectively from a wild-type template. PCR products were annealed and mutant Tudor domain was re-amplified with **SMNTUs** and **SMN4** primers. Presence of the mutation was verified by digestion with the MnlI enzyme, whose single restriction site (5’-CCTC-3’) was deleted in the mutant 5’-CCTT-3’ construct and by DNA sequencing.

tud_mut1: 5’-*TATGGAAATAGAAAGGAGCAAAATCTGTCCGAT*-3’

tud_mut2: 5’-*CAGATTTTGCTCCTTTCTATTTCCATATCCAGT*-3’

For the human U2AF65-RRM3 and the human SF1 proteins, plamids encoding full-length DNA sequences of the respective proteins, were kindly provided by Prof. Angela Kramer from the University of Geneva, Switzerland and served as templates for PCR reactions to amplify portions of these proteins. Primers were again designed to contain 5’ *NcoI* and 3’ *KpnI* sites plus a C-terminal STOP-codon which would terminate translation of the desired protein constructs.

U2AF1s: 5’-TTTCC***ATG***G*GCGGCCACCCGACTG*-3’

U2AF2s: 5’-TTTCCATG*GACGAGGAGTTATGAGGAGATC*-3’

U2AF3r: 5’-TTTGGTACC**TTA***GTCGGGGTCACAGTATTTTG*-3’

U2AFr_1: 5’-TTTGGTACC**TTA***GTCCCGGCGGTGATAAGA*-3’

U2AFr_2: 5’-TTTGGTACC**TTA***GCGGTGATAAGAGTCGGG*-3’

U2AF4r: 5’-TTTGGTACC***TTA***CCAGAAGTCCCGGCGGTGA*-3’

Only the primer combination **U2AF1s** and **U2AF4r** (aa 367-475) corresponds to the complete fold of the last RRM-domain of U2AF^{65} (aa 367-475), whereas all C- and N-terminal shorter poly-peptides truncate the RRM fold and are unstable in solution.

SF1s: 5’-TTTCC***ATG***GCGACCGGAGCGAACGC*-3’

SF1_25r: 5’-TTTGGTACC**TTA***GTCTTGGTTCCAGCGGCTCC*-3’

SF1_110r: 5’-TTTGGTACC**TTA***ATCCGGATTGAGTGCAACCAT*-3’

SF1_142r: 5’-TTTGGTACC**TTA***TTTATCACTCACACGTGTTGC*-3’

SF1_146r: 5’-TTTGGTACC**TTA***CTCATCTTGTGGAATCATGAC*-3’

SF1_166r: 5’-TTTGGTACC**TTA***GCGCAGTTTACGAGTCAGG*-3’

SF1_189r: 5’-TTTGGTACC**TTA***CTCGCTATTGTAGATGGGC*-3’

The nomenclature of the reverse primers (denoted with an **r** in their name) indicates the last amino acid residues of the resulting poly-peptide when PCR amplified in combination with **SF1s**. For the U2AF^{65}-RRM3/SF1 protein-peptide complex, the smallest peptide construct corresponding to the first 25 residues of human SF1 was sufficient for complex formation and subsequently used in the study described above. Additionally to these recombinantly produced SF1 peptides, the following peptides were synthetically produced and purchased from MWG™-Biotech.

SF1_15mer: NH2-**DFPSKKRKRSRWNQD**-COOH (aa 11-25)

SF1_25mer:NH2-**MATGANATPLDFPSKKRKRSRWNQD**-COOH (aa 1-25)

SF1_21mer: NH2-**DFPSKKRKRSRWNQDTMEQKT**-COOH (aa 11-31)

All PCR reactions were carried out under standard conditions (as described in Current Protocols, CP), products were gel-purified (Quiagen gel purification kit), digested with the respective restriction enzymes (according to manufacturers instructions) , re-purified on an argarose gel (see above) and ligated into modified Novagen pET-24d and pET-9d (conferring *Kanamycin (Kan)* antibiotic resistance) vectors to yield in-frame versions of protein expression vectors in which portions of the desired protein domain are N-terminally fused to either a histidine- (His_6) or a Histidine-GST (Glutathion-S-transferase) (His_6-GST) tag. Methods for primer design, PCR amplification, gel-purification and ligation followed standard molecular biology procedures described in CP. Protein expression constructs carry a TEV-cleavage site engineered between the fusion-tag and the protein of interest, allowing for subsequent TEV (tobacco etch virus) protease cleavage and removal of the tag moiety. DMSO (dimethyl-sulfoxide) competent *E.coli* cells (preparation according to the standard procedure described in CP) of the DH5α strain were transformed by standard heat-shock methods, plated onto *Kan* (used at 20-40 μg/ml concentrations depending on plasmid type) containing LB-agar (Luria Broth agar and plate pouring as described in CP) plates and incubated over-night at 37°. Colonies were picked and grown while shaking in 1ml of LB/*Kan* at 37° for 12 hours and plasmids purified according to the manufactures instructions using the Quiagen mini-prep kit. Plamids were re-transformed into DMSO competent *E.coli* cells of the BL21 strain and plated and grown as described above. Integrity of all plasmids was confirmed by DNA sequencing methods and subsequently produced at large quantities by maxi-prep protocols using Quiagen kits.

4.2. Recombinant Protein Expression

Smale scale expression studies were performed in 1ml of LB/*Kan* medium and cells were grown to an optical density at 500 nm of ~0.6 (OD_{500}=0.6), recombinant protein expression was induced by addition of 0.1-1.0 mM IPTG (isopropylthio-β-D-galactoside, which functions as an organic lactose analogue binding to the *LacZ* promoter and driving T7-polymerase expression which transcribes the pET-DNA sequences downstream the pET-T7 promoter and hence produces recombinant protein) and allowed to pursue for 2 hours at 37° C. Cultures were centrifuged at 13 $x10^3$ rpm in a table-top centrifuge, to sediment bacterial cells, re-suspended in

lysis-buffer (50 mM PO_4 pH 8.0, 150 mM NaCl, 5-10 mM β-mercaptoethanol, 5 mM PMSF, 0.5% v/v TX-100) and sonicated by a micro-tip ultra-sonic device to break-up the bacterial cell-walls. The solution was centrifuged at 13 $x10^3$ rpm to pellet the cell debris and samples of the clear supernatant, containing soluble proteins, were separated by conventional SDS- PAGE (sodium dodecyl sulfate- poly-acrylamide gel electrophoresis). Gels were stained and de-stained (according to CP) with Coomassie to visualize proteins of interest, as described. Once a protein expression clone producing recombinant protein at sufficiently high-levels had been identified by that method, protocols were scaled up to obtain large quantities (10-20 mg) of the respective poly-peptides. For this, an over-night culture from a freshly transformed LB/*Kan* plate was inoculated 1:100 into a chosen volume of LB/*Kan* (usually 1-5l) and grown at 37° C until an OD_{500}~0.6. Recombinant protein expression was induced by addition of 0.1-1 mM IPTG and allowed to pursue for 3-6 hours. For some protein expression constructs, the temperature was lowered to 25°-28° C during this period because yields of soluble proteins were higher at lower temperatures and expression was carried out over-night. Cultures were centrifuged at 6 $x10^3$ rpm to sediment cells, supernatant was discarded and cells resuspended in lysis-buffer (see above). Sonication was carried out essentially as described, only the micro-tip device was replaced by a large-scale preparative tip. The solution was cooled on wet ice during the preparation, collected and centrifuged in a high speed ultra-centrifuge at 45 $x10^3$ rpm. The clear supernatant was passed through a 0.2-0.4 μm filter, DNAse1 and PMSF (phenymethyl-sulfonyl fluoride) were added to a final concentration of 10 and 5 mM, respectively. For production of isotope enriched recombinant poly-peptides for structure determination by NMR, cultures of *E.coli* cells were grown in minimal M9 medium, supplemented with ^{15}N-NH_4Cl (5 g/l) or ^{15}N-NH_4Cl (5 g/l) and ^{13}C-Glucose (2-4 g/l) as the sole nitrogen or nitrogen and carbon source. The respective recombinant proteins are either ^{15}N or ^{15}N, ^{13}C-doubly labeled. Protocols for expression and purification were identical for unlabeled or labeled proteins, otherwise.

4.3. Recombinant Protein Purification

Initially all His_6-GST-tagged recombinant proteins were purified in batch on Ni-TA Sepharose columns (Quiagen). 1-5 ml of Ni-TA Sepharose was equilibrated in a

gravity-flow column with lysis-buffer omitting detergent (50 mM PO_4 pH 8.0, 150 mM NaCl, 5-10 mM β-mercaptoethanol, 5 mM PMSF) and added to the cleared bacterial lysate (see above). The suspension was mixed for 2-4 hours in the cold-room by slow stirring, transferred to gravity-flow columns and sedimented therein. The flow-trough lysate was kept until presence of desired recombinant His-tagged proteins on the column was confirmed. The column material was washed with 2 column-volumes of **1.** 50 mM PO_4 pH 8.0, 150 mM NaCl, 10 mM Imidazole, β-mercaptoethanol , PMSF, plus 5%(v/v) of TritonX-100, **2.** same as for step 1 but omitting detergent, **3.** same as step 2 but 300 mM NaCl, **4.** same as step 2 but 30 mM of Imidazole and **5.** eluted in step 2 buffer containing 300 mM Imidazole. Presence of desired fusion proteins was confirmed with the Pierce 'Coomassie Plus Protein Assay' reagent and quantified by measuring absorbance at 280 nm.

4.3.1. TEV Protease Purification and Usage

A His-tagged version of TEV protease was recombinantly produced from an pET-vector based expression clone in a pLys-S/BL21 *E.coli* strain. Recombinant TEV protease was purified with the essentially the same Ni-chelating protocol as described above, protein concentration was adjusted to 1 mg/ml and 20%(v/v) of glycerol was added, samples were aliquoted in 100 µl batches and stored at –80°. For removal of the fusion moiety from the desired protein domains, ~100 µg of recombinant TEV protease was added to the elution-fraction of the Ni-TA purification and incubated over-night at room temperature. Completeness of cleavage was assessed by comparing aliquots of undigested with over-night cleaved samples on a SDS-PAGE and subsequent Coomassie staining. Usually cleavage was complete after 2-3 hours of incubation at room temperature. The elution fraction, in which the TEV cleavage reaction was carried out (containing 300mM of Imidaxzole) was diluted 10 times with 50mM PO_4 pH 8.0, 10 mM NaCl, plus reducing agent and protease inhibitor to yield a solution with < 30 mM Imidazole, which was re-applied on a pre-equilibrated Ni-TA Sepharose column. This time, the flow-through fractions (containing un-tagged domains of interest) were collected whereas the eluted material (see above for procedure), containing the fusion-moiety and the His-tagged TEV protease, was discarded.

4.3.2. Ion Exchange and Size Exclusion Chromatography

Flow-through fractions containing either negatively charged SMN Tudor- (calculated pI=4.9) and U2AF65-RRM3 (calculated pI=4.6) domains were loaded onto and bound to Pharmacia anion exchange ResourceQ columns (at pH8.0), the positively charged recombinant SF1 peptide fractions (calculated pI for SF1 (aa 1-25) =10.2) were applied onto Pharmacia cation exchange ResourceS columns (at pH 8.0). A linear salt gradient (0-1 M NaCl, flow-rate 1ml/min) was applied and proteins of interest were eluted at ionic-concentrations corresponding to their respective pIs. Samples were concentrated to 0.5ml sample volumes in Amicon filter-centrifugation devices using appropriate molecular-weight cut-off (MWCO) pore sizes. For the small SF1 (aa 1-25) peptide, the ion-exchange fraction corresponding to the SF1 elution peak was lyophilized and resuspended in 0.5 ml of water. Concentrated protein samples were loaded on a size-exclusion gel-filtration column (Pharmacia Superdex75), pre-equilibrated with 30 mM PO_4 pH 6.4, 20 mM NaCl, 5 mM DTT (Dithiothreitol), for final purification. Elution peaks were collected and re-concentrated to a final volume of 0.5-0.6ml (for conventional NMR-tubes) and to 0.2-0.3 ml for samples to be used in Shigemi-tubes. SF1 peptide elution fractions were lyophilized, resuspended in H_2O, dialyzed against H_2O with appropriate MWCO dialysis-tube pore-sizes, re-lyophilized and re-suspended in NMR buffer. For NMR measurements 10%(v/v) of D_2O was added to these NMR samples.

4.4. NMR Experiments

All NMR spectra for structure determination were recorded at 27 º C on Bruker DRX 500, DRX 600 and DRX 800 NMR spectrometers. Spectra were processed with NMRPIPE and analyzed using XEASY. A summary of standard triple-resonance NMR experiments employed for doubly-labeled protein samples of Project 1 and Project 2 is given below. Non-standard experiments are specifically discussed in the respective chapters.

	SMN Tudor domain	**U2AF^{65}-RRM3**	**SF1**
Backbone Assignment			
	CBCA(CO)NH	CBCA(CO)NH	CBCA(CO)NH
	HNCA	HNCA	HNCA
	HNCAB	HNCAB	HNCAB
Side-chain Assignment			
	(H)CC(CO)NH	(H)CC(CO)NH	(H)CC(CO)NH
	HCC(CO)NH	HCC(CO)NH	HCC(CO)NH
Distance Restraints			
	^{15}N NOESY	^{15}N NOESY	^{15}N NOESY
	^{13}C NOESY	^{13}C NOESY	^{13}C NOESY

4.5. Biochemical Interaction Studies

4.5.1. SMN/Sm Interaction Experiments

Approximately 50 µl of glutathione-Sepharose (Amersham-Pharmacia) was incubated for 1 hour at 4°C with 1 µg of purified recombinant GST-fusion protein. Fragments fused to GST and used for binding assays include the tails of human Sm D_1 and Sm D_3 (residues 68-119 and 72-126) and the Tudor domain of the wild type and mutated (E134K) human SMN sequence (residues 36 to 160). For protein binding, 3 µl of ^{35}S-labeled, *in vitro* translated proteins were incubated with the matrix-coupled GST-fusion proteins for 1 h at 4° C. The resin was subsequently washed extensively with binding buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$). Bound protein was eluted with SDS-sample buffer and analyzed by

SDS-PAGE followed by autoradiography using Amplify (Amersham). Translates used include full-length human SMN and Sm D_1, core Sm D_1 (residues 1-68) and the Tudor domain of SMN in its wild-type and mutant (E134K) form.

4.5.2. U2AF^{65}/SF1 Interaction Experiments

Approximately 50 µl of glutathione-Sepharose (Amersham-Pharmacia) was incubated for 1 hour at 4°C with 1 µg of purified recombinant wild-type/mutant U2AF^{65}-RRM3 (aa 367-475) GST-fusion protein. 3-5 µg of His-tagged SF1 protein (wild-type/mutant, aa 2-320) was incubated with the matrix-coupled GST-fusion proteins for 1 h at 4° C. The resin was subsequently washed extensively with binding buffer (300 mM NaCl, 50 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$). Bound protein was eluted with SDS-sample buffer and analyzed by SDS-PAGE. Transferred onto Nitrocellulose for Western-Blotting and analyzed by standard Western-Blotting procedures (see CP) using α-His and α-GST antibodies, respectively.

PUBLICATIONS

'SMN Tudor domain structure and its interaction with the Sm proteins.'

Selenko P., Sprangers R., Stier G., Buhler D., Fischer U. and Sattler M.

Nature Structural Biology; 2001 Jan(8): 27-31

'Definition of domain boundaries and crystallization of the SMN Tudor domain.'

Sprangers R., Selenko P., Groves M., Sinning I. and Sattler M.

Submitted for publication, September 2002

'Structural basis for the molecular recognition between human splicing factors U2AF^{65} and SF1.'

Selenko P., Gregorovic G., Sprangers R., Stier G., Rahni Z., Kramer A. and Sattler M.

Submitted for publication, October 2002

CURRICULUM VITAE

Name	***Philipp Selenko***	
Place and Date of Birth	Graz, AUSTRIA	24/05/1972
Citizenship	Austrian	
Education	1978-1982	Elematary School, Vienna
	1982-1990	High School, Vienna
	1990-1992	Study of Chemistry University of Vienna
	1992-1996	Study of Chemistry and Biochemistry University of Vienna
	1997-1998	Diploma Work Institute of Molecular Pathology (IMP), Vienna
	1998-2002	PhD Work European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

ACKNOWLEDGEMENTS

First and above all, I would like to thank Michael Sattler for having me in his group and for giving me the opportunity to learn NMR, pursue my own scientific ideas and for generously providing me with the intellectual freedom to come up with my own concepts and research plans. I truly appreciated this sort of scientific environment and hope to be able to continue working in such a stimulating atmosphere in the future. The whole Sattler group was a great family over the past 4 years and I would like to thank all present and past members for their scientific, social, emotional, political, musical, philosophical, visual, alcoholical, poetical, and friendly support. Special thanks to Gunter for many of the above things but especially for establishing evenings of MAXimum 'intellectual' pleasures, for many jazzanovatic lessons in high-fidelity audio and for becoming a friend. Sprangers, for saving my life (repeatedly and in many ways), for teaching me all of the little I know (about….), for making the synthetic_residence become real and for being such a great friend and companion. A big kiss to the girls (sorry Remco) Silke, Katia (who moved out of our joint office-home because I was talking too much), Ana I and An(n)a II (only you know which is which). Thanks to the 'Sattler-boys'-party-crashers and beer-session terminators, Fabi 'pink CSU' Baby, Phini 'you know these things' Weany and Christian 'who has the ball' E. Which of course leads me to acknowledge the football team (that never made it to a real pitch) and EMBL for keeping silent about all that broken glass. Thanks (!) to members of the EMBL who were, of course, 'one great family'…. but then again families are not always……

Thanks to the many non-EMBL friends in Heidelberg and Mannheim for making my 'country-side life' a bit more tolerable.

Finally, and with all my heart, thanks to M. for the last 11 years and for all those things unmentionable in scientific correspondences. > Zoe, we love you.