NMR Structural Investigations of the U4 snRNA kink-turn and of a lariat-forming ribozyme

INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der Naturwissenschaftlich-Mathematischen Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

vorgelegt von
Dipl.-Ing. Biotechnologie
Melanie Falb
aus Heilbad Heiligenstadt
(Eichsfeld)

Heidelberg, 2009
Tag der mündlichen Prüfung: 27.01.2010

NMR Structural Investigations of the U4 snRNA kink-turn and of a lariat-forming ribozyme

Gutachter: Dr. Anne-Claude Gavin

Prof. Dr. Irmgard Sinning

Betreuer: PD Dr. Teresa Carlomagno

Für meine Familie

Erwirb Dir Gold so viel du brauchst
und Weisheit so viel du kannst.

Kuppelinschrift der Moschee in Schwetzingen

Contents

Summary

Three-dimensional fold of the U4 5'stem loop snRNA in its unbound form

The spliceosome, which catalyzes the splicing of eucaryotic pre-mRNAs, consists of five uridine-rich small nuclear ribonucleoprotein particles (U snRNPs) and numerous other factors. To activate the spliceosome and to enable the first step of splicing, the paired U4 and U6 snRNAs of the U4/U6 snRNP complex need to dissociate from each other. An initial step leading to this dissociation of the U4/U6 snRNP complex is the binding of the protein 15.5K to the 5' stem loop of U4 snRNA (U4 5' SL). Subsequently, further proteins are recruited, in particular protein hPrp31, which forms a ternary complex with the U4 5' SL and 15.5K. Upon binding to the 15.5K protein, the U4 5' SL folds into a characteristic structural motif, called the kink-turn (k-turn), in which the phosphodiester backbone presents a particular sharp turn. In this work the three-dimensional conformation of the free U4 5' SL RNA in solution was investigated in order to elucidate whether the k-turn structural element is already present in the absence of protein binders.

By using NMR spectroscopy, the structure of the unbound U4 5' SL was solved at a precision of 0.6 Å. The use of different NMR experiments enabled the assignment of 88% of the base and ribose protons chemical shifts. Furthermore several restraints were extracted for structure calculations, including 623 NOE-based inter-proton distances as well as 80 dihedral angles and 12 hydrogen-bond restraints. Long range information was provided by measuring 79 residual dipolar couplings (RDCs), necessary to determine the relative orientation of the two helical stems of the U4 5' SL to each other, namely the canonical and the non-canonical stem elements. In the RDC refined structure, the canonical as well as the non-canonical stem of the U4 5' SL present a well-defined helix fold. The U4 5' SL lacks the characteristic sharp turn of the k-turn motif reported for the protein bound form. Instead, the free U4 5' SL presents a more opened, extended conformation. The two non-canonical G-A base-pairs found in the k-turn structure are already formed in the unbound RNA, but the three unpaired residues of the internal loop (AAU) are stacked differently with respect to the k-turn motif.

In this work it was shown that the free U4 k-turn RNA is prevalently found in the extended conformation in solution. Thus, the consensus sequence of the k-turn does not *per se* code for the sharp bent in the RNA backbone. Instead, the structural k-turn element is highly disfavoured in solution and needs to be stabilized by protein binding, which favors the thesis of a protein-assisted mechanism for k-turn RNA folding.

Structural investigations of a spliceosomal related lariat-forming ribozyme

In eukaryotic cells, pre-mRNAs are processed by the spliceosome in a way that internal non-coding regions (introns) are excised and the remaining segments (exons) are joined together. The active spliceosomal core consists of paired U2 and U6 snRNAs, which juxtapose the splice site residues of the pre-mRNA substrate. While the RNA-RNA interactions during splicing have been well-characterized in the past, the description of tertiary structure of the spliceosomal catalytic center remains a challenge due to the large size and dynamics of the complex.
In our laboratory a model ribozyme is investigated, which undergoes a transesterfication reaction with striking similarity to the first step of splicing by forming a 2'-5' lariat. The lariat formation has the same sequence specificity as for the pre-mRNA, including an adenosine at the branch point site, a guanosine at the 5' splice site and the phylogenetically highly conserved ACAGAGA box that is essential for catalytic activity in the spliceosome of higher eukaryotes. Consequently, this selected ribozyme represents an ideal model system to gain insights into the splicing reaction and the specific function of the conserved ACAGAGA motif in lariat formation.
In this work, extensive mutational studies were carried out to identify specific sites that can be either linked to folding of the ribozyme or to their function in the catalytic activity of the RNA molecule. With these studies it could be shown that the residues A29, G32, A33 and A35 are functionally the most important residues within the $A^{29}CAGAGA^{35}$-segment. Furthermore, a well defined helical moiety in the 5' region of the ribozyme (residue 1 to 27) with an unusual high content of non-canonical base-pairs (5 out of 11) was revealed by combining information of the mutational studies with NMR derived distance constraints.
Another aim of this work was to gain first insights into the three-dimensional fold of the ribozyme prior to catalysis. Although the lariat-forming ribozyme is a minimal RNA construct for the first splicing reaction, it represents a challenging system for NMR spectroscopic investigation due to its dynamic behavior and the existence of several non-regular secondary structural motifs. Extensive NMR investigation and analysis were performed, with several differentially ^{13}C, ^{15}N-labeled NMR samples, in order to retrieve numerous complementary NMR spectra for the resonance assignment of the ribozyme. However, although various labeling schemes were applied, the assignment remained ambiguous for some nucleotides, due to both spectral overlap and a conformational exchange process that was detected for half of the resonances. To overcome the spectral complexity of the complete ribozyme, eight RNA mutants were constructed, which enabled

a complete, unambiguous resonance assignment of all nucleotides of the lariat-forming ribozyme. This resonance assignment is a perquisite for the collection of structural restraints, mainly of NOE peak intensities. NOE signals reflect direct distances of neighboring atoms in a molecule and suggest a compact fold, presumably containing both a ribose zipper motif and a pseudoknot motif, for the lariat-forming ribozyme.

The findings derived from the comprehensive mutational studies as well as from the collected NMR data, comprising in particular distance restraints derived from NOESY spectra, permitted first structural calculations. These are in a preliminary stage for the helical, non-exchanging moiety of the ribozyme in its linear form (prior to catalysis) in our laboratory.

All these data represent an excellent starting point to explore the complete three-dimensional structure of the lariat-forming ribozyme and will facilitate its understanding in terms of functionality.

Zusammenfassung

Dreidimensionale Struktur der U4 snRNS 5′ Stammschleife in ungebundener Form

Das Spleißosom katalysiert das Spleißen von eukaryotischer Präkursor Boten-Ribonukleinsäure (prä-mRNS) und besteht aus fünf Uridin-reichen kleinen nukleären Ribonukleoprotein-Partikeln (U snRNP) sowie vielzähliger anderer Faktoren. Um das Spleißosom zu aktivieren und den ersten Schritt der Spleißreaktion zu ermöglichen, ist es erforderlich, dass die basengepaarten U4 und U6 snRNS Moleküle des U4/U6 snRNP Komplexes auseinander dissoziieren. Ein initialer Schritt der Dissoziation des U4/U6 snRNP Komplexes ist die Bindung des Proteins 15.5K an die 5′ Stammschleife der U4 snRNS (U4 5′ SL). Anschließend binden weitere Proteine u.a. das hPrp31 Protein, das einen ternären Komplex mit der U4 5′ SL und dem 15.5K Protein formt. Aufgrund der Bindung des 15.5K Proteins bildet die U4 5′ Stammschleife ein charakteristisches Strukturmotiv (Kink-turn) aus, bei dem das Phosphatrückgrat der RNS stark geknickt wird. Im Rahmen dieser Arbeit wurde die dreidimensionale Konformation der ungebundenen U4 5′ SL RNS in Lösung untersucht, um aufzuklären, ob sich das Kink-turn Strukturelement bereits in Abwesenheit von bindenden Proteinen faltet.

In der vorliegenden Dissertation wurde mit Hilfe der NMR-Spektroskopie die Struktur der U4 5′ Stammschleife mit einer Auflösung von 0,6 Å aufgeklärt. Durch Anwendung verschiedener NMR-Experimente konnten 88 % der chemischen Verschiebungen ihren jeweiligen Basen- und Ribose-Protonen zugeordnet werden. Weiterhin wurden für die Strukturrechnung nötigen Randbedingungen extrahiert, u.a. 623 NOE-basierte Protonenabstände, 80 Diederwinkel sowie Daten zu Wasserstoffbrückenbindungen (12 Abstände). Die Bestimmung von Informationen über eine weitere Entfernung (long range) als NOE- basierter Abstände, erfolgte durch die Messung von dipolaren Restkopplungen (RDC, 79 Kopplungen), um so die relative Orientierung der beiden helikalen Abschnitte der U4 5′ SL (kanonischer und nicht-kanonischer Stamm) zueinander zu ermitteln. In der durch RDC verfeinerten Tertiärstruktur liegen der kanonische sowie der nicht-kanonischen Stamm in einer definierten, helikalen Faltung vor. Der U4 5′ SL fehlt jedoch der charakteristische Knick des Kink-Turn Strukturmotivs, der in der proteingebunden Form vorliegt. Stattdessen liegt die freie U4 5′ Stammschleife in einer offenen, auseinander gestreckten Konformation vor. Während zwei

nicht-kanonische G-A Basenpaarungen, die für die bekannte Kink-Turn Struktur in gebundener Form beschrieben wurden, bereits im ungebundenen RNS Molekül ausgebildet sind, weisen die drei ungepaarten Nukleotide der internen Schleife unterschiedliche Anordnungen in der gebundenen bzw. der ungebundenen Struktur auf.
In dieser Dissertation konnte gezeigt werden, dass die freie U4 Kink-Turn RNS in Lösung in einer ausgestreckten Konformation vorliegt, was bedeutet, dass die Konsensussequenz des Kink-Turn Motivs nicht *per se* für den scharfen Knick im RNS Phosphatrückgrat kodiert. Stattdessen liegt das Kink-Turn Strukturelement in Lösung nur zu einem geringen Anteil vor und benötigt die Stabilisierung durch Proteinbindung.

Strukturelle Untersuchungen eines lariat-formenden Ribozym, einem Modellsystem der ersten eukaryontischen Spleißreaktion
In Eukaryonten findet der Prozess des Spleißens von prä-mRNS im Spleißosom statt. Dabei werden nichtkodierende Sequenzen (Introns) aus der prä-mRNS geschnitten und die verbleibenden kodierenden Sequenzen (Exons) zusammengefügt. Die aktiven Komponenten des Spleißosoms sind die basengepaarten U2 und U6 snRNS Moleküle, welche die Spleißstellen der prä-mRNS einander annähern. Während RNA-RNA Interaktionen im Spleißosom im Detail detailliert beschrieben sind, bleibt Aufklärung der Tertiärstruktur des aktiven Zentrums des Spleißosoms aufgrund Größe und Dynamik des Komplexes eine Herausforderung.
In unserer Arbeitsgruppe wird ein Modell-Ribozym untersucht, das eine Transesterifikation ähnlich der ersten Reaktion des Spleißosoms katalysiert und dabei ein Lariat ausbildet. Das Ribozym enthält das phylogenetisch hoch konservierte ACAGAGA-Element, das wichtig für die katalytische Aktivität des Spleißosoms höherer Eukaryonten ist. Des Weiteren folgt die Spleißreaktion des Ribozyms der gleichen Sequenzspezifität der prä-mRNS, d.h. das gebildete Lariat weist einen Adenosinrest an der Verzweigung auf sowie ein Guanosin an der 5' Spleißstelle auf. Aufgrund dieser Merkmale stellt das ausgewählte Ribozym ein ausgezeichnetes Modellsystem dar, und ermöglicht es, die Spleißreaktion sowie die spezifische Funktion des ACAGAGA-Motivs bei der Lariatbildung näher zu untersuchen.
In Rahmen dieser Doktorarbeit wurden umfangreiche Mutationsstudien durchgeführt, um spezifische Reste zu ermitteln, die in Faltung oder Katalyse des Ribozyms involviert sind. Als Ergebnis dieser Studien konnten die Nukleotide

A29, G32, A33 und A35 als funktional wichtige Reste des $A^{29}CAGAGA^{35}$-Motivs ermittelt werden. Weiterhin zeigte sich durch Kombination der Daten aus den Mutationsstudien sowie der Abstandsrandbedingungen aus verschiedenen NMR Messungen, dass die 5′ Region des Ribozyms (Nukleotide 1-27) einen wohldefinierten helikalen Bereich bildet, welcher einen ungewöhnlich hohen Anteil an nicht-kanonischen Basenpaarungen (5 von 11 Paarungen) enthält.
Ein weiteres Ziel dieser Arbeit bestand darin, erste Einblicke in die dreidimensionale Struktur des Ribozyms vor der katalytischen Reaktion zu erhalten. Obwohl das lariatbildende Ribozym ein Minimalkonstrukt der ersten Spleißreaktion darstellt, repräsentiert es aufgrund seines dynamischen Verhaltens und mehrerer ungewöhnlicher Sekundärstrukturen, ein herausforderndes Molekülsystem für spektroskopische Untersuchungen mit Hilfe von NMR. Daher wurden umfangreiche NMR-Messungen wurden mit unterschiedlichen ^{13}C, ^{15}N-markierten RNS-Proben durchgeführt, um eine Vielzahl an komplementären NMR-Spektren zu erhalten und so eine Resonanzzuordnung der einzelnen Ribozymatome zu ermöglichen. Obwohl zahlreiche Isotopenmarkierungsmuster angewendet wurden, war die eindeutige Zuordnung der gemessenen Resonanzen zu den jeweiligen Nukleotiden aufgrund von Resonanzüberlagerungen und Konformationswechsel des Ribozyms nicht möglich. Um diese spektrale Komplexität des Ribozymmoleküls zu überwinden, wurden acht RNS-Mutanten erzeugt. Diese ermöglichten schließlich die eindeutige und vollständige Resonanzzuordnung aller Nukleotide des lariatbildenen Ribozyms. Die Resonanzzuordnung ist eine Vorraussetzung für die Ermittlung von strukturellen Randbedingungen, im Besonderen die Intensitäten von NOE-Peaks, die direkte Abstände benachbarter Atome in einem Molekül reflektieren. Im Falle des lariatbildenen Ribozyms deuten die NOE-Signale auf eine mögliche dreidimensionale Faltung des Moleküls in ein Ribose-Reißverschluss-Motiv und in ein Pseudoknoten-Motiv hin.
Die Daten der Mutations- und NMR-Studien, insbesondere die durch die Analyse von NOESY-Spektren ermittelten Abstandsrandbedingungen, ermöglichten erste Strukturrechnungen des linearen Ribozymmoleküls (vor der katalytischen Reaktion).
Diese vorgestellten Daten stellen einen exzellenten Startpunkt für die Aufklärung der dreidimensionalen Struktur des lariatbildenen Ribozyms dar und erleichtern das Verständnis in Bezug auf dessen Funktionalität.

1 Introduction

1.1 Biological system

1.1.1 Pre-mRNA splicing

Most genes in higher eukaryotes are interrupted by non-coding sequences (introns) that must be precisely excised from the transcribed pre-messenger RNA (pre-mRNA) molecules in order to place the coding sequences (exons) in the correct reading frame. The process, known as pre-mRNA splicing (reviewed in [7] and [154]), is part of a series of post-transcriptional modification steps also including 5′ capping and 3′ polyadenylation that lead to mature mRNA. The intron removal and ligation of exons during splicing take place through two sequential transesterification steps via a branched intermediate, catalyzed by a large and dynamic macromolecular RNP complex, the spliceosome. Following their maturation, mRNAs are transported from the nucleus to the cytoplasm where they undergo translation and, ultimately, decay [73].
Introns reveal common consensus motifs near their 5′ and 3′ ends that are recognized by spliceosomal components and are required for spliceosome formation (see figure 1.1). The 5′ splice site (ss) in higher eukaryotes is represented by the eight-nucleotide long consensus sequence AG/GURAGU, where the splicing site is represented by the slash sign and the nucleotides GU correspond to the beginning of the intron [42]. The area of the 3′ ss is defined by three sequence elements. The branch point sequence (BPS) is usually located 18-40 nucleotides upstream of the 3′ splice junction and is charaterized by the consensus sequence YNYURAY, in which the adenosine represents the highly conserved branch-point adenosine [110]. Further downstream follows a 10- to 12- nucleotide long pyrimidine stretch

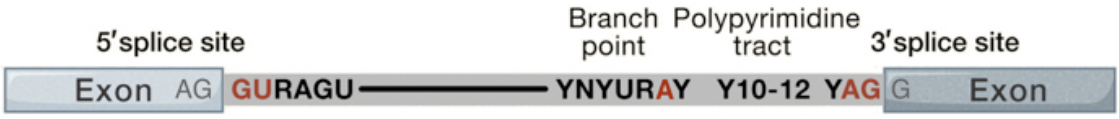


Figure 1.1: *Conserved sequence elements of pre-mRNAs in higher eucaryotes.* | Exons are separated by an intron sequence (gray). The consensus sequences at the 5′ splice site, around the branch point, and the 3′ splice site are indicated with N, R and Y representing any nucleotide, a purine and a pyrimidine, respectively. The polypyrimidine tract is a pyrimidine-rich stretch located between the branch point sequence and the 3′ splice site [154].

which is necessary for branch-site selection. The 3′ ss itself consists of the consensus sequence YAG/G, where the nucleotides AG represent the end of the intron [92].

1.1.2 The splicing reaction

The splicing reaction entails the removal of introns and the ligation of exons by two consecutive transesterification reactions (reviewed in [6] [96]). In order to remove introns precisely, the spliceosome recognizes certain motifs in the mRNA precursor, including the 5′ ss, the 3′ ss and the branch site (see figure 1.1). In the first transesterfication step, the 2′-hydroxyl group of the branch point adenosine attacks the phosphate at the 5′-end of the intron, and through a typical nucleophilic substitution [90][82] releases the 5′-exon while the intron forms a lariat structure [14][30] (see figure 1.2). This lariat intermediate consists of an RNA circle with a 2′-5′ linkage at the branch site adenosine, and a 3′-tail attached to the 3′-exon. Presumably due to a subsequent rearrangement within the active site of the spliceosome, the 3′-OH of the derived 5′ exon is able to attack the 3′ ss in the second splicing step. In the subsequent second reaction the 5′- and 3′-exon are concatenated resulting in the release of the lariat intron. Although the chemistry of the transesterfication reactions is well known [90][82][30], the constitution of the catalytic center in the spliceosome has not been determined yet due to the difficultly in investigating the conformational changes of the involved snRNAs during the splicing process as well the relative inaccessibility of probes to the catalytic center of the spliceosome [143].
Remarkably, the chemical pathway by which the complex spliceosome of higher eukaryotes catalyzes the excision of nuclear introns is similar to the one used by group II introns found in organelles of eucaryotes like fungi, plants and protists as well as in bacteria [102][148]. Group II introns are self-splicing ribozymes, which catalyze their own excision from pre-cursor-mRNAs (reviewed in [74] and which are considered to be the ancestral progenitors of nuclear introns and of the spliceosomal machinery of higher eukaryotes. Recently a high resolution structure of an intact group II intron from *Oceanobacillus iheyensis* has given first structural insights in the catalytic domain of this type of ribozymes [141]. However, there are only few sequence and structural similarities between snRNAs and self-splicing group II introns, thus comparisons between both the two systems although possible but may be misleading.

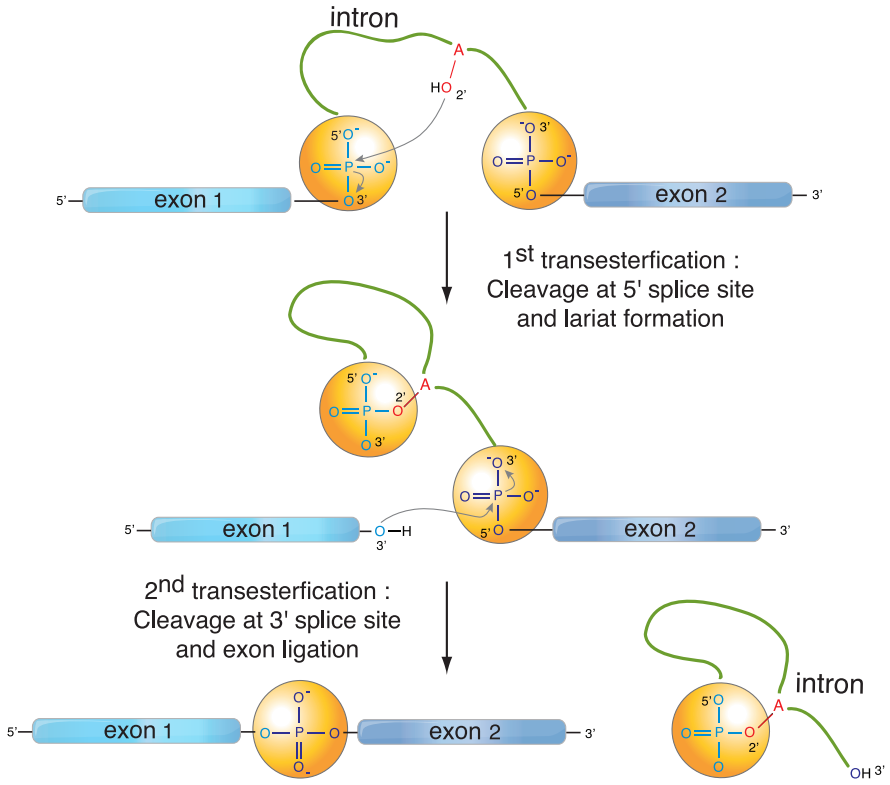


Figure 1.2: *The two steps of pre-mRNA splicing.* | The 2′ OH of the branched point adenosine (red) and the phosphate groups (yellow circles) involved in the reactions are depicted. gray arrows indicate the two nucleophilic substitutions leading to the splicing process. In the first step, the 2′-OH of the branch point adenosine residue attacks the 5′ splice site resulting in a free exon 1 and the lariat intermediate. In a second step, the 3′-OH of exon 1 attacks the 3′ splice site resulting in ligation of exon 1 and 2 and an excised intron in the lariat form.

1.1.3 The splicing machinery

The assembly of the spliceosome around its pre–mRNA substrate molecule is an ordered process that involves five uridine-rich small nuclear RNA molecules (U snRNA). These are associated with proteins and together form U1, U2, U4, U5 and U6 small nuclear ribonucleoprotein particles (snRNPs) [56]. In contrast to ribosomal subunits, none of these particles is able to sustain catalysis alone [154]. Instead, the catalytically active particles are assembled through complicated pathways that include formation of different multi-particle complexes and large rearrangements through a multitude of RNA-RNA, RNA-protein and protein-protein interactions [7] [135].

In the tri-snRNP model of the spliceosomal assembly pathway [7] (see figure 1.3), the first steps of spliceosomal assembly include the recognition of the 5′ splice site and the BPS. Thereby U1 and U2 snRNPs bind to the respective sequence leading to base-pair formation with intronic sequences of the pre–mRNA and, thus,

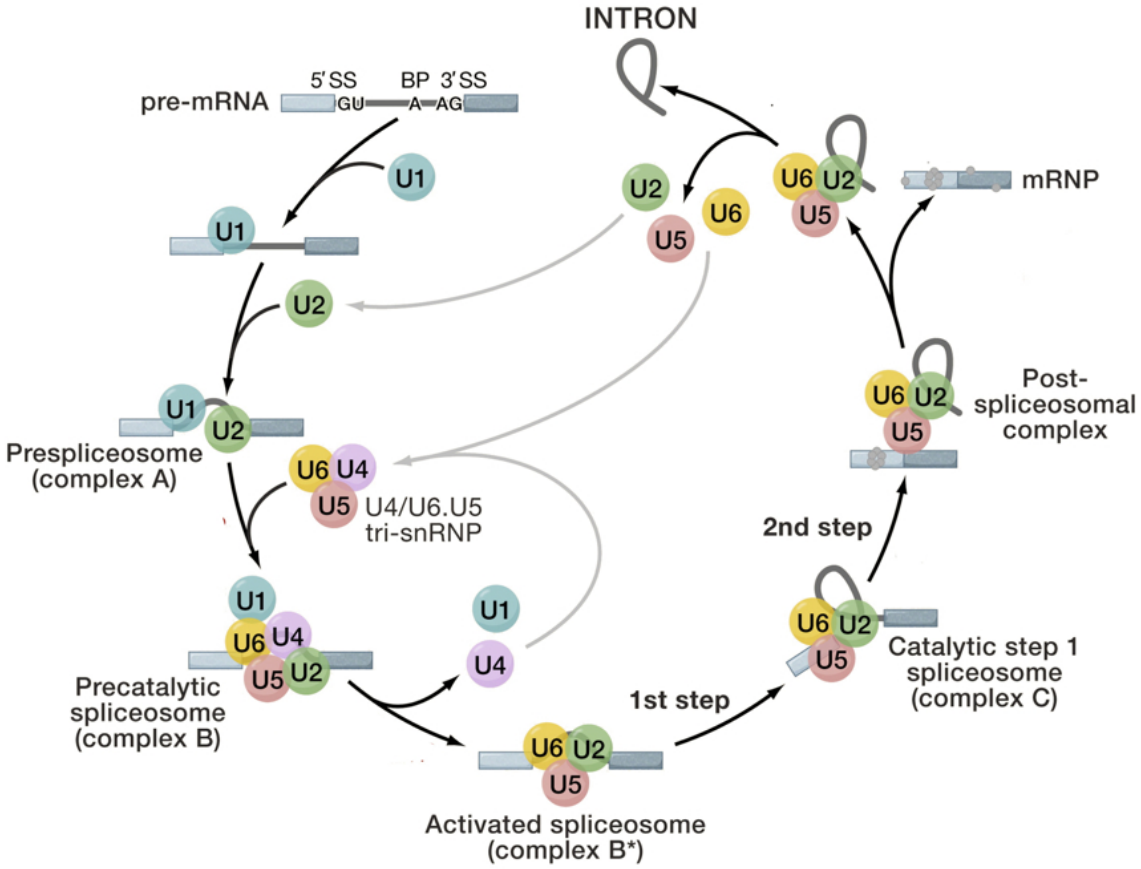


Figure 1.3: *Assembly and disassembly cycle of the major spliceosome [154].* The stepwise interaction of the spliceosomal snRNPs (colored circles) in the removal of an intron from a pre-mRNA with two exons is depicted. Only those spliceosomal complexes that have been biochemically identified in mammalian splicing extracts are shown.

to initiation of the pre-spliceosome, named complex A. The assembly pathway is then continued by association of the pre-assembled U4/U6.U5 tri-snRNP [135] to complex A. Although all snRNPs are present in the formed complex B, it remains catalytically inactive. Major conformational and compositional rearrangements are required at this stage, accompanied by the release of the U1 and U4 snRNPs and leading to an activated complex B*. Subsequently, the first transesterfication step of the splicing reaction takes place in complex B* resulting in the formation of complex C [154]. Prior to the second catalytic step, where the intron is finally excised, additional rearrangements occur in the spliceosome [64]. Following the second transesterfication step, the intron is released in form of a lariat structure and the spliced mRNA in form of a messenger ribonucleoprotein particle (mRNP) [73]. After the splicing process snRNPs are recycled for new rounds of splicing.

1.2 State of research

1.2.1 The U4/U6 snRNP specific proteins 15.5K and hPrp31

Within the splicing cycle (see Figure 1.3) the U4 and the U6 snRNA molecules are associated closely through bp interactions forming a phylogenetically highly conserved Y-shaped U4/U6 interaction domain (see Figure 1.4). The conserved U4/U6 domain contains two intermolecular helices (Stem I and Stem II) of U4 and U6 snRNA [5][50] [112] and an intramolecular 5′ stem loop (5′ SL) of U4 snRNA [6]. It is thought that the U4 snRNA acts as a kind of chaperone that delivers the U6 snRNA to other spliceosomal particles in a repressed state, by masking the catalytic residues of U6 snRNA.
Mutational analyses of U4 snRNA demonstrated that its 5′ SL is required for the transition from B- to C-complex spliceosomes during which catalytic activity occurs [149] [157]. It was suggested that the 5′ SL functions at a stage in the spliceosome assembly subsequent to U4/U6 snRNP formation, since the stem loop is not needed for U4/U6 base pairing *in vitro* [157]. Studies of Nottrott *et al.* revealed that the function of the U4 5′ SL is the recruitment of snRNP proteins

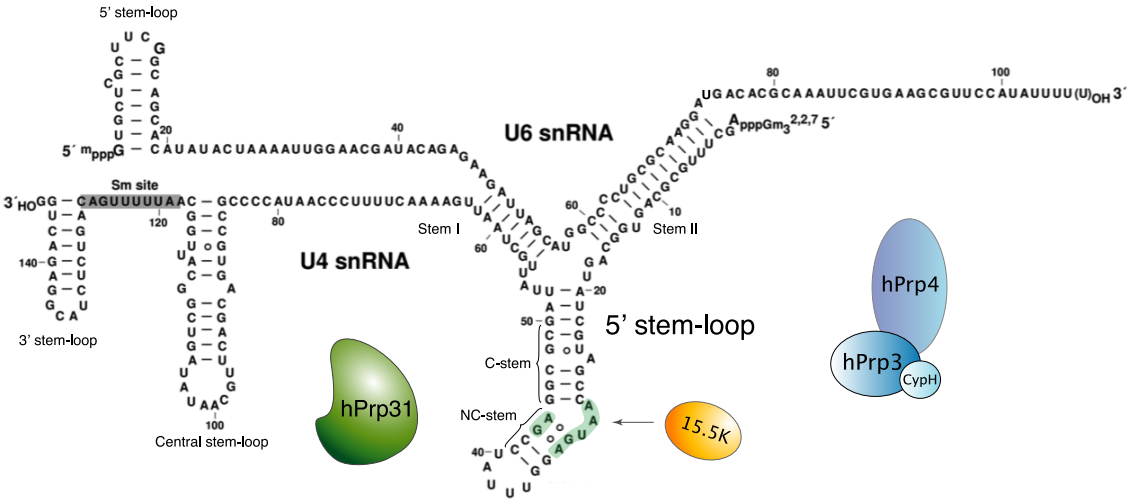


Figure 1.4: *Schematic representation of the secondary structures of the U4/U6 snRNA* | U4 and U6 RNA molecules are associated through a series of base-pair interactions that form two contiguous stems (Stem I and Stem II). U4 comprise the protein binding site for the 15.5K protein, a two-opposite-five region (green) within the internal 5′ stem loop. 15.5K is found to be the nucleation factor which is required for the association of other U4/U6 specific proteins namely hPrp31 and the CypH/ hPrp4/ hPrp3 protein complex. The binding site for Sm proteins (gray), which belong to the common U snRNP proteins, is located near the 3′ end of the U4 snRNA (modified from [6] [149] [157] [120]).

(details below) rather than the interaction with other RNAs [98]. The U4 5′ SL was found to serve as a binding site for the highly conserved 15.5K protein, which was identified as a nucleation factor for further di-snRNP assembly by recruiting the hPrp31 protein as well as the CypH/hPrp4/hPrp3 protein complex [99] (see figure 1.4). The 5′ SL reveals a conserved stem-internal loop-stem structure an element that is found also in RNA molecules of large [3] and small ribosomal subunits [119][160] as well as in the box C/D snoRNAs [156].

The U4/U6 snRNP specific protein 15.5K

The 15.5K protein was the first U4/U6-specific protein identified and it was shown to bind specifically to the 5′ SL of U4 snRNA [98]. The protein structure of 15.5K was solved by X-Ray in complex with a 22 nucleotide long RNA construct of the U4 5′-SL [152] (see figure 1.5). The 15.5K protein folds into a single, compact globular domain of alternating α-helices and β-sheets to form a α-β-α sandwich structure, which is a common protein fold.

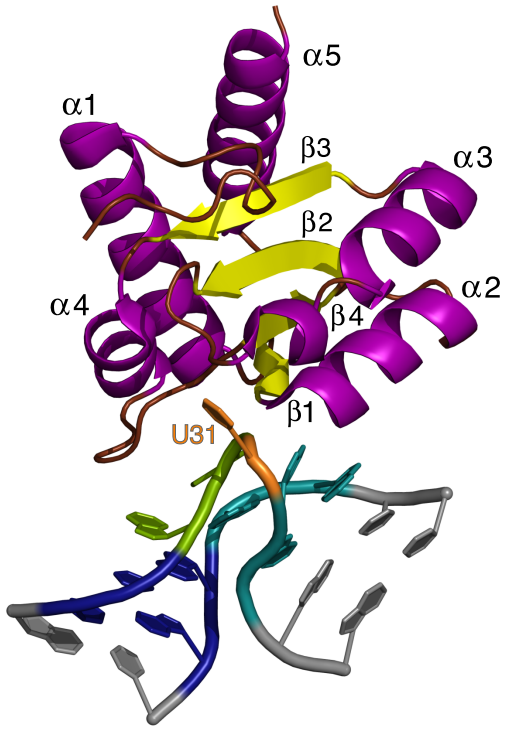


Figure 1.5: *Three-dimensional structure of the 15.5K protein associated to a fragment of the U4 snRNA 5′stem-loop* | In the ribbon plot of the 15.5K protein α-helices are depicted in purple and β-sheets in yellow. Conserved k-turn residues of the canonical-stem and non-canonical-stem of the RNA are presented by blue and light blue colors, respectively. Interloop residues are colored in green and the prominent bulged nucleotide U31 in orange. Detailed information regarding the k-turn motif is given in chapter 1.2.2.

The 15.5K protein belongs to a family of homologous RNA-binding proteins that includes the ribosomal S12, L7ae, L30 and NHP2 proteins [65]. The RNA-binding surface consists of residues located in two α-helices (α2 and α4), a β-strand (β1), and three loop regions (β1-α2, β2-α3, and α4-β4). These residues interact predominantly with the purine-rich (two-opposite-five) internal loop region and the phosphate backbone of the non-canonical stem of U4 5′ SL (see figure 1.4). Through protein binding the nucleotide U31 bulges out into the protein binding pocket of 15.5K, where it is tightly bound. The RNA fold of the U4 5′ SL in complex with 15.5K protein belongs to the the family of kink-turn (k-turn) structural motifs, which is characterized by a sharp kink in the phosphate backbone [63]. As the k-turn is among the objects of investigations presented in this dissertation, it will be described in detail in chapter 1.2.2.

The U4/U6 snRNP specific protein hPrp31

In the hierarchical assembly pathway of the U4/U6 snRNP particle, the 15.5K/ U4 5′ SL complex is strictly required for subsequent binding of the secondary binding protein, the human pre–mRNA processing factor 31 (hPrp31; [99]). The protein hPrp31 comprises 499 amino acids with a molecular mass of 55.4 kDa and is evolutionary conserved. The hPrp31 protein possesses a so-called NOP domain in its central region (see figure 1.6), which is homologous to a domain of the box C/D snoRNP-associated proteins NOP56 and NOP58 [37][153].

The structure of a C-terminally truncated hPrp31$^{78-333}$ protein in complex with 15.5K/ U4 5′ SL has recently been determined using X-ray crystallography and NMR spectroscopy data [71][70] (see figure 1.7). The structural analysis demonstrated that the conserved Nop domain of hPrp31 recognizes a composite RNA-protein binding surface of the U4 5′ SL-15.5K complex.

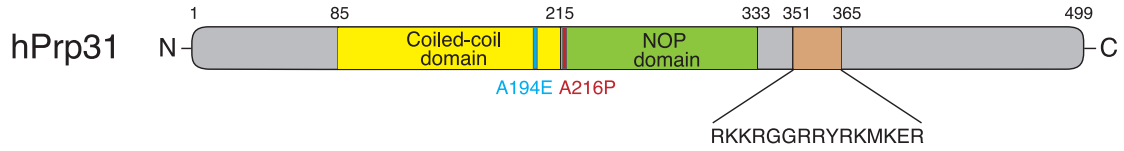


Figure 1.6: *Schematic representation of hPrp31* | The central NOP domain which is also present in boxC/D snoRNP associated proteins is indicated in green, the N-terminal coiled-coil domain in yellow. A third domain, given in brown, contains a series of positively charged amino acids that are predicted to be involved in binding to the canonical stem of the U4 snRNA 5'stem-loop. Mutation sites linked to the autosomal dominant form of *Retinitis pigmentosa* (adRP) are indicated in light blue and red.

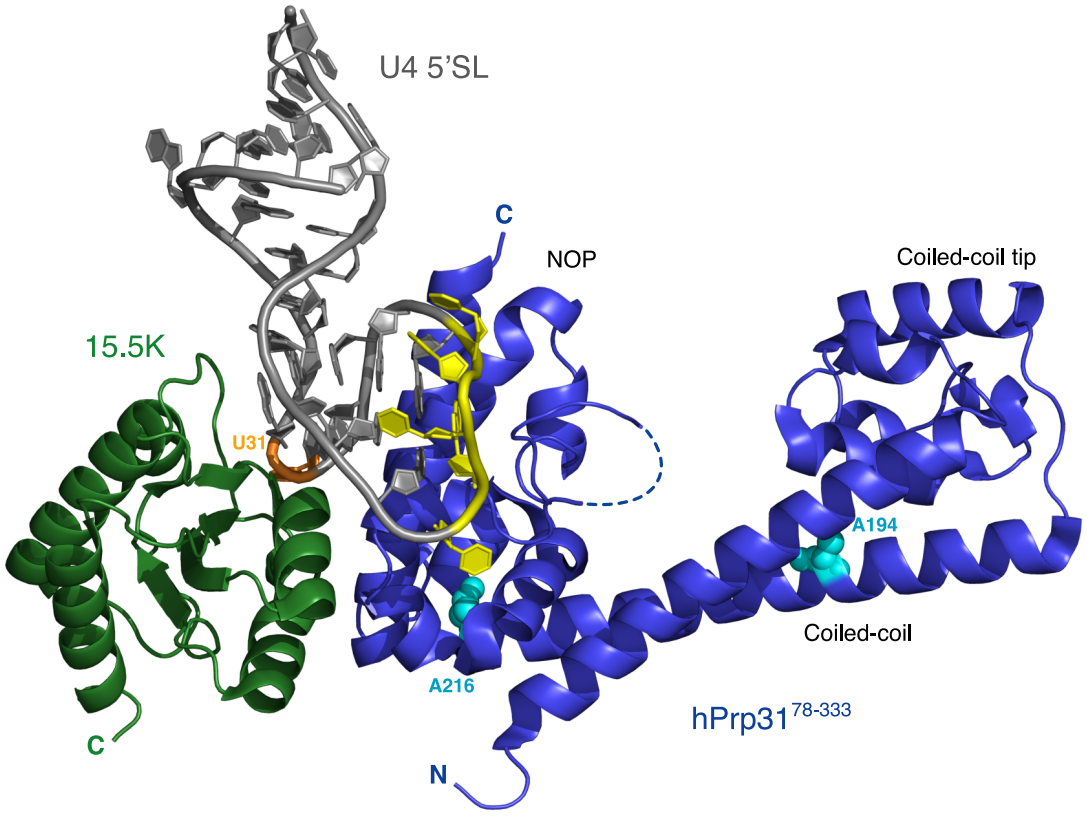


Figure 1.7: *Three-dimensional structure of the hPrp3$^{78-333}$-15.5K-U4 snRNA complex* | hPrp31$^{78-333}$, blue; 15.5K, green; U4 5′ SL, gray, pentaloop residues in yellow and bulged out U31 of the kink-turn motif in orange. The dashed line in hPrp31$^{78-333}$ schematically presents a disordered loop. Alanines 194 and 216, at which missense mutations have been linked to the autosomal dominant form of retinitis pigmentosa are indicated in light blue. The conformation of the core of the 15.5K-RNA complex is mainly unaffected by the binding of hPrp31$^{78-333}$. Structural changes in the RNA are confined to the RNA pentaloop, which becomes ordered through contacts with the NOP domain of hPrp31.

The novel RNA-binding domain sandwiches the U4 5′ SL between both proteins and contacts of hPrp31 stabilize the RNA pentaloop UUUAU, however mostly via sequence-independent interactions. This finding explains the high tolerance of hPrp31 towards changes in the RNA sequence in this region, as previously observed in biochemical studies [120]. These studies investigated a number of different pentaloop mutants including mutants with open pentaloops; all mutants were found to be competent for binding hPrp31. While changes in the pentaloop were tolerated, any elongation of the NC-stem of the 5′ SL lead to abolishment of hPrp31 binding. This observation shows that the hPrp31 NOP domain can act as a molecular ruler measuring the length of the 5′ SL between the k-turn and the pentaloop.

Binding studies with the full-length hPrp31 protein also revealed that it requires the complete C-stem of the U4 5′ SL for binding [120]. This observation and the occurrence of conserved residues in the C-stem of the U4 5′ SL [93] indicate the existence of further contacts between the RNA and the hPrp31 which are not adressed within the structure of the ternary complex of U4 5′ SL-15.5K-hPrp31$^{78-333}$. It is likely that the C-terminal domain of hPrp31 (residue 334-499), which was truncated in the crystallization trials due to its high flexibility, is involved in binding the C-stem of U4 5′ SL. In particular the highly positive charged stretch of amino acids (residue 351 to 365, see figure 1.6), might be suggested as the binding partner of the C-stem of U4 5′ SL. Thus, further structural analysis with the full-length hPrp31 protein are required to understand the role of hPrp31 in spliceosomal assembly.
Beyond described biochemical and structural studies, the hPrp31 gene could be correlated with the autosomal dominant form of retinitis pigmentosa (adRP), a disorder that leads to the degeneration of photoreceptors of the eye [86][153][15]. Two missense mutations (A194E, A216P, see figure 1.6) could be correlated to the human hPrp31 gene (PRPF31). Alanine 194 maps to the second helix of the coiled-coil domain, whereas alanine 216 lies in a short loop connecting the coiled-coil to the Nop domain (see figure 1.7). Although neither of the two residues interacts directly with 15.5K or the U4 5′ SL, the A194E substitution most likely disturbs properties of the coiled coil domain [71], whereas for the A216P substitution a conformational change has been suggested, that negatively affects the stability of the protein [55]. The molecular mechanism of adRP associated with mutations in the protein hPrp31 and further spliceosomal proteins, hPrp3 and hPrp8p, is not yet understood. It has been suggested that since photoreceptors present in the retina have a high demand for splicing mRNAs coding for relevant molecules such as opsin, defects in spliceosomal proteins lead to a disease in this system.

Role of the U4/U6.U5 tri-snRNP in the spliceosomal assembly mechanism
Protein hPrp31 fulfills a dual function in the tri-snRNP complex. On the one hand, hPrp31 binds to the U4 snRNA in the U4/U6 di-snRNP [99]; on the other hand, hPrp31 forms a bridge between the U4/U6 di-snRNP and the U5 snRNP by binding to the U5-specific protein hPrp6 (see figure 1.8)[78].
The assembly of the U4/U6.U5 tri-snRNP complex requires further U4/U6 specific proteins (hCypH/hPrp4/hPrp3), which are also dependent on previous 15.5K binding to the U4 5′ SL [99]. The formation of the tri-snRNP complex most likely takes place in the Cajal bodies [118], where the bridging function of

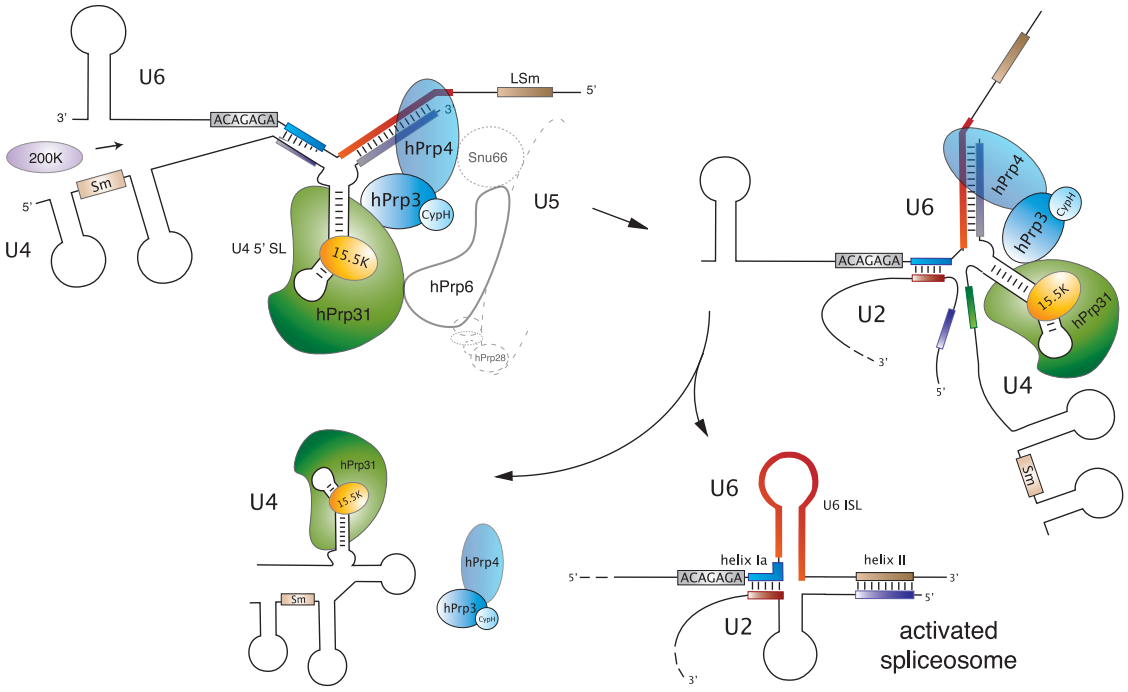


Figure 1.8: *Proposed mechanism for the disruption of the U4/U6 snRNA during activation of the spliceosome.* | Before activation of the spliceosome, the U6 snRNA pairs with the U4 snRNA, forming stem I and II of the U4/U6 snRNA duplex. The model proposes the opening of stem I of the U4/U6 base pairing regions prior to the dissociation of stem II, which is catalyzed by the helicase 200K. However, at present a simultaneous opening of both stems cannot be ruled out. Spliceosome activation is accompanied by the dissociation of U4 snRNA from the U4/U6 snRNP duplex allowing U6 snRNA to pair with U2 snRNA forming helix Ia and II as well as the intramolecular hairpin loop (ISL) of the U6 snRNA. Subunits of the U5 snRNP within the U4/U6.U5 tri-snRNP complex are only indicated (gray circles) as these are not directly involved in the disruption process. The U5 snRNP is essentially required in structural rearrangements of the RNA-RNA network within the spliceosome during its activation.

the hPrp31 protein as well as further protein-protein interactions of U5- and tri-snRNP-specific proteins [78] allow the formation of a stable U4/U6.U5 tri-snRNP. For the next in the spliceosomal cycle, the formation of complex B it is proposed, that the prior stable and preformed U4/U6.U5 tri-snRNP complex is destabilized, which is presumably only possible in the presence of the U2 snRNP. It is thought, that the disruption of the U4/U6 base pairing interaction is performed by U5-specific protein [99], the DExD/H box ATPase (200K) [61][122], which releases the U6 snRNA. This U4/U6 unwinding engages the binding of U2 snRNA and thus the formation of the active spliceosome and the release of U4 snRNP complex and further tri-snRNP proteins (see figure 1.8).

At present, there is little known about the number of discrete RNP remodeling events and the number of structurally distinct spliceosome intermediates during splicing [154]. Furthermore, the spliceosomal model is contradictory discussed in literature (reviewed in [6]). Another model suggests a "holospliceosome" with a simultaneously formation of the penta-snRNP with U1/U2/U4/U5/U6 snRNP [136] instead of the tri-snRNP model described here.

1.2.2 The Kink-turn motif of the U4 snRNA

Kink-turn (k-turn) motifs are universal RNA structural elements present in rRNA mole-cules [63], i.e. snRNAs [152], untranslated regions of mRNAs [79][158] and snoRNAs [156] of Archaea, Prokarya and Eukarya. They represent ubiquitous motifs that function as protein binding platforms within various RNP complexes, such as ribosome, spliceosome and RNA processing enzymes. The k-turn element is a two-stranded helix-loop-helix motif (see figure 1.9). The internal asymmetric loop usually consists of three nucleotides on one strand with no complementary residue on the other strand. The loop is flanked by two stems; the NC-stem containing two non-canonical base-pairs next to the loop region, usually sheared G-A base-pairs, and the C-stem starting with two canonical C-G base-pairs. The k-turn motif owes its name to the sharp bend in the phosphodiester backbone observed when the RNA is bound to its cognate proteins (see figure 1.5) [152] [63] [71]. This sharp bend arranges the two NC- and C-stems at an angle of 120° to each other. The nucleotide on the 5' site of the asymmetric loop (usually a purine) stacks on the C-stem, while the middle one (commonly also a purine) stacks on the NC-stem, and the third nucleotide at the 3' site of the loop protrudes into the protein binding pocket. The k-turn fold is stabilized by long-range contacts between the minor groove edges that are part of the G-A base-pairs and the minor groove of the C-stem (A-minor interaction) [97].
The characteristic sharp kink in the phosphate backbone has been observed for k-turn RNAs in complex with proteins or in the context of large RNP complexes. However, structural information at atomic resolution is not available for unbound k-turn consensus RNA molecules. Thus, the question arises whether the consensus sequence of the k-turn motif shown in figure 1.9 does *per se* codes for the k-turn structural element or whether the kinked backbone conformation of the RNA molecule is induced and stabilized by electrostatic interactions with the positively charged side-chains of the cognate protein.
Both, experimental and theoretical approaches, have been previously used to characterize the structural preferences of k-turn RNAs. An extensive

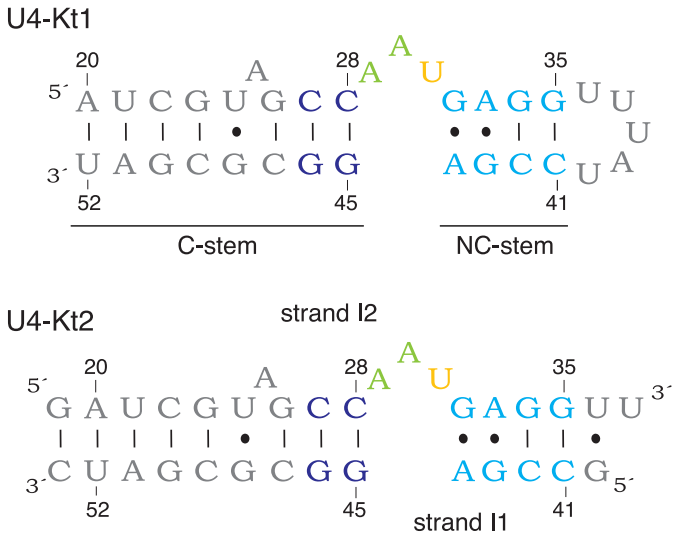


Figure 1.9: *Secondary structure of U4 constructs* | U4 5′ SL of human U4/U6 snRNA (U4-Kt1, [6]) and open loop construct derived from the U4 5′ SL (U4-Kt2) used in this work. Consensus residues of the k-turn motif are color coded, residues in the C-stem and NC-stem are colored blue and light blue, respectively. The bulged nucleotide is labeled orange and further interloop nucleotides are given in green.

experimental study applied gel mobility and fluorescence assays to the ribosomal Kt-7 construct, which is part of helix 7 in the 23S rRNA [63]. This study assessed the presence of the k-turn structural motif in the absence of protein binders and in dependence of the concentration of divalent cations [40]. This work suggested that k-turn RNAs interconvert between an extended and a kinked conformation; in the absence of divalent cations, the ratio between the populations of the extended and kinked conformations was found to be close to 65:35, while in the presence of high magnesium concentrations the population ratio reverts to 30:70. Interestingly, even with a magnesium concentration of 50 mM, a large percentage of extended conformation persists. The study concludes that protein binders are presumably required to fully stabilize the k-turn fold in RNA molecules.

The stability of the k-turn fold in the absence of proteins has been assessed by *in silico* studies using molecular dynamics (MD) simulations for the spliceosomal U4 k-turn RNA [22][23]. The MD trajectories of the unbound RNA, which started from the protein-bound kinked conformation, show a transition to a more extended conformation, where the k-turn is disrupted. The transition from kinked to extended conformation was found to compete with the opening of the G-A base-pairs and ultimately with the disruption of the whole NC-stem. According to Cojocaru *et al.*, the NC-stem of the RNA folds only upon binding to the cognate protein. In another MD study [109] in the group of Šponer, the k-turn fold was

found to be stable also in the absence of protein binders when simulating structures of ribosomal Kt-38, Kt-42 and Kt-58 and for the spliceosomal U4 k-turn. In these simulations dynamics were found to be restricted to hinge-motions of limited amplitude ($\pm \sim 8^{\circ}$) centred on the k-turn region. Long-range A-minor tertiary interactions that stabilize the k-turn fold were found to be preserved during the simulations. Later, the same group performed an extensive implicit solvent conformational search on Kt-38 and identified the presence of a second cluster of structures with a significantly larger angle ($> 140^{\circ}$) between the two stems [109], which is in better agreement with the results of Cojocaru *et al.* [23].
Although these *in silico* studies provided deep insights into the stability of the k-turn motif in unbound RNA molecules, they lack experimental structural data at an atomic level. At present it seems clear that the k-turn fold is stabilized by the interaction with proteins. However, the question whether the consensus sequence of figure 1.9 codes for a kinked, tightly structured RNA motif in the absence of binding partners (and if so to which extent) remains open. Experimental structural information for k-turn RNAs in the absence of binding partners is essential to establish the k-turn fold as a constitutive RNA structural element and, more broadly, to define the conformational space accessible to the consensus sequence of figure 1.9 in solution.

1.2.3 The role of U2, U5 and U6 snRNA in the first step of splicing

In the assembled spliceosome (complex B* and C, see figure 1.3), U2 and U6 snRNAs form base pairing regions through several helices [26][137][162]. A reconstituted U2/ U6 complex with *in vitro* synthesized human U2 and U6 snRNAs has been shown to catalyze an intramolecular reaction similar to the first step of splicing in the presence of the branch site region of the intron [147]. This evidence strongly suggests that these RNA components constitute the core of the active spliceosome.

Current models of RNA-RNA interactions in the activated spliceosome
Splicing catalysis consists of two successive transesterification reactions (see chapter 1.1.2). In the first step, the 2′ hydroxyl group of the branch site (BS) nucleotide nucleophilically attacks the 5′ splice site (ss) to yield a lariat intermediate and a free 5′ exon. In the second step, the free exon nucleophilically attacks the 3′ ss, producing mRNA and an excised lariat intron. In recent years models for each

splicing reaction had been proposed, since the catalytic core of U2 and U6 seems to undergo a conformational switch resulting in remodeling of RNA-RNA interactions and protein composition between the two transesterfication reactions.
First, a three-way junction model for the splicing reaction was established [77] [135][143]. In the activated spliceosome, U2 and U6 snRNAs pair and form a three-way junction (see figure 1.10B) constituted by helix I, II and the internal stem loop (ISL). In this model, the highly conserved AGC triad of the U6 snRNA, located in helix Ib, is associated with the U2 snRNA. In recent years, a four-way junction model was suggested for the first step of splicing (see figure 1.10A), in which the U2 and U6 snRNAs form four helices and where the AGC triad is integrated in the ISL helix. With the latter model, it was proposed that the catalytic core of U2/U6 performs a catalytic switch prior to the second step of splicing, causing the disruption of helix I which is accompanied by formation of new base-pairs for the AGC-triad.
Several studies confirmed a probable change in RNA-RNA interactions during the splicing event. First, it could be shown, that a conformational change involving the 3′ ss of the pre-mRNA occurs prior to the second catalytic step [115]. In this study the 3′ ss was protected from nuclease degradation after the first step of splicing, but is accessible prior to the reaction [124]. This finding was substantiated by other studies showing that many spliceosomal components

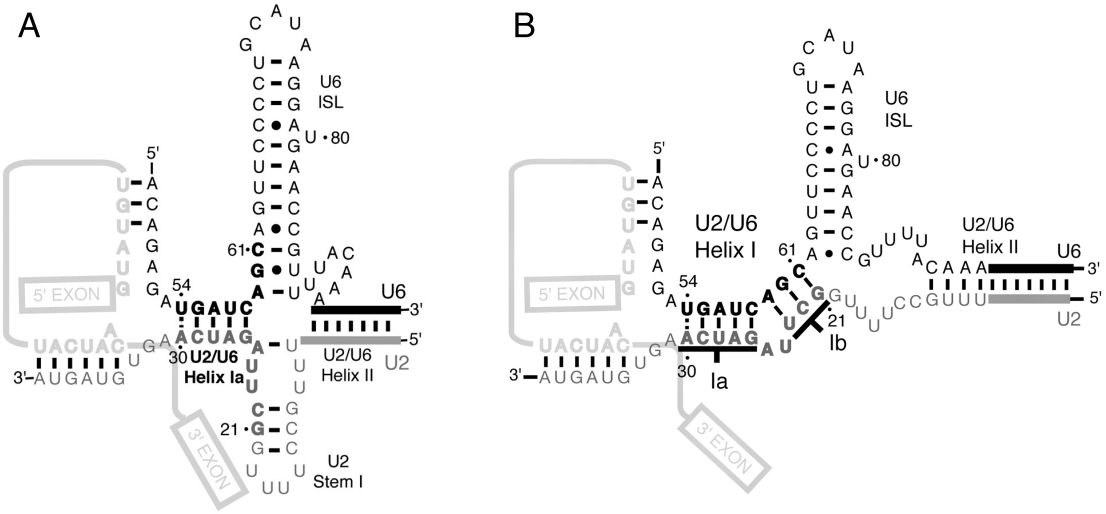


Figure 1.10: *Proposed models of RNA:RNA interactions within the splicesome.* | (A) Four-way junction model of complex B* performing the first step of splicing and (B) Three-way junction model of complex C for the second step of splicing (figures from [87]).

involved in the second step are dispensable for the first [17][18][60][131]. Finally structural insights of the U6 internal stem loop (ISL) proposed a change in base-pair formation of the U2/U6 snRNA catalytic core domain of the spliceosome after the first splicing step [116].

RNA-RNA interactions during first step of splicing

The activated spliceosomal complex B* (see figure 1.3), which performs the first step of splicing, contains four RNA molecules, the U2, U5 and U6 snRNAs as well as the precursor messenger RNA (pre-mRNA) [146]. The active spliceosomal core consists of the U6 snRNA and the 5′-terminal domain of U2 snRNA which are conserved across diverse species [47]. Cross-linking and mutational studies mapped the U6 snRNA to intron sequences near the 5′ ss of the pre-mRNA, whereas the U2 snRNA was cross-linked to the BPS (see figure 1.11) [117][155][58][68]. The most critical U6 snRNA residues cluster in two regions, the ACAGAGA-box and the AGC-triad [25]. Both elements are critical for 5′ ss selection [155][58][68] and have been proposed to assist directly in catalysis of the splicing reaction. The proposed RNA-RNA interactions in the catalytic center of complex B* (see figure 1.11) show a four-helical junction for the U2-U6 complex [137][116], in which the catalytically AGC triad base-pairs within U6. Following the conformational switch U2 and U6 RNA reposition into the three-way junction fold for the second catalytic step.

The U5 snRNP has two major roles in the spliceosome. The particle is essentially involved in structural rearrangements of the RNA-RNA network within the spliceosome during its activation form complex B to B*(see figure 1.8) as well as in the stabilization of the activated spliceosome, where it positions the exons for the ligation step [94][164][95]. The conserved loop sequence in the U5 snRNA (GCCUUUUAC of loop 1) contacts the 5′ exon before and after the first step of splicing (see figure 1.11), whereas interactions with 3′ exon are only detectable after the first step of splicing [131].

Magnesium binding of U2 and U6 snRNA in the activated spliceosome

Interactions of the U2 and U6 snRNA complex with divalent metal ions at specific sites, in particular with magnesium ions, are essential for RNA folding and catalytic activity [41][132]. The first magnesium-binding site of the U6 snRNA could be identified at position U80 in yeast (corresponding to U74 in human) by sulfur substitution experiments. If the substituted phosphoryl oxygen is crucial for function sulfur enables to switch the metal specificity from magnesium to manganese [165]. The second metal binding site has been detected in the region of the

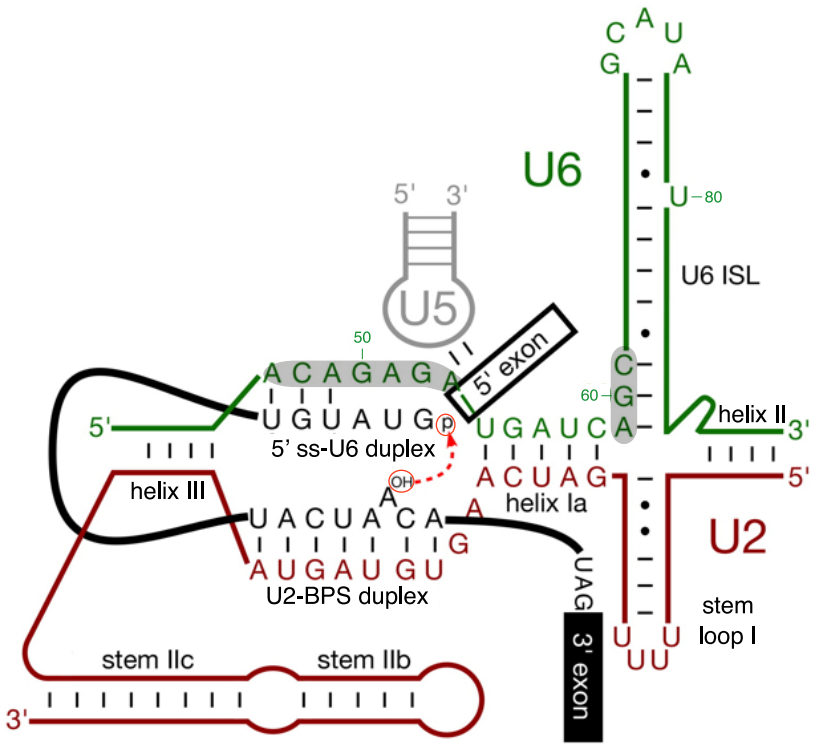


Figure 1.11: *Schematic representation of RNA:RNA interactions that contribute to the first step of splicing according to Smith et al. [130]* | Pre-mRNA is shown in black, U2 snRNA is shown in red, U5 snRNA is shown in gray, and U6 snRNA is shown in green; numbering corresponds to *S. cerevisiae* snRNAs. The first catalytic step of the splicing reaction, the nucleophilic attack of the 5′ ss by the branch point adenosine, is shown.

ACAGAGA-box by the combination of mutational studies and the Förster resonance energy transfer (FRET) method [166]. Further, it was proposed in this FRET study that there is also a magnesium binding site in the region of the four-way junction, which could not be assigned to a specific location yet. The binding sites for metal ions were found to be pH-dependent, with no ion binding observed below a pH-value of 6 to 7, suggesting an involvement in the regulation of the spliceosomal activity.

Tertiary interactions in the activated spliceosome

In order to obtain information for higher-order interactions of the U2/U6 complex prior to and after the first step of splicing, RNA-RNA tertiary studies were performed using hydroxyl-radical probes [59][113]. Before the first splicing reaction, a number of highly conserved RNA structural elements were found in direct vicinity (∼10 Å) of nucleotide U10 of the intron (see figure 1.12). These include (i) the U6-intramolecular stem-loop (ISL), which positions the important metal-

binding nucleotide of U6 snRNA, (ii) the BPS-U2 snRNA duplex, which holds the branch point in position for the first splicing step, and (iii) the highly conserved ACAGAG box, which tethers the 5′ splicing region to the spliceosome. Since the U2 snRNA cleavage sites, caused by the hydroxyl-radical, mapped to the branch-point-binding region, also the pre-mRNA was analyzed revealing no additional cleavages in the branch-point region, the polypyrimidine tract, and the 3′ ss of the pre-mRNA.

The mapping of interactions after the first splicing reaction revealed a differed pattern. Most of the nucleotides around the metal-bindig site of U6-ISL are no longer accessible to hydroxyl radicals and nucleotides of the U2 snRNA are not effected anymore. These observations could be explained by the interaction of further splicing factors prior to the second step of splicing [145][123]. Further, the results may also reflect the RNA rearrangements within the spliceosome [137] [116] as described above. The 5′ ss/U6 snRNA duplex was essentially unchanged

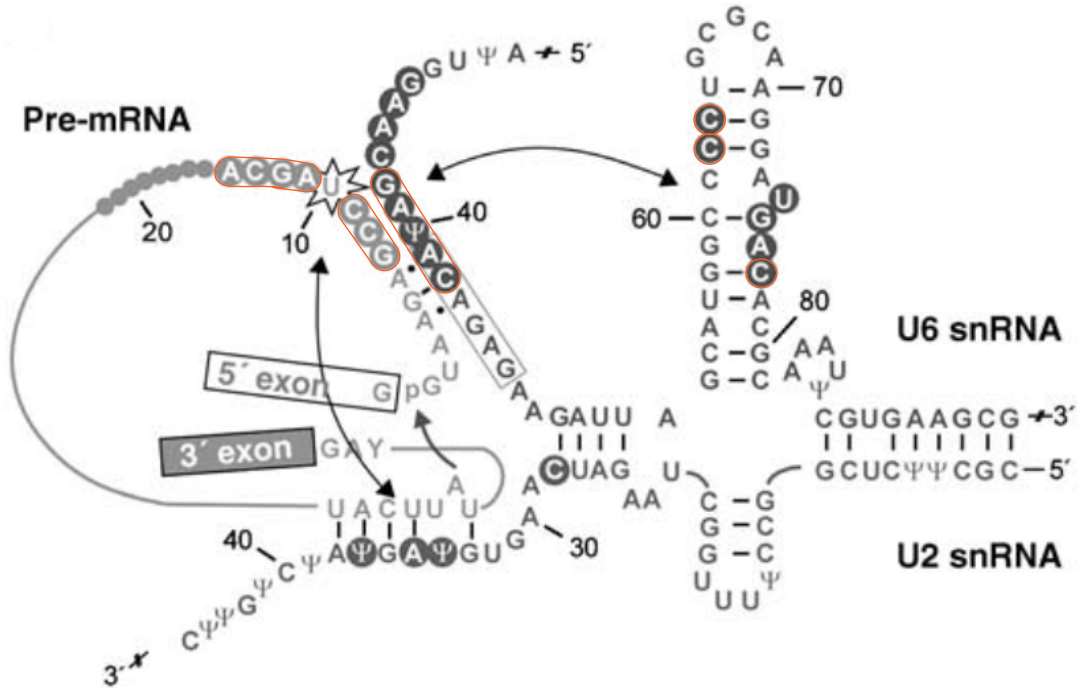


Figure 1.12: *Secondary structure model of the human U2/U6/pre-mRNA before the first step of splicing according to Sun and Manley [137]* | Hydroxyl-radical probing induced cleavage sites in the secondary-structure model of the activated spliceosome. Hydroxyl-cleavage sites in U6, U2 and the pre-mRNA are indicated by circles, prior the first splicing step in gray and after in orange. The double-headed arrows indicate proximity relationships found between U10 (probe site) of the pre-mRNA and U2 snRNA as well as U6-ISL. The attack of the branch point A at the 5′ ss is indicated by an arrow (picture adapted from [113]).

when comparing the spliceosomal complex before and after the first splicing reaction, suggesting that the proposed RNA rearrangement prior to the second step of splicing does not involve this structural element.

1.2.4 The lariat-forming ribozyme - a model system for the first splicing step

Whereas changes in snRNA conformation at the level of secondary RNA-RNA interactions are well known (see figure 1.11), elucidation of the tertiary structure of the spliceosome catalytic center remains a challenge due to the large size of the spliceosomal complex *in vivo*. Thus, model systems small enough to be amenable for structural investigation in solution are very valuable. Tuschl *et. al.* [143] identified a 2'-5'-branch-forming ribozyme (58 nucleotides long), which undergoes a transesterfication reaction with striking similarity to the first step of splicing (see figure 1.13). The ribozyme contains the conserved ACAGAGA box that is essential for catalytic activity in the spliceosome. Moreover, the branching formation is based on the same sequence specificity as in pre-mRNA, including the nucleophile attack of the 2'OH group of an internal adenosine to the phosphate group of a 5'terminal guanosine that leads to a 2'-5' branched lariat. The ribozyme has been selected by *in vitro* selection from a RNA library of 2 x 10^{14} different sequences based on the U6 snRNA sequence. The selected prototype sequence belonged to a major class of sequences, in which 26 residues were conserved (shown in red in figure 1.13). To minimize the ribozyme length, truncated versions of the prototype sequence were designed and tested for the ability to form a lariat with a fast reaction rate.
The final sequence of the model ribozyme (see figure 1.13), was further characterized by mutaional studies of the ACAGAGA-box. The most critical positions regarding lariat formation were A29, G32 and A33. In addition, mutations G34C and A35C showed a tremendous reduction in the reaction rate. The remaining positions of the ACAGAGA box were found to be less important, though all mutants were less active compared to the unmodified ribozyme. Mutations also confirmed the specific need for an adenosine residue at the branch point position 48 as well as for a guanosine at position 1.
Tuschl *et al.* [143] also investigated metal binding of the selected ribozyme by sulfur substitutions. In contrast to the U6 snRNA, no metal binding specifity was found for the conserved ACAGAGA box, whereas residues at the reaction sites G1-A3 (5'end) and C47-C49 (BPS) as well as nucleotides A7-A9, C11/U22,

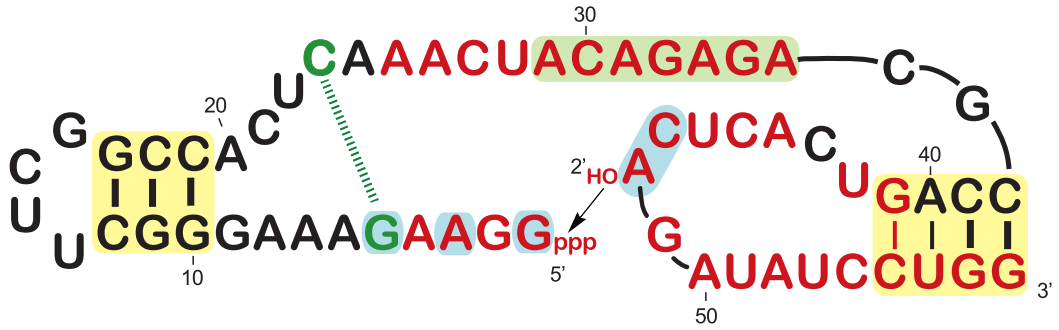


Figure 1.13: *Secondary structure of 2'-5' lariat-forming ribozyme [143]* | Invariant and covarying residues are shown in red and green, respectively. Residues marked in blue indicate nucleotides with similarities to the pre-mRNA in the spliceosomal complex B* and residues in green show the ACAGAGA-box which is highly conserved in the spliceosomal U6 snRNA sequence, whereas yellow labeled nucleotides resemble the spliceosomal U2/U6 helices Ia and III (Figure 1.11). The nucleophilic attack of the 5' end by the branch point adenosine is indicated by an arrow.

A24/A25 seem to be sensitive to metal binding, since substitutions at these positions interfered with lariat formation. Interestingly, the replacement of the terminal 5' triphosphate of G1 showed an eight times greater effect than expected from the the decreased reactivity of the triphosphate upon sulfur substitution. Consequently, the substituted atom, which is located in the reaction center, is either involved in binding a metal ion or other ribozyme residues. Sulfur substitution studies gave first insights into higher-order interactions of the selected ribozyme. However, more interactions need to be elucidated to understand the overall mechanism of catalysis completely. Thus mapping the tertiary structure of the catalytic core of the ribozyme by structural investigations should be the next step for this model system.

1.2.5 NMR Spectroscopy in RNA research

Multidimensional NMR spectroscopy and the development of computational methods led to an increase of structure determination of biological molecules, mainly of proteins ($\sim$ 56.000 structures in PDB, see figure 1.14). Although RNA molecules fulfill a wide repertoire of biological function, in particular regulatory functions as discovered in recent years (miRNAs, snoRNAs, siRNAs, piRNAs, long ncRNAs) [84], there are currently are only 750 RNA structures in the PDB. This is due to the less favourable spectroscopic properties of RNA molecules and to difficulties in RNA synthesis [27]. Although RNA consists only of four individual building blocks (see figure1.15), it is nevertheless an extremely versatile

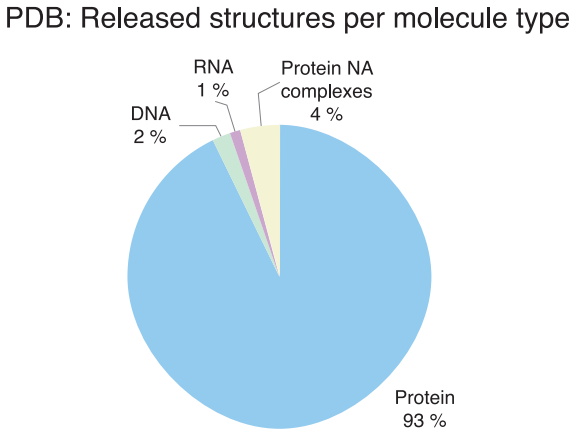


Figure 1.14: *Number of solved 3D structures deposited in the protein data bank [101].*

biomolecule, which makes its structural investigation particularly challenging. In the early 1990s technological advances in chemical and enzymatic synthetic methodologies opened up the possibility of investigating RNA structure by NMR spectroscopy [151]. Numerous NMR techniques were developed to study RNA structure, but in the late 1990s, further progress stalled at a molecular mass limit of 15 kDa (ca. 50 nucleotides). Novel isotopic labeling schemes [125] and development of methods to obtain long range structural information by residual dipolar couplings (RDCs) [105] currently enable the elucidation of RNA structures up to 30 kDa. Nevertheless, the structural investigation of RNA molecules by NMR techniques remains a challening task [144].

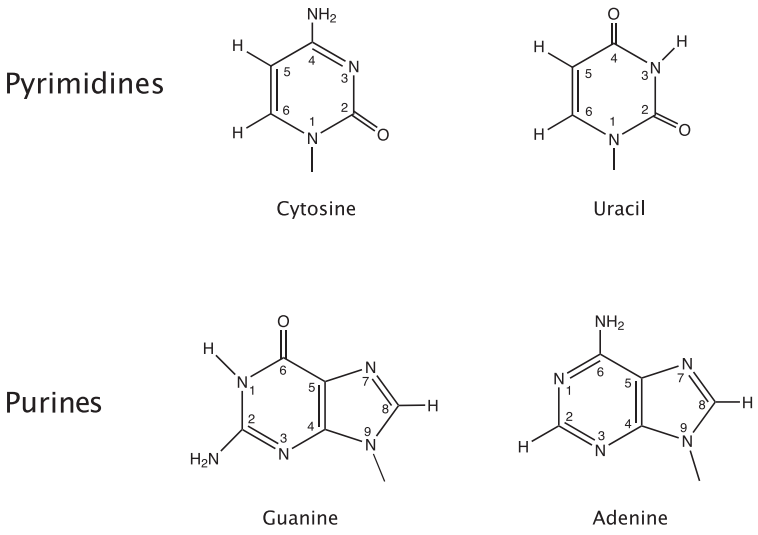


Figure 1.15: *Structure and numbering of nucleic acid bases of RNA*

1.3 Aims of this work

The U4 5'stem loop snRNA

The scope of this project is to investigate the conformation of the U4 kink-turn RNA in solution in the absence of binding proteins and cofactors in order to determine whether the structural kink-turn element is already present in the free RNA molecule.

1. For structural investigations by NMR spectroscopy reasonable amounts (in mg scale) of RNA are required. Therefore a suitable RNA construct for *in vitro* T7 RNA polymerase transcription should be found and the transcription reaction should be optimized for the incorporation of ^{13}C, ^{15}N-labeled NTPs.

2. Several NMR experiments of the U4 5' stem loop RNA construct should be recorded and analyzed to accomplish resonance assignments as well as structural restraints, i.e. NOE-based inter-protons distances, dihedral angles, hydrogen-bond restraints and RDC-based long-range information.

3. In order to test the effects of divalent ions on the conformation of the k-turn region, a magnesium ion titration should be performed and monitored by NMR spectroscopy.

4. For obtaining a second independent measure of the three-dimensional shape of the unbound U4 5' stem loop RNA construct, small angle scattering (SAS) measurements should be performed, for which RNA sample preparation should be conducted within the scope of this work.

5. To gain further insights into the role of U4 snRNA in spliceosomal assembly, the binding of 15.5K and hPrp31 to the U4 5' stem loop RNA construct should be investigated. In particular, the binding site of the C-terminal domain of hPrp31 to the U4 RNA should be studied.

Studies of a lariat-forming ribozyme:
A model RNA construct for the first splicing step of eukaryotic systems

The aim of this project is to investigate the conformation of a 2'-5' branch forming ribozyme, which undergoes a transesterfication reaction with striking similarity to the first step of splicing in the spliceosome of higher eukaryotes. In this work following steps for the investigation of the ribozyme should be approached:

1. Mutational studies for the determination of specific sites that can be either linked to the secondary structure/ folding of the ribozyme or to the catalytic activity of the RNA molecule.

2. The collection and analysis of isotope-filtered and -edited NMR spectra of differentially ^{13}C and ^{15}N-labeled ribozyme samples (A, C, G, U) should be performed in order to accomplish the assignment of individual resonance peaks of the NMR spectra to the respective nuclei within the molecule.

3. To obtain the complete resonance assignment, further NMR experiments of active point-mutants of the ribozyme are required, for which RNA samples should be prepared and isotope-filtered and -edited NMR spectra should be recorded.

4. Direct distance constraints (NOEs) should be determined to gain first insights into the three-dimensional structure of the ribozyme prior to catalysis, which might be analogous to complex B* in the spliceosomal cycle.

5. In order to observe structural changes arising upon the transesterfication reaction, the ribozyme should be studied in its lariat form, which might resembles complex C of the spliceosome after the first splicing reaction.

2 Methods

A list of materials and the description of standard methods in molecular biology are enclosed in the Appendix section (see A.1 & A.2).

2.1 RNA synthesis

In vitro transcription by T7 RNA polymerase is currently the most widely used method to obtain RNA in multi-milligram quantities for structural studies [89]. This is mainly due to the simple preparation of the polymerase and the almost unlimited range of sizes and sequences of RNA molecules that can be synthesized. Moreover the efficiency and accuracy of the synthesis step as well as the more efficient usage of ^{13}C-, ^{15}N-labeled or ^{2}H-deuterated nucleotide triphosphates make this method favorable compared to chemical synthesis. Major limitations restricting this method are the poor transcription efficiency of non-G-rich initial

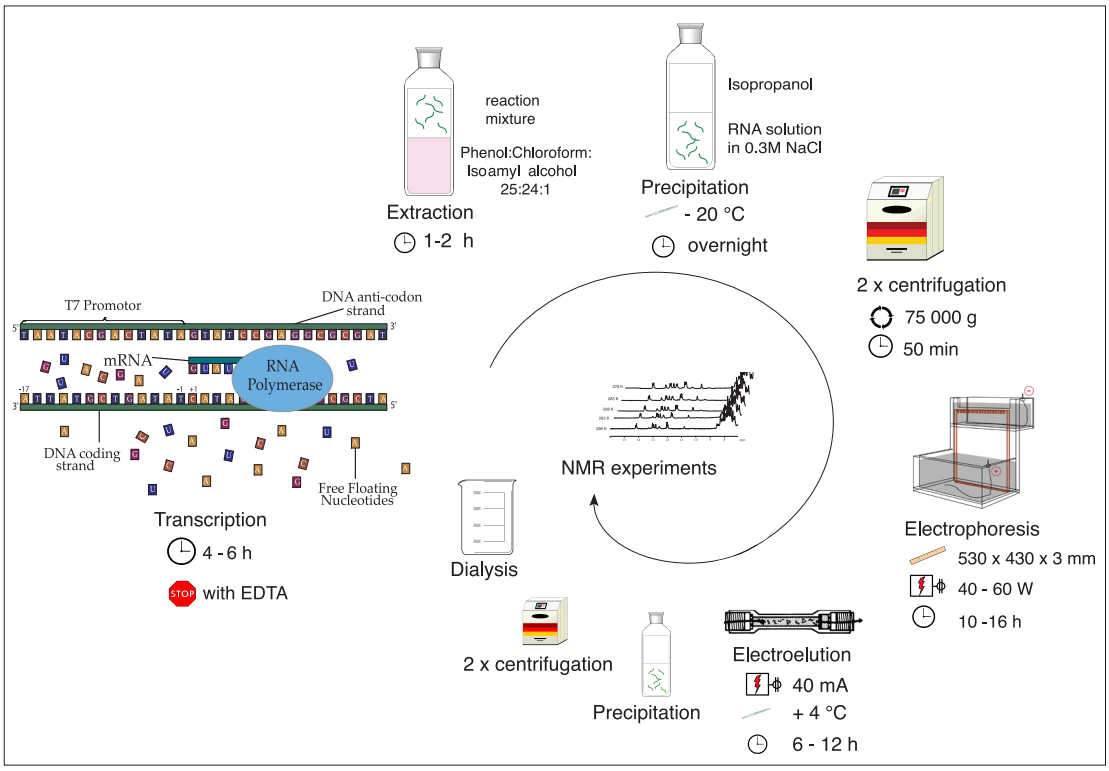


Figure 2.1: Schematic work flow of the synthesis of isotope-labeled RNA molecules by *in vitro* transcription for NMR sample preparation.

sequences and the 3´-end heterogeneity of the transcripts [89][104], since the polymerase may add one or more nucleotides at the 3'-end. In this work this undesirable 3' heterogeneity had been overcome by *trans*-hammerhead ribozyme cleavage (see chapter 2.1.7). Furthermore the formation of undesired secondary structures of the DNA template was prevented by the use of completely double stranded templates. The formation of pyrophosphate, which accumulates during transcription and inhibits T7 Polymerase by sequestering magnesium ions, was prevented by the use of pyrophosphatase [24].

2.1.1 T7 polymerase expression and purification

E. coli Bl21(DE3) cells with His-tagged T7 RNA polymerase were grown in LB media with 100 μg/ml ampicillin to an OD_{600nm} of 0.5 to 0.6. After induction with 0.5 mM IPTG, the cells were incubated for additional 4 hours before being harvested at 4000 x *g* for 10 min. Pellets were stored over night at -80°C. The following purification steps were performed on ice or in the 4°C cold room. First each cell pellet was resuspended in 25 ml of lysis buffer (see table A.9). Then a 30 min centrifugation step followed at 30000 x *g* and 4°C and subsequently imidazole was added to a final concentration of 20 mM. The soluble extract was purified over Ni-NTA agarose beads using a 20 ml economy column. After binding the supernatant on the agarose under gravity flow, the beads were washed with 5 column volumes (CV) of lysis buffer containing 4 mM imidazole followed by 10 CV washing buffer (see table A.9). The His-tagged T7 RNA polymerase was eluted from the agarose using 5 CV elution buffer (see table A.9). The column was equilibrated again with washing buffer and the supernatant was purified a second time. A buffer exchange to the long term storage buffer (see table A.9) was performed over night by dialysis using a dialysis membrane of 10 kDa MWCO. Then 100 % glycerin was added in equimolar ratio before the T7 Polymerase was stored at -20°C.

2.1.2 Transcription optimization

In order to reach maximal RNA yields an empirical determination of each factor required for transcription is crucial. This comprises Mg^{2+}, NTP-, T7 polymerase and template concentrations as well as the reaction time. Since a small change in the template sequence may lead to drastic changes in the optimal conditions,

the optimization procedure need to be repeated for every sequence. The same applies to different production charges of NTPs or T7 Polymerase.
The NTP stocks for the transcription were prepared from lyophilized material dissolved in nuclease free water and 1 M NaOH for pH adjustment (pH 8). The optimization was performed in small scale reactions of 20 μl for unlabeled NTPs and 10 μl reactions working with ^{13}C-, ^{15}N-labeled NTPs or ^{2}H -H3',H4', H5',H5" deuterated NTPs. In a first trial a combined adjustment of Mg^{2+}- and NTP-concentrations was performed by varying concentrations from 10 to 50 mM in 10 mM steps. In case the transcription of the respective DNA template resulted in the desired RNA transcript, the conditions were further optimized in 5 mM steps. Subsequently the T7-Polymerase concentration was varied from 5 % to 25 % (v/v) in 5 % steps and DNA template concentration from 0.5 μM to 10 μM using DNA oligomers or from 2.5 ng/μl to 40 ng/μl using plasmid DNA.

2.1.3 Large scale RNA synthesis

After optimizing the *in vitro* transcription conditions in small scale, the upscale was performed in 1 ml reaction volumes. This step is necessary to estimate the final scale which then should yield RNA in milligram amounts. The final volume was varied from 5 to 40 ml.
For constructs where the 3' heterogeneity was removed by hammerhead cleavage 20 ml reaction volumes were used for transcription. In table 2.1 an example reaction for varying volume is shown. The reaction was usually stopped after six hours by the addition of EDTA (pH 8) to a final concentration of 50 mM.

Table 2.1: Transcription scheme for varying volumes of transcription reaction

Name	Stock concentration	End concentration
DNA template	0.1 mM	0.5 μM to 10 μM
10 x transcription buffer	10 x	1 x
$MgCl_2$	1 M	10 to 50 mM
NTPs		10 to 50 mM
each NTP (ATP, CTP, UTP, GTP)	100 - 300 mM	2.5 to 12.5 mM
PEG 8000	80 ng/ml	0.8 mg/ml
DTT	1 M	5 mM
inorganic Pyrophosphatase	20 U/μl	3 U/ ml
T7 polymerase		1/4 to 1/20
H_2O		

2.1.4 Protein and DNA removal in RNA samples

Phenol extraction

For phenol extraction a phenol:chloroform:isoamyl alcohol mix is added in equimolar ratio to the RNA solution. In this mixture phenol dissociates proteins from DNA, chloroform denatures the protein and maintains the separation of the organic and aqueous phase, and isoamyl alcohol prevents foaming of the solution. Following this treatment, the denatured protein forms an opaque layer between the phases. Within the pH range of 7 to 8 DNA remains in the aqueous phase while it moves to the organic phase when phenol is adjusted to pH 5 to 6.
The RNA solution was mixed with the phenol:chloroform:isoamyl alcohol mix (25:24:1), the mixture was shaken until an emulsion resulted. A phase separation was accomplished by centrifuging the mixture for 5 min at 3000 x *g* and 4°C. To remove phenol traces from the water phase chloroform:isoamyl alcohol (24:1) was added for solvent extraction followed by centrifugation. Finally sodium chloride was added to the RNA solution to yield a final concentration of 0.3 M and the RNA was precipitated with isopropanol (see chapter 2.1.5).

Ion exchange purification

Alternatively to the widely used phenol extraction, the ion exchange resin diethylaminoethyl (DEAE) can be applied for protein and DNA removal. This method is based on the interaction between negatively charged phosphates of the RNA backbone and positively charged DEAE used as weak anion exchanger. RNA binds to anionic DEAE by using the low ionic strength buffer 0.1 M sodium acetate. By increasing the salt concentration in a stepwise manner up to 2 M in the elution buffer, first the template DNA and then RNA were eluted (see figure 2.2). The eluted RNA pool consists, next to the RNA of interest, of all T7 Polymerase products, such as the abortive initiation products or products with 3' overhangs. The RNA of interest was obtained by gel electrophoresis. Prior to electrophoresis, salt was removed from the RNA solution and the RNA was concentrated by isopropanol precipitation (see chapter 2.1.5).

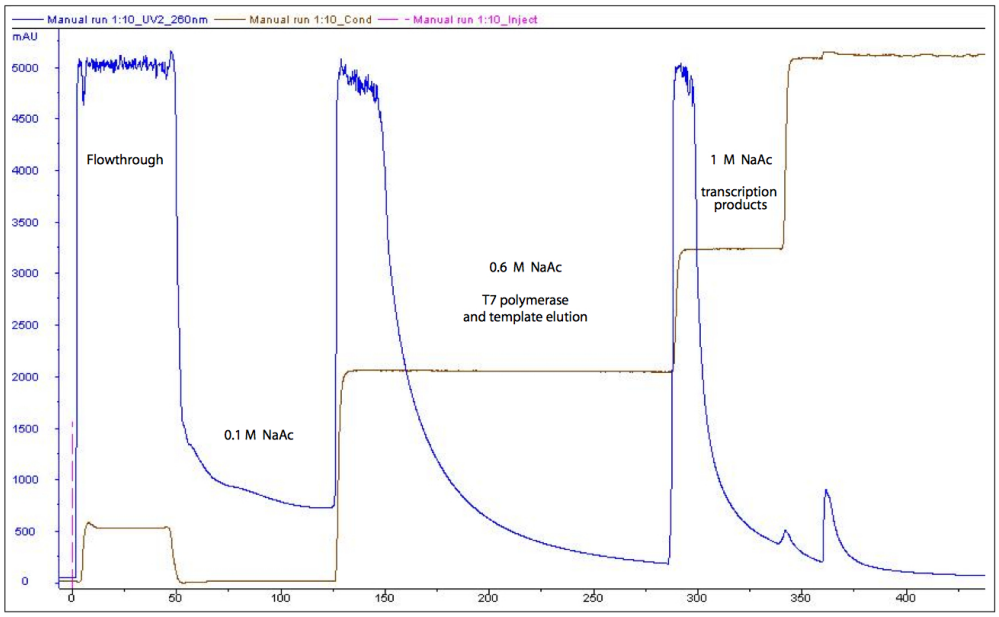


Figure 2.2: *Chromatogram of an DEAE FF 5 ml column run* | The injected sample was a reaction mix of a 20 ml *in vitro* RNA synthesis of HH-I1. Buffer A: water, buffer B: 2M sodium acetate. The binding of RNA was performed with 0.1 M NaAc while the elution was carried out in a stepwise manner using gradients of 30%, 50% and 100% buffer B.

2.1.5 RNA purification

Isopropanol precipitation

Nucleic acid precipitation was used to concentrate and purify nucleic acids; in this procedure addition of monovalent cations and lowering the temperature was used to decrease the solubility of RNA. In this work sodium chloride was added to the RNA solution to a final concentration of 0.3 M followed by one volume of cold isopropanol. After the mixture was stored at -20°C over night, RNA was centrifuged down at 30000 x *g* for 1 h and 4°C or at 50000 x *g* for 50 min. The RNA pellet was washed with 70 % cold ethanol and centrifuged again. Following the careful removal of the supernatant, the pellet was dried in a vacuum concentrator before resuspension in RNase-free water or a mix of water and denaturing loading buffer (see table A.9).

Denaturing PAGE

Purification of the desired RNA product was achieved by denaturing polyacrylamide gel electrophoresis (PAGE) (530 x 430 x 3 mm, see table A.9) and subsequent elution from the gel matrix. Depending on the size of the RNA, the electrophoresis was set for 10 to 15 h at 45 to 60 W and 1 x TBE as running buffer (see table A.9).

Electroelution
Electroelution was used to recover the RNA band of interest from large polyacrylamide gels. The gel slices containing the RNA were excised under UV illumination. The slices were placed in Elutrap systems which were then placed into horizontal gel electrophoresis chambers filled with 1 x TBE buffer. Depending on the RNA size and the length of the gel slices a constant current of 5 to 10 mA was applied for a period of 6 to 12 h, in which molecules migrated from the gel slice into a trap area formed by a semi-permeable and a non-permeable membrane. The RNA was collected every 1 to 2 h and the pooled eluates were concentrated by isopropanol precipitation.

Crush & Soak
For small scale transcriptions, gel pieces were too small after electrophoresis to recover the RNA by electroelution; instead the crush-and-soak method was used. The polyacrylamide gel bands were crushed in smaller pieces and soaked in 0.3 M NaCl solution. The mixture was placed over night on a rotational wheel at 4°C. Afterwards gel pieces were centrifuged down and the RNA-containing solution was precipitated with isopropanol.

2.1.6 Determination of nucleic acid concentration

The concentration of the RNA or DNA samples was determined by UV spectrophotometry. The bases in nucleotides have an absorption maximum around 260 nm, which enables efficient quantification of nucleic acids down to 2.5 ng/ μl. The linear relationship between absorbance and concentration of an absorbing nucleotide species is described by the Lambert Beer law:

$$A = \epsilon \times b \times c \tag{2.1}$$

where ϵ is the wavelength-dependent molar absorptivity coefficient with units of M^{-1} cm^{-1}, b is the path length, and c is the analyte concentration. To calculate the extinction coefficient at the wavelength of λ= 260 nm, the following equation was used:

$$\epsilon_{Oligo,260nm} = \epsilon_A \times n_A + \epsilon_C \times n_C + \epsilon_G \times n_G + \epsilon_U \times n_U \tag{2.2}$$

where ϵ_A = 15.02 M^{-1} cm^{-1}, ϵ_C = 7.07 M^{-1} cm^{-1}, ϵ_G = 12.08 M^{-1} cm^{-1}, ϵ_U = 9.66 M^{-1} cm^{-1} and n describes the number of the respective nucleotides in the oligomer [13].

2.1.7 Hammerhead cleavage

Hammerhead design

The design of hammerhead (HH) RNA was based on the work of Shields *et al.* [127], where amongst other *trans*-cleaving HH systems a fast-cleaving variant was studied. Fast cleaving hammerheads were first described by Clouet-d'Orval & Uhlenbeck [20][21]. In figure 2.3 the hammerheads designed in this work are depicted. Here the secondary sequence of hammerhead I1 and I2 (HH-I1 and HH-I2) bound to their substrates U4 snRNA I1 and I2 are shown in figure 2.3A and 2.3B, respectively. The hammerhead RNA, the substrate RNA, and the product RNA were all designed in such a way that they differ in length by at least five nucleotides to accomplish the purification on polyacrylamide gels.

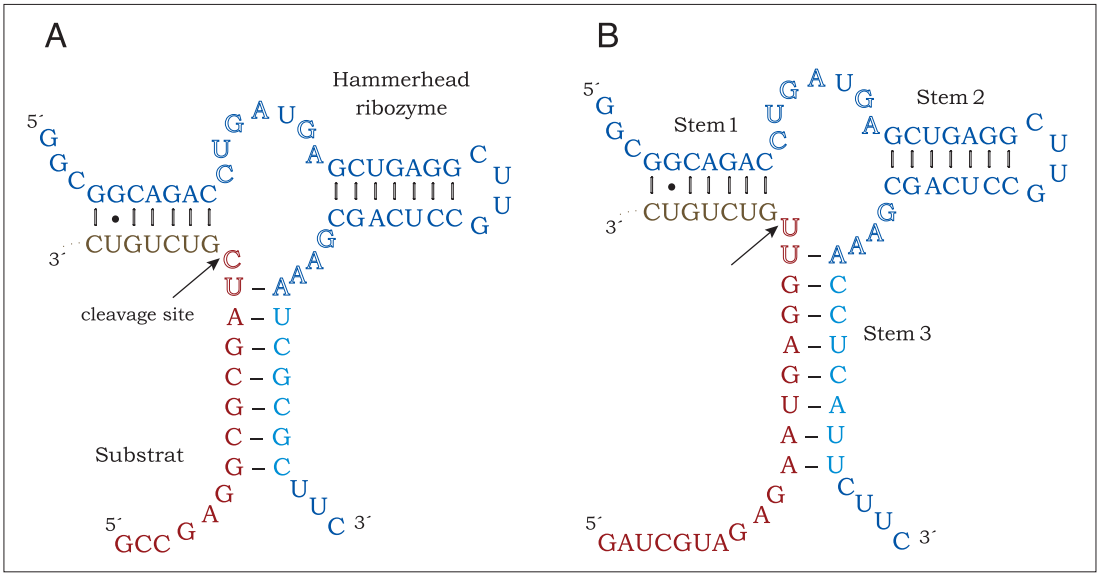


Figure 2.3: *Hammerhead design* | *In vitro* transcribed hammerheads composed of 48-nucleotides for hammerhead I1 cleavage (A) and of 50-nucleotides for hammerhead I2 cleavage (B). The cleavage of the 32-nucleotide I1 substrate yielded the 14-mer (red) I1 construct and the cleavage of the 48-nucleotide I2 substrate yielded the 19-mer (red) I2 construct used for NMR and SANS studies. The arrow indicates the site of cleavage. The hammerhead consensus nucleotides [85] [114] are shown in empty letters. The hammerhead domain is composed of a conserved central core held together by three stems indicated respectively. blue = Hammerhead, light blue = part of HH designed for the respective substrate, red & brown = substrate, red = product, brown = cleaved overhang.

Hammerhead DNA templates
To create HH DNA templates for RNA synthesis, DNA templates (see table A.5) were introduced into the vector *pUC19-EcoRI-HindIII* followed by large scale plasmid purification and *PstI* digestion (see chapter A.2). RNA synthesis of the HH RNA was performed as described in chapter 2.1.

Hammerhead cleavage reaction
To reach complete conversion of substrate RNA to hammerhead-cleaved product RNA, an optimization of Mg^{2+} was performed, testing concentrations of 10 mM, 25 mM, 50 mM, 75 mM and 100 mM. Since the amount of enzyme is limited, the optimization procedure was extended to determine the minimum ratio of enzyme to substrate that varied from 0.5:1 to 2:1.
Reactions were accomplished with 40 mM Tris pH 7.4, 100 mM NaCl, 2 mM EDTA (pH 8). First an annealing step at 85°C was carried out for 3 to 5 min depending on the reaction volume. Following 30 min incubation at room temperature, the reaction was triggered by adding 50 mM Mg^{2+} to the RNA mix, which was then incubated at 37°C over night. The reaction status was checked with an aliquot of the mix using a denaturing RNA gel. In case the reaction was incomplete, the annealing step was repeated and 25 mM of $MgCl_2$ was added. In case of a complete reaction, the hammerhead RNA and the RNA products were separated by electrophoresis using a large scale denaturing polyacrylamide gel. The product and the hammerhead RNAs were then electroeluted and concentrated by isopropanol precipitation.

2.2 Effects of RNA mutations upon lariat formation

Extensive mutational studies (see figure 2.4) were performed in order to determine which nucleotide residues are required for lariat formation of the 2'-5' lariat-forming ribozyme and to investigate the base pairing pattern in the RNA molecule. Lariat formation was tested for each mutation by observing a retardation of effected RNA molecules while moving through a gel. In case the mutated RNA molecules are still able to form the lariat, an additional band can be detected on the gel above the band of the linear form.
To observe whether the introduced mutations (see figure 2.4) effect the lariat-forming ribozyme, its ability for lariat formation was tested. RNA constructs were prepared by 0.1 ml *in vitro* transcriptions. Synthesized RNA constructs were separated by electrophoresis with 20 x 20 x 0.3 cm denaturing polyacrylamide

gels and the band of interest was sliced out. The RNA was recovered by using the crush & soak method (described in chapter 2.1.5) with 3 ml of 0.3 M NaCl over night and after isopropanol precipitation the RNA pellet was dissolved in 40 μl RNase-free water. The RNA concentration was determined by adsorption at 260 nm and 10 to 25 μg of RNA was used for lariat formation tests.
Since the transesterfication reaction that leads to the lariat formation is induced by Mg^{2+} addition, the purified RNA was incubated over night at 30°C in 100 mM KCl, 25 mM $MgCl_2$, 30 mM Tris pH 7.6. As described above the ribozyme in its lariat form can be clearly seperated from the linear form on a denaturing gel. Movement of the lariat form is hindered compared with the linear form, since the ribozyme is more compact in its lariat form. Thus the lariat form migrate slower through the gel pores than the linear form delivering a straight forward monitoring if mutations effect the transesterfication reaction of the ribozyme.

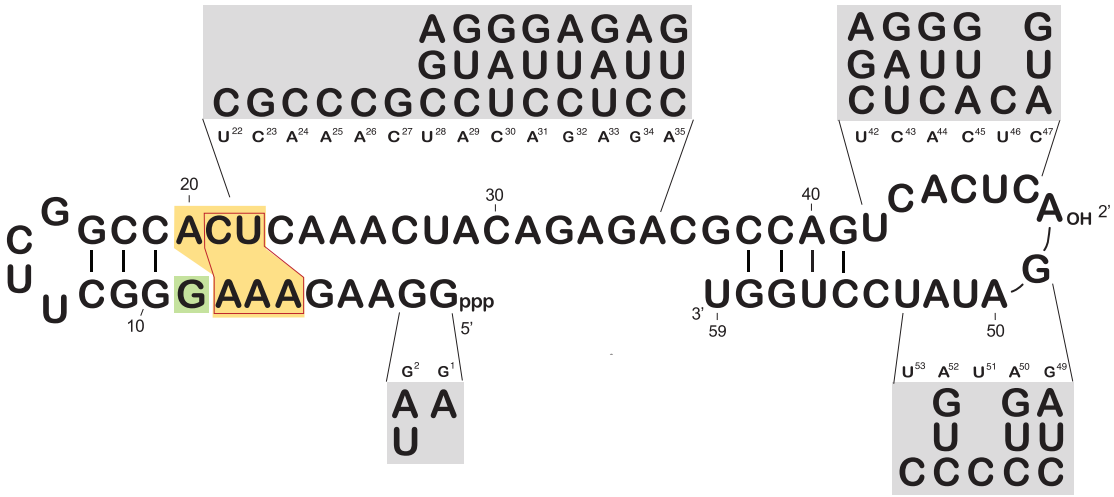


Figure 2.4: *Mutations performed with the lariat-forming ribozyme* | In gray boxes all 60 substitution mutations, which had been performed to investigate nucleotides required for lariat formation, are given in gray boxes. Furthermore three deletion mutations were performed which are indicated with colored boxes and with a red line. The seven double mutations which were performed are not included in the figure, but described in table 5.1.

2.3 Protein synthesis

2.3.1 Expression of unlabeled and ^{15}N-labeled proteins

E. coli cells with the respective plasmid for protein synthesis were grown on LB-plates with the appropriate antibiotic over night. A single colony was picked to inoculate a 2 ml pre-culture, which was grown, transferred to a 50 ml overnight pre-culture, and subsequently used to inoculate one liter of respective media with the appropriate antibiotic (see table A.10). The cell culture was performed at a temperature of 37°C and monitored until the desired absorption at 600 nm was reached, namely an OD_{600nm} of 0.5 to 0.6 for the (^{15}N-) GST-15.5K protein and 0.6 to 0.8 for the (^{15}N-) His-15.5K and the His-hPrp31 proteins. For the growth in deuterated media an OD_{600nm} value of 1.1 was required (^{2}H-GST-15.5K and ^{2}H-His-15.5K proteins). Protein expression was induced with 1 mM IPTG and cells were placed at 22 °C. The incubation time was 12 to 16 hours for unlabeled and ^{15}N-labeled protein expression. For ^{2}H-labeled expression the stationary phase was usually reached at an OD_{600nm} of 1.5 to 1.6 after 24 hours. Cells were harvested at 4000 x *g* for 10 min and pellets were stored at -80°C.

2.3.2 Expression in 99% deuterated media

Since deuterium oxide (D_2O) inhibits both cell growth and cell division *E. coli* cells need to be gradually adapted to deuterated media. For deuteration all buffers and stock solutions were prepared in 99.9% D_2O and the media was enriched with 10% v/v of Silantes-rich growth media to improve cell growth (see table A.10). Stepwise adaption was achieved by growing pre-cultures with increasing percentage of D_2O from 20%, 33 %, 50 %, 70 % to 99.9 %. 2 ml cultures were grown for 8 to 16 hours and 200 μl cell suspension were centrifuged for 2 min at 3000 x *g* before being used for inoculation in the next step. After cells adapted to 99 % D_2O, they were plated on a deuterated agar plate with the appropriate antibiotic. A single colony was picked for a 5 ml culture to yield material for a glycerol stock as well as for a QIAgen Miniprep plasmid isolation.

2.3.3 Purification of the GST-tagged 15.5K protein

The cell pellet of a one-liter expression was resuspended on ice with 25 ml of 1 x PBS (see table A.11) and incubated with 1 μg/ml lysozyme, 20 U/ml RQI DNase, 10 mM DTT and 0.6 % Triton X-100 for 30 min at 4°C on a rocking platform. Cells were sonicated and the lysate was centrifuged at 30000 x g at 4°C for 1 h. The supernatant was purified using a 5 ml GSTrap FF column on an ÄKTA Prime system at 4°C. The ÄKTA system, the column and the whole tubing was cleaned prior and after use with 6 M guanidine hydrochloride to remove possible RNase contaminations. After loading the supernatant, the column was washed with 5 to 100 CV of 1 x PBS and the GST-tagged fusion protein was eluted with 20 mM Tris/HCl pH 8.0, 150 mM NaCl, containing 10 mM reduced glutathione. The peak fractions were combined and concentrated to 1 mg/ml. The GST-tag was cleaved first for 5 to 6 hours at room temperature and then at 4°C over night using 8 units thrombin protease per mg GST-15.5K. The 15.5K protein was then purified from the GST-tag and thrombin using gel filtration chromatography. Therefore a Superdex 200 26/60 column was equilibrated with 20 mM Hepes pH 7.6, 120 mM NaCl and 5 mM DTT and the run was performed with a flow rate of 2 ml/ min.

2.3.4 Purification of the His-tagged 15.5K protein

The cell pellet of a one-liter expression was used. 10 ml of lysis buffer (see table A.8) was used for resuspending two gram of cells on ice. The suspension was incubated for 30 min at room temperature on a rocking platform with 2 μg/ml lysozyme, 40 U/ml RQI DNase, 8.5 mM β-mercaptoethanol and one tablet of EDTA free proteinase inhibitor. Cells were disrupted on a high pressure homogenizer and the lysate was centrifuged at 30000 x g and 4°C for 40 min. The supernatant was filtrated over an Millex-HV (PVDF, 0.45 μm) before being loaded on a 5 ml His trapFF column using a ÄKTA purifier system at room temperature. The ÄKTA system, the column, and the whole tubing was cleaned prior and after use with 6 M guanidine hydrochloride to remove possible RNase contaminations. After binding the supernatant several washing steps followed; initially a wash step of 10 CV of lysis buffer and subsequently 2 to 5 CV of lysis buffer with 40 mM imidazole, then 10 to 20 CV of high salt washing buffer, then 2 to 5 CV of LiCl washing buffer and finally 2 to 5 CV of lysis buffer (see table A.8). The His-tagged 15.5K was eluted with 2 to 6 CV elution buffer. The column was cleaned with 6 M guanidine hydrochloride and equilibrated again with lysis buffer. The elution pool was buffer exchanged to lysis buffer with PD10 columns before the

His-15.5K protein was purified a second time on the His trapFF column. Affinity purification was repeated until the protein was free from RNase contamination, which was checked with the help of the RNase Alert Kit. When this assay showed no more RNase activity, another test was performed by incubating an RNA with the protein solution for 24 hours, 3 days, 1 week up to 4 weeks. The longer incubation tests were used to define, how long the RNA would be stable in NMR samples. If the protein/RNA mixture showed to not degrade the RNA after 3 days the purification was continued by exchanging the protein buffer to lysis buffer. A concentration of 0.5 to 1 mg/ ml was achieved by diluting or concentrating the protein solution. His-tag cleavage was performed with one unit enterokinase per mg of His-tagged protein, first for 5 to 6 hours at room temperature and then at 4°C over night. The His-tag removal was performed with Ni-NTA agarose and the purified 15.5K protein was kept at 4°C in lysis buffer for long term storage.

2.3.5 Purification of the His-tagged hPrp31 protein

The cell pellet of a one-liter expression was used and resuspended in lysis buffer (see table A.8) on ice (10 ml buffer for 2 g of cell pellet). This suspension was incubated for 30 min at room temperature on a rocking platform with 2 μg/ml lysozyme, 40 U/ml RQI DNase, 8.5 mM β-mercaptoethanol and one tablet of EDTA free proteinase inhibitor. Cells were disrupted on a high pressure homogenizer and lysate was centrifuged at 30000 x g and 4°C for 40 min. All further steps were performed at 4°C. The supernatant was loaded on 5 ml Ni-NTA agarose by gravity flow. After protein binding, several wash steps followed. First the beads were washed with 50 to 100 CV of washing buffer 1 (see table A.8), then with 10 to 20 CV of LiCl buffer (see table A.8), followed by 10 CV of washing buffer 2 (see table A.8) and 5 to 10 CV of washing buffer 3. The His-tagged fusion protein was then eluted from the agarose beads using 10 CV of elution buffer (see table A.8). His-hPrp31 was dialyzed against the NMR buffer (see table A.8) using a dialysis membrane with 10 kDa MWCO.

2.3.6 Determination of protein concentration

In order to quantify the amount of protein in solution, UV absorbance at 280 nm is usually applied. The absorption of radiation mediated by the proteins depends mainly on their tyrosine and tryptophan content. Thus the OD_{280nm} value varies considerably between different proteins and specific absorption values need to be determined depending on the amino acid sequence of the protein. Further

it should be noted that the extinction coefficient of nucleic acid is close to the 280 nm region and up to 10 times higher than that of proteins. This means that although this method is simple and the sample is recoverable, it is not suitable for proteins contaminated with RNA or protein-RNA complexes.
Another simple procedure for the determination of protein concentration in solution is the Bradford protein assay, which relies on the binding of the Coomassie Blue G-250 dye to the protein, mostly to the arginine and lysine residues. The quantity of protein is estimated by determining the amount of dye in its blue ionic form which is achieved by measuring the absorbance at 595 nm. In this work the Bradford assay was selected as method of choice since the proteins of interest possessed a high content of arginine and lysine residues and a small content of tyrosine and tryptophan.

2.3.7 Electrophoretic Mobility Shift Assay

The Electrophoretic Mobility Shift Assay (EMSA) also referred as gel retardation assay or gel shift assay, is a common technique to characterize protein-RNA interactions. EMSA is based on the principle that complexes of protein and RNA migrate through a non-denaturing polyacrylamide gel more slowly than free RNA fragments. The samples were incubated at 4°C for 30 to 60 min and loaded onto a 12 % native polyacrylamide gel (see table A.9). A total amount of at least 1 μg of RNA per lane was used to reach the detection limit of ethidium bromide staining. Electrophoresis was performed with a constant voltage of 150 V for 4 to 5 hours.

2.4 NMR sample preparation

Generally RNA samples were prepared by dialyzing the RNA against water to remove salt impurities. The RNA was completely dried in a vacuum concentrator or lyophile and dissolved in $\sim 300\ \mu$l of the desired buffer to reach a concentration of ≥ 0.3 mM. In order to form the correct tertiary structure an annealing step was applied by heating the RNA sample at 85°C for 2 min followed by cooling to room temperature within 30 minutes. When a solvent exchange from H_2O to D_2O was necessary, a similar procedure consisting of drying, dissolving in buffer and annealing of the sample was carried out.

2.4.1 Preparation of U4 snRNA samples

In this work an open loop variant of the U4 snRNA 5' stem loop was used, which is formed by annealing both synthesized RNA strands. Due to the fact that the RNA consists of two strands, several samples with different labeling schemes could be prepared. Initially two samples were done, where one strand was ^{13}C-, ^{15}N-labeled and the other strand was unlabeled and vice versa. An excess of unlabeled RNA was used to ensure that each molecule of the labeled strand, visible by NMR, was bound to one molecule of unlabeled strand.
For some NMR experiments a fully ^{13}C-, ^{15}N-labeled sample was needed. A RNA sample with equimolar ratio was prepared by mixing both labeled RNA strands and annealing them at 85°C. A size exclusion purification step was performed with a Superdex75 prep column and NMR buffer (see table A.9) as running buffer. The ÄKTA system, the column and all tubing was cleaned prior to use with 6M guanidine hydrochloride to remove possible RNase contaminations. The RNA complex fractions were pooled and concentrated by using vacuum concentration. Before salt concentration could exceed a concentration larger then 500 mM NaCl a dialyzing step against NMR buffer was done. Concentrating and dialyzing was performed until an end volume of 300 μl was reached. To exchange the solvent from H_2O to D_2O the sample was stepwise diluted with D_2O-buffer and concentrated again by centrifugation until the D_2O concentration reached 99.9% and the sample volume 300 μl.

U4 snRNA samples for SANS

The preparation of samples for SANS experiments is similar to the NMR sample preparation described in the previous paragraph. Differences to the NMR sample preparation were the amount of RNA used for size exclusion purification since a minimum concentration of 5 mg/ml was necessary for SANS studies in a sample volume of 200 μl in aqueous buffer.

U4 snRNA samples for residual dipolar coupling measurements

For RDC measurements, the isotropic sample of U4 snRNA was aligned with the filamentous phage Pf1, which was purchased in a suspension of 10 mM potassium phosphate buffer pH 7.6 and 0.05 % sodium azide. For solvent exchange, phages were washed twice with NMR buffer and centrifuged down using ultra-centrifugation at 400.000 x g for one hour. The phages were then added to the isotropic samples to yield an end concentration of 10 to 15 mg/ml. The sample was homogenized by gentle but extensive mixing and was then aligned under the magnetic field of the spectrometer prior to experiments.

2.4.2 Dimeric complex U4-15.5K

The U4 snRNA complex was prepared as described above for the SANS sample. This RNA complex was incubated for 30 min at 4°C in a 1:1 ratio with RNase-free ^{15}N-15.5K. The complex was purified by size exclusion and concentrated using vacuum concentration and dialysis until an appropriate injection volume was reached. Afterwards U4-15.5K fractions were pooled and the complex was concentrated again by alternating vacuum concentration and dialysis until an end volume of 300 μl had been reached. In two dialyzing steps, first against buffer in 99% D_2O and second against buffer in 99.96% D_2O, an exchange of solvent from H_2O to D_2O was reached.

2.4.3 Lariat-forming ribozyme

Due to the lack of regular secondary structure information, its size and dynamic behavior, the lariat-forming ribozyme (see figure 2.5) represents a challenging RNA construct for NMR spectroscopic investigation. A typical approach used to characterize this kind of structure involves uniform-labeling of a single nucleotide type with ^{13}C and ^{15}N, while the remaining nucleotides are unlabeled. This enables the collection of isotope-filtered and -edited NMR spectra to accomplish the resonance assignment. In addition, the use of partially deuterated NTPs diminishes the spectral overlap and improves the relaxation properties of the molecule.

Therefore both RNA samples containing deuterated (^{2}H- H3′, H4′, H5′, H5″) or ^{13}C, ^{15}N- labeled nucleotides, were prepared by *in vitro* transcription (see table 2.2b,c).
RNA synthesis and HH-cleavage was performed as described in chapter 2.1. Subsequent magnesium ions were removed by dialyzing the RNA constructs against a high salt buffer containing 1M NaCl and 20 mM phosphate buffer (pH 6.6) for at least 24 hours and subsequently against the NMR buffer (see table A.9).
Due to the extensive resonance overlaps in the 3′ terminal region of the ribozyme

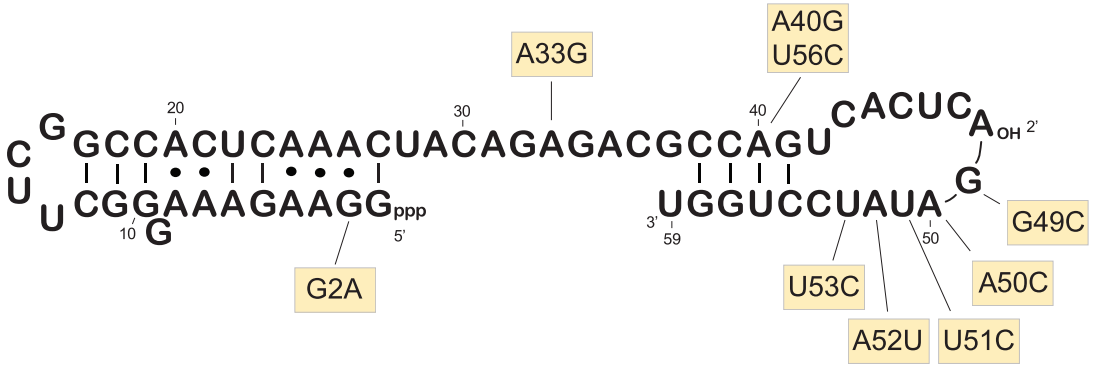


Figure 2.5: *Sequence of the lariat-forming ribozyme* | RNA constructs of selected mutation sites (yellow boxes) were synthesized by in vitro transcription to facilitate the resonance assignment of these nucleotides.

(nucleotides 49-53), additional NMR experiments were performed using RNA constructs that facilitated assignment of these nucleotides (see table 2.2d). Mutations were selected by using results of the mutational studies investigating the effect of mutations on lariat formation (see chapter 2.2). If a certain mutation still enabled the formation of the lariat form of the ribozyme it can be assumed that the tertiary structure remains the same compared to the original form. Thus, NMR spectroscopy investigations are reasonable on RNA constructs with such mutations. The NMR samples analyzed within the scope of this PhD thesis are listed in table 2.2d and depicted in figure 2.5.
In order to further study the helical region in the N-terminal part of the RNA molecule (figure 2.5), a truncated version of (residue 1-29) of the ribozyme was constructed. The helical region was confirmed through mutational studies (see chapter 2.2) and an unlabeled version of this truncated version was used to perform SANS experiments.
Samples for analyzing the lariat form of the ribozyme (see table 2.2f) were prepared by the addition of magnesium ions to linear RNA constructs followed by over night incubation at 30°C and dialysis to remove the ions.

Table 2.2: NMR samples of lariat-forming ribozyme synthesized by *in vitro* transcription and processed by hammerhead cleavage

Name	Ribozyme form	Isotope label	Concentration of NMR sample [mM]	Mutation
(a) unlabeled sample				
spliclin	linear	-	0.3	-
(b) ^{13}C, ^{15}N-labeled samples				
spliclin-G	linear	^{13}C, ^{15}N- G	0.3 & 0.3	-
spliclin-C	linear	^{13}C, ^{15}N- C	0.5	-
spliclin-A	linear	^{13}C, ^{15}N- A	0.4	-
spliclin-U	linear	^{13}C, ^{15}N- U	0.3	-
spliclin-AU	linear	^{13}C, ^{15}N- A, U	0.5	-
spliclin-full	linear	^{13}C, ^{15}N- A, G, C, U	0.67 & 0.35	-
(c) Deuterated samples				
spliclin-deut	linear	^{2}H- A,G,C,U	0.2	-
spliclin-C-deut-AU	linear	^{13}C,^{15}N- C & ^{2}H- A,U	0.1	-
(d) Ribozyme constructs with mutations				
spliclinG_2A*	linear	^{13}C, ^{15}N- G	0.9	G2A
spliclinA_33G*	linear	^{13}C, ^{15}N- A	1.2	A33G
spliclinA_40G-56C*	linear	^{13}C, ^{15}N- A	0.4	A40G & U56C
spliclinG_49C*	linear	^{13}C, ^{15}N- G	1.0	G49C
spliclinA_50C*	linear	^{13}C, ^{15}N- A	0.7	A50C
spliclinU_51C*	linear	^{13}C, ^{15}N- U	1.0	U51C
spliclinA_52C*	linear	^{13}C, ^{15}N- A	0.3	A52C
spliclinU_53C*	linear	^{13}C, ^{15}N- U	0.9	U53C
(e) Truncated RNA construct for SANS measurements				
spliclin29*	linear	-	0.2, 0.5, 0.9	Δ30-59
spliclin29-full*	linear	^{13}C, ^{15}N- A, G, C, U	1.0	Δ30-59
(f) Lariat form				
spliclar-G	lariat	^{13}C, ^{15}N- G		-
spliclar-C	lariat	^{13}C, ^{15}N- C		-
spliclarA_40G-56C*	lariat	^{13}C, ^{15}N- A	0.2	A40G & U56C

* RNA constructs marked by an asterisk were designed without an overhang for hammerhead cleavage.

2.5 NMR methods

The following section describes RNA experiments with isotope-labeled ribonucleotides (NTPs) enabling the assignment of NMR resonances and collection of RNA structure restraints. A table listing the NMR parameters set for various RNA experiments is given in the Appendix section (see chapter A.3).

2.5.1 Spectral assignment of isotope-labeled RNA

In NMR spectroscopy a high-frequent magnetic field B_0 is applied to probes containing magnetically active nuclei (with odd atomic numbers or odd mass, e.g. ^{1}H, ^{13}C). Caused by the magnetic field nuclei start spinning. The observed resonance frequencies of individual nuclei depend on the externally applied magnetic field but also on the small internal magnetic field induced by the orbiting electrons of the spinning nuclei. The superposition of external and the induced internal magnetic fields is measured as chemical shift of the observed nuclei. Depending on the local electronic environments of magnetically active nuclei, different chemical shifts can be observed for individual atoms of a molecule. The assignment of chemical shifts to each atom is a crucial step for structure elucidation of molecules by NMR spectroscopy.
Figure 2.6 shows the chemical shifts of ^{1}H, ^{13}C-nuclei for one of the studied RNA molecules. In order to distinguish individual protons in the NMR spectrum, a second dimension is introduced showing the ^{13}C-frequency resonances of carbon nuclei directly attached to the protons. The figure illustrates that resonance peaks of nuclei situated in similar local electronic environments (e.g. of sugar moieties in individual RNA nucleotides) cluster within the spectrum. For molecular systems of increasing size more and more resonance peak overlaps are observed within the distinct spectral regions. Even for the small 5kDa RNA construct shown in figure 2.6, spectral regions of the ribose protons H2', H3', H4' and H5'/H5'' (see Figure 2.7) show overlapping peaks hampering assignment of these protons. However, spectral regions for ribose protons H1' and the base protons H6 (purines) and H8 (pyrimidines) (see Figure 2.7) reveal a well dispersed pattern of resonance peaks for the individual nucleotides. Therefore, the peak pattern of the spectral regions H1'C1' and H6C6/H8C8 can be used as fingerprint of the RNA molecule, since they are unique for a given RNA sequence.

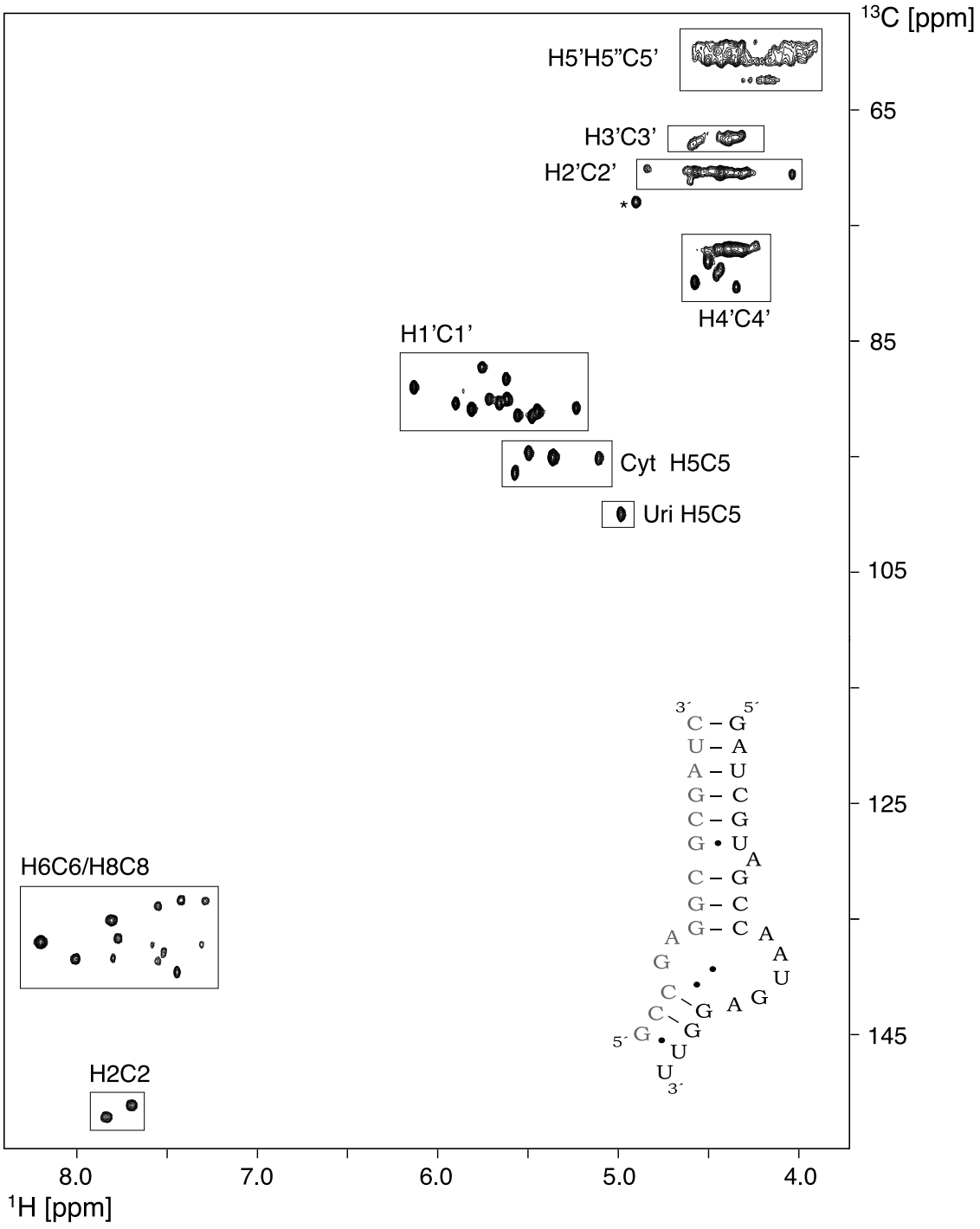
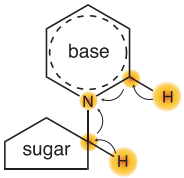
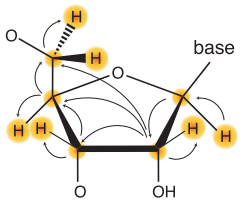
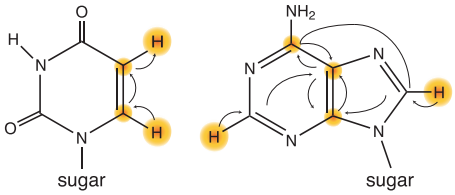
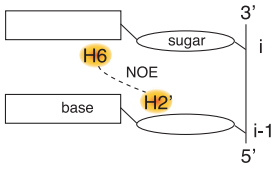
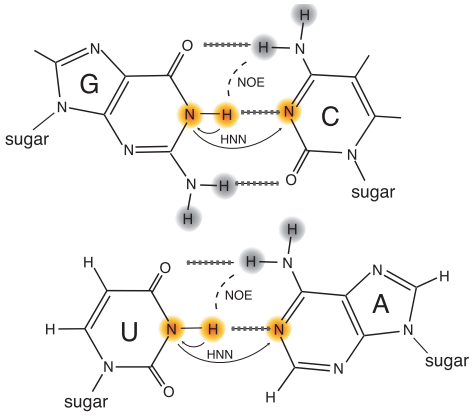


Figure 2.6: *HC-plane of the 14-mer strand of U4 snRNA* | Protons with similar local electronic environments cluster in distinct spectral regions (boxes). The peak annotated by an asterix belongs to a H3'C3' correlation. (The 2D plane was taken from a ^{13}C-edited, ^{13}C-filtered NOESY spectrum. Since the spectrum was recorded with a spectral width of 54 ppms in the ^{13}C-dimension, it needed to be refolded to illustrate the full resonances dispersion. For further simplification the water signal at 4.7 ppm was removed from the spectrum.)

Table 2.3: Schematic drawing indicating the magnetization transfer in standard double and triple resonance experiments required for RNA assignment. The nuclei for which correlations can be observed are indicated by yellow circles. Magnetization transferred by scalar couplings is indicated with arrows and the one transferred by NOE (see chapter 2.5.6) with dashed lines. Py - pyrimidines. Pu - purines.

Experiment	Observed Correlation	Magnetization transfer
*Base to sugar correlation**		
$H_bC_bN_b$	H6-C6-N1 (Py)	
	H8-C8-N9 (Pu)	
$H_sC_sN_b$	H1′-C1′-N1/N9 (Py/Pu)	
*Sugar assignment**		
COSY-TOCSY	H1′C1′, H2′C2′ ... H5″C5′	
	C1′C2′, C2′C3′, C3′C4′, C4′C5′	
	H1′C2′, H2′C3′, H3′C4′, H4′C5′	
*Base assignment**		
H5H6-COSY	H5-C5-C6-H6 (Py)	
H2H8-COSY	H2-C2-C5-C4/C6 (A)	
	H8-C8-C6-C5 (Pu)	
	H8-C8-C4-C5 (Pu)	
Sequential (seq) assignment in A-form RNA		
NOESY	H6/H8(i)-H2′(i-1) (strong)	
	H6/H8(i)-H3′(i-1) (medium)	
	H6/H8(i)-H1′(i-1) (weak)	
Verification of Watson-Crick base-pairs and solvent exchangeable protons		
HNN-COSY	H1(G)-N1(G)-N4(C)	
	H3(U)-N3(U)-N6(A)	
NOESY in water	NH-NH	
	seq: weak	
	intra-bp: strong (only GU)	
	$NH\text{-}NH_2$	
	seq: weak	
	intra-bp: strong → picture	
	$NH_2\text{-}NH_2$	
	seq: weak	

* Assignment of these spectra is verified with data from ^{13}C-edited NOESY experiments.

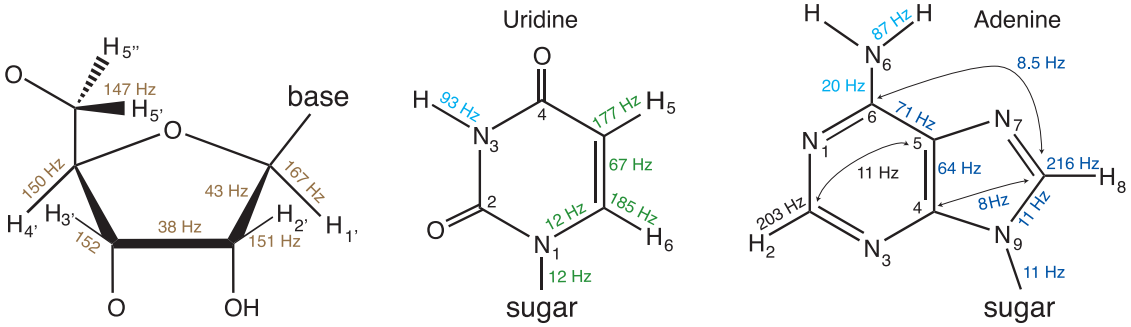


Figure 2.7: *J-coupling constants between ^{1}H, ^{13}C and ^{15}N nuclei in nucleotides (in Hz) based on Ref. [159]* | The given J-couplings are mean values over all nucleotide sugars (brown), all pyrimidines (green), all purines (blue), adenines (black) and all solvent-exchangeable protons (light blue). The J-coupling over one bond (^{1}J) have the largest values for HC bonds, then for HN and CC bonds and the smallest values for CN bonds. J-couplings over several bonds (^{2}J or ^{3}J) are indicated by an arrow.

To assign the chemical shifts of an RNA molecule several spectra need to be recorded (see table 2.3). A common feature of most of the two- or three-dimensional experiments is that the magnetization transfer rely on the use of scalar couplings (see box) between ^{1}H-, ^{13}C- and ^{15}N-nuclei of the RNA (see figure 2.7). A detailed description of these experiments can be found in the following chapters. To assign chemical shifts of an RNA molecule several spectra need to be recorded (see table 2.3). A common feature of most of the two- or three-dimensional

experiments is that the magnetization transfer rely on the use of scalar couplings (see box) between ^{1}H, ^{13}C and ^{15}N nuclei of the RNA (see figure 2.7). A detailed description of these experiments can be found in the following chapters.

Scalar coupling

The scalar coupling arises from the polarization transfer between two magnetically active nuclear spins mediated by the electrons participating in the bond connecting the nuclei. The effect is a split of the energy of the spins into two or more resonance lines depending on the number of nuclei around effecting the observed nucleus. Since scalar coupling is mediated by chemical bonds, increasing the number of bonds between the coupled spins, decreases the coupling, which is usually not measurable anymore over more than three covalent bonds. The scalar coupling is also referred to as spin-spin coupling or J-coupling.

continued on next page

The splitting of resonances caused by scalar coupling is usually suppressed by decoupling the nuclei effecting the observed nuclei during the delay the chemical shift of the observed nucleus evolves. However, in cases scalar coupling is wanted and used to transfer magnetization, delays are build into the pulse programs in which the scalar coupling can evolve. The step of transferring magnetization by exploiting the scalar coupling is named INEPT-step (Insensitive Nucleus Enhanced by Polarization Transfer) [91].

2.5.2 Identification of base to sugar spin system

To identify the resonances of an RNA (see figure 2.6), first a ^{1}H-^{13}C hetero-nuclear Single Quantum Coherence (HSQC) spectrum [4] is recorded. The HSQC experiment is the basic hetero-nuclear two-dimensional NMR method where the coupling between a ^{13}C (or ^{15}N) and the proton directly attached to it is used to transfer magnetization. In this experiment the ^{1}J coupling is used which is large enough (100 - 200 Hz) to enable a fast transfer of coherence without much magnetization loss arising from relaxation, even for large molecules with short relaxation times [100]. After this INEPT transfer step, the chemical shift of ^{13}C evolves in t1 (evolution time) and finally the magnetization of ^{13}C is transferred back to ^{1}H for the evolution of its chemical shift in t2. In the resulting spectrum each single peak corresponds to the chemical shift of one proton of the RNA molecule in the first dimension and the chemical shift of the directly attached carbon nuclei is plotted on the second dimension.

2.5.3 Correlation of base to sugar spin system

Base and sugar moieties of RNA are connected via the C-N glycosidic bond. Hence triple resonance experiments are used to enable the correlation of the sugar proton H1′ to its base resonance H6 for pyrimidines (Py) or to H8 for purines (Pu) via the connecting nitrogen N1 (Py)/ N9 (Pu) resonances (see figure 2.8).
Unfortunately the triple resonance experiments rely on extended evolution periods which arise due to the relatively small coupling constants employed for the magnetization transfer. This results in an loss of magnetization during the evolution time by fast spin–spin relaxation and hence to low sensitivity of these experiments especially for large molecules [159].
Several improvements for this type of experiments have been introduced by the group of Sklenář [33][128] to correlate H1′ to H6/H8 even for large RNA

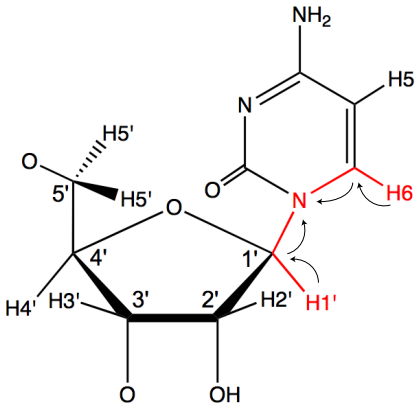


Figure 2.8: *Exemplary magnetization transfer ways for H_sCN_b and H_bCN_b experiments for a cytidine molecule.*

molecules. In the H_sCNCH_b experiment [33] the magnetization is transferred directly from the sugar proton H1′ (H_s) to the respective base proton (H_b) by using the respective coupling constants (transfer way: H1′ → C1′ → N9/1 → C8/6 → H8/6). The correlation of the two protons can also be achieved by recording two spectra, a H_sCN_b [33] and a H_bCN_b [88] (see figure 2.8).

2.5.4 Assignment of the sugar spin system

Compared to DNA, the identification of sugar protons in RNA is difficult due to the crowding of the sugar region, where all sugar protons, with the exception of H1′ resonate in the range of 4 to 5 ppm (see figure 2.6). Nevertheless, with the help of multidimensional NMR experiments complete assignment is possible for RNA up to 60 nucleotides. The larger the system the more difficult it is to distinguish the different sugar protons and to assign all six sugar proton resonances.
The assignment of sugar protons is mainly accomplished by the use of the 3D-HCCH-COSY-TOCSY experiment [54][121]. In this experiment, the ^{1}H and ^{13}C atoms of the ribose ring are correlated by the combination of a COSY (COrrelation SpectroscopY) and a TOCSY (TOtal Correlation SpectroscopY) step. The magnetization is first transferred from a sugar proton to its attached carbon atom by an INEPT step similar as described above for the HSQC. In the following COSY step the magnetization is transferred to its neighboring carbon. This transfer delivers the information for the first 2D plane. The TOCSY step further transfers the magnetization of the COSY-derived ^{1}H, ^{13}C- correlation in the entire sugar spin system (see figure 2.9) which is depicted within the second 2D plane of the three-dimensional experiment. This enables the correlation of the known H1′ resonances to the neighbor H2′ nuclei, then to the H3′ and further to the H4′ up to the H5′ and H5″ of the ribose nuclei.

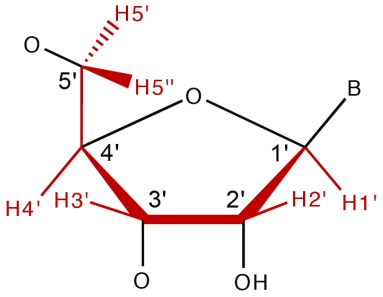


Figure 2.9: *Magnetization transfer in 3D-HCCH-COSY-TOCSY experiment*

2.5.5 Assignment of base spin system

The assignment of the H5-H6 correlation in pyrimidines is achieved by a HCCH-COSY experiment (see figure 2.10A). In this work a 2D-(H)C(C)H-COSY was used [38], where the magnetization is transferred in three steps, first from the proton to its directly attached ^{13}C nucleus via $^{1}J_{CH}$ coupling, then from the carbon to its neighbors via $^{1}J_{CC}$ coupling, and finally from the ^{13}C nucleus back to its attached proton via $^{1}J_{CH}$ [38].
The H2-H8 correlation for adenosines can be determined by the use of HCCH-TOCSY [72][81] or the more sensitive TROSY-relayed HCCH-COSY [126]. In the latter method the magnetization is transferred simultaneously from H2 and H8 to the three aromatic carbons C4, C5 and C6 (see figure 2.10B). Thus, when the H8 resonance is known, the assignment of the three measured carbons becomes possible and their correlation to the H2 proton can be determined. A drawback of the method is the poor resolution arising from the small chemical shift ranges of the C4, C5 and C6 nuclei ranging within 1 to 3 ppm.

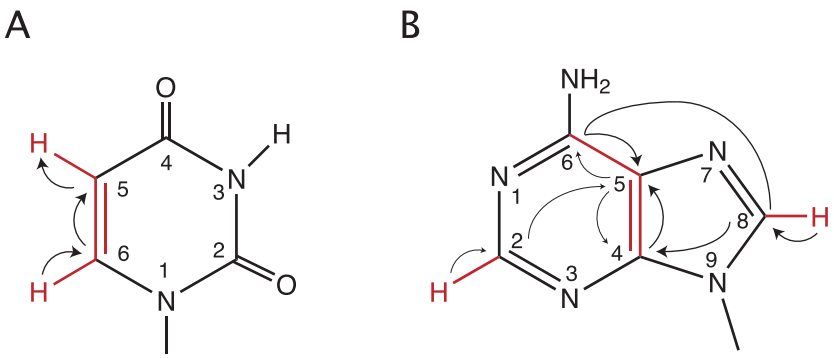


Figure 2.10: *Magnetization transfer in HCCH-COSY experiments* | (A) for correlating pyrimidine H5/H6 resonances (B) for correlating adenosine H2/H8 resonances

2.5.6 Sequential assignment

The assignment of measured resonances to a certain residue in the oligonucleotide sequence is accomplished by the use of through-space information derived from NOESY (Nuclear Overhauser Effect SpectroscopY) experiments. By NOESY (see box) all protons within a distance of 5 Å can be correlated. Because of spectral overlap in the sugar region of RNA oligonucleotides, three-dimensional NOESY experiments are required for a complete assignment.

Dipolar coupling and nuclear overhauser effect
The Nuclear Overhauser Effect (NOE) is caused by dipolar cross-relaxation of neighboring spins causing a resonance line intensity change of the observed spin. This is based on dipolar coupling between through space neighbored nuclear spins. Additionally to the applied magnetic field B_0, the two magnetic dipoles of the nuclei mutually induce a local magnetic field B_{loc}. The magnetic dipole coupling decreases with the third power of the distance between the two dipoles. But since this coupling is anisotropic, the brownian molecular movement averages the dipole interaction in solution NMR measurements.
However, if such a system of coupled dipoles is perturbed, it relaxes over the dipolar interaction. This changes the populations of both nuclei spins leading to a change of polarization and thus to an intensity change of the resonance line. Like the dipolar coupling the nuclear overhauser effect strongly depends on the distance between the interacting nuclear spins ($NOE \propto r^{-6}$). Subsequently the derived data is used for structure determination since they deliver atomic distances information.

The sequential assignment can be obtained by well established NOE correlations. In a well defined A-form helix, with 3'endo sugar conformation, the base proton H6/H8 has a strong NOE signal with the H2' of the preceding nucleotide, a medium NOE with H3' and weak with the H1' proton. Furthermore a weak NOE signal is seen between H5 and the H6 resonance of the preceding nucleotide. For sugars in C2'-endo conformation the NOE correlations differ. Instead of a strong H8-H2'(i-1) NOE, a H8-H3'(i-1) NOE can be observed.
A typical approach used to characterize large RNA structures involves uniform labeling of one of the components with ^{15}N and ^{13}C, while the other component remains unlabeled. This enables the collection of isotope-filtered and -edited NMR experiments which makes it possible to select magnetization. In this work

the 3D ^{13}C-edited ^{12}C-filtered NOESY and the 3D ^{13}C-edited ^{13}C-filtered NOESY experiment [170] were applied on RNA molecules where only one nucleotide was labeled while the other three nucleotides remained unlabeled. If magnetization was transferred between protons of the labeled component to proximal protons attached to unlabeled hetero-atoms, inter-nucleotide NOEs were visualized. While the magnetization transfer between protons both attached to labeled hetero-atoms can deliver both intra-nucleotide and inter-nucleotide NOEs depending on the arrangement of nucleotides. In the case two consecutive nucleotides are labeled (e.g. AA or CC) both NOE transfers are possible at a time while the case a labeled nucleotide is preceded and followed by an unlabeled nucleotide allows either intra-nucleotide (^{13}C-edited ^{13}C-filtered NOESY) or inter-nucleotide (3D ^{13}C-edited ^{12}C-filtered NOESY) NOEs depending on the selected NMR experiment.

2.5.7 Assignment of exchangeable protons

Signals of imino and amino resonances are only observed when the protons are protected from fast exchange with water. This is the case for protons involved in base pairing [31]. Exchangeable protons provide information regarding base pairing and give rise to the majority of cross-strand NOE distance constraints. Even in a 1D experiment it is possible to derive the number of base-pairs, at least for small RNA oligonucleotides, by counting the number of imino resonances between 10 and 15 ppm [35].
The chemical shifts of imino proton resonances in RNA are strongly dependent on the chemical environment. In Watson Crick base-pairs the imino proton of guanine bases resonate between 12 and 13.5 ppm and the imino protons of uracil bases resonate between 13 and 15 ppm. A non-canonical base-pair yields to a clear shift of the resonance of the involved non-exchangeable proton. For example a GU wobble base-pair leads to upfield shifted resonances between 10 to 12 ppm for both the U and the G imino protons.
The chemical shifts of amino protons resonated between 7.5 to 9 ppm and thus are less dispersed. Cytosine amino resonances are the most easily observable amino resonances. Guanosines and adenosines are more difficult to observe since their amino group rotate in a time scale, that leads to an broadening of the resonance lines. However, by lowering the temperature the probability to observe these signals increase [151].
The assignment of imino protons is mainly accomplished by the use of ^{15}N-HSQC, 2D-NOESY and HNN-COSY experiments. The AU base-pairs are identified by a

strong NOE cross-peak between the H2 proton of adenine and the uracil H3 imino proton. The GC base-pairs show strong NOE contacts from the guanine H1 imino proton to the the amino protons of the cytosine.

2.5.8 Restraints for RNA structure determination

Structure determination of biomolecules by NMR is mainly dependent on the collection of interproton distance constraints derived from the analysis of NOESY spectra (see chapter 2.5.6). NOEs are the primary source of structure information since they reflect direct distances of neighboring atoms in a molecule and thus give a dense network of distance restraints.
Furthermore experimental data of the six backbone torsion angles and the glycosidic angle are desirable, since they describe the relative position of sugar to base in the RNA conformation. Due to the low density of protons in the phosphodiester backbone, the dihedral angles α, β, γ, ϵ and ζ are generally not well defined by interproton distance constraints [1][151]. For this reason dihedral angle constraint data can improve the precision and local accuracy of the RNA backbone conformation [150]. Torsion angles can be mainly determined by measuring scalar couplings (see chapter2.5.1). The magnitude of scalar couplings can mainly be related to certain ranges of backbone angles by the use of Karplus equations:

$$J(\theta) = Acos^2(\theta) + Bcos^2(\theta) + C \tag{2.3}$$

where J is the ^{3}J coupling constant, θ is the dihedral angle, and A, B, and C are empirical derived parameters which are optimized for various types of couplings and residues based on the best fit between the measured ^{3}J values and the corresponding value calculated with equation 2.3 for known RNA structures [1][16][108][111]. The observation of homonuclear scalar couplings is performed by measuring E.COSY (Exclusive COrrelation SpectroscopY) patterns [167], where the couplings are measured from the displacement of multiplet resonances separated by large single-bond ^{1}H-^{13}C couplings.

A third type of restraint for structure determination are residual dipolar couplings (RDCs, see box) [140][139]. In RNA the alignment is normally achieved by adding bacteriophages Pf1 particles to the sample [49][43]. A greater impact in studies of nucleic acids than in proteins is assumed for the long-range structural information provided by RDCs, because of the extended secondary structure and the limited number of NOE constraints in DNA and RNA oligomers, due to the smaller proton density in nucleic acids [67].

Residual dipolar coupling
RDCs arise from incomplete averaging of the dipolar interaction (see chapter 2.5.6) in partially aligned molecules. RDCs are used to compute order tensors to describe the average alignment of a part of a molecule to the applied magnetic field [105]. This information help defining the relative orientation of molecular fragments (e.g. RNA helices) and thus deliver information on the relative position of molecular domains.

3 Results: Structural studies of the U4 snRNA 5′ stem loop

3.1 RNA synthesis and hammerhead cleavage

Transcription trials (see chapter 2.1.2) were performed in our laboratory to synthesize the U4 snRNA 5′ stem loop (U4 SL) based on the known native sequence (see chapter 1.2.2) and referred to as U4-Kink-turn 1 (U4-Kt1). However, only low yields of the 33 nucleotide long RNA stem loop could be obtained, presumably due to the uridine-rich loop region [46].
Studies performed by Schultz *et al.* on different penta-loop mutants of U4 SL revealed that open loop constructs were competent for binding of both, 15.5K and hPrp31 protein [120]. Therefore an open loop variant lacking some of the uridine residues was designed for *in vitro* transcription. As transcription start site for T7 RNA Polymerase [89] a guanine was added at the 5′ end of the open loop and a GC pair at the termini of the C-stem. The newly designed RNA construct is referred to as U4 Kink-turn 2 (U4-Kt2) in this work and is depicted in figure 3.1. U4-Kt2 comprises all conserved residues to form the kink-turn motif (see chapter 1.2.2), the base-pairs of the canonical stem (C-stem) and non-canonical stem (NC-stem) as well as the internal loop residues. The protein binding ability of this open loop construct was verified by gel shift assays.

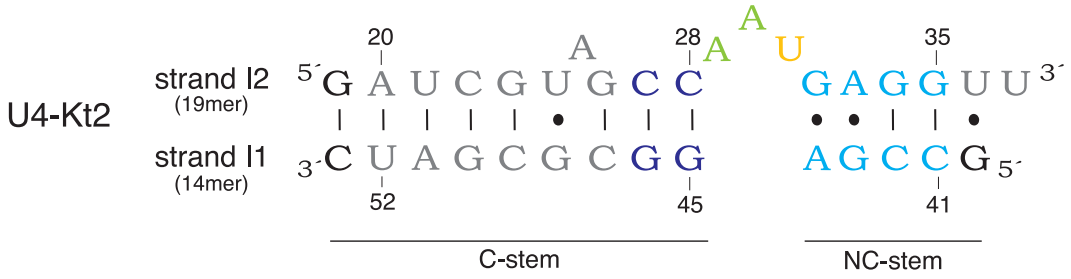


Figure 3.1: *Secondary structure of the U4 construct used in this work* | Consensus residues of the k-turn motif are color coded, residues in the C-stem and the NC-stem are labeled in blue or light blue, respectively. The bulged-out nucleotide and further interloop nucleotides are marked by orange and green coloring, respectively. Residues differing from the native sequence are shown in black.

For the initial transcription trials of the open loop construct U4-Kt2, DNA templates encoding for the two U4 strands I1 and I2 were cloned into a *pUC19* plasmid and then tested for transcription, which lead to RNA products of differing lengths. To avoid this 3′ heterogeneity of RNA transcripts, a hammerhead (HH) overhang was introduced into the template DNA (see chapter 2.1.7) by cloning into *pUC19*. However, RNA transcription from the newly designed plasmid DNA resulted in insufficient RNA amounts, while transcription tested from the original DNA templates (prior to cloning) resulted in satisfactory RNA yields. As a consequence, all following transcription reactions for U4-Kt2 were performed with purchased DNA oligomers as DNA templates.
Transcription reactions for the I1 strand resulted in a 32-mer which could directly be used for HH cleavage reactions, yielding the RNA of interest with 14 nucleotides length and a 18-mer long cleaved rest. The DNA template used in the transcription of the U4-I2 RNA (19mer) needed a further optimization, since the HH cleavage of the transcribed RNA yielded two RNA molecules of similar length, the desired 19-mer U4-Kt2 and a 18-mer RNA, which could not be separated from each other by gel electrophoresis. To overcome this problem, which arised from the fact that the DNA oligomer was initially designed for cloning rather than for *in vitro* transcription, the cloning overhang was removed from the DNA sequence. Unfortunately, this trial of shortening the DNA template resulted again in poor transcription yields and, thus, the template was elongated to yield a 48-mer, that could be further processed to the 19-mer U4-Kt2 strand I2. Both strands, I1 and I2, were then further purified (see chapter 2.1.7) and mixed (see chapter 2.4.1) to form the U4-Kt2 (see figure 3.1).
Parameters used for large scale transcription are summarized in table 3.1. The optimization of the hammerhead cleavage reaction for both RNA strands was optimized, yielded a optimum enzyme-to-substrate ratio of 1.5 to 1, over night at 37°C with 50 mM $MgCl_2$.

Table 3.1: Parameters used in large scale *in vitro* transcription reactions

	[NTP in mM]	[$MgCl_2$ in mM]	[template]	T7 RNAP (v/v)
I1	24 - 30	20 - 30	1.25 - 2.7 μM	1/10 - 1/5
I2	10 - 19.2	20 - 30	3.3 μM	1/5
HH-I1	20	30	25 μg/ml *	1/5
HH-I2	30	30	20 μg/ml *	1/5

* as template plasmid DNA was used (*pUC19* with HH-insert after *PstI* digestion)

3.2 Spectral assignments of the U4 snRNA 5′ stem loop construct

The prerequisite for the structural elucidation of the U4 5′ stem loop construct by NMR spectroscopy is the assignment of individual resonance peaks of the recorded NMR spectra to the respective nuclei within the RNA molecule which is described in detail in the following sections.

3.2.1 Identification of base and sugar protons

Characteristic resonances for RNA molecules are the correlations of H1′C1′ and H6C6 (Py)/ H8C8 (Pu), which are recorded in HSQC spectra (see chapter 2.5.2). For both strands of the studied U4-Kt2 construct, resonance peaks were found to be well dispersed in the recorded spectra (see figure 3.2). In the ^{13}C-HSQC spectra of H1′C1′ a single overlay of C22 and G34 resonances for the 19-mer strand was observed. In the ^{13}C-HSQC of the H6C6/H8C8 region the resonances of G43 and G45 for the 14-mer RNA as well as the resonances of C28 and U36 of the 19-mer RNA could not be distinguished from one another.

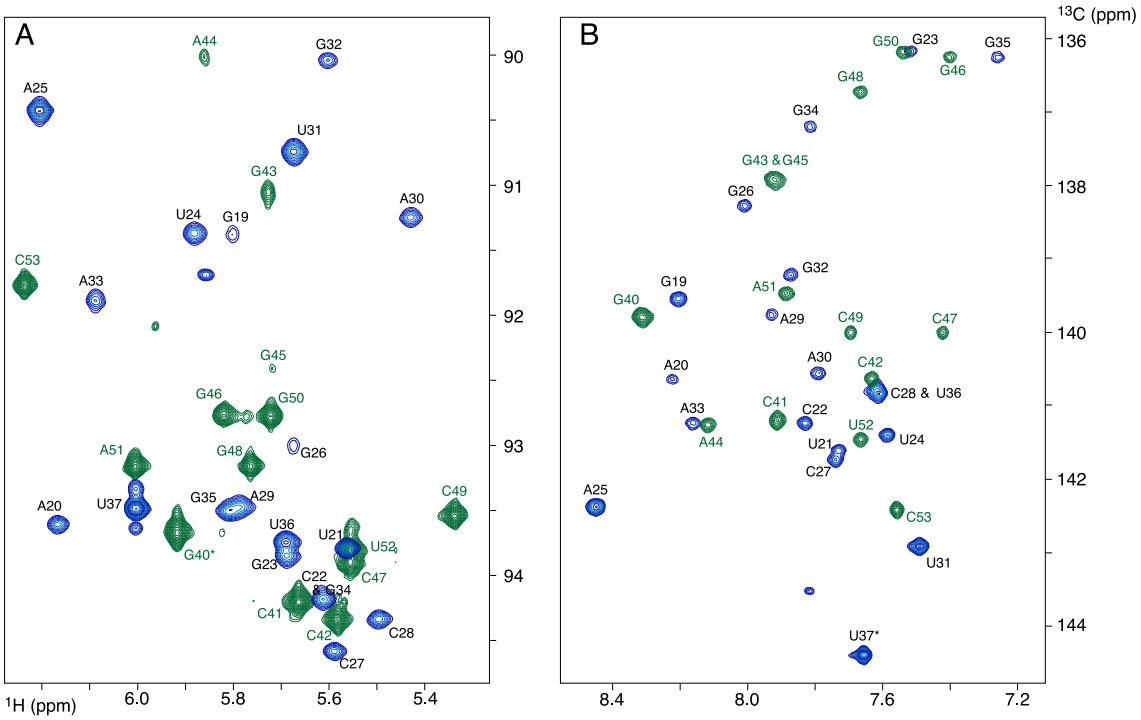


Figure 3.2: *H-C correlations of both strands of the U4-Kt2 construct* | (A) Peak pattern of H1′C1′ ribose resonances and (B) peak pattern of H6C6 (Py) / H8C8 (Pu) base resonances. Resonance peaks of the 19-mer and the 14-mer of U4-Kt2 are given in blue and green, respectively.

Resonance peaks of the remaining sugar protons H2′, H3′, H4′, H5′ and H5″ were identified by measuring a constant time ^{13}C-HSQC (spectra not shown, see table A.14 for used NMR parameters). Pyrimidine H5 protons can be distinguished in the chemical shift region of their C5 nuclei resonating in the range of 100 to 105 ppm for uridine and 95 to 100 ppm for cytidine. The chemical shifts of H5C5 can be detected in the ^{13}C HSQC together with the H1′C1′ correlations. The C2 resonances of adenosine resonate around 155 ppm and could be determined by further analyzing the ^{13}C-HSQC spectra of H6C6/H8C8 correlations.
Following the identification of chemical shifts for protons and their attached nuclei the chemical shifts were assigned to the respective nucleotide of the secondary structure sequence as described in the following sections.

3.2.2 Correlation of the base spin system to the sugar spin system

In isotope-labeled RNA (^{13}C, ^{15}N), the aromatic protons can be correlated to the ribose protons using double or triple resonance experiments. In this work two triple resonance HCN experiments have been carried out for correlating base spin systems to their respective sugar spin systems (see chapter 2.5.3). The H_sCN_b experiment was used to determine through-bond connectivities via H1′-C1′-N9(Pu)/N1(Py), whereas the H_bCN_b experiment was used for through-bond connectivities via H6-C6-N1(Py) or H8-C8-N9(Pu). Exemplarily, two-dimensional slices of both HCN experiments are shown in figure 3.3 for nucleotides A20, C28 and U36 of strand I2. The base protons H6/H8 could be correlated to the corresponding ribose H1′ protons via the common N1/N9 atom.
Aromatic protons can also be correlated to sugar proton resonances by measuring NOE interactions. This method is less reliable, as in general inter-nucleotide NOE interaction are not distinguishable from an intra-nucleotide cross-peak and especially when RNA strands form non-standard secondary structures. Nevertheless, the ^{13}C-edited, ^{13}C-filtered NOESY spectra provide excellent data to verify base to sugar correlations assigned by HCN experiments and to resolve correlations not measurable by HCN experiments due to overlapping resonance peaks or due to the lack of signals. By using HCN and NOESY experiments the correlation of all H1′ protons to their respective H6/ H8 protons could be accomplished for the studied U4-Kt2 construct.

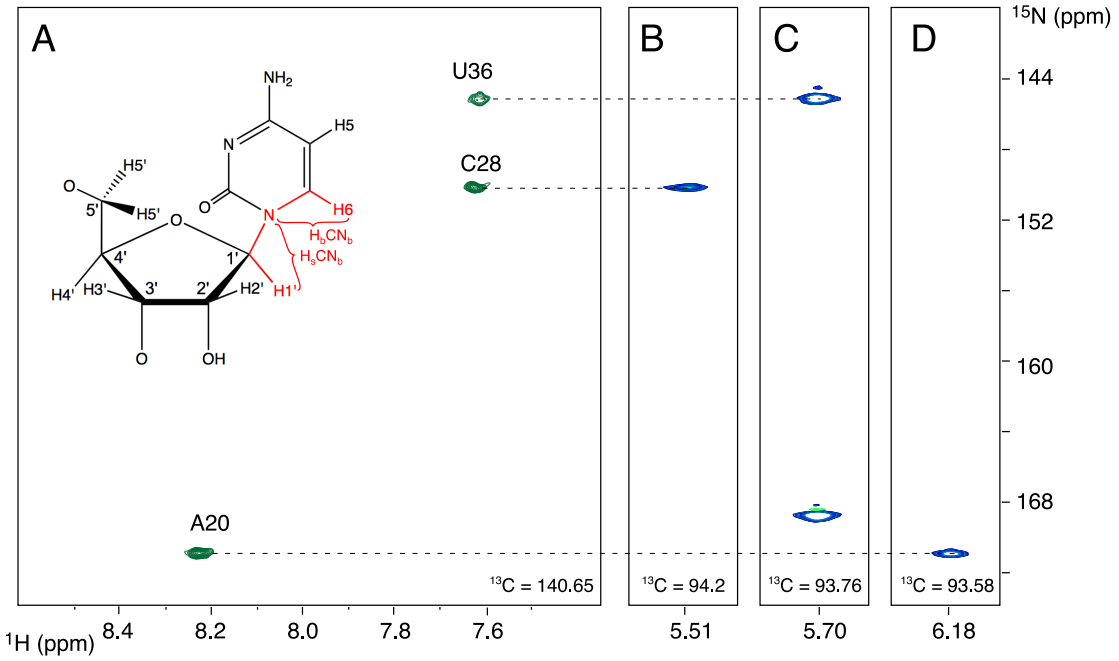


Figure 3.3: *2D slices of the 3D HCN spectra* | The base protons H6/H8 could be correlated to the corresponding ribose H1' protons via the common N1/N9 atom. (A) shows the HN-plane of A20, C28 and U36, measured through the H_bCN_b experiment. In (B), (C) and (D) planes of the H_sCN_b spectra are shown, namely the HN-plane of C28 (B), the HN-plane of U36 (C) and the HN-plane of A20 (D). The magnetization transfer of both experiments is marked red in the depicted nucleotide.

3.2.3 Assigning the sugar spin system

Individual resonances of the sugar spin system can be assigned from the 3D HCCH-COSY-TOCSY experiment as described in chapter 2.5.4 and figure 3.4. In the applied experiment COSY cross-peaks can be observed in the indirect proton dimension, where-as the TOCSY connectivities are found in the direct proton dimension (see figure 3.4B). This allows the correlation of the sugar spin system starting from the H1'C1' region of the spectrum (see figure 3.4C).
Again, assignment of individual ribose protons might be hindered by signal overlap. However, by using ^{13}C-edited ^{13}C-filtered NOESY spectra (data not shown) as an additional data source, assignments could be accomplished and validated for nearly all 198 sugar protons, except for H5' and H5'' of nucleotide A30 and H5'' of nucleotide G32.An HCCH-COSY-TOCSY experiment with longer TOCSY mixing times could have been used as another source of information for assigning ribose protons of the sugar ring. However this was not required, since HCCH-COSY-TOCSY and ^{13}C-edited ^{13}C-filtered NOESY spectra provided all necessary information for an almost complete sugar resonance assignment for both strands of the U4-Kt2 construct.

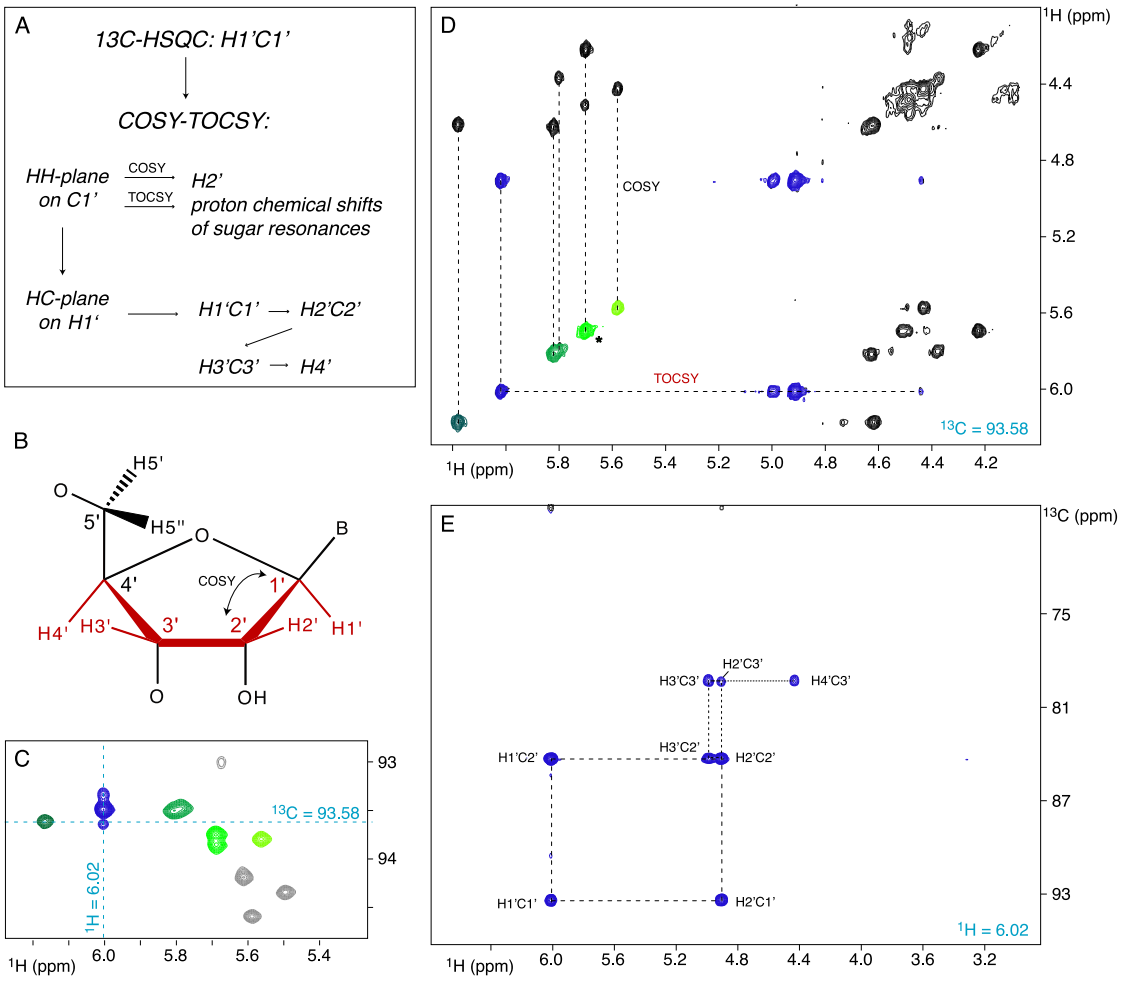


Figure 3.4: *Assignment of sugar resonances from the ^{13}C and ^{15}N-uniformly labeled 19-mer strand of the U4 snRNA stem loop using a 3D HCCH-COSY-TOCSY spectrum.* (A) Data collection scheme for 3D-HCCH-COSY-TOCSY. (B) Ribose moiety. The TOCSY transfer is marked in red and the COSY transfer from H1′ to H2′ is indicated by an arrow. (C) The spectrum of the ^{13}C-edited HSQC to visualize the H1′C1′ resonances, depicted in (D) and (E). (D) HH-slice of the HCCH-COSY-TOCSY spectrum at ^{13}C = 93.58 ppm. The slice shows the H1′ to H2′ correlation via the COSY transfer. The resonances marked with an asterisk could be assigned to two nucleotides due to the COSY transfer. The TOCSY transfer correlates further chemical shifts of the sugar protons to the respective resonance. (E) Exemplary assignment of sugar spin system U37 showing the HC-plane at H = 6.02 ppm, the H1′ chemical shift of U37.

3.2.4 Assigning the base spin system

Nucleotide bases possess only one or two non-exchangeable protons. The pyrimidines cytosine and uridine contain H5 and H6 protons, whereas the purines possess H2 (only Ade) and H8 (Gua, Ade) protons (see figure 1.15). The assignment of H6 and H8 protons is described in chapter 3.2.2. The assignment of H5 in pyrimidines was performed with HCCH-COSY experiments (see chapter 2.5.5 &

shown in figure 3.5A). However, in this constant-time experiment the time for chemical shift evolution of ^{13}C is limited to $t_1 < 1/(4 \times J_{CC})$, to enable maximum sensitivity. The value of the scalar ^{1}J-coupling between C5 and C6 atoms is 67 Hz (see figure 2.7 in method section), resulting in a maximum acquisition time of $\sim$ 7.4 ms do not permit the measurement of a large number of points, thus leading to limited spectral resolution and broad signals in the carbon dimension. (see figure 3.5A). The assignment of the H5 resonances of the 19-mer strand of the U4 snRNA is illustrated in the overlayed COSY/ ^{13}C-edited HSQC-spectra in figure 3.5A, whereas the assignment of the H5 resonances of the 14-mer strand is depicted in the ^{13}C-edited HSQC of figure 3.5B. The assignment of the H5 protons could be accomplished for both RNA strands and could be verified using ^{13}C-edited ^{13}C-filtered NOESY spectra, in which H6 and H5 distance of 2.5 Å results in an intense NOESY cross-peak.

The assignment of the non-exchangeable base protons H2 in adenosines were based on a 3D-TROSY-relayed HCCH-COSY experiment [126] (see chapter 2.5.5). Here, base carbons C4, C5, C6 were assigned, which can subsequently be correlated to H2 and H8, thus, leading to assignment of the base proton H2 by their indirect correlation to the H8s. For the 19-mer strand of the U4 snRNA the H2 proton of A20, A29 and A30 could be assigned, while the correlation of carbons C4, C5, C6 to the respective H2 nuclei could not be accomplished for A25 and A33 due to signal overlap (spectra not shown). The H2 resonances for adenines of the 14-mer strand could already be assigned using the ^{13}C-edited ^{13}C-filtered

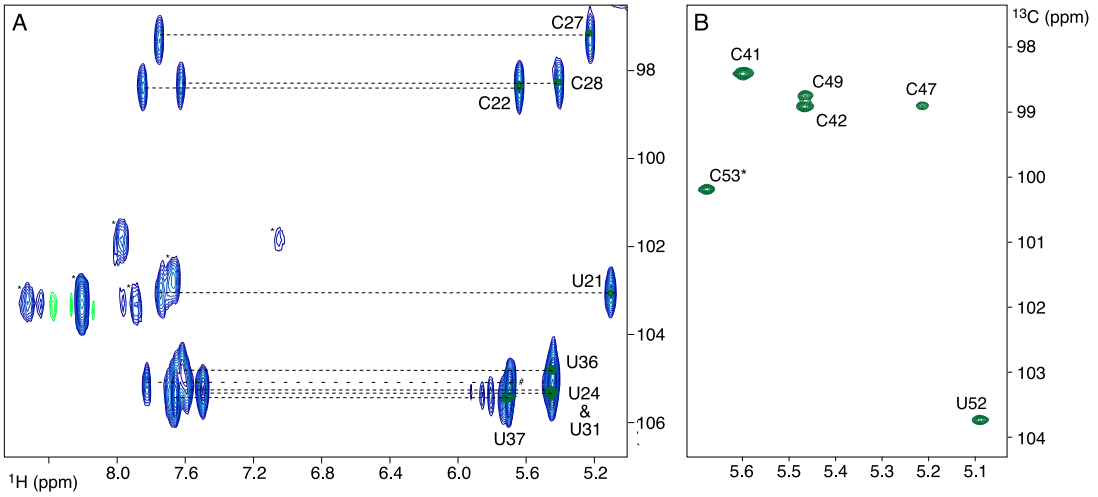


Figure 3.5: *Assignment of H5 protons in pyrimidines* | (A) HCCH-COSY spectra of the 19-mer U4 snRNA strand (blue) overlayed with the ^{13}C-edited HSQC area of the H5C5 correlation (green). Resonances labeled with an asterisk belong to adenine H2 protons. These signals are folded into the spectrum. (B) ^{13}C-edited HSQC of the 14-mer U4 snRNA strand.

NOESY spectra. The H2 proton of A51 exhibit a distinct NOE peak to the H1′ proton of the only uridine residue U52. Hence, the second H2 resonance peak belongs to the other adenine A44.

3.2.5 Sequential assignment

As described in the previous sections the assignment of all non-exchangeable protons of the U4-Kt2 construct could be completed for each nucleotide. However, the identified sugar and base moieties still need to be assigned to specific positions in the RNA sequence. As mentioned in chapter 2.5.6 the sequential assignment can be obtained by well established NOE correlations. In case of a clearly defined A-form helix, base protons H6 and H8 exhibit intense NOE signals to H2′ protons of the preceding nucleotide, a medium NOE signal with the preceding H3′ as well as a weak signal with the preceding H1′ proton. Additionally, a weak NOE signal is seen between H5 and H6/H8 resonances of the preceding nucleotide. It should be noted that NOE contacts for sugars in C2′-endo conformation differ from the described pattern, in the way that instead of a strong H6/H8-H2′(i-1) NOE signal, a strong H6/H8-H3′(i-1) NOE signal can be observed.
An exemplary two-dimensional slice of a ^{13}C-edited, ^{13}C-filtered NOESY spectrum used for sequential assignment of C41 and A44 of the 14-mer strand is shown in figure 3.6. The depicted HH-plane contains chemical shifts of adjacent protons of the two nucleotides. The NOE signals are dispersed in the direct

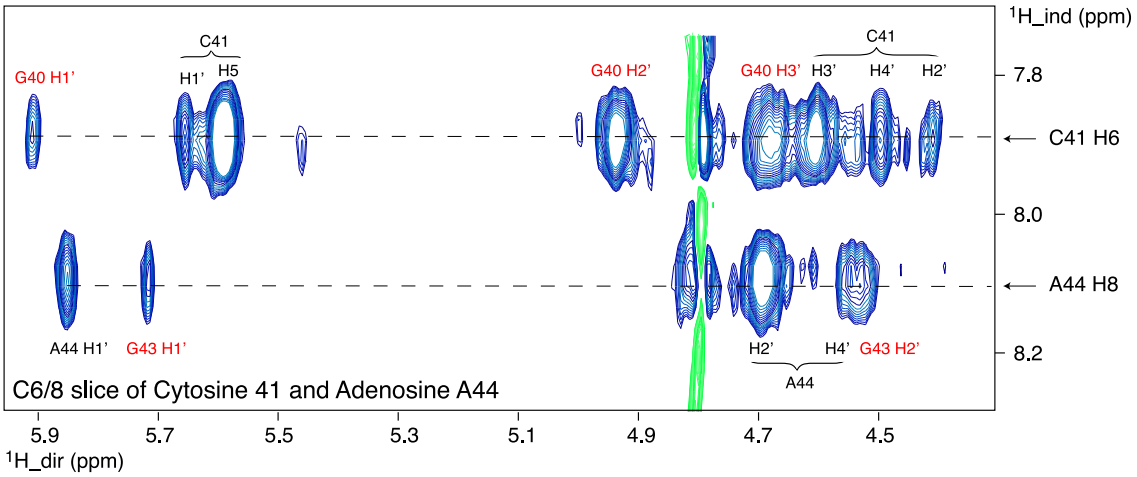


Figure 3.6: *2D slice of ^{13}C-filtered ^{13}C-edited NOESY* | Peaks correspond to intra- (black) and inter-nucleotide (red) connectivities of cytosine 41 and adenosine 44 of the 14 nucleotide strand of the U4 snRNA.

proton dimension, whereas the indirect proton dimension disperses the NOE signals of the respective nucleotides. In the next 2D-plane of the 3D-NOESY spectra the HC-correlation of a certain proton chemical shift can be visualized (H8 of A44 and H6 of C41). In this HC-plane all resonance peaks observed in the direct proton dimension of the HH plane (see figure 3.6), could be assigned to their respective carbon chemical shift. This information deliver the information to which nucleotide the cross-peak, observed in the HH-plane, belongs to. This can be either an intra-nucleotide connection (C41 and A44) or an inter-nucleotide connection (G40 and G43); in the case the NOE arise from a proton of an adjacent nucleotides. Through analysis of NOE spectra the sequential assignment could be completed for both strands of the U4-Kt2 construct (see figure 3.8).

3.2.6 Assignment of exchangeable base protons

Resonances of exchangeable base protons provide crucial information regarding base pairing of RNA strands and enable to measure cross-strand NOE distance constraints used for structure determination [151]. Imino and amino protons were identified by the use of ^{15}N-edited HSQC experiments (see figure 3.7A,B). The imino HSQC delivered eight resonances for the 19-mer and five resonances for the 14-mer of U4 snRNA with one imino resonance of each strand located in the uridine region typical for Watson-Crick base-pairs ($\sim$14 ppm (^{1}H) and 160 ppm (^{13}C)). In the amino HSQC spectra resonances of all 8 cytosines could be observed, whereas only 6 out of 7 adenosine and 3 out of 12 guanosine resonances could be observed due to the rotation of the amino group.
As mentioned in chapter 2.5.7, exchangeable proton resonances can only be observed when protons are protected from a fast exchange with water molecules, which is the case for protons that are involved in base-pairs [31]. Thus, the observation of eight cytosine amino resonances indicates the existence of eight canonical GC base-pairs between the two RNA strands, whereas two cross-strand AU base-pairs are indicated by the imino resonances at 14 ppm and 14.2 ppm.
The assignment of exchangeable protons was performed on the basis of observed NOEs in 2D NOESY experiments with varying mixing times (see figure 3.7C,D). The sequential walk is performed by following imino-imino and imino-amino proton cross-peaks. The imino protons of the wobble base-pair U24/G48 were detected in the 50 ms 2D NOESY due to their strong NOE to each other (see figure 3.7D). The imino protons resonances of U20 and U51 in the canonical stem (C-stem) were easily assignable by their sharp, characteristic NOE to the H2 adenine proton. Furthermore all imino resonances of GC Watson-Crick base-pairs

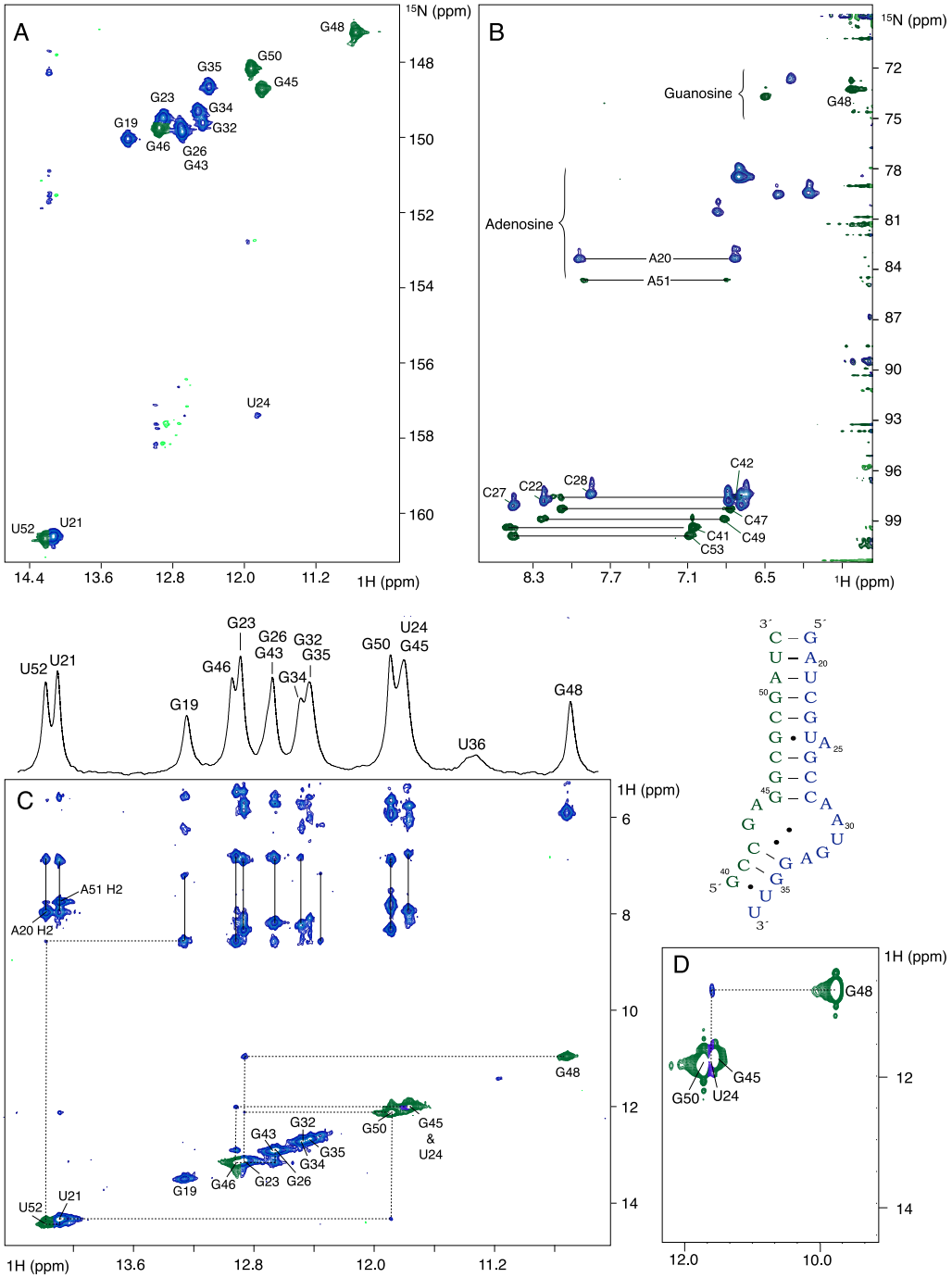


Figure 3.7: *Assignment of exchangeable base protons* | (A) ^{15}N-edited HSQC spectrum of imino resonances H1 (Gua) and H3 (Uri) and (B) ^{15}N-edited HSQC spectrum of amino resonances H61/H62 (Ade), H21/H22 (Gua) and H41/H42 (Cyt). (C) 2D NOESY with 150 ms mixing time with the respective 1D imino spectra given above. Dashed lines indicate the sequential walk and solid lines indicate the intra-basepair cross peaks of the amino protons. U36 is assigned by exclusion (see 1D spectra).(D) Imino-imino cross-peak region of the GU wobble pair in the 2D NOESY with 50 ms mixing time.

could be correlated to the corresponding amino cytosine resonance via their inter-strand NOE in the 150 ms NOESY (see figure 3.7B,C). The assignment of the imino resonances of the GA base-pairs in the non-canonical stem (NC-stem) of the U4 SL construct is based on the observed NOE of the imino proton of G43 to the neighboring amino protons of C42 while G32 is assigned by exclusion.

3.3 NMR data analysis for structure determination of the U4 snRNA 5′ stem loop

Structure determination of biomolecules by NMR mainly depends on inter-nucleotides distance constraints protons derived from the analysis of NOESY spectra described in the pervious chapter. Due to the low density of protons in the phosphodiester backbone, the NMR structure needs to be improved using experimental data of torsion angles. Further, long-range structural information provided by RDCs is required to refine the tertiary structure of the biomolecule.

3.3.1 Structural restraints from Nuclear Overhauser Effect

NOE signals reflect direct distances of neighboring atoms in a molecule and thus give a dense network of distance restraints. The distances between non-exchangable protons were derived from ^{13}C-filtered, ^{13}C-edited NOESY for intra- and inter-nucleotide connectivities and from ^{12}C-filtered, ^{13}C-edited NOESY for inter-strand connectivities as described in the previous chapter. Furthermore 2D NOESY spectra acquired in water were used for distance calculations involving exchangeable protons, which included imino-imino and imino-amino cross-peaks.
The integrated NOE cross-peak volumes were translated into distances using the Felix 2002 program (Accelrys Inc.). For quantification of the cross peak volumes, these have been categorized as weak (1.8 - 5.0 Å), medium (1.8 - 3.4 Å) or strong (1.8 - 2.6 Å). The volume of the pyrimidine H5-H6 cross-peak was used to set the reference distance of 2.45 Å. NOE signals for the inter-nucleotide and inter-strand connections of the two RNA strands are displayed schematically in figure 3.8.

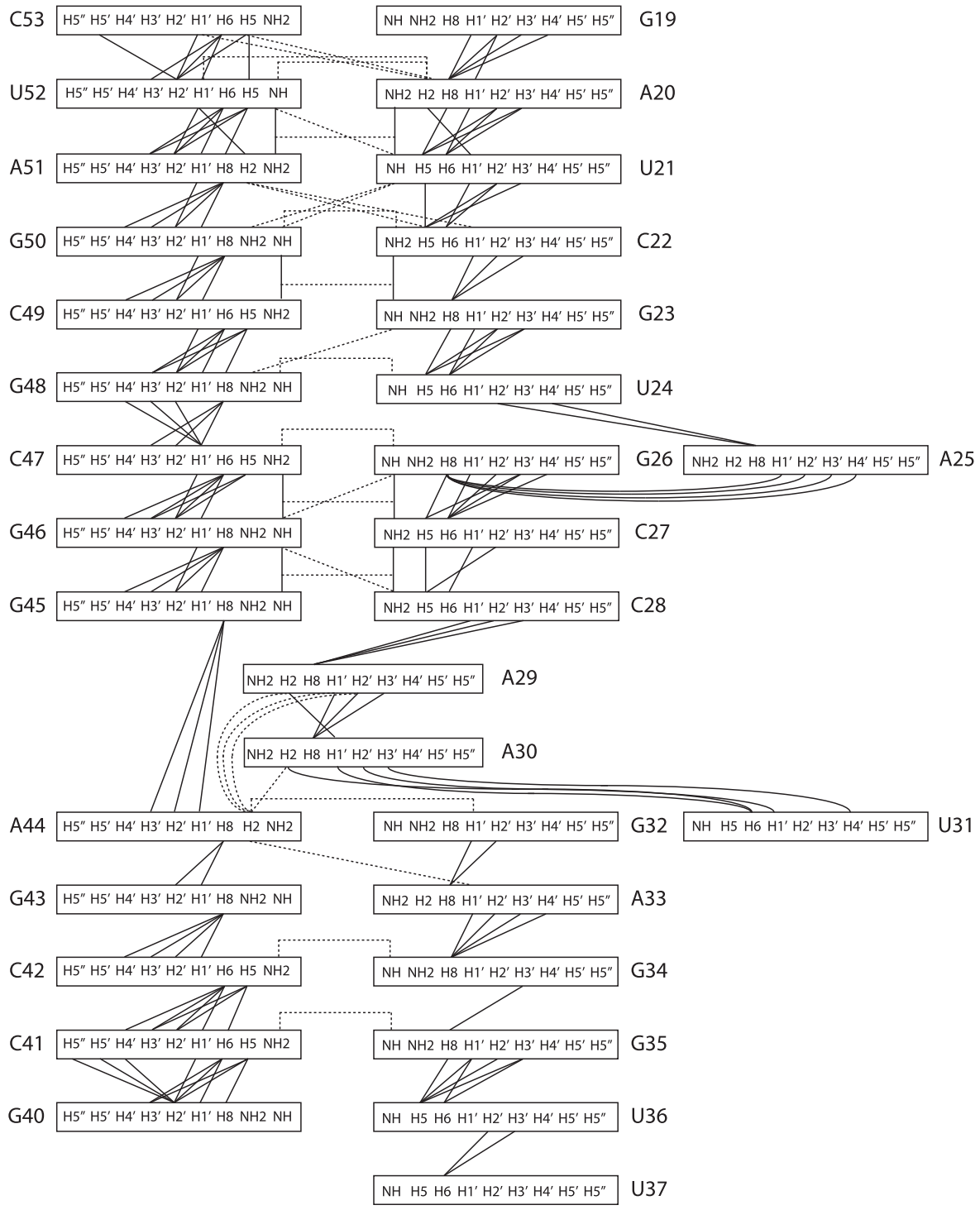


Figure 3.8: *Schematic presentation of internucleotide NOE contacts in the U4-Kt2 construct* | Sequential NOE connectivities are shown with a full line, whereas cross-strand NOE contacts are shown as dotted lines.

3.3.2 Base pairing

The base pairing network of the RNA construct was initially analyzed through imino-imino and imino-amino sequential walks in 2D NOESY spectra (see figure 3.7). The presence of two imino proton resonances between 13.5 and 14.5 ppm indicate the formation of two Watson-Crick AU base-pairs. Further twelve imino resonances could be detected in the 1D imino and 2D NOESY. The resonances at 10.65 and 11.75 ppm are consistent with the GU wobble base-pair, which is present in the C-stem of the U4 snRNA construct (G48 • U24; see figure 3.7D). Eight of the remaining ten resonances could be correlated to Watson-Crick GC base-pairs by imino-amino cross-peaks in the 2D NOESY (see figure 3.7A-C) and subsequently confirmed by the HNN-COSY spectrum (see figure 3.9). In the latter experiment each imino proton H1 (Gua) is correlated with two nitrogen atoms that are involved in hydrogen bond formation of GC base-pairs. The cross-peaks presenting the nitrogen chemical shifts of 145 to 149 ppm belong to the N1 (Gua), which show correlations to the characteristic chemical shifts of N3 (Cyt) found in the range from 194 to 198 ppm.
The two remaining imino resonances belong to the imino protons of the non-canonical GA base-pair present in the NC-stem of U4 snRNA. In the literature

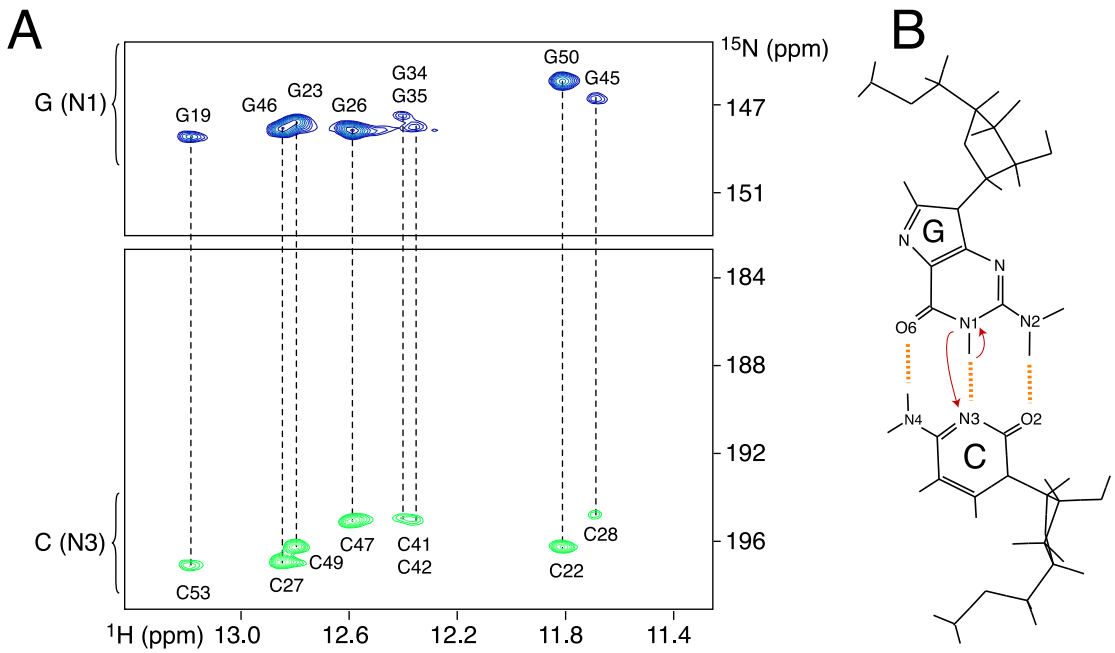


Figure 3.9: *Identification of Watson-Crick GC base-pairs* | (A) HNN-COSY spectrum of fully isotope-labeled U4-Kt2 showing GC base-pairs indicated by dashed lines (B) schematic representation of the magnetization transfer from H1 (Gua) → N1 (Gua) → N3 (Cyt).

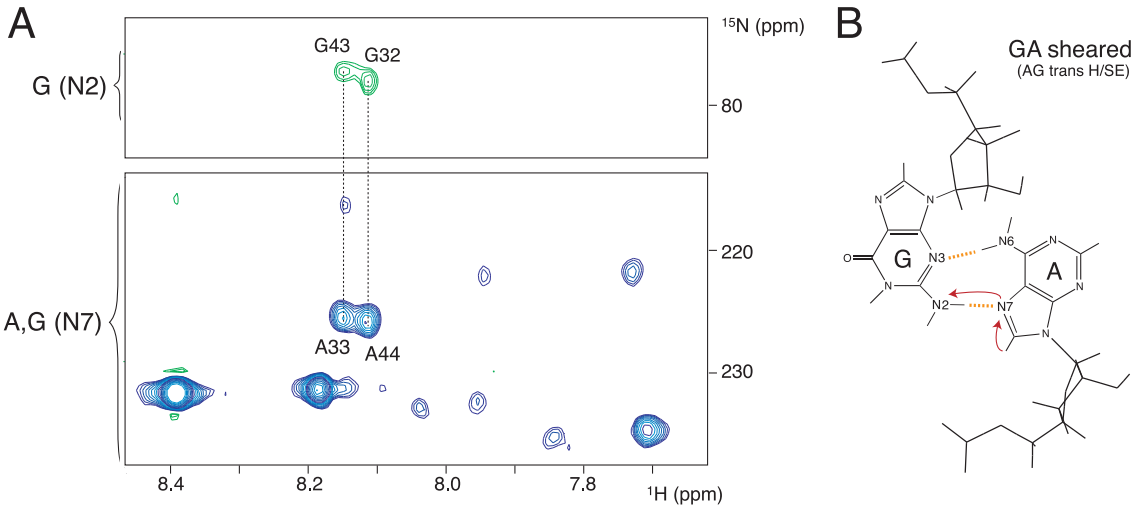


Figure 3.10: *Identification of sheared GA base-pairs* | HNN-COSY spectrum for direct observation of hydrogen bonds A33 • G45 and A44 • G32 (B) schematic representation of the magnetization transfer from H8 (Ade) → N7 (Ade) → N2 (Gua).

.

[23] an opening of the G-A base-pairs is suggested for the protein-unbound form of U4 snRNA 5′ stem loop. It is argued that the NC-stem is unstable in the unbound RNA and an induced fit binding of 15.5K is needed to fold the stem. To analyze whether the GA base-pairs are already formed prior to protein binding, an additional HNN experiment was performed [51]. In this experiment the magnetization is transferred from the H8 to the N7 nuclei of adenosines A33 and A44. From there the magnetization is further transferred via the hydrogen bond to the N2 nuclei of guanine G43 and G32, respectively. The spectra shown in figure 3.10 clearly reveal the existence of the two GA base-pairs. Furthermore the type of GA base-pair could be verified with this experiment, since the transfer of magnetization form A N7 to G N2 is only possible for GA base-pairs in a 'side by side' G (anti): A (anti) conformation, in literature also referred to as sheared GA base-pair [75].

3.3.3 Torsion angles

The conformation of nucleotides is described by phosphate backbone torsion angles, by the endocyclic sugar torsion angles as well as by the relative base to sugar orientation (see figure 3.11). Determination of these dihedral angle constraints is required for accurate description of the RNA backbone. Some of these angles (β, γ, δ, ϵ) can be constrained using semi-quantitative estimates of the magnitude of scalar coupling constants. Karplus equations (see chapter 2.5.8), correlate scalar

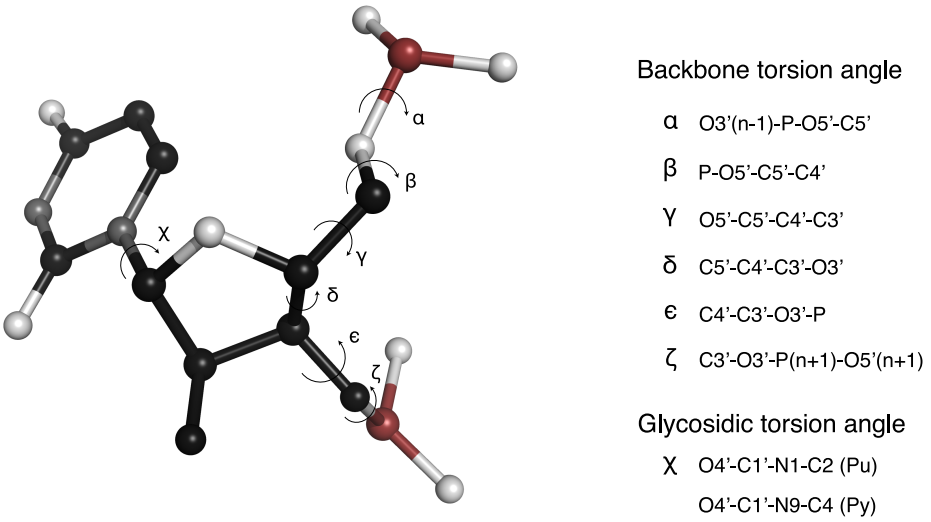


Figure 3.11: *The RNA backbone dihedral angles α to ζ and the glycosidic angle χ.* | The conformation of nucleic acids is mainly determined by six torsion angles defining the conformation of the RNA backbone and the torsion angle χ describing the orientation of the nucleotide base with respect to its ribose moiety.

couplings, and dihedral angles are valued to possible solutions which are consistent with a given scalar coupling value. When several couplings are obtained at the same time, often only a single region of conformational space is consistent with the patterns of observed couplings [1]. For the structure calculation (see chapter 3.4) only those backbone angles were restrained which could be dedicated to a certain conformational space (see figure 3.12).

Furthermore, for the backbone torsion angles α (O3'(i-1)-P-O5'-C5') and ζ (C3'-O3'-P(i+1)-O5'(i+1)) the ^{31}P chemical shifts can be used to restrain the conformational space available to the phosphate dihedrals. When the ^{31}P chemical shifts were comprised between -4 to -5 ppm, the corresponding α and ζ angles were restrained to $\pm$ 120 [1]. Triple resonance HCP spectra were used to constrain 7 out of 19 α and ζ dihedral angles of the 19-mer U4 snRNA strand and 8 out of 14 for the 14-mer strand.

The dihedral angle δ (C5'-C4'-C3'-O3') corresponds to the sugar pseudo-rotation angle defining the ring puckering. In canonical A-form RNA, the ribose sugar adopts a C3'-endo (or North) pucker conformation. In that case the correlation between ribose H1' and H2' protons results in small $^3J_{H1',H2'}$ vicinal couplings (< 2 Hz). In contrast, if a ribose sugar adopts a C2'-endo (or South) pucker, as present in B-form backbone geometry, $^3J_{H1',H2'}$ couplings have higher values (7–8 Hz). Terminal residues in RNA helices often exhibit conformational flexibility switching between different sugar conformations and exhibiting averaged

	ALPHA O3'(i-1) -P-O5'-C5'		GAMMA O5'-C5'-C4'-C3'			ZETA C3'-O3'-P(i+1)-O5'	DELTA C5'-C4'-C3'-O3'		CHI C2-N1-C1'-O4' or C4-N9-C1'-04'	
	^{31}P chemical shift	α	$^{3}J_{H4'-H5'}$	$^{3}J_{H4'-H5''}$	γ	ζ	$^{3}J_{H1'-H2'}$	δ	$NOE_{H6H5'}$ / $NOE_{H6H1'}$	χ
G19*	-3,79	n.d.	6,1	0,4	55 ± 40	0 ± 120	3,3	n.d.	2,2	n.d.
A20	-4,43	0 ± 120				0 ± 120	1,02	82±30	4,4	n.d.
U21	-4,18	0 ± 120				n.d.	0,2	82±30	6,5	190 ± 100
C22	-3,84	n.d.				0 ± 120	0,3	82±30		
G23	-4,60	0 ± 120	5,8	2,1	55 ± 40	n.d.	0,7	82±30	7,0	190 ± 100
U24	-3,43	n.d.	2,4	3,9	55 ± 40	n.d.	6,99	n.d.	5,2	190 ± 100
A25	-3,80	n.d.	2,2	1,4	55 ± 40	n.d.	5,16	n.d.	7,7	190 ± 100
G26	-6,10	n.d.	2.4	5,7	55 ± 40	n.d.	8,19	n.d.	6,0	190 ± 100
C27	-3,83	n.d.				n.d.	0,3	82±30	7,3	190 ± 100
C28	-3,83	n.d.	6,6	3,2	55 ± 40	0 ± 120	1,3	82±30	7,4	190 ± 100
A29	-4,14	0 ± 120				n.d.	2,4	n.d.	4,7	n.d.
A30	-4,06	n.d.				n.d.	5,2	n.d.	3,5	n.d.
U31	-4,06	n.d.	2,2	0,1	55 ± 40	n.d.	6,48	n.d.	3,9	n.d.
G32	-2,87	n.d.				n.d.	6,5	n.d.	3,3	n.d.
A33	-3,77	n.d.				0 ± 120	0,52	82±30	4,5	n.d.
G34	-3,92	0 ± 120	6,1	4,6	55 ± 40	0 ± 120	0,6	82±30		
G35	-4,46	0 ± 120				0 ± 120	2	n.d.	6,2	190 ± 100
U36	-4,28	0 ± 120	7,7	7,4	n.d.	n.d.	0,4	82±30	4,4	n.d.
U37	-3,14	n.d.	5	5	n.d.	n.d.	3,24	n.d.	1,7	n.d.
G40*	-4,67	n.d.	2,2	4,5	55 ± 40	0 ± 120	0,6	82±30		
C41	-3,93	0 ± 120	3,9	7,7	n.d.	0 ± 120	1,8	82±30	3,4	n.d.
C42	-4,13	0 ± 120	*	1,2	n.d.	n.d.	1,2	82±30	4,5	190 ± 100
G43	-3,49	n.d.	5,6	3,1	55 ± 40	n.d.	3	n.d.	2,8	n.d.
A44	-3,51	n.d.	3,7	4,5	55 ± 40	0 ± 120			3,2	n.d.
G45	-3,93	0 ± 120				0 ± 120				
G46	-4,09	0 ± 120	0,9	2,3	55 ± 40	n.d.	0,6	82±30	5,4	190 ± 100
C47	-3,81	n.d.	small*	1,6	55 ± 40	0 ± 120	0	82±30	6,2	190 ± 100
G48	4,50	0 ± 120				n.d.	1,8	82±30	5,1	190 ± 100
C49	-3,83	n.d.	5	3	55 ± 40	0 ± 120	0,6	82±30	5,5	190 ± 100
G50	-4,12	0 ± 120				0 ± 120	1,2	82±30	4,4	190 ± 100
A51	-4,57	0 ± 120	6,2	7,4	n.d.	0 ± 120	1,2	82±30	4,6	190 ± 100
U52	-4,32	0 ± 120	3,3	2,4	55 ± 40	n.d.	2	82±30	3,5	n.d.
C53*	-3,24	n.d.	0	3	55 ± 40	n.d.	4,8	n.d.	3,9	n.d.

Figure 3.12: *Experimentally determined values for torsion angles defining the conformation of the RNA backbone* | n.d. = not defined as torsion angle constraint, * = values could not be determined due to spectral overlaps.

$^{3}J_{H1',H2'}$ couplings. Internal residues within helical regions do not exhibit such dynamic behavior. As listed in figure 3.12, flexible RNA regions around the bulged-out A25 nucleotide as well as the internal loop region (A29 to G32 and G43) can be involved in conformational averaging. The scalar coupling between H1′ and H2′ protons were measured with 3D HCCH-E.COSY spectra.

The γ (O5′-C5′-C4′-C3′) dihedral angles were constrained using semi-quantitative estimates of $^{3}J_{H4'H5'}$ and $^{3}J_{H4'H5''}$ couplings obtained from a 3D HCCH-E.COSY experiment. In the *gauche(+)* conformation found in A-form helices, both couplings are small, whereas in the *trans* or *gauche(-)* conformations $^{3}J_{H4'H5'}$ or $^{3}J_{H4'H5''}$ is large (approximately 10 Hz). When one coupling was less than 5 Hz and the other not > 7 Hz, γ was constrained to the *gauche(+)* conformation (55±40°).

The χ angle (O4'- C1'- N1/N9-C2/C4) defines the relative position of the base to its sugar and is referred to as glycosidic angle. Glycosidic angle constraints can be introduced from interproton distance constraints such as the H1'-H8 NOE cross-peak intensity for purine bases or the Hl'-H6 intensity for pyrimidine bases. H8/H6 resonances from nucleotides in the *syn*-glycosidic conformation are in close contact to the H1' sugar proton, which is revealed through a very strong cross-peak, stemming from a short distance of 2 to 2.5 Å. In the *anti*-conformation, no close contact between H8/H6 and H1' is expected resulting in a weak NOE peak corresponding to a distance of 3.5 to 4.5 Å. Since NOE-based distance information is already present in the NOE constraint list, the glycosidic angle should be sufficiently restrained. Indeed the introduction of defined χ angles in the structure calculation revealed that no explicit torsion angle constraint is required to describe the χ angle.

3.3.4 Residual dipolar couplings

In order to observe residual dipolar couplings (RDCs, see chapter 2.5.8) by solution NMR, partial orientation of the macromolecule in the solution is required. Then the macromolecule adopts a slightly preferred orientation in the magnetic field (see figure 3.13B). Magnetically aligned filamentous bacteriophage were used in this work to induce alignment of the RNA in solution [49]. Residual Dipolar Couplings were determined by measuring the difference between ^{1}H-^{13}C couplings for isotropic and partially aligned samples (see figure 3.13A&C). The splitting of the deuterium signal was used to adjust the phage concentration. For the 14-mer strand of the U4 snRNA a splitting of the deuterium signal of 9.58 Hz was observed whereas the 19-mer strand showed a splitting of 11.3 Hz. The splitting arises from the large deuterium quadrupole moment that is not isotropically averaged for water in prescence of the aligned phage particles. The observed value of the splitted signal varies approximately linearly with the phage concentration and can be used as an indicator for the degree of ordering of the RNA molecules in solution.

RDCs were obtained for the following correlations: H8-C8 (Pu), H6-C6 (Py), H5-C5 (Py), H2-C2 (Ade) for bases and H1'-C1', H2'-C2', H3'-C3', H4'-C4' for sugar moieties by analyzing the ^{13}C-edited HSQC spectrum with help of the Felix program (Accelereys Inc.). As a result 79 RDCs could be derived, i.e. 35 ^{1}H-^{13}C RDCs for bases and 44 for sugar moieties. The values for the rhombic (r) and axial (D_a) components of the alignment tensor were adjusted to D_a = 14.0 and r = 0.2 in the structure calculation protocol and the final low-energy structures were refined using these additional long range restraints.

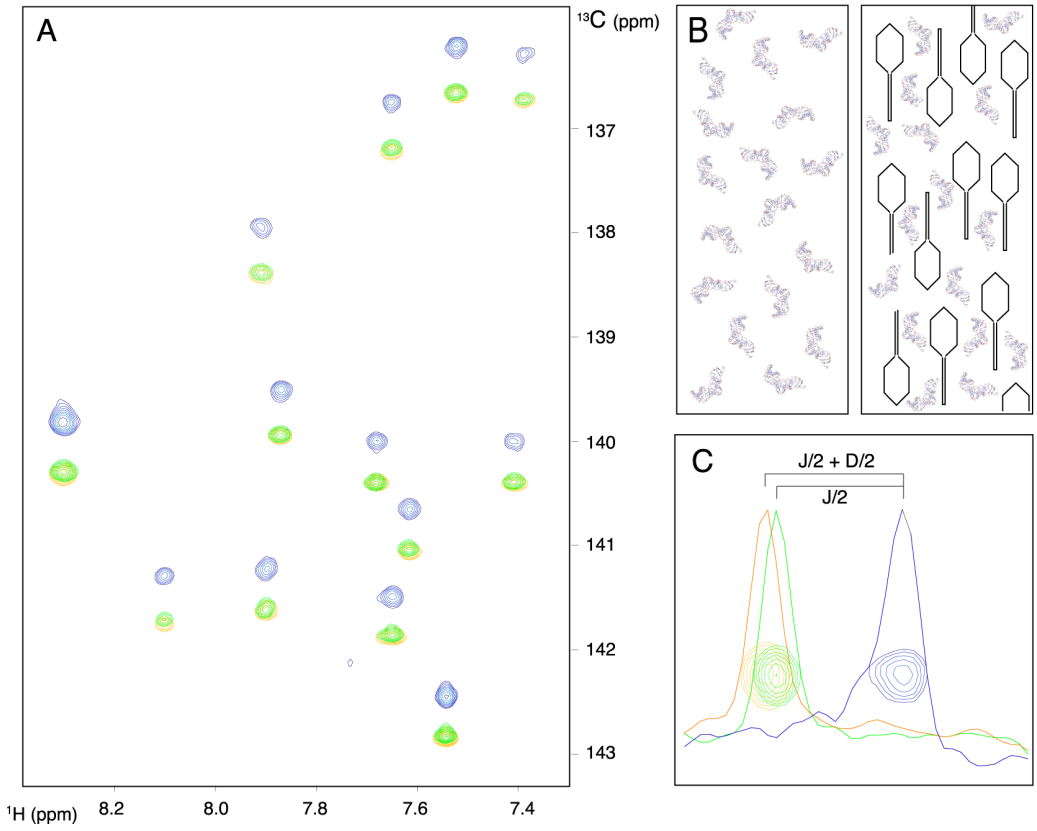


Figure 3.13: *Determination of residual dipolar couplings (RDCs)* | (A) ^{13}C- edited HSQC to measure RDCs for H6C6 and H8C8 correlations. Blue and green peaks present ^{1}H-^{13}C couplings under isotropic conditions, blue and orange peaks under anisotropic conditions (with Pf1 phages). (B) Schematic picture of isotropic and anisotropic behavior, which is induced by phages, of RNA molecules in a NMR tube (C) 1D slice of ^{13}C- edited HSQC spectrum to visualize the determination of the residual dipolar couplings.

3.4 Structure calculation of U4 5′ stem loop

Structure calculations and RDC refinement was performed by Irene Amata within the scope of her PhD thesis. A total number of 100 structures were calculated with a simulated annealing protocol applying the Aria/CNS software [76]. For structure calculations, structural restraints described in the previous sections, 623 NOE-derived distances (see figure 3.8), 80 dihedral angles (see figure 3.12) as well hydrogen bonds (see chapter 3.3.2) were included (see table 3.2).
During the structure calculation step, hydrogen bonds were maintained by distances restraints and their planarity was enforced trough weak planarity restraints (5 kcal mol^{-1} Å^{-2}). In the first step, an extended structure generated by CNS [8] was used to calculate 100 initial structures using NOE, dihedral, hydrogen bonding and planarity restraints.

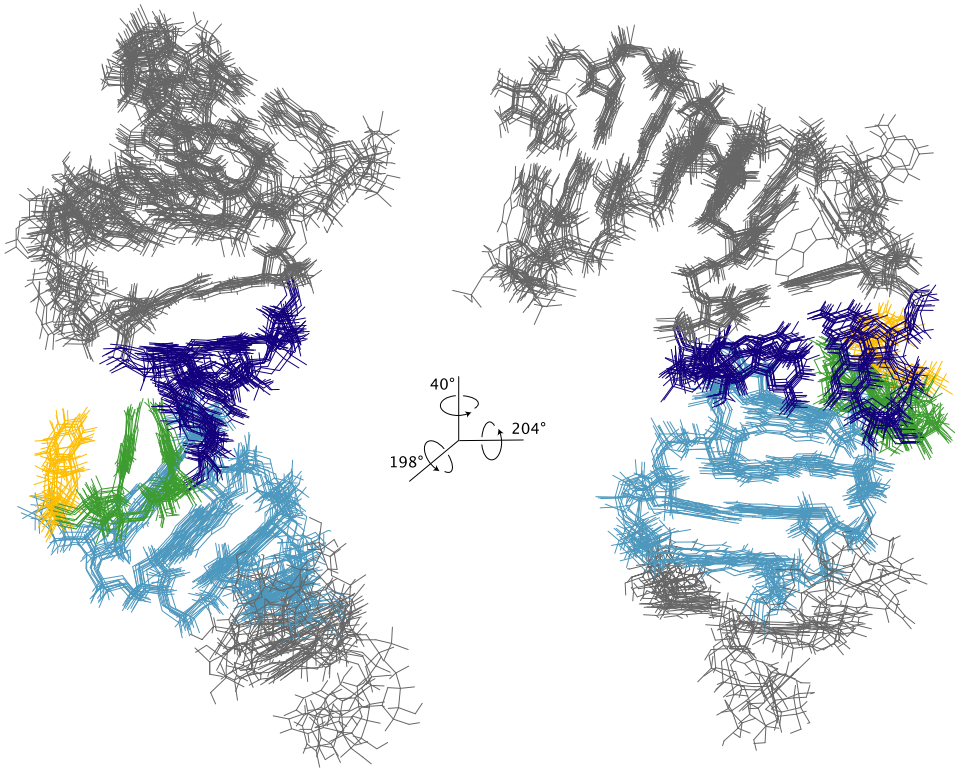


Figure 3.14: *The superposition of the 10 lowest-energy NMR-derived structures for the U4 5′ stem loop construct U4-Kt2* | The structures are superimposed by excluding the highly flexible bulged-out nucleotide A25, the interloop nucleotide U31 (orange), the terminal nucleotide U37 as well as the terminal base-pairs. Conserved nucleotides are color-coded, in blue = C-stem, in light blue = NC-stem, in green and orange = internal loop residues.

The resulting RDC-refined structures were accepted based on their convergence quality and their low overall energy. Further, no significant violations for NOE (> 0.5 Å) and dihedral (> 5°) restraints were found.

The final ensemble of the 10 lowest energy structures prior to RDC refinement shows an excellent convergency for both stems of the U4-Kt2 construct, while the relative position of the two stems was less well defined with a root main square deviation (RMSD) value of 0.98 Å. After refinement including RDC derived long-range structural information and minimization in water, the 10 lowest energy structures converged to 0.61 Å, while the convergency for the NC and C stems are 0.44 Å and 0.50 Å, respectively (see table 3.2). The RMSD was calculated excluding the highly flexible nucleotides of the bulged-out A25 and U31 and the terminal U37 as well as the terminal base-pairs G19-C53 and U36-G40 were excluded. The derived structure of the U4-Kt2 RNA is very well defined not only with respect to the conformation of the two stems, but also with respect to the relative orientation of both stems to each other (see figure 3.14).

Table 3.2: Summary of NMR restraints and structural statistics

A. *Restraint statistics*	
(1) NOE-distance restraints	623
Intra-nucleotide	440
Inter-nucleotide	183
Inter-strands	27
Hydrogen bonds	86
(2) Dihedral angles restraints	80
(3) Refinement by RDC	
Total RDC derived restraints	79
Base	35
Sugar	44
(4) Total constraints	
Total NMR derived number	782
Mean number per residue	$\sim$ 24 /nt
B. *Superposition of structures**	
	RMSD (Å)
10 structures (all nt)	0.61 ($\pm$0.13)
10 structures (C-stem)	0.50 ($\pm$0.17)
10 structures (NC-stem)	0.44 ($\pm$0.08)
Number of NOE violations (>0.5 Å)	0
Number of dihedral angle violations (>5°)	0

* For RMSD evaluations the residues A25, U31 and U37 as well as the terminal base-pairs were excluded from the superimposition. C-stem and NC-stem rmsd were calculated considering the residues ranges [A20-C27, G46-U52] and [G32-G25, C41-A44] for C-stem and NC-stem, respectively.

3.4.1 Structure of the unbound U4-Kt RNA

The structural results of the U4-Kt2 in solution show that the RNA is very well defined both in the conformation of the two stems and in their relative orientation. The structure quality was assessed using the Q-factor calculated from the measured and calculated RDC restraints [19] which resulted an excellent fit of Q = 0.17. The conformation of the U4-Kt RNA in solution is much more extended than that observed in protein bound k-turn motifs. In Cojocaru *et al.* [23] an angle ϕ is defined between the phosphates of C47, U31 and G35 to describe the kink in the RNA backbone. This angle assumes a value of 25° in the protein-bound crystallographic conformation of k-turn RNAs (see figure 3.15B), while it becomes larger for more extended conformations. In the unbound case the ϕ angle equals 69°, which indicates that the k-turn is not formed in solution in absence of cognate proteins (see figure 3.15A). As a consequence of this, no A-minor interaction

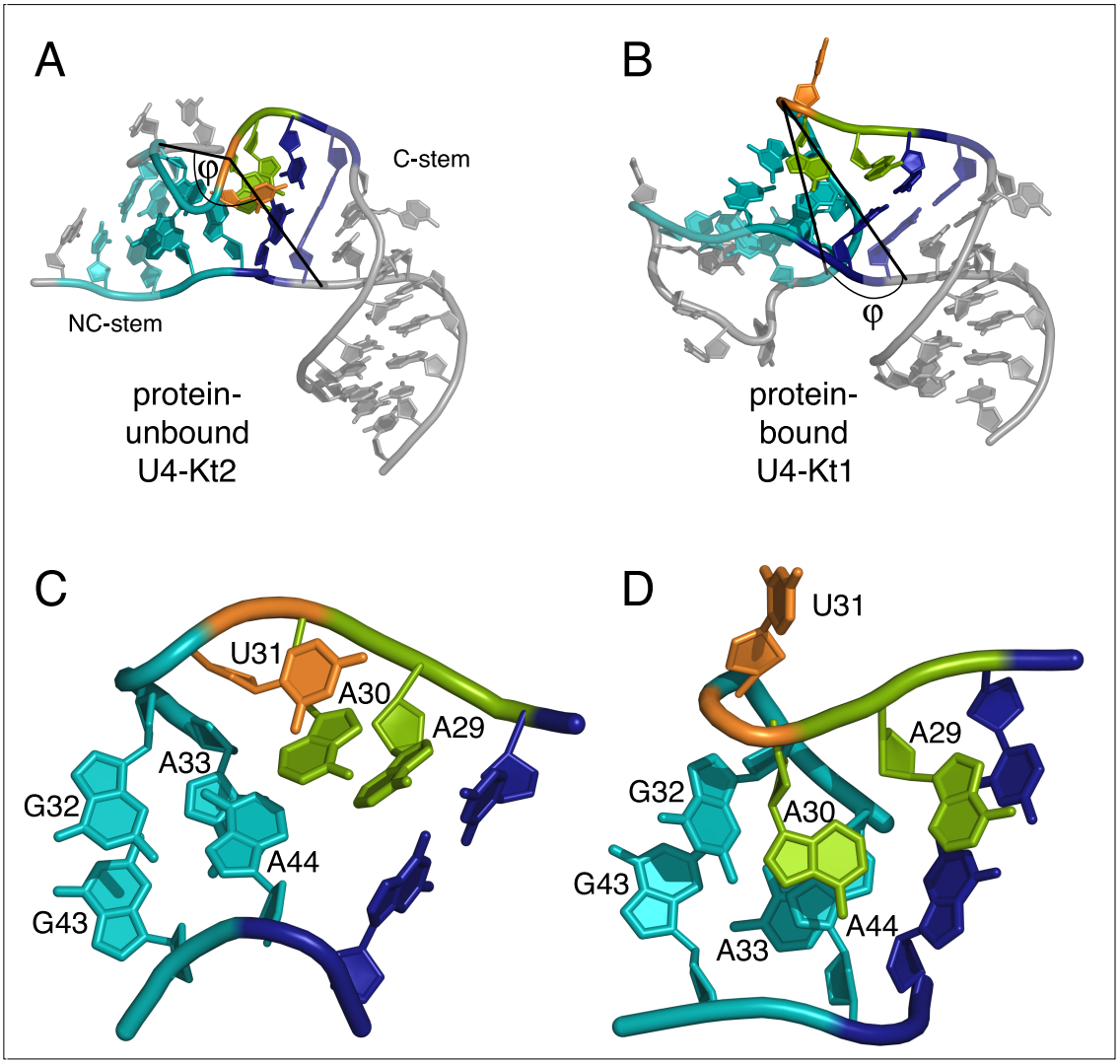


Figure 3.15: *Comparison of the U4 5′ stem loop structures in the free and protein bound state* | (A) Lowest energy NMR structure of the unbound U4-Kt2 and (B) the crystallographic protein-bound structure of U4-Kt1 (2OZB.pdb). The angle ϕ between the P nuclei of C47, U31 and G35 describes the bent between the two stems. The NMR structure of the unbound RNA is quite extended (ϕ = 69°) while the protein-bound structure is sharply kinked (ϕ = 25°). (C and D) The close-up of the internal loop region shows the difference in the location of the unpaired A29 and A30 in the two structures. (C) In the unbound RNA structure A30 stack on the C-stem with A29. (D) In the protein-bound RNA structure A30 stacks on the NC-stem.

is observed between A33 of the NC- stem and G45 of the C-stem, as confirmed by the absence of NOEs between the ribose of G45 and either the ribose or the aromatic protons of A33 (see figure 3.8 and 4.2).

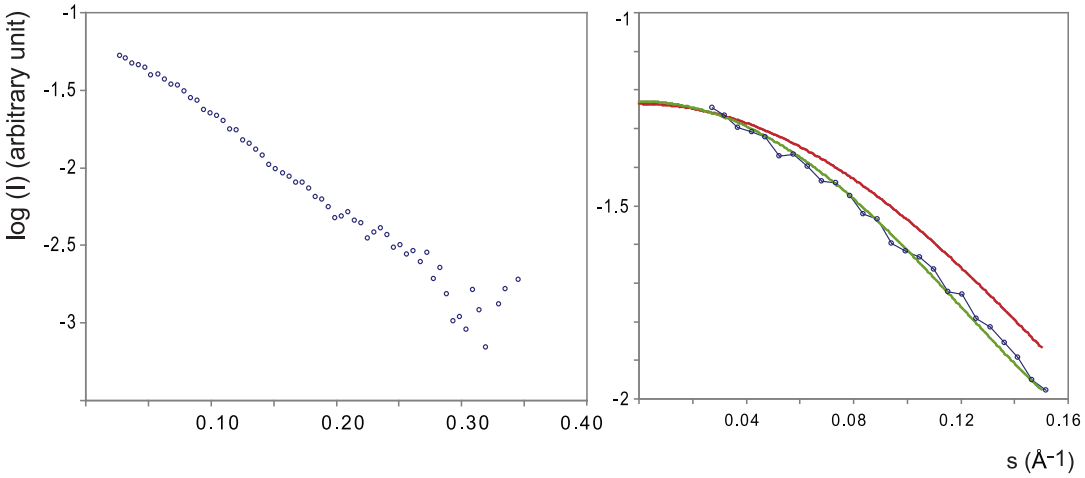
A detailed analysis of the region between nucleotides 28-33 and 43-45 reveals that the k-turn consensus sequence is indeed well-folded in the unbound RNA, but that the pattern of inter-nucleotide interactions substantially differ from that

in complex with proteins. In the unbound structure A30 stacks on A29 (see figure 3.15C), which is situated on the top of C28 of the C-stem. In the sharply kinked structure A30 does not stack on A29 but rather on the NC-stem (see figure 3.15D). The large relocation of A30 from the top of the NC-stem in the protein-bound structure to the top of A29, and thus of the C-stem, in the free RNA solution structure is accompanied by a change in the conformation of the ribose of A30 from C2′-endo to C3′-endo and of the χ dihedral angle form syn to anti. The calculated structural ensemble shows a very well converged C3′-endo conformation for the ribose of A30, which was left unrestrained during the calculations. However, the value of 5.2 Hz measured for the $^{3}J_{H1'H2'}$ scalar coupling of A30 (see figure 3.12) is indicative of a ribose interchanging between the C3′-endo and C2′-endo conformations, suggesting that a certain degree of local structural flexibility is present in the internal loop of the U4-Kt RNA in solution.
However, despite the local dynamics, all NMR data converge to the well-defined major conformation of figure 3.15A. This was verified by the observed line widths of the proton signals which reflect both, structure and dynamics, of nucleotides. For instance, the proton signals for nucleotides in a duplex structure are broader than those in a corresponding single strand or in terminal regions. An intrinsic dynamic is observable if two signals arise for one nucleotide or a considerably broadening of signals is observed showing intermediate exchange on the NMR time-scale between two conformations of the observed molecule. The C-H correlation spectra of both strands of the U4-Kt2 (see figure 3.2) show the presence of one single conformation, thus excluding that the U4-Kt RNA exists in two slowly (> ms) inter-converting conformations that are both substantially populated in solution. Similarly, the sharp line width of the resonances of both the NC- and C-stems and of the internal loop excludes that any of these regions undergoes conformational exchange in the high μs to ms time-scale between two highly populated conformations.

3.4.2 Verification of the NMR structure of U4-Kt2 by SANS

In order to verify the NMR structure described in the previous chapter (see figure 3.14), small angle neutron scattering (SANS) data was obtained to build a low resolution model of the U4 snRNA 5′stem loop. This SANS-derived model could then be compared to the two high resolution structures U4-Kt1 and U4-Kt2. While the crystal structure of the U4-Kt1 construct presents the U4 5′ SL in the protein-bound form [71], the U4-Kt2 NMR structure described above reflects the U4 5′ SL in the protein unbound form.

Table 3.3: Radii of gyration of U4 RNA | The experimental scattering curve of U4-Kt2 (blue) is compared to the scattering curve calculated from the lowest energy NMR structure of the U4 snRNA in the protein unbound form (U4-Kt2, green) and the calculated X-ray structure in the protein bound form (U4-Kt1, red)



	U4-Kt2 experimental	U4-Kt2 (protein-unbound) simulated	U4-Kt1* (protein-bound) simulated
R_g [Å] $^{(a)}$	16.6 ± 0.7	16.2	15.0
χ^2 $^{(b)}$	-	0.81	1.14

(a) The radius of gyration (R_g) defines the size of the molecule and (b) χ^2 describes the fitting between the experimental and simulated scattering curves, from the two different PDBs. * To compare structures, the U4-Kt1 pdb entry was modified by adapting the pentaloop and terminal region to resemble the U4-Kt2 construct.

SANS measurements and analysis were carried out by Irene Amata with the kind help of Frank Gabel (Institute de Biologie Structurale, Grenoble), whereas the sample preparation was in the scope of this work (see chapter 2.4.1). The SANS data were acquired using the small-angle diffractometer D22 at the Institute Laue-Langevin (in Grenoble). A detailed description of SANS measurements and data analysis is given in Falb *et al.* (submitted).

The radius of gyration (R_g), which can be directly derived from SANS data, delivers important quantitative structural information. The R_g value is a size parameter that is defined as the square root of the average squared distance of each scatterer from the particle center [106]. From the experimental scattering curve of the SANS measurements a radius of gyration of 16.6 ± 0.7 Å was retrieved for the U4-Kt2 construct (see table 3.3).

The R_g value of a given molecule can also be predicted from a structural model. Comparison of the experimental with the predicted scattering curve is a powerful tool to discriminate whether the folding of a known molecule presents a reason-

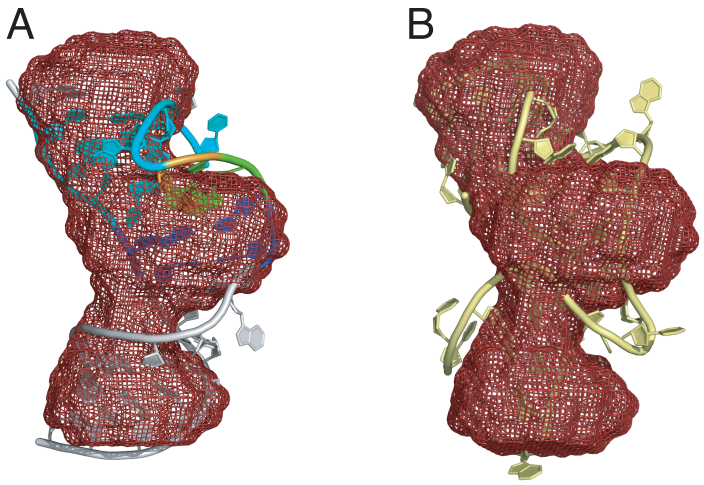


Figure 3.16: *SANS measurements and low resolution 3D modeling* | (A) Overlay of the low resolution 3D model derived from SANS data with the NMR structure of the protein unbound form of the U4 snRNA (U4-Kt2). (B) Overlay of the SANS model with the bound form of the U4 snRNA from the adapted crystal structure (U4-Kt1).

able model for the sample under investigation. The predicted R_g value from the lowest energy NMR-structure is 16.2 Å (see table 3.3), which is in excellent agreement with the experimentally derived value. The fitting of the predicted curve with the experimental curve is expressed by the χ^2 factor which reaches a good value of 0.81 for the simulated curve of the protein-unbound U4-Kt2 construct. To fit the crystal structure of the protein-bound U4 RNA (U4-Kt1) [71] with the SANS data obtained for the open loop construct of U4 snRNA (U4-Kt2), the PDB entry OZ2B of the X-ray structure was modified by removing the pentaloop of the closed loop version of the U4 snRNA and by adding the base-pairs at the end of the RNA stems. For the adapted X-ray structure data, a radius of gyration of 15.0 Å was calculated and a χ^2 value of 1.14 was determined when comparing the simulated with the experimental scattering curves. Thus, the R_g value for the protein-bound U4-Kt1 construct is far away from the experimental range of R_g for U4-Kt2 (table 3.3).

SANS data can further be used to build a low resolution 3D model. The final SANS model was aligned to the lowest energy NMR structure of U4-Kt2 as well as to the adapted X-ray structure (see figure 3.16A,B), revealing a much better fitting of the NMR-structure to the experimental SANS data.

3.5 Ion dependency of the U4-Kt2 fold

Magnesium ions strongly effect RNA folding processes due to their interaction with the negatively charged nucleic acid backbone [32]. Both, NMR and SANS experiments, were performed without magnesium ions in the solutions and led to structural models showing the RNA in an extended conformation. To test the effect of Mg^{2+} on the conformation of the k-turn region of the U4-Kt2 RNA in solution, titrations with $MgCl_2$ were performed (see figure 3.17).
Four titration points in the range of 0 mM to 10 mM Mg^{2+} were recorded for the $^{13}C/^{15}N$ labeled I2 strand paired with the unlabeled I1 strand. The addition of magnesium ions showed dramatic effects on both, the position and line-width of resonances such as those of A29, A30, U31, G32 and A33 and to a minor extent those of G34 and G35. The C8/H8 resonance of A29 shifted and broadened pro-

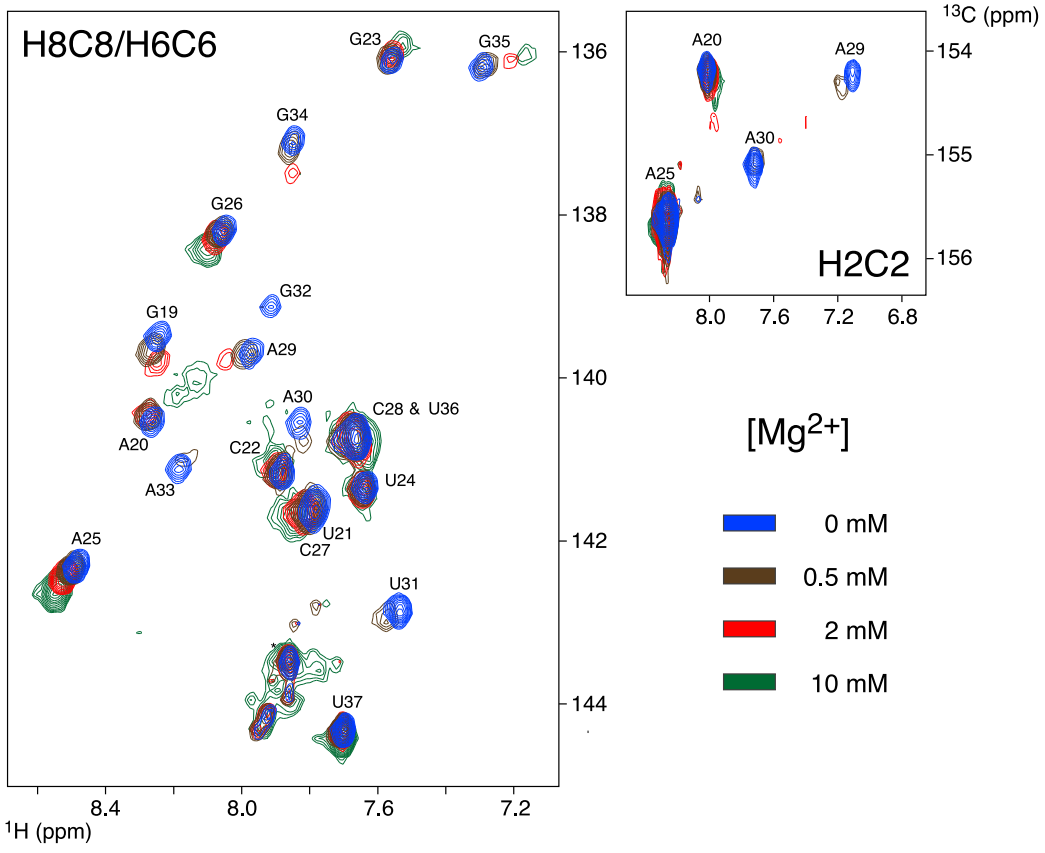


Figure 3.17: *Mg^{2+} titration with U4-Kt2 monitored with ^{13}C-edited HSQC spectra* | Overlay of four titration points in the range of 0 mM to 10 mM Mg^{2+} for the $^{13}C/^{15}N$ labeled I2 strand in complex with the unlabeled I1 strand. The C6/C8/C2-H6/H8/H2 correlations show that the resonances of the A29-G35 region are affected the most upon addition of Mg^{2+}.

gressively upon addition of Mg^{2+}, while the C8/H8 resonances of A30, U31, G32 and A33 completely disappeared at magnesium concentrations of 2 mM. The observed strong line-broadening of resonances is due to an exchange process in the high μs to ms time-scale and effects mainly the internal loop and, to a minor extent, the NC stem. The observed changes to the spectra are likely correlated to the conversion from the extended to the bent RNA conformation triggered by divalent cations, as suggested previously [40] [83].

3.6 Dimeric U4 snRNA-15.5K complex

To gain further insights into the subsequent steps of the U4 snRNA in the spliceosomal assembly, the binding of 15.5K and hPrp31 with the U4-Kt2 RNA construct was investigated. Earlier studies [98][71] of the RNA fold already described the U4 snRNA 5′ stem loop in its protein-bound form and showing the characteristic k-turn motif. This work focussed on the RNA site of the U4 snRNA assembly, whereby the determination of and the interaction of the C-stem of the RNA construct with hPrp31 was planned. Therefore the binding of 15.5K to the U4-Kt2 construct needed to be verified first. The 15.5K protein was expressed using an available vector *pGEX-15.5K* [98] resulting in a GST-tagged version of 15.5K. For the purpose of long-term NMR measurements of the protein in complex with isotope-labeled RNA, the purified protein solution was assessed regarding its RNase activity at the end of the purification steps. Unfortunately all purification attempts showed clear RNase-activity of the 15.5K protein solution, although SDS-PAGE gels showed just one distinct protein band for 15.5K lacking other contamination bands. Purification steps and parameters were varied i.e. by elongating washing steps, by extensively cleaning the used apparatus or by repeating purification steps. However, no progress could be achieved regarding RNase activity. As a consequence a novel constructs needed to be designed and cloned leading to a His-tagged and GST-/ His-tagged versions of 15.5K. By introducing extensive washing steps with high salt during affinity chromatography, RNase-free 15.5K could be produced. Finally, the His-tagged 15.5K protein purification yielded reasonable protein amounts of $\sim$ 10 to 20 mg per one liter expression in minimal media.

The formation of the dimeric protein-RNA complex was first verified by gel shift assays (see chapter 2.3.7) and subsequently investigated by NMR spectroscopy, while working with uniformly ^{15}N-labeled 15.5K. The protein itself was quite unstable and precipitated readily during NMR measurements, while protein stability was greatly improved when it was bound to the U4 RNA. This observation

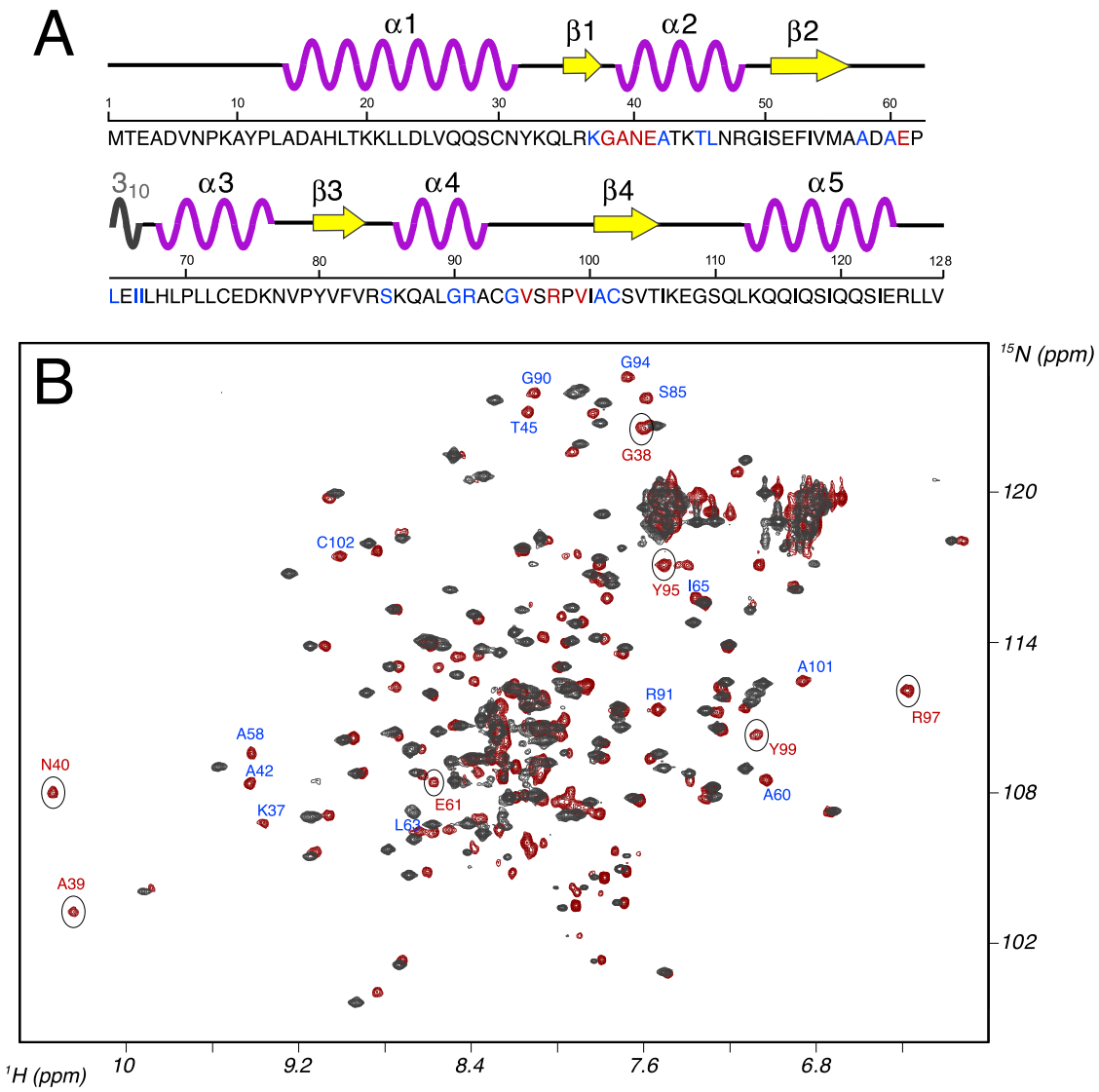


Figure 3.18: *Investigations of labeled ^{15}N-15.5K protein within U4-15.5K complex* | (A) Amino acid sequence of 15.5K used in this study. Secondary structure elements are indicated above. The helix 3_{10} (gray) forms upon RNA binding. Residues of resonances newly appearing upon binding of U4-Kt2 are highlighted in red letters whereas residues with large chemical shift perturbations are shown in light blue letters in A and B. (B) Overlay of 15.5K in RNA-unbound form (gray) and in RNA-bound form (red) recorded with a ^{15}N-edited HSQC experiment. Resonances only appearing upon binding of U4 5′ SL RNA are circled. The assignment of protein resonances [62] was performed in our group by Ping Li. Amino acids marked in red and in light blue have direct RNA contacts in the crystal structure of the dimeric complex [71]. Amino acid residues G38, A39 and N40 belong to helix α2, E61 is located between β2 and helix 3-10 and residues V95, R97 and V99 belong to the loop between α4 and β4.

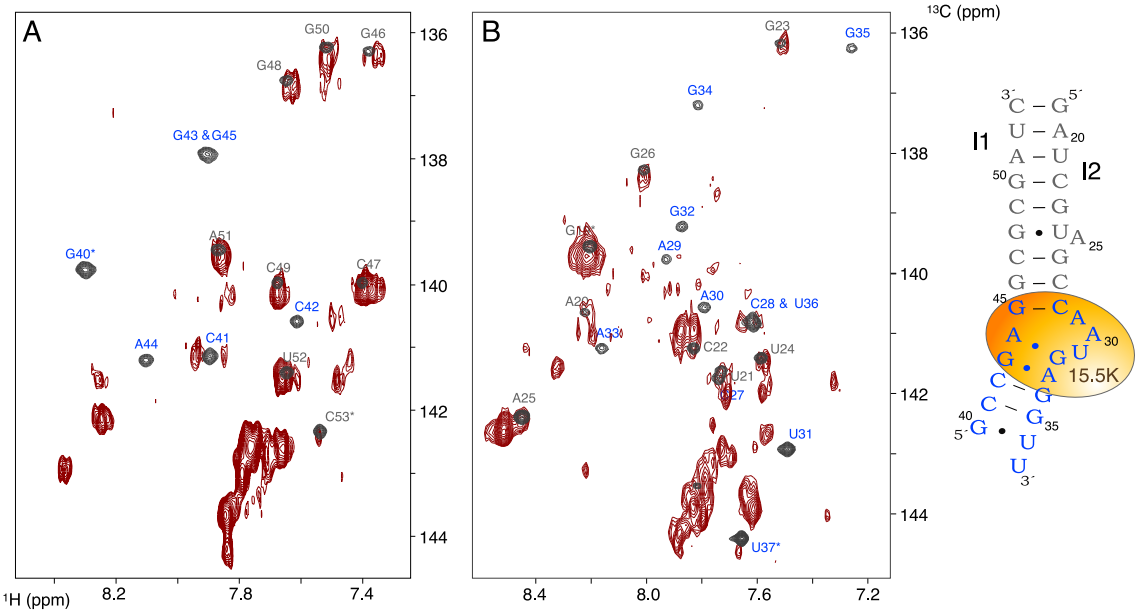


Figure 3.19: *Investigations of labeled ^{13}C,^{15}N-U4-Kt2 RNA in the U4-15.5K complex* | Overlay of U4-Kt2 in the unbound form (gray) and in the protein-bound form (red) Blue letters indicate residues with strong chemical shift perturbation or line broadening upon 15.5K binding, whereas for residues assigned with gray letters retain their chemical shift in the presence of 15.5K. On the left panel the secondary structure of the U4-Kt2 is shown as well as the binding site of 15.5K.

is consistent with the fact that 15.5K protein alone failed to crystallize indicating an intrinsic instability of the free protein in high concentrations. As expected, the HN correlations of 15.5K showed additional resonances upon RNA binding. Figure 3.18 shows the assignment for characteristic resonances for 15.5K based upon Kirkpatrick *et al.* [62] and comprises amino acids with direct RNA contacts as known for the crystal structure [71].

To investigate the changes elicited in the RNA by protein binding, two NMR samples were prepared by dissolving an unlabeled RNA strand and a ^{13}C, ^{15}N-labeled strand. The RNA duplex was mixed with ^{15}N-labeled 15.5K and ^{13}C-HSQC experiments were recorded (see figure 3.19). By comparing the HN correlations of the free and bound form, the resonances of the C-stem of U4-Kt2 exhibit no or only slight chemical shift changes upon protein binding (gray letters in figure 3.19). This is due the fact that the complex is larger than the RNA alone resulting a change in the correlation time shifting the molecule motion to the intermediate exchange range.

Proton resonances within the protein binding site (C28-A33, G43-G45) as well as in the NC-stem (G34-C42) of the bound RNA duplex (assignment in blue letters in figure 3.19) are either severely broadened in the spectra or exhibit extensive

chemical shift perturbations. In the latter case the assignment need to be verified. The application of previously described ^{13}C-filtered NOESY experiment [170] failed to provide NMR data that could be used to assign the newly appearing resonances of the RNA duplex due to the extensive broadening of these resonances. In particular, residue- and sequence-specific assignments for these nucleotides could not be obtained because their intra- and inter-nucleotide NOE cross peaks were either absent or severely broadened in the NOESY spectrum of the RNA.

3.7 Ternary U4 snRNA-15.5K-hPrp31 complex

The aim of the project is to investigate the binding of hPrp31 to the U4 RNA in the context of the U4 snRNP. Hydroxyl radical foot-printing and mutational analyses [120] have revealed direct interactions between hPrp31 and the U4 5′-SL and have delineated RNA secondary structure requirements for hPrp31 binding. Structural studies revealed the binding interactions of the U4-5′ SL in the k-turn region while bound to 15.5K [98] and of the pentaloop region [71] while bound to 15.5K and hPrp31$^{78-333}$. Biochemical data predict further contacts between the C-stem of U4 5′ SL to the C-terminal domain of hPrp31, which have so far not been described in structural studies. NMR spectroscopic investigation of the interaction surface of hPrp31 to the C-stem of U4-Kt2 requires an RNase-free, pure hPrp31 comprising the whole C-terminal domain. Preliminary work performed in our laboratory has shown that the expression of hPrp31 yields poor quantities of purified protein. The protein has poor solubility (0.15 - 0.2 mM) and undergoes rapid degradation in solution. To overcome the expression difficulties a construct of the full-length wildtype sequence was purchased (Geneart, Germany), which was optimized for expression in *E.coli*. After the first purification step of a one-liter expression, protein could be obtained in the reasonable amount of 2-3 mg. However, degradation could be observed while mixing hPrp31 with the dimeric complex U4-15.5K. Mass data revealed a degradation of the flexible C-terminal domain from residue 499 up to 372. Nevertheless, the degraded protein retains the positively charged stretch from 351 to 365, which is expected to bind the C-stem of U4 5′ SL is still present. Further difficulties arised from observed aggregation of hPrp31 to multimeric complexes presumably arising form the N-terminal coiled-coil domain. Further experimental work is in progress to achieve hPrp31 constructs lacking the coiled-coil domain.

4 Discussion: The U4 snRNA 5′ stem loop

NMR analysis confirm the stable folding of the internal loop and the NC-stem
On the basis of previous MD simulations it had been suggested that the NC-stem is highly unstable and unfolds in the absence of protein binders as a consequence of the opening of both the non-canonical G-A and the Watson-Crick (WC) G-C base-pairs [22][23]. Because of these literature reports an investigation of the base-pair patterns of the NC-stems in the U4-Kt2 RNA in solution was needed. The performed NMR experiments clearly showed that both the G-C and the G-A base-pairs are present in the U4-Kt RNA in solution.
The observation of the imino protons G34 and G35 involved in the two WC base-pairs of the NC-stem (see figure 3.7C), indicated the presence of a folded NC-stem. Interestingly, the imino protons of G32 and G43 are visible as well, although they are not involved in hydrogen bonding. This fact points to a reduced solvent accessibility of these sites, as it has been observed previously for G-A base-pairs [57].
With the introduction of ^{15}N-labeling and hetero-nuclear correlation experiments the elucidation of base pairing and more complex hydrogen-bonding patterns in RNA, by NMR spectroscopy, became even more feasible and significant [35]. The so-called HNN correlation experiment directly involves base pairing nuclei by transferring magnetization between the nitrogen nuclei involved in hydrogen bonding, via the $^{2}J_{NN}$ couplings of the hydrogen bonds (see chapter 3.3.2). Accordingly the non-selective HNN-experiments of the U4-Kt2 revealed, and hence confirmed, all WC base-pairs of the C- and NC-stem (see figure 3.9). It should be noted that the resonances of the WC C-G base-pairs of the NC-stem are weaker than the resonances of the C-stem which is expected for terminal base-pairs or for base-pairs belonging to very short helical stems. These nuclei are more easily accessible to the solvent, since the opening rate of those base-pairs is slightly faster [133]. A study determined WC base-pair lifetimes of several RNA duplexes and revealed a range from 2 to 5 ms for base-pairs located on the second outermost position of a terminal region whereas lifetimes of base-pairs within the duplex region range from 26 to 55 ms [133]. However, a rough estimation of the H-bond mediated J_{NN} coupling [29], obtained by quantification of the

cross peaks in the HNN experiment, indicated scalar couplings of similar size for both the C- and the NC-stems, which confirms that the NC-stem is stably formed in solution. Furthermore the presence of the two sheared G-A base-pairs was verified in a selective HNN experiment (see figure 3.10), where the presence of two peaks correlating the H8 nuclei of A33 and A44 with the N2 nuclei of G43 and G32 confirmed that even the G-A base-pairs of the NC-stem are formed in the U4-Kt RNA in the absence of protein binders, in contrast to what proposed on the basis of MD simulations.
In addition analysis of the NOE spectra revealed a tight network of NOEs in the internal loop region (see figure 4.1) including both cross-strand and (i, i+1) sequential NOEs. This is a strong indication that the internal loop assumes a preferred, well-defined, major conformation even in the absence of proteins. However a certain degree of local flexibility is retained around the bulged-out A25 of the C-stem and around the U31 of the internal loop, as indicated by the averaged values of the $^{3}J_{H1'H2'}$ of U24, A25, G26, A30, U31, G32 and G43 (see figure 3.11), which are indicative of riboses inter-converting between the C3′-endo and the C2′-endo conformations. From all NMR data, it can be concluded that both the internal loop and the NC-stem of the U4-Kt RNA are well structured in solution.

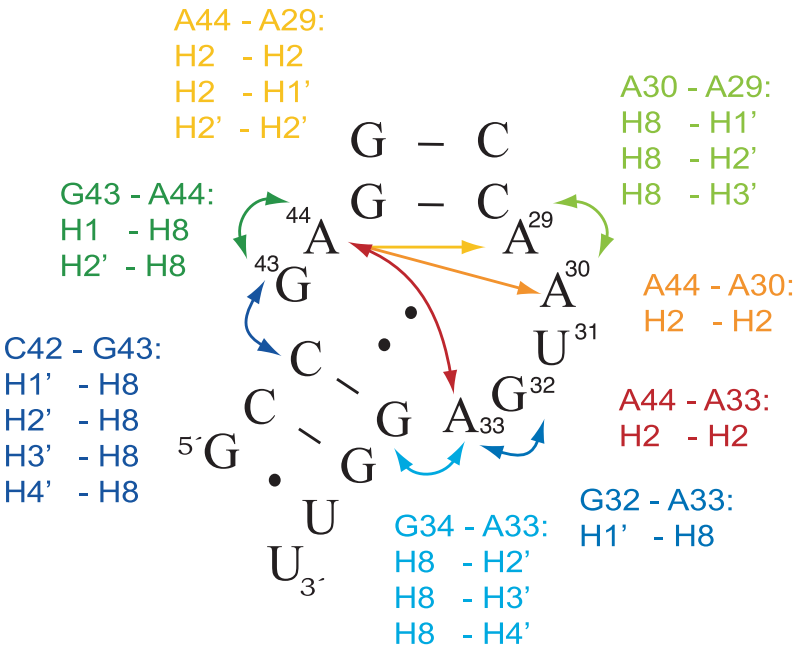


Figure 4.1: *Network of inter-nucleotide NOEs observed at the internal loop site* | The presence of several NOEs confirms that the internal loop is well structured in solution even in absence of proteins.

In the MD-study mentioned above [23], the unfolding of the NC-stem has been proposed to explain the higher accessibility to Kethoxal modification of G32, G34 and G35 with respect to the guanosines of the C-stem in the absence of 15.5K. However, in our opinion there is no need to invoke a complete unfolding of the NC-stem to explain the higher accessibility of those guanosines to Kethoxal modifications. G32 and G35 both belong to terminal base-pairs of the NC-stem and thus their N1 and N2 sites are easily accessible in the free RNA. While the G-C base-pairs of the NC-stem seem to be kinetically less stable than those in the C-stem, as indicated by the lower intensity of their resonances in the HNN spectrum (see figure 3.9), thermodynamically they are as stable as all other G-C base-pairs, as indicated by similar values of the hydrogen bond mediated $^{2}J_{NN}$ coupling constants. Binding of 15.5K results in protection of all three Gs from Kethoxal modification. For G32, this can be attributed to the direct contacts of the side-chain of E41 with the N1 and N2 sites of G32. For G34 and G35, protection from Kethoxal modification in the presence of 15.5K can be attributed to the extensive contacts of the side-chains of K44 and R48 with the phosphate backbone of C42 and C41. These contacts slow down the opening process of the G-C base-pairs of the NC-stem by fixing the helical backbone on the site of the cytosines, thus making Kethoxal modification inefficient.
Taken together all NMR data confirms a well-structured NC-stem beyond any doubt as well as a well-structured internal loop region of the U4-Kt RNA. In the next paragraph it will be discussed whether the tight structural organization of the U4-Kt RNA in solution reflects the typical, sharply bent conformation observed in complex with proteins.

Protein-assisted RNA folding and RNA dynamics enable the kinked conformation

In this work the conformation of the U4-Kt RNA in solution in the absence of protein binders and of cofactors was investigated with the goal of determining whether the k-turn consensus sequence codes for the k-turn structural element in the free RNA. Our combined NMR and SANS study revealed that the U4-Kt2 RNA prefers an extended conformation in solution and that the sharply bent conformation observed in complex with proteins is either not present or very poorly populated, as will be discussed in detail in the following paragraphs. Furthermore it was verified that addition of magnesium ions trigger a conformational change in the internal loop region (discussed in detail on page 95), which is in agreement with the notion that divalent cations induce formation of the kinked

conformation. All in all, the findings indicate that the consensus sequence of figure 1.9 does not code for the sharply kinked structures observed for k-turn RNAs in complex with proteins but rather for a dynamic behaviour of the RNA that allows accessing the kinked conformation in the presence of cofactors. The findings disprove the hypothesis that the k-turn sharply bent structure is coded by the RNA sequence only and favour the thesis of protein-assisted RNA folding. Our study shows that the consensus sequence of figure 1.9 does not per se code for the sharp bent in the RNA backbone, which is highly disfavoured in solution and needs to be stabilized by protein binding. One possible explanation for this is the high-density of negative phosphate backbone charges at the sharp kink, which requires to be compensated through the positively charged protein sidechains at the RNA binding site [152].
Despite the presence of one major populated conformation, the nucleotides close to the internal loop region, and in particular A30, U31, G32 and G43, show a certain degree of local conformational flexibility, as indicated by the rapid interconversion of their riboses between the C3′-endo and the C2′-endo conformations. The local flexibility at the internal loop site might be functionally relevant and is probably required for the protein-induced switch to the kinked conformation. It is tempting to postulate that the interplay between RNA dynamics and protein binding is a general mechanism by which RNA molecules modulate their function. Thus, rather than encoding for well-defined conformations, the four-bases code of the RNA sequence might in some cases determine specific dynamic properties at non-helical sites. Such internal flexibility offers an exquisite possibility of functional regulation in dependence of cofactors or protein binders. One fine example of this principle has been provided recently for the recognition of positively charged co-factors by the HIV-1 TAR-RNA [168]. Hence continuative investigations are currently performed in our laboratory to determine the dynamic behaviour of the internal loop nucleotides of the U4-Kt2 RNA by NMR relaxation data.

Possibility of the presence of a second conformation in the high μs to ms timescale

Both the NMR and SANS data indicated the presence of a major extended conformation for the unbound U4-Kt2 in solution under the conditions of the NMR experiments ([Mg2+] = 0 mM). However, the question arises whether the sharply bent k-turn structure, in spite of not being the preferred conformation, could be populated to a considerable extent, as proposed on the basis of FRET experi-

ments [40]. In this study, k-turn RNAs have been described to be dimorphic with a distribution of populations between the extended and bent conformations of 65:35 in absence of divalent cations. However, as mentioned previously the absence of a second set of NMR lines in the C-H correlation spectra, as well as the absence of conspicuous line-broadening excludes the presence of this considerable high amount of bent conformation in slow to intermediate exchange (high μs < k_{ex-1} < s) with the extended structure [48]. The presence of the bent conformation would only be undetectable if the population of the bent conformation is much less than that of the extended one. Thus, if the U4-Kt RNA is undergoing a conformational exchange between the extended and the bent conformation in the μs to ms time scale, the population distribution must be highly skewed in favour of the extended conformation.
The discrepancy between this and the FRET study can be explained by the much higher length of the RNAs used in the FRET study, which might induce stabilization of the bent conformation through long-range tertiary interactions. Indeed, the sharply bent conformation of the k-turn motif is observed also in the absence of direct protein binders for one of the six k-turns found in the 23S rRNA of *Haloarcula marismortui* [63], suggesting that long range tertiary interactions in the context of large RNP complexes may stabilize the kinked conformation.
An alternative explanation for the discrepancy in the population distribution observed is that the populations of the extended and the bent conformation depend on the exact RNA sequence, as observed previously [83]. In this case the consensus sequence of figure 1.9 would not code for the k-turn structural motif but rather for a dynamic behaviour of the RNA that allows accessing the k-turn motif in presence of cofactors and whose precise dynamics parameters are finely regulated by the specific sequence.

Possibility of the presence of a second conformation considering faster motions

A further possibility, that would escape detection by simple line-width analysis, is represented by an exchange process occurring on the low μs to ns time-scale. In this case, if a considerable amount of bent conformation is present in solution, its occurrence should be detectable by analysis of NOE data. Accordingly NOE connections (NOEs) were predicted for the sharply kinked conformation that are not expected for the extended conformation (see figure 4.2). For example, for the kinked conformation NOEs are expected between G45 and A33, typical of the A-minor interaction, and between A30 and G32 as well as between A44 and A33 in the internal loop (see figure 4.2). Such NOEs are not predicted from the

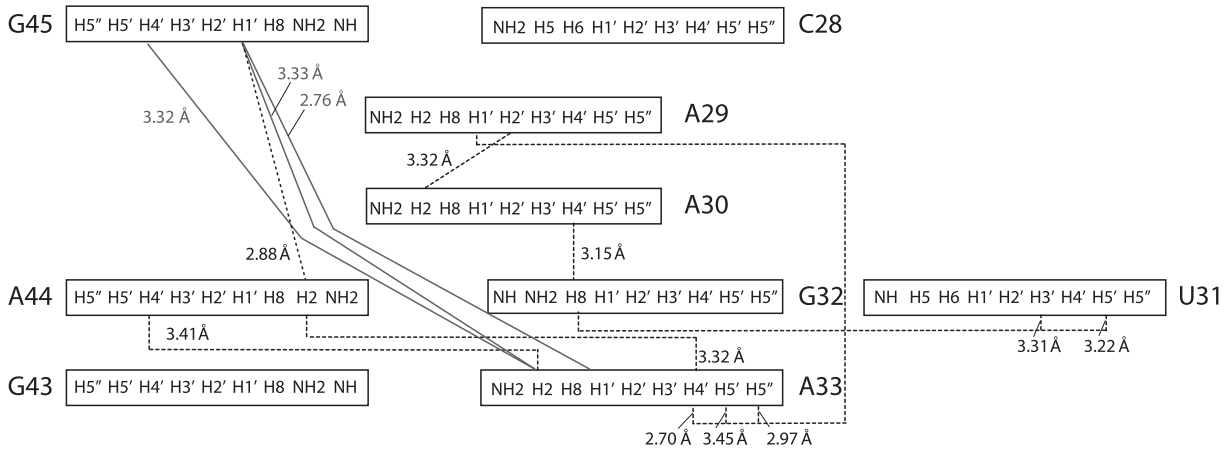


Figure 4.2: *NOE connections expected from the protein-bound conformation of the U4 RNA but not from the unbound extended conformation.* | All these NOEs are not observed in the NOESY spectra of the U4-Kt RNA in solution. Only predicted NOE connections involving either a H1′ or a base proton at a distance < 3.5 Å are depicted. Full and dashed lines indicate the connections expected for the A-minor interaction in the bent conformation and for the internal loop region, respectively.

extended conformation and are indeed not observed in the NOESY spectra in solution (see figure 3.8). In the time regime of low μs to ns time-scale the rate of inter-conversion of a molecule, e.g. between the kinked and the extended conformation, is higher than the overall correlation time of the molecule which in turn depends on the size of the molecule. For the U4-Kt2 the correlation time is ~ 5 ns. In this case NOEs stemming from the kinked conformation would scale linearly with the the mimimum population, and it is possible to estimate the maximum population of the kinked conformation that is compatible with the absence of the characteristic NOEs. Considering a distance of 3 Å between two interacting protons H_A and H_B in the kinked conformation (as for example A44-H2 and A33-H1′ or G45-H1′ and A33-H1′) and assuming that NOEs can be detected up to a distance of 5 Å, the H_A-H_B NOE stemming from the kinked conformation would be visible for a population larger than $\sim 5\%$ [142]. Consequently, if the sharply bent conformation exists in equilibrium with the extended conformation, the former is only transiently occupied, in contrast to the value of 35% proposed on the basis of FRET experiments [40].

At this point it is worth noticing that the population distribution observed for the U4-Kt RNA might not be a universal feature of all k-turns RNA, but might depend on the RNA sequence [83]. In fact, a previous NMR study on the helix-loop-helix portion of the L30 mRNA in the absence of magnesium ions reports on a flexible internal loop (k-turn) with resonances broadened beyond detection

and very few NOEs [80]. This might indicate a higher level of population of the kinked conformation for the k-turn region of the L30 mRNA as compared to the U4-Kt RNA.
Finally an other scenario could be, that the rate of inter-conversion, between the kinked and the extended conformations, were much faster than the inverse of the overall correlation time (k_{ex-1} in the ps range). In this case the NOEs stemming from the kinked conformation would scale quadratically with the population and thus this exchange regime would allow for an increase in the maximum population of the kinked conformation before the characteristic NOE connections become visible up to 22 % [142]. However, as the inter-conversion between the kinked and extended conformation requires the reorientation of helical elements, this process cannot occur on the picoseconds time-scale. The time constant for the motion of helical segments around a hinge region must be at least equal to (or higher than) the correlation time of the helical elements in solution (ns), as it has been observed previously for inter-domain motions in the TAR-RNA [168]. Indeed, a conformational exchange in the μs to ms time-scale is observed for the U4-Kt RNA in the presence of magnesium ions.

Dependence of the RNA conformation on the presence of divalent cations
Upon titration of magnesium ions a severe broadening of the resonances belonging to the bases of the internal loop nucleotides A29, A30, U31, G32 and A33, was observed (see figure 3.17). This suggests the presence of a conformational exchange process occurring on the μs to ms time scale. Unfortunately, the broadening of the NMR lines beyond detection does not permit a detailed structural investigation of the U4-Kt RNA in the presence of magnesium ions. However the observation of an exchange process upon addition of magnesium ions is in agreement with other studies [83][40] which described a stabilization of a second conformation of the U4-Kt RNA with divalent cations. In the previously mentioned FRET experiments [40] the influence of magnesium ions on the population of the kinked conformation were performed and a population distribution of 65:35 between the extended and the kinked conformations has been found in the absence of divalent cations, while the presence of magnesium reverts the distribution to 30:70. The trend observed in the FRET experiments perfectly agrees with our NMR data, even though in the absence of divalent cations the NMR data suggest a more shifted population distribution to the extended conformation.
It can be speculated that the process caused by magnesium addition represents the inter-conversion between the kinked and extended conformations, which becomes visible only after the stabilization of the kinked conformation in the

presence of divalent cations. This would explain all observed NMR data and would confirm the dependence of the population distribution on the concentration of magnesium ions. In a cellular environment, magnesium ions might partially pre-fold the k-turn RNAs prior to protein binding by stabilizing the kinked conformation.

Interdisciplinarity of SANS and NMR data

The combined use of SANS (or Small Angle X-ray Scattering, SAXS) and NMR data is gaining momentum in the study of multi-domain proteins and protein complexes [36][44], where SANS and SAXS data provide an NMR independent measurement of the relative position of protein domains or of globular proteins in a complex. Furthermore SANS and SAXS data are easily accessible and represent a precious addition to the long-range structural information obtained by NMR through the analysis of residual dipolar couplings. The SANS data of the U4-Kt2 in solution completely support the NMR analysis and independently confirm that the sharply bent k-turn motif is not formed in solution in absence of proteins.
In general for structural study of RNA molecules, which inherently consist of several helical domains and whose structure determination by NMR suffers from the paucity of long-range NOE restraints, the combination of SANS or SAXS with NMR data is likely to be even more relevant than for proteins. A recently published pilot study in this direction [45] and the work presented here demonstrate that small angle scattering data together with NMR RDC data allow determining the relative orientation of RNA helices in complex RNA molecules, therefore facilitating the difficult structural investigation of isolated RNAs.

Features of the U4 5′ SL-15.5K and U4 5′ SL-15.5K-hPrp3 complexes

Earlier structural studies already characterized the k-turn motif of the U4 snRNA 5′ SL when bound to the 15.5K protein [152] and when subsequently bound to the hPrp31 protein forming a ternary U4-15.5K-hPrp31 complex [71]. Since only the N-terminal and NOP-domain of the hPrp31 protein was used for the ternary complex, it still needs to be investigated whether the C-terminal domain of hPrp31 and the canonical stem (C-stem) of the U4 5′ SL also establish interactions in the complete ternary complex. In order to extend the knowledge about the role of U4 snRNA in spliceosomal assembly, NMR studies of the dimeric and ternary complex containing the full length hPrp31 protein were designed in this work.
However, NMR investigations of the dimeric U4-15.5K complex containing ^{13}C,^{15}N-labeled U4-Kt2 construct revealed that protein binding induced the RNA

dynamics at the binding site in the intermediate time regime. Thus an assignment of nucleotides of the U4 5′ SL in its 15.5K bound form was not possible for the NC stem and the loop region. It should be noted, that a different behavior was observed in an earlier NMR study of a similar RNA-protein complex of yeast consisting of L30 protein and mRNA, which also adopts a kink-turn motif [80]. In this study, the free RNA exhibited broadening of signals within the internal loop region, while upon protein binding the signal broadening was reduced enabling RNA assignment. Consequently, dynamic properties of the k-turn motif seems to be a unique feature of each RNA molecule that comprises this RNA motif.
Although NMR analysis was not feasible for the dimeric U4-15.5K complex, NMR transfer saturation experiments were intended for the ternary complex in order to reveal nucleotides of the C-stem of the U4 snRNA that are involved in binding the hPrp31 C-terminal domain. Thus, a comprehensive picture would have been established describing the binding surface of the U4 snRNA in the ternary complex. During a saturation experiment, magnetization is transferred from one binding partner to the other one, thus revealing those nuclei that are in close proximity to each other. Unfortunately these studies were prevented so far due to the aggregation of the expressed wildtype hPrp31 protein prior to the ternary complex formation. Most likely the aggregation is caused by the N-terminal coiled-coiled domain of the hPrp31 protein. Furthermore a degradation of the C-terminal domain has been observed while purifying the hPrp31 protein. Thus, further effort is needed in the future to gain a comprehensive picture of the U4 5′ SL binding properties.

5 Results: Structural studies of the lariat-forming ribozyme

Due to the lack of secondary structure data, its size and dynamic behavior the lariat-forming ribozyme presents a challenging RNA construct for NMR spectroscopic investigations. To achieve secondary structure information for the ribozyme construct extensive mutational studies were performed, which are described in the following paragraph.
The strategy to achieve the three-dimensional picture by NMR spectroscopy requests the resonance assignment of the RNA by identifying which 'spin-system' belongs to which nucleotide in the chain. Resonance assignment of RNA is greatly facilitated if the molecule is isotopically labeled, which is nowadays easily done by using commercially available ^{13}C,^{15}N-labeled or deuterated NTPs. The first help to record several multidimensional NMR experiments, which are designed to take advantage of 1-bond (^{1}J), 2-bonds (^{2}J) or 3-bond (^{3}J) coupling constants, in order to correlate the resonances within a nucleotide (see chapter 2.5.1), whereas the use of partially deuterated NTPs diminishes the proton caused spectral overlap and improves the relaxation properties of the molecule. Therefore both, RNA samples containing deuterated or ^{13}C,^{15}N-labeled nucleotides, were prepared and the results are summarized in the following paragraph.

5.1 Effect of RNA mutations upon lariat formation

RNA molecules fold into characteristic secondary and tertiary structures that account for their diverse functional activities. The basis for analyzing RNA three-dimensional structures requires characterization of its secondary structure which is mainly defined by base pairing patterns. Several tools had been developed for *in silico* prediction of RNA secondary structure such as mfold [169], RNAfold [52] and RNAalifold [53]. However, these tools do not adequately capture the various non-canonical pairings or they rely on RNA alignment protocols. Alignment protocols are less useful for RNA since RNA base pairing is more conserved than RNA sequence.

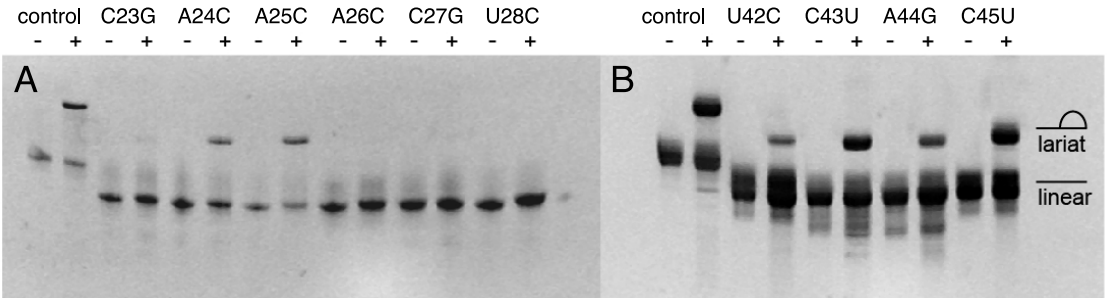


Figure 5.1: *Mutational studies to observe effects on lariat formation.* Gels exemplarily show mutations in the sequence range of nucleotides 23-28 (A) and 42-45 (B) of the lariat RNA which were prepared by *in vitro* transcription. Lariat formation is induced by addition of magnesium ions over night. A comparison of lariat formation of the original ribozyme sequence (*lane 2*) with that of a control omitting magnesium treatment (*lane 1*) shows that the lariat form of the ribozyme is retarded by moving into the polyacrylamide gel. *Lane* 3 and 4 show the lariat formation experiment for mutation C23G, where the mutation caused a complete blocking of the transesterfication reaction. The transcribed control RNAs shown in line 1 and 2 in both gels were performed with a 6 nt longer DNA template due to hammerhead overhang at the 3′ end. Lanes labeled with (-) show linear RNA constructs (no Mg^{2+}) and (+) the lariat form (with Mg^{2+}).

To investigate conserved residues for lariat formation and the secondary structure of the lariat-forming ribozyme, 60 nucleotide substitutions, 3 nucleotide deletions (see figure 5.2b) and 7 double mutations (see table 5.1) were introduced into the original RNA sequence and their effects on lariat formation had been observed. Thereby, reduced or missing bands of the lariat form in the gel shift assay (see chapter 2.2) is an indicator for changes in base-pair formation, either by opening of a base-pair or formation of a new base-pair.

Gelshift results for ten important mutations are shown in figure 5.1. The effect of all performed mutations upon lariat formation is plotted in figure 5.2b as well in table 5.1 for double nucleotide mutations and nucleotide deletions. Gel A shows that for mutations C23G, A26C and C27G the transesterfication reaction leading to the lariat form, is completely blocked, while mutations A24C and A25C only slightly affect lariat formation. Blocking of lariat formation when substituting residue 23 is in agreement with the results of an earlier study by Tuschl *et al.* [143] (see figure 5.2a). This study also found C23 to be covariant with G5, which could be verified by a double mutation (table 5.1). The strong effect of mutations at positions 23, 24 and 26-29 on lariat formation indicates involvement of these residues in base pairing. The combination of NMR derived distance constraints and the results derived from mutational studies allowed us to propose a helical structure involving G1 to C12 and G17 to C27 (figure 5.2b).

Table 5.1: *Nucleotide exchange of double and deletion mutations* | ++ = no inhibition, - = blocking of lariat formation

mutation 1	mutation 2	effect on lariat formation
G1A	C27U	++
G2C	C27G	-
G5C	C23G	++
A6G	U22C	++
U28C	A44G	-
U28G	G49C	-
A40G	U56C	++
Δ6-8	Δ21-22	-
Δ6-9	Δ20-22	-

The proposed helix features several base pairings unknown in the previous secondary structure model of Tuschl (figure 5.2a), namely two canonical and five non-canonical base-pairs. The latter also explain the limitation of standard tools in predicting the helix moiety of the lariat-forming ribozyme.
To investigate the required length of the new determined helix 1-27 for lariat formation, we analyzed sequence variants with deletion of Δ6-8 and Δ21-22 as well as Δ6-9 and Δ20-22. For both deletions the helices were too short to lead to lariat formation. However, it can not be ruled out that the lack of the specific base-pairs might inhibit lariat formation rather than the length of the helix. Furthermore, it was shown by a deletion mutation at position 9, that the nucleotide G9 is not necessary for ribozyme activity. which was confirmed by a deletion that did not effected the lariat formation.
The ACAGAGA-box is a phylogenetically conserved and functionally important pattern of the U6 snRNA. Due to the importance in the spliceosome each nucleotide of the ACAGAGA-motif of the studied lariat-forming ribozyme was substituted with all the other nucleotides (21 single-nucleotide substitutions, figure 5.2b). The most critical residues for lariat formation within the conserved pattern were A29, G32, and A33; except for A to G substitution, all mutations at the three sites lead to reduced or blocked lariat formation. At the remaining positions, only G34C and A35C mutations negatively effected lariat formation.
Nucleotide exchange studies downstream the ACAGAGA box showed that mutations at sites 42 to 45 (gel B in figure 5.1) had only slight (C45A, U42C) or no effect upon lariat formation, while the two nucleotides directly preceding the branch-point A48 negatively effect lariat formation in case of mutation.

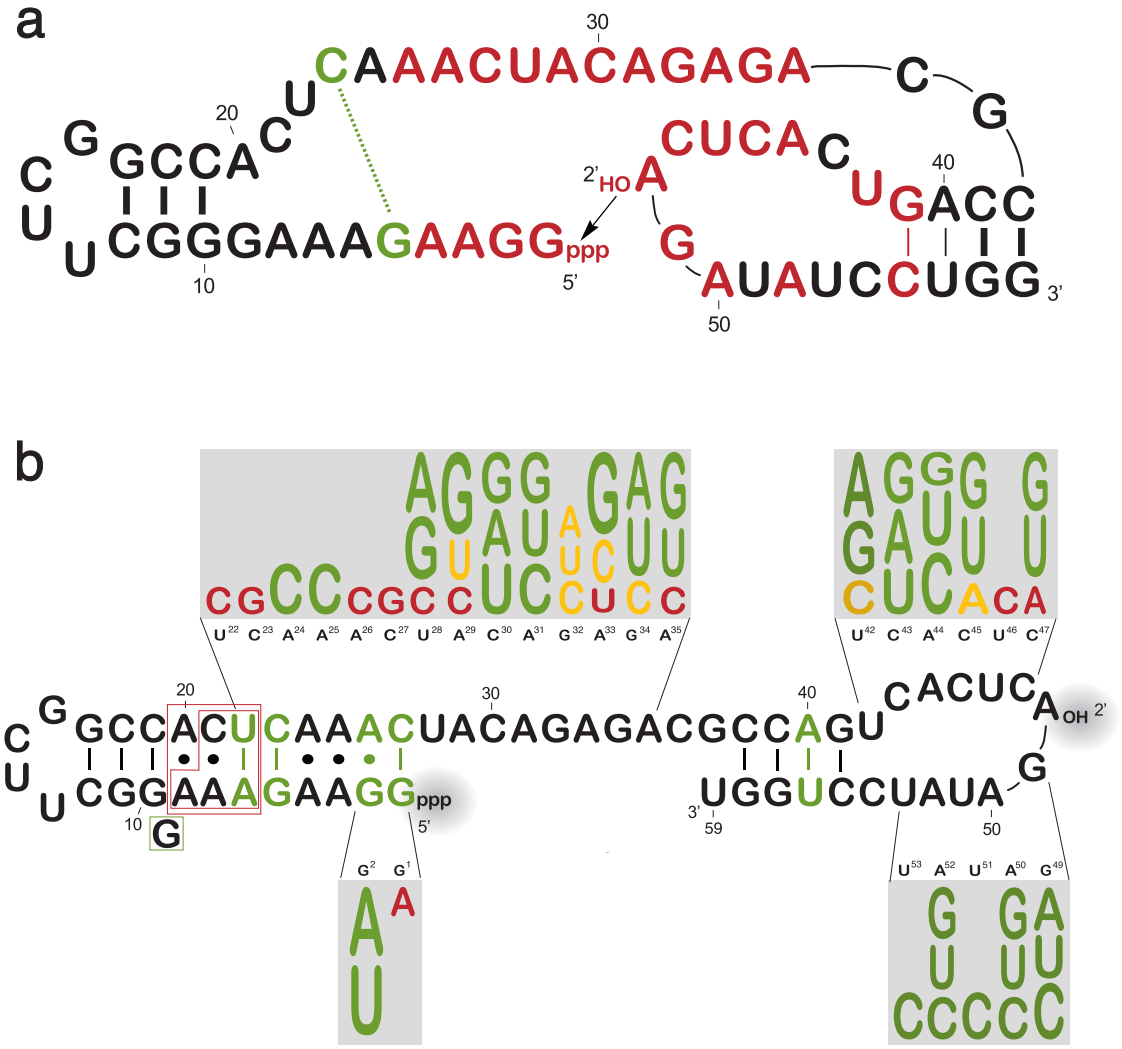


Figure 5.2: *Secondary structure models of the lariat-forming ribozyme* | **(a)** Previous secondary structure proposal of the lariat-forming ribozyme by Tuschl *et al.* [143]. Invariant residues are indicated in red, and the co-varying residues are in green. base-pairs are symbolized by vertical lines and bold lines show when covariation or additional experiments support base pairing. **(b)** Refined secondary structure model based on series of mutations (gray boxes) and NMR derived distance restraints. Colors denote the effect of the mutation upon lariat formation. Red letters indicate mutations leading to complete inhibition of the trans-esterfication reaction. It can be assumed that these mutation causes a change in tertiary structure disabling the branch-point A48 to attack the triphosphate group at the 5′ end (gray circles). Mutations labeled in yellow show an reduced lariat formation, whereas mutations labeled in green seem to have no effect or even an increase in the lariat formation, which is reflected by the size of the letters. Deletions of nucleotides are encoded by green or red boxes depending on the effect upon lariat formation.

5.2 Spectral assignments of the ribozyme in the linear form

The first step in structural analysis of RNA and any other macromolecular structure by NMR methods is the assignment of each observed resonance to the nucleus from which it arises. The lariat-forming ribozyme presents a challenging molecule for NMR studies due to the lack of secondary structure data, its size of 59 nucleotides and its dynamic behavior, especially in 3'-terminal region (see below 5.2.2). However, the resonance assignment could be accomplished by selective labeling of specific nucleic acids (see table 2.2b,c) and NMR sample preparation of certain mutated RNA constructs (see table 2.2d,e). The assignment was accompanied by extensive mutagenesis studies (see chapter 5.1) leading to the secondary structure model.
The assignment procedure of isotope-labeled RNA is already described in detail for the U4 snRNA construct (see chapter 3.2). The following chapter will focus solely on special features of structural studies of the lariat-forming ribozyme. At this point it should be pointed out, that the complete assignment of the ribozyme was a cooperative work performed in our laboratory. In particular, my personal contribution to the assignment work was the intra-nucleotide assignments of the differentially ^{13}C, ^{15}N-labeled samples G, C, AU, A and U (see table 2.2b) whereas the inter-residue assignment was performed in our laboratory.

5.2.1 NMR sample preparation of the lariat-forming ribozyme

The approach used to characterize the structure of the lariat-forming ribozyme is based on selective labeling of specific nucleic acids accompanied by mutagenesis. Therefore several samples had been prepared (see chapter 2.4.3 and table 2.2). The transcription of all RNA constructs had been successfully accomplished. In table 5.2 the transcription results of selected RNA ribozyme constructs are listed.
To overcome the 3' heterogeneity the hammerhead RNA, HH-spliclin, was used to remove the hammerhead overhangs from the original RNA constructs of the 59-nucleotide long lariat-forming ribozyme (see table 2.2b,c). The hammerhead cleavage was not performed for the samples with mutations, thus, their DNA templates had been designed without the hammerhead overhang.

Table 5.2: Parameters used in large scale *in vitro* transcription reactions

	isotope label	[NTP] in mM	[$MgCl_2$] in mM	[template]	T7 RNAP (v/v)
HH-spliclin	-	20	10	2.1 μg/ml*	1/4
spliclin	A,G,C,U	17	10	3.5 μg/ml*	1/5
spliclin29	-	20	20	1.2 μM	1/5
spliclin29	A,G,C,U	20	20	1.2 μM	1/5
spliclinG-2A	G	30	40	0.2 μM	1/5
spliclinA-33G	A	20	30	0.35 μM	3/20
spliclinA-40G-56C	A	20	30	0.2 μM	1/5
spliclinG-49C	G	20	30	0.2 μM	3/20
spliclinA-50C	A	30	20	0.2 μM	1/5
spliclinA-52C	A	30	20	0.2 μM	1/5

spliclinA = linear form of the spliceosome related ribozyme with ^{13}C, ^{15}N- labeled adenosines and
spliclinG = linear form of the spliceosome related ribozyme with ^{13}C, ^{15}N- labeled guanosines.
* plasmid DNA was used as template, whereas all other transcriptions were performed with DNA oligomers.

5.2.2 Conformational exchange of the lariat-forming ribozyme

An unusual structural feature identified form NOESY spectra is the existence of a second conformation of the linear ribozyme. The conformational exchange has been observed for residues ranging from C30 to the 3'end of the ribozyme by detecting exchange signals next to the diagonal signals. Instead, the 29 residues at the 5' end, which form the helical region, do not show this type of motion. When distinct new resonances are observed in an NMR spectra, as seen for the 3' region, the molecule is in slow exchange between two or more available conformations on the NMR time scale. For the 3' region the average residence time in one conformation can be estimated to more than a millisecond. Since for the 5' end terminal region only one resonance line is visible, this moiety seems to exists only in one conformation.
Exemplarily for the observed exchange in conformation, signals for A40 and A48 are depicted in figure 5.3. Dynamic behavior of the ribozyme was observed in the tested temperature range from 5°C to 25°C as well as in the pH-range from 6.0 to 7.4. Although two resonances are observed for the two conformations, these are distinguishable from each other. Only one conformation gives rise to detectable sequential NOE signals, whereas the second conformation solely shows intra-nucleotide NOE contacts.

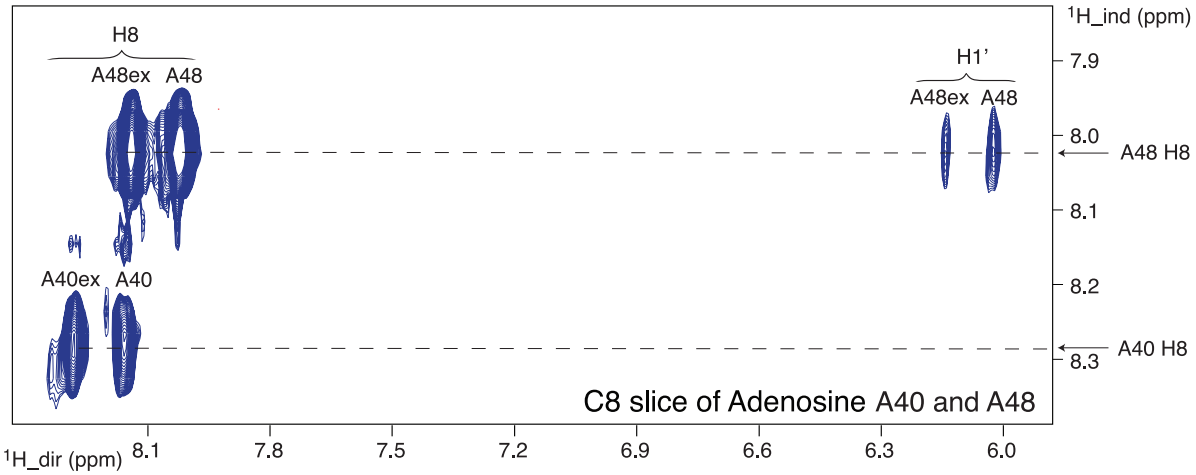


Figure 5.3: *Conformational exchange of the lariat-forming ribozyme.* | The depicted spectral region of the 2D slice of the ^{13}C-filtered ^{13}C-edited NOESY spectrum shows the diagonal signals of adenosines 40 and 48 and the cross-peaks of the exchange conformation of these nucleotides.

5.2.3 Assigning the spin systems of the lariat-forming ribozyme within specific residues

To accomplish the resonance assignment of the large and dynamic ribozyme (59-mer) several NMR spectra were recorded (for detailed explanation see chapter 2.5) on differentially labeled ^{13}C,^{15}N-labeled RNA samples (see table 2.2). The following NMR experiments were measured:

2D ^{13}C-edited HSQCs [4]

2D ^{15}N-edited HSQCs [4]

3D H_sCN_b [33] and 3D H_bCN_b [88]

3D HCCH-COSY-TOCSY [54]

2D HCCH-COSY [38]

3D ^{13}C -edited NOESY [138]

3D TROSY-relayed HCCH-COSY [126]

3D ^{13}C-filtered ^{13}C-edited NOESY [170]

3D ^{12}C-filtered ^{13}C-edited NOESY [170]

3D ^{13}C-filtered ^{13}C-edited base-selective NOESY [170] [9]

2D HNN-COSY [51]

Identification of base and sugar protons

Figure 5.4 shows the 'fingerprint' regions for the H1'C1' and H6C6/H8C8 correlations for the linear ribozyme recorded with the ^{13}C-edited HSQC experiment (see chapter 2.5.2). To overcome the resonance overlap that occurs in a completely

labeled ^{13}C, ^{15}N-sample (data not shown), we used residue-type-specific labeled samples, as depicted in figure 5.4. Residue-type-specific labeling is a powerful method enabling direct assignment of a residue to its specific nucleotide type and thereby reducing possible misassignments. However, although resonance overlap could be decreased by differential labeling, especially in the central part of the H1'C1' region, there are still severe overlapping regions (see 5.4A). For the uridine-labeled sample (red color) up to 16 resonance peaks are expected, since the sequence contains ten uridine residues, of which six are in conformational exchange. However, in the H1'C1' region twelve resonances are clearly distinguishable and in the H6C6 region only four resonances are nicely dispersed whereas all further uridine H6C6 resonate in the region of $\sim$ 7.5 to 7.7 ppm in the proton dimension and $\sim$ 139 to 140 ppm in the carbon dimension. The residual nucleotides resonate at similar frequencies and are observed as overlapping resonances. The same is true for the cytosine-, adenosine- and guanosine- specific labeled sample, for which resonances can also be partly distinguished in the ^{13}CH-correlation spectra.

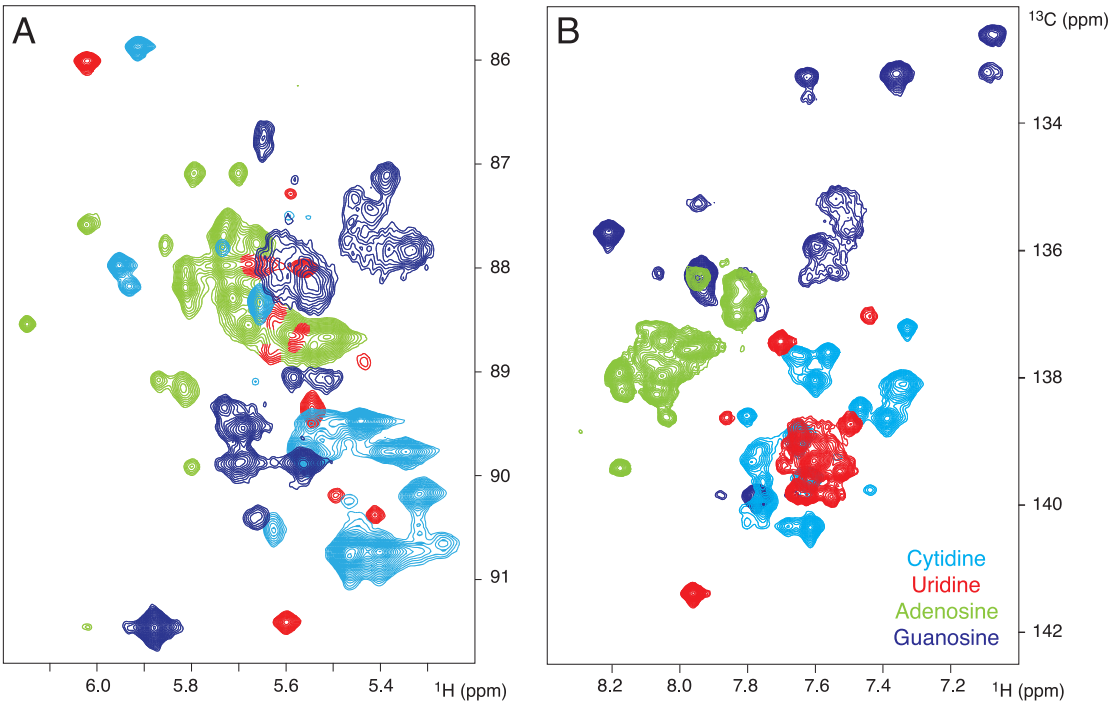


Figure 5.4: *Overlay of ^{13}C-HSQC spectra of the lariat-forming ribozyme (linear form)* | (A) H1'C1' ribose resonances and (B) H6C6 (Py) / H8C8 (Pu) base resonances. Peaks of the sample with ^{13}C, ^{15}N-labeled cytosines are given in light blue. The peaks of samples with labeled uridines, adenosines and guanosines are color coded in red, green and blue, respectively.

Correlation of base to sugar protons

The assignments of H1' to H6 (Py)/H8 (Pu) were carried out by using HCN spectra (see chapter 2.5.3) and ^{13}C-filtered ^{13}C-edited NOESY spectra. With these spectra 55 of 59 sugar-base correlations (93 %) were assigned with 5% ambiguity. For the chemical shifts of the N1(Py)/ N9(Pu) nuclei, 48 of 59 nitrogens (81 %) were assigned from the HCN-spectra with 14% ambiguity. Finally all sugar-base correlations could be accomplished using sequential NOE data (see chapter 2.5.6).

Assigning the sugar and base spin system

The next step was the correlation of H1' protons with the remaining sugar protons (H2', H3', H4', H5', H5''). Sugar proton assignments were based on 3D HCCH-COSY-TOCSY spectra (5.4 ms, 10.8 ms and 16.2 ms mixing time, see chapter 2.5.4), and confirmed by ^{13}C-filtered ^{13}C-edited NOESY spectra. The system size and the observed conformational exchange of the ribozyme (see chapter 5.2.2) resulted in extensive overlaps and assignment difficulties. Thus 82 % of the sugar spin system could be assigned, 90 % of the more rigid moiety (nucleotide 1 to 29) and 74 % of the dynamic 3' end part of the molecule (residue 30 to 59).

For the bases almost complete assignments were accomplished. For the H5 base protons of the pyrimidines, 25 of 26 protons could be correlated to the respective H6 proton. The adenosine H2 protons were assigned to 55 % (10 of 18 protons) by the use of ^{13}C-filtered ^{13}C-edited NOESY spectrum.

Analysis of resonances by deuterium labeling of the ribozyme

A major problem in assigning the lariat-forming ribozyme by NMR was the severe overlap of the sugar proton resonances [163], which occur in a narrow range of around 1.0 ppm. In order to tackle this difficulty, NMR samples were prepared where deuterium is incorporated in the sugar moiety. The deuterium isotope of hydrogen has a different magnetic moment and spin compared to protons (^{1}H) and thus is invisible in a spectrometer tuned to protons. Labeling with deuterium reduces the spectral crowding by suppressing the specific NMR resonances that are exchanged from ^{1}H to ^{2}H. Further features of the applied deuterium approach is the reduction of NOE spin diffusion as well as the reduction of line-broadening that is associated with ^{1}H dipolar relaxation.

A deuterated NMR sample was prepared by *in vitro* transcription (see table 2.2c and figure 5.5), in which the commercially available, partially-deuterated nucleotides (^{2}H- H3',H4', H5', H5'') were incorporated.

However, the used nucleotides are not simultaneously ^{13}C-labeled, thus, enabling only homo-nuclear NMR experiments with the remaining ^{1}H- nuclei, namely 2D-^{1}H, ^{1}H-NOESY and an 3D-^{1}H,^{1}H,^{1}H-TOCSY-NOESY experiments. These two

experiments verified the readily known sugar-base correlations between H1′, H2′, H6 and H8 protons but new assignments could not be accomplished, which is due to the missing dispersion of resonances in an additional hetero-nuclear dimension (^{13}C or ^{15}N). In figure 5.5 the spectral region of the H2′-H8 correlations (left) and the H2′-H1′ correlations (right) are displayed for the 2D-^{1}H, ^{1}H-NOESY spectrum. In a non-deuterated sample these regions would be overlapped with NOE cross peaks arising form the sugar protons H2′ to H5″, since all these protons resonate around 4-5 ppm (vertical axis in the figure 5.5). Due to the deuteration of all sugar protons, apart from H1′ and H2′, all cross peaks peaks in the region of 4 to 5 ppm arise from H2′ protons. However, the system is too large to overcome the crowding within the H2′ spectral region. Also the introduction of a third dimension by introducing a TOCSY step as performed by recording the 3D-^{1}H,^{1}H,^{1}H-TOCSY-NOESY (data not shown) only partially improved the visualization for certain nucleotides.

It can be concluded that the reduction of observed proton resonances by deuteration is a reasonable strategy for smaller RNA systems, where no further ^{13}C or ^{15}N-labeling is required. For larger RNA systems, like the lariat-formimg ribozyme, simultaneously deuteration and ^{13}C or ^{15}N-labeling is indispensable in order to facilitate RNA assignment. However, such nucleotides are not yet available commercially.

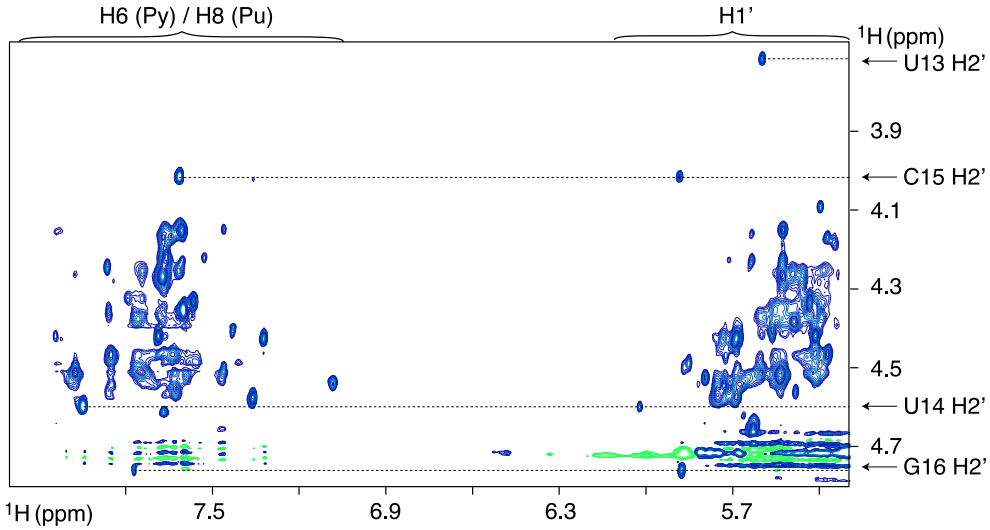


Figure 5.5: *2D-NOESY spectrum of a deuterated ribozyme sample (H3′, H4′, H5′, H5″-^{2}H)* | The spectral region of H2′-H6/H8 and H2′-H1′ cross peaks is shown. Assignment of the characteristic resonances of the UUCG tetra-loop are depicted by dotted lines. Cross-peaks within the H6(Py) /H8(Pu) region (left) and within the H1′ region (right) show severe overlaps due to the missing resolution from hetero-nuclear labeling.

5.2.4 Sequential assignment and introduction of selected mutations

Sequential resonance assignment requires the analysis of proton-proton NOE correlations mainly between the sugar protons H1′ and H2′ and the aromatic base protons H6 and H8 of neighboring nucleotides (see chapter 2.5.6). Due to the size of the lariat-forming ribozyme, RNA samples were prepared with specific ^{13}C, ^{15}N labeling for each nucleotide type. Connectivities between neighboring nucleotides were accomplished by analyzing ^{12}C-filtered ^{13}C-edited NOESY spectra (performed in our laboratory, see figure 5.6 and 5.8).
For neighboring nucleotides with the same base moiety (e.g the G16-G17) sequential connections were determined from ^{13}C-filtered ^{13}C-edited NOESY spectra of the specifically labeled sample (e.g. guanosine-labeled sample), which was necessary for 13 out of 58 sequential connectivities. However, since nucleotide-type-specific spectra comprise intra- and inter-nucleotide NOEs, only few sequential connections could be determined for neighboring nucleotides of the same base type. Underrepresentation of sequential NOE connections between nucleotides of the same base-type can be seen in the overall NOE network (see figure 5.8).

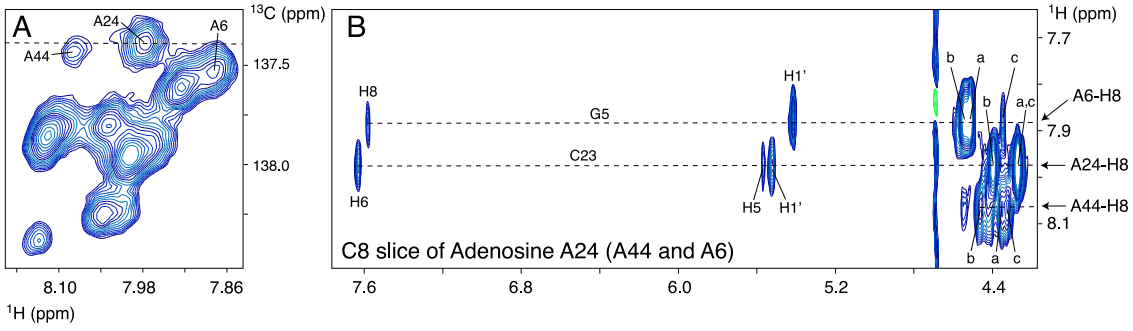


Figure 5.6: *Sequential assignment of the lariat-forming ribozyme* | (A) Spectrum of the ^{13}C-edited HSQC to visualize the H8C8 resonances of A6, A24 and A44 (B) 2D slice of a ^{12}C-filtered ^{13}C-edited NOESY spectrum recorded on the adenosine-specific labeled sample of the ribozyme. The carbon dimension correspond to the chemical shift of the C8 nuclei of adenosine 24, which also enables to see the nearby C8 nuclei of adenosine A44 and A6. The inter-residue cross peak between H8 and either H6/H8 or H5 or H1′ or H2′ (a) or H3′ (b) or H4′ (c) are connected by lines, with the numbering of the residues being indicated in the label.

Sequential assignment by selected mutations of the lariat-forming ribozyme

The sequential assignment could not be accomplished with full confidence by the strategy described above, especially in the dynamic moiety of nucleotide 30 to 59. In order to resolve assignment ambiguities eight, catalytically-active point-mutants, G2A, A33G, A40G/U56C, G49C, A50C, U51C, A52U, U53C were prepared (see table 2.2d). Nucleotide substitution leads to disappearance of a specific resonance peak in the spectra of the corresponding selectively labeled RNA and, thus, enables to distinguish the resonance peak in the spectrum of the original RNA construct.

However, the substitution of one base of the lariat-formig ribozyme resulted not only in disappearance of the resonance signal of the mutated residue, but

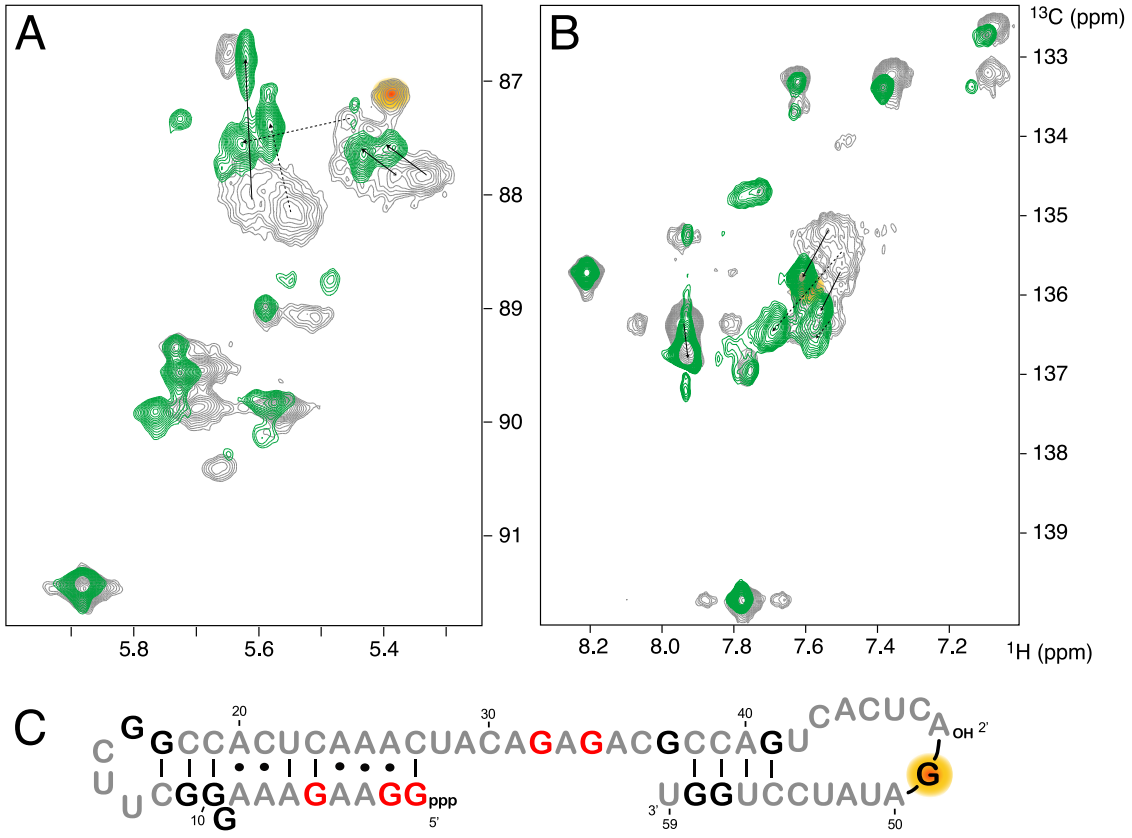


Figure 5.7: *Overlay of H1'C1' (A) and H8C8 (B) correlations of the original ribozyme (gray) and the G49C mutant (green).* | ^{13}C-HSQC spectra were recorded based upon ^{13}C,^{15}N guanosine-specific labeled RNA sample. The substituted residue 49 (yellow circle) could be unambiguously identified in the guanosine-specific labeled ^{13}C-filtered ^{13}C-edited NOESY spectrum (not shown). Peaks with arrows mark chemical shift perturbations that were also elucidated by the help of NOESY spectra. Dashed arrows indicate ambiguous peak assignments. (C) Secondary structure of the lariat-forming ribozyme. Residues with large chemical shift changes are color-coded in red. Unlabeled residues are given in gray.

also in chemical shift perturbations of several residues that are close to the mutation in the RNA construct. Therefore, in addition to the ^{13}C-edited HSQC spectra of H1'C1' and H6C6/H8C8 spectra, ^{13}C-filtered ^{13}C-edited and ^{12}C-filtered ^{13}C-edited NOESY spectra needed to be recorded for each mutation sample in order to unambiguously assign all nucleotides. The spectra of the G49C mutation are exemplarily shown in figure 5.7. In these spectra the disappearance of G49 (yellow circle) was complicated by the chemical shift changes of G1, G2, G5, G32 and G34, which are effected by the substitution of G49 to C49 and, thus, are in proximity to the mutated site G49.
Sequential assignment of the ACAGAGA region (residues 29-35, see figure 5.7C) was particularly challenging, due to the high density of purines and the conformational exchange in this part of the RNA molecule (double resonance signals). With the help of the A33G, mutation nucleotides A33 and A35 could be distinguished from each other. The repetition of the AU stretch from residue 50 to 53 also hampered a straightforward sequential assignment. The strategy of selected mutations finally allowed a complete sequential assignment of all nucleotides of the ribozyme.

Sequential assignment by deuterium labeling

Instead of introducing mutations to elucidate ambiguous sequential assignments, an alternative approach using deuterated samples have been applied, since deuteration leads to a reduced set of proton resonances. However, as discussed above, deuteration needs to be combined with hetero-nuclear labeling for larger systems. Ideally the four nucleotides with ^{2}H, ^{13}C,^{15}N-labeling would be used for this approach, which are currently not commercially available, though. Therefore, an NMR sample was prepared by *in vitro* transcription that contained deuterated adenosines and uridines, ^{13}C,^{15}N-labeled cytosine and protonated guanosines. Due to the deuterated nucleotides only low transcription yields were reached. Spectra of the low concentrated sample (0.1 mM) lead to the verification of prominent resonances, but did not elucidate challenging overlapping resonance regions. Therefore this cost-intensive approach was abandoned in favor of the mutation strategy discussed above.

5.3 Structure determination of the lariat-forming ribozyme

The following chapter describes the ongoing structure determination of the lariat-forming ribozyme. I will focus on the structural restraints used for structure calculation and the characteristics of base pairing patterns in the RNA molecule. Furthermore the preliminary NMR structure of the helical moiety (nucleotides 1 to 28) will be presented as well as first information achieved from SANS data.

Distance constraints for structure calculation

RNA structure determination by NMR mainly depends on distance constraints of inter-nucleotide protons derived from the analysis of NOESY spectra as described in the methods (chapter 2.5.6). NOE signals reflect direct distances of neighboring atoms in a molecule and thus establish a dense network of distance restraints. Distances between non-exchangable protons were derived from ^{13}C-filtered ^{13}C-edited NOESY and from ^{12}C-filtered ^{13}C-edited NOESY experiments for inter-nucleotide and inter-strand connectivities (see figure 5.8). At this point it should be pointed out, that the complete assignment of the ribozyme was a cooperative work performed in our laboratory. In particular, my personal contribution to the assignment work was the intra-nucleotide assignments of the differentially ^{13}C, ^{15}N-labeled samples G, C, AU, A and U (see table 2.2b) resulting in the collection of the intra-nucleotide NOEs. Instead, the inter-residue assignment was performed in the laboratory and my personal contribution to this second part of the assignment was the measuring and processing of the NMR spectra used for the sequential assignment and the NOE data collection shown in figure 5.8.

The integration of NOE volumes and the calibration of the distances were performed by an internal routine of Felix 2002 program (Accelrys Inc.). The quantification of the cross peak volumes was performed by categorizing them as: very weak (3.6 - 8.0 Å), weak (1.8 - 5.5 Å), medium (1.8 - 3.6 Å) or strong (1.8 - 3.0 Å). The volume of the pyrimidine H5-H6 cross-peak was used to set the reference distance of 2.45 Å. Adenosine volumes were indirectly referenced via NOE contacts to pyrimidines whereas guanosine volumes were referenced by using well established distances of the UUCG tetraloop.

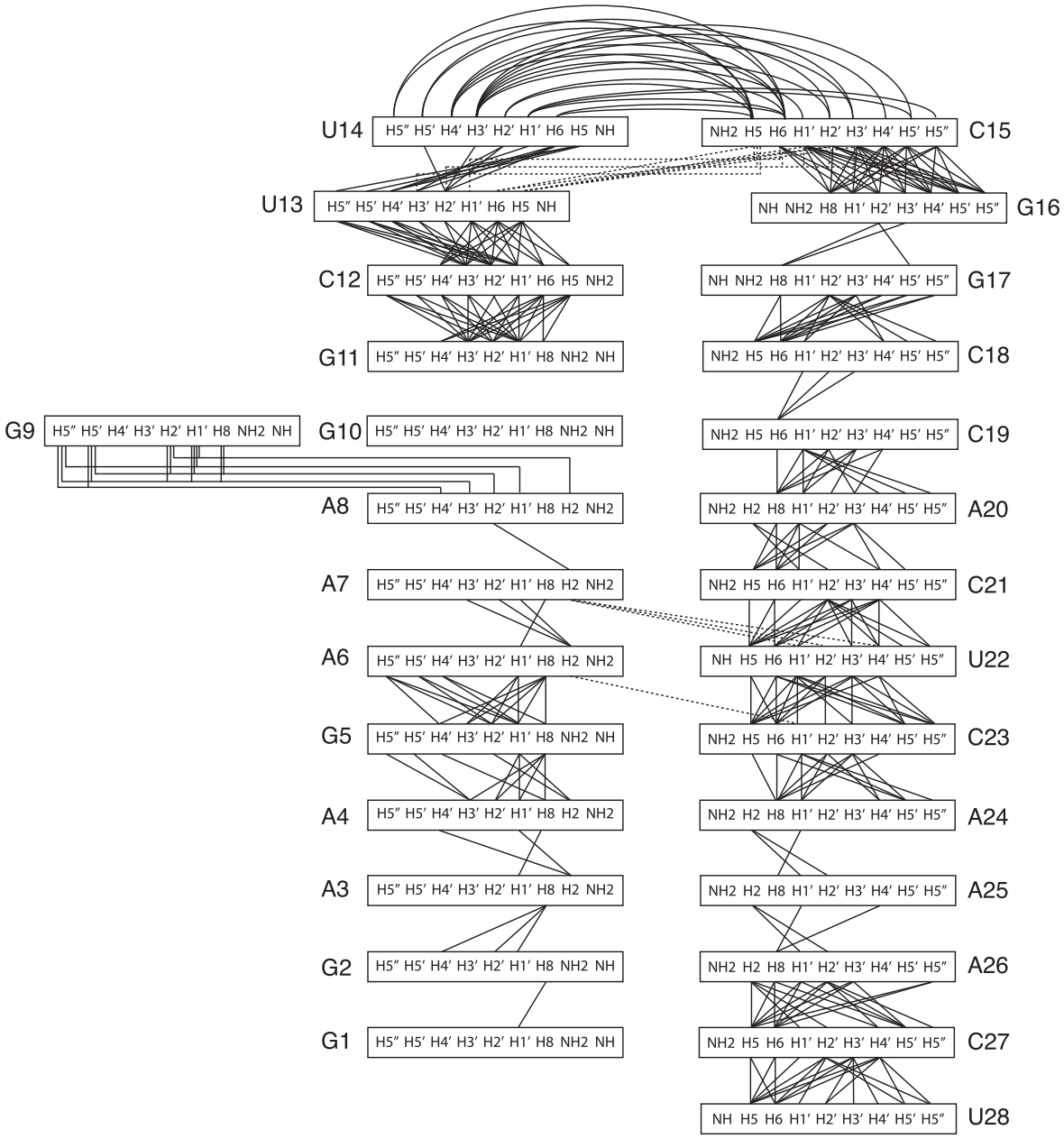


Figure 5.8: *Schematic presentation of internucleotide NOE contacts in the helical region 1 to 28 of the lariat-forming ribozyme* | Sequential NOE connectivities are presented by a full line, whereas long-range NOE contacts are shown as dotted lines.

The ribozyme comprises the conserved UUCG tetra-loop (residue 13-16) which delivers a dense network of NOE distance constraints, as visualized in figure 5.8. The motif was used as a starting point for the sequential assignment procedure, since the C(UUCG)G region displayed nearly identical chemical shifts to those previously reported [2][10][150]. As shown in figure 5.8 sequential assignment, could be accomplished for all nucleotides of the helical moiety

(residues 1-28) but distance constraints are unequally distributed. The characterization of neighboring nucleotides with the same base type (e.g. G1 and G2 or A7 and A8) delivers fewer inter-nucleotide restraints, due to the overlap of inter-nucleotide (sequential) NOEs (figure 5.8) and intra-nucleotide NOEs in the ^{13}C-filtered ^{13}C-edited NOESY spectra. Amongst most nucleotides, NOE connections between H6(Py)/ H8 (Pu) base protons and sugar protons H2′ and H1′ of the preceding residue could be observed, except that for G10 to G11 no connections could be seen, since all proton resonances of theses nucleotides have very similar chemical shifts and thus the NOE connections are highly overlapped with each other. However, the possible sequential walk through the sequence, together with the inter-strand NOEs between the H2 protons of A6 and A7 and the nucleotides C23 and U22, were strong indicators of the existence of a helix from nucleotide 1 to 27 in the lariat-forming ribozyme. The existence of the helical moiety was then verified by the HNN-COSY spectra (see next section) and mutational studies (see chapter 5.1). Analysis of the NOE spectra revealed strong sequential constraints between the H2 proton and the H1′ proton of the following nucleotide as well as weak NOE connections to H2′, H3′, H4′ protons. The distances from H2 protons in A-form helices usually range from 3.5 to 4 Å (H1′), 5Å to 7Å (H2′), 6.5Å to 8.5Å (H3′) and 6.5Å to 8Å (H4′), thus, are only poorly (H1′) or not observable (H2′ to H4′) in NOE spectra. Here, H2 protons of A3, A4, A6, A7, A20, A24, A25 and A26 revealed NOE signals to the H2′ of the following nucleotide (see figure 5.8). Moreover the H2 proton of A6 showed a NOE signal to the H3′ proton of A7 and the H2 protons of A3 and A26 are close to the H4′ proton of the following nucleotide. These unusual sequential NOEs of the H2 protons of adenines, especially to the ones of H3′ and H4′ sugar protons, indicated the formation of non-canonical base-pairs within the helix.
Sequential and long-range distances have also been constrained for the dynamic part of the molecule (residues 30 to 59), but since secondary structure elements remain to be elucidated by NMR studies and structure calculations for this moiety, a visualization of those NOE contacts is momentarily not practicable.

Elucidation of base pairing patterns

Resonances of exchangeable base protons provide crucial information regarding base pairings of RNA strands [151]. Identification of exchangeable imino and amino protons was performed using ^{15}N-edited HSQC experiments. The assignment of exchangeable protons was performed on the basis of a published assignment of a tetra-loop RNA segment with the same sequence GGC(UUCG)GCC [2] as present in the ribozyme and from the HNN-COSY

experiment (see figure 5.9B). Imino proton resonances of this segment (nucleotide 10 to 19) were clearly visible in the HNN-COSY spectrum.
The assignment of the G5 and U22 residues was based on comparison of imino proton resonances in the ^{15}N-HSQC spectrum of the original ribozyme (59-mer) versus the truncated (29-mer) RNA construct (spectrum not shown). Since imino resonances of these two nucleotides were present in both spectra, these belong to

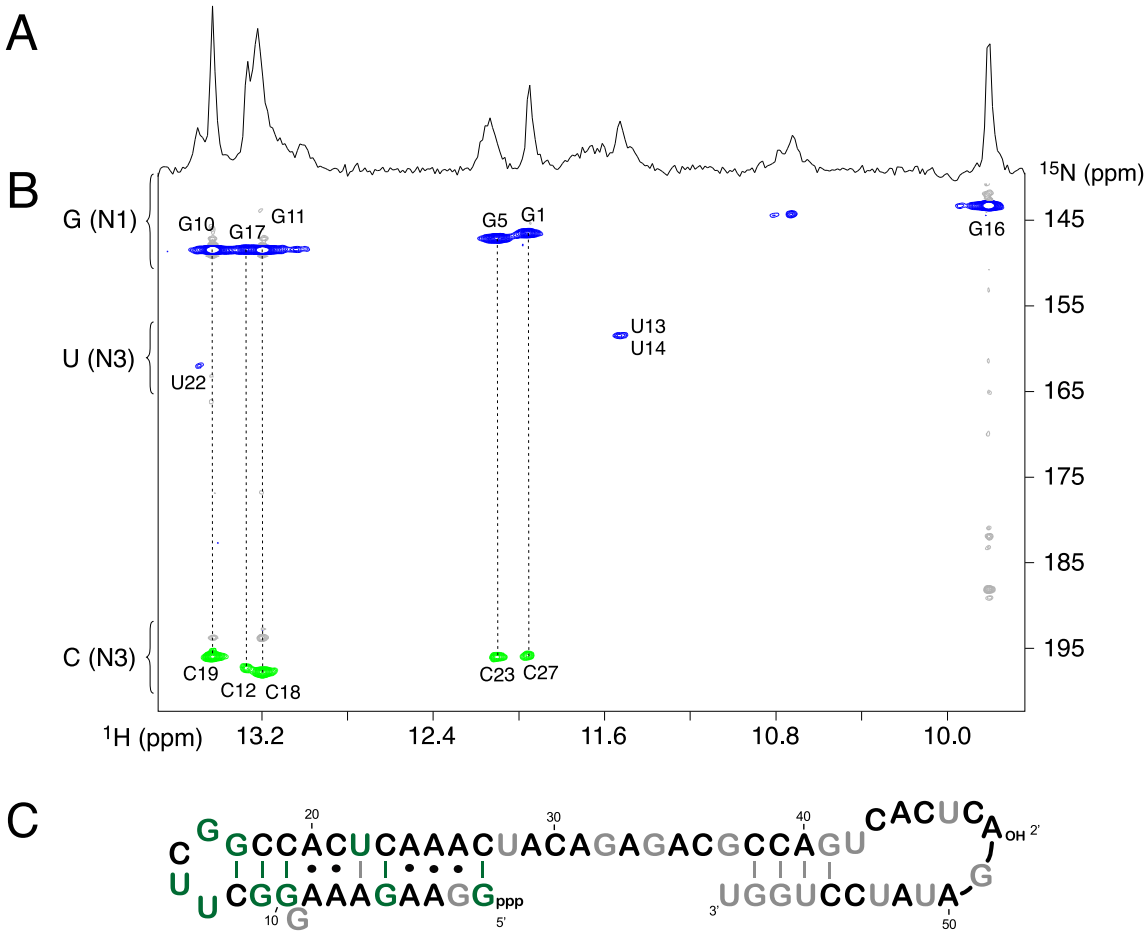


Figure 5.9: *Identification of base pairings with 1D imino spectrum (^{1}H) and a HNN-COSY experiment at 600 MHz and 298 K* | (A) 1D spectrum for the imino protons is shown on top. (B) base pairing interactions of GC Watson Crick base-pairs of the lariat-forming ribozyme are indicated by dashed lines. The transfer of magnetization occurs between N1 of guanines and N3 of cytosines via a $^2J_{NN}$ scalar coupling mediated by the hydrogen bond. Resonance assignment is based on literature values of the stable GUUC tetraloop [2]. Unassigned imino protons at 10.81 ppm and 10.74 ppm belong to the ribozyme moiety of nucleotide 30 to 59, since they are not observed in the sample of the helical region 1 to 29. (C) Secondary structure of the lariat-forming ribozyme. Uridine and guanidine residues with resonance assignment as well as verified base-pair correlations are color-coded in green. Residues with unassigned imino protons are shown in gray colors.

the helical moiety of the ribozyme (residue 1 to 29). Of the four uridines in this molecule part, U13/U14 resonances were already assigned to the tetra-loop and U28 is not part of the helix structure. The guanine resonance is part of the stable Watson-Crick (WC) base-pairs (see figure 5.9B), thus, can be assigned to G5. The imino resonance assigned to G1 is not visible in the spectra of the truncated NMR sample of the the ribozyme. However, the assumption can be made, that the terminal G1-C27 base-pair is stabilized by long-range contacts from the dynamic moiety (residue 30-59) and, thus, is only observable in the full-length construct.
Interestingly, the resonance of U22 displayed no cross peak in the HNN spectra, which would be expected for an A-U WC base-pair, suggesting that U22 is forming a non-canonical base-pair to A6. It should be noted that the resonance peak of U22 is quite weak compared to the guanidine signals. Thus, it might be possible that the U22-A6 pairing is of classic WC type, but that the A6 N1 peak is just not detected in the recorded spectra. In this case the weak, not detectable signal would arise from the dynamic behavior of the imino proton, which exchanges faster with water than a less dynamic A-U WC base-pair.
The unassigned imino protons of guanines at 10.81 ppm and 10.74 ppm belong to dynamic 3' end moiety of the ribozyme (residue 30 to 59), since they are not observed in the NMR sample of the truncated ribozyme version from residue 1 to 29. Since no cross peak to cytosines are observed in the HNN-COSY, a base-pair assignment is currently not possible.
The 1D ^{1}H NMR spectrum for the imino protons on top of the HNN-COSY spectra (figure 5.9A) indicates the existence of further imino proton resonance peaks (around 11.7 ppm and 13 ppm), which could not be resolved in multidimensional spectra, HNN-COSY and ^{15}N-edited HSQC. Probably, these imino protons belong to nucleotides that are part of non-canonical base-pairs, in which imino protons are not directly involved. Thus the exchange with water prevents detection of these imino protons in multidimensional spectra. Another explanation might be, that these nuclei are involved in base pairing, but they are more easily accessible to the solvent, since the opening rate of those base-pairs is slightly faster [131] and thus the exchange of the imino protons with water prevent their detection by NMR spectroscopy. This effect might also account for the fact that previously described WC base-pairs G41-C55, G57-C39 and G58-C38 were not observed in the recorded HNN-COSY. Furthermore this hypothesis is in agreement with the conformational exchange observed for the moiety of residues 30 to 59 (see chapter 5.2.2).

To further characterize the proposed non-canonical base-pairs A3-A25, A4-A24 and A8-A20 several further spectra were recorded, including versions of HNN-COSY [51], H2H8-TOCSY [72], H2H8-COSY [126] and ^{15}N-NOESY spectra, which were selective for amino protons [129]. However, none of these experiments enabled the assignment of those H2 protons that could not be easily assigned from the ^{13}C-filtered ^{13}C-edited NOESY spectrum, or the observation of magnetization transfer over the hydrogen bond of AA base-pairs. To verify non-canonical AA base-pair connections, that are clearly indicated by the combination of NOESY data (see figure 5.8) and mutational studies (see figure 5.2), further NMR measurements on the truncated version (see table 2.5e) of the lariat-forming ribozyme are currently in progress in our group.

Tertiary structure determination of the helical moiety from the ribozyme

The NOE data described above constituted the basis for structure calculations of the helical moiety (nucleotides 1-28) of the lariat-forming ribozyme. Structure calculations were performed by Irene Amata within the scope of her PhD thesis using 833 NOE-derived distances (see figure 5.8). A total number of 100 structures were calculated with a simulated annealing protocol from the Aria/CNS software [76].

The final ensemble of the 10 lowest energy structures (see figure 5.10) show an excellent convergency with a root main square deviation (RMSD) value of 0.6 Å.f For the stable tetra-loop region the structures converges to 0.34 Å. However,

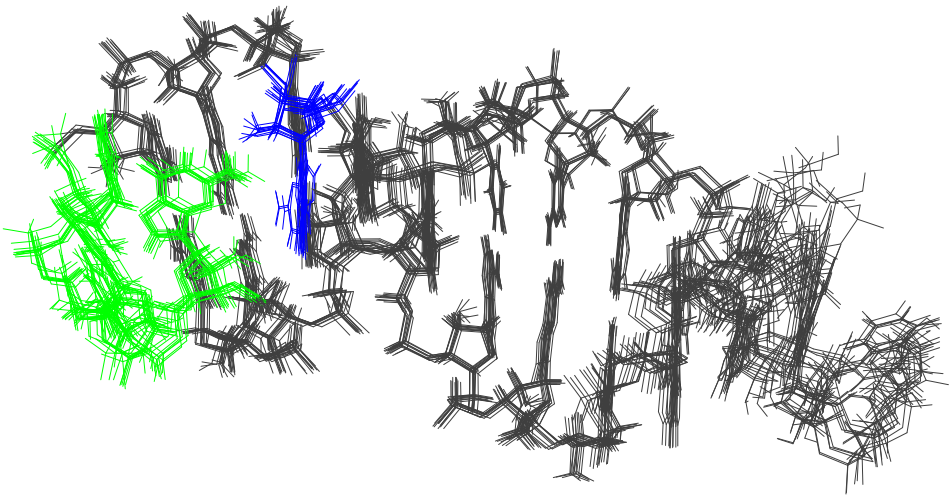


Figure 5.10: *The superposition of the 10 lowest-energy NMR-derived structures for the helix 1 to 28 of the lariat-forming ribozyme* | The structures are superimposed by excluding the G9 (blue) and U28 as well as the terminal base-pair G1-C27. Tetra-loop nucleotides are color-coded in green.

the hydrogen bond restraints used for the described structure calculations are determined by an extensive search of various combinations of non-canonical base-pairs, which need to be verified by NMR, as descibed in the previous chapter.

Measurements of SANS data of lariat-forming ribozyme
To obtain a description of the overall shape of the lariat-forming ribozyme, we performed a small angle neutron scattering analysis (SANS). The scattering curve strictly depends on the spatial coordinates of the scattering nuclei and therefore on the shape of the molecule. Three samples were prepared in context of this work, comprising the full length 59-mer RNA construct in its linear and lariat form as well as the truncated RNA construct of 29 nucleotide length. SANS measurements and analysis were carried out by Irene Amata with the help of Frank Gabel (Institute de Biologie Structurale, Grenoble). The experimental curve of the linear form of the ribozyme provided a radius of gyration of 20 Å when applying the Guinier approximation, whereas the lariat form of the ribozyme showed a value of 22.3 Å. Interestingly the truncated 29-mer RNA construct provide a similar R_g value of 19.9 Å as the full-length ribozyme. The similarity of R_g values from the experimental scattering curves for the 29-mer and 59-mer suggests that the stable helix (residue 1 to 29) of the ribozyme mainly contributed to the overall shape of the ribozyme molecule.

5.4 Studies of the ribozyme in the lariat form

The lariat-forming ribozyme is a model system with intriguing similarity to the first splicing reaction of the human spliceosome. To gain structural information of the model system and the subsequent lariat formation step, NMR trials were performed investigating the lariat form of the ribozyme (see figure 5.11). The initiation of the transesterfication reaction was induced by addition of magnesium chloride (see method chapter 2.4.3). After complete transesterfication, magnesium-ions were removed from the RNA construct by dialysis and the lariat samples were used to record ^{13}C-HSQC spectra (see figure 5.11).
A general feature of all recorded spectra of the lariat form of the ribozyme was the observation of extensive line broadening for all resonances except for those of the (GGC)UUCG(GCC) tetra-loop region and the WC-base-pairs C38-G58 and C39-G57 (figure 5.11). The reduction of NMR signals through line broadening suggests a dynamic behavior of the lariat form within a time window of microseconds.

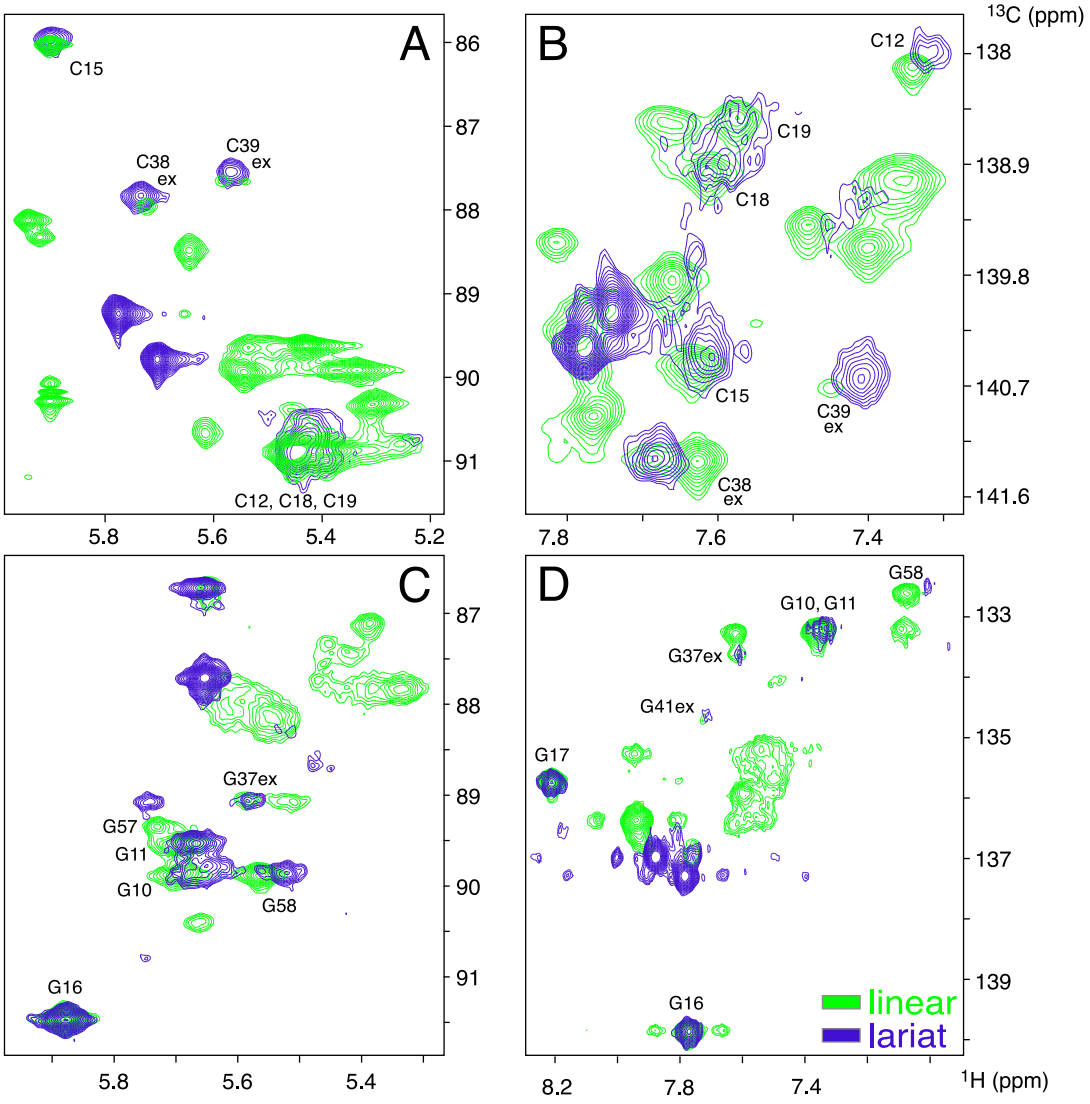


Figure 5.11: *Overlay of ^{13}C-edited HSQC spectra for the linear (green) and lariat (blue) form of the ribozyme* | H1'C1' (A) and H6C6 -correlation (B) of C-only labeled sample and H1'C1' (C) and H8C8-correlation (D) of G-only labeled sample. Resonances not effected by chemical shift perturbation were assigned. In the lariat form the chemical shifts of the exchange conformation is observed for residues G37, C38, C39 and G41. The broadening of most resonances of the lariat form indicates a conformational exchange of the ribozyme in the μs to ms time scale.

An increase of the exchange rate (fast exchange) would result in a single narrow line and, thus, would enable improved investigation of the lariat form by NMR. Also a decrease of the exchange rate would be desirable, since NMR signals of both conformation would remain detectable (see chapter 5.2.2). Subsequently, to influence the exchange rate, a trial was performed by addition of magnesium ions which are known as strong effectors of dynamic processes in RNA molecules (see figure 5.12).

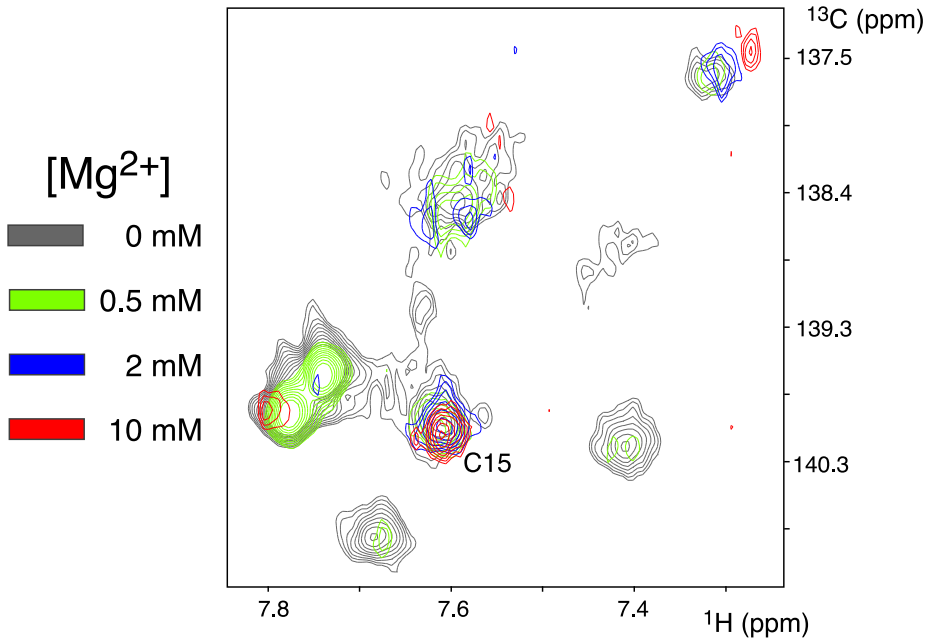


Figure 5.12: *Magnesium titration of the cytosine-specific ^{13}C,^{15}N-labeled sample of the lariat form of the ribozyme* | The H6C6 correlation is given for four magnesium-concentrations ranging from 0 to 10 mM. Except for C15 within the UUCG tetra-loop all nucleotides exhibit strong line broadening upon addition of magnesium ions.

Spectra of the cytosine-specific ^{13}C, ^{15}N-labeled lariat form were recorded for six magnesium concentrations, 0 mM, 15 μM, 50 μM, 0.5 mM, 2mM and 10 mM. The results of four titration points are depicted in figure 5.12 to show that the conformational exchange process could not be shifted to a time-regime, that would make lines observable in NMR spectra. The addition of magnesium even increased the line broadening effects. Subsequently, a structural investigation of the lariat form of the ribozyme by NMR spectroscopy is currently not feasible.

6 Discussion: The lariat-forming ribozyme

Characterization of the lariat-forming ribozyme

Structure determination of RNA molecules by NMR spectroscopy is usually based on the primary sequence and secondary structure information, in particular base pairing patterns suggested by biochemical studies. In the case of the ribozyme, activity studies can determine specific sites and patterns that can be either linked to folding or to their function in the catalytic activity of the RNA molecule. In Tushl *et al.* [143] the conservation of certain residues of the selected lariat-forming ribozyme (see figure 5.2) was previously elucidated by rounds of *in vitro* selection. The study focussed on functional important residues of the ACAGAGA region as well as the branch-point adenosine, which were further analyzed by assaying the activity of single-point mutation constructs. However, the residues with significance to the folding mechanism of the ribozyme have not been analyzed. In this work, site-specific mutagenesis studies were performed for the lariat-forming ribozyme to gain further insights into its secondary structure organization and to elucidate residues required for proper folding of the molecule. Figure 5.2B summarise the single substitutions as well as the deletion mutants investigated in this study and depicts the refined secondary model of the ribozyme.

Formation of a helical stem. A major achievement of this study has been the definition of the secondary structure in the 5′ region of the ribozyme, the first 27 nucleotides. Taking the results of the mutational studies and of the NMR analysis into account, the ribozyme revealed a well defined helical stem with an unusual high content of non-canonical base-pairs (5 out of 11). The importance of the length of the stem for the ribozyme activity was analyzed by deletion mutants; the functionality of the ribozyme was abolished when truncating the 5′ stem by 3 base-pairs. Extensive substitution studies in the helical segment further revealed the importance of WC base-pairs G5-C23 and A6-U22 for ribozyme function, whereas base-pair G2-A26 can be only mutated into a non-canonical base-pair or an U-A WC base-pair but not in a G-C base-pair. Consequently these observations indicate that an adenosine at position 26 is absolutely critical.

The necessity of a guanosine residue at the 5' end, which is attacked by the branch pointed adenosine, could not be confirmed. Although the mutation of G1A prevent lariat formation, the double mutation G1A/C27U restore the ribozyme activity and enabled the formation of lariat. Consequently these results suggest that a formation of a base-pair between the position 1 to 27 is required. When mutating U28, the first nucleotide downstream of the helical stem, lariat formation was impaired by the substitution U28C. We hypothesized that this substitution enables the formation of a WC C-G base-pair with G49, which prevents the proper orientation of the proximal catalytic center and, thus, block the nucleophilic attack of A48 to the α phosphate of G1. This hypothesis is confirmed by the fact, that the double mutation U28G/G49C blocks the lariat formation as well, whereas the single mutations U28G and G49C are tolerated. However, although the formation of a C-G base-pair is absolutely prohibited, the formation of a U-A base-pair might be possible, since the mutation of G49A is tolerated.
When comparing the data for the helical moiety derived within this work with the findings of the previous *in vitro* selection study [143], we find that the identity of the residues G2, A25 and U28 is not essential for ribozyme function in contrast to Tuschl *et al.*. In conclusion, extensive mutational analysis and NMR data enabled the refinement of the secondary structure in the N-terminal region of the ribozyme and revealed a well-defined helical region from residue 1 to 27.

Importance of the ACAGAGA segment. The ACAGAGA-box is a phylogenetically conserved and functionally important region within the U6 snRNA, which is also present in the lariat-forming ribozyme investigated here, which resulted from an *in vitro* selection experiment as a model system for the first transesterfication reaction occurring during processing of pre-mRNA [143]. Within the study of Tuschl *et al.* the conserved motif was tested by assaying the ribozyme activity for all possible single-nucleotide substitutions. The ACAGAGA segment was also included into the comprehensive mutational studies presented here, since NMR analysis performed in this work suggested a folding of the ribozyme in which the 3' end of the ribozyme is in proximity to the branch point sequence (BPS) and thus probably as well near the conserved ACAGAGA region. In general, our findings are consistent with the previous mutational studies confirming the importance of the ACAGAGA box for a functional ribozyme. However, in our experimental approach the amount of formed lariat was analyzed over a longer period of time compared to the previous study. This could mean that substitutions

previously found less active over a short period of time still deliver the same or even higher amounts of lariat over a long reaction period. This might explain why all substitutions of C30 and A31 as well as A29G and A33G do not inhibit the lariat formation in this study, while reducing ribozyme activity in the study of Tuschl *et al.*.
In this work the most critical mutations were found to be A29C, A33U and A35C, which completely blocked the transesterfication reaction. Another mandatory site is G32, where substitutions still allowed lariat formation, but with a tremendous decrease in lariat yields. The same is true for mutations A29U, A33C and G34C. However, since there is no comprehensive view of the secondary and tertiary structure of the ribozyme yet, the specific role of each nucleotide within the molecule remains unclear.

Nucleotides in proximity of the branch point site. In order to test whether nucleotides around the branch point adenosine exert an auxiliary role for the transesterfication reaction nucleotides 42 to 47 and 49 to 53 were systematically mutated. G49 and A50, located directly downstream from the branching adenosine were not essential for lariat formation. Thus, G49 and A50 are likely involved only in stacking interactions (not base-specific) but not directly in stabilizing the reaction center with specific functional groups. Downstream from the branch-point adenosine a higher tendency for conserved residues was observed, in particular when substituting U46 and C47, suggesting a more distinct role in the catalytic reaction. Nevertheless, it cannot be excluded that C47A and U46C substitutions might be prohibited due to a disruption of possible base-pairs.
By comparing the conserved patterns derived by previous *in vitro* selection studies [143] with the results from distinct mutations, the necessity to conserve the nucleotides identity in the stretch from A44 to A52 could not be confirmed. The discrepancy might be explained by the different experimental approaches. The previous study allows multiple sequence substitution at the same time, while monitoring the reaction rate; this study enables observing the effect of a distinct mutation upon lariat formation, thus, allowing more refined data regarding the need of specific residues. Here we rather asked the question whether a certain mutation construct can still perform the lariat-forming reaction in a reasonable time. Since only few base substitutions in the region around the BPS prevent the catalytic reaction of the ribozyme, the performed mutational studies give no hints regarding possible base-pair interactions. Thus, a secondary structure model for this ribozyme segment could not be established so far and requires more structural data.

In conclusion the findings of the performed mutational studies as well as the collected NMR data (mainly distances restraints derived from NOE connections) permitted first structural calculations, that are momentarily taking place for the helical, non-exchanging moiety of residue 1 to 28. The presented data derived from the comprehensive mutation studies represent an excellent starting point to explore the complete three-dimensional structure of the lariat-forming ribozyme and will facilitate its understanding in terms of functionality.

Observation of conformational exchange for the lariat-forming ribozyme

The analysis of NOESY spectra revealed a conformational exchange within the ribozyme starting from the ACAGAGA segment and reaching until the 3′ end. For the helical moiety in the 5′ region of the RNA molecule no conformational exchange could be observed. Since the switch of conformations is detectable by the presence of two sets of NMR lines, the rate of conversion is in the slow exchange regime on the NMR time-scale and thus can be confined to the high μs to s range.
The observed conformational switch explains the failure of crystallization trials previously performed in our laboratory [34]. However, the conformational exchange of the ribozyme also makes the assignment of NMR resonances quite challenging. The system size of 59 nucleotides itself present difficulties in the analysis of NMR data, since secondary structure elements are not well defined except for the helix in the 5′ region. Considering the slow conformational switch, the ribozyme constitutes 88 (30+29x2) assignable nucleotides. Thus, for the complete assignment of the ribozymes numerous differentially labeled samples and eight mutants were required.

First insights of the tertiary structure

Formation of the helical stem with numerous non-canical base-pairs. The preliminary 3-dimensional NMR structure of the 5′ stem of the ribozyme (see figure 5.10) converges to a well defined helix. Nonetheless, there are still some violations of the NOEs data mainly arising from intra-nucleotide distances, so that the data input and the structure outcomes need to be further refined and validated in the future. An interesting feature of the present structure is that G9 forms a base triple with G10 and C19, previously identified in structures of tRNAs, named GGC amino-carbonyl WC base triple [107]. However, mutational studies indicated that G9 can be removed, without loosing ribozyme activity. Consequently, the base triple is not necessary for the transesterfication reaction.

Assumptions regarding the overall fold of the ribozyme. Due to its system size and dynamic behavior the lariat-forming ribozyme presents a challenging system for NMR spectroscopic investigations. This is impressively underlined by the huge network of long-range interactions, observed by NOESY spectra analysis. The derived NMR data has been used as a basis for building a model of the overall fold of the ribozyme presented in figure 6.1. The suggested model visualizes the proximity of certain residues within the ribozyme as established by NOE data. So far, specific base pairing could only be established for the 5′ helix, but not for the 3′ part of the ribozyme, except for the base-pairs from C38 to G41. The reason for that could be the observed conformational exchange of the ribozyme, which might prevent the measurement of existing base-pairs as well as complicated tertiary folds, which might involve base triples or base quadruples. Such motifs are common in RNA structures and have been found in multitude in the ribosomal structure.

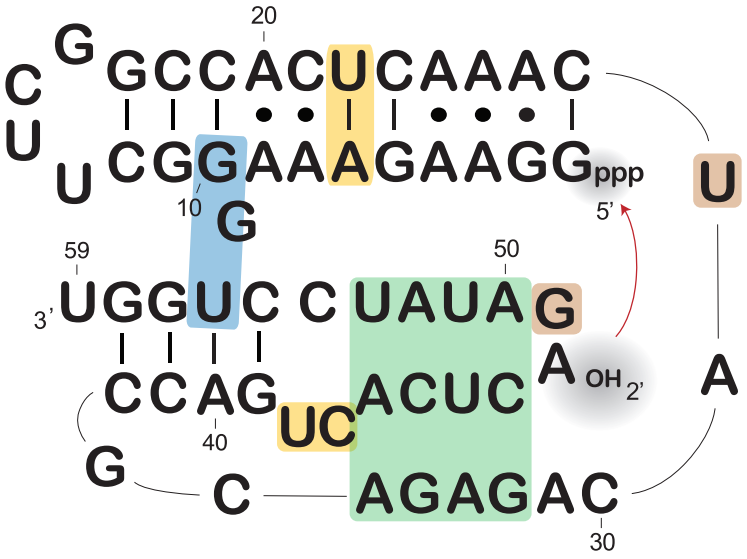


Figure 6.1: *Model of the 2′ 5′ lariat-forming ribozyme* | Colored boxes represents the proximity of the included residues established by analyzing NOE connections. The nucleophilic attack of the 5′ end by the hydroxyl group of the branch point adenosine is indicated by an arrow.

In particular, an RNA motif should be mentioned, which has been identified 46 times in the large subunit of the ribosome of *H. marismortui*, in which two of the found motifs bridge the small with the large subunit. The motif is named ribose zipper and is characterized by consecutive hydrogen-bonding interactions between ribose 2′-hydroxyls from different regions of an RNA chain. The RNA motif presents a suitable candidate for the fold of the lariat-forming ribozyme, since NOE data revealed several contacts between segments of the ribozyme

(green and yellow boxes in figure 6.1), which might be explained by the ribose zipper motif. Interestingly, ribose zippers are found in the structures of group I [11] and group II [141] introns, which are catalytically active RNA molecules like the studied ribozyme.
Another interesting fold, which might be adopted by the ribozyme is the pseudoknot fold. This structural RNA motif is minimally comprised of two helical elements connected by two single-stranded loops, thereby providing a way, in which the single-strand loop of the RNA can fold back on the stem region [134]. Due to variations of loop and stem lengths as well as the type of interactions between them, pseudoknots represent a diverse group of RNA structures. Pseudoknot motifs have been described for diverse RNA functions including catalytic cores of various ribozymes and self-splicing introns (reviewed in [39]). Features of the pseudoknot can be transfered to the presented ribozyme model, where the 3′ end region and parts of the ACAGAGA segment could serve as stem and loop region.
So far the number of published structures of RNA molecules is very limited compared to the number of protein structures; considering the high versatility of RNA molecules a vast variety of RNA motifs and folds is yet to be discovered. Thus, ongoing structure calculations of the full length ribozyme need to be awaited to understand the complete tertiary fold of the lariat-forming ribozyme and to analyze the role of the ACAGAGA segment in the transesterfication reaction.

Insights from the lariat form of the ribozyme?
The 2′ 5′ lariat-forming ribozyme, which is similar to the first transesterfication reaction in pre-mRNA splicing, can be used as a model system of the splicing process. Thereby the ribozyme fold is of interest both prior to the catalytic reaction and after lariat formation. In this work the transesterfication reaction of the lariat-forming ribozyme was initiated upon addition of magnesium ions and the formed lariat was investigated by NMR spectroscopy. Unfortunately the ribozyme in its lariat form exhibits motion in the intermediate range on the NMR time scale, which translates into excessive line broadening of almost all nucleotides. NMR titration experiments by adding magnesium ions were performed to shift the observed motion into a more favorable time scale, which would allow observation of NMR signals. However, the addition of magnesium during NMR measurements even amplified the mentioned line broadening. In conclusion, a structural investigation of the lariat form of the ribozyme by NMR spectroscopy is currently not feasible.

Nevertheless, the measured NMR data revealed that next to the $U^{13}UCG^{16}$-tetra-loop segment of the helical moiety and its three adjacent base-pairs, the base-pairs C38-G58 and C39-G57 as well as the nucleotides G37 and G41 are still present in the lariat form of the ribozyme. Interestingly, the nucleotides G37, C38, C39 and G41 were present with the chemical shifts that represent the second conformation of the ribozyme in its linear form, whereas G57 and G58 still exist in the first conformation. This observation suggests that parts of the second conformation of the linear ribozyme, namely the stretch of G37 to G41, already resemble the arrangement of nucleotides of the lariat form. In the samples of the linear ribozyme form, no magnesium ions are present and thus the reaction cannot be performed causing the conformation switches back to the more favored first conformation.

In addition to the NMR experiments SANS measurements were performed for the RNA construct in its linear and lariat form, thus, enabling a comparison of the ribozyme prior and after the transesterfication reaction. From the linear to lariat form an increase of the radius of gyration has been observed, which implies a slight change in the overall shape of the ribozyme mediated through the branch formation. The measured SANS data will be used for *ab inito* modeling in the future in order to reconstruct the overall shape of the system in both conformations.

Abbreviation and acronym list

1D, 2D, 3D	1-, 2-,3- **D**imensional
A	**A**denosine
Å	**Å**ngström
ARIA	**A**mbiguous **R**estraints for **I**terative **A**ssignment
ATP	**A**denosine **T**ri**p**hosphate
BPS	**B**ranch **p**oint **s**equence
C	**C**ytosine
CNS	**C**rystallography & **N**MR **S**ystem
COSY	**Co**rrelated **S**pectroscop**y**
C-stem	**c**anonical stem
CTP	**C**ytosine **T**riphosphate
CV	**C**olumn **V**olume
DEAE	**Di****e**thyl**a**mino**e**thyl
DNA	**D**eoxyribo**n**ucleic **a**cid
ds	**d**ummy **s**can
DTT	**Di****th**io**th**reitol
E-COSY	**E**xclusive **Co**rrelation **S**pectroscop**y**
EDTA	**E**thylene **d**i-amine **t**etra **a**cetic acid
FF	**F**ast **F**low
FRET	**F**luorescence **R**esonance **E**nergy **T**ransfer
g	gravitation constant
G	**G**uanosine
GARP	**G**lobally optimized **A**lternating phase **R**ectangular **P**ulse
GST	**G**lutathione-**S**-**t**ransferase
GTP	**G**uanosine **T**riphosphate
h	**h**ours
Hepes	4-(2-**H**ydroxy**e**thyl)-1-**p**iperazine**e**thane **s**ulfonic acid
HH	**H**ammer**h**ead
hPrp	**h**uman **p**re-m**R**NA **p**rocessing factor
HSQC	**H**eteronuclear **S**ingle **Q**uantum **C**oherence
Hz	**H**ert**z**

INEPT	**I**nsensitive **N**uclei **E**nhancement by **P**olarisation **T**ransfer
IPTG	**I**so-**p**ropyl-β-D-**t**hio-**g**alactopyranoside
ISL	**I**nternal **s**tem **l**oop
$^{n}J_{ab}$	scalar coupling constant through n bonds between nucleus a and b
K	**K**elvin
kDa	**k**ilo **Da**lton
k_{ex-1}	rate constant
KT	**K**ink-**t**urn
k-turn	**K**ink-turn
LB	**L**ysogeny **b**roth or **L**uria-Bertani **b**roth
M	**M**olar
MD	**M**olecular **D**ynamic simulation
min	**min**ute
miRNA	**mi**cro **r**ibo**n**ucleic **a**cid
MHz	**M**ega**h**ert**z**
mM	**m**illi**m**olar
mRNA	**m**essenger **r**ibo**n**ucleic **a**cid
mRNP	**m**essenger **R**ibo**n**ucleo**p**article
MWCO	**M**olecular **W**eight **C**ut **O**ff
N any	**N**ucleotide
NC-stem	**n**on-**c**anonical stem
ncRNA	**n**on-**c**oding **r**ibo**n**ucleic **a**cid
Ni-NTA	**Ni**ckel-**n**itrilo**t**ri-**a**cetic acid
nm	**n**ano**m**eter
NMR	**N**uclear **M**agnetic **R**esonance
NOE	**N**uclear **O**verhauser **E**ffect
NOESY	**N**uclear **O**verhauser **E**ffect **S**pectroscop**y**
ns	**n**umber of **s**cans
nt	**n**ucleo**t**ide
NTP	**N**ucleotide **t**ri**p**hosphate
01p, 02p, 03p	carrier frequency of channel 1,2 or 3 in ppm
OD	**O**ptical **d**ensity
PAGE	**P**olyacryl**a**mide **g**el **e**lectrophoresis
PBS	**P**hosphate **B**uffered **S**aline
PCR	**P**olymerase **C**hain **R**eaction
PDB	entry of the **P**rotein **D**ata **B**ank
PEG	**P**oly**e**thylene **g**lycol
piRNA	**pi**wi **r**ibo**n**ucleic **a**cid

ppm	**p**arts **p**er **m**illion
pre-mRNA	**p**re **m**essenger **r**ibo**n**ucleic **a**cid
Pu	**Pu**rine
Py	**Py**rimidine
r	distance
R_g	**R**adius of **g**yration
R	Purine
RDC	**R**esidual **D**ipolar **C**oupling
RMSD	**R**oot **M**ean **S**quare **D**eviation
rRNA	**r**ibosomal **r**ibo**n**ucleic **a**cid
SANS	**S**mall **A**ngle **N**eutron **S**cattering
SAXS	**S**mall **A**ngle **X**-ray **S**cattering
SDS	**S**odium **D**odecyl **S**ulfate
siRNA	**s**mall **i**nterfering **r**ibo**n**ucleic **a**cid
SL	**S**tem **L**oop
snoRNP	**s**mall **n**ucle**o**lar **r**ibo**n**ucleic **a**cid
snRNA	**s**mall **n**uclear **r**ibo**n**ucleic **a**cid
snRNP	**s**mall **n**uclear **r**ibo**n**ucleo**p**rotein particle
ss	**s**plice **s**ite
STATES-TPPI	
T7 RNAP	*T7* **RNA P**olymerase
t1, t2, t3	chemical shift evolution time 1, 2 or 3
TBE	**T**ris **B**orate **E**DTA
TOCSY	**T**otal **Co**rrelation **S**pectroscop**y**
Tris	Tris(hydroxymethyl)aminomethane
TROSY	**T**ransverse **R**elaxation **o**ptimized **S**pectroscop**y**
tRNA	**t**ransfer **r**ibo**n**ucleic **a**cid
μl	**micro**liter
μs	**micros**econd
U	**U**ridine
U1, U2, U4, U5, U6	Uridine-rich snRNP 1, 2, 4, 5 or 6
UTP	**U**ridine **T**riphosphate
UV	**U**ltra**v**iolet
V	**V**olt
W	**W**att
WC	**W**atson **C**rick
WURST	**W**ideband **U**niform **R**ate **S**mooth **T**runcation pulse
Y	P**y**rimidine

Bibliography

[1] Allain, F. H. and Varani, G. (1995), *J. Mol. Biol.*, **250**, 3:333–353

[2] Bae, S.-H., Cheong, H.-K., Lee, J.-H., Cheong, C. *et al.* (2001), *PNAS*, **98**(19), 10602-10607

[3] Ban, N., Nissen, P, Hansen, J., Moore, P. B. *et al.* (2000), *Science*, **289**, 905–20

[4] Bodenhausen G. and Ruben D. J. (1980), *Chem. Phys. Lett.*, **69**, 1:185–189

[5] Bringmann, P., Appel, B., Rinke, J., Reuter, R. *et al.* (1984), *EMBO J.*, **3**, 1357- 1363

[6] Brow, D. and Guthrie, C. (1988), *Nature*, **334**, 213-218

[7] Brow, D.A. (2002), *Annu Rev. Genet*, **36**, 333-360

[8] Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L. *et al.* (1998), *Acta Crystallogr.* **D54**, 905-921

[9] Brutscher, B., Boisbouvier, J., Kupce, E., Tisne, C. *et al.* (2001), *J. Biomol. NMR*, **19**(2), 141–151

[10] Butcher, S.E., Dieckmann, T. and Feigon, J. (1997), *EMBO J.*, **16**, 7490-7499

[11] Cate, J. H., Gooding, A. R., Podell, E., Zhou, K. *et al.* (1996), *Science*, **273**, 1678-1685

[12] Carlomagno, T., Hennig, M. and Williamson, J. R. (2002), *J. Biomol. NMR*, **22**, 65-81

[13] Cavaluzzi, M.J. and Borer P.N. (2004), *NAR*, **32**, e13

[14] Cech, T. R. (1987), *Science*, **236**, 1532-1539

[15] Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L. *et al.* (2002), *Mol. Genet*, **11**, 87-92

[16] Chen, G., Kennedy, S. D., Qiao, J., Krugh, T. R. *et al.* (2006), *Biochemistry*, **45**, 6889-6903

[17] Chiara, M. D. and Reed, R. (1995), *Nature*, **375**, 510-513

[18] Chua, K. and Reed, R. (1999), *Nature*, **402**, 207-210

[19] Clore, G.M. and Garret, D.S. (1999), *JACS*, **121**, 9008-9012

[20] Clouet-d'Orval, B. and Uhlenbeck, O.C. (1996), *RNA*, **2**, 483-491

[21] Clouet-d'Orval, B. and Uhlenbeck, O.C. (1997), *Biochemistry*, **36**, 9087-9092

[22] Cojocaru, V., Klement, R. and Jovin, T.M. (2005), *NAR*, **33**, 3435- 3446

[23] Cojocaru, V., Nottrott, S., Klement, R. and Jovin, T.M. (2005), *RNA*, **11**, 197-209

[24] Cunningham, P.R., Ofengand, J. (1990), *Biotechniques*, **9**, 6, 713-714

[25] Datta, B. and Weiner, A.M. (1993), *Mol. Cell Biol.*, **13**(9), 5377-5382

[26] Datta, B. and Weiner, A.M. (1991), *Nature*, **352**, 821–824

[27] Dayie, K.T. (2008), *Int. J. Mol. Sci.*, **9**, 1214-1240

[28] Dingley, A.J., Nisius, L., Cordier, F. and Grzesiek, S. (2008), *Nat. Prot.*, **3**(2), 242-248

[29] Dingley, A.J. and Grzesiek, S. (1998), *JACS*, **120**(33), 8293

[30] Domdey, H., Apostol, B., Lin, R. J, Newman, A. *et al.* (1984), *Cell*, **39**(3), 611–621

[31] Dornberger, U., Leijon, M., Fritzsche, H. (1999), *J.Biol.Chem.*, **274**(11), 6957-6962

[32] Draper, D.E., Grilley, D. and Soto, A.M. (2005), *Annu. Rev. Biophys. Biomol. Struct.*, **34**, 221-243

[33] Fiala, R., Jiang, F and Sklenář, V. (1998), *J. Biomol. NMR*, **12**, 373-383

[34] Fohrer, J. (2007), NMR-spektroskopische Untersuchungen an katalytischer RNA Frankfurt (Main), Univ., Diss., URN: 30-56362

[35] Fürtig, B., Richter, C., Wöhnert, J. and Schwalbe, H. (2003), *ChemBioChem*, **4**, 936-962

[36] Gabel, F., Simon, B., Nilges, M., Petoukhov, M. *et al.* (2008), *J. Biomol. NMR*, **41**, 199-208

[37] Gautier, T., Berges, T., Tollervey, D. and Hurt, E. (1997), *Mol. Cell. Biol.*, **17**, 7088-7098

[38] Gehring, K. and Ekiel, I. (1998), *J. Magn. Reson.*, **135**, 185-193

[39] Giedroc, D.P., Theimer, C.A., and Nixon, P.L. (2000), *J. Mol. Biol.*, **298**, 167-185

[40] Goody, T.A., Melcher, S.E., Norman, D.G. and Lilley, D.M. (2004), *RNA*, **10**, 254-264

[41] Gordon,P.M., Sontheimer,E.J. and Piccirilli,J.A. (2000), *RNA*, **6**, 199-205

[42] Green, M.R. (1986), *Annu. Rev. Genet.*, **20**, 671-708

[43] Grishaev, A., Ying, J. and Bax, A. (2006), *J. Am. Chem.*, **128**, 10010-10011

[44] Grishaev, A., Tugarinov, V., Kay, L.E., Trewhella *et al.* (2008), *J. Biomol. NMR*, **40**, 95-106

[45] Grishaev, A., Ying, J., Canny, M.D., Pardi, A. *et al.* (2008), *J. Biomol. NMR*, **42**, 99-109.

[46] Groebe, D.R. and Uhlenbeck, O.C.(1988), *NAR*, **16**, 24, 11725-11735

[47] Guthrie, C. and Patterson, B. (1988), *Annu Rev. Genet.*, **22**, 387-419

[48] Gutowsky, H. S., and Holm, C. H. (1956), *J. Chem. Phys.*, **25**, 1228-1234

[49] Hansen, M.R., Mueller, L. and Pardi, A. (1998), *Nat. Struct. Biol.*, **5**, 1065-1074

[50] Hashimoto, C. and Steitz, J.A. (1984), *NAR*, **12**, 3283-3293

[51] Hennig, M. and Williamson, J.R. (2000), *NAR* **28**, 1585

[52] Hofacker, I.L., Fontana, W., Stadler, P., Bonhoeffer, L. *et al.* (1994), *Monatshefte Chemie*, **125**, 167-188

[53] Hofacker, I.L., Fekete, M. and Stadler, P. (2002), *JMB*, **319**, 1059-1066

[54] Hu, W., Kakalis, L. T., Jiang, L., Jiang, F. *et al.* (1998), *J. Biomol. NMR*, **12**, 559-564

[55] Huranova, M., Hnilicova, J., Fleischer, B., Cvackova, Z. *et al.* (2009), *Hum. Mol. Genet.*, **18**, 2014-2023

[56] Jurica, M.S. and Moore, M.J. (2003), *Mol. Cell.*, **12**(1), 5-14

[57] Katahira, M., Kanagawa, M., Sato, H., Uesugi, S. *et al.* (1994), *NAR*, **22**(14), 2752-2759

[58] Kandels-Lewis, S. and Séraphin, B. (1993), *Science*, **262**, 2035-2039

[59] Kent, O. A. and A. M. MacMillan (2002), *Nat. Struct. Biol.*, **9**, 576-581

[60] O'Keefe,R.T., Norman,C. and Newman,A.J. (1996), *Cell* **86**, 679-689

[61] Kim, D.H. and Rossi, J.J. (1999), *RNA*, **5**(7), 959-971

[62] Kirkpatrick, J.P., Li, P. and Carlomagno, T. (2009), *ChemBioChem*, **10**(6), 1007-1014

[63] Klein, D.J., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2001), *Embo J.*, **20**, 4214-4221

[64] Konarska M.M., Vilardell J. and Query C.C. (2006), *Mol. Cell.*, **21**, 543-553

[65] Koonin, E.V., Bork, P. and Sander, C. (1994), *NAR*, **22**, 2166-2167

[66] Koradi, R., Billeter, M. and Wuthrich, K. (1996), *J. Mol. Graph.*, **14**, 51-55

[67] Latham, M.P., Brown, D.J., McCallum, S.A. and Pardi, A. (2005), *ChemBioChem*, **6**, 1492-1505

[68] Lesser, C.F. and Guthrie, C. (1993), *Science*, **262**, 1982-1988

[69] Linge, J.P., Habeck, M., Rieping, W. and Nigels, M. (2003), *Bioinformatics*, **19**, 315-316

[70] Li, P., Kirkpatrick, J. and Carlomagno, T. (2009), *J. Mol. Biol.*, **388**(2), 283-98

[71] Liu, S. and Li, P., Dybkov, Nottrott, S. *et al.* (2007), *Science*, **316**, 115-120

[72] Legault, P., Farmer II, B.T., Müller, L. and Pardi, A. (1994), *JACS*, **116**, 2203-2204

[73] Le Hir, H., Moore, M.J. and Maquat, L.E. (2000) *Genes Dev.*, **14**, 1098-1108

[74] Lehmann, K. and Schmidt, U. (2003), *Crit. Rev. Biochem. Mol. Biol.*, **38**, 249-303

[75] Leontis, N.B. and Westhof, E. (2001), *RNA*, **7**, 499-512

[76] Linge, J.P., Habeck, M., Rieping, W. and Nilges, M. (2003), *Bioinformatics*, **19**, 315-316

[77] Madhani, H.D. and Guthrie, C. (1994), *Annu Rev. Genet.*, **28**, 1-26

[78] Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher *et al.*(2002), *EMBO J.*, **21**, 1148-1157

[79] Mao, H.Y., White, S.A. and Williamson, J.R. (1999), *Nat. Struct. Biol.*, **6**, 1139-1147

[80] Mao, H.Y. and Williamson, J.R. (1999), *NAR*, **27**, 4059-4070

[81] Marino, J.P., Prestegard, J.H. and Crothers, D.M. (1994), *JACS*, **116**, 2205-2206

[82] Maschhoff, K. L. and Padgett, R. A. (1993), *NAR*, **21**, 5456

[83] Matsumura, S., Ikawa, Y. and Inoue, T. (2003), *NAR*, **31**, 5544-5551

[84] Mattick, J.S. and Makunin, I.V. (2006), *Hum Mol Genet.*, **15**(1), R17-29

[85] McKay, D.B. (1996), *RNA*, **2**, 395-403

[86] McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E.*et al.* (2001), *Hum. Mol. Genet.*, **10**, 1555-1562

[87] Mefford, M.A. and Staley, J.P. (2009), *RNA*, **15**(7), 1386-97

[88] Van Melckebeke H., Pardi A., Boisbouvier J., Simorre J. P. *et al.* (2005), *J. Biomol. NMR*, **32**, 4:263-271

[89] Milligan, F.M., Groebe, D.R., Witherell G.W. and Uhlenbeck O.C. (1987), *NAR*, **15**, 8783-8798

[90] Moore, M.J. and Sharp, P.A. (1993), *Nature*, **23**(365), 364-368

[91] Morris, G.A. and Freeman, R. (1979), *JACS*, **101**, 760-762

[92] Mount, S. M. (1982), *NAR*, **10**, 459-72

[93] Myslinski, E. and Branlant, C. (1991), *Biochimie*, **73**, 17-28

[94] Newman, A. and Norman, C. (1991), *Cell*, **65**(1), 115-123

[95] Newman, A. (1997), *EMBO J.*, **16**(19), 5797-5800

[96] Nilsen, T.W. (2003), *Bioessays*, **25**, 1147-1149

[97] Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. *et al.* (2001), *PNAS*, **98**, 4899-4903

[98] Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H.*et al.* (1999), *Embo J.*, **18**, 6119-33

[99] Nottrott, S., Urlaub, H. and Lührmann, R. (2002), *Embo J.*, **21**, 5527-5538

[100] Oh, B. H., Westler, W. M., Derba, P. and Markley, J. L. (1988), *Science*, **240**, 90821

[101] Research Collaboratory for Structural Bioinformatics (RCSB) - Protein Data Bank (PDB), www.rcsb.org/pdb

[102] Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L. *et al.* (1986), *Cell*, **44**(2), 213-223

[103] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch *et al.* (2004), *J Comput Chem*, **25**(13), 1605-12

[104] Pleiss, J.A., Derrick, M.L. and Uhlenbeck O.C. (1998), *RNA*, **4**, 1313-1317

[105] Prestegard, J.H. (1998), *Nature, Struc. Biol.*, **NMR Suppl**, 517-522

[106] Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007), *Q. Rev. Biophys.*, **40**(3), 191-285

[107] Quigley, G.J. and Rich, A. (1976), *Science*, **194**, 796-806

[108] Ramos, A., Bayer, P. and Varani, G. (1999), *Biopolymers*, **52**, 181-196

[109] Razga, F., Zacharias, M., Reblova, K., Koca *et al.* (2006), *Structure* **14**, 825-835

[110] Reed R. (1989), *Genes Dev.*, **3**(12B), 2113-2123

[111] Reiter, N. J., Maher, L. J. and Butcher, S. E. (2008), *NAR*, **36**, 1227-1236

[112] Rinke, J., Appel, B., Digweed, M. and Lührmann, R. (1985), *J. Mol. Biol.* **185**, 721-731

[113] Rhode, B.M., Hartmuth, K., Westhof, E. and Lührmann, R. (2006), *EMBO J.*, **6**(25), 2475-2486

[114] Ruffner, D.E, Storno and G.D. and Uhlenbeck, O.C. (1987), *Biochem*, **29**, 10695-10702

[115] Rymond, B.C. and Rosbash. M. (1985), *Nature*, **317**, 735-737

[116] Sashital, D.G., Cornilescu, G., McManus, C.J., Brow *et al.* (2004), *Nat. Struct. Mol. Biol.*, **11**, 1237-1242

[117] Sawa, H. and Abelson, J. (1992), *PNAS*, **89**(23), 11269-73

[118] Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. *et al.* (2004), *EMBO J.*, **23**, 3000-3009

[119] Schluenzen, F., Tocilj, A., Zarivach, R., Harms *et al.* (2000), *Cell*, **102**, 615-23

[120] Schultz, A., Nottrott, S., Hartmuth, K. and Lührmann, R. (2006), *J. Bio. Chem.*, **281**, 2878-2886

[121] Schwalbe, H., Marino, J. P., Glaser S. J., Griesinger, C. (1995), *JACS* **117**, 7251-7252

[122] Schwer B. (2000), *Nature Struct. Biol.*, **8**, 113-116

[123] Schwer, B. and Gross, C.H. (1998), *EMBO J.*, **17**(7), 2086-2094

[124] Schwer, B. and Guthrie, C. (1992), *EMBO J.*, **11**(13), 5033-9

[125] Scott, L.G., Tolbert, T.J., Williamson, J.R. (2000), *Methods Enzymol.*, **317**, 18-38

[126] Simon, B., Zanier, K., Sattler, M. (2001), *J. Biomol. NMR*, **20**, 173-176

[127] Shields, T.P., Mollova, E.T., Marie, L.S., Hansen *et al.* (1999), *RNA*, **5**, 1259-1267

[128] Sklenář, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993), *J. Biomol. NMR*, **3**, 721-727

[129] Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993), *J. Mag. Res. Series A*, **102**, 241-245

[130] Smith DJ, Query CC, Konarska MM (2008), *Mol. Cell*, **30**(6), 657-66

[131] Sontheimer, E. and Steitz, J. (1993), *Science*, **262**, 1989-1996

[132] Sontheimer, E.J., Gordon, P.M. and Piccirilli, J.A. (1999), *Genes Dev.*, **13**(13), 1729-1741

[133] Snoussi, K. and Leroy, J.L. (2001), *Biochem.*, **40**, 8898-8904

[134] Staple, D.W. and Butcher, S.E. (2005), *PLoS Biol.*, **3**(6), e213

[135] Staley, J.P. and Guthrie, C. (1998), *Cell*, **92**, 315-326

[136] Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E. *et al.* (2002), *Mol. Cell*, **9**, 31-44

[137] Sun, J.S. and Manley, J.L. (1995), *Genes Dev.*, **9**(7), 843-854

[138] Talluri S. and Wagner G. (1996), *J. Mag. Res Series B*, **112**(2), 200-205

[139] Tjandra, N. and Bax, A. (1997), *Science*, **278**, 1111-1114

[140] Tolman, J. R., Flanagan,J. M., Kennedy, M. A. and Prestegard, J. H. (1995), *PNAS*, **92**, 9279

[141] Toor, N., Keating, K.S., Taylor, S.D. and Pyle, A.M. (2008), *Science*, **320**, 77-82

[142] Tropp J. (1980), *J. Chem. Phys.*, **72**, 6035-6043

[143] Tuschl, T., Sharp, P.A. and Bartel, D. P. (2001), *RNA*, **7**, 29-43

[144] Tzakos, A.G, Grace, C.R.R., Lukavsky, P.J. and Riek, R. (2006), *Annu. Rev. Biophys. Biomol. Struct.*, **35**, 319-42

[145] Umen, J.G. and Guthrie, C. (1995), *RNA*, **1**, 869-885

[146] Valadkhan, S. and Manley, J.L. (2001), *Nature*, **413**, 701-707

[147] Valadkhan, S. and Manley, J.L. (2000), *RNA*, **6**, 206-219

[148] Van der Veen,R., Arnberg, A. C., Vanderhorst, G., Bonen *et al.* (1986), *Cell*, **44**(2), 225-234

[149] Vankan, P., McGuigan, C. and Mattaj, I.W. (1992), *EMBO J.*, **11**, 335-343

[150] Varani, G., Cheong, C. and Tinoco, I. (1991), *Biochem*, **30**(13), 3280-3289

[151] Varani, G., Aboul-ela, F., Allain, F.H.T. (1996), *Prog. NMR Spectrosc.* **29**, 51-127

[152] Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R. *et al.* (2000), *Mol. Cell*, **6**, 1331-1342

[153] Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A. *et al.* (2001), *Mol. Cell*, **8**, 375-381

[154] Wahl, M.C., Will, C.L. and Lührmann, R. (2009), *Cell* **136**(4), 701-718

[155] Wassarman, D. A. and Steitz, J. A. (1992), *Science*, **257**, 1918-1925

[156] Watkins, N.J., Segault, V., Charpentier, B., Nottrott, S. *et al.* (2000), *Cell*, **103**, 457-466

[157] Wersig, C. and Bindereif, A. (1992), *Mol. Cell. Biol.*, **12**, 1460-1468

[158] White, S.A., Hoeger, M., Schweppe, J.J., Shillingford, A. *et al.* (2004), *RNA*, **10**, 369-377

[159] Wijmenga, S.S. van Buuren B.N.M. (1998), *Prog. NMR Spectrosc.*, **32**, 287–387

[160] Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr, Morgan-Warren, R.J. *et al.* (2000), *Nature*, **407**, 327–39

[161] Wozniak, A., Nottrott, S., Kühn-Hölsken, E., Schröder, G. *et al.* (2005), *RNA*, **11**, 1545-1554

[162] Wu,J.A. and Manley,J.L. (1991), *Nature*, **352**, 818-821

[163] Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids, Wiley*

[164] Wyatt J R; Sontheimer E J; Steitz J A. (1992), *Genes Dev.*, **6**(12B), 2542-2553

[165] Yean, S.L., Wuenschell, G., Termini, J., Lin, R.J. (2000), *Nature*, **408**, 881-884

[166] Yuan, F., Griffin, L., Phelps, L., Buschmann, V. *et al.* (2007), *NAR*, **35**, 2833-2845

[167] Zimmer, D. P., Marino, J. P. and Griesinger, C. (1996), *Magn. Res. Chem.*, **34**, S177-S186

[168] Zhang, Q., Stelzer, A., Fisher, C. K. and Al-Hashimi, H.M. (2007), *Nature*, **450**, 1263-1267

[169] Zuker, M. and Stiegler, P. (1981), *NAR*, **9**, 133-148

[170] Zwahlen, C., Legault, P., Vincent, S. J. F., Greenblatt, J. *et al.* (1997), *JACS*, **119**, 6711-6721

A Appendix

A.1 Materials

A.1.1 Materials, Chemicals and Kits

Table A.1: Chemicals, materials and enzymes

Materials	Company
Superase-In, RNaseAlert Lab Test Kit^{TM}	Ambion, USA
ECL Western blotting reagents	Amersham Biosciences
Pf1 NMR co-solvent, RNase-free	ASLA^{biotech} , Latvia
centrifuge Bottle 50 ml Polyallomer with Caps, Bottle 1000ml PP	Beckman Coulter
acrylamid bisacrylamid solution 40% (19:1), ammonium heptamolybdat, bromphenol blue sodium salt, cobalt(II) chloride, coomassie brilliant blue (G250, R250), dithiothreitol (DTT), ethylene diamine tetraacetic acid (EDTA), ethidium bromide solution (1%), isopropyl-β-D-thiogalactopyranoside (IPTG), glutathione reduced, sodium dodecylsulfate (SDS), spermidine, tetra methyl ethylene diamine (TEMED), thiamine Hydrochloride, xylene cyanol FF	AppliChem, Germany
extra thick blot paper, gel filtration standard, Precision Plus Protein^{TM}, Quick Start^{TM} bradford dye reagent	Biorad, USA
ammonium chloride (^{15}N, 99%), d8-glycerol (99%)	Eurisotop, Germany
dNTP Set, T4 DNA ligase	Fermentas
Phusion High-Fidelity DNA Polymerase	Finnzymes, Finland
Tube-O-dialyzer MWCO 1000	G Biociences, USA
illustra GFX PCR DNA and Gel Band Purification Kit, PD-10 desalting column	GE Healtcare, UK
GSTrapFF (1ml, 5ml), HisTrapFF (1ml, 5ml), HiTrap DEAE FF (1ml, 5ml), HiTrap SP (1ml), Superdex 200 10/30 GL, HiLoad 16/60 Superdex 75 pg, HiLoad 26/60 Superdex 200 pg	GE Healthcare Biosciences
Thrombin	GE Healthcare Life science
E.coli DNA ligase, enterokinase	invitrogen, USA
acetic acid, ammonium chloride, ammoniumperoxodisulfat, boric acid, calcium chloride, dipotassium hydrogen phosphate, disodium hydrogen phosphate ethanol, isopropanol, glycine, glycerol, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), lysozyme, magensium chloride, methanol, potassium dihydrogen phosphate, potassium chloride, sodium chloride, urea, water (HPLC grade)	Merck, Germany

continued on next page

continued from previous page

Materials	Company
Amicon Ultra4 & Ultra15 (MWCO 3 kDa, 10 kDa), Microcon (YM3, YM10), Millipore Express^{TM} Plus, 0.22 μm, Millex-HV (PVDF, 0.45 μm)	Millipore, Ireland and USA
DNA ladder (100bp, 1kb), *EcoRI*, XhoI, PstI, *HindIII*, NdeI, NcoI, inorganic pyrophospatase,	New England Biolabs, USA
RNase-free tubes (for ÄKTA)	Neolabs
pET30 EK/LIC vector kit, pET41 EK/LIC vector kit	Novagen, USA
complete ^{TM} EDTA-free Protease Inhibitor Tablets	Roche Diagnostics
RQ1 RNase-free DNase	Promega, USA
Ni-NTA Agarose, Penta-His HRP conjugate, polypropylene column, QIAprep spin miniprep kit, QIAEX II gel extraction kit, Qiagen plasmid mega kit	Qiagen, Germany
ATP, CTP, GTP, UTP, copper chloride, lithium chloride, milk powder, Roti-chloroformisoamyl alcohol (24:1), Roti-aqua-phenol, Rotiphorese Gel 30 (37.5:1), Spectra/POR dialysis membran (MWCO 1kDa, 3kDa, 10kDa), Tris hydrochloride, zinc sulfate	Roth, Germany
agarose, ampicillin, β mercapto ethanol, bovine serum albumin, deuterium oxide (99.9%, 99.96%), ferrum(II) sulfate, guanidin hydrochloride, igepal CA-630, imidazole, kanamycine, manganese(II) chloride, magnesium sulfate, trizma base, triton X-100, tween 20	Sigma, Germany
E.coli OD 2 D, 2H labeled (98%) media	Silantes, Germany
Shigemi	Shigemi Inc.
deuterium oxide (99%), ^{13}C-,^{15}N- rATP (98%), ^{13}C-,^{15}N- rCTP (98%), ^{13}C-,^{15}N- rGTP (98%), ^{13}C-,^{15}N- rUTP (98%)	Spectra Stable Isotopes, USA
Slide-A-Lyzer (2K, 3.5K, 10K)	Thermo Scientific, USA
NMR tubes WG-5-7E, 5mm NMR tube caps	Wilmad Lab Glass, USA
Elutrap BT1, BT2 membranes	Whatman, USA

A.1.2 Instruments

Table A.2: Laboratory equipment

	Name	Company
centrifuge	Avanti J-30I with JLA 8.1000, JA 30.50 Ti	Beckmann-Coulter
	Allegra 64R centrifuge with F0650	Beckman-Coulter
	centrifuge 5415D	eppendorf
	centrifuge 5804	eppendorf
HPLC	ÄKTApurifer ÄKTAprime	GE Healthcare life science
power supply for electrophoresis	PowerPac 1000 & 3000 & universal & HV	Bio-Rad
	Gibco BRL PS9009	Gibco
	EPS600 & ECPS 3000/150P	Pharmacia Biotech
lyophile	Christ Alpha 2-4 Christ beta 1-8	B. Braun Biotech Int.
concentrator	concentrator 5301	eppendorf
incubator	Certomat R with Certomat HK	B. Braun Biotech Int.
	ISF-1-W incubator	Kühner shaker
	B15	Thermo Scientific
PCR	HYBAID PCRSprint Thermal Cycler	Thermo Scientific
	Primus25advanced	Peqlab, Germany
UV spectrophotometer	DU-70	Beckman
	Ultraspec 2100pro	Amersham Biosciences
	nanodrop ND-1000	Peqlab, Germany
heating	Heizblock neoBlock 1 2-2503	NeoLab
	Heizblock Typ SON-DA	Gebrüder Liebisch
	Thermomixer compact	eppendorf
	water bath Haake DC 10	Thermo electron corporation
pH meter	pH-meter 766 Calimatic	Knick
	pH-meter Seveneasy	Mettler Toledo
	with InLab 423 pH electrode	Mettler Toledo
	with inLab Micro	Mettler Toledo
cell disruption	Branson sonifer W-250	G.Heinemann
	Emulsi-Flex-C3 high pressure homogenizer	Avestin
water supply	Milli-Q-water supply apparatus	Millipore, USA
balances	L2200S-D R160 analytical	Sartorius, Germany
shaker	KL2 shaker	Edmund Bühler GmbH
pipetts	P2N, P10N, P100N, P200N, P1000N	Gilson
	multipipette plus	eppendorf
	pipetboy acu	IBS integra bioscience

Table A.3: NMR spectrometers

	Instrument	Company
NMR	900 MHz spectrometer (21,1 Tesla) with avance console	Bruker
	HCN cryo-probehead	Bruker Biospin
	800 MHz spectrometer (18,8 Tesla) with DRX console	Bruker
	HCN cryo-probehead	Bruker Biospin
	800 MHz spectrometer (18,8 Tesla) with avance III console	Bruker
	HCN cryo-probehead	Bruker Biospin
	700 MHz spectrometer (16,4 Tesla) with avance console	Bruker
	HCP cryo-probehead	Bruker Biospin
	HCN probehead with xyz-gradients	Bruker Biospin
	600 MHz spectrometer (14,0 Tesla) with avance console	Bruker
	HCN cryo-probehead	Bruker Biospin
	600 MHz spectrometer (14,0 Tesla) with avance III console	Bruker
	HCN cryo-probehead	Bruker Biospin
	600 MHz spectrometer (14,0 Tesla) with DRX console	Bruker
	with HCN probehead with z-gradients	Bruker Biospin
	500 MHz spectrometer (14,0 Tesla) with DRX console	Bruker
	with HCN probehead with xyz-gradients	Bruker Biospin

A list of software packages used for processing and analyzing NMR experiments, structural calculation, structure viewing and chromatography is given in table A.4.

Table A.4: Software packages

Name	Company
Topspin 1.3 / 2.0 / 2.1	Bruker, Karlsruhe, Germany
X-WINNMR 3.5	Bruker, Karlsruhe, Germany
FELIX-ND (2000.1 / 2007)	Accelrys, San Diego, USA
ARIA	[76], www.pasteur.fr/recherche/unites/Binfs/aria/
CNS	[8], http://cns.csb.yale.edu/v1.1
UCSF Chimera	[103], www.cgl.ucsf.edu/chimera/
PyMOL	DeLano Scientific LLC
MOLMOL	[66]
Unicorn^{TM}	GE Healthcare Life science

A.1.3 Oligonucleotides

All DNA oligonucleotides used for cloning, PCR and transcription were ordered from IBA (Göttingen, Germany) and are presented in Table A.5.

Table A.5: *DNA templates used for cloning, PCR and RNA synthesis.* | Oligonucleotides used directly as DNA template for transcription reactions are named 'RNA synthesis'. For mutational studies of the lariat-forming ribozyme more than 50 mutations were done. Here an example mutation sequence of *spliclin_A50C* is illustrated. Other mutations are depicted in figure 5.2b and in table 5.1. Oligonucleotides cloned into *pUC19* to yield self-produced DNA templates for transcription are named 'cloning'. Oligonucleotides used for PCR reactions are named 'PCR primer'. (HH = hammerhead, bot = bottom, letters in blue: transcription start; in red: mutation site)

Method	Name	Sequence
RNA synthesis	U4 I1 top	5′ TAATACGACTCACTATAGCCGAGGCGCGATCGTCTGTCCTGCAGAAGCT 3′
	U4 I1 bot	5′ AGCTTCTGCAGGACAGACGATCGCGCCTCGGCTATAGTGAGTCGTATTA 3′
	U4 I2 top	5′ TAATACGACTCACTATAGATCGTAGCCAATGAGGTTGTCTGTCCTGCAGA AGCTTCCCGGGGATC 3′
	U4 I2 bot	5′ GATCCCCGGGGAAGCTTCTGCAGGACAGACAACCTCATTGGCTACGATCT ATAGTGAGTCGTATTA 3′
	spliclin_A50C_top	5′ TAATACGACTCACTATAGGAAGAAAGGGCTTCGGCCACTCAAACTACAG AGACGCCAGTCACTCAGCTATCCTGGT 3′
	spliclin_A50C_bot	5′ ACCAGGATAGCTGAGTGACTGGCGTCTCTGTAGTTTGAGTGGCCGAAGC CCTTTCTTCCTATAGTGAGTCGTATTA 3′
	spliclin_29mer_top	5′ TAATACGACTCACTATAGGAAGAAAGGGCTTCGGCCACTCAAACTA3′
	spliclin_29mer_bot	5′ TAGTTTGAGTGGCCGAAGCCCTTTCTTCCTATAGTGAGTCGTATTA 3′
cloning	HH spliclin top	5′ AATTCTAATACGACTCACTATAGGCAGACCTGATGAGCTGAGGCTTGCCT CAGCGAAACCAGGACTGCAGA 3′
	HH spliclin bot	5′AGCTTCTGCAGTCCTGGTTTCGCTGAGGCAAGCCTCAGCTCATCAGGTCT GCCTATAGTGAGTCGTATTAG 3′
	spliclin_59mer top	5′ AATTCTAATACGACTCACTATAGGAAGAAAGGGCTTCGGCCACTCAAAC TACAGAGACGCCAGTCACTCAGATATCCTGGTCTGCCTGCAGA 3′
	spliclin_59mer bot	5′ AGCTTCTGCAGGCAGACCAGGATATCTGAGTGACTGGCGTCTCTGTAGT TTGAGTGGCCGAAGCCCTTTCTTCCTATAGTGAGTCGTATTAG 3′
	HH-I1 top	5′ AATTCTAATACGACTCACTATAGGCGGCAGACCTGATGAGCTGAGGCTT GCCTCAGCGAAATCGCGCTTCTGCAGA 3′
	HH-I1 bot	5′ AGCTTCTGCAGAAGCGCGATTTCGCTGAGGCAAGCCTCAGCTCATCAGG TCTGCCGCCTATAGTGAGTCGTATTAG 3′
	HH-I2 top	5′ AATTCTAATACGACTCACTATAGGCGGCAGACCTGATGAGCTGAGGCTT GCCTCAGCGAAACCTCATTCTTCTGCAGA 3′
	HH-I2 bot	5′ AGCTTCTGCAGAAGAATGAGGTTTCGCTGAGGCAAGCCTCAGCTCATC AGGTCTGCCGCCTATAGTGAGTCGTATTAG 3′
PCR primer	15.5K-f-LIC	5′ GACGACGACAAGATGACTGAGGCTGATGTGA 3′
	15.5K-r-LIC	5′ GAGGAGAAGCCCGGTTAGACTAAGAGCCTTTCAATG 3′
	hPrp31^{215}-NdeI	5′ CT CTA TCG CAT ATG ATT GCA CCG AAT CTG AGC ATT 3′
	hPrp31^{499}-XhoI	5′ CA GTA TCG CTC GAG GGT GCT CAT CAG A 3′
	hPrp31^{216}-NcoI	5′ CT CTA TCG CCA TGG GCA CCG AAT CTG AGC ATT ATT A 3′
	hPrp31^{209}-NcoI	5′ CT CTA TCG CCA TGG GAA AGC CGC ATG AGC TTT A 3′

A.1.4 Bacterial strains and plasmids

Table A.6 listed all strains that were used in this work. DH5α was used for transformations of pET30-15.5K and pET41-15.5K. Plasmids were purified from DH5α and subsequently transformed into protein expression strains Bl21(DE3) and RosettaII. Although both strains lead to sufficient protein yields, the expression system Rosetta II was used in this work for further studies. All other vectors listed in table A.7 were transformed into Bl21(DE3). Nova blue giga cells were used for transformations after Ligation independent cloning.

Table A.6: Bacterial strains

Name	Company	Selection	Source
DH5α	non-expression host, general purpose cloning, plasmid propagation	none	Life Technologies
Nova blue giga cells	non-expression host, general purpose cloning, high-efficiency transformation	none	Novagen
Bl21(DE3)	general purpose expression host	none	Novagen
Rosetta II	expression host; expression of genes with rare codons	Cam	Novagen

Table A.7: Plasmids used in this work

Name	Description	Selection	Source	Restriction sites
pUC19	high copy number plasmid	Amp	NEB	
pET-30 EK/LIC	expression vector with N-(EK) & C-His Tag	Kan	Novagen	
pET-41 EK/LIC	expression vector with N-GST Tag, N-(EK) & C-His Tag	Kan	Novagen	
pET-28a	expression vector with N-(Thr) & C-His Tag	Kan	Novagen	
pETM11	expression vector with N-(TEV) & C-His Tag	Kan	EMBL G.Stier	
pUC19-U4-I1	template for RNA synthesis of U4 snRNA I1 (see ...)	Amp	this work	*EcoRI* & *HindIII*
pUC19-U4-I2	template for RNA synthesis of U4 snRNA I2 (see ...)	Amp	this work	*EcoRI* & *HindIII*
pUC19-U4-HH I1	template for RNA synthesis of hammerhead RNA for U4 snRNA I1 cleavage	Amp	this work	*EcoRI* & *HindIII*
pUC19-U4-HH I2	template for RNA synthesis of hammerhead RNA for U4 snRNA I2 cleavage	Amp	this work	*EcoRI* & *HindIII*

continued on next page

continued from previous page

Name	Description	Selection	Source	Restriction sites
pGEX-15.5K	expression of GST-15.5K-wt cleavage with Thrombin	Amp	S.Nottrott [98]	
pET30-15.5K	expression of His-15.5K-wt cleavage with Enterokinase	Kan	this work	ligation independent cloning
pET41-15.5K	expression of GST-His-15.5K-wt cleavage with Enterokinase	Kan	this work	ligation independent cloning
pET41-hPrp31	expression of wt-hPrp31-His, sequence is optimized for expression in *E.coli*	Kan	GENEART	*NdeI* & *XhoI*

A.1.5 Buffer and solutions

Prepared solutions were autoclaved or sterile filtrated.

Table A.8: Buffers and solutions for purification of proteins

Name	Description	Selection	Source	Remark
purification of His-15.5K	lysis buffer	50 mM 500 mM 10 mM 2 mM	Tris HCl pH 7.5 NaCl Imidazole β-mercaptoethanol	pH 7.5
	high salt washing buffer	50 mM 1 M 10 mM	Tris HCl pH 7.5 NaCl Imidazole pH 8.0	pH 7.5
	LiCl washing buffer	50 mM 2 M	Tris HCl pH 7.5 LiCl	pH 7.5
	elution buffer	50 mM 500 mM 300 mM 2 mM	Tris HCl pH 7.5 NaCl Imidazole pH 8.0 β-mercaptoethanol	pH 7.5
	NMR buffer	20 mM 120 mM 1 mM	HEPES pH7.6 NaCl DTT	pH 7.6
purification of His-hPrp31	lysis buffer	20 mM 150 mM 0.2 % (w/v) 10 mM	Tris HCl pH 8.0 NaCl Igepal Imidazole	
	washing buffer 1	20 mM 1 M 10 mM	Tris HCl pH 8.0 NaCl Imidazole pH 8.0	
	washing buffer 2	20 mM 150 mM 10 mM	Tris HCl pH 8.0 NaCl Imidazole pH 8.0	
	washing buffer 3	20 mM 150 mM 50 mM	Tris HCl pH 7.5 NaCl Imidazole pH 8.0	
	elution buffer	20 mM 150 mM 250 mM	Tris HCl pH 7.5 NaCl Imidazole pH 8.0	
	NMR buffer	20 mM 120 mM 1 mM	HEPES NaCl DTT	pH 7.6

Table A.9: Buffer and solutions for *in vitro* RNA synthesis

Method	Name	Amount	Chemical	Remark
purification of T7 Polymerase	lysis buffer	50 mM 300 mM 2 mM up to 250 ml	$NaPO_4$ pH 8.0 NaCl β-mercaptoethanol H_2O	
	washing buffer	50 mM 300 mM 10 mM 2 mM up to 250 ml	$NaPO_4$ pH 8.0 NaCl Imidazole pH 8.0 β-mercaptoethanol H_2O	
	elution buffer	50 mM 300 mM 250 mM 2 mM up to 50 ml	$NaPO_4$ pH 8.0 NaCl Imidazole pH 8.0 β-mercaptoethanol H_2O	
	storage buffer	40 mM 200 mM 0.2 mM 0.2 mM up to 1 l	$NaPO_4$ pH 7.7 NaCl EDTA pH 8.0 DTT H_2O	
RNA synthesis	10 x transcription buffer	400 mM 10 mM 50 mM 0.1% (v/v) up to 100 ml	Tris-HCl spermidine DTT Triton X-100 H_2O	pH 8.0 stored at -20°C
	10 x TBE buffer	108 g 55 g 20 ml up to 1 l	Tris base boric acid 0.5M EDTA pH 8.0 H_2O	
	2 x loading buffer	0.5% 0.5% up to 50 ml	bromphenol blue xylene cyanol FF H_2O	
	20% den. PAGE	420.42 g 500 ml 100 ml up to 1 l	urea acrylamide bisacrylamide solution 40 % (19:1) 10x TBE buffer H_2O	ca. 4 weeks stable
	15% den. PAGE	420.42 g 375 ml 100 ml up to 1 l	urea acrylamide bisacrylamide solution 40 % (19:1) 10x TBE buffer H_2O	ca. 4 weeks stable
	RNA gel mix for 10 ml	10 ml 80 μl 4 μl	den. PAGE mix 10 % APS TEMED	example in cm 20x20x0.1 = 40 ml 53x43x0.3 = 700 ml
	12% native PAGE	15 ml 5 ml 35 ml 360 μl 18 μl	acrylamide bisacrylamide solution 40 % (19:1) 10x TBE buffer H_2O 10 % APS TEMED	for 20x20x0.1
NMR buffer	U4 snRNA	20 mM 120 mM	HEPES NaCl	pH 7.6
	lariat-forming ribozyme	20 mM	$NaPO_4$-buffer	pH 6.6

Table A.10: Buffer and solutions for expression of unlabeled, ^{13}C-, ^{15}N- and ^{2}H-labeled proteins

Method	Name	Amount	Chemical	Remark
protein expression	LB media (unlabeled proteins)	10 g 5 g 5 g up to 1 l	bacto tryptone yeast extract NaCl H_2O	pH 7.2 made by EMBL kitchen
	L-Agar	10 g 5 g 10 g up to 1 l 15 g	bacto tryptone yeast extract NaCl H_2O Agar	pH 7.2 made by EMBL kitchen
	ampicillin	100 mg/ml	ampicillin sodium salt	stored at -20°C
	kanamycin	50 mg/ml	kanamycin	stored at -20°C
	chloramphenicol	34 mg/ml	chloramphenicol	stored at -20°C
	lysozyme	100 mg/ml	lysozyme	stored at -20°C
	IPTG	1M	IPTG	stored at -20°C
	10 x M9 salts	67.8 g 30 g 5 g up to 1 l	Na_2HPO_4 KH_2PO_4 NaCl H_2O	pH 7.4, sterile filtrate
	100 x trace elements	600 mg 115 mg 80 mg 70 mg 30 mg 2 mg 25 mg 500 mg	$FeSO_4$*7H_2O $MnCl_2$*4H_2O $CoCl_2$*4H_2O $ZnSO_4$*7H_2O $CuCl_2$*2H_2O H_3BO_3 $(NH_4)_6MO_7$*2H_2O stir for 10 min EDTA stir over night until golden	step by step in 100 ml H_2O or D_2O, sterile filtrate
	minimal media, in H_2O (labeled proteins)	858 ml 10 ml 6 ml 100 μl 2 ml 20 ml 4 ml 100 ml	H_2O 100x trace elements 5 mg/ml thiamin-HCl 1M $CaCl_2$ 1M $MgCl_2$ 20 % (^{13}C)-glucose 0.25 g/ml (^{15})NH_4Cl 10x M9 salts	sterile filtrate
	minimal media in 99 % D_2O (deuterated proteins)	704 ml 9 ml 5.4 ml 90 μl 1.8 ml 18 ml 3.6 ml 90 ml 100 ml	H_2O 100x trace elements 5 mg/ml thiamin-HCl 1M $CaCl_2$ 1M $MgCl_2$ 20 % d8-glycerol 0.25 g/ml NH_4Cl 10x M9 salts sterile filtrate E.coli OD 2 D 2H labeled (98 %) media	all stocks in D_2O

Table A.11: Buffer and solutions for standard biochemical methods

Method	Name	Amount	Chemical	Remark
SDS-PAGE	separating gel 15 %	12.5 ml	rotiphorese gel 30	for 6 gels
		6.25 ml	Tris HCl pH 8.8	
		5.75 ml	H_2O	
		250 μl	10 % SDS	
		250 μl	10 % APS	
		10 μl	TEMED	
	separating gel 20 %	16.7 ml	rotiphorese gel 30	for 6 gels
		6.25 ml	Tris HCl pH 8.8	
		1.55 ml	H_2O	
		250 μl	10 % SDS	
		250 μl	10 % APS	
		10 μl	TEMED	
	stacking gel 5 %	2 ml	rotiphorese gel 30	for 6 gels
		1.5 ml	Tris HCl pH 6.8	
		8.26 ml	H_2O	
		120 μl	10 % SDS	
		120 μl	10 % APS	
		6 μl	TEMED	
	4 x protein loading buffer	7.5 ml	1M Tris HCl pH 6.8	
		17 ml	10 % SDS	
		23 ml	glycerin	
		50 mg	brom phenol blue	
		500 μl	β-mercaptoethanol	
	10 x running buffer	151 g	Tris base	pH should be around 8.3
		720 g	glycine	
		23 ml	glycerin	
		500 ml	10 % SDS	
		up to 5 l	H_2O	
	staining solution	2 g	Coomassie brilliant blue R250	
		0.5 g	Coomassie brilliant blue G250	
		50 ml	methanol	
		450 ml	ethanol	
		100 ml	acetic acid	
		400 ml	H_2O	
	destaining solution	1.25 l	isopropanol	
		0.5 l	acetic acid	
		3.25 l	H_2O	
Western blot	transfer buffer	0.37 g	SDS	
		2.92 g	glycine	
		5.85 g	Tris base	
		100 ml	methanol	
		up to 1 l	H_2O	
	1 x PBS	0.2 g	KCl	make by EMBL kitchen
		1.15 g	Na_2HPO_4	
		0.2 g	KH_2PO_4	
		8	NaCl	
		up to 1 l	H_2O	
	blotting buffer	50 ml	1 x PBS	make fresh
		2.5 g	milk powder	
	washing buffer	1 l	1 x PBS	
		1 g	Tween 20	

A.2 Molecular standard methods

A.2.1 PCR amplification

Polymerase Chain Reaction (PCR) is a very sensitive method of amplifying specific nucleic acids. The reaction was performed enzymatically by using short oligo primers and free deoxynucleotides in replicative cycle of denaturing, annealing and extending. In this work PCR was used as tool to produce DNA templates in sufficient amounts for later performed plasmid construction. Forward and reverse primers were designed to yield annealing temperatures around 55 to 60°C (see table A.5). For ligation dependent cloning compatible restriction enzyme sites were incorporated and additional bases were added at the 5′ ends to allow efficient digestion of the restriction enzymes. For ligation independent cloning (LIC) the very specific 13 to 14 nucleotide single-stranded overhangs were used for primer design (see table A.5). As thermostable DNA polymerase the highly accurate and processive Phusion Polymerase has been chosen. The PCR reaction mixtures were:

10 μl	5 x HF buffer
1μl	20 mM dNTPs
0.5μl	Phusion
1μl	10 pmol/μl Primer 1
1μl	10 pmol/μl Primer 2
up to 50 μl	with nuclease free water

and cycling conditions as follows:

$$\begin{array}{rl} 98°C & 4\ \text{min} \\ 30\ \text{x} \left\{ \begin{array}{ll} 98°C & 10sec \\ 56°C & 30sec \\ 72°C & 1min \end{array} \right. & \\ 72°C & 7\ \text{min} \end{array}$$

A.2.2 Agarose gel electrophoresis and DNA fragment isolation

For analysis and purification of plasmid DNA and PCR products a 1 % agarose gel was used with 1 kb and 100 bp DNA ladder as marker. Electrophoresis was performed with 5 V/ cm and 1 x TBE buffer (see table A.9) was used as running buffer. The gel was stained afterwards with ethidium bromide and fragments of interest were visualized with UV light, excised and purified from the gel matrix by the use of QIAEX II gel extraction kit, illustra GFX PCR DNA & gel band purification kit or QIAprep spin miniprep kit. DNA concentration was determined by agarose gel or with the nanodrop spectrophotometer.

A.2.3 Enzyme digestion

Vector and PCR-product for ligation dependent cloning were digested with the respective restriction enzymes for 3 to 12 hours at 37°C to generate sticky ends. Thereby 2 to 10 units enzyme per μg plasmid were used with a total DNA concentration of about 0.1 to 0.2 μg /μl. DNA was purified over night on a 30 cm long 1 % agarose gel with a constant voltage of 80 V.
Plasmids for RNA synthesis were digested with *PstI*. The *PstI* site was directly incorporated at the end of the transcription region forcing T7 polymerase to stop transcription at this position. The digestion was performed with 2 units enzyme per μg plasmid for about 3 hours at 37°C. The total concentration of DNA in the reaction mixture were 0.5 to 1 μg /μl. DNA was purified by phenol extraction and precipitated afterwards with 0.9 times isopropanol and 0.3 M sodium acetate.

A.2.4 Ligation

Using the ligation dependent cloning method DNA pieces generated before with compatible DNA ends are joined together by DNA ligase. For the ligation of DNA templates into *pUC19-EcoRI-HindIII* (see table A.7), 40 pmol of hybridization mix (see table A.5) was combined in a 20 μl reaction with 5 U of T4 DNA ligase, 10 mM ATP and 50 ng of linearized vector.
Ligation-independent cloning (LIC) was developed for directional cloning of PCR products without restriction enzyme digestion. Thereby the LIC method takes advantage of the 3′ $\rightarrow$ 5′ exonuclease activity of T4 DNA polymerase to create very specific 13 to 14 nucleotide single-stranded overhangs in the vector and the insert. Afterwards annealing delivers an insert insertion into the vector of interest. The method was performed as described in the Novagen EK/LIC cloning Kit user protocol.

A.2.5 Competent cells

For the uptake of extracellular DNA cells are treated with cold $CaCl_2$ to make their cell membrane transiently permeable to DNA. Therefore 50 ml LB media is inoculated with 500 μl of an over night culture. Cell were grown until an OD_{600nm} of 0.8 were reached and then harvest at 1000 x g and 4°C for 10 min. Cell pellet was carefully resuspended with 40 ml ice cold 100 mM $CaCl_2$. Incubation of cells for 30 to 40 min on ice followed before centrifugation was performed at 1000 x g at 4°C for 10 min. Pellet was resuspended in 2.5 ml of cold 100 mM $CaCl_2$ and the end volume of the suspension was determined. By the use of ice cold 86 % glycerol an end concentration of 15 % glycerol was adapted to the cell suspension which was stored at -80°C in 200 μl aliquots.

A.2.6 Transformation by heat shock

Competent cells were thawed on ice and DNA or ligation reaction to be transformed was added to the competent cells. The cell mixture was incubated on ice for 20 to 30 minutes and then at 42°C for 90 seconds for heat shock. 900 μl of LB medium was added and cells were shaken at 37°C for one hour before spreading on LB plates containing appropriate antibiotic. The plates were incubated at 37°C overnight.

A.2.7 Plasmid purification & sequencing

The plasmid DNA is obtained by miniprep procedure using QIAprep spin miniprep kit. Therefore a single colony was inoculated into 3 ml LB liquid medium containing appropriate antibiotic and grown overnight at 37°C with shaking. 2 ml of overnight culture was used to extract plasmid as described in the manufacturers protocol.
Sequencing of purified plasmid DNA was performed via the extended Hot Shot DNA sequencing service of Seqlab (Göttingen, Germany) or the sequence service of the genomic core facility (EMBL, Heidelberg). Furthermore a -80°C glycerol stock was prepared by mixing 500 μl cell culture with same volume of 40 % glycerol.
For the Hammerhead RNA synthesis performed in this work the template DNA was obtained by growing 1 l cultures over night at 37°C. The purification of plasmid was done with plasmid Maxi Kit, followed by the *PstI* digest (see above).

A.2.7.1 SDS-PAGE

Denatured SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used to analytically separate proteins according their size on 15 % or 20 % polyacrylamide gels (see table A.11) of 1 mm thickness. Before loading, the protein samples were boiled with protein loading buffer at 95 °C for 2 min to ensure a complete denaturation. Electrophoresis was performed with 1x SDS running buffer (see table A.11) and 25 mA per gel in a Biorad Mini Protean 3 chamber.

A.2.7.2 Westernblot

Protein blotting is an analytical method that involves the immobilization of proteins on membranes before detection using an antibody probe. In western blotting also known as immunoblotting, prior to protein immobilization on a nitrocellulose membranes, sample proteins are separated using SDS-PAGE.
The gel, membrane and sheets of extra thick blot paper were soaked in transfer buffer (see table A.11) and placed in the respective order into the Trans Blot SD Semi-dry transfer cell. The transfer of the proteins to the membrane in the blotting step was performed for 40 min at 10 V. The use of pre-stained MW standards indicated a successful transfer. To decrease non-specific binding of antibody, the membrane was blocked with freshly prepared blocking buffer (see table A.11) for one hour at room temperature on a rocking platform. Afterwards the membrane was washed 3 times with washing buffer (see table A.11) before incubation with the antibody against the protein of interest followed for one hour. The dilution of the antibody penta His conjugate was set to 1:2000 in blocking buffer. The membrane was again washed 3 times and then developed with enhanced chemiluminescence (ECL) detection reagents. Equal volumes of detection reagents 1 and 2 were mixed and applied to the membrane. After incubation for 60 to 90 seconds at room temperature, the membrane was drained, wrapped in plastic film and exposed to X-ray film for 5 seconds to 5 minutes.

A.3 Pulse programs and NMR parameters

This following section contains pulse programs as well as necessary acquisition parameters for the 2D- and 3D- NMR experiments (see table A.12 to A.15) which had been used in this work on the various RNA samples.

——————————————— H1′C1′ HSQC———————————————

```
;hsqc_sug_h1c1.txt
;hsqc with watergate
;J(C5-C6) refocused with inversion shaped pulse.
;For the determination of evolution
; of coupling to C during watergate (3.37*d19),
;see: V.Sklenar et col,
:     JMR Ser.A, 1993, 102, 241-245

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=4u"
"d4=1s/(cnst2*4)"  ;172 Hz (C1r and C5)
"d11=30ms"
"d12=20u"
"d20=d0*2-4u"
"d23=p3-p1"
"d14=d4-p16-d16-50u"

"CEN_CN1=(p12-p22)/2"
;specific for decoupled version:
"CEN_HC1=(p12-p2)/2"

1 ze
  20u pl3:f3 pl12:f2
2 d11 do:f3 do:f2
  4u pl9:f1
  d1 cw:f1
  20u do:f1
    50u UNBLKGRAD
  4u pl1:f1

;####### INEPT
  (p1 ph1)
  d14 pl2:f2
      50u
      p16:gp3
      d16
  (d23 p2 ph1):f1  (p4 ph6):f2
      50u
      p16:gp3
      d16
  d14
  (p1 ph2)
      50u
      p16:gp1
      d16
  (p3 ph3):f2

;####### Evolution 13C

  d0
  d12 pl20:f2

;p2 = specific for decoupled version:
  (CEN_HC1 p2 ph1):f1
   (CEN_CN1 p22 ph1):f3
    (p12:sp2 ph1):f2

  4u
  (p14:sp4 ph1):f2
  d0
  (p13:sp3 ph5):f2  ; on C1',C5
  d20
  (p12:sp2 ph1):f2 ; on C6
  4u
  (p14:sp4 ph1):f2 ; on C2'-C5'
  4u
  d12 pl2:f2

;####### INEPT

  (p3 ph4):f2
      50u
```

```
      p16:gp4
      d16
  (p1 ph1)
   d14 pl2:f2
      50u
      p16:gp2
      d16
  (d23 p2 ph1):f1  (p4 ph6):f2
      46u
      p16:gp2
      d16
      4u BLKGRAD
      d14 pl12:f2

  go=2 ph31 cpd2:f2
  d11 do:f3 do:f2 mc #0 to 2 F1PH(ip3, id0)
exit

ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=0 0 2 2
ph6=0
ph12=0
ph31=0 2 2 0

;default power levels:
;pl1 : f1 channel
;pl2 : f2 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel
;pl20: f2 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree

;p16: homospoil/gradient pulse
;cnst2 : J(CH) = 160 Hz for base C6/C8
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I / O [30 ms]
;d13: short delay [3 usec]
; (e.g. to compensate delay line)
;d16: delay for homospoil/gradient recovery
;d24: 1/(4J)XH for XH
;     1/(6J)XH for all multiplicities
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2 (for old topspin versions)
;FnMODE: States-TPPI

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;      [GARP for this HSQC]

;gradient: SINE.100
;gpz1: 40%
;gpz2: 30%
;gpz3: 25%
;gpz4: 50%
;most of the gpz can be altered
;because they are not related among them

;notes:
;Notice that d0 initial is not calculated in order to
; have controlled phase in F1.
;Instead, it is compensated by the pi pulse plus d20,
; which make vanish the net evolution of the M
; in the first increment.

;The shaped pulses allow the refocusing of the J-CC.
;The third sp hits C5 but does not do anything.
;Simply compensates the Bloch-Siegert phase shifts.
; Possibly, it inverts the M on C5,
; but this M will not become detectable.

;Calculation of d20.
; set equal delays before and
; after the inversion pulse preceeding d20.
; before: d0 + d12 + p12 + d0
; after: d20 + p12 + 4u + d12
```

H6C6/H8C8 HSQC

```
;hsqc_base.txt
;adapted from vshchsqcwg
;hsqc with watergate

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"in0=inf1/2"

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=4u"
"d4=1s/(cnst2*4)"  ;180 Hz for base in RNA
"d11=30ms"
"d13=20u"
"d20=d0*2-4u"
"d23=p3-p1"
"d14=d4-p16-d16-50u"

"CEN_HC1=(p12-p2)/2"

1 ze
  20u pl3:f3 pl12:f2
2 d11 do:f3 do:f2
  4u pl9:f1
  d1 cw:f1
  20u do:f1
      50u UNBLKGRAD
  4u pl1:f1
  4u pl2:f2

;####### INEPT

  (p1 ph1)
  d14 pl2:f2
      50u
      p16:gp3
      d16
  (d23 p2 ph0):f1  (p4 ph6):f2
      50u
      p16:gp3
      d16
  d14
    (p1 ph2)
      50u
      p16:gp1
      d16
  (p3 ph3):f2

;####### Evolution 13C

  d0
  d12 pl20:f2
  ;(CEN_CN1 p22 ph1):f3
  (p12:sp2 ph0):f2 (CEN_HC1 p2 ph0)
   d0
  (p13:sp3 ph5):f2 ; on C6,C8
  d20
  (p12:sp2 ph0):f2 ; on C5
  4u
  d12 pl2:f2

;####### INEPT

  (p3 ph4):f2
      50u
      p16:gp4
      d16
  (p1 ph0)
   d14 pl2:f2
      50u
      p16:gp2
      d16
  (d23 p2 ph0):f1  (p4 ph0):f2
      46u
      p16:gp2
      d16
      4u BLKGRAD
      d14 pl12:f2
  (p1 ph13):f1

  go=2 ph31 cpd2:f2
  d11 do:f3 do:f2 mc #0 to 2 F1PH(ip3, id0)
exit

ph0=0
ph1=0 0 0 0 2 2 2 2
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=0 0 2 2
ph6=0
ph12=0
ph13=2
ph31=0 2 2 0 2 0 0 2

;default power levels:
;pl1 : f1 channel
;pl2 : f2 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree

;p16: homospoil/gradient pulse
;cnst2 : J(CH) = 160 Hz for base C6/C8
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I / O [30 ms]
;d13: short delay      [3 usec]
; (e.g. to compensate delay line)
;d16: delay for homospoil/gradient recovery
;d24: 1/(4J)XH for XH
;     1/(6J)XH for all multiplicities
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2 (for old topspin versions)
;FnMODE: States-TPPI

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;      [GARP for this HSQC]

;gradient: SINE.100
;gpz1: 50%
;gpz2: 60%
;gpz3: 15%
;gpz4: 40%
;most of the gpz can be altered
;because they are not related among them

;notes
;Notice that d0 initial is not calculated in order to
; have controlled phase in F1.
;Instead, it is compensated by the pi pulse plus d20,
; which make vanish the net evolution of the M
; in the first increment.

;Calculation of d20.
; set equal delays before and
; after the inversion pulse preceeding d20.
; before: d0 + d12 + p12 + d0
; after: d20 + p12 + 4u + d12
```

H_sCN_b

```
;hcnsug.txt
;avance-version (00/10/16)
;HCN on ribose
;3D inverse correlation for triple
; resonance using multiple
;   inept transfer steps
;
;   F1(N,t1) -> F2(C1',t2) -> F3(H1')
;
;phase sensitive
;with decoupling during acquisition
;
;R. Fiala, F. Jiang, V. Sklenar
;   J. Biomol. NMR 12, 373 - 383 (1998)

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"
"in25=inf2/4"
"in22=in25"

aqseq 312

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=3u"
"d4=1.50m"  ; J(HC)=170Hz, d4=1/4J
"d11=30m"
"d13=p3"
"d15=d0*2+p14"
"d22=DELTA"
"d23=22.6m" ; J(CN)=11-12Hz, d23=1/4J
"d25=DELTA"
"d26=p12/2-p3"
"d27=p13/2-p3"

"DELTA=d23/4-d16-p16-p11/2"
"DELTA1=d4-p16-d16-4u"

"CEN_CN1=(p12-p13)/2"
"CEN_CN2=(p13-p17)/2"
"CEN_HC1=(p4-p2)/2"
"CEN_HC2=(p14-p2)/2"

1 ze
   d11 pl12:f2 ;pl16:f3
2 d11 do:f2 do:f3
3 d1 pl3:f3 pl2:f2 pl1:f1

;####### INEPT

  (p1 ph1)
  d13
  d4
  (CEN_HC1 p2 ph1):f1 (p4 ph1):f2
  d4 UNBLKGRAD
  (p3 ph2):f2

;####### Transfer to Nitrogen

  d26
  DELTA pl20:f1
  p16:gp1
  d16 pl21:f2 pl22:f3
  (p11:sp1 ph11):f1
  d16
  p16:gp1
  DELTA
  (p12:sp2 ph1):f2 (CEN_CN1 p13:sp3 ph1):f3
  DELTA
  p16:gp1
  d16 pl2:f2
  d13
  (p11:sp1 ph11):f1
  d16 pl3:f3
  p16:gp1
  DELTA pl1:f1
  d26
  (p3 ph3):f2
  (p1 ph4):f1
  20u
  p16:gp2
  d16

;####### Evolution 15N

  (p21 ph5):f3
    d0
   (CEN_HC2 p2 ph1):f1 (p14:sp4 ph1):f2
    d0
    (p22 ph1):f3
    d15
  (p21 ph6):f3
  20u
  p16:gp3
  d16 pl2:f2

;####### Transfer to Carbon

  (p1 ph7):f1
  (p3 ph8):f2

;####### Evolution 13C

  d27
  d22 pl20:f1
  p16:gp4
  d16 pl21:f2 pl22:f3
  (p11:sp1 ph11):f1
  d16
  p16:gp4
  d22
  (CEN_CN2 p17:sp7 ph1):f2 (p13:sp3 ph1):f3
  ;(p12:sp2 ph1):f2 (CEN_CN1 p13:sp3 ph1):f3
  d25
  p16:gp4
  d16 pl2:f2
  (p11:sp1 ph11):f1
  d16 pl3:f3
  p16:gp4
```

```
  d25 pl1:f1
  d27

;####### INEPT

  (p3 ph9):f2
   4u
   DELTA1
   p16:gp5
   d16
   (CEN_HC1 p2 ph1):f1 (p4 ph1):f2
   d16
   p16:gp5
   DELTA1 pl12:f2 ;pl16:f3
   4u  BLKGRAD
   go=2 ph31 cpd2:f2 ;cpd3:f3
   d11 do:f2 do:f3 mc #0 to 2
     F1PH(ip5,id0)
     F2PH(rd0 & ip8, id22 & dd25)
exit


ph1=0
ph2=0 2
ph3=1
ph4=1
ph5=0 0 2 2
ph6=0
ph7=1
ph8=1
ph9=0 0 0 0 2 2 2 2
ph11=0
ph31=0 2 2 0 2 0 0 2

;default power levels:
;pl0 : 120db
;pl1 : f1 channel
;pl2 : f2 channel
;pl3 : f3 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel
;pl16: f3 channel
;shaped pulse 180 degree:
;sp0 : f2 channel (modulated)
;sp3 : f2 channel (C6/C8 on resonance)
;sp5 : f2 channel (C1' off resonance)
;sp9 : f3 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree

;p14: f2 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse [1 ms]
;p28: f1 channel - trim pulse [1 ms]
;p30: f3 channel - 180 degree shaped pulse
;    for inversion
;d0 : incremented delay (2D) [3 us]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(HC) (C1')) [1.6 ms]
;d11: delay for disk I/O  [30 ms]
;d13: short delay  [4 us]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;     d19 = (1/(2*d)),
;     d = distance of next null (in Hz)

;d22: 1/(4J(CN))
;        C1':  [18 ms]
;d23: 1/(4J(CN))
;        C1' and C6/C8: [17 ms]
;d25: 1/(4J(CN))
;        C6/C8: [16 ms]
;d27: 1/(4J(HC)) (C6/C8)
;        for H6/H8:  [1.25 ms]
;in0: 1/(2 * SW(H)) =  DW(H)
;nd0: 2 (for old topspin versions)
;NS: 8 * n
;DS: > = 32
;td1: number of experiments in F1
;FnMODE: States-TPPI (or TPPI)

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;cpd3: according to sequence defined by cpdprg3
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel
;pcpd3: f3 channel

;for z-only gradients:
;gpz1: -13.5%
;gpz2: 70%
;gpz3: 45%
;gpz4: 15%
;gpz5: 20%

;use gradient files:
;gpname1: SINE.100
;gpname2: SINE.100
;gpname3: SINE.100
;gpname4: SINE.100
;gpname5: SINE.100

;calculate pulselength according to:
;
;   (DeltaOmega * DeltaT) /
;   (width of region[ppm] *
;   SFOn{MHz})
;
;for p14 use q3 pulse
;(DeltaOmega * DeltaT = 3.448)
;   to cover 17.13ppm
;   (pulselength: 2ms at 400.13 MHz)
;   sp3:   C6/C8 on res.
;   sp5:   C1' off res.
;   sp0:   twofold modulated pulse
;      at 90ppm (C1' off res.)
;             and 137ppm (C6/C8 on res.)
;
;for p30 use iburp2 pulse
;(DeltaOmega * DeltaT = 4.53)
;   to cover 55.86ppm (N1/N9)
;   (pulselength: 2ms at 400.13 MHz)
;
;$Id: hcnchgpjrphsp,v 1.5
;  2000/10/16 11:47:21 ber Exp $
```

H_bCN_b

```
;hcnbase.txt
;avance-version (00/10/16)
;HCN on base
;3D inverse correlation for triple
;   resonance using multiple
;   inept transfer steps
;
;   F1(N,t1) -> F2(C6/C8,t2) -> F3(H6/H8)
;
;phase sensitive
;with decoupling during acquisition
;
;R. Fiala, F. Jiang, V. Sklenar
;   J. Biomol. NMR 12, 373 - 383 (1998)

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"
"in25=inf2/2"
"in10=in25"

aqseq 321

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=3u"
"d10=4u"
"d11=30m"

"d4=1.25m-p19-d16" ;selective for base
   ;J(HC)=200Hz, d4=1/4J
"d5=d4-4u"

"d15=p12+6u"
"d22=24m-4u"
"d25=24m-p19-d16-d15-p22-4u"

"CEN_HC1=(p4-p2)/2"
"CEN_HC2=(p3-p1)/2"
"CEN_HC3=(p12-p2)/2"


1 ze
  d11 pl12:f2 pl16:f3
2 d11 do:f2 do:f3
3 d11 pl9:f1
  d1  cw:f1
  d11 do:f1
  d11 pl3:f3 pl2:f2 pl1:f1

;####### INEPT

  (p1 ph1)
  d4 UNBLKGRAD
  p19:gp1
  d16
  (CEN_HC1 p2 ph1):f1 (p4 ph1):f2
  d16
  p19:gp1
  d4
  (p1 ph3):f1
 ; 4u pl11:f1
 ; p11 ph11
  50u
  p16:gp6
  d16 pl1:f1
  (p3 ph2):f2

;####### Start Evolution 15N (States-Tppi)

  d10
  (p13:sp3 ph1):f2
  d22 ;pl20:f2

;####### Evolution 13C (Echo-Antiecho)

  (p21 ph5):f3
  d0
  (p12:sp2 ph1):f2
  d0
  (p22 ph1):f3
  d15
 (p21 ph6):f3

  d25 ;pl2:f2
  p19:gp2*EA
  d16

;####### finish Evolution 15N (States-Tppi)

  (p13:sp3 ph1):f2  ; suppresses C5-coupling to C6/C8
                    ; when pyrimidines are labeled
  4u
  (p1 ph7):f1
  4u
  d4 pl2:f2
  p19:gp3
  d16
  (CEN_HC1 p2 ph1):f1 (p4 ph1):f2
  d16
  p19:gp3
  d4 pl2:f2
  4u

;####### INEPT

  (CEN_HC2 p1 ph10):f1 (p3 ph1):f2
  4u
  d4 pl2:f2
  p19:gp4
  d16
  (CEN_HC1 p2 ph1):f1 (p4 ph1):f2
  4u
  d16
  p19:gp5
  d5 pl12:f2 pl16:f3
```

```
  4u BLKGRAD
  ;p1 ph11
  go=2 ph31 cpd2:f2 ;cpd3:f3

   d11 do:f2 do:f3 mc #0 to 2
     F1PH(rd10 & rd25 & ip5, id0)
     F2EA(igrad EA & ip10*2,id10 & dd25)
exit

ph1=0
ph2=0 2
ph3=1
ph5=0 0 2 2
ph6=0
ph7=0
ph10=1
ph11=3
ph31=0 2 2 0

;default power levels:
;pl0 : 120db
;pl1 : f1 channel
;pl2 : f2 channel
;pl3 : f3 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel
;pl16: f3 channel
;shaped pulse 180 degree:
;sp0 : f2 channel (modulated)
;sp3 : f2 channel (C6/C8 on resonance)
;sp5 : f2 channel (C1' off resonance)
;sp9 : f3 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree


;p14: f2 channel - 180 degree shaped pulse
;p16: homospoil/gradient pulse [1 ms]
;p19: homospoil/gradient pulse [0.5 ms]
;p28: f1 channel - trim pulse  [1 ms]
;p30: f3 channel - 180 degree shaped pulse
   for inversion
;d0 : incremented delay (2D) [3 us]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(HC) (C1') [1.6 ms]
;d11: delay for disk I/O [30 ms]
;d13: short delay  [4 us]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;     d19 = (1/(2*d))
;       d = distance of next null (in Hz)
;d22: 1/(4J(CN))
;        C1': [18 ms]
;d23: 1/(4J(CN))
;        C1' and C6/C8: [17 ms]
;d25: 1/(4J(CN))
;        C6/C8: [16 ms]
;d27: 1/(4J(HC)) (C6/C8)
;        for H6/H8: [1.25 ms]
;in0: 1/(2 * SW(H)) =  DW(H)
;nd0: 2
;NS: 4 * n
;DS: >= 32
;td1: number of experiments in F1
;FnMODE_15N(F1): States-TPPI (or TPPI)
;FnMODE_13C(F2): E/AE

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;cpd3: according to sequence defined by cpdprg3
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel
;pcpd3: f3 channel

;shaped pulses for 600MHz:
;p13@sp3 iburp2.1000, p13=2100us, spoffs=-7000Hz (C5)
;p12@sp2 reburp1000, p12=2500us, spoffs=0Hz

;shaped pulses for 700MHz:
;p13@sp3 iburp2.1000, p13=2100us, spoffs=-7000Hz (C5)
;p12@sp2 reburp1000, p12=2300us, spoffs=0Hz

;use gradients:
;   gp 1 : gp 2 : gp 3 : gp 4 : gp 5 : gp 6
;     20 :   80 :   25 :  15 :   35 :  -60

;use gradient files:
;gpname1: SINE.50
;gpname2: SINE.50
;gpname3: SINE.50
;gpname4: SINE.50
;gpname5: SINE.50
;gpname6: SINE.100

;calculate pulselength according to:
;
;   (DeltaOmega * DeltaT) /
;   (width of region[ppm] *
;   SFOn{MHz})
;
;for p14 use q3 pulse
;   (DeltaOmega * DeltaT = 3.448)
;   to cover 17.13ppm
;   (pulselength: 2ms at 400.13 MHz)
;   sp3:   C6/C8 on res.
;   sp5:   C1' off res.
;   sp0:   twofold modulated pulse
;    at 90ppm (C1' off res.)
;          and 137ppm (C6/C8 on res.)
;for p30 use iburp2 pulse
;   (DeltaOmega * DeltaT = 4.53)
;   to cover 55.86ppm (N1/N9)
;   (pulselength: 2ms at 400.13 MHz)

;$Id: hcnchgpjrphsp,v 1.5
;2000/10/16 11:47:21 ber Exp $
```

HCCH-COSY-TOCSY

```
;hcchcosytocsy for nucleic acids
;avance-version (00/12/08)
;HCCH-COSY-TOCSY
;3D sequence with
;   inverse correlation using multiple inept
;   transfer and C-C DIPSI3 spinlock
;
;   F1(H,t1) -> F2(C,t2) -> F2(C') -> F1(H',t3)
;

;phase sensitive (t1)
;phase sensitive (t2)
;
;Hu et al JBNMR 12, 559 (1998)
;Differences to the paper:
;  1.-DELTA6 and DELTA7 are a bit different
;     from tau-c and tau-d.
;     DELTA6 is shortened as a compromise
;     for CH and CH2 [*0.6]
;  2.-Use 15N instead of 31P.
;     here only decouple N or P,
;
;*****************************************
;ATTENTION!!!!
;This program is an example of bad practice.
;  The definition of increments and
;  delays is confusing:
;      d0, d20, d10 are a semi-CT on 1H
;      d23, d25 are the CT on 13C
;SOLUTION to the big mess:
;  Do not believe the increments and
;  SW in the EDA window.
;  What dominate are the values of the
;    increments as defined in the ASED
;    window and in the loops of the
;    program. The corresponding SW
;    are 1/(2*in) when st-ttpi mode is used.
;  When processing in Felix, input the
;    correct SW_hz calculated by hand
;    instead of believing what Felix
;    gets automatically from xwinnmr.
;*****************************************
;1sw=0.5*in10
;2sw=0.5*in23

prosol relations=<triple>

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p4=p3*2"
"p22=p21*2"
"p2=p1*2"

"DELTA1=1.5m"    ;1/[1J(CH)]  160 Hz
"DELTA2=1.0m"    ;0.67 DELTA1,
 ;compromise for CH & CH2

"d0=3u"
"d20=3u"
"d10=DELTA1 + 6u + p4"
"d11=30m"
"d12=20u"
"d23=3u"
"d25=3.2m"

"p20=1m"    ;trim pulse
"p28=500u"  ;trim pulse
            ;Although a trim pulse can be up to 2ms
            ;to 2ms, preference to use less length

"p16=500u"   ;grad
"p29=300u"   ;grad
"p30=2m"     ;grad
"p31=1.25m"  ;grad

"DELTA4=3.2m-p16-50u"   ;refocus J(CC)
       ;2*DELTA4 is duration of COSY (?) -see paper
"DELTA3=DELTA4-DELTA2-d16+50u"
"DELTA5=3.2m-p22-3u"

;CT is 2*DELTA5 = 3.2ms =0.58/[1J(CC)]  45 Hz
;Though 1/(2*J)=11ms, we evolve for ca. 6ms as
;a compromise for C-C and C-C-C carbon atoms
;in the ribose

"DELTA6=1.0m-p29-d16-50u"
"DELTA7=1.6m-p29-50u-d16"
"DELTA8=DELTA7-p1"

"CEN_HN2=(p22-p2)/2"
"CEN_HC2=(p4-p2)/2"

"in0=inf1/2"
"in23=inf2/2"

aqseq 312

1 d11 ze
  d11 pl12:f2
2 d11 do:f2
3 d1 do:f2
  50u UNBLKGRAD
  d12 pl1:f1
  d12 pl2:f2
  d12 pl3:f3

 ; (p3 ph0):f2
 ;    50u
 ;    p16:gp1
 ;    d16
 ; (p3 ph2):f2
 ;    50u
 ;    p16:gp1
 ;    d16

;####### INEPT & Evolution of 1H (semi-CT)

  (p1 ph1)
  DELTA1
  d0
  (p4 ph0):f2
  d20
  (p2 ph0)
  d10
  (p1 ph2)
     50u
     p16:gp2
     d16
```

```
  (p3 ph3):f2

;####### COSY

  DELTA2
  (p2 ph0)
  DELTA3
     p16:gp3
     d16
  (p4 ph0):f2
     50u
     p16:gp3
  DELTA4

;####### Evolution of 13C (CT)

  (p3 ph4):f2
  d23
  (p22 ph0):f3 (CEN_HN2 p2 ph0)
  DELTA5
  (p4 ph0):f2
  d25

  (p3 ph5):f2
      50u
     p30:gp4
     d16

;####### TOCSY mixing

  (p1 ph0)
     50u
     p31:gp4
     d16
  (p3 ph6):f2

  d12 pl15:f2

  (p20 ph0):f2     ;this p20 trim pulse is part
                   ;of the spin-lock
                   ;begin DIPSI3
9 (p9*2.722 ph7):f2
  (p9*4.389 ph9):f2
  (p9*2.778 ph7):f2
  (p9*3.056 ph9):f2
  (p9*0.333 ph7):f2
  (p9*2.556 ph9):f2
  (p9*4.000 ph7):f2
  (p9*2.722 ph9):f2
  (p9*4.111 ph7):f2
  (p9*3.778 ph9):f2
  (p9*3.889 ph7):f2
  (p9*2.889 ph9):f2
  (p9*3.000 ph7):f2
  (p9*0.333 ph9):f2
  (p9*2.500 ph7):f2
  (p9*4.050 ph9):f2
  (p9*2.830 ph7):f2
  (p9*4.389 ph9):f2
  (p9*2.722 ph9):f2
  (p9*4.389 ph7):f2
  (p9*2.778 ph9):f2
  (p9*3.056 ph7):f2
  (p9*0.333 ph9):f2
  (p9*2.556 ph7):f2
  (p9*4.000 ph9):f2
  (p9*2.722 ph7):f2
  (p9*4.111 ph9):f2
  (p9*3.778 ph7):f2
  (p9*3.889 ph9):f2
  (p9*2.889 ph7):f2
  (p9*3.000 ph9):f2
  (p9*0.333 ph7):f2
  (p9*2.500 ph9):f2
  (p9*4.050 ph7):f2
  (p9*2.830 ph9):f2
  (p9*4.389 ph7):f2
  (p9*2.722 ph9):f2
  (p9*4.389 ph7):f2
  (p9*2.778 ph9):f2
  (p9*3.056 ph7):f2
  (p9*0.333 ph9):f2
  (p9*2.556 ph7):f2
  (p9*4.000 ph9):f2
  (p9*2.722 ph7):f2
  (p9*4.111 ph9):f2
  (p9*3.778 ph7):f2
  (p9*3.889 ph9):f2
  (p9*2.889 ph7):f2
  (p9*3.000 ph9):f2
  (p9*0.333 ph7):f2
  (p9*2.500 ph9):f2
  (p9*4.050 ph7):f2
  (p9*2.830 ph9):f2
  (p9*4.389 ph7):f2
  (p9*2.722 ph7):f2
  (p9*4.389 ph9):f2
  (p9*2.778 ph7):f2
  (p9*3.056 ph9):f2
  (p9*0.333 ph7):f2
  (p9*2.556 ph9):f2
  (p9*4.000 ph7):f2
  (p9*2.722 ph9):f2
  (p9*4.111 ph7):f2
  (p9*3.778 ph9):f2
  (p9*3.889 ph7):f2
  (p9*2.889 ph9):f2
  (p9*3.000 ph7):f2
  (p9*0.333 ph9):f2
  (p9*2.500 ph7):f2
  (p9*4.050 ph9):f2
  (p9*2.830 ph7):f2
  (p9*4.389 ph9):f2
  lo to 9 times l1

  (p20 ph0):f2     ;this p20 trim pulse is part
                             ;of the spin-lock
                             ;end DIPSI3
  d12 pl2:f2

  (p3 ph8):f2

  (p28 ph0)
  (p28*2 ph2)
  50u
  p30:gp4
  d16

;####### INEPT

  (p1 ph0)
  50u
  p31:gp4
  d16
  (p3 ph5):f2
  50u
  p29:gp5
  d16
  DELTA6
  (CEN_HC2 p2 ph0) (p4 ph0):f2
```

```
  50u
  p29:gp5
  d16
  DELTA6
  (p3 ph0):f2
  50u
  p16:gp6
  d16

;####### INEPT

  (p1 ph0)
  50u
  p29:gp7
  d16
  DELTA7
  (CEN_HC2 p2 ph0) (p4 ph0):f2
  50u
  p29:gp7
  d16 pl12:f2
  DELTA8 BLKGRAD
  (p1 ph10)

  go=2 ph31 cpd2:f2
  d11 do:f2 mc #0 to 2
   F1PH(ip1, id0 & id20 & dd10)
   F2PH(rd0 & rd20 & rd10 & ip3 & ip4,id23 & dd25)
exit


ph0=0
ph2=1
ph1=0 2
ph3=0 0 2 2
ph4=0
ph5=3
ph6=1
ph7=1
ph9=3
ph8=1
ph10=2
ph31=0 2 2 0


;default power levels:
;pl0 : 120dB
;pl1 : f1 channel
;pl2 : f2 channel
;pl3 : f3 channel
;CPD/BB decoupling power level:
;pl12: f2 channel
;TOCSY-spinlock decoupling power level:
;pl15: f2 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree
;low power pulse - TOCSY:
;p9 : f2 channel -  90 degree
;TOCSY of 7.4 kHz at 700 MHz   [32-35u]

;p16: homospoil/gradient pulse         [500 usec]
;p19: gradient pulse 2                 [2 msec]
;p20: f2 channel - trim pulse          [2 msec]
;p28: f1 channel - trim pulse          [1 msec]
;p29: gradient pulse 3                 [300 usec]
;p30: gradient pulse 4                 [5 msec]
;p31: gradient pulse 5                 [4.4 msec]
;d0 : incremented delay (F1 in 3D)     [3 usec]
;d1 : relaxation delay; 1-5 * T1       [1.8 sec]
;d10: incremented delay (F2 in 3D)     [3 usec]
;d11: delay for disk I/O               [30 msec]
;d12: delay for power switching        [20 usec]
;d16: delay for homospoil/gradient recovery
;d21: 1/(6J'(CH)) - tau c              [1.1 msec]
;d23: tau b                            [475 usec]
;l1: loop for DIPSI cycle:
;       mixing time =
;       ((p9*54.33*4) * l1) + (p20) [12 msec]
;in0: 1/(2 * SW(H)) = DW(H)
;nd0: 2
;in10: 1/(2 * SW(C)) = DW(C)
;nd10: 2
;NS: 16 * n
;DS: 256, set a high value due to DIPSI

;td1: number of experiments in F1
;td2: number of experiments in F2
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: States-TPPI (or TPPI) in F2
;decoupling:
;cpd2: according to sequence defined
;      by cpdprg2 [GARP]
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel

;1J(CH): 160 Hz in ribose
;1J(CC): 45 Hz in ribose
;The compromise for CX and CX2
;    is ca. 0.67*... [0.5-0.6]

;DELTA2: refocus HC ... 1/(8*J)
;DELTA3: refocus CC
;DELTA4*2: time for CC refocus [3.2ms]

;for z-only gradients:
;gpz2: 50%
;gpz3: 10%
;gpz4: 30%
;gpz5: 25%
;gpz6: 60%
;gpz7: 18%

;use gradient files:
;gpnam2: SINE.50
;gpnam3: SINE.50
;gpnam4: SINE.100
;gpnam5: SINE.50
;gpnam6: SINE.50
;gpnam7: SINE.50

;st-TPPI on both
;TD: 1k - 64 - 128
;SW: 6 - 28 - 4 ?
;in: dominates sw -- 13c in23=100u
;                   -- 1h  in0=180u
;o?p : 4.7 - 79.0 - 158.0 ppm

;$Id: hcchdigp3d,v 1.8
;2000/12/11 15:56:19 ber Exp $
```

HCCH-COSY

```
;hcchcosy.txt
;
;from vshcchcosySE2d for nucleic acids
;avance-version (00/12/08)
;(H)CCH-COSY-SE
;2D version for C5-C6 of C and U in RNA.

;2D version of 3D sequence with
;inverse correlation using multiple
;inept transfer and
;
;  F1(C,t1) -> F2(C',t2) -> F2(C') -> F1(H',t3)
;
;phase sensitive (t1)
;phase sensitive (t2)
;
;K. Gehring et al JMR 135, 185 (1998)

;In the 2D version, only the first CT is evolved.
;In the States-TPPI manner.
;The second CT is not t2 evolved, so the E/AE
;is not feasible and the Sensitivity Enhancement
;either, in contrast with the 3D version.

;Check delays for delayed acquisition
;and ph-corr=0,0

;When refocusing heteronuclear J coupling,
; correct 1/4J with a 0.62 factor as a
; compromise for nuclei bound either to
; one or to two neighbours (3D version).
;In the 2D version for the C5-C6 of RNA
;  C and U (HCC'H'), use instead 1/8J
;  in order to have magnetization both
;  in C and C' (incomplete transfer).
;  Notice that, besides, some M will go
;  to C4 and will be lost there.
;  It is not possible to use shaped
;  pulses for selective refocusing of
;  C5-C4 coupling due to the fact that
;  the bands of C4 and C6 are too close.

prosol relations=<triple>

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=3u"
"d4=1/(4*cnst2)"
       ;1J(XH) pyrimidine = 175-185Hz =180Hz
       ;1J(XH)     ribose = 145-170Hz =160Hz

"DELTA1=d4-50u-p16-d16"
       ;paper says grad not possible

"d5=0.5* 1/(4*cnst3)"
       ;CT/2
       ;1J(CC) pyrimidine = 67Hz
       ;1J(CC)     ribose = 40Hz
       ;Atten. by 0.5 to keep M both in C and C',
       ;i.e., incomplete transfer from C to C'.

"DELTA3=2*d4"
       ;Total evoln. of 1J(CH) during the CT
       ;    -- refocus.
       ;*0.62: compromise for refocusing 1J(CH)
       ;of CH and CH2 in the paper (for protein).
       ;In RNA bases, full evoln. is possible.

"DELTA4=d5-d4"
       ;the paper calls it DELTA
       ;"DELTA4=d5-d4-p29-d16" -corrected for grad
       ;"DELTA4=d5-0.62*d4-p29-d16"
       ;if also CH2 to refocus.

"DELTA5=d4-d0-p22"
       ;this is CT/2-DELTA4=0.5*DELTA3
       ;"DELTA5=0.62*d4" if also CH2 to refocus.

"d10=d5+p2"
       ;p2 added for good phase
       ;p29 shorter than p16 to allow
       ;longer evolution
       ;"d10=d5-50u-p29-d16+p2" with grad

"d20=3u"
"d30=d5+p2"   ;identical to d10

"DELTA7=d4-50u-p16-d16-4u"
       ;refocus 1J(CH)
       ;1.1ms in the paper vs 1.7m DELTA1.
       ;"DELTA7=0.65*d4-50u-p16-d16"
       ;if also CH2 to refocus.

"DELTA8=50u+p16+d16"
       ;compensation of last grad

"d11=30m"
"d12=20u"

"CEN_HN2=(p22-p2)/2"
"CEN_HC2=(p4-p2)/2"
"CEN_HC1=(p3-p1)/2"

;aqseq 312

1 d11 ze
  d11 pl12:f2
2 d11 do:f2
3 d1 do:f2
  50u UNBLKGRAD
  d12 pl1:f1
  d12 pl2:f2
  d12 pl3:f3

;####### purging of C natural magnetization

  (p3 ph0):f2
      50u
      p16:gp1
      d16

;####### INEPT - paper says grad not possible

  (p1 ph1)
  DELTA1
      50u
      p16:gp12
      d16
  (CEN_HC2 p2 ph2) (p4 ph1):f2
      50u
      p16:gp12
      d16
  DELTA1
  (p1 ph2)
      50u
      p16:gp2
      d16
  (p3 ph3):f2     ;increment for St-TPPI
```

```
;####### CT evolution t1: transfer via 1J(CC)
;
; As J=ca 67 Hz, the CT should be 1/2J=7.4m.
; But it's shortened to 1/2 in order to have
; M both in C and C'.
; This makes the CT too short, so
; that gradients can not be used.
; This happens also in the paper.
; Instead, the pi pulse on C is
; exorcycled -that happens to be more efficient

  d0
  (p22 ph1):f3
  DELTA5
  (p2 ph2)
  DELTA4
      ;p29:gp3
      ;d16
  (p4 ph4):f2
      ;50u
      ;p29:gp3
      ;d16
  d10
  (p3 ph5):f2

;####### purging pulses
;        for residual antiphase H-C terms

  (p1 ph6)
  (p1 ph1)

;####### second CT evolution -
;        disabled in the 2D version
  d20
  (p22 ph1):f3
  DELTA5
  (p2 ph2)
  DELTA4
      ;p29:gp5
      ;d16
  (p4 ph8):f2
      ;50u
      ;p29:gp6   ; here the 3D has
                 ; EA concerted with ph7
      ;d16
  d30
  (CEN_HC1 p1 ph1)  (p3 ph1):f2
        ;spoil grad not usable here with E/AE.
        ;taco recommends here simultaneous
        ;90deg pulses.
;Sensitivity enhancement disabled
;in the 2D version

;;####### Refocus antiphase M as in a usual HSQC

  DELTA7
      4u
      50u
      p16:gp7
      d16
  (CEN_HC2 p2 ph1) (p4 ph1):f2
      50u
      p16:gp7
      d16
  DELTA7 pl12:f2
      4u BLKGRAD

  go=2 ph31 cpd2:f2
     d11 do:f2 mc #0 to 2
     F1PH(ip3, id0 & dd10)
exit

;Phase cycle of the 2D version.
;gp3 and gp4 disabled and pi pulse exorcycled
ph0= 0
ph1= 0
ph2= 1
ph3= 0
ph4= 0 1 2 3 0 1 2 3
ph8= 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
     3 3 3 3 3 3 3 3 0 0 0 0 0 0 0 0
ph5= 0 2 2 0 2 0 0 2
ph6= 0 0 2 2 0 0 2 2 2 2 0 0 2 2 0 0
     ;ph6 does not contribute to
     ;the detectable signal (?)
ph31=0 2 0 2 0 2 0 2 2 0 2 0 2 0 2 0

;default power levels:
;pl1 : f1 channel
;pl2 : f2 channel
;pl3 : f3 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree

;p16: homospoil/gradient pulse       [500 usec]
;p19: gradient pulse 2               [2 msec]
;p29: gradient pulse 3               [300 usec]
;p30: gradient pulse 4               [5 msec]
;p31: gradient pulse 5               [4.4 msec]
;d0 : incremented delay (F1 in 3D)   [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(CH)) - tau a             [1.6 msec]
;d10: decremented delay (F1 in 3D)   [3 usec]
;d11: delay for disk I/O             [30 msec]
;d12: delay for power switching      [20 usec]
;d16: delay for homospoil/gradient recovery
;in0: 1/(2 * SW(H)) = DW(H)
;nd0: 2
;in10: in0
;nd10: 2
;NS: 16 * n
;DS: 32
;cnst2: 1J(CH)=180Hz in C and U -RNA
;cnst3: 1J(CC)=67Hz for C5-C6 in RNA
;FnMODE: States-TPPI (or TPPI) in F1
;FnMODE: States-TPPI (or TPPI) in F2 -- E/AE??

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel

;for z-only gradients
;gpz1: 20%
;gpz2: 40%
;gpz7: 25%
;gpz12: 15%
;
;gpz3: 25%
;gpz4: = gpz3
;gpz5: -26.8%
;gpz6: -26.8, 0, -67

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.50
;gpnam3: SINE.100
;gpnam4: SINE.100

;$Id: hcchdigp3d,v 1.8
;2000/12/11 15:56:19 ber Exp $
```

H2H8-COSY

```
; bsADE.3d
; TROSY relayed HCCH-COSY experiment
; for correlating adenine H2/H8 resonances
; in uniformly 13C-labeled RNA molecules
; lit: B. Simon, K. Zanier, M. Sattler
;      J. Bio. NMR 2001, 20(173)

; F1 1H 4.7ppm
; F2 13C 140ppm
; F3 15N 150ppm


#include <Avance.incl>
#include <Grad.incl>

"p2=p1*2"
"p4=p3*2"
"p6=p5*2"

"p16=600u" ; gradient PW
"d16=300u" ; gradient recovery

"d4=1.25m"
         ; 1/4J(H,C) RNA aromatic
"d23=28.4m"
         ;  4*1/2J(C5,C6) in U and C
         ; for C2-C5 and C8-C4/C6
         ; long range COSY transfer
"d11=7.43m-p16-d16"
         ;  1/2J(C5,C6) in U and C
         ; for suppression of pyrimidine signals
"d25=2.0m/2-p16-d16"
         ; 1/8J(C5,C6) for C5<->C6
         ;  relayed COSY transfer
         ;  JC5C6=66.5Hz, JC4C5=74.4Hz
"d24=d4"
         ;  1/4J(H,C)
"d22=d23/2-p16-d16"
         ;  constant time out
"d20=d23/2-p16-d16-d24"
         ; constant time back

"d0=3u"
"d12=d22-d11-p17-3u"
"d13=d22-p17-23u"
"d10=in10/2-1.27324*p3"
;"d5=d23/2-p16-d16-d24+p1-d0-1.2733*p3"
         ; without p6 on N
"d5=d23/2-p16-d16-d24+p1-d0-1.2733*p3-p6"
         ; with p6 on N
"d26=p3/2-p1/2"
"d27=p4/2-p2/2"
"d28=d4-(p16+d16+50u)"
"d8=p1"
"d9=p3"

"l10=(td1/2)"      ;d10
"l0=(td2/2)"        ;ct d0 d20

aqseq 312

1 ze
2 d1 do:f3 do:f2
  5m
3 30m
4 30m
5 5m
6 20u pl1:f1
  20u pl2:f2
  20u pl3:f3
;  20u pl7:f1
;  (p29 ph0):f1
  20u pl1:f1
  (p1 ph0):f1
   ; H2-C2 and H8-C8 INEPT  transfer
  d4 ; 1.25m
  (p4 ph0):f2 (d27 p2 ph10)
  d4
  (p1 ph11):f1
  100u UNBLKGRAD
  p16:gp1
  d16
  (p3 ph7):f2
   ; C2-C5 and C8-C4/C6 INEPT transfer
  p16:gp4
  d16
  d11 pl11:f2
  (p17:sp11 ph20):f2
    ; EBURP2 1500us at 100ppm
 ; (-40ppm=-6000Hz) selective 90 on C5 of U and C
  3u
  d12 pl2:f2
  (p4 ph23):f2
  p16:gp4
  d16 pl11:f2
  d13
  (p17:sp11 ph1):f2
   ; EBURP2 Bloch-Siegert Phasecompensation
  3u
  20u pl2:f2
  (p3 ph21):f2
   ; C5-C4/C6 and C4/C6-C5 realyed INEPT transfer
  d25
  p16:gp5
```

```
d16
(p4 ph23):f2
p16:gp5
d16
d25
(p3 ph13):f2
   d10
 ; cs evolution of C2,C4,C5,C6,C8
(p3 ph3):f2
 ; C4/C6-C5 and C5-C4/C6
 ; realyed INEPTBACK transfer
d25
p16:gp6
d16
(p4 ph0):f2
p16:gp6
d16
d25
(p3 ph22):f2
 ; C5-C2 and C4/C6-C8 INEPTBACK transfer
d0
 ; id0 CT cs evolution of C2,C8
(p6 ph0):f3
 ; if no 15N decoupling take p6 out
 ; and change d5 in header
d5
p16:gp2*EA
d16 pl2:f2
(p4 ph0):f2
p16:gp2*-1*EA
d16
d20
 ; dd20 CT cs evolution of C2,C8
(p1 ph5):f1
 ; half TROSY BACK tranfser C8-H8, C2-H2
d24
(p4 ph24):f2 (d27 p2 ph1):f1
d24
(p3 ph25):f2 (d26 p1 ph4):f1
d4
(p4 ph24):f2 (d27 p2 ph0):f1
d28
p16:gp3
d16 pl12:f2
50u BLKGRAD
go=2 ph31 cpd2:f2
d1 do:f2 wr #0 if #0 zd
5m ip5*2 igrad EA
lo to 3 times 2    ; E/AE
5m id0   ; IN0=IN20 = 1/(2*SWH)
5m dd20   ; l0 < d20/IN20
5m ip24*2
5m ip25*2
  5m ip31
  5m ip31
  lo to 4 times l0        ; C8,C2
  5m rd0
  5m rd20
  5m ip7
  5m ip13
  5m ip21
  5m ip23
  lo to 5 times 2    ; States-TPPI
  5m id10
  lo to 6 times l10    ; C2,C4,C5,C6,C8
exit


ph0=0
ph1=1
ph11=1
ph10=0  ; x add 13C magnetization,
        ; y subtract 13C magnetization
ph3=1
ph13= 1 3
ph23= 0 2
ph21=0 2
ph22=0
ph4=2
ph5=3  ; y TROSY -y ANTI-TROSY on C2/C8
ph7=0 2
ph20= 0 0 2 2
ph24=0
ph25=1
ph31=0 2


;ph10 x Add 13C magnetization,
;ph10 y subtract 13C magnetization
;ph5 y TROSY   ph5 -y ANTI-TROSY
;IN0 =IN20 = 1/(2*SWH)
;l0 < d20/IN20
;p1 ,p2  1H hard 90,180 pulses (~11/22us)
;p3 ,p4  13C hard 90,180 pulses (~13/26us)
;p6  15N hard 180 puls (~70us)
;p17   EBURP2 1500us at 100ppm
; (-40ppm=-6000Hz) slective 90 on C5 of U and C
;gpz1 70, strength z-gradient 1 zz
;gpz2 45, strength z-gradient 2 (E/A)
;gpz3 22.6, strength z-gradient 3 (E/A)
;gpz4 13, strength z-gradient 4 180
;gpz5 -7, strength z-gradient 5 180
;gpz6 9, strength z-gradient 6 180
```

^{13}C-edited - ^{13}C/^{12}C-filterted NOESY

```
;13Ceditednoesy_12C.txt
;lab_unlab Version
;pulse program for 13Ceditednoesy_13C.txt in ( )
;
;filtered NOESY to detect NOEs
; between 13C/12C-bound protons
;(filtered NOESY to detect NOEs
;  between 13C/13C-bound protons)
;
;Filter described in C Zwahlen,
;   JACS 1997, 119, 6711-6721
;   Based on adiabatic pulse at
;   a given rate plus some
;   fitting of coupling and c.s.

; 3D version

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"
"in10=inf2/2"

"p2=p1*2"
"p4=p3*2"
;"p6=p5*2"
"p24=500u"

;"d0=(in0*2-p3*4/3.1416-p2)/2"
 ; for ph = (+180, -360)
"d0=(in0-p3*4/3.1416-p2)/2"
 ; for ph = (+90, -180)
;delayed AQ only in 3D version

"d4=1.4m"               ;1/(4*cnst2)
"d14=d4-p19-50u-p24/2"
"d22=1.7m-p25/2"
     ;numbers 1.7m and 1.8m
     ;taken from paper
"d23=1.8m-p26/2"
"d10=3u"
"d11=30m"
"d12=3u"
"d13=d4-p19-50u-d16"
"d15=d4-p19-50u-p24-d10-d12-d16"

  ;delayed AQ in semi-CT:
  ; equal delay on both sides
  ; of p2 -- no evoln. of H c.shift.
  ; ph_corr=0,0

"d8=150m" ; NOESY mixing time
"DELTA=d8-50u-p16-d16*2"

"CEN_HC1=(p24-p2)/2"
"CEN_HC2=(p25-p2)/2"
"CEN_HC3=(p26-p2)/2"

"l1=td1/2"
"l3=td2/2"
"in13=d13/l3"
"in12=in10-in13"
;in10=1/(2*SWH(1H))

aqseq 312

1 ze
2 d11 do:f2 do:f3
  d11
3 d11
4 d11*2
5 d11*3
6 d1
  10u pl1:f1
  10u pl2:f2
  10u ;pl3:f3
  (p3 ph0):f2
     50u UNBLKGRAD
     p16:gp1
     d16*2
  p1 ph0
     50u
     p19:gp2
  d14  pl18:f2  ;optional - to 120dB
    (CEN_HC1 p2 ph0)(p24:sp4 ph0):f2
     50u
     p19:gp2
  d14 pl2:f2
   ;put power up after adiabatic pulse
  p1 ph12
     50u
     p16:gp3
     d16 ;pl13:f3
  4u ;cpd3:f3
  (p3 ph1):f2
  d0
  (p2 ph0)
  d0
  (p3 ph0):f2
  4u ;do:f3
     50u
     p16:gp4
     d16
  (p1 ph2):f1
     46u
     p19:gp5
     d16 ;pl13:f3
  4u ;cpd3:f3
  d13 pl18:f2
     ;IN13=d13/l3; l3 number of real points in F2
  (p2 ph0)
  d12
     ;IN12=IN10-IN13
  (p24:sp4 ph0):f2
  d10
     ;IN10=1/(2*SWH(1H))
  d15
  4u ;do:f3
     46u
     p19:gp5
     d16
  (p1 ph3):f1
  DELTA      ;mixing time
     46u
     p16:gp6
     d16
  d16 ;pl13:f3  ;perhaps set it to pl19 120dB
  4u            ;d16 is too long for changing pl13
  (p1 ph4):f1

;######### Block that filters out H bound to 13C

  d22
  (CEN_HC2 p2 ph0) (p25:sp5 ph0):f2
    ;delay required before changing pl2
  d22
  (p1 ph5):f1
```

```
  20u pl2:f2
  (p3 ph0):f2
   ;(exclude (p3 ph0):f2 for 13C-filtered version)

     46u
     p16:gp7
     d16 pl18:f2   ;perhaps set it to pl18 120dB
     4u
  (p1 ph0):f1
  d23
  (CEN_HC3 p2 ph0) (p26:sp6 ph0):f2
  d23
  (p1 ph0):f1
   ;(use ph3 in 13C-filtered version)

   46u
     p16:gp8
     d16 pl12:f2
     4u BLKGRAD
   (p1 ph0):f1

  go=2 ph31 ;cpd2:f2 cpd3:f3
  ; (include cpd2 for 13C-filtered version)

  d11 do:f2 do:f3 wr #0 if #0 zd
     d11 ip1
  lo to 3 times 2
     d11 id0
  lo to 4 times l1
     d11 rd0
     d11 ip2
  lo to 5 times 2
     d11 id10
     d11 id12
     d11 dd13
  lo to 6 times l3
exit

    ph0 = 0
    ph1 = 0 2
    ph2 = 0 0 2 2
    ph3 = 1
    ph4 = 1
    ph5 = 1 1 1 1 3 3 3 3
    ;(ph5 = 0 0 0 0 2 2 2 2
    ;  in 13C-filtered version)

    ph11= 1
    ph12= 1
    ph31= 0 2 2 0 2 0 0 2


;For the 2D version, the other loop
;must be deleted. Besides, the 13C
;c.s. evoln. period (d0) must deleted, too.
;The 2D version is recorded with no
; decoupling to better identify residual
; signals from H bound to 13C that
; are not totally refocused in the last
; inept-like period due to the fact that
;1J(HC) in RNA are too different.

;shaped pulses on the 900MHz:
;p24 smoothed cirp size 1000,
;  tot sweep width 53000, 500usec, 20%, -1
;p25 wurst size 1000, tot sweep width 90000,
;  1865usec, ad power index 20, +1
;p26 wurst size 1000, tot sweep width 90000,
;  2389usec, ad power index 20, +1
;p25 & p26 offset -14000Hz,
;  power of 6000Hz := 41,7usec

;shaped pulses on the 800MHz:
;p24 smoothed cirp size 1000,
;  tot sweep width 50000, 500usec, 20%, -1
;p25 wurst size 1000, tot sweep width 80000,
;  1859usec, ad power index 20, +1
;p26 wurst size 1000, tot sweep width 80000,
;  2358usec, ad power index 20, +1
;p25 & p26 offset -12000Hz,
;  power of 6000Hz := 41,7usec

;shaped pulses on the 700MHz:
;p24 smoothed cirp size 1000,
; tot sweep width 48000,  500usec, 20%, -1
;p25 wurst size 1000, tot sweep width 70000,
;  1852usec, ad power index 20, +1
;p26 wurst size 1000, tot sweep width 70000,
;  2341usec, ad power index 20, +1
;p25 & p26 offset -10500Hz,
;  power of 6000Hz := 41,7usec

;shaped pulses on the 600MHz:
;p24 smoothed cirp size 1000,
;  tot sweep width 45000, 500usec, 20%, -1
;p25 wurst size 1000, tot sweep width 60000,
;  1844usec, ad power index 20, +1
;p26 wurst size 1000, tot sweep width 60000,
;  2339usec, ad power index 20, +1
;p25 & p26 offset -9600Hz
;power of 5000Hz =50usec

;p25 calculated y=1804.4+0.068*x
;p26 calculated y=2173.8+0.239*x

;p16 : gradient pulse, long   SINE.100  [1000u]
;p19 : gradient pulse, short  SINE.50   [300u]

;pl12: f2 low power level
;pl13: f3 low power level
;pl18: f2 low power level -- use it before sp
;pl19: f3 low power level -- use it before sp

;Increments in semi-CT
;in13 : d13/l3
;in10 : 1/(2*SW_h)    -- SW_h of 1H
;in12 : in10 - in13
; for older tospin versions:
;nd0 : 2   for 13C evoln
;nd10: 2
;  for 1H evoln  - why is it 2 in semi-CT? the
;      value of nd10 is not relevant - the value of
;      in10 is what matters to define the actual SW

;d4 : 1/(4*cnst2)
;cnst2: 1J(HX)
;d8 : mixing time, ms

;l3 : td1/2 in the 2D version

;Gradient strength
;gpz1: 15%
;gpz2: 10%
;gpz3: 30%
;gpz4: 40%
;gpz5: 10%
;gpz6: 35%
;gpz7: 50%

;o2p: 114p, centre of the observable 13C range
;   in RNA,13C: observe 60 -160 ppm
;                         decouple 60-220 ppm
;o3p: 165p, centre of the 15N range - decouple
```

^{15}N-HSQC for iminos groups

```
;imino_hsqc.txt
;HMQC -- 2D H-1/X correlation via heteronuclear
;zero and double quantum coherence
;phase sensitive
;with decoupling during acquisition
;water suppression using 3-9-19
;  pulse sequence with gradients
;V. Sklenar, M. Piotto, R. Leppik & V. Saudek,
; JMR, Series A 102, 241 -245 (1993)

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"p27=p1"
"p6=p5*2"

"d0=3u"
"d11=30m"
"d12=20u"
"d18=d0*2+p14"
"d26=1s/(cnst4*4)"

"DELTA=d19-p6/2"
"DELTA1=d26-p16-d16-p27*2.385-d19*5+p6/2"
"DELTA2=
 d26-p16-d16-p27*2.154-p27*0.231-d19*5+p6/2-8u"

"CEN_HN1=(p5-p1)/2"
"CEN_HN2=(p6-p2)/2"
"CEN_HC2=(p14-p2)/2"

1 d11 ze
   d11 pl13:f3
2 d1 do:f3
3 d12 pl1:f1

;####### INEPT

  (p1 ph1)
  d26 pl3:f3
  (CEN_HN2 p2 ph1) (p6 ph1):f3
  d26 UNBLKGRAD
  (p1 ph2)
  50u
  p16:gp2
  d16
  (p5 ph3):f3

;####### Evolution 15N

  d0
  (CEN_HC2 p2 ph1) (p14:sp2 ph1):f2
  d0
  (p6 ph1):f3
  d18
  (p5 ph4):f3
  p1 ph1
  DELTA1
  p16:gp1
  d16
  p27*0.231 ph7
  d19*2
  p27*0.692 ph8
  d19*2
  p27*1.462 ph9
  DELTA
  (p6 ph1):f3
  DELTA
  p27*1.462 ph10
  d19*2
  p27*0.692 ph11
  d19*2
  p27*0.231 ph12
  4u
  p16:gp1
  d16
  4u BLKGRAD
DELTA2 pl13:f3
  go=2 ph31 cpd3:f3
  d11 do:f3 mc #0 to 2 F1PH(ip3, id0)
exit

ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph7=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph8=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph9=0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3
ph10=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph11=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph12=2 2 2 2 3 3 3 3 0 0 0 0 1 1 1 1
ph31=0 2 2 0 2 0 0 2

;default power levels:
;pl1 : f1 channel
;pl3 : f3 channel
;CPD/BB decoupling power levels:
;pl12: f3 channel
;3-9-19-pulse (watergate)
;pl18: f1 channel

;high power pulse:
;p0 : f1 channel -  90 degree
  ;use for fine adjustment
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree

;p16: homospoil/gradient pulse
;p27: f1 channel -  90 degree pulse at pl18
;d0 : incremented delay (2D)        [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O            [30 msec]
;d12: delay for power switching     [20 usec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;     d19 = (1/(2*d)),
;       d = distance of next null (in Hz)
;d26 : 1/(4J)YH
;cnst4: = J(YH)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC

;decoupling:
;cpd3: according to sequence defined by cpdprg3
;90 degree pulse for decoupling sequence:
;pcpd3: f3 channel

;for z-only gradients:
;gpz1:  70%
;gpz2: -40%

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100

;$Id: inv4f2gpph19,v 1.2
;2000/05/08 11:40:05 eng Exp $
```

^{15}N-HSQC for aminos groups

```
;amino_hsqc.txt
;inv4f3gpph19
;avance-version (00/04/28)
;HMQC
;2D H-1/X correlation via heteronuclear
;  zero and double quantum coherence
;phase sensitive
;with decoupling during acquisition
;using f3 - channel
;water suppression using 3-9-19
; pulse sequence with gradients
;M. Piotto, V. Saudek & V. Sklenar,
;J. Biomol. NMR 2, 661 - 666 (1992)
;V. Sklenar, M. Piotto, R. Leppik & V. Saudek,
; JMR, Series A 102, 241 -245 (1993)

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p2=p1*2"
"p4=p3*2"
"p22=p21*2"

"d0=3u"
"d11=30m"
"d12=20u"
;"d14=d0*2+p2+p11*2+8u"
"d14=d0*2+p4"
"d26=99u"

"CEN_HN1=(p21-p1)/2"
"CEN_HN2=(p22-p2)/2"
"CEN_HC2=(p4-p2)/2"

1 ze
  d11 pl13:f3
2 d1 do:f3
3 d12 do:f3
  d12 pl1:f1 pl3:f3
  (p1 ph1)
  d26  UNBLKGRAD
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
   d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph13) (p22 ph13):f3
d26
d26
(CEN_HN2 p2 ph14) (p22 ph14):f3
d26
d26
(CEN_HN2 p2 ph13) (p22 ph13):f3
d26
d26
(CEN_HN2 p2 ph14) (p22 ph14):f3
d26
d26
(CEN_HN2 p2 ph14) (p22 ph14):f3
d26
d26
(CEN_HN2 p2 ph13) (p22 ph13):f3
d26
d26
(CEN_HN2 p2 ph14) (p22 ph14):f3
d26
d26
(CEN_HN2 p2 ph13) (p22 ph13):f3
d26
d26
(CEN_HN2 p2 ph11) (p22 ph11):f3
d26
d26
(CEN_HN2 p2 ph12) (p22 ph12):f3
d26
d26
(CEN_HN2 p2 ph11) (p22 ph11):f3
d26
d26
(CEN_HN2 p2 ph12) (p22 ph12):f3
 d26
(p1 ph2)
4u pl11:f1
(p11 ph24)
 d12
 d12
 p16:gp1
 d16 pl2:f2
(p11 ph1)
50u
 p16:gp3
d16
 p11 ph2
50u
p16:gp4
 d16 pl1:f1

(p21 ph3):f3

  d0              ;pl11:f1
                  ;p11 ph15
                  ;4u pl1:f1
   (CEN_HC2 p2 ph5):f1 (p4 ph1):f2
                  ;4u pl11:f1
                  ;p11 ph15
  d0              ;pl1:f1
     (p22 ph1):f3
   d14

(p21 ph4):f3
  50u
  p16:gp2
  d16
 (p1 ph1)
```

```
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
   d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph13) (p22 ph13):f3
  d26
  d26
  (CEN_HN2 p2 ph14) (p22 ph14):f3
  d26
  d26
  (CEN_HN2 p2 ph13) (p22 ph13):f3
  d26
  d26
  (CEN_HN2 p2 ph14) (p22 ph14):f3
  d26
  d26
  (CEN_HN2 p2 ph14) (p22 ph14):f3
  d26
  d26
  (CEN_HN2 p2 ph13) (p22 ph13):f3
  d26
  d26
  (CEN_HN2 p2 ph14) (p22 ph14):f3
  d26
  d26
  (CEN_HN2 p2 ph13) (p22 ph13):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26
  (CEN_HN2 p2 ph12) (p22 ph12):f3
  d26
  d26
  (CEN_HN2 p2 ph11) (p22 ph11):f3
  d26
  d26 BLKGRAD
  (CEN_HN2 p2 ph12) (p22 ph12):f3
   d26   pl13:f3
  p12 ph0
  go=2 ph31 cpd3:f3
  d1 do:f3 mc #0 to 2 F1PH(ip3, id0)
exit
```

```
ph0=(360) 182
ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
ph5=0 0 0 0 2 2 2 2
ph11=0
ph12=1
ph13=2
ph14=3
ph15=2
ph24=3
ph31=0 2 2 0

;default power levels:
;pl1 : f1 channel
;pl3 : f3 channel
;CPD/BB decoupling power levels:
;pl13: f3 channel
;3-9-19-pulse (watergate):
;pl18: f1 channel

;high power pulse:
;p0 : f1 channel -  90 degree
  ;use for fine adjustment
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree

;p16: homospoil/gradient pulse
;p27: f1 channel -  90 degree pulse at pl18
;d0 : incremented delay (2D)       [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O           [30 msec]
;d12: delay for power switching    [20 usec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;     d19 = (1/(2*d)),
;       d = distance of next null (in Hz)
;d26 : 1/(4J)YH
;cnst4: = J(YH)
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC

;decoupling:
;cpd3: according to sequence defined by cpdprg3
;90 degree pulse for decoupling sequence:
;pcpd3: f3 channel

;for z-only gradients:
;  gp 1 : gp 2 : gp 3 : gp 4
;    60 :  -40 :   15 :  40

;use gradient files:
;gpnam1: SINE.100

;$Id: inv4f3gpph19,v 1.2
;2000/05/08 11:40:05 eng Exp $
```

Imino-NOESY

```
;noesyfpgpphjrrs
;avance-version (07/04/04)
;NOESY
;2D homonuclear correlation via dipolar
;coupling, dipolar coupling may be due
;to noe or chemical exchange.
;phase sensitive
;water suppression using 1-1 echo
;pulse sequence with radiation damping
;suppression using gradients in t1
;
;V. Sklenar, J. Magn. Reson. A114, 132-135 (1995)
;
prosol relations=<triple>

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"in0=inf1/2"

"p2=p1*2"
"d12=20u"

#   ifdef LABEL_CN
"p22=p21*2"
#   else
#   endif /*LABEL_CN*/

#   ifdef LABEL_CN
"d0=6u"
"DELTA1=d0*2+larger(p14,p22)+4u"
#   else
"d0=in0/2-p1*2/3.1416-1u"
#   endif /*LABEL_CN*/

"DELTA=d8-p16-d16"

1 ze
#   ifdef LABEL_CN
2 d1 do:f2 do:f3
  d12 pl2:f2 pl3:f3
#   else
2 d1
  d12
#   endif /*LABEL_CN*/
3 d12
  50u UNBLKGRAD
  (p1 ph1)
#   ifdef LABEL_CN
  d0 gron0
  2u groff
  (center (p14:sp3 ph2):f2 (p22 ph2):f3 )
  d0 gron0*-1
  2u groff
  (p2 ph7)
  DELTA1
#   else
  d0 gron0
  d0 gron0*-1
  2u groff
#   endif /*LABEL_CN*/
  (p1 ph2)
  DELTA
  p16:gp1
  d16
  (p1 ph3)
  d19
  (p1 ph4)
  4u
  p16:gp2
  d16
  (p1 ph5)
  d19*2
  (p1 ph6)
  p16:gp2
#   ifdef LABEL_CN
  d16 pl12:f2 pl16:f3

  4u BLKGRAD
  go=2 ph31 cpd2:f2 cpd3:f3
  d1 do:f2 do:f3 mc #0 to 2 F1PH(ip1, id0)
#   else
  d16
  4u BLKGRAD

  go=2 ph31
  d1 mc #0 to 2 F1PH(ip1, id0)
#   endif /*LABEL_CN*/
exit

ph1=0 2
ph2=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph3=2 2 0 0 1 1 3 3
ph4=0 0 2 2 3 3 1 1
ph5=0
ph6=2
ph7=1 1 1 1 1 1 1 1 3 3 3 3 3 3 3 3
ph31=0 2 2 0 1 3 3 1 2 0 0 2 3 1 1 3

;default power levels:
;pl1 : f1 channel
;pl2 : f2 channel
;pl3 : f3 channel
;shaped pulse 180 degree (adiabatic dec):
;sp13: f2 channel
;spnam13: crp45,1,20.1

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p21 : f3 channel -  90 degree
;p22 : f3 channel - 180 degree
;shaped pulse for inversion (adiabatic):
;p14 : f2 channel - 180 degree
;     = 1ms for crp30,1,20.1

;p16: homospoil/gradient pulse
;d0 : incremented delay
; (2D, min >= 6usec) =
;    in0/2-p1*2/3.1416-1u or 6u
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d12: delay for power switching     [20 usec]
;d16: delay for homospoil/gradient recovery
;d19: delay for binomial water suppression
;     d19 = (1/(2*d)),
;       d = distance of next null (in Hz)
;inf1: 1/SW = 2 * DW
;in0: 1/(2 * SW) = DW
;nd0: 2
;NS: 8 * n
;DS: 16
;FnMODE: States-TPPI, TPPI, States or QSEQ

;for z-only gradients:
;gpz0: 2%
;gpz1: 50%
;gpz2: 30%

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100

   ;preprocessor-flags-start
;LABEL_CN: for C-13 and N-15 labeled
;samples start experiment with
   ;preprocessor-flags-end

;$Id: noesygpphjrrs,v 1.3
;2007/04/11 13:34:31 ber Exp $
```

HSQC for sugar protons

```
;hsqc_sug_full.txt
; interleaved version to measure decoupled
; and coupled version simultaneously
; used for RDC measurements
;inviprst
;avance-version
;2D H-1/X correlation via double inept transfer
;phase sensitive using States-TPPI method
;with decoupling during acquisition
;G. Bodenhausen & D.J. Ruben,
;Chem. Phys. Lett. 69, 185 (1980)

#include <Avance.incl>
#include <Grad.incl>

"p2=p1*2"
"p4=p3*2"
"p8=p3*59.4/90"
"p9=p3*298/90"

"d0=3u"
"d4=1s/(cnst2*4)"
"d10=12.5m"
"d11=30m"
"d12=20u"
"d13=3u"
"d21=12.5m-3u"
"d23=12.5m-p2-3u"

"l3=(td1/4)"
"l0=0"

1 ze
  d11 pl12:f2
2 d11 do:f2 do:f3
  3m
3 3m
4 d11*2
5 d12 pl9:f1
  d12 pl2:f2
  d1 ;cw:f1 ph29
  d13 do:f1
  d12 pl1:f1
  p1 ph1
  d4 UNBLKGRAD
  (center (p2 ph2):f1
    (p8 ph20 p9 ph22 p8 ph20):f2)
  d4
  (p1 ph3)
  50u
  p16:gp1
  d16
  (p3 ph6):f2

if "l0 %2 == 1" goto 100

;coupled part

   d0
   d21
   (p4 ph16):f2
   d10
   goto 101

;decoupled part

100 d0
    (p2 ph8)
    d23
    (p4 ph16):f2
    d10

101 (p3 ph7):f2
    50u
    p16:gp2
    d16
    p1 ph4
    d4 BLKGRAD
    (center (p2 ph17):f1
      (p8 ph20 p9 ph22 p8 ph20):f2)
    d4 pl12:f2
    go=2 ph31 cpd2:f2
    d11 do:f2 do:f3 wr #0 if #0 zd
    3m iu0
    lo to 3 times 2
    3m ip6
    lo to 4 times 2
    d11 id0
    d11 dd10
    lo to 5 times l3
   exit


ph1=0
ph2=0 0 2 2
ph3=1
ph4=1
ph5=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph6=0 2
ph7=0 0 0 0 2 2 2 2
ph17= 0 0 0 0 2 2 2 2
ph8=1 1 3 3
ph9=0 0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph20=0
ph22=2
ph29=0
ph16=0
ph31=0 2 0 2 2 0 2 0

;default power levels:
;pl1 : f1 channel
;pl2 : f2 channel
;CPD/BB decoupling power levels:
;pl12: f2 channel
;power level for presaturation:
;pl9 : f1 channel

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p3 : f2 channel -  90 degree
;p4 : f2 channel - 180 degree

;d0 : incremented delay (2D)       [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J)XH
;d11: delay for disk I/O           [30 msec]
;d12: delay for power switching    [20 usec]
;d13: short delay                  [3 usec]
;cnst2: = J(XH)
;l3: loop for phase sensitive 2D using
;States-TPPI method: l3 = td1/2
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 4 * n
;DS: 16
;td1: number of experiments
;MC2: States-TPPI

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel
```

E-COSY

```
; hcch_ecosy.txt
; 3D HCCH-E.COSY  for RNA
; Schwalbe ... Griesinger (1994),
;  JBN 4, 631-44
; Zimmer ... Griesinger (1996),
;  Magn. Res. Chem. 34,S177-S186

#include <Grad.incl>
#include <Delay.incl>
#include <Avance.incl>

;############ IMPORTANT ############:
;; => set PHCOR1 = PHCOR11 = PHCOR21 = 44 degree
;; => constant time: l10 <= d30/in30

"in0=inf1/2"
"in10=inf2/2"

;for H4'-H5':
"TAU=6.15m" ; 3/4Jcc  or 1/2Jcc
"TAU1=12.2m" ; 1/4Jcc

;for H1'-H2':
; "TAU=12m"
; "TAU=6.1m"

"d0=3u"
"d20=3u"
"d2=3.0m"   ; DELTA 1/2J(H,C) sugars in RNA
"d4=d2/2"   ; DELTA/2
"d12=d2+p3"   ; DELTA/2
"d5=d4+6u+p4"   ; DELTA/2

"d8=TAU/4-d4-p2"  ; id8
"d9=TAU/4+d4-3u"  ; dd9
"d10=3u"   ; id10
"d30=TAU/4"   ; tau=3/4Jcc (for t1-evolution)

"d26=TAU1/2"   ; 2tau1=1/4Jcc
"d25=d26-d2-p2*2"
; this was wrong before: "d26=d25-d4-20u"

;"p10=1000u"

"l0=td1/2"
"l10=td2/2"

"in5=d5/l0"
"in20=in0-in5"
"in8=in10/2"
"in9=in10/2"
"in30=in10/2"

aqseq 321

1 ze
2 d1 do:f2
  5m do:f3
3 20m
4 20m
5 15m
6 20u ;pl13:f3
  20u pl7:f1
;3u
  20u UNBLKGRAD
;18 (p18:sp0 ph0):f1
;  lo to 18 times 100
;  3u
;  (p29 ph0):f1
  20u pl1:f1
  20u pl2:f2
  (p1 ph1:r):f1
  d0 ; in0  IN0=1/2SWH
  d4
  (p4 ph0):f2
  d20 ; in20 IN20=IN0-IN5
  (p2 ph21:r):f1
  d5 ; dd5 IN5=d5/l0
 ; (p10 ph21:r):f1 ; trim pulse
 ; 3u
  (p1 ph11:r):f1
  4u pl8:f1
 ;  (p12 ph12:r):f1
  4u
   p16:gp1 ; spoil gradient
    d16 pl1:f1
  (p3 ph2):f2 ;-----------
  d8 ; tau/4-delta/2+t2/4
  (p2 ph21:r):f1
  d9 ; tau/4+delta/2-t2/4
  (p2 ph21:r):f1
  d10 ; t2/2 IN10=1/2SWH
  (p4 ph0):f2
  d30 ; tau/4-t2/4
  (p2 ph21:r):f1
  d30 ; tau/4-t2/4
  (p3 ph3):f2 ;------------ C -> C

  d26
  (p4 ph0):f2
  d25
  (p1 ph21:r):f1 ; beta
  3u ; pulse
  (p1 ph0):f1 ; implementation
  d2
  (p3 ph0):f2
  (p2 ph0):f1
  d12 pl12:f2 BLKGRAD
  go=2 ph31 cpd2:f2
  d1 do:f2 do:f3 wr #0 if #0 zd
  5m ip2
  lo to 3 times 2 ; States-TPPI 13C
  5m id8 ; in8 =1/4SWH
  5m dd9 ; in9  =1/4SWH
  5m id10 ; in10 =1/2SWH
  5m dd30 ; in30 =1/4SWH
  lo to 4 times l10 ; 13C
  5m ip1 rd8
  5m rd9
  5m rd10
  5m rd30
  lo to 5 times 2 ; States-TPPI 1H
  5m id0
  5m id20
  5m dd5
  lo to 6 times l0 ; 1H
exit

ph0=0
ph1=0 2
;ph1=1 3
ph2=0 0 2 2
ph11=1
ph12=3
ph21=0
ph3=0
ph31=0 2 2 0

;l10 <d30/in30
;in5 <d5/l0
;in8 1/4SWH(13C)
;in9 1/4SWH(13C)
;in30 1/4SWH(13C)
;in10 1/2SWH(13C)
;in0 1/2SWH(1H)
;in20 1/2SWH(1H)
```

```
;hnn_cosy_water.txt
;in water
;HNN-COSY - Pulsprogramm zur Bestimmung
;von kanonischen Wasserstoffbrueckenbindungen
;M. Hennig & J.R. Williamson,
;NAR 2000, 28, 1585-1593

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>


"p2=p1*2"
;"p4=p3*2"
"p22=p21*2"
"p8=p21*59.4/90"
"p9=p21*298/90"
"p28=p1"

"in0=inf1/2"

"d11=30m"
"d12=20u"
"d15=d19"
"d16=500u"

;"d30=50u"
;"d0=(in0-p21*4/3.14159-p2)/2"
     ;ph = (+90, -180)
"d0=(in0*2-p21*4/3.14159-p2)/2"
     ;ph = (+180, -360)

"d20=2.5m"
"d21=2.5m"
"d22=20m - d16 - p16"
  ;note:delays for NN tranfers 45ms;
  ;45ms/2=22.5ms
"d23=20m - d30 - p16"
 ;"d28=14m - d16 - p16"
"d28=2.5m - d16 - p16-4u-p28*2.385-d19*5"
 ;"d29=14m - d30 - p16"
 ;"d29=2.5m - d16 - p16-4u-p28*2.385-d19*5"


1  d11 ze
   d11
2  d1 do:f3 do:f2
   d11*4
3  d11
4  d12 do:f3 do:f2

   d12 pl1:f1
   d12 pl3:f3
   d12 pl12:f2

   (p1 ph1):f1

   d20 UNBLKGRAD
   (center (p2 ph0):f1
    (p8 ph20 p9 ph22 p8 ph20):f3)
   d21

   (p1 ph2):f1
   d30
   p16:gp1 ;G1
   d16
   4u cpd2:f2
   (p21 ph3):f3
   d22
   p16:gp2 ;G2
   d16

   (p8 ph11 p9 ph12 p8 ph11):f3

   d30
   p16:gp2 ;G2
   d23
   (p21 ph4):f3

   d0 ;t1/2
   (p2 ph2):f1
   d0

   (p21 ph2):f3
   d22
   p16:gp3 ;G3
   d16
   (p8 ph20 p9 ph22 p8 ph20):f3
   d30
   p16:gp3 ;G3
   d23
   (p21 ph1):f3
   4u do:f2
   d30
   p16:gp4 ;G4
   d16
   (p1 ph6):f1
   4u
   d28
   p16:gp7 ;G5
   d16

;   (center (p2 ph0):f1
;(p8 ph20 p9 ph22 p8 ph20):f3)

;####### reverse inept transfer starts

;binomial watergate pulse substitutes
;the non-selective 180 pulse on H)

   p28*0.231 ph5
   d19*2
   p28*0.692 ph5
   d19*2
   p28*1.462 ph5
   d15
   (p8 ph20 p9 ph22 p8 ph20):f3
   d15
   p28*1.462 ph7
   d19*2
   p28*0.692 ph7
   d19*2
   p28*0.231 ph7
   4u
   d16
   ;d30
   p16:gp7          ;G5
   d28 pl16:f3 pl12:f2
   4u BLKGRAD
   go=2 ph31 cpd3:f3 cpd2:f2
   d1 do:f3 do:f2 mc #0 to 2
     F1PH(ip3 & ip4 & ip11 & ip12, id0)

exit
```

```
ph0=0
ph1=0
ph2=1
ph3=0 2
ph4=1 1 3 3
ph5=0
ph6=2
ph7=2
ph11=0
ph12=2
ph20=0
ph22=2
ph31=0 2


;f1=Protonen 1H
;f2=Carbon 13C
;f3=Stickstoff 15N

;high power pulse:
;p1 : f1 channel -  90 degree
;p2 : f1 channel - 180 degree
;p21: f3 channel -  90 degree
;p22: f3 channel - 180 degree

;decoupling:
;cpd2: according to sequence defined by cpdprg2
;cpd3: according to sequence defined by cpdprg3
;90 degree pulse for decoupling sequence:
;pcpd2: f2 channel
;pcpd3: f3 channel

;gradients: all SINE.100 with p16=1000ms
; gpz1=10%
; gpz2=7%
; gpz3=-17%
; gpz4=15%
; gpz7=30%

;purine: J(H8-N7/N9)=10Hz
;        J(H2-N1/N3)=15Hz
;pyrimidines: J(H5-N1)=4.5Hz
;             J(H6-N1)=3.5Hz
;             J(H5-N3)=2.5Hz

;d1 : relaxation delay; 1-5 * T1
;d16: delay for homospoil/gradient recovery
;d19: delay for binomail water suppression
; d19 = (1/2*d),
;   d = distance of next null (in Hz)
```

Tab. A.12: General NMR parameters I

	H1'C1'-HSQC	H6C6/H8C8-HSQC	H_sCN_b	H_bCN_b
pulseprogram	hsqc_sug_c1h1.txt	hsqc_base.txt	hcnsug.txt	hcnbase.txt
data points	2k (t2) x 512 (t1)	2k (t2) x 512 (t1)	2k (t3) x 72 (t2)x 128 (t1)	2k (t3) x 64 (t2) x 64 (t1)
sweep width ω_1, ω_2, ω_3	^{13}C: 23.5 ppm, ^{1}H: 10 ppm -	^{13}C: 30 ppm, ^{1}H: 10 ppm -	^{15}N: 34 ppm, ^{13}C: 10 ppm, ^{1}H: 6 ppm	^{15}N: 34 ppm, ^{13}C: 16 ppm ^{1}H: 12 ppm
carrier frequency o1p, o2p, o3p	^{1}H: on water, ^{13}C: 97 ppm, ^{15}N: 165 ppm	^{1}H: on water, ^{13}C: 146 ppm, ^{15}N: 165 ppm	^{1}H: 5.7 ppm, ^{13}C: 89 ppm, ^{15}N: 157 ppm	^{1}H: on water, ^{13}C: 136 ppm, ^{15}N: 157 ppm
ns & ds	16 to 32 & 32	16 to 32 & 32	16 to 24 & 128	36 - 44 & 128
repetition delay	1.4 s	1.4 s	1.7 s	1.7 s
delay	INEPT: 1.45 ms	INEPT: 1.4 ms	for $J_{C1'H1'}$: Δ = 1.5 ms for J_{CN}: δ = 22.6 ms	INEPT: 1.25 ms
band selective pulses	180° Gaussian Q3 pulse cascade: on ^{13}C for C1'/C5 600 MHz: 800 μs	180° Gaussian Q3 pulse cascade: on ^{13}C for C6/C8 600 MHz: 500 μs	REBURP on ^{1}H, 4 ms (600 MHz) on ^{13}C, 3 ms (600 MHz)	REBURP on ^{13}C, 2.5 ms (600 MHz)
			IBURP-2 on ^{15}N, 1.5 ms (600 MHz)	IBURP-2 on ^{13}C, 2.1 ms (600 MHz) only if pyrimidines labeled
off resonance selective pulses	180° Gaussian Q3 pulse cascade on ^{13}C at 153 ppm for C6/C8 600 MHz: 450 μs	180° Gaussian Q3 pulse cascade on ^{13}C at 97 ppm for C1'/C5 600 MHz: 800 μs	smoothed chirp on ^{13}C at 113 ppm during ^{15}N evolution 600 MHz: 1.5 ms	
	180° Gaussian Q3 pulse cascade on ^{13}C at 69 ppm for C2' to C5' 600 MHz: 1ms			
decoupling	^{13}C: GARP with 6,9 kHz	^{13}C: GARP with 6,9 kHz	^{13}C: GARP with 4,1 kHz	^{13}C: GARP with 4,1 kHz
z-gradients [G/cm]	SINE.100 with 0.7 ms: z1 = 20, z2 = 15, z3 = 12.5, z4 = 25	SINE.100 with 0.7 ms: z1 = 25, z2 = 30, z3 = 7.5, z4 = 20	SINE.100 with 0.7 ms: z1 = 7, z2 = 35, z3 = 22.5, z4 = 7.5, z5 = 10	SINE.50 with 0.5 ms: z1 = 10, z2 = 40, z3 = 12.5, z4 = 7.5, z5 = 17.5 SINE.100 with 1 ms: z6 = -30

Tab. A.13: General NMR parameters II

	HCCH-COSY-TOCSY	HCCH-COSY	H2H8-COSY	^{13}C-edited - ^{13}C/ ^{12}C-filterted NOESY
pulseprogram	hcchcosytocsy.txt	hcchcosy.txt	bsADE.3d	13Ceditednoesy_12C.txt
data points	2k (t3) x 64 (t2) x 128 (t1)	2k (t2) x 70 (t1)	2k (t3) x 60 (t2) x 120 (t1)	2k (t3) x 118 (t2) x 170 (t1)
sweep width ω_1, ω_2, ω_3	^{1}H: 4 ppm, ^{13}C: 28.6 ppm, ^{1}H: 10 ppm	^{13}C: 52.6 ppm, ^{1}H: 9 ppm -	^{1}H: 10 ppm, ^{13}C: 71 ppm, ^{13}C: 71 ppm	^{13}C: 54 ppm ^{1}H: 7.7 ppm, ^{1}H: 10 ppm
carrier frequency o1p, o2p, o3p	^{1}H: on water, ^{13}C: 79 ppm, ^{15}N: 159 ppm	^{1}H: on water, ^{13}C: 119 ppm,	^{1}H: on water, ^{13}C: 140 ppm, ^{15}N: 150 ppm	^{1}H: on water, ^{13}C: 114 ppm, ^{15}N: 165 ppm
ns & ds	16 to 32 & 32	400 & 128	20 & 64	8 to 16 & 128
repetition delay	1.5 s	1.2 s	1.5 s	1.2 s
delay	INEPT: 1.5 ms for J_{CC} in COSY: Δ = 3.2 ms for $J_{CH/CH2}$ after COSY: δ = 1 ms	INEPT: 1.4 ms for J_{CC} in COSY: Δ = 1.866 ms	INEPT: 1.25 ms for J_{CC} in COSY: Δ = 28.4 ms	INEPT: 1.4 ms mixing time: 150 ms and 300 ms
selective pulses	DIPSI-2 mixing for 5.44 to 16.22 ms on ^{13}C for magnetisation transfer between all sugar carbons		EBURP2 on ^{13}C, 1.2ms (800 MHz) offset: -8 kHz	chirp on ^{13}C during INEPT with 9 kHz (800 MHz) ^{13}C: Wurst pulses for filter Γ = 1859 ns and Ξ = 2358 sweep width: 80 kHz offset: -12 kHz
decoupling	^{13}C: GARP with 4.8 kHz	^{13}C: GARP with 6.9 kHz	^{13}C: GARP with 9.0 kHz	^{15}N: chirp with 400 Hz (800 MHz) ^{13}C: chirp with 2 kHz (800 MHz)
z-gradients [G/cm]	SINE.50 with 0.5 ms: z2 = 25, z3 = 5, z6 = 30 SINE.50 with 0.3 ms: z5 = 12.5, z7 = 9 SINE.100 with 2 ms & 1.25ms: z4 = 15	SINE.50 with 0.5 ms: z1 = 10, z2 = 20, z7 = 12.5, z12 = 7.5	SINE.100 with 1 ms: z1 = 35, z2 = 22.5, z3 = 11.3, z4 = 6.5, z5 = -3.5, z6 = 4.5	SINE.50 with 0.3 ms: z2 = 5, z5 = 5, SINE.100 with 1 ms: z1 = 7.5, z3 = 15, z4 = 20, z6 = 17.5, z7 = 25

Tab. A.14: General NMR parameters III

	^{15}N-HSQC for Iminos:	^{15}N-HSQC for Aminos:	imino-NOESY	HSQC for sugar protons
pulseprogram	imino_hsqc.txt	amino_hsqc.txt	imino_noesy_jr.txt	
data points	2k (t2) x 512 (t1)	2k (t2) x 192 (t1)	4k (t2) x 640 (t1)	2k (t2) x 912 (t1)
sweep width ω_1, ω_2	^{15}N: 20 ppm, ^{1}H: 22 ppm	^{15}N: 34 ppm, ^{1}H: 13 ppm	^{1}H: 22 ppm, ^{1}H: 22 ppm,	^{13}C: 40 ppm ^{1}H: 11 ppm,
carrier frequency o1p, o2p, o3p	^{1}H: on water, ^{13}C: 135 ppm, ^{15}N: 155 ppm	^{1}H: on water, ^{13}C: 130 ppm, ^{15}N: 85 ppm	^{1}H: on water, ^{13}C: 114 ppm, ^{15}N: 153 ppm	^{1}H: on water, ^{13}C: 75 ppm, ^{15}N: 159 ppm
ns & ds	32 & 32	24 & 32	64 & 32	32 & 48
repetition delay	1.4 s	1.5 s	1.6 s	1.5 s
delay	INEPT: 2.77 ms binomial water suppression: d19 = 54 μs (600 MHz)		INEPT: ms binomial water suppression: d19 = 50 μs (600 MHz)	INEPT: 1.655 ms
selective pulses	chirp on ^{13}C during ^{15}N evolution with 4.5 kHz (600 MHz)		chirp on ^{13}C during ^{15}N evolution with 4,9 kHz (600 MHz)	
decoupling	^{15}N: GARP with 4,8 kHz	^{15}N: GARP with 5 kHz	^{15}N: GARP with 4 kHz ^{13}C: chirp with 6 kHz (600 MHz)	^{13}C: GARP with 4.1kHz
z-gradients [G/cm]	SINE.100 with 1 ms: z1 = 35, z2 = -20	SINE.100 with 1 ms: z1 = 30, z2 = -20, z3 = 7.5, z4 = 20	SINE.100 with 1 ms: z1 = 35, z2 = 15, z3 = 25	SINE.100 with 0.9 ms: z1 = 34, z2 = 26

Tab. A.15: General NMR parameters IV

	E-COSY	HNN-COSY for WC bp
pulseprogram	hcch_ecosy.txt	hnn_cosy_water.txt
data points	H1'H2': 2k (t3) x 92 (t2) x 64 (t1) (H4'H5': 2k (t3) x 72 (t2) x 96 (t1))	1k (t2) x 116 (t1)
sweep width $\omega_1, \omega_2, \omega_3$	^{1}H: 3 ppm, ^{13}C: 30 ppm (26.5 ppm) ^{1}H: 8 ppm	^{15}N: 100 ppm, ^{2}2H: ppm
carrier frequency o1p, o2p, o3p	^{1}H: on water, ^{13}C: 82 ppm (72 ppm) ^{15}N: 165 ppm	^{1}H: on water, ^{13}C: 140 ppm ^{15}N: 189 ppm
ns & ds	32 & 128	512 & 128
repetition delay	1.4 s	1.2 s
delay	INEPT: 1.5 ms for J_{CC} (in H1H2): TAU = 12 ms, TAU1 = 6.1 ms for J_{CC} (in H4H5): TAU = 6.15 ms, TAU1 = 12.2 ms	INEPT: 2.5 ms NN-transfer: 22,5 ms binomial water suppression: d19 = 60 μs (600 MHz)
decoupling	^{13}C: GARP with 12 kHz	^{13}C: GARP with 6,7 kHz ^{15}N: GARP with 5 kHz
z-gradients [G/cm]	SINE.100 with 1 ms: z1 = 40	SINE.100 with 1 ms: z1 = 5, z2 = 3,5 z3 = -8,5, z4 = 15

Publication and conference list

Publications

Falb, M., Carlomagno, T., NMR assignment of the U4 5' stem loop RNA, in preparation

Jehle, S., **Falb, M.**, Kirkpatrick, J., Oschkinat, H., van Rossum, B.-J., Althoff, G. and Carlomagno, T. (2009), Intermolecular protein-RNA interactions revealed by 2D ^{15}N–^{31}P magic angle spinning solid-state NMR spectroscopy. *PNAS*, submitted

Falb, M., Amata, I., Gabel, F., Simon, B. and Carlomagno, T. (2009), Structure of the K-turn U4 RNA in the absence of protein binders: a combined NMR and SANS study. *RNA*, under review

Previous publications

Rosowski, M., **Falb, M.**, Tschirschmann, M. and Lauster, R. (2006), Initiation of Mesenchymal Condensation in Alginate Hollow Spheres - A Useful Model for Understanding Cartilage Repair? *Artificial Organs* **30**(10), 775-784

Conferences

1st International Symposium on Structural Systems Biology (CSSB), 2009, Hamburg, talk

13th Annual Meeting of the RNA Society, 2008, Berlin, poster presentation

GDCH Magnetic Resonance Division, 29th Annual Discussion Meeting: Magnetic Resonance in Biophysical Chemistry, 2007, Göttingen, poster prize

12th Annual Meeting of the RNA Society, 2007, Madison, poster presentation

Courses

International School of Biological Magnetic Resonance (ISBMR): 8th Course, Biophysics and the Challenges of Emerging Threats, 2007, Erice, poster presentation

Acknowledgements

The work presented in this PhD thesis has been carried out in the NMR Group of the European Molecular Biology Laboratory (EMBL) in Heidelberg and in the NMR department at the Max-Planck-Institut für biophysikalische Chemie (MPIbpc) in Göttingen.

I would like to thank my supervisor PD Dr. Teresa Carlomagno for accompanying this PhD work with her scientific interest and guidance, for her generous support as well as for helpful discussions and encouragement.

I would like to especially thank my thesis advisory committee, which consisted of Dr. Anne-Claude Gavin, Dr. Andreas Ladurner from the EMBL and Prof. Irmgard Sinning from the University of Heidelberg. I am very grateful for their ideas and recommendations to my work.

I would like to thank Prof. Thomas Rausch from the University of Heidelberg for participating in my defense committee.

During the last four years, I was fortunate to interact and work with many people at EMBL and MPIbpc, to all of which I am indebted for their help, advice and friendship. I would like to thank particularly Irene Amata for working with me on the U4 5' stem loop and lariat-forming ribozyme projects and Dr. Jörg Fohrer for working with me on the lariat-forming ribozyme project as well as for his excellent introduction to the field of NMR spectroscopy and both for their encouragement and friendship. I am also grateful to Claudia Schwiegk for her introduction to RNA synthesis methods and for her helpful advice in the first part of my PhD.

Moreover I would like to thank all past and present members of the Carlomagno group. Due to their presence, I have greatly enjoyed working in Heidelberg and Göttingen. I will not mention any names here to minimize the danger to accidently forget someone of my colleagues.

I am especially thankful to Dr. Teresa Carlomagno, Dr. Michaela Falb, Dr. Jörg Fohrer and Irene Amata for proofreading my thesis or parts thereof.

Dankbar bin ich auch meinen Freunden für ihr Interesse und ihren seelischen Beistand. Ganz besonders aber möchte ich meiner Familie und Markus danken, die mich in den vergangenen Jahren jederzeit und auf vielfältige Weise unterstützt und mir stets Mut gemacht haben, dass ich diese Doktorarbeit erfolgreich beenden werde.